Unfolding Studies of the Protease Domain of Urokinase-Type Plasminogen Activator: The Existence of Partly Folded States and Stable Subdomains†

Ursula K. Nowak,‡ Alan Cooper,§ Derek Saunders,‖ Richard A. G. Smith,⊥ and Christopher M. Dobson*,‡

Oxford Centre for Molecular Sciences and New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, U.K., Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, U.K., Grünenthal GmbH, D-52220 Stolberg, Germany, and SmithKline Beecham Pharmaceuticals, Coldharbour Road, The Pinnacles, Harlow, Essex CM19 5AD, U.K.

Received October 18, 1993; Revised Manuscript Received December 28, 1993®

ABSTRACT: The domain structure and the stability against thermal and chemical denaturation of urokinase-type plasminogen activator (u-PA) have been investigated by NMR spectroscopy and differential scanning calorimetry (DSC). At least five structurally autonomous regions of this three-domain protein have been found to exist. Two of these are the EGF-like and the kringle domains; the others are all within the third domain, which is a serine protease. The latter undergoes three unfolding transitions in its enzymatically active form. Reaction with a specific affinity label (L-Glu-L-Gly-L-Arg-chloromethyl ketone) to produce an inactivated protein results in a stabilization of the structure involved in two of these transitions, and an increase in cooperativity to give a domain which unfolds in two, not three, distinct steps. These are attributed to the denaturation of the two major subdomains of the protease structure. One of the subdomains has exceptional stability, being unfolded only under extreme conditions such as 75 °C at pH 2.5 or 4 M GuDCl at pH 4.5 and 29 °C. This region has been identified by isolation and characterization of a fragment (residues Ile-159 to Thr-277) obtained by limited proteolysis with thermolysin under conditions where the protease domain was partly unfolded. The NMR data are consistent with this stable region being at the N-terminus of the protein and indicate that its structure and stability are similar to those of the corresponding region of the native protein. These results support the idea that the u-PA protease domain has structural resemblance to the digestive serine proteases, but that stabilizing interactions within the structure can differ significantly between a group of homologous proteins.

Urokinase-type plasminogen activator (u-PA, urokinase)[1] is a protein consisting of three domains: an N-terminal growth factor-like domain (EGF-like domain, residues 1–45, 5 kDa), a kringle domain (residues 46–135, 10 kDa), and a C-terminal serine protease domain (residues 136–411, 30 kDa). The protease domain contains the "catalytic triad" (His-204, Asp-255, and Ser-356) found in all serine proteases (Steffens et al., 1982), but requires cleavage at the proteolytically sensitive Lys-158 to Ile-159 bond before becoming fully active. It is connected to the kringle domain by a 16-residue linker peptide but can be isolated following cleavage at Lys-135 on limited proteolysis with plasmin. The resulting two-chain low molecular weight u-PA (LMW u-PA) is fully active and is present along with intact u-PA in body fluids. u-PA itself is synthesized as a single-chain glycoprotein (single-chain u-PA, scu-PA, prourokinase), which is likewise converted to a fully active high molecular weight two-chain form (HMW u-PA) by cleavage between the Lys-158 and Ile-159 bond by plasmin (Holmes et al., 1985); the two chains remain connected by a disulfide bridge. Together with t-PA (tissue-type plasminogen activator), u-PA is one of the physiologically important activators of plasminogen. It not only is an important fibrinolytic agent but also is involved in tissue degradation in normal and pathological processes (Danø et al., 1985).

One of the most striking properties of u-PA and other fibrinolytic proteins is their multidomain structure. Although no complete structural information is available for u-PA, sequence homology leads to the conclusion that the domains of u-PA are similar to proteins of known structure (Günzler et al., 1982). For one domain, namely, the kringle domain, experimental data are available (Li et al., 1992; Bokman et al., 1993; Li et al., 1994) which show it to be closely similar to kringle domains of other proteins, especially to t-PA kringle 2 (Byeon & Llinás, 1991; de Vos et al., 1992). The EGF-like domain is related to human and murine growth factors; the structures of these (Cooke et al., 1987; Montelione et al., 1992) and of the first EGF-like domain of factor IX (Baron et al., 1992) are known. The protease domain is homologous to serine proteases for which several structures have been solved, including those of trypsinogen and chymotrypsinogen (Bode et al., 1976; Kossiakoff et al., 1977; Wang et al., 1985). Serine proteases have a distinct bilobal structure, each lobe or subdomain having a β-barrel-type topology (Richardson, 1981).

A comparison of spectra of intact u-PA and its fragments (protease domain, kringle domain, and EGF–kringle domain pair) suggests that the isolated kringle and protease domains are fully structured and not disrupted by the absence of the neighboring domains. This has made a significant number of assignments in the ^{1}H NMR spectrum of u-PA possible despite the large size of the protein (Oswald et al., 1989; Nowak et

† This is a contribution from the Oxford Centre for Molecular Sciences, which is supported by the Science and Engineering Research Council (SERC) and the Medical Research Council of the United Kingdom. The biological microcalorimetry facilities in Glasgow are supported by the SERC.

* Author to whom correspondence should be addressed.
‡ University of Oxford.
§ University of Glasgow.
‖ Grünenthal GmbH.
⊥ SmithKline Beecham Pharmaceuticals.

® Abstract published in *Advance ACS Abstracts,* February 1, 1994.

[1] Abbreviations: NMR, nuclear magnetic resonance; DQF COSY, double quantum-filtered correlated spectroscopy; ppm, parts per million; u-PA urokinase-type plasminogen activator; HMW u-PA, high molecular weight u-PA; LMW u-PA, low molecular weight u-PA; t-PA, tissue-type plasminogen activator; EGF, epidermal growth factor; 1-D, one dimensional; 2-D, two dimensional.

0006-2960/94/0433-2951$04.50/0

al., 1993). Assignments were obtained by comparison of the spectra of the intact protein with those of the isolated domains. The spectrum of the kringle domain has been virtually completely assigned (Li et al., 1992), which has allowed assignments of resonances in the spectrum of u-PA to specific residues in the kringle domain to be made. Assignments to specific domains, but not specific residues, has been possible for a large number of resonances arising from the protease and the EGF-like domains (Nowak et al., 1993).

An interesting feature of the domain structure of u-PA concerns the extent of segmental mobility experienced by the different domains. This mobility is revealed by the high-resolution nature of the NMR spectra of u-PA, despite its molecular weight of 46 kDa, and particularly by the observation that resonances from the kringle and the EGF-like domains give rise to narrower lines than those from the protease domain (Oswald et al., 1989). This arises because of faster molecular tumbling of smaller domains if linkers do not substantially constrain them. The domain interactions have been studied in detail by simulations of regions of 1-D and 2-D NMR spectra (Nowak et al., 1993) to obtain line widths of individual resonances. These were then used to define the correlation times for the individual domains isolated and in the intact protein.

A different and complementary approach to studying domain structure and interactions involves unfolding studies. By this means, the segmental unfolding of domains in the intact molecule can be investigated and their stability can be compared with the stability of the isolated domains. This was originally reported for u-PA by Bogusky et al. (1989), who studied inactivated human urinary HMW u-PA, the isolated kringle domain, and the noninactivated protease domain by NMR. Novokhatny et al. (1992, 1993) carried out an unfolding study of u-PA and its domains by DSC which, like the NMR studies, showed the unfolding of the domains of u-PA to be independent and largely unaffected by isolation from the intact protein. Other multidomain fibrinolytic proteins have been studied with regard to their thermal unfolding behavior; both calorimetric and NMR studies of t-PA (Radek et al., 1988; Novokhatny et al., 1991; Teuten, 1991) and plasminogen (Novokhatny et al., 1984; Teuten et al., 1991, 1993) have indicated that at least some of the domains in these proteins have independent unfolding behavior.

In this paper, results from 1-D and 2-D NMR studies of recombinant u-PA and its fragments are reported, along with data from DSC studies of their thermal unfolding. Previously obtained assignments for the NMR spectra have made possible a study of the thermal stability of the domains of u-PA, with special emphasis on the protease domain. The unfolding transitions of noninactivated and inactivated LMW u-PA have been analyzed in detail. Of particular interest has been the characterization of a very stable part of the protease domain which has been identified as the N-terminal subdomain by isolation and characterization of a fragment obtained by limited proteolysis with thermolysin.

MATERIALS AND METHODS

Protein Samples. Recombinant nonglycosylated high molecular weight two-chain u-PA (HMW u-PA, residues 1–411) and low molecular weight u-PA (LMW u-PA, protease domain, residues 135–411), both containing cleavages at the activation site (Lys-158 to Ile-159), were kindly provided by Grünenthal GmbH, Aachen, Germany. These were derived from single-chain nonglycosylated human u-PA expressed in

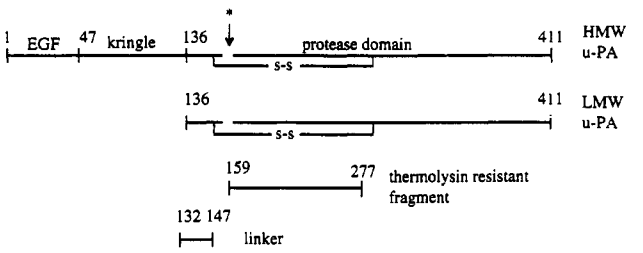

FIGURE 1: Schematic view of the sequences of u-PA indicating its domain structure and fragments studied in this work. Only the disulfide bridge connecting the two chains is shown; an asterisk (*) indicates the activation site.

Escherichia coli, the two-chain protein and fragments being produced by proteolytic cleavage. Recombinant two-chain HMW u-PA and LMW u-PA (protease domain) were prepared in a 0.1 M Tris, 0.15 M NaCl, 20% v/v glycerol containing buffer at pH 7.4 at a concentration of 10 mg/mL and inactivated with a 3-fold molar excess of EGRcmk (L-Glu-L-Gly-L-Arg-chloromethyl ketone, Cambridge Bioscience, Cambridge, U.K.) (20 mM in ethanol) at 25 °C until the activity against S-2444 (pyro-Glu-Gly-Arg-*p*-nitroanilide) was reduced to less than 0.5% of that of the untreated protein. The samples were buffer exchanged into 0.1 M NH_4HCO_3 at pH 8.5 on Sephadex G-25M (PD-10) gel filtration columns at 4 °C and lyophilized. The N-terminal fragment of the protease domain was isolated following limited proteolysis by thermolysin. Two millimolar LMW u-PA was incubated with 3 μM thermolysin at pH 4.0 and 70 °C for 15 min, prior to addition of EDTA to terminate the digestion. The resulting material was purified by FPLC on a MonoS column and an at least 90% pure fragment corresponding to residues Ile-159 to Thr-277 could be isolated. A schematic view of the sequences is given in Figure 1.

The protein samples for the NMR experiments were dissolved in D_2O (99.9% D) at pH 4.5 and purified from small molecules and salts by five cycles of centrifugation at 4 °C using Centricon-3 and Centricon-10 concentration cells with a 3-kDa and 10-kDa cutoff membrane, respectively (Amicon, Gloucester, England). The final NMR samples were 0.4 to 1.2 mM in D_2O at pH 4.5. The pH was adjusted by adding dilute DCl and NaOD (4 and 0.4%). NMR samples for the thermal unfolding experiments were buffer exchanged into 100 mM d_4-acetic acid at pH 4.5, 3.5, and 2.5.

NMR Spectroscopy. NMR spectra were recorded on a Bruker AM600 spectrometer with a proton resonant frequency of 600.13 MHz. Experiments were carried out over a range of temperatures between 29 and 74 °C. 1-D spectra were recorded using 4K data points over spectral widths of 7812.5 and 8928.5 Hz; 200–500 scans were required for good signal-to-noise ratios. Resolution enhancement was achieved using the Lorentzian–Gaussian transformation (GB 0.15, LB-10) prior to zero-filling to 8K. DQF COSY spectra (Rance et al., 1983) were acquired over 2K data points and 512–800 t_1 increments in the absorption mode with time proportional phase incrementation (TPPI) for quadrature detection in the t_1 dimension. Water saturation was achieved by low-power irradiation during the relaxation delay introduced between scans, and 80 or 96 transients were collected for each t_1 increment. 2-D spectra were processed on a Sun 4 workstation using the Felix program provided by Dr. D. R. Hare (Felix version 1.1, Hare Research Inc.). The data set was resolution enhanced by double-exponential and trapezoidal multiplication in t_2, and a combination of double-exponential, shifted sinebell-squared and trapezoidal multiplication in t_1, prior to zero-

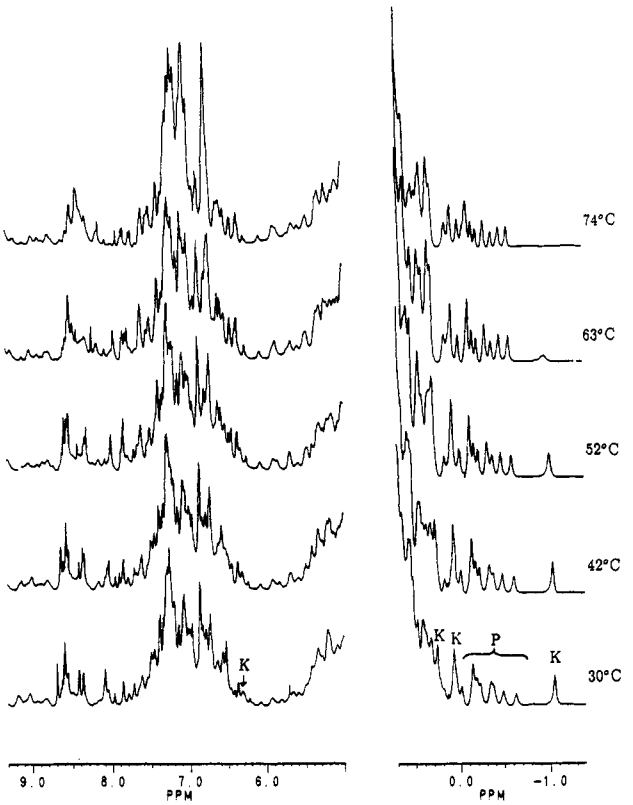

FIGURE 2: 1-D ^{1}H NMR spectra (600 MHz) of the thermal unfolding of inactivated two-chain HMW u-PA at pH 4.5. The unfolding transition is reversible under the conditions used.

filling in both dimensions. After zero-filling, the digital resolution was 3.8 Hz/point in both dimensions. All spectra were referenced to a 1,4-dioxane internal standard at 3.74 ppm.

Differential Scanning Calorimetry. The sample was dissolved in 50 mM sodium citrate buffer, pH 4.5, to give a final concentration of 2.30 mg/mL, determined by UV absorbance using A_{280} (1 mg/mL) = 1.36. The sample was degassed briefly before loading into a Microcal MC-2D calorimeter. The thermogram was recorded between 5 and 105 °C at a scan rate of 60 °C/h. Thermodynamic parameters for protein unfolding were obtained after normalization and base-line correction by standard procedures (Privalov & Potekhin, 1986; Sturtevant, 1987) using the Microcal ORIGIN software package.

RESULTS

Inactivated Recombinant u-PA. The spectra of recombinant two-chain HMW u-PA, inactivated with EGRcmk, at pH 4.5 between 30 and 74 °C shown in Figure 2 are closely similar to those reported earlier for the human urinary protein (Bogusky et al., 1989). However, the spectra are of significantly higher quality, partly as a consequence of the much better homogeneity of the recombinant material compared with human urinary protein, and partly because the pH and ionic strength of the protein solutions have been optimized for the NMR experiments (Nowak et al., 1993). In the upfield region of the 1-D spectrum at 30 °C (Figure 2), there are three resolved methyl group resonances of the kringle domain as assigned by comparison with the spectrum of the isolated kringle domain (Li et al., 1992), namely, Leu-96(47), Val-123(74), and Leu-80(31) (consensus kringle numbering in brackets), marked "K". Methyl group resonances of the protease domain, marked "P", are not yet assigned to specific residues. As the temperature is increased, the decreasing intensity and disappearance of resonances in the NMR spectrum can be interpreted in terms of unfolding of structural regions. Between 30 and 42 °C, few changes can be observed in the spectra of u-PA (Figure 2) other than the narrowing of the resonances anticipated from the shorter correlation times for molecular tumbling of the still folded protein as a consequence of the increased thermal energy. Above 52 °C, however, although resonances from the protease domain narrow still further, those from the kringle domain broaden and disappear.

The broadening is an indication of rates of exchange between the folded and the unfolded states on the order of 10^2–10^3 s^{-1}. For the determination of the unfolding temperature, spectra were recorded at different temperatures, increasing the temperature in 3 or 5 °C steps. Peak areas of resolved resonances were measured, and the transition temperature midpoint was determined as the temperature at which the area is reduced to half the area of this resonance in the fully structured protein. The reversibility of the unfolding process was shown by the regeneration of the original spectrum after reducing the temperature following the unfolding experiment.

The midpoint of the unfolding transition (T_m) of the kringle domain is 63 ± 2 °C; by 69 °C the kringle domain is effectively fully unfolded. The unfolding transition is furthermore associated with an increase in intensity between 0.9 and 1.2 ppm characteristic of methyl group resonances in an unstructured protein. Most of the upfield shifted resonances of the protease domain, which at lower temperature are poorly resolved, are increasingly well resolved at higher temperatures. This can be attributed to the narrowing of the resonances of the folded domain as the correlation time is reduced by increased thermal motion at higher temperatures. It is possible that the correlation time of the protease domain may be reduced at these temperatures relative to that at lower temperature as a consequence of the unfolding of the kringle domain. Although it is clear that there is substantial motional independence between the domains, there is sufficient interaction to increase the line width relative to those in the isolated domains (Nowak et al., 1993).

At temperatures above ca. 70 °C, the intensities of some of the resonances of the protease domain decrease as unfolding of this domain takes place. The midpoint of this transition cannot be seen clearly here because the samples have not been heated to temperatures higher than 76 °C in order to minimize irreversible denaturation of the protein that occurrs from prolonged exposure to high temperatures. The midpoint is, however, estimated to occur at 76 ± 3 °C. It will be shown later that this transition corresponds to the unfolding of only part of the protease domain. The only resonance of the EGF-like domain in this part of the spectrum (at 0.4 ppm) is not well enough resolved to allow the determination of the unfolding temperature.

In the low-field region of the 1-D spectrum, identification of resonances from individual domains is limited by the large degree of resonance overlap. Because of this, and because of small changes in chemical shift positions with increasing temperature, the transition related to the unfolding of the kringle domain is difficult to follow from resonances in this region of the spectrum. Although resonances of most of the aromatic ring protons of the kringle domain have been assigned in the DQF COSY spectrum of HMW u-PA, only one resonance, corresponding to Tyr-101(52), is well resolved in the 1-D spectrum; this has been labeled in Figure 2 and decreases in intensity in accord with the unfolding temperature

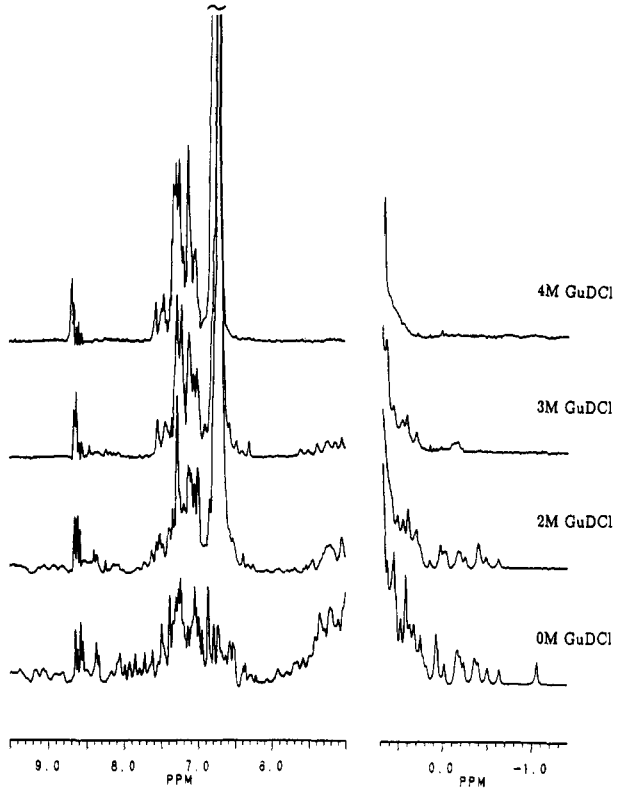

FIGURE 3: ^{1}H NMR spectra (600 MHz) of inactivated recombinant two-chain HMW u-PA at 29 °C, pH 4.5, and 0, 2, 3, and 4 M GuDCl. The resonance at 6.8 ppm is caused by residual protons in GuDCl.

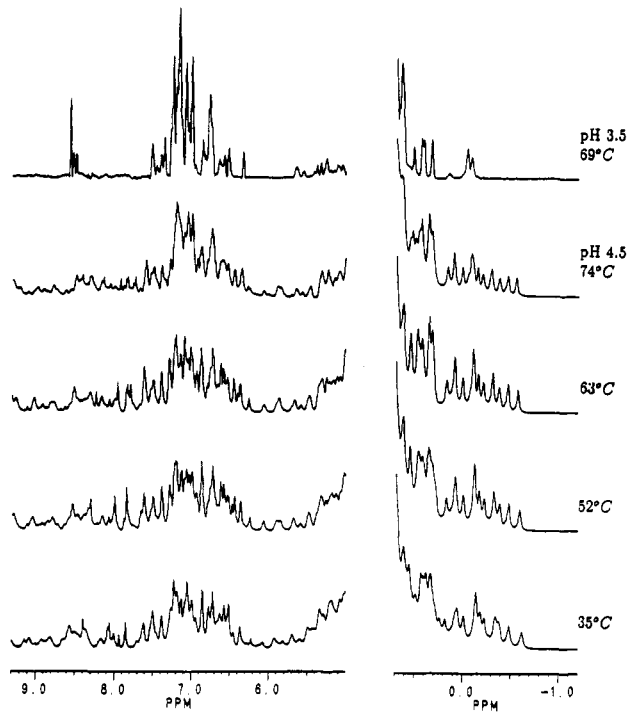

FIGURE 4: ^{1}H NMR spectra (600 MHz) of the unfolding of recombinant inactivated two-chain LMW u-PA at pH 4.5; the top trace shows for comparison the spectrum of the high-stability region of recombinant inactivated LMW u-PA observed at pH 3.5 and 69 °C.

of the kringle domain. The most obvious change in the aromatic region of the spectrum associated with the transition, however, is an increase in intensity in the region from 6.8 to 7.3 ppm; this results from the ring protons of the aromatic residues resonating in the region which is typical for protons of aromatic groups in unstructured proteins. This increase in intensity becomes even more prominent as the protease domain starts to unfold above 70 °C. The unfolding of u-PA is fully reversible under the conditions used, and the spectrum recorded after lowering the temperature to 30 °C is closely similar to that recorded before the unfolding experiment.

The unfolding of HMW u-PA has also been investigated in guanidine deuteriochloride where complete unfolding can be achieved at room temperature. Spectra are shown at GuDCl concentrations of 0, 2, 3, and 4 M GuDCl at 29 °C and pH 4.5 (Figure 3); it is possible to observe clearly three unfolding transitions. Unfolding of the kringle domain is apparent between 0 and 1 M GuDCl. A major unfolding of the protease domain takes place between 2 and 3 M GuDCl. Nevertheless, the spectrum of HMW u-PA in 3 M GuDCl is not one characteristic of a completely unfolded protein. In particular, some resonances of the protease domain remain in positions characteristic of resonances in folded proteins, showing that there exists a stable region of this domain. It will be shown (see below) that a similar result can be obtained by recording spectra at high temperature. The spectrum of the stable region of the protease domain (in 3 M GuDCl or at high temperature) is characterized particularly by a number of downfield shifted $C_\alpha H$ proton resonances between 5.3 and 5.7 ppm indicating the presence of β-sheet structure. In addition, several methyl group resonances appear upfield of the main aliphatic region (between 0.0 and 0.7 ppm) and resonances of a number of aromatic ring protons are resolved in the region of 6.4–6.7 ppm, both indicating the presence of globular structure. Thus, both secondary and tertiary structures persist under these conditions. In 4 M GuDCl the spectrum is typical for a globally denatured protein, with no resonances upfield of 0.7 ppm or between 5.0 and 6.8 ppm.

Inactivated LMW u-PA. In order to perform a more detailed analysis of the unfolding transitions of the protease domain, this domain has been studied following isolation from the intact protein. The changes observed in the spectra of two-chain LMW u-PA, inactivated with EGRcmk, in the temperature range studied (35–74 °C, Figure 4), are closely similar to those of the protease domain as observed in the spectrum of the intact protein (Figure 2) and are reversible. Up to 62 °C, no changes other than the narrowing of resonances occur in the spectra, but the decreasing intensity of resolved resonances at higher temperatures indicates an unfolding transition with a midpoint of 76 ± 3 °C, the same as one of the transitions attributed to the protease domain in the intact protein. Again this transition is not complete at 74 °C under the conditions used. However, by lowering the pH to 3.5, the protein is destabilized and the unfolding transition can be observed at lower temperature. Under these conditions only the more stable part of the protease domain remains folded at 69 °C and the spectrum observed (Figure 4, top trace) is similar to that observed for u-PA in 3 M GuDCl.

Further evidence for the existence of a highly stable part of the protease domain comes from DSC studies. LMW u-PA inactivated with EGRcmk was studied at pH 4.5; the thermogram is shown in Figure 5. Two clear transitions with midpoints at 76.0 ± 0.1 °C (as seen in the NMR study) and 92.3 ± 0.3 °C can be observed (ΔH = 610 ± 40 kJ mol^{-1} and 290 ± 40 kJ mol^{-1}, respectively). Repeat scans were noisy and did not show reversibility, probably due to aggregation of the protein in its denatured state at the very high temperature required for complete unfolding; the sample was turbid after removal from the DSC.

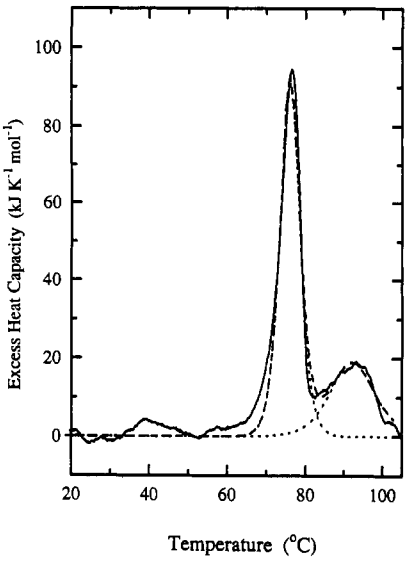

FIGURE 5: DSC thermogram of inactivated recombinant two-chain LMW u-PA in 50 mM sodium citrate buffer, pH 4.5. The solid lines show experimental excess heat capacity data after control base line subtraction and concentration normalization. The dotted lines show theoretical fits obtained by deconvolution of these data in terms of two independent, two-state transitions with T_m = 76.0 and 92.3 °C, and ΔH = 610 and 290 kJ mol^{-1}, respectively. The dashed line shows the overall fit.

Noninactivated LMW u-PA. In contrast to the intact protein, the isolated protease domain can be studied in a noninactivated state without fear of degradation, because the main cleavage sites of autolysis of u-PA are not within this domain but in the linker regions between the EGF, kringle, and protease domains. The spectra of active recombinant two-chain LMW u-PA are closely similar to those of the inactivated protein. This is seen particularly clearly in the 2-D DQF COSY spectra (Figure 6a,b). Because of the size of the protein, full analysis of the entire spectrum is not feasible. Three regions of the spectra are, however, well enough resolved to allow a more detailed analysis, namely, the upfield region containing ring current shifted resonances of methyl groups, the aromatic region, and the region of the spectrum containing downfield shifted $C_\alpha H$ protons. In the upfield region of the 2-D spectrum, for example, several resolved cross peaks can be seen. Six of these, most likely arising from three leucine or valine spin systems (labeled VII, XII, and XIII in Figure 6; in the absence of sequence specific assignments Roman numerals have been used to label resonances) are in positions identical within ±0.02 ppm in the two proteins. Some changes in chemical shifts do, however, occur for other resonances and can be detected in 2-D spectra; these range from very small changes (±0.06 ppm) for the resonances of I, VIII, and IX to larger changes for the resonance labeled X (0.4 ppm). Other cross peaks, including those labeled II, III, IV, V, and XI in Figure 6, cannot be observed clearly in the spectrum of the noninactivated protein because of the lower signal-to-noise ratio. In the aromatic region of the spectrum (see Figure 6), the four tryptophan spin systems have been identified, and another (Trp-V) has been tentatively assigned. Three of these are in nearly identical positions in the spectra of inactivated and noninactivated LMW u-PA (Trp-III, Trp-IV, and Trp-V, within ±0.02 ppm). The other two (Trp-I and Trp-II), the most upfield shifted resonances of the tryptophan residues, are in contrast shifted and are found slightly closer to the random coil values in the spectrum of the noninactivated protein than in the inactivated analogue (the chemical shift differences of the H4 resonances of Trp-I and Trp-II in the two samples are 0.10 and 0.07 ppm, respectively). Most tyrosine cross peaks are closely similar in both spectra; Tyr-I is, however, again slightly less shifted from the random coil position in the noninactivated protein (by 0.10 ppm). Tyr-II is split into two cross peaks most likely as a consequence of small structural changes on inactivation that cause two cross peaks to overlap. The cross peaks of Phe-I and Phe-II are in closely similar chemical shift positions (±0.05 ppm) in the two proteins.

Chemical inactivation of LMW u-PA therefore causes only minor changes in the spectrum at low temperature and, hence, indicates that the overall fold of the protein at least is essentially identical under these conditions. At higher temperatures, however, the spectra are no longer closely similar. This can be attributed to significant differences in the thermal unfolding behavior of noninactivated and inactivated LMW u-PA. Two unfolding transitions can be observed for the noninactivated protein in the temperature range between 30 and 74 °C (Figure 7). The first transition occurs with a midpoint of 53 ±2 °C (see Table 1); this can be seen most directly in the disappearance of the two most upfield shifted resonances at −0.46 and −0.60 ppm and in the aromatic region by the disappearance of resonances between 5.6 and 6.4 ppm. The next transition occurs with a midpoint of 62 ± 2 °C and is particularly evident in the loss of the majority of the upfield shifted methyl resonances (Figure 7). This unfolding behavior is consistent with the two transitions in this temperature range that have been observed in DSC studies (Novokhatny et al., 1992). Even at 74 °C one region of the protein, however, remains folded; the spectrum under these conditions is characterized by several resonances upfield of the main aliphatic region between 0.0 and 0.7 ppm, and between 6.3 and 6.7 ppm, and by downfield shifted $C_\alpha H$ proton resonances between 5.0 and 5.7 ppm. These features are closely similar to those described above for u-PA, partly unfolded either thermally or by intermediate concentrations of GuDCl.

High-Stability Region of the Protease Domain. To gain further insight into the high-stability region of the protease domain, a 2-D NMR study has been performed on inactivated recombinant two-chain LMW u-PA at pH 3.5 and 69 °C, conditions under which the less stable region of the domain is fully unfolded whereas the more stable part remains structured (see top trace of Figure 4). In the upfield region of the DQF COSY spectrum (Figure 6c), resonances of the three residues discussed above (labeled VII, XII, and XIII) are observed at chemical shift positions very close (less than 0.1 ppm difference) to corresponding resonances in the spectrum of the fully folded protein at pH 4.5 and 29 °C (Figure 6b). Other resonances are, however, missing, and no cross peaks can be observed upfield of −0.1 ppm; these are attributed to residues in the less stable region of the protease domain which has unfolded under the conditions used. Part of the protein structure is, however, preserved under these conditions.

In the aromatic region of the spectrum (Figure 6c) assignment of spin systems of the partly folded state by comparison with those of the fully folded protease domain is more difficult because of spectral overlap. This is also a consequence of the large number of cross peaks near to the diagonal in this region of the spectrum of the partly unfolded LMW u-PA. This correlates with an increase in intensity between 6.8 and 7.3 ppm as seen in the 1-D spectra, and results from resonances of residues in the unfolded segment of the protein. Resonances of ring protons of at least one tyrosine and one tryptophan residue can, however, be seen at

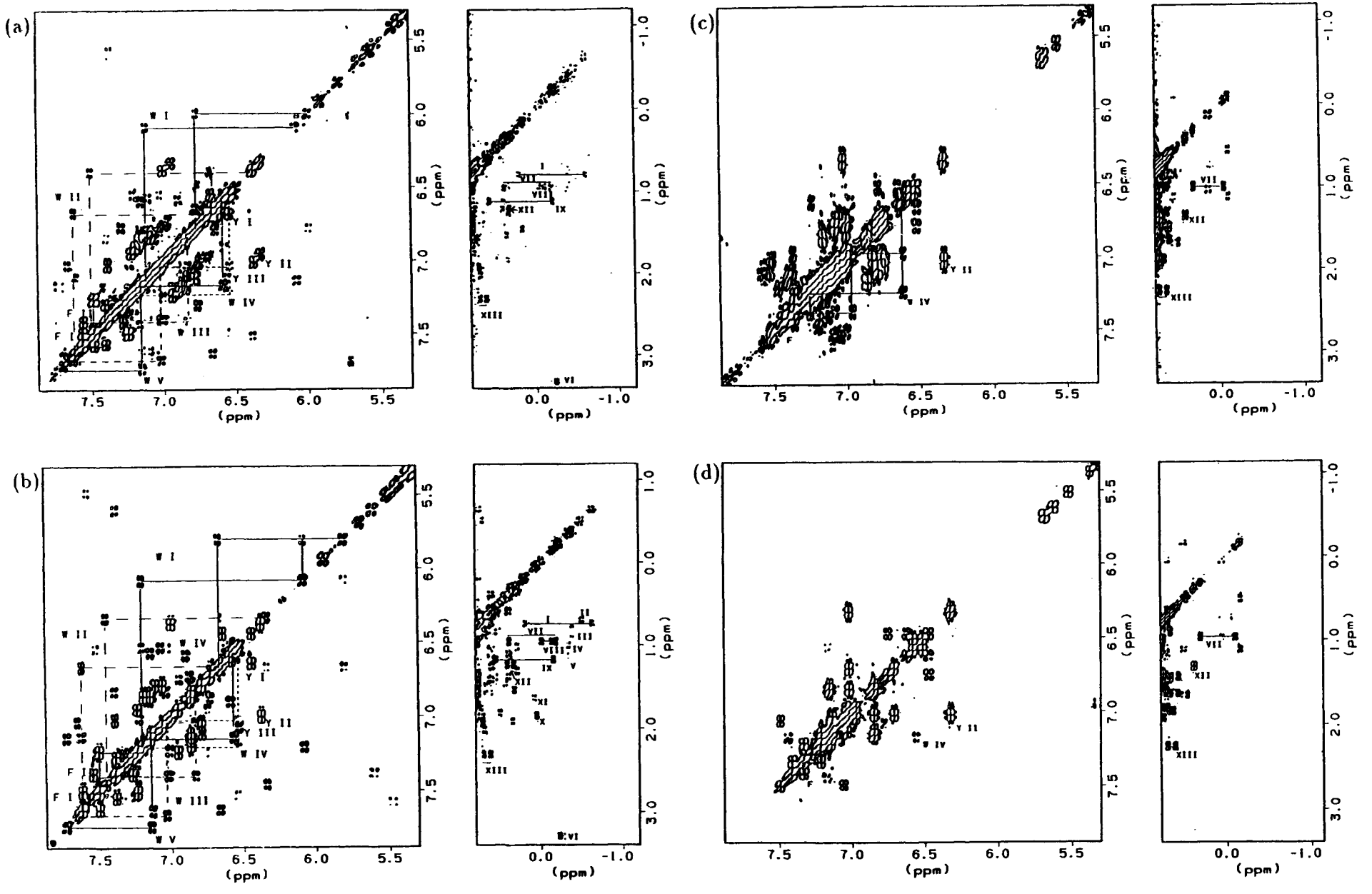

FIGURE 6: Upfield and aromatic regions of the 600-MHz DQF COSY spectrum of noninactivated recombinant two-chain LMW u-PA at 29 °C and pH 4.5 (a), of inactivated two-chain LMW u-PA at 29 °C and pH 4.5 (b) and at 69 °C and pH 3.5 (c), and of the fragment corresponding to the N-terminal subdomain at pH 4.5 and 47 °C (d). In the absence of sequence specific assignments, Roman numerals have been used to label resonances.

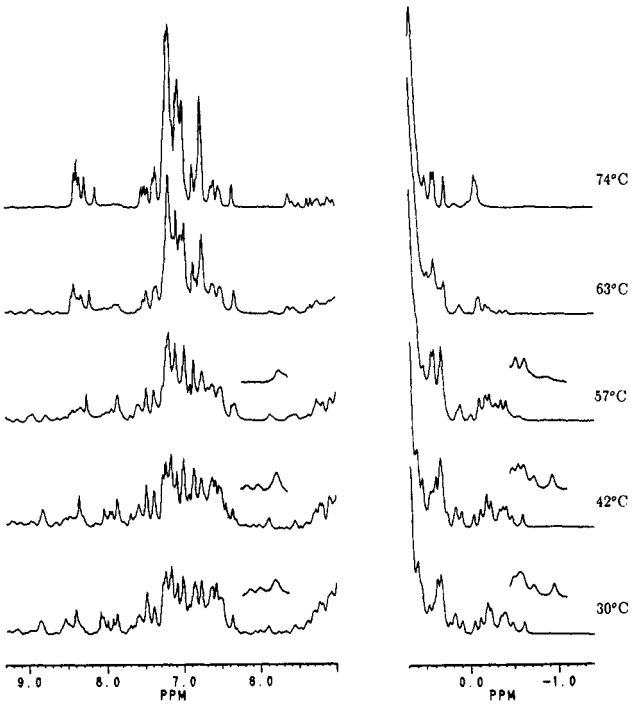

FIGURE 7: ^{1}H NMR spectra (600 MHz) of a temperature study of noninactivated recombinant two-chain LMW u-PA at pH 4.5. As this sample had been dissolved in H_2O and only then been transferred into deuterated solvent, several amide NH resonances are present downfield of 8.5 ppm. Most of the amide NH protons are exchanged for deuterons in samples when spectra are recorded above ca. 60 °C. This is, however, attributable not only to the unfolding but also to accelerated hydrogen exchange rates at higher temperature.

Table 1: Unfolding Midpoint Temperatures (°C) of Two-Chain HMW u-PA and Its Fragments[a]

	K	SP-C	SP-N
HMW u-PA, inactivated, pH 4.5	62	76	>76
LMW u-PA, inactivated, pH 4.5		76	92[b]
LMW u-PA, noninactivated, pH 4.5		53, 62	>74
human urinary kringle, pH 4.5	60[c]		
LMW u-PA, inactivated, pH 2.5		45	71
thermolysin resistant u-PA fragment, pH 2.5			58

[a] "K" is used as the abbreviation for the kringle domain; "SP-C" and "SP-N" are used for the C-terminal and the N-terminal regions of the protease domain. The lower limit set for SP-N indicates the temperature range of the experiment and not the thermal stability of the domain. The transition temperature of the EGF-like domain could not be determined in these experiments. The errors in the transition temperatures are ±2 °C. [b] Data from DSC; all other data are taken from the NMR studies. [c] Bogusky et al., 1989.

chemical shift positions similar to those of resonances of Tyr-II and Trp-IV in the spectrum of the fully folded protein (within 0.04 for Tyr-II and a maximum chemical shift difference of 0.08 ppm for the ring protons of Trp-IV). Resonances of another three tyrosine residues can be observed at chemical shift positions very similar to those in the fully folded protein; these are, however, close to random coil shift positions and would not be expected to experience a large change in chemical shift position even on unfolding.

The third part of the spectrum that provides evidence for the existence of a structured region of the protease domain under these conditions is that downfield of 5 ppm. Resonances in this region arise from $C_\alpha H$ protons and indicate the existence of β-sheet structure. Fourteen $C_\alpha H$–$C_\beta H$ cross peaks with $C_\alpha H$ chemical shift positions downfield of 5 ppm can be detected; this compares to 41 cross peaks in this region in the fully folded protease domain under optimal conditions at 50 °C.

Further analysis of the high-stability region of the protease domain was made possible by its isolation from the intact domain by controlled digestion with thermolysin. Noninactivated LMW u-PA was partly unfolded at pH 4.0 by heating the sample to 70 °C and exposed to thermolysin which retains proteolytic activity under these conditions. A fragment with a molecular weight of 13.5 kDa was isolated, with a sequence corresponding to residues Ile-159 to Thr-277 of HMW u-PA. Figure 6 shows a comparison of the DQF COSY spectrum of thermolysin resistant u-PA at pH 4.5 and 47 °C (d) with that of inactivated LMW u-PA at pH 3.5 and 69 °C (c). In the upfield region, a clear similarity between the two spectra can be seen; resonances of the three residues labeled VII, XII, and XIII can be observed in chemical shift positions similar within ±0.06 ppm in both dimensions of the corresponding resonances in partly unfolded LMW u-PA. In the aromatic region of the spectrum, the cross peaks corresponding to Tyr-I and Tyr-II resonate within ±0.02 ppm in the two spectra. Only part of the spin system of Trp-VI can be observed in the spectrum of the fragment. A striking difference between the spectra of the fragment and that of partly unfolded LMW u-PA is a number of resonances in the latter which occur close to random coil shift positions and are attributed to ring protons of residues in the unfolded part of the protein. Furthermore, several downfield shifted $C_\alpha H$ resonances typical of β-sheet structure are present in the spectrum of the isolated fragment between 5.0 and 5.7 ppm. Ten $C_\alpha H$–$C_\beta H$ cross peaks can be observed in the DQF COSY spectrum downfield of 5 ppm with chemical shift positions very close to those seen in the spectrum of the partly unfolded LMW u-PA (within ±0.05 ppm for nine of the cross peaks, ±0.07 ppm for the other).

Examination of the temperature dependence of the NMR spectrum of thermolysin resistant u-PA shows that the isolated fragment, like the same region in the intact protein, is highly stable. At low temperature, resonances of several slowly exchanging amide protons can be observed between 8.0 and 10.3 ppm indicating the persistence of globular structure. The fragment unfolds with a transition midpoint of 58 ± 2 °C at pH 2.5 (Figure 8), and this process is again reversible. This compares to a transition midpoint of 71 ± 2 °C at pH 2.5 observed for this region of the structure in inactivated LMW u-PA. The unfolding transition can be seen most clearly by the disappearance of the most upfield shifted resonances and of resonances between 5.0 and 6.8 ppm in a manner similar to that found for the intact protein (see above); the spectrum at 74 °C is typical of that for a fully unfolded protein.

DISCUSSION

The unfolding behavior of u-PA and its isolated domains has been analyzed in detail from the changes observed in the NMR spectra on heating or addition of GuDCl, together with the results of the DSC studies. The unfolding temperatures are summarized in Table 1. All the unfolding transitions studied by NMR were reversible. The results for inactivated and noninactivated LMW u-PA, and for the kringle domain, show many similarities, and closely similar transition temperatures, to those from a DSC study published by Novokhatny and co-workers (Novokhatny et al., 1992), although our data show slightly higher transition temperatures under similar conditions (by up to 6 °C). The conclusion of Novokhatny et al. that the inactivated protease domain unfolds cooperatively is not, however, consistent with our data. This is not,

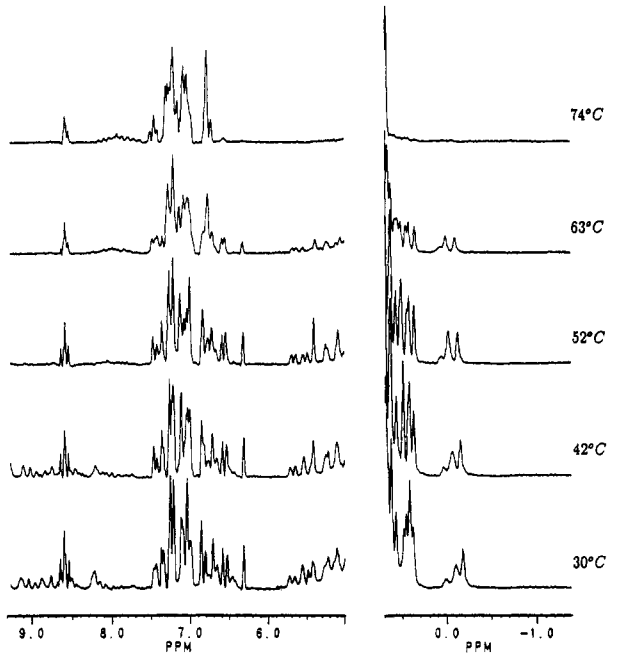

FIGURE 8: 1-D ^{1}H NMR spectra (600 MHz) of the unfolding study of a thermolysin resistant u-PA fragment (residues 159–277) at pH 2.5.

however, due to inconsistencies within the experimental data. Novokhatny and co-workers did not observe the high-stability region of the protease domain because of the solution conditions and the temperature range chosen for their studies; the proteins were not heated to high enough temperatures to observe the final unfolding transition. Furthermore, studies using chemical denaturation did not detect residual structure because they were performed in 4 and 6 M GuHCl at pH 4.5, conditions under which the protein is fully denatured at 29 °C, as shown by the NMR studies reported here.

Having observed two unfolding transitions for inactivated LMW u-PA, it has been our aim to identify the structural regions associated with these transitions. The region of highest stability can be assigned to the N-terminal region of the protease domain by comparison of the spectra of the protease domain with those of the N-terminal fragment. The spectra of the partly unfolded LMW u-PA and the fragment isolated by digestion with thermolysin are closely similar and give evidence for similar secondary and tertiary structure. This can be seen by the presence in both spectra of a number of downfield shifted $C_\alpha H$ proton resonances, indicating β-sheet structure, and resonances of side chain protons shifted away from their random coil chemical shift positions to a comparable extent, indicating similar contacts within the globular fold. The fragment unfolds cooperatively and can be refolded in the absence of the C-terminal subdomain. All of these features lead to the conclusion that this region of the protein when isolated adopts a globular structure similar to that of partly folded LMW u-PA. Spectra of the latter show, however, a subset of resonances observed for the fully folded protein. This indicates that both partly folded LMW u-PA and the isolated N-terminal fragment have close structural similarity to the corresponding region in the native protein.

The proteases of fibrinolytic proteins have considerable homology to the digestive serine proteases. A sequence alignment of chymotrypsin and the protease domain of u-PA (Greer, 1990) shows 33% identical residues and another 40% conserved and similar residues. Comparative modeling studies of u-PA and t-PA have shown that a superposition of the conserved regions is possible and that insertions can be accommodated without difficulty in the chymotrypsin framework (Strassburger et al., 1983; Nienaber et al., 1992). In the chymotrypsinogen structure, Figure 9 (Wang et al., 1985), the regions of the protein forming the two lobes of the structure have no main chain interactions between them. It is therefore reasonable that such lobes could exist as discrete structural units when isolated from each other. Indeed, the isolation of subdomains of serine proteases has been reported for elastase (Ghelis et al., 1978) and for coagulation factor IX (Kisiel et al., 1985; Vysotchin et al., 1993). The NMR spectrum of the highly stable N-terminal region of u-PA is consistent with such a situation, in that this lobe contains only β-sheet structure, and has one tryptophan residue in the core region of the structure. Furthermore, an overlay of the sequence of the isolated fragment of u-PA with that of the homologous region of the chymotrypsinogen structure indicates that, except for the N-terminal residues, the rest of the sequence forms a well-defined six-stranded β-barrel structure (Figure 9).

Having assigned the high-stability region to the N-terminal lobe of the protease domain, we conclude that the less stable region is a subdomain corresponding to the C-terminal lobe. That the resonances assigned to this subdomain correspond to about half the resonances observed in the spectra of LMW u-PA is consistent with this conclusion. This region of the protease domain again has extended β-sheet structure, hence the downfield shifted $C_\alpha H$ proton resonances in the NMR spectrum, and a globular core giving rise to substantial ring current shifted resonances. The sequence of the protease domain of u-PA has three tryptophan residues in the region of the protein corresponding to the C-terminal lobe of digestive serine proteases; resonances attributed to these were identified in the NMR spectrum of this region. Whereas the N-terminal subdomain has been shown to refold to a native-like structure even in the absence of the C-terminal region, we do not have any data on the C-terminal region in the absence of the N-terminal subdomain. Under the conditions used, thermolysin digests the C-terminal region to small peptides.

In the inactivated protein the C-terminal region unfolds cooperatively. In the absence of the specific inhibitor, i.e., in the noninactivated protein, this is no longer the case, and two unfolding transitions can be observed. It is not clear yet which structural region relates to this additional transition. One possibility is that it might be associated with helical and loop regions outside the main β-barrel topology. Another is that it could be related to the activation domain proposed to exist in digestive serine proteases. This is a region that appears from X-ray data to become fully structured only after activation and, in trypsin, correspond to residues 16–18, 142–152, 184–193, and 216–222 (chymotrypsinogen numbering), all of which are in loop regions of the structure (Huber & Bode, 1978; Huber & Bennett, 1983). Support for this idea in the u-PA protease comes from the observation that two tryptophan residues fall into the corresponding region of the sequence of LMW u-PA. Resonances from two tryptophan residues can be seen in spectra of the active u-PA (labeled Trp-I and Trp-II in Figure 6) but not in those of the zymogen (scu-PA) and are furthermore affected by inactivation of LMW u-PA (Nowak, 1992). The increased stability of the inactivated protein might be caused by a stabilization of interdomain contacts via the cross-linking of the two subdomains, which has been postulated for the EGRcmk reaction product (Bode et al., 1989).

These findings for u-PA are interesting to compare with the results of studies of the protease domain of plasmin. DSC

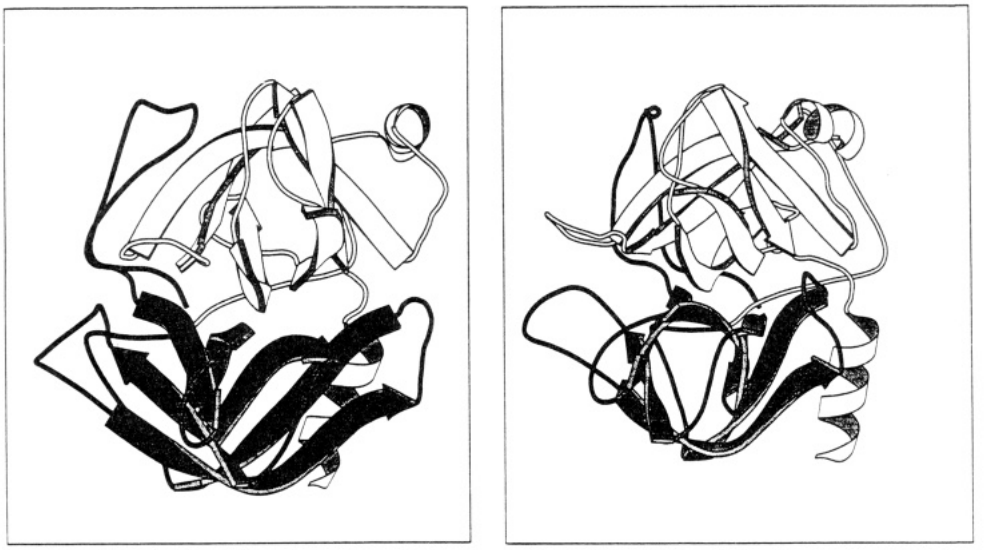

FIGURE 9: Two views of the X-ray structure of bovine chymotrypsinogen (Wang et al., 1985) showing the two lobes of the protein corresponding to residues 1–120 (in black) and 121–245 (in white) and their β-barrel topologies. The structures are presented using the MOLSCRIPT program (Kraulis, 1991). The two lobes would correspond to residues 148–277 and 278–411 of the u-PA sequence.

and NMR studies show that the latter unfolds in a single cooperative transition with a midpoint of 72 °C at pH 4.0 (Teuten et al., 1993) and does not contain a highly stable region as observed for u-PA. This is similar to the cooperative unfolding observed for chymotrypsin (Privalov, 1979) and different from the behavior observed here for u-PA. Two transitions can, however, be observed for the protease domain of plasmin after inactivation (with TLCK, N_α-*p*-tosyl-Lys-chloromethyl ketone, or VPLCK, Val-Phe-Lys-chloromethyl ketone). Part of the inactivated protease domain unfolds independently with a transition temperature that is lower than that of the rest of the protease domain, and which is dependent on the affinity label used. In comparison with plasmin, the region of lowest stablity of u-PA is of somewhat higher stability as it is already structured at 29 °C and pH 4.5 prior to inactivation, but might again be associated with the proposed activation domain of digestive serine proteases.

The results in this paper provide clear evidence for the existence of regions of different stability within the serine protease domain of u-PA. Under specific conditions, e.g., intermediate temperatures and concentrations of GuDCl, therefore, the protein can be obtained in at least two stable partly folded states as well as in the fully folded and fully unfolded ones. As stable partly folded states have been shown for a number of proteins to be related to kinetic intermediates on folding pathways (Dobson, 1992), it will be interesting to see in the case of serine proteases whether this is the case here; no data are presently available to test this possibility. The NMR data indicate that the partly folded states of u-PA contain well-defined "subdomains" of persistent native-like structure. They differ, therefore, from the compact but globally denatured "molten globule" states now being observed and characterized for an increasing number of proteins, such as α-lactalbumin, cytochrome *c*, and myoglobin (Baum et al., 1989; Jeng et al., 1990; Hughson et al., 1990). Such molten globule states have persistent secondary structure but lack the specific tertiary interactions and close packing of side chains characteristic of native proteins (Dobson, 1992; Haynie & Freire, 1993). We suggest that subdomains with persistent and well-ordered structure are particularly likely to exist for well-defined supersecondary structural motifs such as the β-barrel topologies of the lobes of serine protease structures. In the particular case of β-sheets, the specificity of the main chain tertiary interactions involving interstrand hydrogen bonds may be an important factor contributing to this stability. Further studies on stable regions of proteins such as that observed here should shed light on these issues and, perhaps, assist in the design of novel proteins based on specific structural motifs such as the six-stranded β-barrel topology found in the protease fragment isolated in this work.

ACKNOWLEDGMENT

We thank Christel Hentschel for excellent technical assistance in preparing the thermolysin-resistant fragment and Dr. J. Boyd and N. Soffe for assistance with NMR experiments. U.K.N. is supported by an Austrian Erwin Schrödinger Scholarship and is a Guy Newton Junior Research Fellow at Wolfson College, Oxford, U.K. The research of C.M.D. was supported in part by an International Research Scholars Award from the Howard Hughes Medical Institute.

REFERENCES

Baron, M., Norman, D. G., Harvey, T. S., Handford, P. A., Mayhew, M., Tse, A. G. D., Brownlee, G. G., & Campbell, I. D. (1992) *Protein Sci. 1*, 81–90.

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

Bode, W., Fehlhammer, H., & Huber, R. (1976) *J. Mol. Biol. 106*, 325–335.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R., & Hofsteenge, J. (1989) *EMBO J. 8*, 3467–3475.

Bogusky, M. J., Dobson, C. M., & Smith, R. A. G. (1989) *Biochemistry 28*, 6728–6735.

Bokman, A. M., Jimenéz-Barbero, J., Llinás, M. (1993) *J. Biol. Chem. 268*, 13858–13868.

Byeon, I. L., & Llinás, M. (1991) *J. Mol. Biol. 222*, 1035–1051.

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature 327*, 339–341.

Danø, K., Andreasen, P. A., Grøndahl-Hansen, J., Kristensen, P., Nielsen, L. S., & Skriver, L. (1985) *Adv. Cancer Res. 44*, 139–266.

de Vos, A. M., Ultsch, M. H., Kelley, R. F., Padmanabhan, K., Tulinsky, A., Westbrook, M. L., & Kossiakoff, A. A. (1992) *Biochemistry 31*, 270–279.

Dobson, C. M. (1992) *Curr. Opin. Struct. Biol. 2*, 6–12.

Ghelis, C., Templete-Gaillourdet, M., & Yon, J. M. (1978) *Biochem. Biophys. Res. Common. 84*, 31–36.
Greer, J. (1990) *Proteins 7*, 317–334.
Günzler, W. A., Steffens, G. J., Ötting, F., Kim, S.-M. A., Frankus, E., & Flohé, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem. 363*, 1155–1165.
Haynie, D. T., & Freire, E. (1993) *Proteins 16*, 115–140.
Holmes, W. E., Pennica, D., Blaber, M., Rey, M. W., Günzler, W. A., Steffens, G. J., & Heynecker, H. L. (1985) *Bio/Technology 3*, 923–929.
Huber, R., & Bode, W. (1978) *Acc. Chem. Res. 11*, 114–122.
Huber, R., & Bennett, W. S. (1983) *Biopolymers 22*, 261–279.
Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.
Jeng, M.-F., Englander, S. W., Elöve, G. A., Wand, A. J., & Roder, H. (1990) *Biochemistry 29*, 10433–10437.
Kisiel, W., Smith, K. J., & McMullen, B. A. (1985) *Blood 66*, 1302–1308.
Kossiakoff, A. A., Chambers, J. L., Kay, L. M., & Stroud, R. M. (1977) *Biochemistry 16*, 654–664.
Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure, *J. Appl. Crystallogr. 24*, 946–950.
Li, X., Smith, R. A. G., & Dobson, C. M. (1992) *Biochemistry 31*, 9562–9571.
Li, X., Bokman, A. M., Llinás, M., Smith, R. A. G., & Dobson, C. M. (1994) *Biochemistry* (in press).
Mazar, A. P., Buko, A. M., Petros, A. M., Barnathan, E. S., & Henkin, J. (1992) *Fibrinolysis 6*, Suppl. 1, 49–55.
Montelione, G. T., Wüthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D., & Scheraga, H. A. (1992) *Biochemistry 31*, 236–249.
Nienaber, V. L., Young, S. L., Birktoft, J. J., Higgins, D. L., & Berliner, L. J. (1992) *Biochemistry 31*, 3852–3861.
Novokhatny, V. V., Kudinov, S. A., & Privalov, P. L. (1984) *J. Mol. Biol. 179*, 215–232.
Novokhatny, V. V., Ingham, K. C., & Medved, L. V. (1991) *J. Biol. Chem. 266*, 12994–13002.
Novokhatny, V., Medved, L., Mazar, A., Marcotte, P., Henkin, J., & Ingham, K. (1992) *J. Biol. Chem. 267*, 3878–3885.
Novokhatny, V. V., Medved, L. V., Mazar, A., & Ingham, K. C., (1993) *J. Biol. Chem. 268*, 17211–17218.
Nowak, U. K. (1992) Ph.D. Thesis, University of Vienna, Austria.
Nowak, U. K., Li, X., Teuten, A. J., Smith, R. A. G., & Dobson, C. M. (1993) *Biochemistry 32*, 298–309.
Oas, T. C., & Kim, P. S. (1988) *Nature 336*, 42–48.
Oswald, R. E., Bogusky, M. J., Bamberger, M., Smith, R. A. G., & Dobson, C. M. (1989) *Nature 337*, 579–582.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.
Privalov, P. L., & Potekhin, S. A. (1986) *Methods Enzymol. 131*, 4–51.
Radek, J. T., & Castellino, F. J. (1988) *Arch. Biochem. Biophys. 267*, 776–786.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Sottrup-Jensen, L., Claeys, H., Zajdel, M., Petersen, T. E., & Magnusson, S. (1978) *Prog. Chem. Fibrinolysis Thrombolysis 3*, 191–209.
Steffens, G. J., Günzler, W. A., Ötting, F., Frankus, E., & Flohé, L. (1982) *Hoppe Seyler's Z. Physiol. Chem. 363*, 1043–1058.
Strassburger, W., Wollmer, A., Pitts, J. E., Glover, I. D., Tickle, I. J., Blundell, T. L., Steffens, G. J., Günzler, W. A., Ötting, F., & Flohé, L. (1983) *FEBS 157*, 219–223.
Sturtevant, J. M. (1987) *Annu. Rev. Phys. Chem. 38*, 463–488.
Teuten, A. J. (1991) D.Phil. Thesis, University of Oxford.
Teuten, A. J., Smith, R. A. G., & Dobson, C. M. (1991) *FEBS Lett. 278*, 17–22.
Teuten, A. J., Cooper, A., Smith, R. A. G., & Dobson, C. M. (1993) *Biochem. J. 293*, 567–572.
Vysotchin, A., Medved, L. V., & Ingham, K. C. (1993) *J. Biol. Chem. 268*, 8436–8446.
Wang, D., Bode, W., & Huber, W. (1985) *J. Mol. Biol. 185*, 595–624.