Local Structural Preferences of Calpastatin, the Intrinsically Unstructured Protein Inhibitor of Calpain[†,‡,§]

Robert Kiss,[||] Dénes Kovács,[⊥] Péter Tompa,[⊥] and András Perczel*[,||,@]

Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Budapest, Hungary, Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary, and Protein Modeling Group MTA-ELTE, Institute of Chemistry, Eötvös Loránd University, H-1538 Budapest, P.O. Box 32, Hungary

Received February 4, 2008; Revised Manuscript Received May 10, 2008

ABSTRACT: Calpain, the calcium-activated intracellular cysteine protease, is under the tight control of its intrinsically unstructured inhibitor, calpastatin. Understanding how potent inhibition by calpastatin can be reconciled with its unstructured nature provides deeper insight into calpain function and a more general understanding of how proteins devoid of a well-defined structure carry out their function. To this end, we performed a full NMR assignment of hCSD1 to characterize it in its solution state. Secondary chemical shift values and NMR relaxation data, R_1, R_2, and hetero-NOE, as well as spectral density function analysis have shown that conserved regions of calpastatin, subdomains A and C, which are responsible for calcium-dependent anchoring of the inhibitor to the enzyme, preferentially sample partially helical backbone conformations of a reduced flexibility. Moreover, the linker regions between subdomains are more flexible with no structural preference. The primary determinant of calpain inhibition, subdomain B, also has a non-fully random conformational preference, resembling a β-turn structure also ascertained by prior studies of a 27-residue peptide encompassing the inhibitory region. This local structural preference is also confirmed by a deviation in chemical shift values between full-length calpastatin domain 1 and a truncated construct cut in the middle of subdomain B. At the C-terminal end of the molecule, a nascent helical region was found, which in contrast to the overall structural properties of the molecule may indicate a previously unknown functional region. Overall, these observations provide further evidence that supports previous suggestions that intrinsically unstructured proteins use preformed structural elements in efficient partner recognition.

Calpain, a Ca^{2+}-activated cysteine protease, is common in eukaryotic cells and participates in basic cellular functions such as regulation of cell division, differentiation, and cell motility (*1*). In accordance, calpain activity is tightly controlled by the level of free Ca^{2+} and its endogenous protein inhibitor, calpastatin (*2*). Vertebrates contain multiple isoforms of this enzyme; some, like μ- and m-calpain, are ubiquitously expressed, while others, such as p94 and p82, are more tissue specific. Calpastatin is also ubiquitously expressed, usually at levels exceeding that of the enzyme(s) (*3*). Calpastatin is composed of four equivalent inhibitory domains, each of which is capable of inhibiting a separate calpain molecule (*4*) (Figure 1). Within each domain, three short, conserved segments or subdomains are primarily responsible for the observed inhibitory effect. Subdomains A (S^{12}−G^{30}) and C (S^{87}−C^{105}) (using the numbering of Figure 1) bind the enzyme in a Ca^{2+}-dependent manner and potentiate inhibition (*5*–*7*), whereas subdomain B (M^{50}−R^{70}) binds in the active site region of the enzyme and carries inhibitory potential in itself (*7, 8*). The physical separation

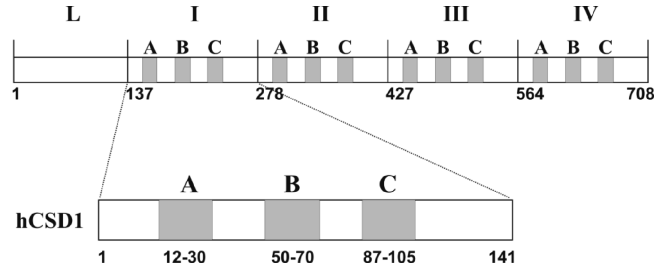

FIGURE 1: Domain structure of human calpastatin. Each of the four inhibitory domains (I, II, III, and IV) contains three conserved subdomains, A, B, and C, primarily responsible for inhibition of calpain. The numbering of full-length calpastatin is given from position 1 of the N-terminus of the L domain, which does not inhibit calpain but has some other function. The inset shows domain I (hCSD1), characterized in this work, enlarged and numbered in a manner independent from the rest of the molecule. This numbering is used throughout the rest of this paper.

[†] This research was supported by grants from the Hungarian Scientific Research Fund (OTKA K60694, K72973, NI-68466, and TS049812). P.T. acknowledges the support of Wellcome Trust International Senior Research Fellowship ISRF 067595.

[‡] The chemical shifts and relaxation parameters have been deposited in the Biological Magnetic Resonance Data Bank (BMRB) as entry 15766.

[§] We dedicate this work to Hedvig Medzihradszky-Schweigeron on the occasion of his 80th birthday.

* To whom correspondence should be addressed: Institute of Chemistry, Eötvös Loránd University, H-1538 Budapest, P.O. Box 32, Hungary. Telephone: 36-1-209-0555, ext. 1653. Fax: 36-1-3722-620. E-mail: perczel@chem.elte.hu.

[||] Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University.

[⊥] Hungarian Academy of Sciences.

[@] Protein Modeling Group MTA-ELTE, Institute of Chemistry, Eötvös Loránd University.

10.1021/bi800201a CCC: $40.75
Published on Web 06/07/2008

of these different motifs most likely ensures very fast and extremely specific interaction with the enzyme. The ensuing effective inhibition of calpain is particularly important for its proper function, as witnessed under a variety of pathological conditions that result from the loss of control of the calpain−calpastatin system (*9, 10*). Uncovering the mechanism of binding and inhibition of calpain by calpastatin, however, is important not only for understanding calpain function and malfunction but also for meeting the challenge of structurally characterizing the newly recognized class of intrinsically unstructured proteins (IUPs).[1] IUPs lack a well-defined three-dimensional (3D) structure under native conditions but still carry out diverse and specific functions often intimately linked with structural disorder (*11–13*). These proteins prevail in various proteomes, and their frequency increases with an increase in the complexity of the organism (*14, 15*). Flexibility and structural adaptability of their polypeptide chain provide specific functional advantages such as increased speed of interaction, specificity without excessive binding strength, the capacity of binding multiple binding partners, and harboring distinct activities within the same region of the protein (*11–13*). In general, recognition of the existence and efficient functioning of IUPs demands the structure−function paradigm of proteins to be extended and partially re-assessed (*16, 17*).

Calpastatin belongs to this class of proteins. It lacks appreciable tertiary structure or even well-defined secondary structure elements in isolation (*16, 19, 20*), and its binding to calpain, accompanied by a large-scale induced folding, is extremely fast. Our working hypothesis, which might resolve the potentially contradicting demands of high speed and extensive structural reorganization, is that calpastatin is not fully disordered prior to binding but possesses transient structural elements critical in partner recognition and subsequent local folding. In keeping with previous suggestions that IUPs exploit limited structural organization, in the form of preformed structural elements (*20*), primary contact sites (*16*), or molecular recognition elements (*15*), for effective partner recognition we assumed calpastatin binding also relies on these structural−functional devices. As subdomains A and C adopt helical structures when bound to calpain (*21*) (Figure 2) and also display significant helical tendencies when studied in forms of isolated peptides (*22*), we hypothesized that in the full-length inhibitor they do retain certain features typical of the bound state. Two-dimensional ^{1}H NMR experiments on a 27-residue peptide encompassing subdomain B (*23*) suggested that this region adopts multiple conformations except for the T^{64}−R^{70} segment which strictly forms a type I β-turn. Furthermore, Ala-scanning mutagenesis studies revealed that L^{58}−G^{59} and T^{64}−P^{66} conformations (L^{11}−G^{12} and T^{17}−P^{19}, respectively, in the original numbering) are essential for preserving the proper inhibitory function of this peptide (*24*).

To trace residual structural features of full-length human calpastatin domain 1 (hCSD1), suitable triple-resonance NMR experiments on ^{15}N and on doubly ^{15}N- and ^{13}C-labeled proteins were completed. It is to be stressed here that full assignment and analysis of an IUP are rather demanding due to rapid fluctuations over a structural ensemble of the protein. Such a detailed analysis of an IUP has only been carried out in a few cases, such as CREB (*25*), p27 KID (*26*), FnBP (*27*), α-synuclein (*28*), p53 TAD (*29*), FlgM (*30*), and tau repeat domain (*31*). By analyzing NMR relaxation and chemical shift data, we show here that both subdomains A and C dominantly sample helical states in the conformational ensemble. The region separating these two subdomains behaves as a flexible linker with some residual structural features within subdomain B. This information about structure could also be ascertained by comparing secondary chemical shift values of full-length human calpastatin domain I (hCSD1) and its C-terminal half, hCSD1(67−141). These analyses have also revealed significant structural preferences within the region on the C-terminal side of subdomain C (E^{118}−T^{131}). Correlation of these observations with biochemical data suggests that this region may also have an important role in calpastatin function. The possibility of considering this region as a separate functional region of the molecule is discussed.

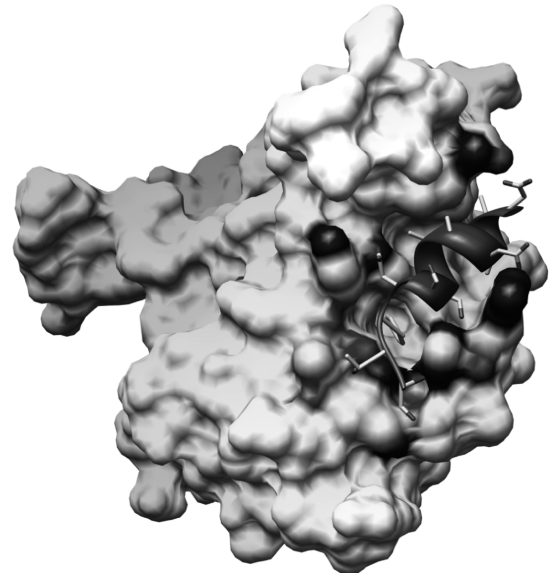

FIGURE 2: Three-dimensional structure of calpastatin subdomain C bound to calpain domain VI. X-ray structure of the calmodulin-like domain VI of the porcine m-calpain small subunit, binding a 19-residue peptide corresponding to the C subdomain of human calpastatin domain I (PDB entry 1nx0).

MATERIALS AND METHODS

Sample Preparation. ^{15}N-labeled and ^{15}N- and ^{13}C-labeled full-length hCSD1 [corresponding to A^{137}−K^{277} of human calpastatin, SwissProt entry P20810 (cf. also Figure 1)] were produced by using the pEThCSD1 vector (courtesy of M. Maki, Nagoya University, Nagoya, Japan) and *Escherichia coli* BL21 strain. Transformed cells were grown on a minimal medium (5 g of glucose, 6 g of Na_2HPO_4, 3 g of KH_2PO_4, 1 g of NH_4Cl, 0.5 g of NaCl, 0.12 g of $MgSO_4$, and 0.01 g of $CaCl_2$ per liter of medium) at 37 °C to an OD_{600} of ∼0.4−0.5, induced with 0.5 mM isopropyl β-D-thiogalactopyranoside and grown overnight at 30 °C to allow protein expression. The C-terminal half of calpastatin (the position in domain 1, P^{67}−K^{141}; the position in whole calpastatin,

[1] Abbreviations: SCS, secondary chemical shift; hCSD1, domain I of human calpastatin; hCSD1(67−141), C-terminal 67-residue part of human calpastatin domain I; IUP, intrinsically unstructured protein; DIA, calpastatin domain I A subdomain; DIC, calpastatin domain I C subdomain; sdf, spectral density function; RC, random coil.

$P^{204}-K^{277}$) was constructed with regular cloning techniques, using PCR, and was ligated into a pET22b-based vector. The ^{15}N-labeled and ^{15}N- and ^{13}C-labeled protein were expressed in a similar manner. Both hCSD1 and N-terminally truncated hCSD1 were initially purified as previously described (*16*) and then were further purified on a C-18 column using a HPLC system. The final purity of lyophilized samples was between 96 and 98%.

The purity control of hCSD1 prior to each NMR experiment was achieved by mass spectrometry. ESI-MS was performed on a Bruker (Bremen, Germany) Esquire 3000 Plus ion trap mass spectrometer, operating in continuous sample injection at a flow rate of 4 μL/min. Samples were dissolved in 50% acetonitrile and 50% water containing 0.01% acetic acid. Mass spectra were recorded in positive mode in the *m/z* 200−1500 range. Mass spectra of hCSD1 are presented in the Supporting Information.

NMR Assignment and Relaxation Measurements. NMR samples contained ~1 mM protein dissolved in a 90% H_2O/10% D_2O mixture. All NMR data were collected on a BRUKER DRX 500-MHz instrument and were referenced to internal DSS at pH 6.07 in the case of hCSD1 and pH 6.17 in the case of hCSD1(67−141). Resonance assignments of backbone carbon and nitrogen atoms were obtained from $^1H-^{15}N$ HSQC, HNCA, HN(CO)CA, HNCACB, HN(CO)-CACB, and C(CO)NH experiments (2048 complex points in the direct ^1H dimension and 128 and 64 complex points in the ^{15}N and ^{13}C dimensions, respectively). Typical spectral widths were 10 ppm in the direct dimension, 28 ppm in the nitrogen dimension, 60 ppm in the Cα and Cβ dimensions, and 32 ppm in the Cα dimension. Proton assignments were accomplished using 3D $^1H-^{15}N$ TOCSY-HSQC (~70 ms spin-lock time) (1024, 64, and 220 complex points) and 3D $^1H-^{15}N$ NOESY-HSQC experiments (150 ms mixing time) (1024, 64, and 220 complex points).

The pH dependence of both proteins, full-length hCSD1 and its C-terminal half, was investigated by measuring $^1H-^{15}N$ amide cross-peak intensities from the appropriate HSQC spectra collected at 298 K and at pH 4.3, 5.23, and 6.17 for hCSD1(67−141) as well as pH 3.85, 5.53, 6.07, and 7.25 for hCSD1. The temperature dependence of the same type of resonances was measured at 280, 300, and 320 K in aqueous solution for hCSD1(67−141). To maintain conditions for quantitative comparison of peak heights, the same processing protocol was used for all spectra. All data were processed using nmrPipe (*32*). Spectra were apodized with a 90°-shifted cosine-squared function and zero-filled before Fourier transformation. Polynomial baseline correction was applied in the direct ^1H dimension only. Linear prediction was applied by duplicated complex points in the ^{15}N dimensions. Resonance assignments were completed using Sparky (*33*).

A total of 121 N−H resonances (out of 126) were adequately resolved for R_1, R_2, and hetero-NOE type relaxation studies of hCSD1. Chemical shifts were obtained from triple-resonance and HSQC experiments as described above. Secondary chemical shifts reported here are the differences in the measured ^{13}C or ^1H chemical shifts with respect to the random coil values for each type of amino acid residue. For the calculation of the secondary chemical shift, we used the random coil (RC) values determined by Wishart and Skyes (*34*). These RC values were collected at pH 5, by using a 1 M urea solution of peptide GGXAG or GGXPG at a low concentration. Among the different RC data available in the literature, the experimental conditions published in ref *35* are the closest to those used in this study.

For comparison of ^{15}N and ^1H resonance frequencies of the amide groups along full-length hCSD1 and hCSD1(67−141), the following expression was used (*35*):

$$\Delta\Delta\delta = \{[\delta(^1H)_{hCSD1} - \delta(^1H)_{hCSD1(65-141)}]^2 + 0.17[\delta(^{15}N)_{hCSD1} - \delta(^{15}N)_{hCSD1(65-141)}]^2\}^{1/2}$$

where $\delta(^1H)_{hCSD1}$ and $\delta(^1H)_{hCSD1(67-141)}$ are the appropriate measured resonance frequencies (data are from HSQC type spectra recorded at pH 6.07 and 298 K).

The $^3J_{HNH\alpha}$ scalar coupling constants were determined with 3D HNCA-E.COSYcosy (*36*) type experiments.

NMR ^{15}N relaxation data were acquired by using a spectral width of 10 ppm [2048 complex points in the direct and 256 complex points in the nitrogen dimension (23 ppm)]. R_1 and R_2 data were collected for a total of 18 h each. Spectra used for R_1 analysis were collected by using the following relaxation delay times: 1.2, 51.2, 101.2, 251.2, 751.2, 1001.2, 2001.2, and 3251.2 ms. R_2 data were measured using the following relaxation delays: 17.59, 35.18, 52.77, 70.36, 105.54, 158.31, 211.08, 316.62, and 386.98 ms. For the determination of R_1 (spin−lattice) and R_2 (spin−spin) relaxation rates, resonance heights were extracted and fitted as a function of the relaxation delay time by using a two-parameter exponential fitting routine of Sparky (*33*). The steady-state NOE data for hCSD1, with a relaxation delay of 5 s, were collected at 298 K and pH 6.17 using the pulse sequence described by Palmer et al. For measuring NOE enhancement, peak intensities were used.

Relaxation data were analyzed using the reduced spectral density mapping approach as described by Lefevre et al. (*37*). Spectral densities at given frequencies were calculated form R_1, R_2, and hetero-NOE values of the amide proton and nitrogen atoms.

RESULTS

Resonance Assignments. Thermal fluctuations and the lack of a unique chemical environment make the NMR resonances of IUPs fall rather close to the average or so-called random coil values. The ensuing low signal dispersion (Figure 3) requires double labeling even for molecules smaller in size (>10 kDa). Selected triple-resonance measurements were therefore used to obtain an almost complete ^{13}C resonance assignment of both full-length human calpastatin domain I, hCSD1, and its C-terminal fragment, hCSD1(67−141). ^1H chemical shifts were obtained from ^{15}N-edited TOCSY and NOESY spectra. For hCSD1(67−141), 58 spin system of the 67 non-proline residues were determined. In the case of hCSD1, as many as 139 spin systems were identified, exceeding the total number of non-proline residues (126). However, only 121 amino acid residues were finally assigned, because of the lack of well-defined sequential information encoded in the spectra identified above. The flexible nature of the backbone leads to narrower, lines resulting in an adequate signal-to-noise ratio. The purity of the samples was controlled by MS measurements and gel prior to each experiment. We found that the samples were free of

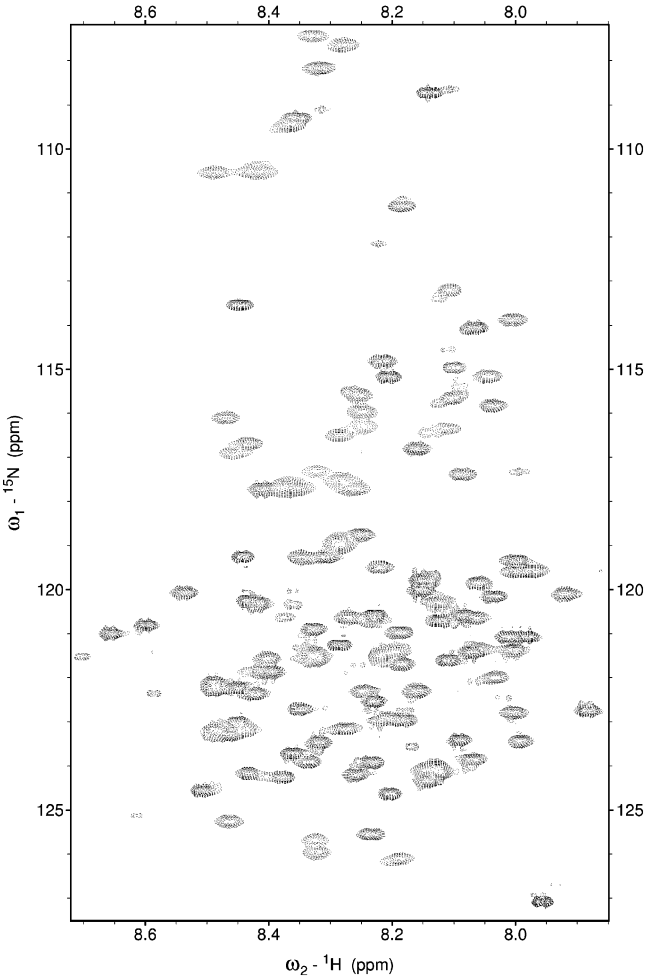

FIGURE 3: ^{1}H−^{15}N HSQC spectrum of hCSD1 recorded at 298 K and pH ~6.1. IUPs typically exhibit poor chemical shift dispersion especially in the proton dimension. Nevertheless, the low half-width of the resonances at acceptable sensitivity make possible the spectral assignment.

degradation as well as of aggregation. Thus, we suppose that the presence of low-intensity peaks could be the consequence of the cis−trans isomerization in the vicinity of proline residues, although in ^{13}C-based experiments the intensities of minor comformers are not good enough to accomplish a full NMR resonance assignment. The fact that minor conformers seem to correspond to residues neighboring prolines supports the idea that besides the more populated trans form cis isomers are also present.

Temperature and pH Dependence of Spectra. hCSD1-(67−141) did not exhibit any significant chemical shift change in the 280−320 K temperature range. For most amide proton resonances, as recorded for ^{1}H−^{15}N HSQC spectra, the temperature-induced chemical shifts are <0.01 ppm/K. For example, even the N−H shift of Gly^{91}, one of the most affected resonances, has a $\Delta\delta/\Delta T$ value of only 2.3 ppb/K. Moreover, HSQC spectra were recorded at pH 4.3, 5.23, and 6.17 for hCSD1(67−141) and at pH 3.85, 5.53, 6.07, and 7.25 for hCSD1 to investigate the influence of pH on chemical shift and peak intensities. Since the physiological pH is 7, measurements were carried out to approach this value. We found that an increasing pH (from 4.3 to 6.4) resulted in insignificant peak drifting with rather similar line shape characteristics. However, at pH >6.5, changes in both Hα and H_N chemical shifts stopped but the extent of line broadening due to chemical exchange increased dramatically. As a consequence of these observations, we have carried out our measurements at 298 K and pH 6.1.

Secondary Chemical Shifts and $^3J_{HNC\alpha}$ Coupling Constants. As described by Dyson et al. (*38*), $^3J_{HNC\alpha}$ coupling constants are only moderately informative for IUPs. In our case, most $^3J_{HNC\alpha}$ values are between 5 and 7 Hz and do not show any site specific pattern. As the typical helical $^3J_{HNC\alpha}$ value is 4 Hz, the observed values correspond to the mainly random character of hCSD1 (data not shown).

Selected resonance frequencies (e.g., CO, Cα, and Hα) of proteins are highly sensitive to the local molecular environment. Although efforts to quantitate such dependence are in progress (*39*), currently the most common approach to interpreting chemical shift data relies on empirical structure−chemical shift correlation values.

Because of their largely disordered character, SCS values of IUPs are expected to be close, or equal, to zero. Upon closer inspection, however, some IUPs turn out to contain various, and often significant, amounts of function-related structural elements (*16*) which can be effectively monitored by SCS values (*40*). To characterize structural propensities of hCSD1, a secondary chemical shift analysis was completed for ^{13}CO, ^{13}Cα, and ^{1}Hα nuclei (Figure 4).

Analysis of SCSs along the sequence shows evidence for residual helices in several regions. On the basis of Hα and Cα SCSs, short segments with limited helical propensity can be recognized in a time-averaged manner between residues D^{18} and I^{25}, S^{51} and G^{59}, R^{70} and K^{75}, and T^{121} and R^{133} (see Table 1).

On the basis of Cα SCSs, it is immediately functionally relevant that subdomains A and B, two characteristic binding and functional sites of the inhibitor, have some helical character. Cα secondary chemical shifts of residues in subdomain C, G^{91}−T^{104}, also exhibit some helical propensities. As mentioned above, a significant helical preference can also be ascertained for the T^{121}−R^{133} region on the basis of Cα and Hα data. The SCS data analysis of carbonyl carbons, CO (*35*), suggests a helical tendency for the T^{121}−R^{133} region (0.21 ± 0.09). Moreover, these contiguous sequential units, incorporating 7−10 residues, are sufficiently long to accommodate several helical turns, corresponding to transiently populated helical backbone conformers.

The T^{121}−R^{133} region has never been implicated in calpastatin action, but its nonrandom structural behavior suggests some role in calpain inhibition. By aligning the three inhibitory domains of human calpastatin, this region appears to be rather well conserved, which also suggests functional significance (Figure 5). Exon−intron boundaries of the calpastatin gene roughly correspond to the functional units of calpastatin, i.e., subdomains A, B, and C, which attests to the primordial assembly process of the inhibitor.

The inhibitory region of calpastatin, i.e., subdomain B, also deviates from random coil values (Figure 4). Furthermore, selected structural information can also be retrieved from the literature on subdomain B, established by two-dimensional ^{1}H NMR experiments (*23*) based on the analysis of a 27-residue model peptide. These data suggest that P^{20}−R^{23} (corresponding to P^{67}−R^{70}, respectively, in hCSD1) adopt a type I β-turn. This region falls within the highly conserved TIPPXYR consensus motif within subdomain B, long proved to be primarily responsible for calpain inhibition

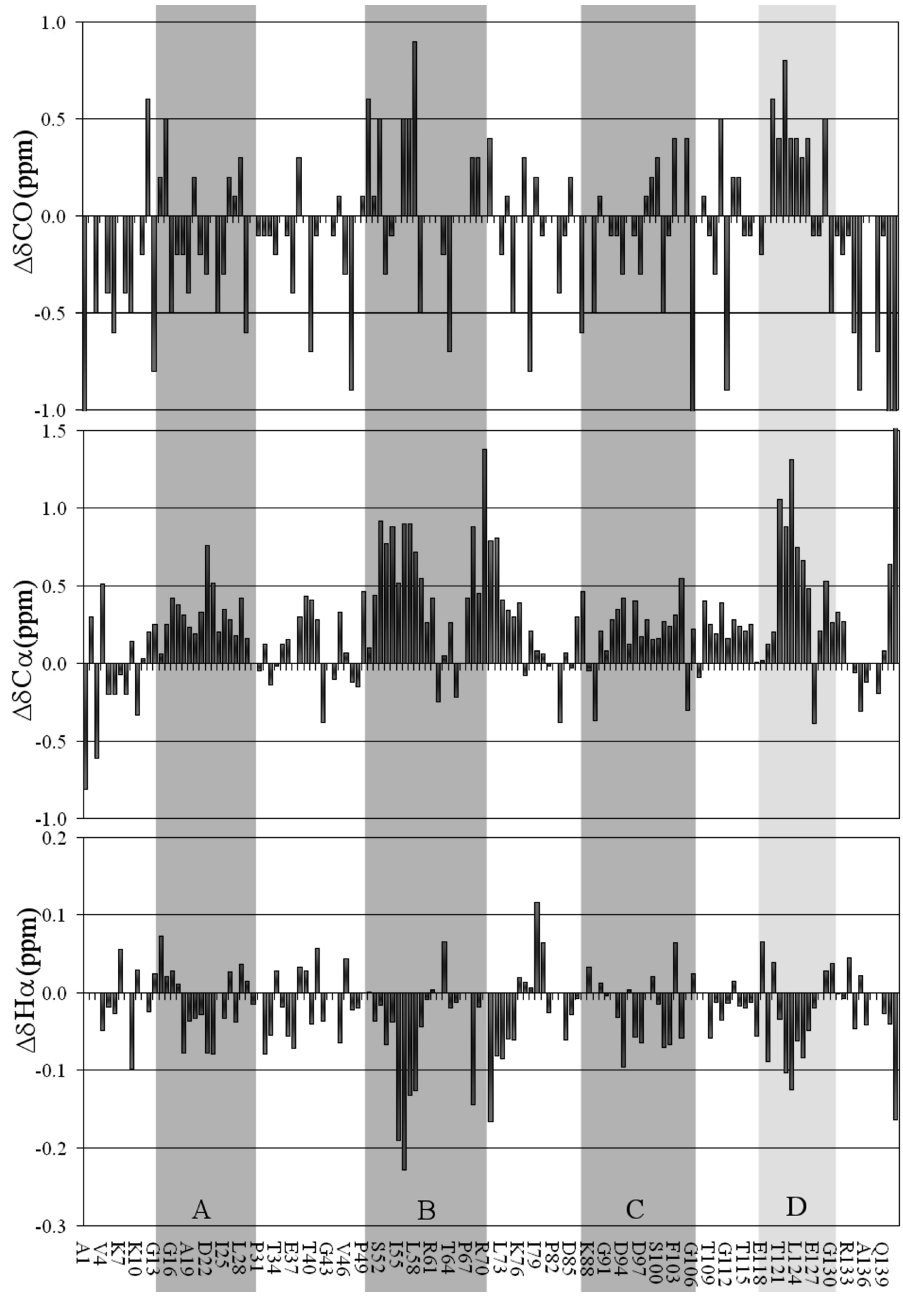

FIGURE 4: Secondary chemical shifts of hCSD1. Secondary chemical shifts of CO (top panel), Cα (middle panel), and Hα (bottom panel) of hCSD1 as a function of the primary sequence. Regions with positive deviations in Cα and negative deviations in Hα SCS indicate some helical preference. Subdomains A, B, and C are highlighted in gray. The newly suggested structural motif in the C-terminal region of the protein is highlighted in light gray.

Table 1: Average Secondary Chemical Shifts of the Functionally Important Regions of hCSD1[a]

	A	B	C		helical motif
sequence	$D^{18}-I^{25}$	$S^{51}-G^{59}$	$R^{70}-K^{75}$	$G^{91}-T^{104}$	$T^{121}-R^{133}$
SCS (Cα)	0.36 ± 0.19	0.71 ± 0.19	0.53 ± 0.14	0.27 ± 0.07	0.56 ± 0.26
SCS (Hα)	0.05 ± 0.03	-0.10 ± 0.09	-0.09 ± 0.05	0.03 ± 0.05	-0.03 ± 0.02

[a] From the average Cα values of subdomains A, B, and C, the existence of some helixlike secondary structural element seems to be supported. See also Figure 4.

by Ala-scanning mutagenesis (*24*). It is obvious from both Hα and Cα SCSs that these residues indeed break the helical tendency of residues surrounding them in hCSD1, namely, the $S^{51}-G^{59}$ and $R^{70}-K^{75}$ segments.

Truncation Changes Chemical Shift Values. Even though N^{HN} and H^{HN} resonance frequencies are not the primary choice for analyzing the secondary structure content of proteins, their unique sensitivity to the chemical environment can be used to monitor even subtle structural changes in the protein (*35*). Besides pH, temperature, etc., the chemical shift values of both $^{15}N^{HN}$ and $^{1}H^{HN}$ nuclei of an amino acid residue within a protein depend on the residue type itself, on neighboring residues, and on the secondary or tertiary structure of the molecule (*34*). Our construct corresponding

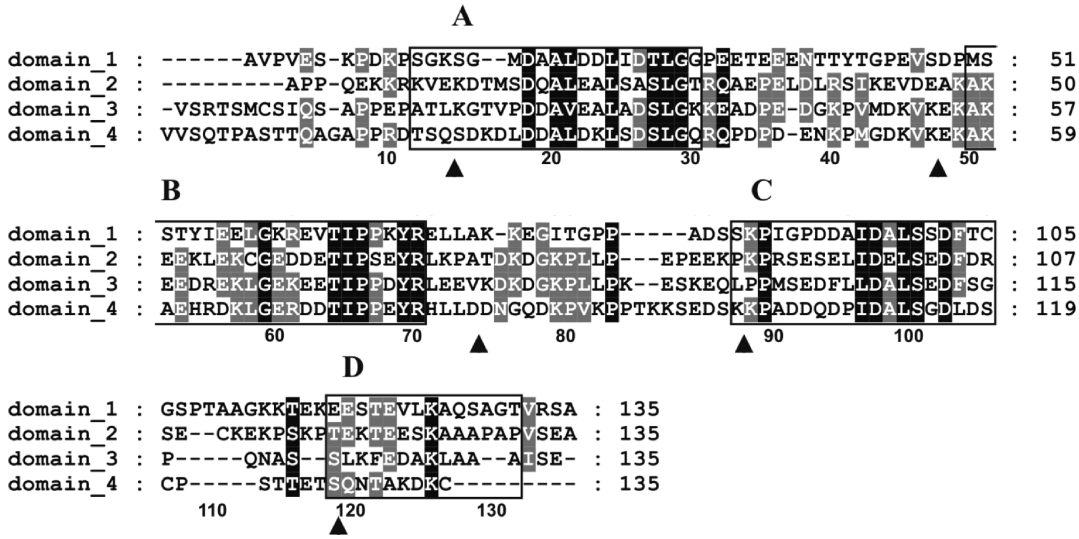

FIGURE 5: Alignment of human calpastatin inhibitory domains. Human calpastatin inhibitory domains have been aligned with ClustalW (*41*) and are shown here with identical residues highlighted in black and strongly similar residues in gray. Subdomains A, B, and C as well as the helical structural motif are boxed. Exon−intron boundaries, as they apply to domain 1, are denoted with large arrowheads to indicate units that contributed to the assembly of the calpastatin gene. (Please note that the numbering due to differences in the lengths of the inhibitory domains differs from that in the text.)

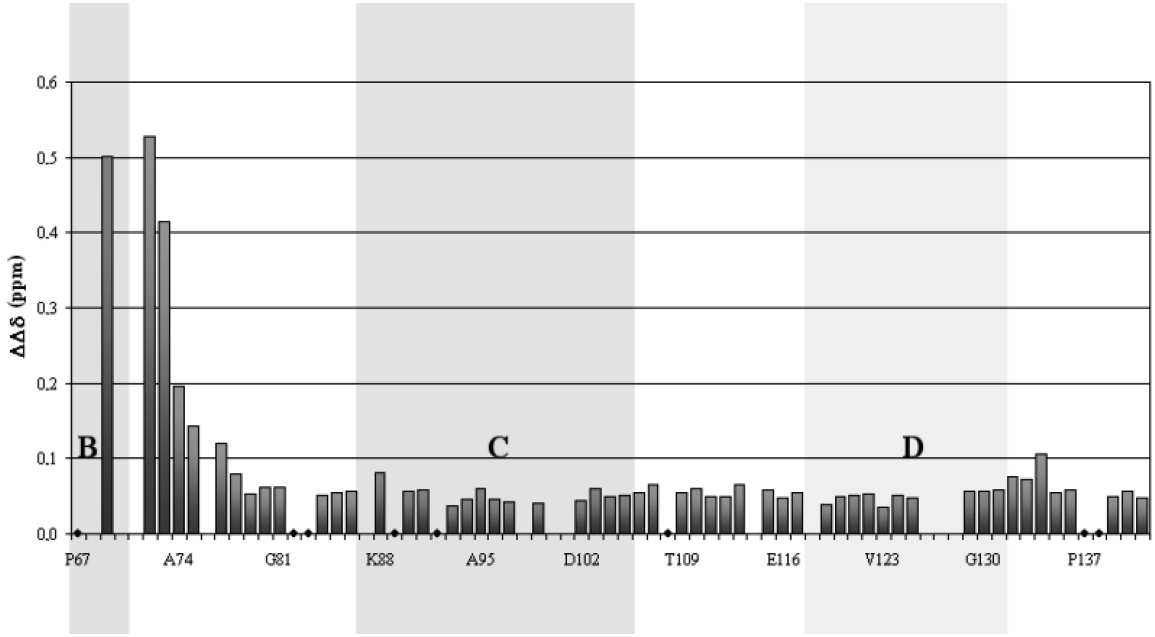

FIGURE 6: Comparison of secondary chemical shift values of full-length hCSD1 and its C-terminal half. The combined N^{HN} and H^{HN} chemical shift differences of full-length hCSD1 and its C-terminal half, hCSD1(67−141), derived from the $^{1}H-^{15}N$ HSQC spectra, are shown. Values close to zero (0.04 ppm) are clear indications of a similar type of secondary structure and/or IUP character of both proteins ($I^{79}-K^{141}$). The larger deviation form zero toward the N-terminal end ($P^{67}-G^{78}$) suggests that truncated hCSD1 must have an increased flexibility in this region, reflecting indirectly the fact that here in full-length hCSD1 some sort of time-average residual structure (e.g., β-turns) could be present.

to the C-terminal half (residues 67−141) of hCSD1, which consists of subdomain C of calpastatin cut in the middle of the key inhibitory TIPPXYR segment, enables us to test signs of transient long-range interactions within calpastatin, suggested by the nonadditivity of the CD spectra of its two halves previously (*22*). By plotting the combined N^{HN} and H^{HN} chemical shift differences of hCSD1 and its C-terminal fragment, hCSD1(67−141), we find values for the 62-amino acid C-terminal part ($I^{79}-K^{141}$) of hCSD1 are uniformly close to zero (0.02 ± 0.012 ppm) for all residues located between I^{79} and K^{141} (Figure 6). This strong similarity between the appropriate N−H chemical shifts of the two molecules could be the consequence of either identical "high-order structure" or the total lack of recognizable long-range tertiary interactions.

On the other hand, combined N^{HN} and H^{HN} chemical shift differences for the N-terminal dodecapeptide ($P^{67}-G^{78}$) are significantly different form zero (0.3 ± 0.03 ppm), with residues closest to the N-terminus showing the largest deviations (Figure 6). This effect extends rather long, which probably indicates both a remarkably increased local flexibility elicited by the new terminus generated by the truncation and also that hCSD1 at subdomain B must have some residual structure eliminated via removal of its sequential neighbors. The observed CS difference between

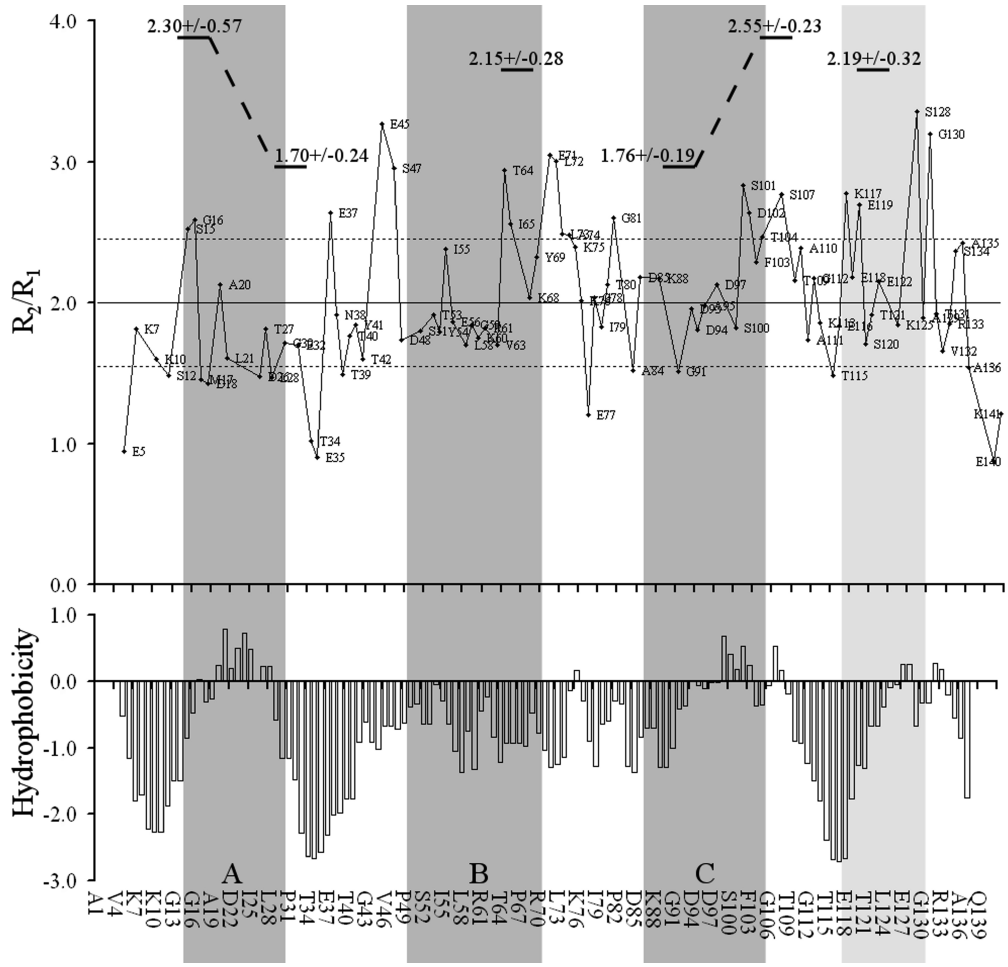

FIGURE 7: R_2/R_1 ratio (top panel) and hydrophobicity indexes (bottom panel) of hCSD1 as a function of its amino acid sequence. R_2/R_1 values of 1 reveal motions occurring on a faster time scale. The solid line shows the average ($\langle\rangle$) and the dashed line the $\langle\rangle + \sigma$ and $\langle\rangle - \sigma$ values, where σ stands for the standard deviation. The hydrophobicity plot is based on the Kyte and Doolittle scale (*43*). Data for subdomains A, B, and C are highlighted in gray. The average R_2/R_1 values of the five N- and C-terminal residues of subdomains A and C, plus five additional central residues of subdomain B and the C-terminal structural motif, are depicted at the top.

the truly unfolded N-terminal tail ($P^{67}-G^{78}$) and the $I^{79}-K^{141}$ fragment of hCSD1 and hCSD1(67−141) also suggests that the latter sequence unit in hCSD1 must have at least some residual structural features while the N-terminal tail of hCSD1(67−141) is unstructured to a greater degree.

Relaxation Measurements and Reduced Spectral Density Function Analysis. As for globular proteins, NMR relaxation experiments have the potential to provide valuable insights into the internal molecular motion of the unfolded and/or partly folded states. Variations of backbone dynamics are pursued for detection of propensities of local structural constraints of the polypeptide chain which result in the restriction of backbone motions (*38*).

Typical ^{15}N R_2 values of residues situated in a globular protein similar in size (~16 kDa) at a given magnetic field strength (e.g., 500 MHz) are 13.8 s^{-1} (*42*). The significantly lower spin−spin relaxation values ($R_{2av.} = 3.89 \pm 0.94$ s^{-1}) measured for hCSD1, in conjunction with the all-negative hetero-NOE data (see the Supporting Information), are clear indications that calpastatin is a predominantly unstructured protein.

For depiction of relaxational propensities of the molecule, the R_2/R_1 ratio proved to be an apposite tool. When higher-frequency motions (e.g., ω_N) dominate the ^{HN}N relaxation, the values of both R_1 and R_2 have a relative ratio of 1. However, for slower time scale motions, R_2 is determined more and more by slower motions pushing the R_2/R_1 value over 1. The apparently measured R_2 is further increased by chemical exchange (R_{ex}) since low-frequency molecular motions (microseconds to milliseconds) can influence transversal relaxation. When the R_2/R_1 values are analyzed in terms of the primary sequence of full-length hCSD1 (Figure 7), a significant fluctuation (standard deviation of 0.56) around the average value ($\langle R_2/R_1\rangle = 2.02$) is noticed. The high fluctuation of R_2/R_1 has a weak chance of a common global motion and/or different R_{ex} rates for the amino acid residues.

The spectral density function (sdf) (*37*) describes the relative amount of N−H bond vector fluctuation, allowing a more direct evaluation of the relative degree of motion at three different frequencies, namely, at ω_N, ω_H, and zero. $J(\omega_N)$ is the most sensitive to R_1 but does not discriminate well enough between faster and slower motions. $J(0)$ is influenced by low-frequency motions (nanoseconds) along with some fluctuations occurring on the millisecond to microsecond time scale arising form chemical exchange (*38*). Values of $J(0.87\omega_H)$ do not change much along the primary sequence except for a few residues, signaling that high-frequency fluctuations occur throughout the whole polypeptide chain. For hCSD1, $J(\omega_N)$ values of most residues are

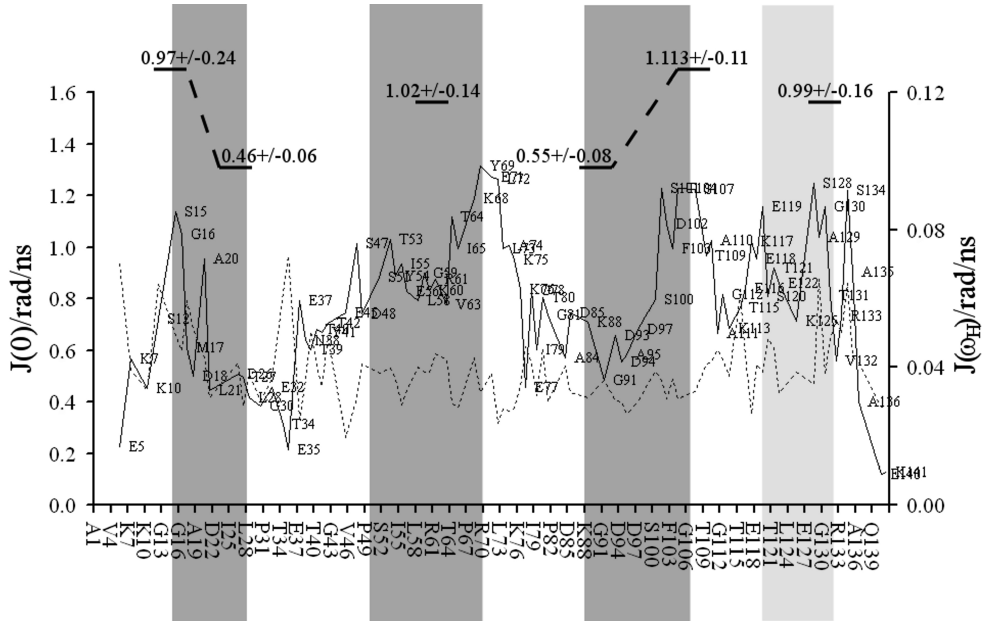

FIGURE 8: Spectral density function of hCSD1 at given frequencies [$J(0)$ and $J(\omega_H)$] as a function of the primary sequence. $J(0)$ values are indicated with a solid line and $J(\omega_H)$ values with a dashed line.The average $J(0)$ values of the five N- and C-terminal residues of subdomains A and C, plus five additional central residues of subdomain B, are depicted at the top.

quasi equal [$J(\omega_N)_{av} = 0.303 \pm 0.07$], while those of $J(0)$ and $J(\omega_H)$ vary along the polypeptide chain, revealing sequence specific mobility information (Figure 8).

Restricted motions on a subnanosecond time scale indicated by larger than average $J(0)$ values are observed for $G^{13}-M^{17}$, $K^{68}-L^{72}$, $S^{101}-C^{105}$, and $S^{128}-V^{132}$ (Figure 8). These residues of restricted mobility also present some residual local structural features highlighted both by secondary chemical shifts, SCS, and by their hydrophobicity pattern (Figure 7). For example, the $G^{13}-M^{17}$ segment is located in a region where hydrophilicity decreases while hydrophobicity increases markedly and where SCSs (e.g., $\Delta\delta C\alpha$) forecast some residual helicity (Figure 4).

Regions with lower than average $J(0)$ values are $D^{26}-G^{30}$, $L^{58}-E^{62}$, $I^{90}-D^{94}$, and $T^{121}-K^{125}$ (Figure 8). For these residues, except $T^{121}-K^{125}$, SCSs (e.g., $\Delta\delta C\alpha$) forecast that not even traces of helicity can be detected. Therefore, amide groups of the latter residues can be characterized by faster and local molecular motions showing almost no sign of cooperativity.

Interestingly, most of the sequential regions with characteristic dynamics properties discussed above coincide with the boundaries of the subdomains of hCSD1. For example, $K^{68}-L^{72}$, $S^{101}-C^{105}$, and $S^{128}-V^{132}$ residues, all presenting signs of synchronized mobility, are located at the C-termini of subdomains B and C and the C-terminal structural motif, respectively. Furthermore, both $D^{26}-G^{30}$ and $I^{90}-D^{94}$ subunits are positioned at the edges of subunits A and C.

DISCUSSION

Uncovering the mechanism of action of the intrinsically unstructured inhibitor of calpain is important for gaining a better understanding of the regulation of calpain function, and also for extending our rather limited structural knowledge of IUPs. IUPs exist and function without a well-defined 3D structure, which has called for the re-assessment of the classical structure−function paradigm (*17*). In practical terms, this re-assessment demands detailed structural and dynamic characterization of distinct IUPs in both solution and partner-bound states, to elucidate the special functional features their unstructured nature provides. In particular, a recent surge of reports has underlined the notion that IUPs are not fully unstructured but harbor function-related residual structure in the form of preformed structural elements (*20*), primary contact sites (*16*), or MoREs, i.e., molecular recogition elements (*15*). The basic idea behind these diverse concepts is that local residual structure resembles the bound state of the IUPs and thus limits the necessary conformational search for achieving the partner-bound state. This, in turn, enables specific and kinetically and/or thermodynamically effective interaction with the partner. A current endeavor in IUP research is the experimental characterization of such residual structural elements.

Calpastatin, the inhibitor of calpain, is a prime example for these foregoing considerations. It is a fully unstructured protein (*16, 18, 19*) with short recognition segments, known as subdomains A, B, and C, that are primarily responsible for the binding and effective inhibition of calpain (*4*). Subdomains A and C adopt helical structures in the bound form (*21*) and display a significant helical tendency when studied as isolated peptides (*22*), while subdomain B incorporates a type I β-turn substructure (*23*). We therefore hypothesized that in the free state of full-length inhibitor I, each subdomain retains some of these structural features to guarantee extremely fast inhibition of the enzyme. To trace residual structural features of full-length human calpastatin domain 1 (hCSD1), suitable triple-resonance NMR experiments on ^{15}N-labeled and on doubly ^{15}N- and ^{13}C-labeled proteins were completed. We have managed a nearly complete assignment and analysis of backbone resonances (121 of 126 non-prolines), which enabled a thorough structural characterization of calpastatin in the solution state.

Deviation of SCSs from typical random coil values suggests some transient secondary structural preferences

within regions noted for evolutionary conservation and interaction with calpain. Regions of helical propensity by SCS correspond to noted functional elements of CST: $D^{18}-I^{25}$ region to subdomain A ($S^{12}-G^{30}$), the consecutive $S^{51}-G^{59}$ and $K^{68}-K^{75}$ sequences in subdomain B ($M^{50}-M^{70}$), and the fourth sequential unit, $G^{91}-T^{104}$, within subdomain C ($S^{87}-C^{105}$). On the basis of previous work and data reported here, the $T^{64}-R^{70}$ region is suggested to adopt a β-turn structure. This structural feature could also be ascertained by comparing secondary chemical shift values of full-length human calpastatin domain I (hCSD1) and its C-terminal half, hCSD1(67−141), extending to the middle of inhibitory subdomain B. Combined N^{HN} and H^{HN} chemical shift differences within the novel N-terminal dodecapeptide generated by the truncation are significantly different from zero, which suggests that this region in full-length calpastatin must have some residual structure. A further intriguing and unexpected feature of calpastatin is a significant helical propensity established for the C-terminal region, $E^{118}-T^{131}$, which has not been implicated in calpain inhibition.

Analysis of relaxation parameters also suggests a functionally significant deviation from a fully random behavior of selected sequence units. Whereas the segments corresponding to subdomains A and B and perhaps the C-terminal half of subdomain C have a lower-than-average flexibility, the linker regions between them exhibit higher local plasticity. Calpastatin from this perspective appears to behave as "beads on the string", in which regions of local transient order are connected by highly flexible linkers. This mode of structural organization can be rationalized in terms of calpain structure and the likely mode of interaction of the inhibitor with the enzyme (*21, 44*). The structure of calpain is best approximated as an extended oval, and key interactions with calpastatin subdomains A and C occur at the CaM-like domains of the enzyme but with subdomain B at the catalytic site on the opposite end of the enzyme (*21*). This mode of binding assumes an almost fully extended main chain structure of calpastatin, which has to wrap around and contact the enzyme at three distinct regions. Flexibility of the linker regions described above is most likely instrumental in enabling this mode of binding. The difference between subdomains A and C in this regard is also worth noting.

Overall, this qualitative analysis provide some insight into the motional properties of calpastatin in action, in conjunction with previous suggestions that IUPs use short, transiently preformed recognition elements connected by more flexible linkers for effective partner recognition. An additional issue is the role of the C-terminal region, $E^{118}-T^{131}$, which also shows strong signs of transient local structure and decreased flexibility, reminiscent of the behavior of subdomains A and C. This region is less conserved among eukaryotic species than subdomains A and C, but the level of sequence identity is still 60%, which is markedly high, which makes it suitable for preservation of the secondary structure characteristics. Additionally, this region is not wholly represented among calpastatin domains (Figure 5); hence, without functional studies, it cannot be marked as an unexceptionable subdomain. One possible scenario would implicate this region as a recognition element akin to subdomains A and C. This, however, is not easily reconciled with current structural data, which leaves hardly any space for this region to bind to calpain. Furthermore, subdomains A and C are likely to act as amphipathic helices, which use their conserved hydrophobic side for calpain binding (*21*), suggested to be a rather general feature of IUP MoREs (*15*). Amino acid residues conserved in $E^{118}-T^{131}$, however, are typically hydrophilic and/or charged. Another scenario we deem more likely is that in which this conserved region acts as a spacer between adjacent inhibitory domains. Typical calpastatins have four inhibitory domains, each of which is capable of inhibiting a calpain on its own (*4*). This, however, raises serious steric problems, as the full inhibitor (708 amino acids) has to wrap around four separate calpain molecules for successful interaction and inhibition. A relatively rigid spacer between the inhibitory units is definitely of use in this complicated molecular adaptation process. Such a mode of action has been described in the case of another well-studied IUP, p27 KID, for which specific staplelike recognition by two rather disordered terminal regions is facilitated by a rigid, helical middle segment (*26, 45*). Whether this scenario also applies to calpastatin inhibition remains to be seen, but due to the structural and evolutionary considerations and the appealing analogy with another IUP, we suggest considering the $E^{118}-T^{131}$ region of calpastatin a new separate structural motif.

ACKNOWLEDGMENT

We thank Peter Rüdinger and Peter Dvortsak for their kind support and scientific advice. Special thanks for MS spectra to Ildikó Szabó and Katalin B. Bai.

SUPPORTING INFORMATION AVAILABLE

Hetero-NOE data as well as gel and MS spectra proving sample purity. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Goll, D. E., Thompson, V. F., Li, H. Q., Wei, W., and Cong, J. Y. (2003) The calpain system. *Physiol. Rev. 83*, 731–801.
2. Wendt, A., Thompson, V. F., and Goll, D. E. (2004) Interaction of calpastatin with calpain: A review. *Biol. Chem. 385*, 465–472.
3. Farkas, A., Tompa, P., and Friedrich, P. (2003) Revisiting ubiquity and tissue specificity of human calpains. *Biol. Chem. 384*, 945–949.
4. Emori, Y., Kawasaki, H., Imajoh, S., Minami, Y., and Suzuki, K. (1988) All 4 repeating domains of the endogenous inhibitor for calcium-dependent protease independently retain inhibitory activity: Expression of the CDNA fragments in *Escherichia coli*. *J. Biol. Chem. 263*, 2364–2370.
5. Ma, H., Yang, H. Q., Takano, E., Lee, W. J., Hatanaka, M., and Maki, M. (1993) Requirement of different subdomains of calpastatin for calpain inhibition and for binding to calmodulin-like domains. *J. Biochem. 113*, 591–599.
6. Ma, H., Yang, H. Q., Takano, E., Hatanaka, M., and Maki, M. (1994) Amino-terminal conserved region in proteinase inhibitor domain of calpastatin potentiates its calpain inhibitory activity by interacting with calmodulin-like domain of the proteinase. *J. Biol. Chem. 269*, 24430–24436.
7. Takano, E., Ma, H., Yang, H. Q., Maki, M., and Hatanaka, M. (1995) Preference of calcium-dependent interactions between calmodulin-like domains of calpain and calpastatin subdomains. *FEBS Lett. 362*, 93–97.
8. Tompa, P., Mucsi, Z., Orosz, G., and Friedrich, P. (2002) Calpastatin subdomains A and C are activators of calpain. *J. Biol. Chem. 277*, 9022–9026.
9. Branca, D. (2004) Calpain-related diseases. *Biochem. Biophys. Res. Commun. 322*, 1098–1104.
10. Zatz, M., and Starling, A. (2005) Mechanisms of disease: Calpains and disease. *N. Engl. J. Med. 352*, 2413–2423.

11. Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci. 27*, 527–533.
12. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry 41*, 6573–6582.
13. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol. 6*, 197–208.
14. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol. 337*, 635–645.
15. Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry 44*, 1989–2000.
16. Csizmok, V., Bokor, M., Banki, P., Klement, T., Medzihradszky, K. F., Friedrich, P., Tompa, K. A., and Tompa, P. (2005) Primary contact sites in intrinsically unstructured proteins: The case of calpastatin and microtubule-associated protein 2. *Biochemistry 44*, 3955–3964.
17. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol. 293*, 321–331.
18. Uemori, T., Shimojo, T., Asada, K., Asano, T., Kimizuka, F., Kato, I., Maki, M., Hatanaka, M., Murachi, T., Hanzawa, H., and Arata, Y. (1990) Characterisation of a functional domain of human calpastatin. *Biochem. Biophys. Res. Commun. 166*, 1485–1493.
19. Konno, T., Tanaka, N., Kataoka, M., Takano, E., and Maki, M. (1997) A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. *Biochim. Biophys. Acta 1342*, 73–82.
20. Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol. 338*, 1015–1026.
21. Todd, B., Moore, D., Deivanayagam, C. C. S., Lin, G. D., Chattopadhyay, D., Maki, M., Wang, K. K. W., and Narayana, S. V. L. (2003) A structural model for the inhibition of calpain by calpastatin: Crystal structures of the native domain VI of calpain and its complexes with calpastatin peptide and a small molecule inhibitor. *J. Mol. Biol. 328*, 131–146.
22. Mucsi, Z., Hudecz, F., Hollosi, M., Tompa, P., and Friedrich, P. (2003) Binding-induced folding transitions in calpastatin subdomains A and C. *Protein Sci. 12*, 2327–2336.
23. Ishima, R., Tamura, A., Akasaka, K., Hamaguchi, K., Makino, K., Murachi, T., Hatanaka, M., and Maki, M. (1991) Structure of the active 27-residue fragment of human calpastatin. *FEBS Lett. 294*, 64–66.
24. Betts, R., Weinsheimer, S., Blouse, G. E., and Anagli, J. (2003) Structural determinants of the calpain inhibitory activity of calpastatin peptide B27-WT. *J. Biol. Chem. 278*, 7800–7809.
25. Radhakrishnan, I., Perez-Alvarado, G. C., Dyson, H. J., and Wright, P. E. (1998) Conformational preferences in the Ser(133)-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett. 430*, 317–322.
26. Sivakolundu, S. G., Bashford, D., and Kriwacki, R. W. (2005) Disordered p27(Kip1) exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J. Mol. Biol. 353*, 1118–1128.
27. Penkett, C. J., Redfield, C., Jones, J. A., Dodd, I., Hubbard, J., Smith, R. A. G., Smith, L. J., and Dobson, C. M. (1998) Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from *Staphylococcus aureus*. *Biochemistry 37*, 17054–17067.
28. Eliezer, D., Kutluay, E., Bussell, R., and Browne, G. (2001) Conformational properties of α-synuclein in its free and lipid-associated states. *J. Mol. Biol. 307*, 1061–1073.
29. Vise, P. D., Baral, B., Latos, A. J., and Daughdrill, G. W. (2005) NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic Acids Res. 33*, 2061–2077.
30. Bahar, I., and Jernigan, R. L. (1998) Vibrational dynamics of transfer RNAs: Comparison of the free and synthetase-bound forms. *J. Mol. Biol. 281*, 871–884.
31. Barre, P., and Eliezer, D. (2006) Folding of the repeat domain of tau upon binding to lipid surfaces. *J. Mol. Biol. 362*, 312–326.
32. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPIPE: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.
33. Kneller, D. G., and Kuntz, I. D. (1993) UCSF SPARKY: An NMR display annotation and assignment tool. *J. Cell. Biochem.,* 254–254.
34. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) H-1, C-13 and N-15 random coil NMR chemical-shifts of the common amino-acids. 1. Investigation of nearest-neighbor effects. *J. Biomol. NMR 5*, 67–81.
35. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear-magnetic-resonance chemical-shift and protein secondary structure. *J. Mol. Biol. 222*, 311–333.
36. Weisemann, R., Ruterjans, H., Schwalbe, H., Schleucher, J., Bermel, W., and Griesinger, C. (1994) Determination of H(N),H-α and H(N), C′ coupling constants in C-13, N-15 labeled proteins. *J. Biomol. NMR 4*, 231–240.
37. Lefevre, J. F., Dayie, K. T., Peng, J. W., and Wagner, G. (1996) Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry 35*, 2674–2686.
38. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol. 12*, 54–60.
39. Wishart, D. S., and Sykes, B. D. (1994) Chemical shift as a tool for structure determination. *Methods Enzymol. 239*, 363–392.
40. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., and Wright, P. E. (1988) Folding of immunogenic peptide-fragments of proteins in water solutions. 1. sequence requirements for the formations of a reverse turn. *J. Mol. Biol. 201*, 161–200.
41. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL-W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res. 22*, 4673–4680.
42. Lee, A. L., and Wand, A. J. (2001) Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature 411*, 501–504.
43. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydrophatic character of a protein. *J. Mol. Biol. 157*, 105–132.
44. Strobl, S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa, K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H., Suzuki, K., and Bode, W. (2000) The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *Proc. Natl. Acad. Sci. U.S.A. 97*, 588–592.
45. Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L. M., Weiss, S., Hengst, L., and Kriwacki, R. W. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol. 11*, 358–364.

BI800201A