Primary Contact Sites in Intrinsically Unstructured Proteins: The Case of Calpastatin and Microtubule-Associated Protein 2†

Veronika Csizmók,‡ Mónika Bokor,§ Péter Bánki,§ Éva Klement,|| Katalin F. Medzihradszky,||,⊥ Peter Friedrich,‡ Kálmán Tompa,§ and Peter Tompa*,‡

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Post Office Box 7, H-1518 Budapest, Hungary, Research Institute for Solid State Physics and Optics, Hungarian Academy of Sciences, Post Office Box 49, H-1525 Budapest, Hungary, Mass Spectrometry Facility at the Biological Research Center, Hungarian Academy of Sciences, Post Office Box 521, H-6701 Szeged, Hungary, and Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California 94143−0446

Received October 12, 2004; Revised Manuscript Received January 4, 2005

ABSTRACT: Intrinsically unstructured proteins (IUPs) exist in a disordered conformational state, often considered to be equivalent with the random-coil structure. We challenge this simplifying view by limited proteolysis, circular dichroism (CD) spectroscopy, and solid-state ^{1}H NMR, to show short- and long-range structural organization in two IUPs, the first inhibitory domain of calpastatin (CSD1) and microtubule-associated protein 2c (MAP2c). Proteases of either narrow (trypsin, chymotrypsin, and plasmin) or broad (subtilisin and proteinase K) substrate specificity, applied at very low concentrations, preferentially cleaved both proteins in regions, i.e., subdomains A, B, and C in CSD1 and the proline-rich region (PRR) in MAP2c, that are destined to form contacts with their targets. For CSD1, nonadditivity of the CD spectra of its two halves and suboptimal hydration of the full-length protein measured by solid-state NMR demonstrate that long-range tertiary interactions provide the structural background of this structural feature. In MAP2c, such tertiary interactions are absent, which points to the importance of local structural constraints. In fact, urea and temperature dependence of the CD spectrum of its PRR reveals the presence of the extended and rather stiff polyproline II helix conformation that keeps the interaction site exposed. These data suggest that functionally significant residual structure exists in both of these IUPs. This structure, manifest as either transient local and/or global organization, ensures the spatial exposure of short contact segments on the surface. Pertinent data from other IUPs suggest that the presence of such recognition motifs may be a general feature of disordered proteins. To emphasize the possible importance of this structural trait, we propose that these motifs be called primary contact sites in IUPs.

Our view of proteins is dominated by the notion that a well-defined three-dimensional structure is a prerequisite for their function. In a recent surge of papers, however, the generality of this paradigm has been challenged by showing that a lot of proteins and protein domains, termed intrinsically unstructured (*1*), intrinsically disordered (*2*), natively unfolded (*3*), or pliable (*4*), function without a well-defined folded structure. It appears that such proteins/domains are common in living organisms, corresponding to about 20−30% of all of the residues found in various proteomes (*2*, *5*−*7*).

Physicochemical data indicate that these intrinsically unstructured proteins (IUPs)[1] have a noncompact and highly flexible structure, which resembles the denatured state(s) of globular proteins. Their structure may be approximated by a random coil (*8*, *9*); notwithstanding, a true random coil practically never exists (*10*), and a fully featureless state is hardly compatible with the highly effective functioning of IUPs (*1*, *2*, *4*). In fact, pertinent data suggest residual structure, such as permanent or transient secondary structural elements and/or long-range interactions, in many IUPs (reviewed in refs *3* and *11*). Because the rapid and specific binding of these large IUPs to their partners and the coupled

† This work was supported by Grants T 29059, T 32360, T 34255, TS 040723, and T 037196 (the latter for K.F.M. and E.K.) from the Hungarian Scientific Research Fund, FKFP 0100/2000 and NKFP 1/010 from the Ministry of Education, and ISRF GR067595 from the Wellcome Trust (for P.T.). P.T. and M.B. acknowledge the support of a Bolyai János Scholarship.

* To whom correspondence should be addressed: Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, P.O. Box 7, H-1518 Budapest, Hungary. Telephone: (361) 279-3143. Fax: (361) 466-5465. E-mail: tompa@enzim.hu.

‡ Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences.

§ Research Institute for Solid State Physics and Optics, Hungarian Academy of Sciences.

|| Mass Spectrometry Facility at the Biological Research Center, Hungarian Academy of Sciences.

⊥ University of California San Francisco.

[1] Abbreviations: BSA, bovine serum albumin; CBB-R250, Coomassie Brilliant Blue R250; CD, circular dichroism; CBP, CREB-binding protein; CREB, cyclic AMP response element binding protein; EDTA, ethylenediaminetetraacetic acid; CST, calpastatin; FID, free-induction decay; CSD1, human CST domain 1; IUP, intrinsically unstructured protein; MAP2(c), microtubule-associated protein 2(c); MS, mass spectrometry; NPC, nuclear pore complex; Nup, nucleoporin; PCS, primary contact site; PPII, polyproline II helix; PRR, proline-rich region; PVDF, polyvinilidene difluoride membrane; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

10.1021/bi047817f CCC: $30.25
Published on Web 02/18/2005

folding toward their final state are related to the classical Levinthal paradox of protein folding (*12*), such a residual structure may play a crucial mechanistic role in their function. In short, their binding and induced folding would take too long if the IUP had to search through all of the conformational space from a fully unstructured initial conformational state. *A priori*, this problem can be overcome if certain amount of structural order within the IUP ensured exposure of dedicated region(s) for effective initial interaction (*3*, *13*) and/or the conformation preferentially sampled by the IUP would presage the final structure attained upon binding to its partner. Because this latter mechanism has already been shown to be of general importance (*13*), here, we look for the presence of exposed regions serving as potential initial interaction sites.

In essence, the extension of the structure−function paradigm to IUPs(*1*) requires detailed characterization of transient conformational elements in IUPs to find their relation(s) with function. Best suited for this purpose is the application of high-resolution multidimensional NMR, which provides sequence-specific structural and dynamic information (*14*, *15*). When the size of the protein precludes this approach, however, further physicochemical techniques may yield relevant information. In this work, we have undertaken to compare the unbound structure of calpastatin (CSD1) and microtubule-associated protein 2c (MAP2c), two functionally well-characterized IUPs (*3*, *4*). Calpastatin (CST), the 76.5-kDa specific inhibitor of calpain, is composed of four equivalent inhibitory domains, each capable of inhibiting the enzyme on its own (*16*−*18*). It functions by rapid and specific inhibition of calpain by interaction via three binding regions, subdomains A, B, and C. Of these three, subdomain B is the inhibitory region, whereas subdomains A and C potentiate binding in a Ca^{2+}-dependent manner (*19*). MAP2 is a 202-kDa neuron-specific accessory protein, which regulates microtubule assembly and spacing (*20*, *21*). Its C-terminal microtubule-binding region consists of three or four 31 amino acid repeats, flanked by a proline-rich region (PRR) and a fifth, partial repeat (*22*, *23*). The N-terminal region of the molecule projects away from the surface of microtubules and serves as a spacer between adjacent microtubule fibers (*24*, *25*). MAP2c, the 49-kDa juvenile splice variant, has essentially the same domain organization and function. The application of limited proteolysis at very low protease concentrations, combined with circular dichroism (CD) spectroscopy and solid-state NMR relaxation studies, reveals that both proteins expose short segments critical for their interaction with their partners. A compilation of related data for several other IUPs underlines the possible generality of this phenomenon. In light of the foregoing considerations, we suggest that such preferentially accessible interaction regions, ensured by the residual structure, serve as primary contact sites (PCSs) and folding centers in the effective functioning of many IUPs.

MATERIALS AND METHODS

DNA Constructs. The full-length clones of human CST domain 1 (CSD1, Ala^{137}-Lys^{277} of CST) and rat MAP2c were kindly provided by Prof. M. Maki from Nagoya University (Nagoya, Japan) and Prof. A. Matus from the Friedrich Miescher Institut (Basel, Switzerland). Clones encoding for shorter fragments were generated by PCR from these full-length clones by high-fidelity Pfu polymerase. The recombinant halves have been separated within functionally important regions, i.e., subdomain B of CSD1 and PRR of MAP2c (cf. Figure 4). In the case of CSD1, the oligonucleotide primers used were TAGGCTGCATATGGCTGTGCCAGTTGAATCTAAACC and CTGCTCGAGTCAAGGAATTGTGACTTCTCTTTTACC for the N-terminal half (Ala^{137}-Pro^{202}) and TCGGCCGCATATGCCAAAATATAGGGAACTATTGG and ATGCTCGAGTCACTTCTCTTGGGGTGGA for the C-terminal half (Pro^{203}-Lys^{277}). In the case of MAP2c, TAGGCTGCATATGGCTGACGAGAGGAAAGAC and ATTCTCGAGTCATCCAGGGGTGCCTGGGGTAC were used for the N-terminal half (Met^{1}-Gly^{261}) and TATGCAGCATATGACCCCGAGCTATCCCAGG and ATGCTCGAGTCACAAGCCCTGCTTAGCGAGCG for the C-terminal half (Thr^{262}-Leu^{467}). The PRR of MAP2c (Pro^{186}-Pro^{297}) has been generated by the primers TAGGCTGCATATGCCAAGACCTTCCTCCATCC and ATTCTCGAGGGGAGTAGCTGGGGACTTTGG. The cDNA products obtained by PCR were cloned into the *Nde*I−*Xho*I sites of the expression vector pET22b (Novagene).

Protein Purification. Full-length human CSD1 and its two halves were expressed in the *E. coli* strain BL21 and purified on a DEAE-cellulose anion-exchange column as described for CSD1 (*26*). Full-length rat MAP2c and its C-terminal half were prepared according to ref *27*; for purifying the N-terminal half, the procedure was slightly modified by replacing the original S-Sepharose column with DEAE-cellulose because the isoelectric point of this half differs considerably from that of the other. The PRR of MAP2c was prepared as described for MAP2c and purified on a Ni-NTA agarose column. The proteins were dialyzed and stored frozen in 20 mM Tris (pH 7.5), 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 5 mM mercaptoethanol. To concentrate CSD1 and MAP2c for the NMR experiments, the original preparations were dialyzed into distilled water, 1 mM benzamidine, and 0.1 mM PMSF and dried down by lyophilization.

Limited Proteolysis. Limited proteolysis of MAP2c and CSD1 was carried out with proteases of both narrow and broad specificity at concentrations much lower than usually applied with globular substrates (at enzyme/substrate ratios of 1:800−1:5000, cf. Figure 1), to adapt to the extreme proteolytic sensitivity of IUPs. The five different enzymes, i.e., trypsin, chymotrypsin, plasmin, subtilisin, and proteinase K, were dissolved in a buffer of 20 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, and 5 mM mercaptoethanol. Digestion was allowed at 25 °C in the same buffer for the time indicated. The reaction was stopped by adding sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) sample buffer and boiling for 3 min.

Protein Sequencing and Mass Spectrometry (MS). The primary site of cleavage of CSD1 has been determined by N-terminal sequencing of the primary fragments generated. The digests were run on SDS−PAGE and electroblotted onto polyvinilidene difluoride membranes (PVDFs), and selected bands were subjected to five rounds of Edman degradation at the Agricultural Biotechnology Center, Gödöllö, Hungary. With MAP2c, a different strategy had to be followed because of difficulty experienced in blotting the fragments on the C-terminal side of the primary cleavage site. To this end, MALDI−TOF analysis of the digests was performed on a

Reflex III mass spectrometer (Bruker) in linear mode. Sinapinic acid (Agilent Technologies) was used as the matrix. Masses were measured in seven independent experiments, and the average is listed in the paper. When abundant enough, the singly and doubly charged ions of the full-length protein were used for internal calibration. Otherwise, horse heart apomyoglobin and bovine serum albumin (BSA) provided a two-point external calibration. ProteinProspector (http://prospector.ucsf.edu) was used to identify the potential cleavage products.

CD Measurements. CD spectra were recorded at 0.1 mg/mL protein concentration in 10 mM Na_2HPO_4 (pH 7.5) and 150 mM NaCl in a 1-mm path-length cuvette on a Jasco J-720 spectropolarimeter in continuous mode with a 1 nm bandwidth, 8 s response time, and 20 nm/min scan speed. All spectra were obtained by subtracting buffer spectra and represent the average of nine separate scans. Spectra are presented in molar units to make constructs of various length directly comparable. The temperature was either maintained at 25 °C or scanned from 10 to 80 °C with a Neslab RTE-111 circulating water bath.

^{1}H NMR Measurements. Solid-state NMR measurements were carried out on a BRUKER SXP 4-100 pulsed spectrometer at 84.2 MHz frequency and 50 mg/mL concentration of full-length CSD1, MAP2c, BSA, and the recombinant halves of CSD1 and MAP2c, basically as described in ref *28*. The protein powders, including BSA (Sigma, fraction V), were dissolved in 20 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, and 5 mM mercaptoethanol to the desired final concentration and gave a stable solution after a 10 min 10000*g* centrifugation. The measured free induction decay (FID) amplitude was normalized to the sample mass, protein and buffer concentrations, 90° pulse length, and temperature dependence of nuclear susceptibility. The time scales of the measurement were 4 μs for the 90° pulse length, 10 μs for the dead time of the spectrometer, a few hundred microseconds for the "bound" length, and about 20 ms for the "free" water FID length, respectively. Cooled nitrogen gas system controlled the sample temperature. The speed of the temperature change has excluded the overcooling of samples. It is to be noted that the phases of ice protons, protein protons, and (unfrozen) water protons are clearly separated in the FID signal by virtue of large differences in the spin−spin relaxation rate. This enables specific recording of FIDs that belong to the hydrate layer of proteins, which makes the absolute determination of the hydrate layer possible.

Other Methods. SDS−PAGE was run according to Laemmli (*29*). PCR products were purified from agarose gels with the Nucleo-Spin extract kit (Macherey−Nagel). Purified DNA samples were sequenced by MWG-AG Biotech (Ebersberg, Germany). The mean hydrophobicity of proteins was determined via the ProtScale server available on the Internet (http://us.expasy.org/cgi-bin/protscale), by using the Kyte and Doolittle hydrophobicity parameter set for amino acids.

Materials. Proteases and all other chemicals were purchased from Sigma Chemical Co. Buffers were made in Millipore MilliQ water.

RESULTS

Limited Proteolysis. One hallmark of IUPs is their extreme sensitivity to proteolysis (*4*), which results in their rapid and complete degradation by proteases even under the usual conditions of limited proteolysis. We reasoned, nevertheless, that if the structure of an IUP is not fully random, proteases at extremely low concentrations could be used to probe their accessible sites by analyzing the very first cleavage event that occurs. Evidently, even under these conditions, proteases would attack all sequentially favored sites with time, because exposure/buriedness of sites is only transient because of the rapid fluctuations of these proteins. However, if there are structured parts in the otherwise random-coil polypeptide, heralding association with *in vivo* partners, these will expose sensitive sites for the test proteases, which would then dominate the degradation pattern. The protease can thus be used to probe the incipient structure, with particular sensitivity to sites exposed for partner interaction. In accordance with this expectation, proteolysis of CSD1 and MAP2c under such conditions results in a picture typical of limited proteolysis, featuring the formation of only a few, relatively stable, fragments.

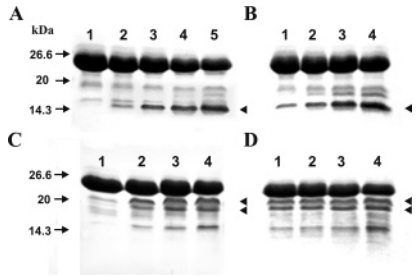

FIGURE 1: Limited proteolysis of CSD1. CSD1 at 2.5 mg/mL was digested by proteases of narrow (trypsin, chymotrypsin, and plasmin; not shown) and broad (subtilisin and proteinase K) substrate specificity at the enzyme/substrate ratios (w/w) specified. Digestion was stopped at the times indicated and visualized by SDS−PAGE and Coomassie Brilliant Blue R250 (CBB-R250) staining. Fragments marked by arrowheads on the right were analyzed by N-terminal sequencing. (A) Trypsin (1:2000), digestion for 0 s, 10 s, 30 s, and 1 and 2 min, respectively. (B) Chymotrypsin (1:800), digestion for 10 s, 30 s, and 1 and 2 min, respectively. (C) Subtilisin (1:5000), digestion for 0 s, 10 s, 30 s, and 2 min, respectively. (D) Proteinase K (1:800), digestion for 10 s, 30 s, and 1 and 2 min, respectively.

Brief digestion of CSD1 with two narrow-specificity enzymes generates only one prominent band, at 14.5 kDa with trypsin (Figure 1A) and at 16 kDa with chymotrypsin (Figure 1B). Although certain other fragments are also apparent on the gel, these are either already present in the sample prior to digestion (e.g., at 19 kDa with trypsin) or appear after the band selected, i.e., result from cleavage at a site that is less exposed (e.g., at 18 kDa with chymotrypsin). In a similar manner, limited proteolysis with plasmin, that has the same substrate specificity as trypsin, results in two major products at 23 and 10 kDa (data not shown). Digestion of CSD1 with two enzymes of broad substrate specificity, i.e., subtilisin (Figure 1C) and proteinase K (Figure 1D), yields products at approximately 20 and 18 kDa, which have the same N termini because of a further cut close to the C-terminal end. An additional, slowly appearing, 15-kDa fragment probably results from a secondary cleavage site between the sites of trypsin and chymotrypsin. The sites of initial cleavage for all five enzymes, determined by N-terminal sequencing (trypsin, $R^{197}E$; chymotrypsin, $Y^{190}I$; plasmin, $K^{150}S$ and $K^{224}P$; and subtilisin and proteinase K, $L^{164}G$), cluster within the functionally important subdomains

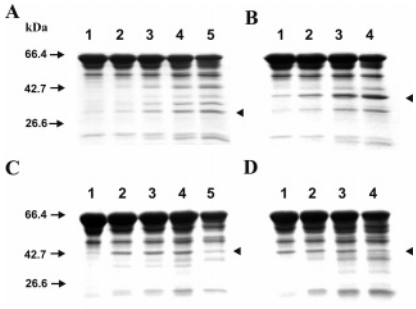

FIGURE 2: Limited proteolysis of MAP2c. MAP2c at 1 mg/mL was digested as CSD1 in Figure 1 and analyzed by MALDI−TOF MS. The primary site of cleavage was identified by looking for cleavage products corresponding in mass to those marked by arrowheads. (A) Trypsin (1:2000), digestion for 0 s, 10 s, 30 s, and 1 and 2 min, respectively. (B) Chymotrypsin (1:4000), digestion for 10 s, 30 s, and 1 and 2 min, respectively. (C) Subtilisin (1:5000), digestion for 0 s, 10 s, 30 s, and 1 and 2 min, respectively. (D) Proteinase K (1:2000), digestion for 10 s, 30 s, and 1 and 2 min, respectively.

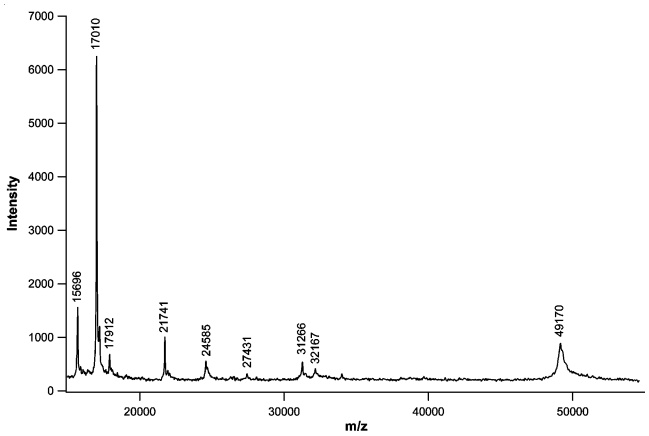

FIGURE 3: MALDI−TOF mass spectrum of the Map2c 10 s chymotryptic digest. Internal calibration was performed on the singly and doubly charged ions of the full-length protein at *m*/*z* values of 49170 and 24585, respectively. The peak at *m*/*z* 15696 was present in the undigested sample as well.

A, B, and C, as shown in Figure 4A. The structural significance of this nonrandom cleavage is underlined by two considerations. First, there are numerous evenly distributed theoretical cleavage sites for the narrow specificity enzymes (16 for trypsin and 12 for chymotrypsin), and even more for the broad specificity enzymes. Because the sensitivity to proteases is not exclusively determined by the P_1 site of the substrate, secondary interactions might distinguish between these sites causing the observed cleavage patterns. When these preferences are taken into consideration, however, the sites actually cleaved are sequentially rather disfavored (*30*), which leaves the only explanation that preferential cleavage ensues from structural exposure of the given sites. Second, the application of completely unrelated proteases also argues against uneven cleavage simply coming from the sequential preference of proteases, because there is no reason to expect that favorable sequences for different proteases are located near one another.

Limited proteolysis of MAP2c (Figure 2) also results in a nonrandom cleavage, demonstrating different accessibility of its large number of potentially vulnerable sites. The unfractionated digests were analyzed by MALDI−TOF MS to determine masses of cleavage products (cf. Figure 3 for

Table 1: Masses Determined by MALDI−TOF MS of MAP2c Fragments Obtained by Limited Proteolysis

$mass_{detected}$	assignment	$mass_{calculated}$	deviation (%)
	Tryptic Digest		
17 753 ± 3[a]	[^{302}Leu -Leu^{467}][b]	17 755	−0.014
19 011 ± 3	[^{2}Ala −Arg^{182}]	19 019	−0.040
22 732 ± 3	[^{2}Ala-Arg^{217}]	22 745	−0.056
22 889 ± 4	[^{2}Ala −Arg^{218}]	22 901	−0.052
23 245 ± 4	[^{2}Ala −Arg^{221}]	23 259[c]	−0.062
34 104 ± 4	[^{2}Ala −Arg^{301}]	31 434	−0.094
	Chymotryptic Digest		
17 007 ± 10	[^{2}Ala-Phe^{164}][d]	17 007	−0.002
17 911 ± 9	[^{301}Arg-Leu^{467}]	17 912	−0.003
21 738 ± 9	[^{2}Ala-Phe^{207}]	21 742[c]	−0.017
27 434 ± 7	[^{208}Ser-Leu^{467}]	27 447[c]	−0.049
31 264 ± 10	[^{2}Ala-Leu^{300}]	31 277	−0.043
32 165 ± 16	[^{165}Lys-Leu^{467}]	32 182	−0.052
	Subtilisin Digest		
22 043 ± 4	[^{2}Ala-Asn^{210}]	22 056	−0.059
22 417 ± 7	[^{2}Ala-Ser^{214}]	22 430	−0.06
	Proteinase K Digest		
22 035 ± 5	[^{2}Ala-Asn^{210}][e]	22 056	−0.095
22 213 ± 5	[^{2}Ala-Ser^{212}]	22 230	−0.077
22 410 ± 8	[^{2}Ala-Ser^{214}]	22 430	−0.091
22 561 ± 9	[^{2}Ala-Ala^{216}]	22 589	−0.12
22 724 ± 6	[^{2}Ala-Arg^{217}]	22 745	−0.091
31 253 ± 4	[^{2}Ala-Leu^{300}]	31 277[c]	−0.078
35 361 ± 6	[^{2}Ala-Lys^{337}]	35 379	−0.051

[a] Masses determined by MALDI−TOF MS are the average of seven measurements. [b] The specificity of trypsin was considered. [c] Fragments corresponding to the primary cleavage event, identified as given in the Results. [d] The specificity of chymotrypsin was considered. [e] Cleavage sites were assigned based on ladder sequences observed in the digest.

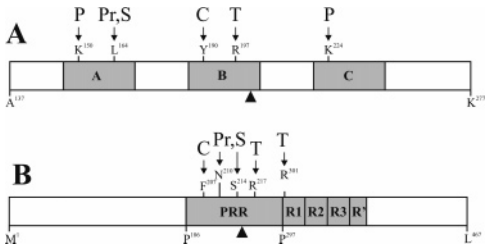

FIGURE 4: Schematic representation of the domain structure of CSD1 and MAP2c and the location of initial proteolytic cleavage sites. (A) Scheme of CSD1, with subdomains A, B, and C shown in gray and the cleavage sites of the five different enzymes, plasmin (P), proteinase K (Pr), subtilisin (S), chymotrypsin (C), and trypsin (T) indicated by arrows. The residue N-terminal to the scissile bond (P_1) is given for each site. (B) Scheme of rat MAP2c with the microtubule-binding region (PRR, the imperfect repeats, R1−R3, and a partial repeat, R′) shown in gray and cleavage sites of four enzymes indicated by arrows. The primary cleavage of subtilisin was taken to be identical to that of proteinase K, because the two enzymes produce the same cleavage pattern (also with CSD1). For the two IUPs, the site separating the recombinant halves is indicated by large arrowheads below.

chymotrypsin). Of all of the fragments observed by MS (Table 1), those corresponding to the initial cleavage events have been identified by looking for a match in mass with the band(s) marked in SDS−PAGE (Figure 2), taking into account the anomalously slow run of IUPs on SDS gels (*4*). Initial cleavage by all of the proteases (trypsin, $R^{217}R$ and $R^{301}L$; chymotrypsin, $F^{207}S$; and subtilisin and proteinase K, $N^{210}S$ and $S^{214}S$) occurs within a short segment of MAP2c, the PRR, and the adjacent R1 repeat (Figure 4B). The structural and possibly functional significance of this finding

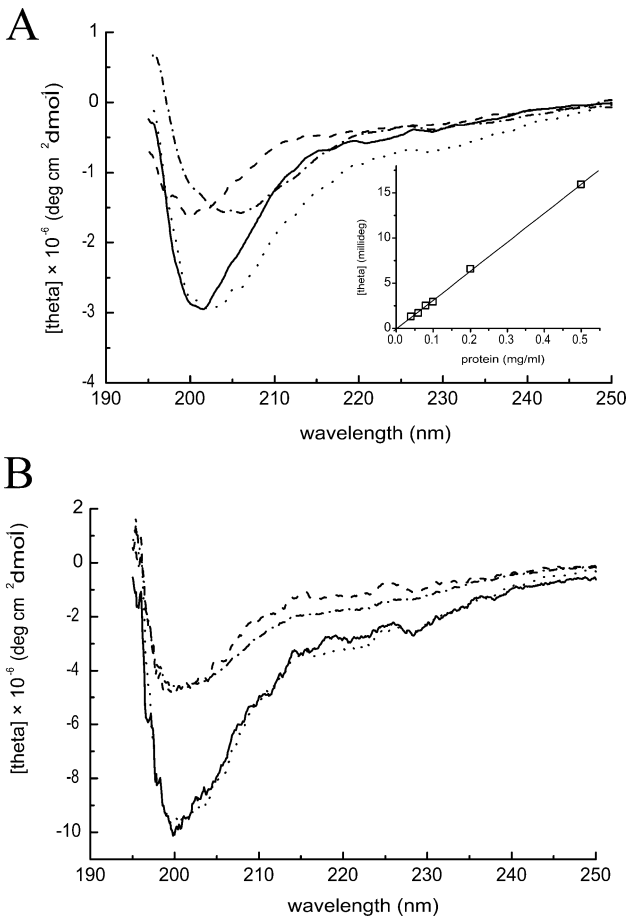

FIGURE 5: Far-UV CD spectra of CSD1, MAP2c, and their recombinant halves. (A) Molar ellipticity of CSD1 (—), its N-terminal half (- - -), C-terminal half (-•-), and the sum of the spectra of the N- and C-terminal halves (•••). The CD spectrum of the C-terminal half was also recorded at several concentrations and the absolute value of ellipticity values at 208 nm is plotted as a function of the protein concentration (inset). (B) Molar ellipticity of MAP2c (—), its N-terminal half (- - -), C-terminal half (-•-), and the sum of the two halves (•••). The position of the two halves within the sequences is shown in Figure 4.

is underlined by the fact that there are an even greater number of theoretical sites than for CSD1 above (63 for trypsin and 40 for chymotrypsin). Further, the limited proteolysis of MAP2, i.e., its 202-kDa adult form, also occurs within the same narrow region (*31*, *32*) confined to a segment of only 7% of the molecule. The amply characterized homologue of MAP2, tau protein, is also cleaved within the same region under similar conditions of limited proteolysis (*33*, *34*).

CD Analysis. To explore the structural background of this nonrandom cleavage, the CD spectra of various fragments of both CSD1 and MAP2c have been recorded and compared to that of the full-length proteins. Recombinant PRR of MAP2c and the two halves of both proteins separated within the middle of their exposed regions, i.e., subdomain B in CSD1 and PRR in MAP2c (cf. Figure 4), have been generated for this purpose.

For CSD1, the CD spectrum of the full-length protein has a minimum at 201 nm, characteristic of a protein in a largely disordered conformation (Figure 5A). The spectra of the two halves of the protein, however, are different, with the minimum of the C-terminal half shifted to about 205 nm. As a result of this difference, the spectra of the two halves are not additive (Figure 5A), which indicates their transient interaction within the CST molecule. Taken at face value, the larger negative ellipticity at wavelengths above 206 nm for the sum of spectra of the two halves are suggestive of more "structure" in the separate fragments, primarily transient β-type interactions in the C-terminal half. When its spectrum is recorded at several protein concentrations covering a wide range (inset of Figure 5A), it is demonstrated that these β-type interactions are inherent to the isolated C-terminal half. The shape of the spectrum does not change over a 10-fold concentration range, which indicates that the β-type interactions are characteristic of the truncated protein and do not result from aggregation/amyloid formation of this fragment. As witnessed by the CD spectrum of full-length CSD1, then, these interactions are overriden by competing native long-range contact(s) within the full-length protein that make it appear more disordered. This more global structural organization may nevertheless account for the exposure of the regions preferentially attacked by proteases under limiting conditions.

The spectrum of MAP2c also shows a characteristic minimum at 200 nm and is compatible with a random-coil conformation (Figure 5B). The spectra of the two halves in this case are very similar, and their sum reproduces the spectrum of the full-length protein within experimental error. This finding is consistent with the overall random structure of MAP2c and the lack of appreciable long-range interactions between its two halves. The structurally exposed nature of PRR in this case probably results from local structural constraints that stem from its high proline content. Such regions in proteins often adopt a polyproline II helix (PPII) conformation (*35*), and our CD data support the presence of this structural element (cf. refs *36*−*38*). First, PRR has a strong negative band at about 203 nm and a negative shoulder at 220 nm (Figure 6A). Second, urea, which usually unfolds structural elements, such as α helics and β sheets, stabilizes the structure, as witnessed by a gradual increase in negative ellipticity with increasing urea concentrations (Figure 6A). Upon heating from 10 to 70 °C, the negative peak at 200 nm undergoes a gradual decrease, indicating the loss of structure without a cooperative transition (Figure 6B). The isodichroic point at 215 nm indicates the equilibrium of two major conformations, PPII and random coil. The linear response to temperature with a slight kink at 63 °C (Figure 6C) also indicates the absence of a sharp structural transition but a biphasic behavior characteristic of PPIIs. All of these observations argue for the prevalence of PPII conformation (*36*−*38*) in the PRR of MAP2c. This extended and conformationally rather stiff local conformation may probably explain the preferential accessibility of this region for two plausible reasons. First, local structural stiffness could reduce excluded volume around a particular site and thus increase its accessibility. Second, the transition from a free to a protease-bound state might be facilitated by limited local order. The productive interaction with the active site of a protease extends to a region of about 10−12 residues (*39*, *40*). The accompanying decrease in entropy might be limited by the PPII "preorganization" (*41*). Further, PPII is fully exposed, lacks intramolecular hydrogen bonds, and progresses easily to other extended conformational states implicated in molecular recognition (*4*, *11*). These considerations are fully

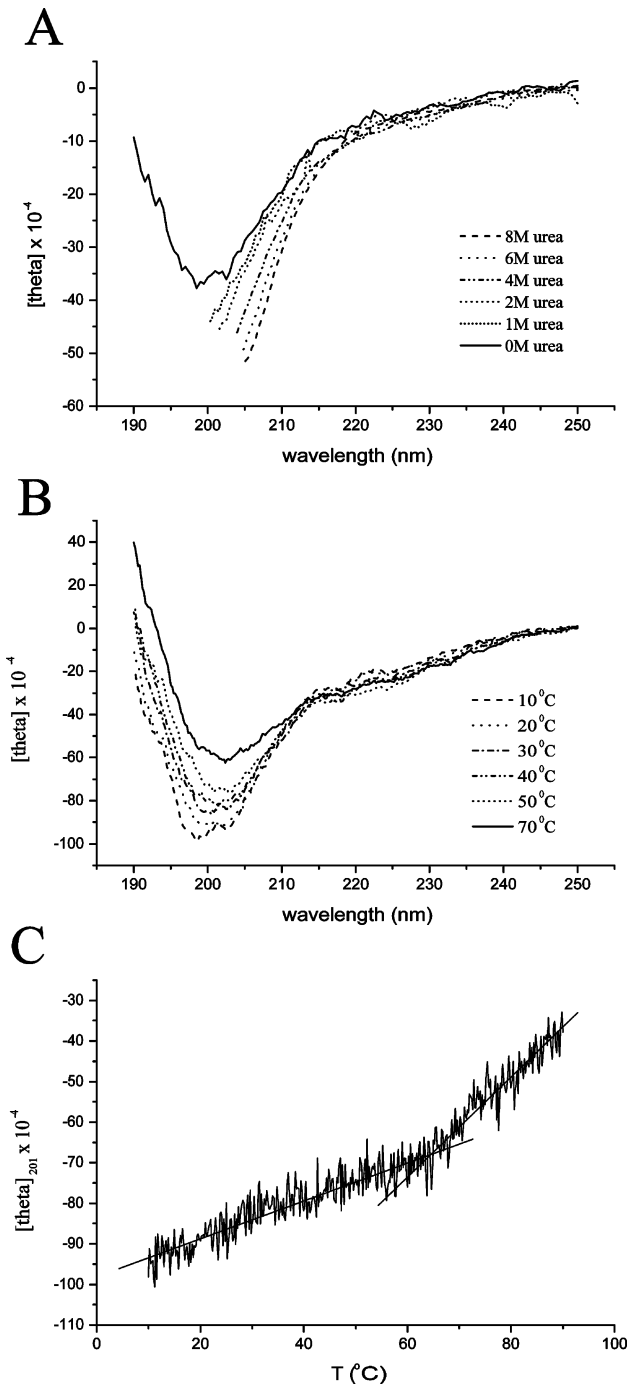

FIGURE 6: PRR of MAP2c adopts a PPII conformation. The structure of the PRR of MAP2c has been studied by far-UV CD spectroscopy. (A) Spectrum of recombinant PRR has been recorded at various urea concentrations ranging from 0 to 8 M. The increase of the negative peak with an increasing urea concentration is indicative of the PPII conformation (please note that the effect is clear although urea absorbance precluded recording for the whole wavelength region). (B) Spectrum of PRR at various temperatures from 10 to 70 °C. (C) Molar ellipticity values at 201 nm as a function of the temperature. The positive slope and slight kink at 63 °C are indicative of some structural order and cooperative structural transition, characteristic features of the PPII conformation.

compatible with the effective interaction of PRR with both an attacking protease and a physiological partner.

^{1}H NMR Measurements. The nature of transient order in the largely disordered structure of IUPs can also be probed by detecting suboptimal hydration of the polypeptide chain or its fragments, as shown previously for CSD1 by calorimetry (*42*). One-dimensional, solid-state ^{1}H NMR is suited for this purpose because it can yield the amount of bound water after freezing bulk water out (*28*).

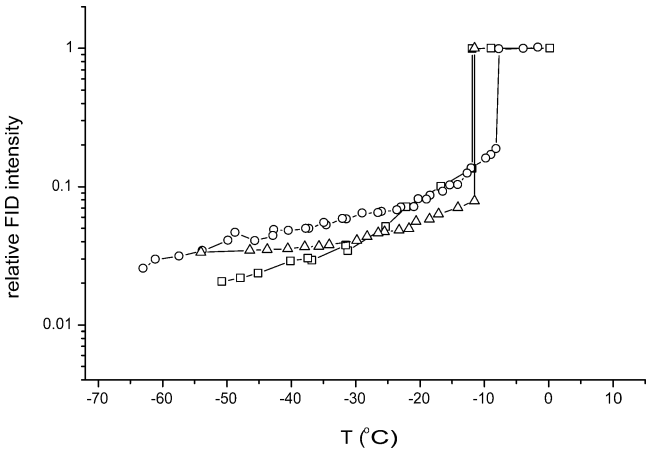

FIGURE 7: Temperature dependence of the FID intensity of CSD1 and MAP2c. Temperature dependence of the NMR FID intensity of water protons in CSD1 (□), MAP2c (○), and BSA (△) was measured at 50 mg/mL protein concentration. The sharp drop between −6 and −12 °C indicates that the mobile ("free") fraction of water freezes out; the remaining signal is indicative of water bound in the hydrate layer of the proteins. Because the FID amplitude is proportional to the number of protons, the intensity values yield the quantity of hydrate water, which is 8.1% for BSA, 11.7% for CSD1, and 16.2% for MAP2c.

We compared the temperature dependence of FID signal amplitude of water protons in CSD1 and MAP2c and in a control globular protein, BSA (Figure 7). Upon cooling the sample, the "free", i.e., unbound, fraction of water freezes out between about −6 and −12 °C, which enables the FID intensity of protons in the bound water to be measured below this temperature. The major part of proton FID signals coming from the protein and water in ice is in the dead time of the spectrometer and does not contribute to the FID signal measured. The amount of water in the hydrate layer of BSA (8.1%) is much smaller than that for any of the two IUPs, showing their open and largely solvent-exposed character. The difference between MAP2c and CSD1, however, is relevant with respect to their differences in structural organization. MAP2c binds significantly more water than CSD1 (16.2% versus 11.7% under identical conditions), and it binds the water stronger, because its hydrate layer freezes out completely at a lower temperature (−64 °C) than that of CSD1 (−52 °C). Because the mean hydrophobicity of MAP2c and CSD1 is practically the same (−0.780 for CSD1 and −0.781 for MAP2c, cf. other methods), the difference in their hydration indicates a substantial difference in the extent of their intramolecular interactions: judged simply by the amount of water that they bind, CSD1 appears more ordered than MAP2c.

The temperature-dependence of FID amplitudes of full-length proteins and their two separate halves (Figure 8) underlines this conclusion. In the case of CSD1, the magnitude of normalized FID intensity for the full-length protein is significantly lower than for its two halves; i.e., the intact CSD1 molecule binds less water than its two separate halves together (Figure 8A). In other words, its two halves interact in the intact molecule and partially replace each other's bound water. In contrast, the hydration of MAP2c can be accounted for by its N- and C-terminal halves

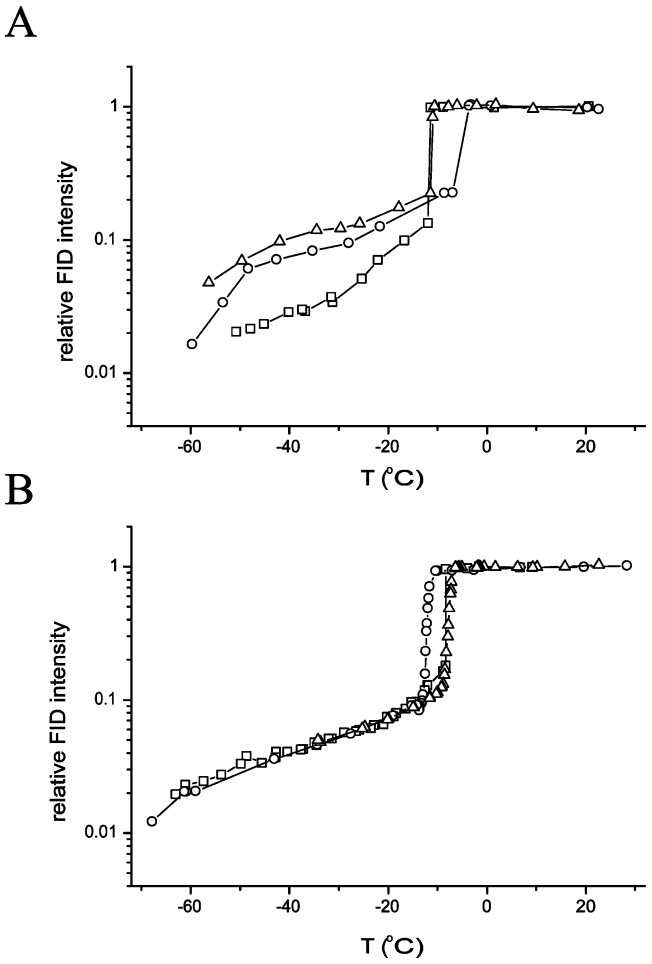

FIGURE 8: Temperature dependence of the FID intensity of CSD1, MAP2c, and their two halves. Temperature dependence of the NMR FID intensity of water protons for (A) CSD1 (□), its N-terminal (○), and C-terminal half (△) and for (B) MAP2c (□), its N-terminal (○), and C-terminal half (△), each at 50 mg/mL. The position of the two halves within the sequences is shown in Figure 4. The suboptimal hydration of full-length CSD1 is to be noted.

(Figure 8B); i.e., there are no long-range interactions between the two halves of the molecule.

DISCUSSION

In this paper, we demonstrated that the nature of the residual structure of IUPs and its relationship with function can be characterized by limited proteolysis, CD spectroscopy, and solid-state NMR, applied to recombinant fragments designed to coincide with functionally important regions of the protein. Our data suggest that the two IUPs studied encompass functionally significant residual structure, manifest as either transient short- and/or long-range structural organization, which ensures the preferential accessibility of their sites destined for productive interaction with their partner.

The spatial exposure of functionally important interaction sites in both CSD1 and MAP2c is shown by limited proteolysis at very low protease concentrations. In the case of globular proteins, primary cleavage is expected at sites that conform to the amino acid preferences of the protease and are sterically accessible and sufficiently flexible for productive interaction with the enzyme (*39*, *40*). Within a fully disordered protein, cleavage is expected to be dictated by amino acid preferences alone; the deviation from this indicates that certain sites are exposed on the time-average, whereas others tend to be buried. In CST, these fall within subdomains A, B, and C, the major determinants of binding with calpain in the enzyme−inhibitor complex (*16*−*18*), which suggests that the unbound inhibitor is structurally primed for rapid interaction with calpain. Cleavage of MAP2c in the PRR and the previously noted similar behavior of MAP2 (*31*, *32*) and tau protein (*33*, *34*) can be rationalized in a similar manner. Traditionally, the imperfect repeats within the microtubule-binding domain of MAPs have been thought to bind microtubules (*20*, *21*). Deletion mutagenesis of MAP2c (*27*) and tau protein (*22*, *23*), however, has shown that the major strength of microtubule binding comes from PRR and R′ flanking the repeats as a pair of jaws (cf. Figure 4). The preferential association with microtubules of PRR, thus, is in perfect agreement with prior binding data.

CD and solid-state ^{1}H NMR reveal the somewhat different structural background of this functionally critical feature in the two proteins. In the case of CSD1, the lack of additivity of the CD spectra of its two halves and suboptimal hydration of the full-length protein seen by NMR (cf. also ref *42*) suggest long-range interactions between the two halves of this largely disordered protein (*43*). Furthermore, its hydration is significantly less than that of MAP2c, which also indicates transient but significant long-range contacts within the molecule. Short-range secondary structural order of subdomains A, B, and C (*44*−*46*) probably also contributes to their effective interaction with calpain. With MAP2c, the situation is slightly different. The CD spectra of its two halves are additive, which points to the absence of long-range tertiary contacts. The NMR data also support this conclusion because MAP2c shows much larger hydration than CSD1 and the hydration of its two halves is additive in the intact protein. Thus, the exposure of PRR is probably best explained by local structural order that stems from its special amino acid composition. PRRs in proteins tend to adopt the left-handed PPII conformation (*35*, *47*), an open, fully hydrated but relatively stiff secondary structural motif often implicated in protein−protein interactions (*35*). The presence of this structural motif probably explains the capacity of this region for an initial, strong interaction with microtubules.

The major conclusion to be drawn from the foregoing is that certain functional sites in the fluctuating structure of IUPs are spatially exposed on the time average. As already mentioned, local structure may generally correspond to the structure adopted by IUPs in the bound state (*13*). It is tempting to speculate that the spatial exposure of interacting region(s) is another common and important functional faculty of these proteins: such regions we suggest to be denoted as PCSs. These are conceptually very similar to molecular recognition elements (MoREs) also suggested for IUPs (*48*) and anchor sites observed in globular proteins (*49*), in that they are structurally primed recognition motifs that dock to the partner and lead to the formation of a native-like encounter complex. PCSs/MoREs/anchor sites probably determine the "on" rate of complex formation and are conceptually different from hot spots also implicated in protein−protein interactions (*50*). A hot spot signifies the region in the molecular interface that contributes the major part of the free energy of binding, whereas a PCS is defined

Table 2: Interaction Sites in Various IUPs[a]

IUP (Swiss-Prot)	partner	PCS segment
CST (P20810)	calpain	Y^{190}IEE**LG**KREV**TIP**PKYRE[b] D^{229}DA**I**DA**L**SSD**F**TCGSP[c]
CREB[d] (P16220)	CBP	S^{121}QKRREILSRRP**S(phos)**YRKI
nucleoporin NSP1 (P14907)	nuclear transport factor 2	E^{545}SKSA**FSFG**SKPTG
MAP2c (P15146)	microtubule	S^{231}GTST**P**TT**P**GSTAI
$p27^{Kip1}$ (P46527)	Cyc A−Cdk 2	A^{28}CRN**LFG**PVDH
FnBP (Q53971)	fibronectin	T^{1044}TEV**ED**SKPK**L**SI**F**DNE**WP**KED
$p21^{Cip1}$ (P38936)	20S proteasome	H^{152}SKRR**LIF**SKRKP
PPI-2 (P41236)	protein phosphatase 1	A^{4}SHRP**IKGI**LKNKT
Tcf-4 (Q9NQB0)	β catenin	D^{11}LGAN**DE**LISFK S^{36}AERD**L**AD**V**KSS**L**VNESET

[a] Structural and mutagenesis data from the literature have been collected to compile the site(s) of IUPs, which make(s) critical contribution to the interaction of the IUP with its partner (for details, see the Discussion); residues shown in bold are the ones crucial for efficient interaction. Above the bold line, proteins are enlisted for which direct or indirect evidence indicates the probable presence of a PCS/anchor site, i.e., a spatially exposed region that makes the kinetically critical first contact with the partner. The sites are mostly hydrophobic, contributed by a short stretch of hydrophobic amino acids (e.g., for $p27^{Kip1}$ and NSP1), an amphipathic helix with hydrophobic amino acids 3−4 positions apart (e.g., CST subdomain C), or a PP II helix with proline side chains facing the partner (MAP2c putative site). Below the bold line are examples for which evidence points to the presence of a hot spot with a critical contribution to the binding energy of the IUP. These sites are also mostly hydrophobic (e.g., $p21^{Cip1}$ and PPI-2), but occasionally charged buttons (e.g., Tcf-4 first segment, FNBP ED motif) may fulfill the same role. It is to be noted that the distinction between the two categories is not absolute, because hot spots may turn out to be PCSs/anchor sites in later kinetic studies. [b] Subdomain B. [c] Subdomain C. [d] Phosphorylated on S^{133}.

in kinetic terms as a recognition element, which forms the initial contact with the partner and serves as a folding center in the subsequent induced folding process. The two may actually coincide for a given protein, but a good deal of kinetic and thermodynamic work will be needed to sort them out. It is also noteworthy that the presence of such site(s) may appear to be unnecessary in light of the original idea that protein disorder confers kinetic advantage on the recognition process per se (*4*). The contradiction, however, can be easily resolved by noting that molecular recognition may in fact be speeded by making use of the folding funnel of the IUP, as suggested in the "fly casting" mechanism (*51*). This mechanism invokes both the greater capture radius of IUPs that advances exploration in configuration space and the mechanistic coupling of the recognition process to folding. Preformed, exposed recognition elements, such as PCSs/MoREs, may be effective mechanistic devices that structurally limit the initial encounter complex, direct it toward the folding funnel, and thus mechanistically link recognition to the effective formation of the final folded state.

In the case of MAP2c, we must resort to indirect evidence for the exact location and nature of its PCS. As mentioned, PRRs are preferred interaction sites of proteins (*35*). The structure of a proline-rich peptide bound to an SH3 domain (*52*) shows that conserved prolines are in direct hydrophobic contact with the partner molecule. Conserved prolines in the PRR of MAP2c and tau may thus have a double role, ensuring exposure of the PCS region by maintaining the extended PPII conformation and providing key residues for interaction. Thus, we suggest the putative PCS of MAP2c as given in Table 2. With CSD1, mutagenesis, and inhibition studies identified Leu and Ile in subdomain B (Table 2) as absolutely crucial for interaction (*53*), this region tends to take a β-turn conformation (*44*), which places these residues in an exposed position so that they meet the criterion of PCS. The crystal structure of subdomain C peptide bound to the CaM-like domain of calpain small subunit has been solved recently (*45*) and also points to the importance of hydrophobic interactions. The peptide winds up to an amphipathic α helix (cf. also ref *46*) with the hydrophobic side chains L, F, and to a lesser extent I in intimate contact with the protein; a similar binding mode was suggested for subdomain A. Because prior mutagenesis studies have demonstrated the critical role of L in both subdomains A and C (*54*), the hydrophobic surfaces created by preformed helices are ideal candidates to be PCS in the interaction with calpain (Table 2).

For the lack of a great deal of relevant kinetic data and detailed structural studies on the free state of IUPs, we may infer the nature and location of PCSs only in a handful of other IUPs (Table 2). The binding of $p27^{Kip1}$ to Cyc A−Cdk 2 is initiated by a conserved LFG motif (*55*) that forms contact with a hot-spot hydrophobic docking site on Cyc A. Simulation of microscopic unfolding of the complex has shown that the LFG motif detaches last from the complex, and thus, it is probably the first to bind the cyclin (*56*). Recent kinetic studies in fact provided clear-cut evidence for this binding mode (*57*). Similar hydrophobic patches have also been implicated in the function of nucleoporins (Nups). The nuclear pore complex (NPC) consists of several dozen Nups, many of which contain extended, structurally disordered FG repeat regions with FxFG or related core consensus repeat motifs (*58*). Rapid and reversible cyclic interactions of the core motif with various carriers, such as the nuclear transport factor 2 (*59*) or importin β (*60*), are crucial to all translocation mechanisms (*61*). This assumes that essential Phe residues are exposed and structurally primed for efficient interaction. A different group as possible PCS has also been implicated in the interaction of the transciption factor cyclic AMP response element binding protein (CREB) with CREB-binding protein (CBP). This interaction critically depends on the phosphorylation of S^{133} within the kinase-inducible domain (KID) of CREB by protein kinase A (*62*). In a manner analogous to our studies here, the spatial exposure and preferred accessibility of this region has been demonstrated in limited proteolysis experiments (*63*).

In some other IUPs, equilibrium binding studies have ascertained hot-spot regions (Table 2). The combination of a charged button and an amphipathic α helix is crucial in the interaction of the unstructured CBD domain of Tcf

transcription factors with β catenin. Mutagenesis studies suggested that $D^{16}E^{17}$ within the central, extended region and an amphipathic helix towards the C terminus, with $L^{41}V^{44}L^{48}$ contributing the critical interactions (*64*−*66*), are the major determinants of binding. A similar combination has been implicated in the binding of a fibronectin-binding protein to fibronectin (*67*, *68*). Intriguingly, in the case of $p21^{Cip1}$, the homologue of $p27^{Kip1}$, the C8 α subunit of the 20S proteasome interacts with a hydrophobic patch within its C terminus distinct from the cyclin recognition site (*69*). Although the identity of this site as a PCS has not been established, this observation raises the possibility that different PCS signals within the same protein may mediate different interactions. Similar hydrophobic interactions dominate the contact of inhibitor 2 with protein phosphatase type 1. With this inhibitor, the major strength of interaction comes from the short IKGI motif located near the N terminus of the molecule; deletion of this "anchoring" segment (*70*) has almost completely crippled the inhibitor.

In summary, our data and all of the foregoing considerations are consistent with the concept that PCSs of IUPs are short, spatially exposed segments, which enable rapid and productive interaction with the partner molecule. Their exposure is ensured by a combination of transient local and global structural organization. The residues critical for interaction are mostly hydrophobic, but charged buttons may also work. Detailed binding studies will be required to clarify the generality and importance of this phenomenon within the class of IUPs.

ACKNOWLEDGMENT

The authors thank Dr. András Patthy (Agricultural Biotechnology Center, Gödöllö, Hungary) for protein sequencing.

REFERENCES

1. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure−function paradigm, *J. Mol. Biol. 293*, 321−331.
2. 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.
3. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics, *Protein Sci. 11*, 739−756.
4. Tompa, P. (2002) Intrinsically unstructured proteins, *Trends Biochem. Sci. 27*, 527−533.
5. Wootton, J. C. (1994) Sequences with "unusual" amino acid compositions, *Curr. Opin. Struct. Biol. 4*, 413−421.
6. Kim, T. D., Ryu, H. J., Cho, H. I., Yang, C. H., and Kim, J. (2000) Thermal behavior of proteins: Heat-resistant proteins and their heat-induced secondary structural changes, *Biochemistry 39*, 14839−14846.
7. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes, *Genome Inform. Ser. Workshop Genome Inform. 11*, 161−171.
8. Schweers, O., Schonbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for β-structure, *J. Biol. Chem. 269*, 24290−24297.
9. Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Muller-Frohne, M., Zirwer, D., Czarnecki, J., and Damaschun, G. (1995) Prothymosin α: A biologically active protein with random coil conformation, *Biochemistry 34*, 13211−13218.
10. Shortle, D. (1996) The denatured state (the other half of the folding equation) and its role in protein stability, *FASEB J. 10*, 27−34.
11. Tompa, P. (2003) The functional benefits of protein disorder, *J. Mol. Struct. 666*−*667*, 361−371.
12. Levinthal, C. (1969) in *Mossbauer Spectroscopy in Biological Systems* (Monticello, I. L., Ed.) pp 22−24, University of Illinois Press, Allerton House, IL.
13. 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.
14. Dyson, H. J., and Wright, P. E. (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states, *Methods Enzymol. 339*, 258−270.
15. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR, *Chem. Rev. 104*, 3607−3622.
16. Emori, Y., Kawasaki, H., Imajoh, S., Minami, Y., and Suzuki, K. (1988) All four 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.
17. Maki, M., Bagci, H., Hamaguchi, K., Ueda, M., Murachi, T., and Hatanaka, M. (1989) Inhibition of calpain by a synthetic oligopeptide corresponding to an exon of the human calpastatin gene, *J. Biol. Chem. 264*, 18866−18869.
18. 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.
19. 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.
20. Matus, A. (1994) in *Microtubules* (Hyams, J. S., and Lloyd, V. W., Eds.) pp 155−166, Wiley-Liss, Inc., New York.
21. Sanchez, C., Diaz-Nido, J., and Avila, J. (2000) Phosphorylation of microtubule-associated protein 2 (MAP2) and its relevance for the regulation of the neuronal cytoskeleton function, *Prog. Neurobiol. 61*, 133−168.
22. Gustke, N., Trinczek, B., Biernat, J., Mandelkow, E. M., and Mandelkow, E. (1994) Domains of tau protein and interactions with microtubules, *Biochemistry 33*, 9511−9522.
23. Preuss, U., Biernat, J., Mandelkow, E. M., and Mandelkow, E. (1997) The "jaws" model of tau-microtubule interaction examined in CHO cells, *J. Cell Sci. 110*, 789−800.
24. Chen, J., Kanai, Y., Cowan, N. J., and Hirokawa, N. (1992) Projection domains of MAP2 and tau determine spacings between microtubules in dendrites and axons, *Nature 360*, 674−677.
25. Mukhopadhyay, R., and Hoh, J. H. (2001) AFM force measurements on microtubule-associated proteins: The projection domain exerts a long-range repulsive force, *FEBS Lett. 505*, 374−378.
26. Rácz, P., Hargitai, C., Alföldy, B., Bánki, P., and Tompa, K. (2000) ^{1}H spin−spin relaxation in normal and cataractous human, normal fish, and bird eye lenses, *Exp. Eye Res. 70*, 529−536.
27. Yang, H. Q., Ma, H., Takano, E., Hatanaka, M., and Maki, M. (1994) Analysis of calcium-dependent interaction between amino-terminal conserved region of calpastatin functional domain and calmodulin-like domain of μ-calpain large subunit, *J. Biol. Chem. 269*, 18977−18984.
28. Ferralli, J., Doll, T., and Matus, A. (1994) Sequence analysis of MAP2 function in living cells, *J. Cell Sci. 107*, 3115−3125.
29. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680−685.
30. Schellenberger, V., Turck, C. W., Hedstrom, L., and Rutter, W. J. (1993) Mapping the S′ subsites of serine proteases using acyl transfer to mixtures of peptide nucleophiles, *Biochemistry 32*, 4349−4353.
31. Joly, J. C., Flynn, G., and Purich, D. L. (1989) The microtubule-binding fragment of microtubule-associated protein-2: Location of the protease-accessible site and identification of an assembly promoting peptide, *J. Cell Biol. 109*, 2289−2294.
32. Wille, H., Mandelkow, E. M., Dingus, J., Vallee, R. B., Binder, L. I., and Mandelkow, E. (1992) Domain structure and antiparallel dimers of microtubule-associated protein 2 (MAP2), *J. Struct. Biol. 108*, 49−61.
33. Aizawa, H., Kawasaki, H., Murofushi, H., Kotani, S., Suzuki, K., and Sakai, H. (1988) Microtubule-binding domain of tau proteins, *J. Biol. Chem. 263*, 7703−7707.
34. Steiner, B., Mandelkow, E. M., Biernat, J., Gustke, N., Meyer, H. E., Schmidt, B., Mieskes, G., Soling, H. D., Drechsel, D., Kirschner, M. W., et al. (1990) Phosphorylation of microtubule-associated protein tau: Identification of the site for Ca^{2+}-

calmodulin dependent kinase and relationship with tau phosphorylation in Alzheimer tangles, *EMBO J. 9*, 3539−3544.
35. Williamson, M. P. (1994) The structure and function of proline-rich regions in proteins, *Biochem. J. 297*, 249−260.
36. Ma, K., Kan, L., and Wang, K. (2001) Polyproline II helix is a key structural motif of the elastic PEVK segment of titin, *Biochemistry 40*, 3427−3438.
37. Gutierrez-Cruz, G., Van Heerden, A. H., and Wang, K. (2001) Modular motif, structural folds, and affinity profiles of the PEVK segment of human fetal skeletal muscle titin, *J. Biol. Chem. 276*, 7442−7449.
38. Bochicchio, B., and Tamburro, A. M. (2002) Polyproline II structure in proteins: Identification by chiroptical spectroscopies, stability, and functions, *Chirality 14*, 782−792.
39. Fontana, A., Polverino de Laureto, P., de Filippis, V., Scaramella, E., and Zambonin, M. (1997) Probing the partly folded states of proteins by limited proteolysis, *Folding Des. 2*, R17−R26.
40. Hubbard, S. J., Beynon, R. J., and Thornton, J. M. (1998) Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures, *Protein Eng. 11*, 349−359.
41. Ferreon, J. C., and Hilser, V. J. (2004) Thermodynamics of binding to SH3 domains: The energetic impact of polyproline II (PII) helix formation, *Biochemistry 43*, 7787−7797.
42. Häckel, M., Konno, T., and Hinz, H. (2000) A new alternative method to quantify residual structure in "unfolded" proteins, *Biochim. Biophys. Acta 1479*, 155−165.
43. Uemori, T., Shimojo, T., Asada, K., Asano, T., Kimizuka, F., Kato, I., Maki, M., Hatanaka, M., Murachi, T., Hanzawa, H., et al. (1990) Characterization of a functional domain of human calpastatin, *Biochem. Biophys. Res. Commun. 166*, 1485−1493.
44. 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.
45. Todd, B., Moore, D., Deivanayagam, C. C., Lin, G. D., Chattopadhyay, D., Maki, M., Wang, K. K., and Narayana, S. V. (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.
46. Mucsi, Z., Hudecz, F., Hollósi, M., Tompa, P., and Friedrich, P. (2003) Binding-induced folding transitions in calpastatin subdomains A and C, *Protein Sci. 12*, 2327−2336.
47. Reiersen, H., and Rees, A. R. (2001) The hunchback and its neighbours: Proline as an environmental modulator, *Trends Biochem. Sci. 26*, 679−684.
48. Bracken, C., Iakoucheva, L. M., Romero, P. R., and Dunker, A. K. (2004) Combining prediction, computation, and experiment for the characterization of protein disorder, *Curr. Opin. Struct. Biol. 14*, 570−576.
49. Rajamani, D., Thiel, S., Vajda, S., and Camacho, C. J. (2004) Anchor residues in protein−protein interactions, *Proc. Natl. Acad. Sci. U.S.A. 101*, 11287−11292.
50. Bogan, A. A., and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces, *J. Mol. Biol. 280*, 1−9.
51. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism, *Proc. Natl. Acad. Sci. U.S.A. 97*, 8868−8873.
52. Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. (1994) Structural basis for the binding of proline-rich peptides to SH3 domains, *Cell 76*, 933−945.
53. 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.
54. 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.
55. Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J., and Pavletich, N. P. (1996) Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex, *Nature 382*, 325−331.
56. Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T., and Rose, P. W. (2003) Simulating disorder−order transitions in molecular recognition of unstructured proteins: Where folding meets binding, *Proc. Natl. Acad. Sci. U.S.A. 100*, 5148−5153.
57. Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L., 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.
58. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L., and Rexach, M. (2003) Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded, *Proc. Natl. Acad. Sci. U.S.A. 100*, 2450−2455.
59. Bayliss, R., Leung, S. W., Baker, R. P., Quimby, B. B., Corbett, A. H., and Stewart, M. (2002) Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats, *EMBO J. 21*, 2843−2853.
60. Bayliss, R., Littlewood, T., and Stewart, M. (2000) Structural basis for the interaction between FxFG nucleoporin repeats and importin-β in nuclear trafficking, *Cell 102*, 99−108.
61. Stewart, M. (2000) Insights into the molecular mechanism of nuclear trafficking using nuclear transport factor 2 (NTF2), *Cell Struct. Funct. 25*, 217−225.
62. Radhakrishnan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions, *Cell 91*, 741−752.
63. Gonzalez, G. A., Menzel, P., Leonard, J., Fischer, W. H., and Montminy, M. R. (1991) Characterization of motifs which are critical for activity of the cyclic AMP-responsive transcription factor CREB, *Mol. Cell. Biol. 11*, 1306−1312.
64. Omer, C. A., Miller, P. J., Diehl, R. E., and Kral, A. M. (1999) Identification of Tcf4 residues involved in high-affinity β-catenin binding, *Biochem. Biophys. Res. Commun. 256*, 584−590.
65. Graham, T. A., Weaver, C., Mao, F., Kimelman, D., and Xu, W. (2000) Crystal structure of a β-catenin/Tcf complex, *Cell 103*, 885−896.
66. Fasolini, M., Wu, X., Flocco, M., Trosset, J. Y., Oppermann, U., and Knapp, S. (2003) Hot spots in Tcf4 for the interaction with β-catenin, *J. Biol. Chem. 278*, 21092−21098.
67. Schwarz-Linek, U., Werner, J. M., Pickford, A. R., Gurusiddappa, S., Kim, J. H., Pilka, E. S., Briggs, J. A., Gough, T. S., Hook, M., Campbell, I. D., and Potts, J. R. (2003) Pathogenic bacteria attach to human fibronectin through a tandem β-zipper, *Nature 423*, 177−181.
68. Schwarz-Linek, U., Pilka, E. S., Pickford, A. R., Kim, J. H., Hook, M., Campbell, I. D., and Potts, J. R. (2004) High affinity streptococcal binding to human fibronectin requires specific recognition of sequential f1 modules, *J. Biol. Chem. 279*, 39017−39025.
69. Touitou, R., Richardson, J., Bose, S., Nakanishi, M., Rivett, J., and Allday, M. J. (2001) A degradation signal located in the C-terminus of p21WAF1/CIP1 is a binding site for the C8 α-subunit of the 20S proteasome, *EMBO J. 20*, 2367−2375.
70. Yang, J., Hurley, T. D., and DePaoli-Roach, A. A. (2000) Interaction of inhibitor-2 with the catalytic subunit of type 1 protein phosphatase. Identification of a sequence analogous to the consensus type 1 protein phosphatase-binding motif, *J. Biol. Chem. 275*, 22635−22644.

BI047817F