THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 20, Issue of May 14, pp. 20775–20785, 2004
Printed in U.S.A.

On the Sequential Determinants of Calpain Cleavage*

Received for publication, December 18, 2003, and in revised form, February 20, 2004
Published, JBC Papers in Press, February 26, 2004, DOI 10.1074/jbc.M313873200

Peter Tompa‡, Peter Buzder-Lantos§, Agnes Tantos, Attila Farkas, András Szilágyi, Zoltán Bánóczi¶, Ferenc Hudecz¶, and Peter Friedrich‖

From the Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, P.O. Box 7, H-1518, Hungary and the ¶Research Group of Peptide Chemistry at Eötvös Loránd University, Hungarian Academy of Sciences, Budapest, P.O. Box 32, H-1528, Hungary

The structural clues of substrate recognition by calpain are incompletely understood. In this study, 106 cleavage sites in substrate proteins compiled from the literature have been analyzed to dissect the signal for calpain cleavage and also to enable the design of an ideal calpain substrate and interfere with calpain action via site-directed mutagenesis. In general, our data underline the importance of the primary structure of the substrate around the scissile bond in the recognition process. Significant amino acid preferences were found to extend over 11 residues around the scissile bond, from P_4 to P_7'. In compliance with earlier data, preferred residues in the P_2 position are Leu, Thr, and Val, and in P_1 Lys, Tyr, and Arg. In position P_1', small hydrophilic residues, Ser and to a lesser extent Thr and Ala, occur most often. Pro dominates the region flanking the P_2-P_1' segment, *i.e.* positions P_3 and P_2'-P_4'; most notable is its occurrence 5.59 times above chance in P_3'. Intriguingly, the segment C-terminal to the cleavage site resembles the consensus inhibitory region of calpastatin, the specific inhibitor of the enzyme. Further, the position of the scissile bond correlates with certain sequential attributes, such as secondary structure and PEST score, which, along with the amino acid preferences, suggests that calpain cleaves within rather disordered segments of proteins. The amino acid preferences were confirmed by site-directed mutagenesis of the autolysis sites of *Drosophila* calpain B; when amino acids at key positions were changed to less preferred ones, autolytic cleavage shifted to other, adjacent sites. Based on these preferences, a new fluorogenic calpain substrate, DABCYL-TPLKSPPPSPR-EDANS, was designed and synthesized. In the case of μ- and m-calpain, this substrate is kinetically superior to commercially available ones, and it can be used for the *in vivo* assessment of the activity of these ubiquitous mammalian calpains.

It is generally held that calpains, intracellular cysteine proteases, play crucial roles in basic physiological (1–3) and pathological (4) processes. This view is partially based on the fact that calpains cleave their substrates in a limited manner, modifying, rather than terminating, their action. So far, more than 100 proteins have been identified as calpain substrates; among them, cytoskeletal/structural proteins, membrane receptors, enzymes, and transcription factors abound (5, 6). Despite this wealth of data, the sequential/structural determinants of calpain cleavage are little understood. Early studies on small fluorogenic peptides (7) and peptide hormones (7, 8) have led to the so-called P_2-P_1 rule, which states that the preferred residues for the ubiquitous forms, μ- and m-calpain, are Leu and Val at position P_2, whereas in P_1 Arg and Lys (and to a lesser extent Tyr) occur most often (9). Additional, weaker, preferences have also been noted for positions P_3 (Phe, Trp, Leu, Val) and P_1' (Arg, Lys, Leu).

An elegant study encompassing all possible point mutations at the P_2 position of the principal cleavage site of fodrin has shown good correlation with the P_2-P_1 rule (10); other studies on protein substrates, however, displayed frequent deviations (11, 12). These observations suggested that calpain cleavage, to a large extent, depends on higher order structural clues. This aspect of substrate recognition was convincingly demonstrated in histone hydrolysis, where cleavage of sites susceptible in the intact protein was restrained in its fragments (11). It was concluded that calpain probably recognizes global structural elements present in the intact protein but absent from its smaller fragments.

This recognition has fueled efforts to uncover these putative higher order structural elements. In the above report (11) it was suggested that cleavage, as in histone, may never occur in the middle of either an hydrophobic or hydrophilic cluster but always at the border separating them. In another, comprehensive treatise including most protein substrates known at the time (5) it was found that calpain substrates often have a calmodulin-binding motif and targeting of the enzyme may occur by an adjacent PEST[1] sequence. This idea was raised in the original proposal of the PEST hypothesis, *i.e.* that the occurrence of region(s) rich in amino acids Pro, Glu(Asp), and Ser/Thr flanked by Arg/Lys residues correlates with the short lifetime of proteins (13, 14). Such a negatively charged region was thought to bind Ca^{2+} and provide both peptide bonds and the necessary co-factor for calpain action. Subsequent experimental testing of these suggestions brought controversial re-

* This work was supported by Grants T 29059, T 32360, T 34255, T 043576, and TS 040723 from OTKA (Hungarian Scientific Research Fund) and Research and Development Fund of Higher Education (FKFP) Grants 0100/2000 and 0053/2001 and National Research and Development Fund (NKFP) Grant 1/010. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by a Bolyai János Scholarship.

§ Current address: IVAX-Institute for Drug Research Ltd., Budapest, Hungary.

‖ To whom correspondence should be addressed: Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1518 Budapest, P. O. Box 7, Hungary. Tel.: 361-466-5856; Fax: 361-466-5856; E-mail: friedric@enzim.hu.

[1] The abbreviations used are: PEST, Pro, Glu(Asp), and Ser/Thr region; DABCYL, 4-(4-dimethylaminophenylazo)benzoic acid; EDANS, 5-[(2-aminoethyl) amino]naphthalene-1-sulfonic acid; FRET, Förster-type resonance energy transfer; LY-AMC, *N*-succinyl-Leu-Tyr-7-amido-4-methylcoumarine; DCM, dichloromethane; DIEA, *N*,*N*-diisopropylethylamine; RP-HPLC, reverse phase high performance liquid chromatography; BOC, tert-butyloxycarbonyl; Fmoc, 9-fluorenylmethoxycarbonyl; DMF, *N*,*N*-dimethylformamide.

sults, though. Purification of μ-calpain by an immobilized peptide containing the calmodulin-binding motif of plasma membrane ATPase supported the involvement of such region(s) (15). Attempts with the PEST motif, however, yielded equivocal results. In some instances mutation of the PEST sequence abolished calpain cleavage (16, 17) but was without effect in other cases (18, 19). Thus, whether a PEST region provides a degradation signal for calpain, is still an open issue.

In general, higher order clues still lay largely hidden in the structure of calpain substrates, and recent reviews on the subject (2, 6, 20) resort to the P_2-P_1 rule unveiled a long time ago. In our view, the major drawback of this point of reference is that it is mostly based on studies with peptides, known to be poor calpain substrates (9). Further, the use of a small number of peptides limits the range of amino acid combinations tested, and thus biases the amino acid preferences observed. An incomparably larger and more complete set of potential substrate sites is represented by the proteome, of which the cleavable ones make their way into the literature as calpain substrates. To advance our understanding of what constitutes a calpain site, we carried out a systematic analysis of these data and constructed a preference matrix. The sequential preferences manifest in this matrix point to some general aspects of the primary structure in substrate recognition by μ- and m-calpain. Furthermore, the versatile applicability of this matrix in engineering substrates of these ubiquitous calpains is also described in this report.

MATERIALS AND METHODS

Data Collection and Analysis—49 calpain substrates with a total of 106 sequentially identified cleavage sites have been selected from the literature. For homologous substrates cleaved at identical sites, only one species (preferably human) was considered. The protein sequences were downloaded from the Swiss-Prot and TrEMBL databases via the ExPASy molecular biology server (www.expasy.ch). The sequences were truncated to 50 amino acids in both the N- and C-terminal direction from the scissile bond and used for further studies. Amino acid frequency matrices and profile scores were calculated by a program written in the Python programming language. The amino acid composition of the complete Swiss-Prot and TrEMBL data base used for normalizing matrices was taken from Swiss-Prot (us.expasy.org/sprot/). Sequential attributes were calculated or predicted by the appropriate servers: PEST-score with the PESTfind algorithm (www.at.embnet.org/embnet/tools/bio/PESTfind), secondary structure with the nnPredict site (www.cmpharm.ucsf.edu/~nomi/nnpredict.html), and hydrophobicity with ProtScale (us.expasy.org/cgi-bin/protscale.pl). These servers are available via ExPASy. For strong and weak PEST regions detected by the PESTfind algorithm, we arbitrarily assigned a score of 3 and 1. For scoring the border of a weak PEST region, its flanking hexapeptide was assigned weights of 1, 2, 4, 4, 2, and 1. For a strong PEST region, weights of 1, 3, 9, 9, 3, and 1 were used; in both cases all other residues were set to 0.

Synthesis of FRET Substrate—The substrate, designed on the basis of the observed amino acid preferences of calpain cleavage, is an 11-mer peptide with the sequence of Thr-Pro-Leu-Lys-Ser-Pro-Pro-Pro-Ser-Pro-Arg, with a fluorescent donor, EDANS (Sigma) attached to its C terminus and a quenching acceptor, DABCYL attached to the N terminus. Since the measurement of calpain activity is based on the increase of donor fluorescence due to cessation of FRET between the EDANS and DABCYL groups upon cleavage (21), this substrate is termed as the FRET substrate throughout the article.

The peptide was synthesized on 0.2 g of Boc-Arg(Mts)-PAM-resin (0.31 mmol/g) (from NovaBiochem, Laufelfingen, Switzerland) by Boc (tert-butyloxycarbonyl) chemistry. Side chains of threonine and serines were protected with benzyl groups, whereas the ϵ-amino group of lysine was blocked by an Fmoc (9-fluorenylmethoxycarbonyl) group. (All Boc amino acid derivatives were obtained from Reanal, Budapest, Hungary.) Boc protection was removed with 33% trifluoroacetic acid in DCM (2 + 20 min) followed by washing with DCM (5 × 0.5 min), neutralization with 10% DIEA in DCM (3 × 1 min), and DCM wash (4 × 0.5 min). The amino acid derivatives and coupling reagents (*N*,*N*,*N*′-dicyclohexyl carbodiimide and 1-hydroxybenzotriazole, at 1:1 molar ratio) dissolved in DCM-DMF (*N*,*N*-dimethylformamide) 4:1 (v/v) were used in 3 molar excess for the resin capacity. The coupling reaction was continued for 60 min at room temperature, and the resin was washed with DMF (2 × 0.5 min) and DCM (3 × 0.5 min). The efficiency of coupling was checked by ninhydrin or isatin tests.

The N-terminal Boc group was removed from the protected peptide prior to cleavage by hydrogen-fluoride. The peptide was removed from the resin with 10 ml of hydrogen fluoride containing 0.5 g of *p*-cresol. The crude product was purified by RP-HPLC.

DABCYL was coupled to the N-terminal amino group of the purified H-TPLK(Fmoc) SPPPSPR-OH peptide in solution. Peptide was dissolved in DMF and reacted with DABCYL succinimide ester (Fluka, Buchs, Switzerland) in the presence of DIEA (DABCYL succinimide/DIEA/peptide = 2:1:1 mol/mol/mol). The crude product was purified by RP-HPLC. The N^{ϵ}-Fmoc group of the purified DABCYL-TPLK(Fmoc)SPPPSPR-OH peptide was removed with 20% piperidine/DMF. The peptide was purified by RP-HPLC. EDANS was attached to the C terminus of the DABCYL-peptide in DMF, using the following molar ratio: peptide/EDANS/EDC×HCl/1-hydroxybenzotriazole/DIEA = 1:2:4.5:6.8:9 (mol/mol). The target peptide was purified by RP-HPLC. The peptides were characterized by analytical HPLC (see below), amino acid analysis using a Beckman (Fullerton, CA) model 6300 amino acid analyzer and by mass spectra recorded on a PerkinElmer Life Sciences API 2000 triple quadrupole mass spectrometer (Sciex, Toronto, Canada).

Analytical RP-HPLC was performed on a Knauer (Bad Homburg, Germany) HPLC system using Phenomenex Synergi MAX-RP C12 column (250 × 4.6 inner diameter) with 5-mm silica (80 Å pore size) as a stationary phase. Linear gradient elution (0 min, 0% B; 5 min, 0% B; 50 min, 90% B) with eluent A (0.1% trifluoroacetic acid in water) and eluent B (0.1% trifluoroacetic acid in acetonitrile/water (80:20, v/v)) was used at a flow rate of 1 ml/min at ambient temperature. Sample was dissolved in eluent B, and peaks were detected at a λ of 220 nm. The crude product was purified on a semipreparative Phenomenex Jupiter C18 column (250 × 10-mm inner diameter) with silica (300 Å pore size). The peptides were checked by amino acid analysis using a Model 6300 amino acid analyzer, and mass spectra were recorded on an API 2000 triple quadrupole mass spectrometer.

Calpain B Mutagenesis and Autolysis Experiments—Calpain B autolysis sites (22) were mutated with the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's recommendations using the following primers: 5′-CGAGTCATGCCGGTGTTCCGTCGTATGCGGGAC-3′ and 5′-GTCCCGCATACGACGGAACACCGGCATGACTCG-3′ for the first major site; 5′-GAAGGTGCCCGAGGCTGTTAACATGTTTTGG-3′ and 5′-CAAAACATGTTAACAGCCTCGGGCACCTTC-3′ for the second major site (22).

Calpain B autolysis reaction was started in calpain buffer (10 mM HEPES, pH 7.5, 150 mM NaCl, and 1 mM EDTA) with 10 mM free Ca^{2+} and run for 10 min. The reaction was stopped by the addition of SDS sample buffer and 5 min of boiling. Samples were run on an SDS-polyacrylamide gel.

Enzyme Activity Measurements—Activity measurements were carried out with m-calpain, calpain A, calpain B, papain, trypsin, chymotrypsin, and cathepsin-B. Enzyme activity was measured with a Jasco FP 777 spectrofluorometer at excitation/emission wavelengths of 380/460 nm for LY-AMC (Sigma) and 320/480 nm for the FRET substrate, in a 3 × 3 mm quartz cuvette. The reaction mixture in 50 μl of calpain buffer (10 mM HEPES, 150 mM NaCl, 1 mM EDTA, 5 mM benzamidine, 0.5 mM phenylmethylsulfonyl fluoride, 10 mM β-mercaptoethanol, pH 7.5) contained various substrate concentrations and 3 mM free Ca^{2+} in the case of m-calpain and 19 mM free Ca^{2+} in the case of calpain A and calpain B. The reaction was started by rapidly mixing the enzyme in the reaction mixture. The enzyme concentrations typically used were as follows: 0.2 μM m-calpain, 4.8 μM calpain A, 1.8 μM calpain B, 1.0 μM papain, 1.0 μM trypsin, 40.0 μM chymotrypsin, and 10.0 μM cathepsin B. Data were analyzed by the MicroCal Origin data analysis software to determine the initial slope of fluorescence change.

Calpain Activity Measurements in Vivo—*In vivo* calpain activity was measured in cultured *Drosophila* S2 cells. The cell cultures were maintained in *Drosophila* SFM (Invitrogen) medium supplemented with 20 mM L-glutamine, 50 units/ml penicillin, and 50 μg/ml streptomycin at 23 °C. Four hours prior to starting the experiments, 200 μl of S2 cells (3×10^6 cell/ml) were lipofected with 200 μl of FRET substrate (giving a final substrate concentration of 280 μM in the medium) either without or with 15 μM calpastatin, using Cellfectin (Invitrogen) reagent. The protocol provided by the manufacturer was followed. Control cells were treated with the same amount of Cellfectin, but without the addition of the FRET substrate.

Before measurement, lipofected cells were collected by centrifugation at 1000 rpm for 4 min at room temperature and were resuspended in

400 μl of phosphate-buffered saline. This washing procedure was repeated twice to remove all the remaining substrate from the medium. 50 μl of cell suspension was transferred to a 3 × 3 mm quartz cuvette for recording fluorescence. Measurements with lipofected and control cells were started by the addition of 10 μM ionomycin either with or without 500 nM E-64 or calpain inhibitor II (both are cell permeable calpain inhibitors). The fluorescence measurements and data analysis were done similarly to the *in vitro* experiments.

COS-7 Cell Extracts—COS-7 cells were grown in Dulbecco's modified Eagle's medium (Sigma) at 37 °C and 5% CO_2 content. Confluent cells were washed three times with phosphate-buffered saline and were scraped in 1 ml of ice-cold phosphate-buffered saline. Cells were collected by centrifugation at 1500 × *g* for 2 min at 4 °C. Collected cells were resuspended in a buffer containing 100 mM Tris-HCl, 5 mM EDTA, 1 mM dithioerythritol, 5 mM benzamidine, 0.5 mM phenylmethylsulfonyl fluoride, and 10 mM β-mercaptoethanol and were sonicated four times for 10 s with 1-min breaks. After sonication, the lysate was centrifuged at 15,000 × *g* for 20 min at 4 °C to remove the cell debris, and the supernatant was used for fluorometric calpain activity measurements.

Proteins—The 80-kDa large subunit and the 21-kDa truncated small subunit of rat m-calpain was expressed in *Escherichia coli* and purified according to the method described in Ref. 23. *Drosophila* calpain A and calpain B were expressed in *E. coli* using the method described in Ref. 24. Human calpastatin domain 1 was prepared as given in Ref. 25. Other enzymes for activity measurements, such as papain (P-3375), trypsin (T-8003), α-chymotrypsin (C-4129), and cathepsin B (C-6286), were purchased from Sigma. These lyophilized enzymes were dissolved in calpain buffer just prior to the experiment and were kept on ice until use.

Other Methods and Materials—SDS-polyacrylamide gel electrophoresis was carried out according to Ref. 26. Calpain inhibitors were purchased from Calbiochem. For N-terminal sequence analysis, samples were run on SDS-PAGE gels and blotted onto polyvinylidene difluoride membranes (Sigma). N-terminal sequence analysis was performed using a modified Edman degradation sequencer program.

RESULTS

The Preference Matrix—The 106 sites of 49 substrates are listed in Table I. The number of proteins known to be calpain substrates is actually significantly higher than this. For most proteins, however, the exact site of cleavage has not been determined. Cleavage by μ- and m-calpain has not been distinguished, because the two ubiquitous isoforms have nearly identical substrate preferences (6, 9), and the quantity of data would not allow such a fine distinction to be made.

The amino acid preferences of calpain in positions P_4 through P_7' of the substrate has been calculated by determining the frequencies for each position in the 106 sites and dividing this percentage with the average amino acid occurrence values for the entire Swiss-Prot and TrEMBL data base (these in percentage are: Ala, 7.64; Gln, 3.94; Leu, 9.55; Ser, 7.06; Arg, 5.20; Glu, 6.47; Lys, 5.96; Thr, 5.56; Asn, 4.34; Gly, 6.87; Met, 2.36; Trp, 1.21; Asp, 5.25; His, 2.25; Phe, 4.09; Tyr, 3.15; Cys, 1.63; Ile, 5.84; Pro, 4.89; Val, 6.62). Therefore, the matrix generated (Table II) is normalized to the chance occurrence and thus represents the preference of calpain for each amino acid in these positions. Outside this region, no significant preference can be seen (not shown).

In the matrix, amino acids that occur at least two times as frequently as by chance are bolded. Apart from values in support of the classical P_2-P_1 rule, most notable is the abundance of Thr in P_2, Ser in P_1', and Pro in positions P_3 and P_2'-P_4'. The paucity of some amino acids in certain places is also conspicuous: the virtual absence of Val and Ile from P_1 and Pro from P_1' seems to be the most significant. Other marked deviations from chance occurrence (*e.g.* the preference for Trp but avoidance of Cys in P_3') are also observable; their significance, however, can not be established due to the limited size of our data base and the low frequency of these amino acids in Swiss-Prot and TrEMBL. A further interesting point, nevertheless, is that positions P_1'-P_7' correspond to the consensus inhibitory subdomain of calpastatin, the specific inhibitor of calpain (TIPPXYR, Ref. 27, underlined in Table II). The similarity, however, breaks down in P_1. Here Val, Glu, and Asp can be found in the four inhibitory domains of human calpastatin. In general, the matrix as a whole is a much better representation of calpain preferences than the classical P_2-P_1 rule, and its use has three important ramifications.

Predicting Calpain Cleavage Sites—The first important use of the extended preferences is that they can be used to score potential cleavage sites within a sequence. This can be done by scoring the sequence with the preference matrix, *i.e.* calculating a profile value for each peptide bond along the chain via adding up the corresponding matrix values for each position in the putative flanking P_4-P_7' 11-mer sequence; a maximum is expected where calpain actually cleaves the protein. To demonstrate this possibility, the scan was carried out for all 106 sites collected and analyzed; to make the score independent of the actual substrate site, the scan was always made with a unique matrix calculated by leaving the actual sequence out (Fig. 1). The average of distributions underscore the significant inclination for calpain to cleave at a position where the score reaches a maximum (Fig. 1*A*). Characteristic plots of 106 scans, for which cleavage occurs in the global maximum (vimentin, Fig. 1*B*), in a local maximum (von Willebrand factor, Fig. 1*C*) or elsewhere (p53, Fig. 1*D*) are also shown. Out of the 106 sites considered, 16 was found to fall into the global maximum, 59 are in local maxima, and there are only 31 cases where calpain cleaves in a site with a low profile score. To emphasize this point further, the profile score plot of a typical substrate with multiple cleavage sites (tau protein) is shown in Fig. 1*E*. This plot is a composite of 9 separate scans, each made with a weight matrix independent of the actual cleavage site. Of the 9 sites, 7 fall into local maxima, and only two are found at positions of low score.

The fact that this scoring scheme cannot unequivocally pinpoint calpain cleavage sites underscores the importance of more distant structural clues in substrate recognition. To test for such clues, various sequential attributes have been scanned and correlated with the site of cleavage. We found no particular correlation with hydrophobicity (data not shown), although cleavage sites were previously claimed to fall within regions separating hydrophobic and hydrophilic clusters (11). Secondary structure, on the other hand, does have an influence on where calpain attack occurs. Although the enzyme cleaves in rather disordered regions (70% of the cases irregular, 8% β-sheet, and 22% α-helix, Fig. 2*A*), in the immediate vicinity of the scissile bond a significant deviation from these values is seen. On the N-terminal side, the β-sheet value goes up to almost 20%. In contrast, within the C-terminal decapeptide both α-helix and β-sheet values drop sharply, and irregular structure occurs almost 90% of all cases. Furthermore, in the very few cases where calpain cleaves in an α-helical segment of a substrate with known three-dimensional structure, this site is preferably located close to the chain terminus, where it can easily undergo local unfolding prior to cleavage (data not shown). In the case of PEST regions, some correlation with calpain cleavage is also apparent. It should not be ignored, though, that the occurrence of a PEST region is not required for calpain cleavage in general, as many calpain substrates have no PEST region at all (19). For calpain substrates with a significant PEST region, nevertheless, cleavage tends to occur at the C-terminal border of PEST regions, as its position coincides with the global maximum of the arbitrary PEST border score (Fig. 2*B*).

Interfering with Calpain Cleavage: the Case of Calpain B Autolysis—The second point of practical importance is that the preference matrix can be used for designing targeted substrate

TABLE I
Cleavage sites in calpain substrates

The sequential positions of 106 cleavage sites in 49 substrates of μ- and m-calpain have been collected from the literature. The sequence of the hexapeptide around the site and the position of the residue at P_1 is given.

Substrate	Species	Cleavage site	Ref.
ABP	Human	Pro-Gln-Tyr^{1761}-Thr-Tyr-Ala	SwissProt P21333
Actin	Human	Val-Gly-Arg^{39}-Pro-Arg-His	48
Annexin I	Bovine	Thr-Val-Lys^{26}-Gly-Ser-Lys	49
Arrestin	Bovine	Phe-Val-Phe^{377}-Glu-Glu-Phe, Gln-Asn-Leu^{385}-Lys-Asp-Ala	50
Calpain 30K	Chicken	Val-Ser-Met^{87}-Val-Asp-Pro	51
Alpain 80K	Chicken	Arg-Leu-Arg^{17}-Ala-Glu-Gly	51
CaMK IV	Mouse	Val-Cys-Gly^{201}-Thr-Pro-Gly, Thr-Glu-Asn^{23}-Leu-Val-Pro	52
CaM-PDE1A2	Bovine	Val-Val-Gln^{126}-Ala-Gly-Ile	53
Caspase-9	Human	Gln-Leu-Asp^{330}-Ala-Ile-Ser, Pro-Glu-Ile^{115}-Arg-Lys-Pro	54
c-Fos	Rat	Ser-Gln-Thr^{90}-Arg-Ala-Pro	55
c-Jun	Rat	Leu-Asn-Leu^{42}-Ala-Asp-Pro, Leu-Leu-Thr^{62}-Ser-Pro-Asp, Ile-Thr-Thr^{90}-Thr-Pro-Thr,[a] Ser-Leu-His^{164}-Ser-Glu-Pro	55
Connexin50	Ovine	Leu-Thr-Glu^{290}-Val-Gly-Met, Pro-Leu-Ser^{300}-Ala-Lys-Pro	56
β-Crystallin A3	Bovine	Glu-Leu-Glu^{11}-Ser-Leu-Pro,	57
		Thr-Thr-Lys^{17}-Met-Ala-Gln	58
dystrophin	Human	Pro-Leu-Glu^{1992}-Ile-Ser-Tyr, Val-Thr-Thr^{690}-Arg-Glu-Gln[b]	59
EGFR	Human	Arg-Leu-Leu^{659}-Gln-Glu-Arg, Trp-Ile-Pro^{709}-Glu-Gly-Glu, Ser-Thr-Ser^{1006}-Arg-Thr-Pro, Ser-Cys-Pro^{1035}-Ile-Lys-Glu, Asp-Thr-Phe^{1062}-Leu-Pro-Val, Ser-Thr-Phe^{1127}-Asp-Ser-Pro, Pro-Asn-Gly^{1161}-Ile-Phe-Lys	60
GluR-1	Human	Ala-Ile-Arg^{855}-Thr-Ser-Thr, Ser-Ile-Asn^{851}-Glu-Ala-Ile	61
a-Hemoglobin	Human	Asn-Val-Lys^{11}-Ala-Ala-Trp	62
b-Hemoglobin	Human	Glu-Glu-Lys^{8}-Ser-Ala-Val	62
Histone H2A	Bovine	Arg-Leu-Leu^{34}-Arg-Lys-Gly	11
Histone H2B	Bovine	Gly-Thr-Lys^{116}-Ala-Val-Thr	11
Histone H3.2	Bovine	Ala-Thr-Gly^{33}-Gly-Val-Lys	11
HMG-CoA reductase	Rat	Pro-Lys-Lys^{363}-Ala-Gln-Asp	63
Integrin β_2	Human	Thr-Val-Met^{762}-Asn-Pro-Lys, Lys-Leu-Lys^{744}-Ser-Gln-Trp, Pro-Leu-Phe^{754}-Lys-Ser-Ala	64
Integrin β_3	Human	Glu-Arg-Ala^{735}-Arg-Ala-Lys, Trp-Asp-Thr^{741}-Ala-Asn-Asn, Pro-Leu-Tyr^{747}-Lys-Glu-Ala, Ser-Thr-Phe^{754}-Thr-Asn-Ile, Ile-Thr-Tyr^{759}-Arg-Gly-Thr	65
Interleukin-1a	Dog	Lys-Pro-Arg^{108}-Ser-Val-Ala,	66
	Human	Lys-Pro-Arg^{112}-Ser-Ser-Pro	67
MAP2c	Rat	Val-Val-Thr^{99}-Ala-Glu-Ala	-[c]
MBP	Bovine	Asn-Ile-Val^{93}-Thr-Pro-Arg, Ala-Ser-Ala^{16}-Ser-Thr-Met, His-Tyr-Gly^{68}-Ser-Leu-Pro, Thr-Pro-Arg^{96}-Thr-Pro-Pro	68
Merlin	Human	Val-Asn-Lys^{294}-Leu-Ile-Leu, Ile-Leu-Gln^{298}-Leu-Cys-Ile	69
MIP	Rat	Ile-Leu-Lys^{238}-Gly-Ala-Arg	70
Myosin-V (brain)	Chicken	Pro-Leu-Arg^{1140}-Met-Glu-Glu, Pro-Leu-Ser^{614}-Arg-Thr-Pro	71
NKEF-B	Human	Lys-Glu-Tyr^{193}-Phe-Ser-Lys, Ser-Asp-Thr^{182}-Ile-Lys-Pro	72
NMDAR 2A	Rat	Leu-Gln-Phe^{1279}-Gln-Lys-Asn, Leu-Phe-Ser^{1330}-Val-Pro-Ser	73
p35	Mouse	Ser-Thr-Phe^{98}-Ala-Gln-Pro	74
p53	Human	Trp-Lys-Leu^{25}-Leu-Pro-Glu	75
pADPRT	Bovine	Ala-Val-His^{384}-Ser-Gly-Pro, His-Leu-Leu^{480}-Ser-Pro-Trp, Lys-Ser-Gly^{502}-Ala-Ala-Pro, Lys-Leu-Thr^{658}-Val-Asn-Pro	76
Phospholipase C-β1	Bovine	Ala-Leu-His^{880}-Ser-Gln-Pro, Glu-Asn-Pro^{1208}-Gly-Lys-Glu	77
Phosphorylase kinase g	Rabbit	Pro-Arg-Gly^{302}-Lys-Phe-Lys	78
PKC-α	Human	Gly-Asn-Lys^{316}-Val-Ile-Ser, Lys-Ala-Lys^{309}-Leu-Gly-Pro, Glu-Asp-Arg^{324}-Lys-Gln-Pro	79
PKC-β	Human	Lys-Ile-Gly^{311}-Gln-Gly-Thr, Glu-Glu-Lys^{320}-Thr-Ala-Asn	79
PKC-γ	Human	Pro-Ser-Ser^{321}-Ser-Pro-Ile, Arg-Cys-Phe^{338}-Phe-Gly-Ala	79
PMCA-2	Human	Gly-Leu-Asn^{1135}-Arg-Ile-Gln, Glu-Leu-Arg^{1124}-Arg-Gly-Gln	80
RyR1	Rabbit	Met-Met-Thr^{1400}-Gln-Pro-Pro, Ile-Ser-Gln^{2844}-Thr-Ala-Gln	81
Spectrin aII	Human	Glu-Val-Tyr^{1176}-Gly-Met-Met	57
Spectrin b	Human	Lys-Ser-Thr^{2058}-Ala-Ser-Trp	57
Talin	Human	Val-Leu-Gln^{433}-Gln-Gln-Tyr	82
Tau	Human	Glu-Val-Met^{11}-Glu-Asp-His, Gly-Leu-Lys^{44}-Glu-Ser-Pro, Val-Val-Arg^{230}-Thr-Pro-Pro, Asp-Leu-Lys^{254}-Asn-Val-Lys, Asn-Val-Lys^{257}-Ser-Lys-Ile, Asn-Leu-Lys^{267}- His-Gln-Pro, Ile-Val-Tyr^{310}- Lys-Pro-Val, Glu-Val-Lys^{340}-Ser-Glu-Lys, Ile-Val-Tyr^{394}-Lys-Ser-Pro	83
Tyrosine 3-hydroxylase	Bovine	Ala-Ile-Met^{29}-Ser-Pro-Arg, Glu-Leu-Asp^{22}-Ala-Lys-Gln	84
Vimentin	Murine	Arg-Leu-Arg^{70}-Ser-Ser-Val, Gly-Ser-Gly^{18}-Thr-Ser-Ser, Gly-Thr-Ser^{20}-Ser-Arg-Pro, Val-Thr-Thr^{32}-Ser-Thr-Arg, Arg-Thr-Tyr^{37}-Ser-Leu-Gly, Ser-Leu-Gly^{40}-Ser-Ala-Leu, Ser-Leu-Tyr^{52}-Ser-Ser-Ser, Val-Thr-Arg^{63}-Ser-Ser-Ala	85
von Willebrand factor	Human	Leu-Leu-Lys^{1913}-Ser-His-Arg, Ser-Lys-Arg^{763}-Ser-Leu-Ser	86

[a] In the original report, the cleavage site is erroneously given as 75–76, but in reality its position in the sequence is 90–91.
[b] In the original study, rabbit dystrophin is reported but only the human sequence is available.
[c] P. Friedrich, M. Pulai, and P. Tompa, unpublished observations.

mutations for either increasing or decreasing sensitivity to calpain attack. This feature is demonstrated by inhibitory mutations in the autolysis sites of *Drosophila* calpain B. Fig. 3 shows that mutations hamper autolytic cleavage of the given site: amino acids were replaced on either one, or both sides of the two major autolytic sites with amino acids not accepted by

TABLE II
Amino acid preferences of calpain around the cleavage site

The occurrence of each residue in positions P4–P7′ for all 106 cleavage sites (Table I) has been counted and normalized to the frequency of the same amino acid in Swiss-Prot and TrEMBL.

	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'	P_5'	P_6'	P_7'
Trp	0.00	**2.34**[a]	0.00	0.00	0.00	0.00	**3.12**[a]	0.79	0.00	**2.38**[a]	0.79
Tyr	0.90	0.00	0.30	**2.69**[a]	0.00	0.30	0.60	0.60	1.81	1.52[b]	0.30
Phe	0.23	0.23	0.23	1.84	0.46	0.46	0.23	1.39	0.00	0.94	1.17
Leu	0.59	0.59	**2.97**[a]	0.59	0.59	0.40	0.20	0.40	0.90	0.60	1.01
Ile	0.32	1.13	0.81	0.16	0.65	0.65[b]	0.97	0.33	0.33	0.33	0.99
Val	0.71	1.57	**2.00**[a]	0.14	0.71	0.71	0.57	1.58	1.01	1.16	0.29
Ala	1.11	0.74	0.12	0.25	1.73	1.36	0.99	1.00	0.75	0.88	0.88
Gly	0.69	0.83	0.14	1.24	0.69	1.24	0.55	1.25	1.53	0.84	1.26
Ser	1.20	1.60	0.93	0.80	**3.33**	1.87	0.80	1.08	1.88	1.36	1.49
Thr	**2.03**[a]	1.02	**2.71**[a]	1.86	1.86[b]	0.68	0.85	1.71	1.03	1.55	0.69
Pro	1.54	**2.32**[a]	0.77	0.58	0.19	**2.89**[a]	**5.59**[a,b]	**2.92**[a,b]	1.75	1.77	1.18
Asp	0.72	0.36	0.54	0.36	0.18	0.72	0.36	0.73	0.73	0.92	0.73
Glu	1.31	1.60	0.73	0.44	0.73	1.31	0.73	0.29	0.88	0.89	0.89
Gln	1.68	0.48	0.72	0.96	1.20	1.68	1.44	0.97	0.48	1.22	0.73
Asn	1.08	0.65	1.30	0.65	0.43	0.65	0.65	1.53	1.75	0.88	0.66
Lys	1.11	1.42	0.47	**2.84**[a]	1.11	1.42	1.42	0.96	0.96	0.81	1.45
Arg	1.45	1.09	0.36	**2.54**[a]	1.82	0.36	0.91	0.55	0.92	0.74	**2.22**[a,b]
His	1.68	0.84	0.00	1.26	0.42	0.42	0.84	0.85	0.85	0.85	0.85
Cys	0.00	0.00	1.75	0.00	0.00	0.58	0.00	0.59	1.18	**2.37**	0.00
Met	0.80	0.40	0.40	1.59	0.80	0.40	1.19	0.00	0.00	0.00	1.22

[a] Values exceeding 2.00 are in bold.
[b] Values corresponding to the consensus calpastatin inhibitory segment in positions P_1'-P_7' (i.e. TIPPXYR) are underlined.

calpains (22). The following changes were made: the first autolytic site Ala-Gln-Asn^{74}-Ala-Ser-Tyr was mutated to Ala-Gly-Val^{74}-Pro-Ser-Tyr; the second autolytic site Pro-Glu-Asn^{223}-Gln-Asn-Met was changed to Pro-Glu-Ala^{223}-Val-Asn-Met. These modified calpain B forms were expressed in *E. coli*, purified, and tested for Ca^{2+}-induced autolysis. As a result of these mutations, autolysis at the original site was arrested, and its site shifted to new, nearby peptide bonds. In the case of site 223, two new sites emerged: one at Phe^{215}–Thr^{216} and one at Gly^{230}–Arg^{231}, as identified by N-terminal sequencing. In the case of site 74, altered autolysis gave rise to low intensity, blurred band(s), which could not be analyzed by sequencing. These new sites were more sluggish than the original ones: the activation lag phase with the "autolysis mutant" was longer than with wild-type calpain B (data not shown).

The FRET Substrate—The third practical application of the preference matrix of calpain cleavage is to design a novel and optimized substrate, offering an improvement in both *in vitro* and *in vivo* experiments. The amino acids with the highest scores in the matrix were selected, and an 11-mer peptide with TPLKSPPPSPR sequence was constructed. A fluorescent donor (EDANS) and quenching acceptor (DABCYL) were attached to its C- and N-terminal ends, respectively, yielding the novel substrate based on the FRET principle. As shown in Fig. 4*A*, before cleavage by calpain, the level of fluorescence is low and calpain cleavage results in an about 3-fold increase in fluorescence intensity. This fluorescence change is enough to be followed in time course measurements. Judging from the kinetic parameters shown in Fig. 4*B*, the FRET substrate is cleaved by m-calpain with a much better efficacy than the widely used fluorescent dipeptide substrate (LY-AMC) for which the K_m and k_{cat} values were 2.21 mM and 0.092 liter/s, respectively (7). These measurements were done with the *Drosophila* calpains, calpain A and calpain B as well giving the following kinetic values: K_m was 0.12 mM, k_{cat} was 0.04 liter/s for calpain A; 0.25 mM and 0.17 1iter/s for calpain B, respectively. The other proteases showed variable activity toward the FRET substrate; the kinetic parameters proved to be the best for m-calpain (Fig. 4*B*, *inset*).

Next we compared the FRET substrate and the dipeptide substrate with respect to their efficiency in detecting calpain activity in COS-7 cell lysates. Substrate cleavage was followed with and without 20 mM Ca^{2+}. As shown in Fig. 5, the dipeptide gave a very high background activity, which led to a marginal (7%) difference between activities measured in the absence and presence of calcium. In contrast, almost no background activity was measured with the FRET substrate and the difference in the fluorescent change was four times higher. This difference is sufficient to follow subtle alterations in calpain activity.

We also tested the applicability of this novel substrate in an *in vivo* system. The Cellfectin lipofection system was used to transfer the substrate into living *Drosophila* S2 cells and to check the emission spectrum of lipofected cells before and after the addition of 10 μM ionomycin (Fig. 6*A*). The spectra are similar to those in the *in vitro* measurements (Fig. 4*A*), thus proving that the method is appropriate for assaying the calpain activity *in vivo*. The applicability of the FRET substrate *in vivo* was further supported by time course measurements (Fig. 6*B*). As shown, the fluorescence intensity of the cells did not increase without the addition of ionomycin, *i.e.* without calcium there is no cleavage of the substrate. The slight decrease of the fluorescence intensity was due to the slow sedimentation of cells. After the addition of 10 μM ionomycin, the fluorescence intensity increased, indicating calpain activation and substrate cleavage. This change was truly due to the activity of the endogenous calpains of *Drosophila* cells, as calpain inhibitors (calpain inhibitor II and E-64) abolished any detectable fluorescence intensity change, when added to the cell suspension either at the start or during the measurement (E-64). No calpain activity could be detected either, when calpastatin was added in 15 μM concentration to the medium during lipofection.

It should be noted that a similar fluorogenic substrate was produced previously for assaying retroviral proteases (21); however, that peptide was only an 8-mer with different sequence. Despite the longer distance between the two fluorophores, the FRET substrate described here proved to be efficient to monitor calpain activity.

DISCUSSION

Calpains play important roles in many different physiological and pathological processes, and the mechanisms underlying their substrate recognition have been studied. Nevertheless it has remained an open question whether local or global sequential or other, higher order, signals are recognized and accepted

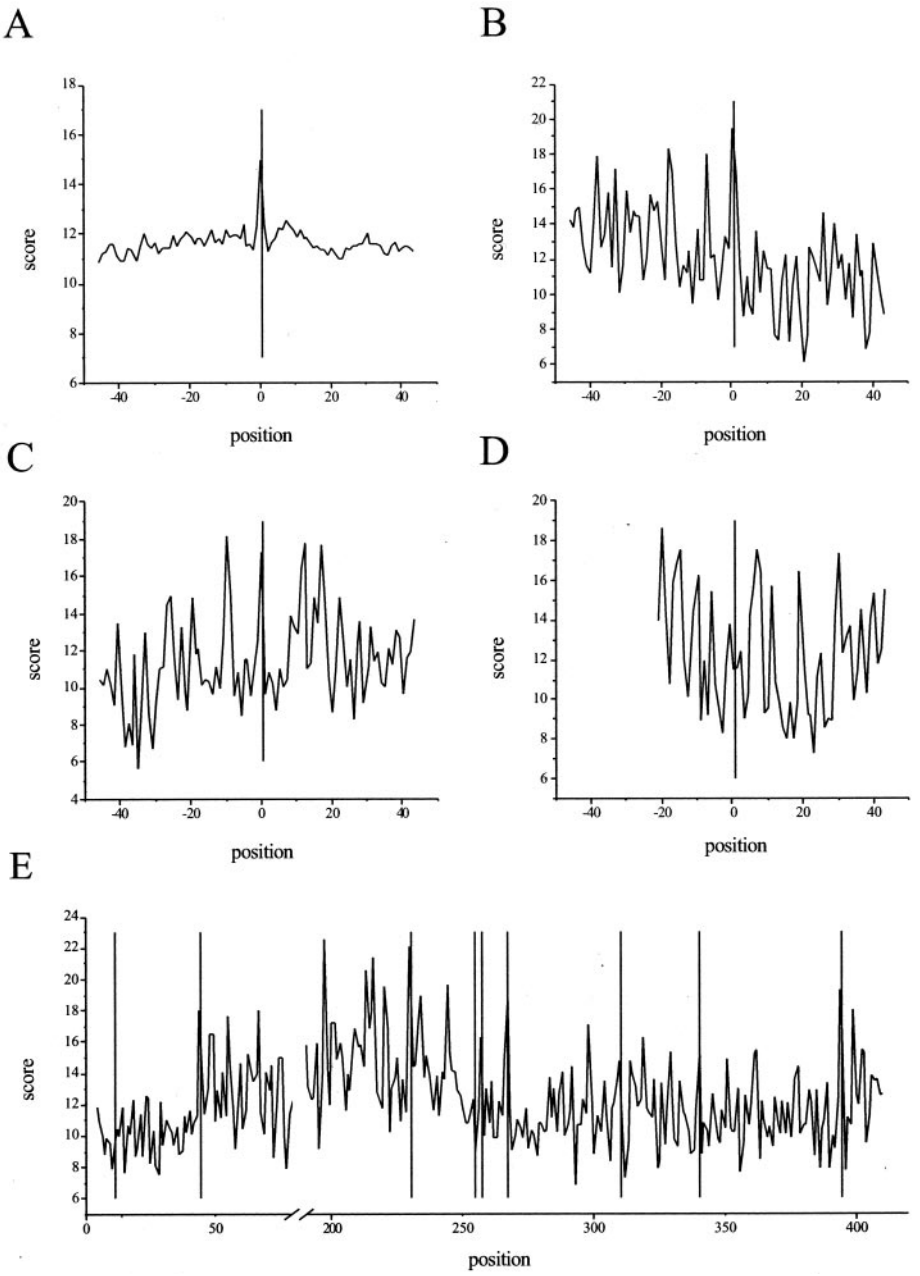

FIG. 1. **Profile scores of selected substrates.** Each substrate sequence has been scored with a preference matrix calculated, such as that in Table II except that the actual substrate was always left out. The resulting profile scores have been averaged (*A*) or are shown for the primary site of vimentin (*B*) and von Willebrand factor (*C*), and for the sole site of p53 (*D*). For tau, a composite profile score covering all 9 cleavage sites has been plotted (*E*). For creating this plot, adjacent scans were truncated halfway between their cleavage sites and fit together. The sites of cleavage are marked by *vertical lines*.

by calpains. Early studies on peptides and a limited number of proteins have led to the concept that calpain shows recognizable sequential preferences only in the P_2 and P_1 position of the substrate (7, 8). These preferences, however, do not apply to all substrates, which points to the importance of global structural clues in recognition (5, 11). Further research has led the researchers to conclude that calpain may need PEST or PEST-like sequences to recognize a protein as its substrate, but a susceptible cleavage site is also necessary for cleavage to occur (5). In this review the authors also referred to the P_2-P_1 rule as the site of calpain cleavage, allowing the possibility that other restrictions might exist. A comprehensive analysis of cleavage sites within potentially physiological substrate proteins has not been carried out.

Our statistics on 106 protein sites shows that amino acid preferences extend over 11 positions around the scissile bond: these data corroborate, correct, and extend the classical P_2-P_1 rule. We found that in positions P_2 and P_1 the prevalent view is correct with the minor correction that Thr is more preferred than Val in P_2. In P_1', a very strong, and so far missed, preference for Ser is seen. An entirely unexpected feature is the frequent occurrence of Pro in positions P_3 and P_2'-P_4': the highest value in the entire preference matrix is 5.59 for Pro in P_3'. Overall, the primary structure around the scissile bond is probably more important in calpain recognition than previously thought. The correlation of cleavage sites with profile values scored by the preference matrix highly supports this conclusion. Distinct aspects of these observations provide novel mechanistic insight into the process of substrate recognition by calpains, as follows.

Because Pro is the poorest α-helix- and β-sheet-forming residue (28), and it tends to dominate intrinsically unstructured/disordered regions of proteins (29, 30), the observed preferences imply an affinity of calpain for disordered regions in proteins, especially on the C-terminal side of the scissile bond. Predictions of the secondary structure also support this view: the predicted α-helix and β-sheet potential drops very low in the C-terminal decapeptide of the cleavage site suggesting that one major structural attribute calpain is likely to recognize is protein disorder. Proteases in general tend to cleave their sub-

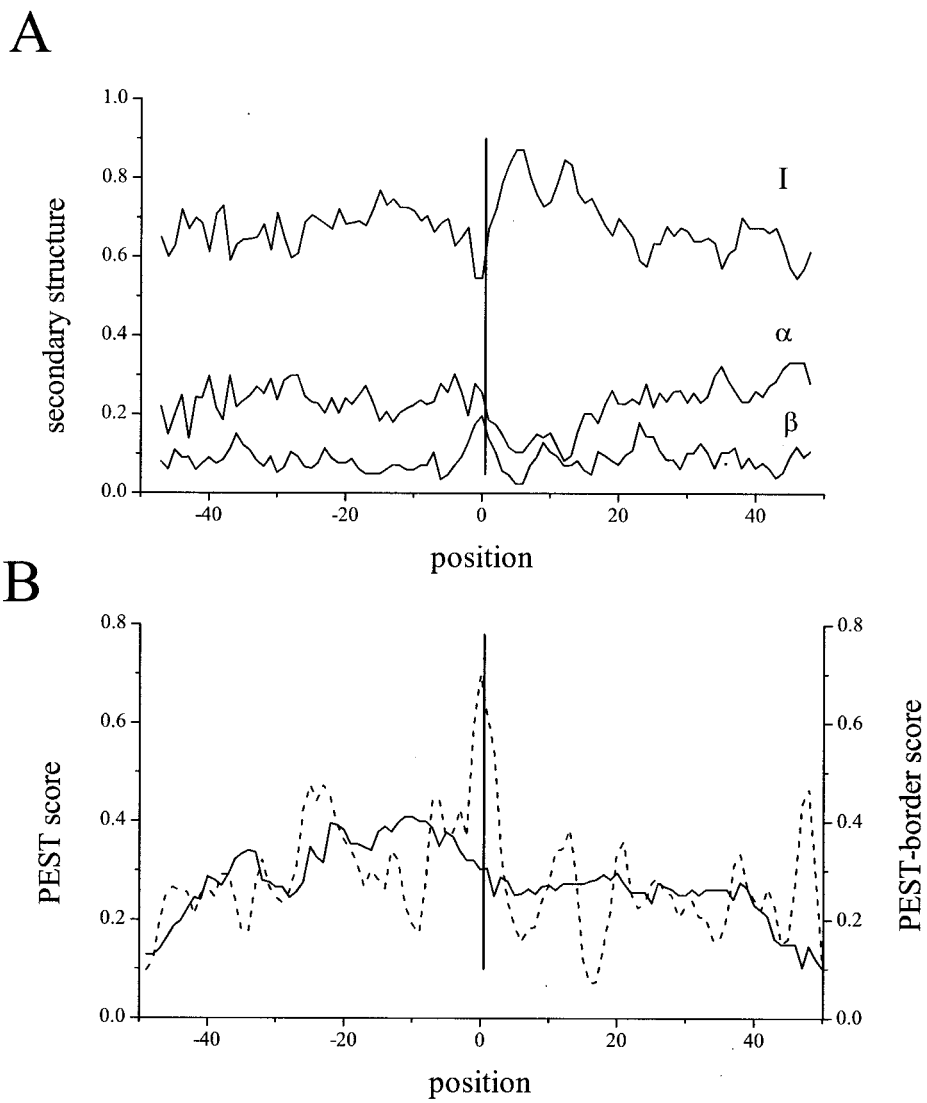

FIG. 2. **Structural clues of calpain cleavage.** Secondary structural elements, *i.e.* α-helix, β-sheet, and irregular (*I*) structure (*A*) and PEST segments (*B*, *continuous line*) have been predicted via online servers and averaged for the 100-residue segments around all 106 calpain cleavage sites. An arbitrary PEST border score has also been calculated and is shown by a *dashed line* (*B*). *Vertical lines* mark the site of cleavage between positions 0 and 1. The URL address of the servers and the definition of the PEST border score are given in "Materials and Methods" under "Data Collection and Analysis."

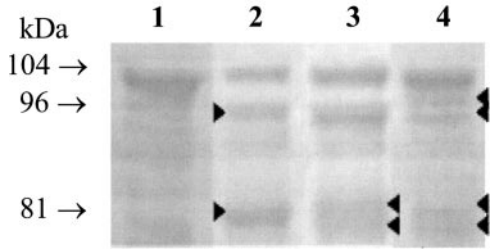

FIG. 3. **Autolysis of wild type and autolysis-site mutant calpain B forms.** *Drosophila* calpain B autolysis reactions were run for 10 min in the presence of Ca^{2+}, and then samples were run on an SDS gel. *1*, calpain B control; *2*, calpain B where the two autolytic products are indicated by *arrowheads*; *3*, calpain B N223Q → A223V mutant (81-kDa autolytic site), the 81-kDa autolytic product split into two bands, as indicated by *arrowheads*; *4*, autolyzed calpain B QN74A → GV74P, N223Q → A223V double site mutant, both autolytic products are split into two bands, as indicated by *arrowheads*.

strates in sterically accessible and flexible regions (31, 32), but this preference seems particularly important for calpain. This may be an important component of its regulatory function, *i.e.* that often it separates regulatory and catalytic domains of enzymes causing activation of the latter (1–3). Besides lowering the occurrence of ordered secondary structural elements, prolines may also play an active role in this enzyme-substrate recognition: Pro-rich regions are known to be directly involved in protein-protein interactions, such as with SH3 domains of regulatory proteins in signal transduction (33, 34). This aspect of the recognition of substrate proteins by calpain, *i.e.* the possible interference of calpain with signal transduction is worthy of consideration. This may also ensue from the high frequency of the (S/T)P dipeptide in the P_1'-P_2' and P_2'-P_3' positions (5–10× the average in Swiss-Prot and TrEMBL) in calpain substrates. This dipeptide is the consensus site of Pro-directed kinases, such as mitogen-activated protein kinase, GSK3, and cyclin-dependent kinases, which play central roles in signal transduction (35–37). This raises the possibility of the regulatory interplay of protein phosphorylation and calpain-mediated activation in signal transduction; phosphorylation by Pro-directed and other kinases does in fact interfere with calpain action (38, 39).

Another notable aspect of the observed preferences is the unmistakable similarity of the conserved inhibitory segment of the endogenous, specific calpain inhibitor, calpastatin (TIP-P*X*YR) to the sequence of the ideal substrate on the C-terminal side of the scissile bond (see Table II). This implies that calpastatin may bind to the S_1'-S_7' pocket of the enzyme as a substrate analog, in perfect agreement with the observed mode of competitive inhibition of the enzyme (40). An old and intriguing observation concerning calpastatin action can also be rationalized via this similarity. Calpastatin has four homologous inhibitory domains, each of which is capable of inhibiting one

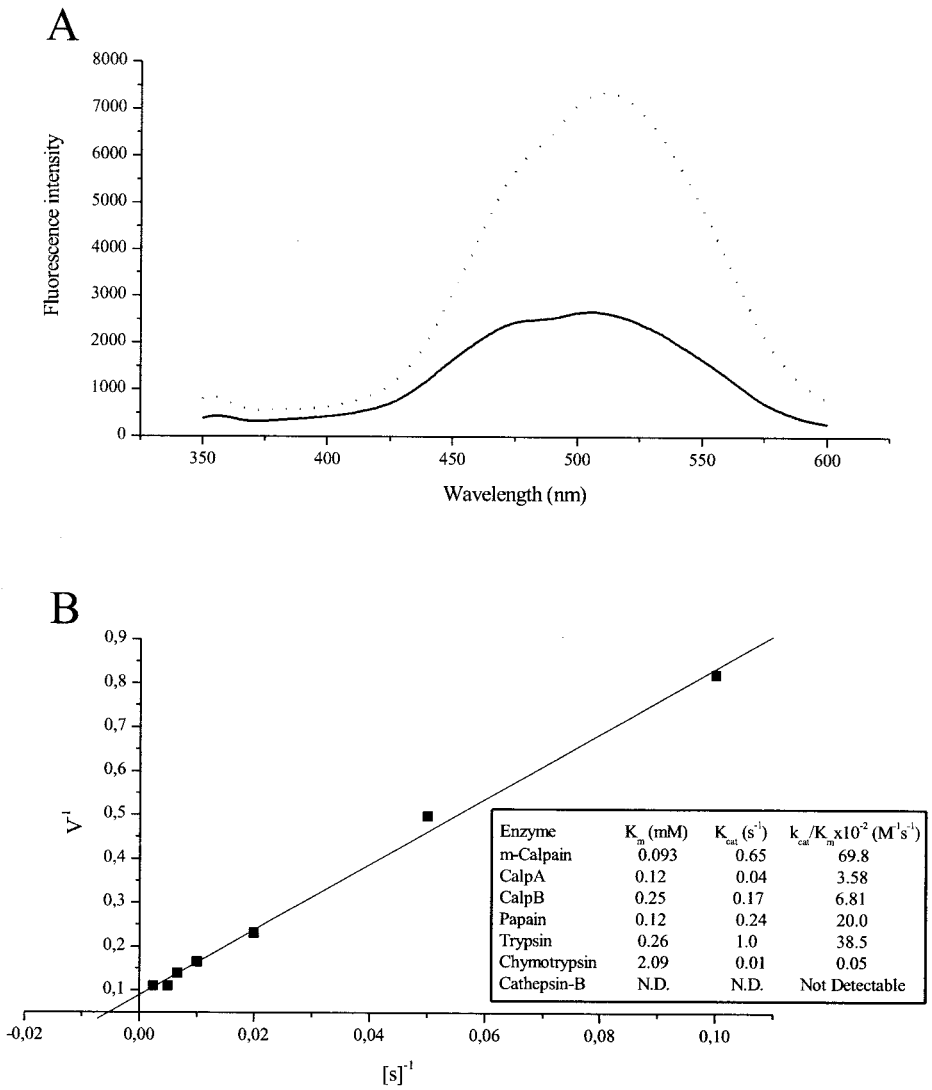

FIG. 4. **Calpain activity measurement with the FRET substrate *in vitro*.** *A*, digestion of the FRET substrate by m-calpain *in vitro*. 400 μM FRET substrate was digested with 0.5 μM m-calpain in calpain buffer for 40 min, and the fluorescence spectrum of the substrate was recorded before (*solid line*) and after (*dashed line*) digestion. The excitation wavelength was 320 nm. *B*, Lineweaver-Burk plot of m-calpain kinetics with the FRET substrate. The FRET substrate was added to the reaction mixture (containing 5 mM free $CaCl_2$ and 0.2 μM m-calpain,) in various concentrations and the initial rate of fluorescence change was determined. This rate was normalized to the total change that was attained at high enzyme concentration and long incubation time, to calculate absolute V_0 and k_{cat} values. *Inset*, comparison of kinetic parameters for some proteases with the FRET substrate. Calcium concentrations used were: 5 mM free $CaCl_2$ in the case of m-calpain, 19 mM free $CaCl_2$ in the case of calpain A and calpain B in calpain buffer and no $CaCl_2$ in the case of the other proteases. The following enzyme concentrations were used: 0.2 μM m-calpain, 4.8 μM calpain A, 1.8 μM calpain B, 1.0 μM papain, 1.0 μM trypsin, 40.0 μM chymotrypsin, and 10.0 μM cathepsin B. Excitation and emission wavelengths were 380 and 460 nm, respectively, in the case of the LY-AMC and 320 and 480 nm for the FRET substrate.

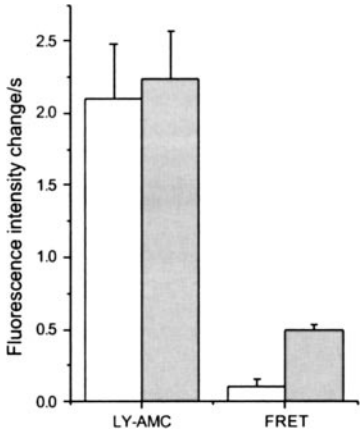

FIG. 5. **Calpain activity in COS-7 cell lysates.** COS-7 cell lysates were prepared as described under "Materials and Methods." Calpain activity was followed by fluorometric measurements. Either 5 mM EDTA (*white columns*) or 20 mM free $CaCl_2$ (*gray columns*) was added to 30-μl lysates. The change in fluorescence intensity was monitored with LY-AMC (1 mM) or the FRET substrate (100 μM). Each column represents the mean values of four independent measurements.

calpain molecule. Still, the inhibition is perfect only up to one equivalent of the enzyme, above which the inhibitor slowly starts degrading, way before the theoretical equivalence point of 4:1 is reached (41, 42). In the four homologous inhibitory domains of human calpastatin, Val (domain I), Glu (domains II and III), and Asp (domain IV) can be found N-terminal to the conserved TIPP*X*YR motif. According to our statistics, Val (0.14) practically never occurs in the P_1 position of calpain substrates, whereas Glu (0.44) and Asp (0.36) are moderately acceptable. Judged by that, one would expect that only domain I of calpastatin acts as a perfect competitive inhibitor, and the other three behave more like a suicide substrate. In fact, separate domains of calpastatin were found not to be equally effective: domain I is the most effective and domain III is the weakest (27).

The role of PEST regions in calpain cleavage may also need a reappraisal in light of our data. As mentioned, this issue is still not settled, as mutations of PEST regions either abrogate substrate sensitivity to calpain (16, 17) or have no effect at all (18, 19). It appears that PEST regions, or rather the PEST score, is not directly linked to calpain action. This also follows from the fact that a significant fraction of calpain substrates have no PEST region at all (19); here we found that close to half the sites have no PEST region within 50 residues in either direction. On the other hand, the position of a PEST region in substrates, where it occurs at all, is not random with respect to the site of calpain cleavage: most often it occurs at the C-terminal end. We reckon that the PEST score does express something important for calpain, perhaps the lack of structure also manifest in the high frequency of prolines, as discussed

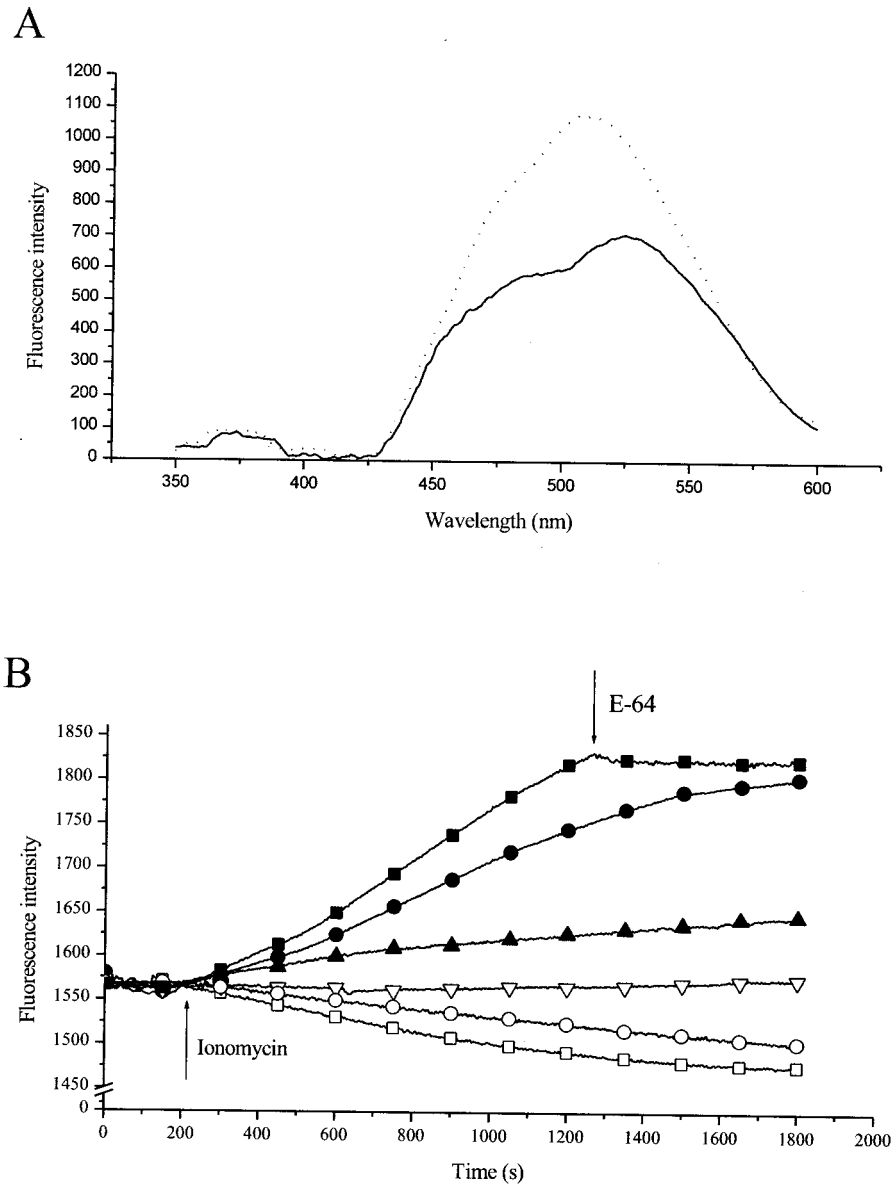

FIG. 6. **Recording calpain activity in *Drosophila* S2 cells.** *A*, emission spectrum of the FRET substrate in S2 cells. S2 cells were suspended in *Drosophila* SFM medium at a 3×10^6 cell/ml cell number and were lipofected with 200 μM FRET substrate. Calpain activity was induced with the addition of 10 μM ionomycin. *Solid line*, difference spectrum of lipofected and non-lipofected cells before the addition of ionomycin. *Dashed line*, difference spectrum of lipofected and non-lipofected cells 40 min after the addition of ionomycin. *B*, time course of calpain activity in S2 cells. S2 cells were suspended in *Drosophila* SFM medium at a 3×10^6 cell/ml cell number and were lipofected with 200 μM FRET substrate. *In vivo* calpain activity was measured in a 3 × 3 mm quartz cuvette with a Jasco FP777 fluorometer. Calpain activity was induced by the addition of 10 μM ionomycin, except for the control measurement, where no ionomycin was added. Cells were measured for 30 min, and each measurement was repeated at least three times. Symbols: □, S2 cells without ionomycin; ●, S2 cells with 10 μM ionomycin; ▽, S2 cells with 10 μM ionomycin and 500 nM calpain inhibitor II (the inhibitor was added to the cells immediately before the start of the measurement); ▲, S2 cells with 10 μM ionomycin and 500 nM E-64 (the inhibitor was added to the cells immediately before the start of the measurement); ○, S2 cells lipofected with calpastatin and the FRET substrate; ■, S2 cells with 10 μM ionomycin and 500 nM E-64 (the inhibitor was added 20 min after starting the measurement).

previously (29). Furthermore, PEST regions have flanking positive charges, whereas calpain shows a strong preference for Lys and Arg in P_1, which is another reason for correlation to occur.

Our statistical analysis also entails important applications in calpain studies. Probably the most important dividend of the preference matrix is that it significantly improves our capacity to manipulate calpain-substrate systems. One of the practical aspects is that it enables us to design point mutations, which interfere with the action of calpain in very specific ways. For example, substituting Val (Ile) at P_1 or Pro at P_1' for the actual residue would probably seriously impair cleavage by calpain, without much interfering with the structure and function of the substrate in any other way. We used this approach to design two mutant calpain B forms, both having changed amino acids at the site of autolysis. As shown by our results, these mutant calpain B forms lacked the ability of normal autolysis, thus proving that it is indeed possible to interfere with calpain cleavage. The fact that new autolysis sites appeared suggests that calpain cleavage site might not be strictly fixed, and as a response to changes in the substrate molecule other, less preferred sites may be promoted to cleavage sites. This point seems to be supported by a recently published work (43) where the researchers observed abnormal calpain cleavage of the fibrillized α-synuclein. *In vitro*, α-synuclein was cleaved after amino acid 53, whereas fibrillized α-synuclein was cleaved in the region of amino acid 120.

In apparent contrast, it was observed that an FF mutation at the m-calpain autolysis site completely arrested autolysis (44). It should be remembered, however, that the N-terminal segment of m-calpain is very short (9 amino acids) compared with the very large (224 amino acids) and disordered N-terminal segment of calpain B (22). It is likely that on the N-terminal segment of m-calpain only a simple site is sterically available to calpain attack.

All these facts seem to support the observation that apart from sequential clues calpains recognize other, higher order structural elements as well. Applying the matrix to other, known calpain substrates allows us to design and synthesize other mutant proteins as well, thus gaining a deeper insight of the calpain-substrate systems.

The second practical aspect of our findings is that by using this matrix we managed to construct a substrate, which is superior to the commercially available ones. Since the widely used fluorescent dipeptide substrate (LY-AMC) is not specific to calpain and is quite poor in kinetic terms, there have been

attempts to create better substrates for specific applications. For the lack of a clear insight into the mechanism of calpain cleavage, researchers have been left with the only possibility of using the sequences of known calpain substrates. One of the preferred choices for this purpose is α-spectrin, a protein that is specifically cleaved by calpain. Vanderklish *et al.* (45) constructed a vector encoding for an 11-amino acid long sequence of α-spectrin flanked by two fluorescent protein domains to serve as an *in vivo* FRET substrate. This construct led to identifying dendritic spines in which calpain is activated during the excitatory synaptic activity. Another group followed a similar approach to design two FRET substrates using the same α-spectrin sequence (46). In a previous attempt we used a different protein substrate of calpain, MAP2, labeled with DTAF to be able to monitor calpain continuously (47). While all three approaches worked well under the given experimental circumstances, they did not provide a substrate, which is easy to construct and handle and is appropriate to measure calpain activity in different environments, *in vivo* and *in vitro*. The new FRET substrate tested in this study is designed on the basis of the analysis of more than 100 calpain cleavage sites and thus presents an ideal calpain cleavage site. This substrate is water soluble, its synthesis is relatively simple, and its use is reliable and accurate. It is to be noted that the substrate primarily targets μ- and m-calpains, the cleavage preferences of which its development has been based on. Its kinetic preference for these mammalian forms *versus* Drosophila calpains (Fig. 4*B*) points to its probable specificity for these ubiquitous calpain forms. Given the conservation of the sequence preference of calpains (9), this result probably translates to μ- and m-calpains from other species. With respect to other mammalian isoforms (*e.g.* p94, calpain 10), the utility of the FRET substrate cannot be ascertained for the lack of sufficient data. Further work will be needed to clarify this point.

Detecting calpain activity in crude cell and tissue extracts has always been a great challenge, because the commercial substrates are cleaved by many other proteases as well, and they are quite poor calpain substrates. These peptides may even escape calpain cleavage, since in a crude extract, other, much better calpain substrates abound and these peptides are incapable of successfully competing with them. All these result in a high background activity, making the correct measurement of the calpain activity very difficult. Although the FRET substrate is not absolutely specific, in particular with respect to trypsin and papain (Fig. 4*B*), it represents a significant development in our capacity of assessing calpain activity *in vitro* an *in vivo*: as demonstrated by COS-7 cell lysates, one may expect much less interference by other proteases than with the commercial dipeptide substrate. In fact, the background activity is close to zero with the FRET substrate. The FRET substrate is a much stronger competitor of the endogenous substrates than the commercial peptide substrates hence it is capable to detect the actual calpain activity more precisely, and to give information about weak calpain inhibition or activation.

In summary, our data show that various functions such as amino acid preferences, secondary structure and PEST segments correlate with the cleavage site selected by calpain. These correlations and the success of applying the preference matrix for practical purposes suggest that the information encoded in the primary structure is important for calpain selecting its site of cleavage and underscore that by carefully placed point mutations we might interfere with calpain action in a very specific manner even in intact cells or tissues.

REFERENCES

1. Sorimachi, H., Ishiura, S., and Suzuki, K. (1997) *Biochem. J.* **328,** 721–732
2. Ono, Y., Sorimachi, H., and Suzuki, K. (1998) *Biochem. Biophys. Res. Commun.* **245,** 289–294
3. Carafoli, E., and Molinari, M. (1998) *Biochem. Biophys. Res. Commun.* **247,** 193–203
4. Huang, Y., and Wang, K. K. (2001) *Trends Mol. Med.* **7,** 355–362
5. Wang, K. K., Villalobo, A., and Roufogalis, B. D. (1989) *Biochem. J.* **262,** 693–706
6. Wang, K. K. W., and Yuen, P.-w. (1999) in *Calpain: Pharmacology and Toxicology of Calcium-dependent Protease* (Wang, K. K. W., and Yuen, P.-w., eds), pp. 77–101, Taylor and Francis, Philadelphia
7. Sasaki, T., Kikuchi, T., Yumoto, N., Yoshimura, N., and Murachi, T. (1984) *J. Biol. Chem.* **259,** 12489–12494
8. Hirao, T., and Takahashi, K. (1984) *J. Biochem.* (*Tokyo*) **96,** 775–784
9. Takahashi, K. (1990) in *Calpain Substrate Specificity* (Mellgren, R. L., and Murachi, T., eds) CRC Press, Boca Raton
10. Stabach, P. R., Cianci, C. D., Glantz, S. B., Zhang, Z., and Morrow, J. S. (1997) *Biochemistry* **36,** 57–65
11. Sakai, K., Akanuma, H., Imahori, K., and Kawashima, S. (1987) *J. Biochem.* (*Tokyo*) **101,** 911–918
12. Banik, N. L., Chou, C. H., Deibler, G. E., Krutzch, H. C., and Hogan, E. L. (1994) *J. Neurosci. Res.* **37,** 489–496
13. Rogers, S., Wells, R., and Rechsteiner, M. (1986) *Science* **234,** 364–368
14. Rechsteiner, M., and Rogers, S. W. (1996) *Trends Biochem. Sci.* **21,** 267–271
15. Molinari, M., Maki, M., and Carafoli, E. (1995) *J. Biol. Chem.* **270,** 14576–14581
16. Noguchi, M., Sarin, A., Aman, M. J., Nakajima, H., Shores, E. W., Henkart, P. A., and Leonard, W. J. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11534–11539
17. Shumway, S. D., Maki, M., and Miyamoto, S. (1999) *J. Biol. Chem.* **274,** 30874–30881
18. Molinari, M., Anagli, J., and Carafoli, E. (1995) *J. Biol. Chem.* **270,** 2032–2035
19. Carillo, S., Pariat, M., Steff, A., Jariel-Encontre, I., Poulat, F., Berta, P., and Piechaczyk, M. (1996) *Biochem. J.* **313,** 245–251
20. Wang, K. K. (2000) *Trends Neurosci.* **23,** 20–26
21. Matayoshi, E. D., Wang, G. T., Krafft, G. A., and Erickson, J. (1990) *Science* **247,** 954–958
22. Farkas, A., Tompa, P., Schad, E., Sinka, R., Jekely, G., and Friedrich, P. (2004) *Biochem. J.* **378,** 299–305
23. Elce, J. S., Hegadorn, C., Gauthier, S., Vince, J. W., and Davies, P. L. (1995) *Protein Eng.* **8,** 843–848
24. Jekely, G., and Friedrich, P. (1999) *J. Biol. Chem.* **274,** 23893–23900
25. Yang, H. Q., Ma, H., Takano, E., Hatanaka, M., and Maki, M. (1994) *J. Biol. Chem.* **269,** 18977–18984
26. Laemmli, U. K. (1970) *Nature* **227,** 680–685
27. Emori, Y., Kawasaki, H., Imajoh, S., Minami, Y., and Suzuki, K. (1988) *J. Biol. Chem.* **263,** 2364–2370
28. Adzhubei, A. A., and Sternberg, M. J. (1993) *J. Mol. Biol.* **229,** 472–493
29. Tompa, P. (2002) *Trends Biochem. Sci.* **27,** 527–533
30. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) *J. Mol. Graph. Model.* **19,** 26–59
31. Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. (1986) *Biochemistry* **25,** 1847–1851
32. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) *Fold Des.* **2,** R17-R26
33. Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. (1994) *Cell* **76,** 933–945
34. Ohba, T., Ishino, M., Aoto, H., and Sasaki, T. (1998) *Biochem. J.* **330,** 1249–1254
35. Kennelly, P. J., and Krebs, E. G. (1991) *J. Biol. Chem.* **266,** 15555–15558
36. Nigg, E. A. (2001) *Nat. Rev. Mol. Cell. Biol.* **2,** 21–32
37. Pearson, G., Robinson, F., Beers Gibson, T., Xu, B. E., Karandikar, M., Berman, K., and Cobb, M. H. (2001) *Endocr. Rev.* **22,** 153–183
38. Litersky, J. M., and Johnson, G. V. (1995) *J. Neurochem.* **65,** 903–911
39. Alexa, A., Tompa, P., Baki, A., Vereb, G., and Friedrich, P. (1996) *J. Neurosci. Res.* **44,** 438–445
40. Maki, M., Takano, E., Osawa, T., Ooi, T., Murachi, T., and Hatanaka, M. (1988) *J. Biol. Chem.* **263,** 10254–10261
41. Mellgren, R. L., Mericle, M. T., and Lane, R. D. (1986) *Arch. Biochem. Biophys.* **246,** 233–239
42. Nakamura, M., Inomata, M., Imajoh, S., Suzuki, K., and Kawashima, S. (1989) *Biochemistry* **28,** 449–455
43. Mishizen-Eberz, A. J., Guttmann, R. P., Giasson, B. I., Day, G. A., 3rd, Hodara, R., Ischiropoulos, H., Lee, V. M., Trojanowski, J. Q., and Lynch, D. R. (2003) *J. Neurochem.* **86,** 836–847
44. Elce, J. S., Hegadorn, C., and Arthur, J. S. (1997) *J. Biol. Chem.* **272,** 11268–11275
45. Vanderklish, P. W., Krushel, L. A., Holst, B. H., Gally, J. A., Crossin, K. L., and Edelman, G. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2253–2258
46. Mittoo, S., Sundstrom, L. E., and Bradley, M. (2003) *Anal. Biochem.* **319,** 234–238
47. Tompa, P., Schád, E., Baki, A., Alexa, A., Batke, J., and Friedrich, P. (1995) *Anal. Biochem.* **228,** 287–293
48. Villa, P. G., Henzel, W. J., Sensenbrenner, M., Henderson, C. E., and Pettmann, B. (1998) *J. Cell Sci.* **111,** 713–722
49. Ando, Y., Imamura, S., Hong, Y. M., Owada, M. K., Kakunaga, T., and Kannagi, R. (1989) *J. Biol. Chem.* **264,** 6948–6955
50. Azarian, S. M., King, A. J., Hallett, M. A., and Williams, D. S. (1995) *J. Biol. Chem.* **270,** 24375–24384
51. Crawford, C., Willis, A. C., and Gagnon, J. (1987) *Biochem. J.* **248,** 579–588
52. McGinnis, K. M., Whitton, M. M., Gnegy, M. E., and Wang, K. K. (1998) *J. Biol. Chem.* **273,** 19993–20000
53. Kakkar, R., Raju, R. V., and Sharma, R. K. (1998) *Arch. Biochem. Biophys.*

358, 320–328
54. Chua, B. T., Guo, K., and Li, P. (2000) *J. Biol. Chem.* **275,** 5131–5135
55. Pariat, M., Salvat, C., Bebien, M., Brockly, F., Altieri, E., Carillo, S., Jariel-Encontre, I., and Piechaczyk, M. (2000) *Biochem. J.* **345,** 129–138
56. Lin, J. S., Fitzgerald, S., Dong, Y., Knight, C., Donaldson, P., and Kistler, J. (1997) *Eur. J. Cell Biol.* **73,** 141–149
57. Seubert, P., Baudry, M., Dudek, S., and Lynch, G. (1987) *Synapse* **1,** 20–24
58. David, L. L., and Shearer, T. R. (1993) *FEBS Lett.* **324,** 265–270
59. Yoshida, M., Suzuki, A., Shimizu, T., and Ozawa, E. (1992) *J. Biochem.* (*Tokyo*) **112,** 433–439
60. Gregoriou, M., Willis, A. C., Pearson, M. A., and Crawford, C. (1994) *Eur. J. Biochem.* **223,** 455–464
61. Bi, X., Chang, V., Molnar, E., McIlhinney, R. A., and Baudry, M. (1996) *Neuroscience.* **73,** 903–906
62. Melloni, E., Salamino, F., Sparatore, B., Michetti, M., and Pontremoli, S. (1984) *Biochim. Biophys. Acta* **788,** 11–16
63. Chin, D. J., Gil, G., Russell, D. W., Liscum, L., Luskey, K. L., Basu, S. K., Okayama, H., Berg, P., Goldstein, J. L., and Brown, M. S. (1984) *Nature* **308,** 613–617
64. Pfaff, M., Du, X., and Ginsberg, M. H. (1999) *FEBS Lett.* **460,** 17–22
65. Du, X., Saido, T. C., Tsubuki, S., Indig, F. E., Williams, M. J., and Ginsberg, M. H. (1995) *J. Biol. Chem.* **270,** 26146–26151
66. Straubinger, A. F., Viveiros, M. M., and Straubinger, R. K. (1999) *Gene* (*Amst.*) **236,** 273–280
67. Kobayashi, Y., Yamamoto, K., Saido, T., Kawasaki, H., Oppenheim, J. J., and Matsushima, K. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 5548–5552
68. Tsubata, T., and Takahashi, K. (1989) *J. Biochem.* (*Tokyo*) **105,** 23–28
69. Kimura, Y., Koga, H., Araki, N., Mugita, N., Fujita, N., Takeshima, H., Nishi, T., Yamashima, T., Saido, T. C., Yamasaki, T., Moritake, K., Saya, H., and Nakao, M. (1998) *Nat. Med.* **4,** 915–922
70. Schey, K. L., Fowler, J. G., Shearer, T. R., and David, L. (1999) *Invest. Ophthalmol. Vis. Sci.* **40,** 657–667
71. Nascimento, A. A., Cheney, R. E., Tauhata, S. B., Larson, R. E., and Mooseker, M. S. (1996) *J. Biol. Chem.* **271,** 17561–17569
72. Schroder, E., Willis, A. C., and Ponting, C. P. (1998) *Biochim. Biophys. Acta* **1383,** 279–291
73. Guttmann, R. P., Baker, D. L., Seifert, K. M., Cohen, A. S., Coulter, D. A., and Lynch, D. R. (2001) *J. Neurochem.* **78,** 1083–1093
74. Lee, M. S., Kwon, Y. T., Li, M., Peng, J., Friedlander, R. M., and Tsai, L. H. (2000) *Nature* **405,** 360–364
75. Pariat, M., Carillo, S., Molinari, M., Salvat, C., Debussche, L., Bracco, L., Milner, J., and Piechaczyk, M. (1997) *Mol. Cell. Biol.* **17,** 2806–2815
76. Buki, K. G., Bauer, P. I., and Kun, E. (1997) *Biochim. Biophys. Acta* **1338,** 100–106
77. Park, D., Jhon, D. Y., Lee, C. W., Ryu, S. H., and Rhee, S. G. (1993) *J. Biol. Chem.* **268,** 3710–3714
78. Pete, M. J., Liao, C. X., Bartleson, C., and Graves, D. J. (1999) *Arch. Biochem. Biophys.* **367,** 104–114
79. Kishimoto, A., Mikawa, K., Hashimoto, K., Yasuda, I., Tanaka, S., Tominaga, M., Kuroda, T., and Nishizuka, Y. (1989) *J. Biol. Chem.* **264,** 4088–4092
80. James, P., Vorherr, T., Krebs, J., Morelli, A., Castello, G., McCormick, D. J., Penniston, J. T., De Flora, A., and Carafoli, E. (1989) *J. Biol. Chem.* **264,** 8289–8296
81. Wu, Y., Aghdasi, B., Dou, S. J., Zhang, J. Z., Liu, S. Q., and Hamilton, S. L. (1997) *J. Biol. Chem.* **272,** 25051–25061
82. Hayashi, M., Suzuki, H., Kawashima, S., Saido, T. C., and Inomata, M. (1999) *Arch. Biochem. Biophys.* **371,** 133–141
83. Yang, L. S., and Ksiezak-Reding, H. (1995) *Eur. J. Biochem.* **233,** 9–17
84. Kiuchi, K., Titani, K., Fujita, K., Suzuki, K., and Nagatsu, T. (1991) *Biochemistry* **30,** 10416–10419
85. Fischer, S., Vandekerckhove, J., Ampe, C., Traub, P., and Weber, K. (1986) *Biol. Chem. Hoppe Seyler* **367,** 1147–1152
86. Nozaki, H. (1987) *Tokai J. Exp. Clin. Med.* **12,** 223–227