Probing the Dimeric Structure of Porcine Aminoacylase 1 by Mass Spectrometric and Modeling Procedures†

Chiara D'Ambrosio,‡,§ Fabio Talamo,‡,§ Rosa Maria Vitale,|| Pietro Amodeo,⊥ Gianluca Tell,# Lino Ferrara,‡ and Andrea Scaloni*,‡

Proteomics and Mass Spectrometry Laboratory, ISPAAM, National Research Council, via Argine 1085, 80147 Napoli, Italy, Department of Environmental Sciences, Second University of Naples, 81100 Caserta, Italy, Institute for Biomolecular Chemistry, National Research Council, 80078 Pozzuoli, Napoli, Italy, and Department of Biochemistry, Biophysics and Macromolecular Chemistry, Center of Excellence in Biocrystallography, University of Trieste, 34127 Trieste, Italy

Received November 20, 2002; Revised Manuscript Received January 30, 2003

ABSTRACT: Aminoacylase 1 is a zinc-binding metalloprotease catalyzing the hydrolysis of N^{α}-acylated L-amino acids; it presents altered expression levels in different renal and small cell lung carcinomas. A description of its redox and oligomerization state was achieved by combined biochemical and mass spectrometric procedures. A topological analysis of the enzyme structural architecture was derived from limited proteolysis and selective chemical modification experiments, using a broad range of proteases and chemical reagents. The analysis of the reaction products by different mass spectrometric techniques identified 26 amino acids as being accessible on the molecular surface, defining polypeptide regions exposed in the structure of the dimeric protein. The nature of the intermolecular contact zone between monomers was investigated by cross-linking reaction and mass mapping experiments. The cross-linked dimer was isolated, and the intermolecular cross-linked peptides were characterized, thus demonstrating the spatial proximity of Lys220 and Lys231 at the dimerization interface. Standard modeling procedures based on automatic alignment on the structure of members of the M20 peptidase family failed to produce a dimeric model consistent with experimental data. Discrepancies were observed mainly at the dimer interface and at loop regions. Therefore, a refined model for this dimeric protease was calculated by selecting the one able to generate a structure fully compatible with experimental findings, among all possible suboptimal sequence alignments. According to this model, each aminoacylase monomer consists of two domains: a globular catalytic subunit (residues 1−188 and 311−399) consisting of a β-sheet sandwiched between α-helices and a second β-sheet located on the surface, and the dimerization domain (residues 189−310) folding into a β-sheet flanked on one side by two α-helices. These results indicate that reliable approaches such as limited proteolysis, selective chemical modification, and cross-linking coupled to mass spectrometry can be used to test and optimize molecular models of multimeric proteins and highlight problems in automatic model building.

Aminoacylase 1 (ACY1),[1] also known as acylase I (N^{α}-acyl-L-amino acid amidohydrolase, EC 3.5.1.14), is a zinc-binding metalloenzyme differently expressed in mammalian tissues catalyzing the hydrolysis of N^{α}-acylated L-amino acids to L-amino acids (*1*, *2*). It has been suggested to be involved in the final steps of the N^{α}-acylated protein's degradation together with acylpeptide hydrolase (APH, EC 3.4.19.1) (*3*−*6*). In fact, APH is able to generate an N^{α}-acyl amino acid from an N^{α}-acyl peptide or N^{α}-acyl protein, whereas ACY1 hydrolyzes the resulting N^{α}-acyl amino acid. Moreover, additional investigations reported that APH could degrade a series of bioactive peptides as the α-melanocyte stimulating hormone (*5*).

Human and porcine ACY1 and APH have been cloned and located on chromosomes 3p21.1 and 3p21.3 in humans, respectively (*7*−*11*); these regions present a consistent loss of heterozygosity in small cell lung and kidney carcinomas (*7*, *12*). In parallel, altered expression of ACY1 and APH, together with absent activities for both enzymes, has been reported in different renal cancer and SCLC samples (*13*−*15*). Several authors already reported in SCLC a high-frequency loss of genetic material involving chromosome 3p(14−23), suggesting that several tumor suppressor genes may be located within this region (*16*, *17*). The coexistence of ACY1 and APH genes on chromosome 3p21 is of interest, as these enzymes are functionally linked in the sequential degradation of acylated peptides and are able to bind each

† This work was supported by grants from the Italian National Research Council and from Regione Campania.

* To whom correspondence should be addressed: Proteomics and Mass Spectrometry Laboratory, ISPAAM, National Research Council, via Argine 1085, 80147 Napoli, Italy. E-mail: A.Scaloni@iabbam.na.cnr.it. Telephone: ++39-81-5966006. Fax: ++39-81-5965291.

‡ Proteomics and Mass Spectrometry Laboratory, ISPAAM, National Research Council.

§ These authors contributed equally to this work.

|| Second University of Naples.

⊥ Institute for Biomolecular Chemistry, National Research Council.

University of Trieste.

[1] Abbreviations: ACY1, aminoacylase 1; APH, acylpeptide hydrolase; BS_3, bis(sulfosuccinimidyl)suberate; CAM, carboxyamidomethyl; CBPG, carboxypeptidase G2; ESI, electrospray; hACY1, human aminoacylase 1; HD, helical region at the dimerization interface; LC, liquid chromatography; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; pACY1, porcine aminoacylase 1; PTH, phenylthiohydantoin; SAS, solvent accessible surface; SCLC, small cell lung cancer.

10.1021/bi0206715 CCC: $25.00
Published on Web 03/29/2003

other's substrate in a competitive fashion. This observation led to the speculation that impairment or loss of the APH/ACY1 deacylating system may result in the accumulation of a yet unknown N^{α}-acylated peptide with growth promoting properties responsible for cell transformation (*15*).

Although a detailed characterization of the biochemical properties was reported for both enzymes, no data on the three-dimensional structure of these proteases have been published to date. This lack of data is undesirable, not only for a final understanding of their substrate specificity, which would explain their functional differences, but also because ACY1 and APH are examples of two distinct classes of proteases, whose X-ray structures were determined only for a single member. In fact, APH belongs to the S9 peptidase family whose structural architecture was determined only in the case of porcine prolyl oligopeptidase (*18*, *19*). This enzyme presents a peptidase domain with an α/β hydrolase fold and an unusual β-propeller domain covering the active site. In contrast, ACY1 belongs to the M20 peptidase family, whose general fold was hypothesized only on the basis of the crystallographic structure of dimeric *Pseudomonas* sp. strain RS16 carboxypeptidase G2 (*20*, *21*). In this protein, each subunit consists of a larger catalytic domain containing two zinc ions at the active site, and a separate smaller domain forming the dimer interface.

On the basis of amino acid sequence alignment, the M20 family has been reported to include ACY1 from different mammals, carboxypeptidase S from *Saccharomyces cerevisiae*, succinyl-diaminopimelate desuccinylase from *Escherichia coli*, *Haemophilus influenzae*, and *Corynebacterium glutamicum*, acetyl-ornithine deacetylase from *E. coli* and *Dictyostelium discoideum*, and CBPG from *Pseudomonas* sp. strain RS16. Most of these metalloexopeptidases are soluble dimeric zinc- or cobalt-dependent enzymes exhibiting similar biochemical properties and three conserved amino acid motifs (*21*). However, recently a trimeric aggregation state has been reported in the case of rat ACY1 (*22*); similarly, monomeric forms of pACY1 and *S. cerevisiae* carboxypeptidase S have also been described (*23*). In addition, a hypothetical disulfide connecting Cys270 and Cys292 has been reported for pACY1 (*24*). These cysteine residues are not conserved in the other mammalian aminoacylases sequenced to date.

To definitely elucidate the redox state of pACY1 and provide clarity on its aggregation state, describing the elements involved in protein oligomerization, a detailed structural characterization of the enzyme was achieved. A topological description of the aminoacylase molecular architecture was derived from limited proteolysis, selective chemical modification, and cross-linking experiments combined with mass spectrometric procedures. These approaches have been previously used by different groups to locate accessible amino acids on the molecular surface of monomeric proteins or to identify interacting regions in protein complexes (*25*−*38*). These data allowed us to propose an experimentally verified model for dimeric pACY1, calculated on the basis of its sequence homology with *Pseudomonas* CBPG.

EXPERIMENTAL PROCEDURES

Materials. Trypsin, chymotrypsin, subtilisin, tetranitromethane, acetic anhydride, α-cyano-4-hydroxycinnamic acid, BS_3, and iodoacetamide were purchased from Sigma. Endoprotease GluC was from Roche. PD10 columns were from Amersham Pharmacia Biotech. All other reagents and solvents were HPLC-grade from Fluka.

pACY1 Purification and Apparent Molecular Mass Determination. pACY1 was purified from porcine kidney according to a procedure previously published (*2*). The final preparation was homogeneous on PAGE in the presence and absence of SDS. The enzyme activity was assayed using furylacryloyl-L-Met as a substrate (*22*). The apparent molecular mass was estimated by analytical gel filtration chromatography using a Superdex 200 HR column (Amersham Pharmacia Biotech) and gel filtration marker proteins. The column was equilibrated and eluted with 150 mM NaCl and 50 mM sodium phosphate (pH 7.0). The effluent was monitored at 280 nm.

Redox State Determination. Native pACY1 was denatured in 0.25 M Tris-HCl and 1.25 mM EDTA containing 6 M guanidinium chloride (pH 7.0) without the presence of reducing agents. The protein sample was quickly alkylated by the addition of iodoacetamide (1.1 M final concentration) at room temperature for 1 min, in the dark (*39*). Carboxyamidomethylated protein was desalted by eluting the reaction mixture through a PD10 column.

Limited Proteolysis. Limited proteolysis experiments with native pACY1 (0.5 nmol) were conducted in 100 μL of a solution containing 0.1 M NH_4HCO_3, 1 mM DTT, and 1 mM $ZnCl_2$ (pH 7.5) at 30 °C by using an enzyme:substrate ratio ranging from 1:25 to 1:2500 (w/w). The extent of digestion was monitored on a time course basis by sampling 10 μL of the reaction mixture at time intervals ranging from 5 to 360 min. Peptide mixtures were directly analyzed by MALDI-MS following a desalting step on a ZipTip C4 device (Millipore).

Selective Acetylation and Nitration. Modification of pACY1 lysine and tyrosine residues was carried out by incubating aliquots of the native protein (2.5 nmol) with the appropriate reagent (200 μL final volume). Acetylation of lysines was performed in 10 mM NH_4HCO_3, 1 mM DTT, and 1 mM $ZnCl_2$ (pH 7.5) at 25 °C for 10 min by adding a 10−5000-fold molar excess of acetic anhydride to pACY1. Nitration of tyrosines was performed in 25 mM sodium phosphate and 1 mM $ZnCl_2$ (pH 7.5) at 0 °C for 60 min by adding a 100−5000-fold molar excess of tetranitromethane to pACY1. All the reactions were stopped by acidification with 1% TFA, and the samples were directly injected onto a Vydac C4 column (250 mm × 2.1 mm, 5 μm) (The Separation Group). The modified species were eluted by means of a linear gradient from 30 to 80% acetonitrile in 0.1% TFA over the course of 25 min, with a flow rate of 0.2 mL/min; elution was monitored at 220 and/or 360 nm. Individual fractions were collected manually.

Cross-Linking Experiments. The cross-linking reaction of native pACY1 was carried out by incubating aliquots of the native protein (1.5 nmol) with BS_3 in 10 mM sodium phosphate, 1 mM DTT, and 1 mM $ZnCl_2$ (pH 7.5) (50 μL final volume). The reagent was added in a 10−2500-fold molar excess to pACY1. Each mixture was incubated for 1 h at 25 °C, and the reaction was quenched by lyophilization. Samples were directly analyzed by 10% SDS−PAGE.

Extensive Enzymatic and Chemical Hydrolysis. The differently modified samples of pACY1 were digested with

trypsin or endoprotease GluC in 50 mM ammonium acetate (eventually containing 1 M urea) (pH 6.0) at 37 °C, overnight, using an enzyme:substrate ratio of 1:50 (w/w). Digestion was alternatively performed by incubating pACY1 samples in 5% formic acid at 55 °C, overnight.

Protein bands from SDS−PAGE analysis were excised, carboxyamidomethylated, and finally digested *in situ* with trypsin as previously reported (*40*). Peptides were extracted by sonication (twice) with 100 μL of a 25 mM 1:1 (v/v) NH_4HCO_3/acetonitrile mixture (pH 8) and lyophilized.

Peptide digests from in solution or *in situ* hydrolysis were directly analyzed by MALDI-MS or LC−ESI-MS.

MALDI-MS Analysis. Matrix-assisted laser desorption ionization mass spectra were recorded using a Voyager DE-PRO mass spectrometer (Applera) as previously reported (*40*); 1 μL of an analyte solution (0.1−50 pmol/μL) and 1 μL of an α-cyano-4-hydroxycinnamic acid solution (5 mg/mL in 50% acetonitrile and 5% formic acid) were applied to the sample plate and dried. Molecular mass standards were eventually added prior to the analysis. Internal mass calibration was performed using the molecular ions from added standards, peptides produced by enzyme autoproteolysis, and the matrix. All masses are reported as average values.

LC−ESI-MS Analysis. pACY1 digests were analyzed using an API-100 single-quadrupole mass spectrometer (Applera) equipped with an atmospheric-pressure ionization source, as previously reported (*39*). Peptide mixtures were separated on a microbore Vydac C18 column (250 mm × 1 mm, 5 μm) (The Separation Group), using a linear gradient from 5 to 70% acetonitrile containing 0.1% TFA, over a period of 65 min, at a flow rate of 0.08 mL/min. The column effluent was split 1:1 into the mass spectrometer connected on-line. The remaining part was spectrophotometrically detected at 220 and/or 360 nm. In this case, peptides were manually collected and lyophilized for further characterization. Spectra were acquired in the range of m/z 250−2000. Data were processed using BioMultiView version 1.3 (Applera). Mass calibration was performed by means of the multiply charged ions from a separate injection of horse heart myoglobin. All masses are reported as average values.

Mass Ion Assignment. The observed mass values were assigned to pACY1 polypeptides using GPMAW 4.23 (Lighthouse Data). This software generated a mass/fragment database output, based on the pACY1 sequence, the selectivity of the protease or chemical reagent used for protein digestion, the nature of the amino acid susceptible to eventual selective chemical modification or cross-linking reaction, and the molecular mass of the modifying groups or cross-linking arm. Mass values were matched to peptide regions using a 0.1% mass tolerance value.

Protein Sequence Analysis. Automated N-terminal degradation of the purified peptides was performed by using a Procise 491 protein sequencer (Applera) equipped with a 140C microgradient apparatus and a 785A UV detector (Applera) for the automated identification of PTH-amino acids. PTH-Tyr-NO_2 was detected according to the method of Zappacosta et al. (*34*).

Homology Modeling. Standard modeling procedures were based on the automated comparative protein server freely available at http://www.expasy.ch/swissmod/swiss-model.html (*41*).

The optimized homology model of pACY1 was built with the program MODELLER 6.2 (*42*), using the pACY1 sequence (SWISS-PROT entry P37111) as a target, and the structure of *Pseudomonas* sp. strain RS16 CBPG (*43*) (chains A and D of PDB entry 1CG2) as a template. The sequence alignments were carried out with CLUSTALW version 1.81 (*44*). Multialignments also included the sequence of hACY1 (SWISS-PROT entry Q03154). Secondary structure information, to avoid gaps and insertions in structured regions, was derived from PHD prediction (http://dodo.cpmc.columbia.edu/predictprotein) for the target and from DSSP analysis (*45*) of the experimental structure for the template. It was introduced in the alignment by performing a profile alignment in CLUSTALW with secondary structure gap penalties, using the program default weighting scheme.

MODELLER calculations included symmetry restraints for the two monomers and, wherever explicitly stated, secondary structure restraints for an elongated helical region in the dimerization domain, with a simultaneous deletion of the corresponding template sequences. Because of discrepancies with the template in the zinc-binding site regions, zinc ions and zinc-binding water molecules from PDB entry 1CG2 were excluded in the template.

Fifty MODELLER models were generated from each alignment. The models that exhibited no topological error (knots) or abnormally large objective function value addition were analyzed with PROCHECK version 3.5.4 (*46*). The best five models were completed by adding all hydrogen atoms and underwent energy minimization with the SANDER_CLASSIC module of AMBER 6 (*47*), using the PARM91 force field (*48*, *49*), reduced charges on Zn ions ($q = 1.2$ atomic units), charged ends and/or residues ($q = \pm 0.2$ atomic unit), a distance-dependent dielectric constant, and a 12 Å cutoff for nonbonded interactions. The minimization protocol included 100 steps of steepest descent, followed by 5000 steps of conjugate gradients or until the average gradient was $\leq 10^{-4}$ kcal mol^{-1} Å^{-1}. Distance restraints on the atoms involved in zinc binding (force constant of 40 kcal mol^{-1} Å^{-1}) were used in the minimization procedure.

A model of the experimentally observed cross-linked dimer was obtained by manually docking a fully extended conformation of the OC-$(CH_2)_6$-CO spacer arm in the intermonomer region between Lys220 and Lys231 of two different monomers. Amide bonds were formed at both ends, by minimal rearrangement of both Lys side chains and cross-linker terminal CH_2−CO torsional angles. The model that was obtained underwent a local energy minimization (limited to the linker and the two Lys side chains) under the same conditions described above for model refinement. Validation of the final models was obtained with PROCHECK version 3.5.4 (*45*) and ProsaII version 3.0.2 (*50*).

The solvent accessible surface (SAS) (*51*) area has been calculated with the "CalcSurf" command of MolMol. It calculates a numerical approximation to the true SAS based on the following approach: if p is the precision given as an argument, it distributes $2^{(2p+2)} + 2$ points over each sphere, and checks which fraction of them are not inside another sphere. The part of the sphere surface corresponding to this fraction is added to the total surface. In our analysis, we used a solvent radius of 1.4 Å and a precision value p equal to 4.

Analyses and pictures were obtained with MOLMOL (*52*).

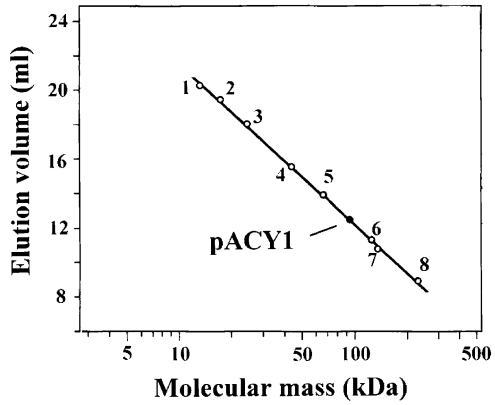

FIGURE 1: Determination of the apparent molecular mass of pACY1 by analytical gel filtration chromatography. The measurement was obtained by loading a mixture of pACY1 and gel filtration calibrants on a Superdex 200 HR column as described in Experimental Procedures. Gel filtration of marker proteins was performed consecutively under the same experimental conditions. The apparent molecular mass of pACY1 is represented by a filled circle. Calibrants were ribonuclease A (1), myoglobin (2), chymotrypsinogen (3), ovalbumin (4), serum albumin (5), serum albumin dimer (6), aldolase (7), and catalase (8).

RESULTS

pACY1 Apparent Molecular Mass and Redox State Determination. An aliquot of freshly purified porcine kidney ACY1 was analyzed by ESI-MS following a desalting step on a RP-HPLC column. The spectrum that was obtained exhibited the occurrence of a single component with a molecular mass of 45259.9 ± 3.8 Da. This value differed (by 42 Da) from the calculated value for the pACY1 deduced from its cDNA sequence (theoretical value of 45216.4 Da), thus confirming the occurrence of an acetyl group at the N-terminus and excluding the occurrence of other post-translational modifications in the polypeptide chain (*8*). Similarly, another sample was analyzed by gel filtration chromatography with the aim of determining the pACY1 aggregation state. The analysis showed an apparent mass of ~90 kDa for the protein (Figure 1), thus definitely demonstrating its dimeric state. No additional peaks indicating the simultaneous occurrence of monomeric or trimeric species were detected.

Furthermore, a freshly prepared pACY1 sample was quickly alkylated with iodoacetamide under not reducing but denaturing conditions, desalted, and finally digested with trypsin. The peptide digest was analyzed by LC−ESI-MS. This analysis allowed the verification of the entire sequence deduced from cDNA (data not shown), demonstrating, at the same time, that the five cysteines occurring in the polypeptide were all present in a reduced state (Table 1). In fact, all signals were assigned to peptides containing cysteine residues as carboxyamidomethylated species; no peaks associated with eventual disulfide-containing peptides were detected. These results definitely proved that pACY1 does not contain disulfide bridges. These data were in contrast with other observations previously reported and deduced by SDS−PAGE analysis (*24*).

pACY1-Limited Proteolysis. Limited proteolysis and SDS−PAGE analysis have long been widely used to define the domain structure organization of proteins or to identify highly flexible polypeptide regions on their molecular surface. More recently, a strategy that combines limited proteolysis experiments with mass spectrometric procedures has been proposed as a rapid, sensitive, and definitive methodology for the topological analysis of monomeric proteins (*25*−*27*). This approach, applied to proteins as isolated species and following eventual complex formation, generates differential peptide maps whose interpretation also can lead to the identification of the interface regions in protein complexes (*28*−*31*).

Table 1: LC−ESI-MS Analysis of Native pACY1 Alkylated under Denaturing and Not Reducing Conditions following Digestion with Trypsin[a]

measured mass (Da)	theoretical mass (Da)	peptide
903.2 ± 0.4	903.1	(43−50)CAM
1130.6 ± 0.7	1130.3	(41−50)CAM
1337.9 ± 0.4	1337.6	(115−125)CAM
1493.5 ± 0.6	1493.7	(115−126)CAM
1658.6 ± 0.4	1658.9	(334−348)CAM
2091.7 ± 0.8	2091.4	(334−351)CAM
2790.3 ± 0.6	2790.1	(283−306)CAM
4187.5 ± 0.9	4187.8	(236−274)CAM

[a] Indicated are only the cysteine-containing peptides. CAM is carboxyamidomethyl.

To define the accessible regions in the dimeric structure of native pACY1, limited proteolysis experiments were conducted with a variety of proteases under strictly controlled conditions. Preliminary experiments were performed in each case to select the optimal enzyme:substrate ratios to be used. The extent of enzymatic hydrolysis was monitored on a time course basis by sampling the incubation mixture at different times followed by MALDI-MS analysis. The assignment of generated peptides was performed on the basis of the pACY1 sequence and enzyme specificity. The identification of cleavage sites was achieved only by recognition of complementary peptides in the mass spectra. In fact, these species were generated by single proteolytic events on pACY1 native structure and were indicative of residue accessibility on the molecular surface.

Figure 1 shows the time course analysis of the endoprotease GluC digestion of native pACY1 carried out by using an E:S ratio of 1:25. Under these conditions, the protein remained mostly undigested, showing that proteolysis occurred at few specific sites. In fact, after 15 min, only the hydrolysis at Glu88 was observed, as demonstrated by the presence of residues 1−88 and 89−406 (theoretical *m*/*z* values of 9739.1 and 35540.3, respectively) (Figure 2A). Two additional peptide pairs were observed after incubation for 360 min (Figure 2B); their molecular mass identified them as peptides 1−255 and 256−406, and peptides 1−309 and 310−406 (theoretical *m*/*z* values of 28711.6, 16567.8, 34698.4, and 10581.0, respectively). These data demonstrated that endoprotease GluC can recognize Glu255 and Glu309 as additional sites also exposed on the dimeric pACY1 structure. As the digestion time increases, fragments originated from single proteolytic events become susceptible to further cleavages, leading to the formation of subdigestion products (Figure 2B). Being uninformative, these species were not considered in our data analysis.

Similar experiments were performed using trypsin as the proteolytic enzyme. In this case, the characterization of the complementary peptides 1−5 and 6−406, 1−22 and 23−406, 1−196 and 197−406, 1−208 and 209−406, and 1−306

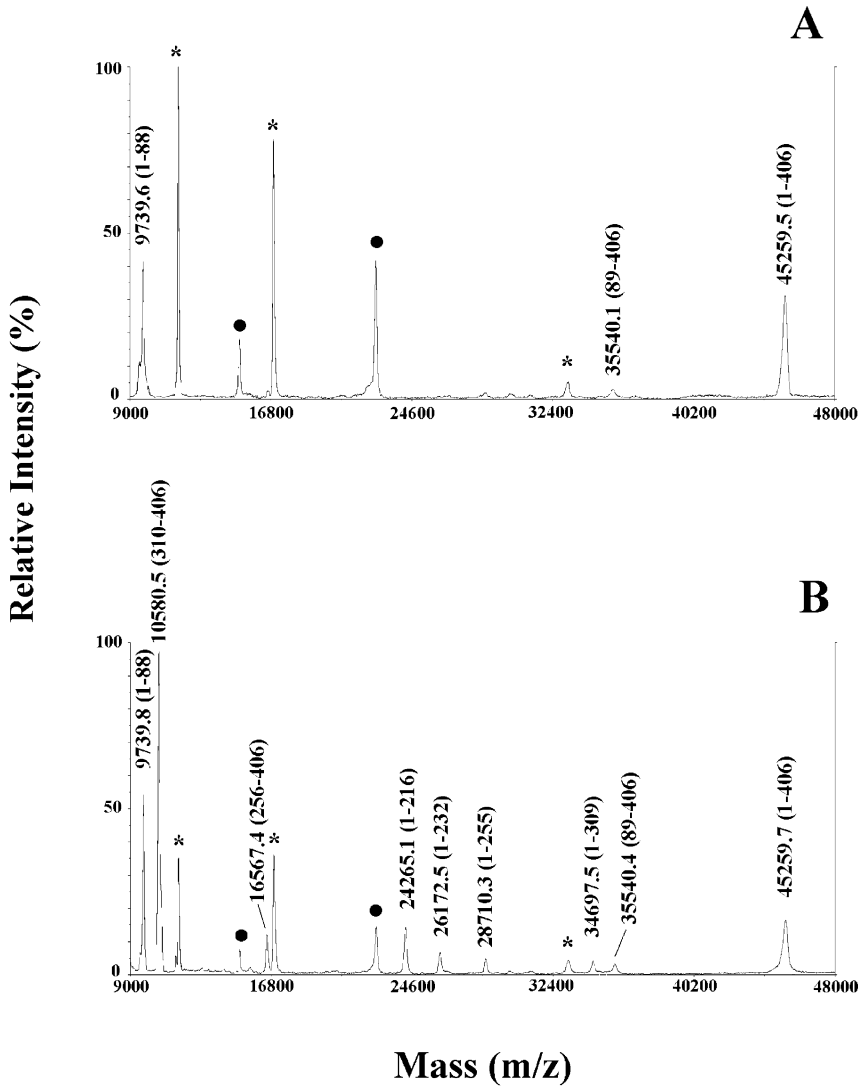

FIGURE 2: Limited proteolysis of native pACY1 with endoprotease GluC. The MALDI-MS time course analysis following incubation with the protease (1:25 E:S ratio) after 15 (A) and 360 min (B) is reported. Mass values and the corresponding polypeptides are indicated. Dots denote doubly and triply charged pACY1 ions; asterisks denote markers used for internal mass calibration.

and 307−406 within time course-sampled aliquots (measured masses of 559.8, 44717.2, 2643.8, 42633.9, 22223.4, 23053.9, 23388.4, 21888.1, 34249.1, and 11027.3 Da, respectively; theoretical masses of 560.5, 44716.8, 2644.0, 42633.4, 22223.2, 23054.2, 23388.5, 21888.9, 34250.9, and 11026.5 Da, respectively) allowed the identification of Arg5, Arg22, Arg196, Arg208, and Lys306 as basic residues exposed on the molecular surface of pACY1.

Incubation of native pACY1 with narrow-specificity proteases resulted in hydrolysis at very few easily identifiable sites but limited the amount of information on the exposed uncharged amino acid residues. Therefore, further experiments were carried out by incubating native pACY1 with broader-specificity proteases such as chymotrypsin or subtilisin. As an example, the exclusive occurrence in the pACY1 chymotryptic digest (digestion time of 15 min, E:S ratio of 1:50) of the complementary peptides 1−90 and 91−406 (measured masses of 10061.2 and 35215.2 Da, respectively; theoretical masses of 10061.5 and 35215.9 Da, respectively) demonstrated that hydrolysis occurred selectively at Trp90. The same fragments were detected also after 30 min, whereas three further complementary peptide pairs (1−209 and 210−406, 1−260 and 261−406, and 1−307 and 308−406) were observed after reaction for 220 min (measured masses of 23534.7, 21740.9, 29157.8, 16119.1, 34436.5, and 10840.6 Da, respectively; theoretical masses of 23535.7, 21741.8, 29158.1, 16119.3, 34437.2, and 10840.2 Da, respectively). These data definitely demonstrated that chymotrypsin hydrolyzes pACY1 at Trp90, Phe209, Tyr260, and Trp307.

Table 2: Preferential Cleavage and Modification Sites in Native pACY1 As Determined by Limited Proteolysis and Selective Chemical Modification Experiments

protease/reagent	cleavage/modification sites
endoprotease GluC	Glu88, Glu255, Glu309
trypsin	Arg5, Arg22, Arg196, Arg208, Lys306
chymotrypsin	Trp90, Phe209, Tyr260, Trp307
subtilisin	Gln311
acetic anhydride	Lys3, Lys50, Lys87, Lys99, Lys128, Lys159, Lys202, Lys217, Lys220, Lys282, Lys306
tetranitromethane	Tyr30, Tyr260
BS_3	Lys128, Lys202, Lys220, Lys231, Lys233

The overall results of the proteolysis experiments are summarized in Table 2. As expected, the cleavage sites recognized by different proteases were grouped within the same regions of the polypeptide chain, thus defining the

accessible areas and demonstrating the consistency of these data. Five exposed regions of the pACY1 structure were depicted: the N-terminus, a short segment comprising Glu88−Trp90, a broader region encompassing Leu196−Phe209, the polypeptide segment of Glu255−Tyr260, and the C-terminal region from Lys306 to Gln311.

pACY1 Chemical Modification. Information complementary to limited proteolysis experiments was obtained by an independent strategy consisting of selective chemical modification at specific residues. This approach was previously used to probe the accessibility of selected amino acids on a protein surface with several reagents (*32*−*35*). In fact, under conditions of limited chemical modification, their relative reactivities depend primarily on accessibility, unless peculiar electrostatic interactions, hydrogen bonding, and other microenvironmental effects take place. A modified strategy was also derived for the topological characterization of protein complexes and used to identify regions occurring at the interface between macromolecules (*31*, *36*).

To further probe the topology of the pACY1 dimer, we used this approach by performing a systematic study of the relative reactivity of its lysine and tyrosine residues. Particular care was taken to select optimal incubation times and stoichiometry for each modifying reagent, under the same experimental conditions used for limited proteolysis. Preliminary experiments were performed using different reagent:protein ratios to limit the extent of modification. Only mono- and dimodified species were taken into account, assuming that an extensive introduction of modifying groups might result in, or might have been due to, the alteration of the dimer conformation.

Acetic anhydride:protein molar ratios of 2000:1 and 4000:1 resulted in the optimal experimental conditions for the selective modification of lysine residues. In the first case, a mixture of unmodified and monomodified pACY1 was observed in the reaction products (measured masses of 45258.9 ± 4.5 and 45301.9 ± 4.1 Da, respectively) following reaction for 10 min, as determined by ESI-MS measurements. In the second case, the increase in reagent concentration resulted in an equimolar mixture of mono- and diacetylated pACY1 (measured masses of 45302.7 ± 3.5 and 45342.4 ± 4.5 Da, respectively). Traces of triacetylated pACY1 were also detected. To identify the modified lysines, half of each sample was digested with endoprotease GluC and the resulting peptide mixtures were directly analyzed by MALDI-MS. Figure 3 and Table 3 show the spectra obtained and the relative mass assignments. The occurrence of satellite signals ($\Delta m = +42$ Da) with respect to lysine-containing peptides, absent in the unreacted ACY1 digest (blank), made acetylated peptides easily identifiable. As some lysine residues in peptides were undetected in the spectra shown in Figure 3, the remaining digests were hydrolyzed with 5% formic acid and the peptide mixtures similarly analyzed by MALDI-MS. The results are reported in Table 4. A comprehensive analysis of all data obtained made location of the N^{ϵ}-acetyllysines straightforward, the pACY1 N-terminus already being acetylated. In the case of a 2000:1 reagent:protein molar ratio, we concluded that acetylation occurred selectively at eight of the 17 lysine residues present in the pACY1 sequence, namely, Lys87, Lys99, Lys128, Lys159, Lys202, Lys217, Lys220, and Lys282. Identical residues were modified in the product generated using a 4000:1 reagent:protein ratio; furthermore, the presence of additional signals (see Figure 3 and Tables 3 and 4) demonstrated that modification also occurred at Lys3, Lys50, and Lys306.

A similar approach was used for the selective radicalic nitration of tyrosines by tetranitromethane. The extent of reaction was verified by monitoring the specific absorbance at 360 nm of 3-nitrotyrosines (*34*). ESI-MS measurements were performed to characterize the modified species by simply verifying the Δm of +45 Da, associated with the addition of the nitro group. In fact, MALDI-MS analysis of nitrotyrosine-containing peptides was reported to be affected by laser-induced photochemical decomposition (*53*). On this basis, an optimal 5000:1 reagent:protein molar ratio was selected. In this case, the reaction yielded a mixture of unmodified and mononitrated species (measured masses of 45258.7 ± 4.1 and 45303.5 ± 4.8 Da, respectively). The identification of the nitrated residues was carried out by the comparison of the peptide mass mapping experiments obtained for this species with respect to a sample of unreacted pACY1, used as a blank. Each tryptic digest was resolved by LC, and the column effluent was split 1:1 into the mass spectrometer and a spectrophotometer, both connected online. In the chromatogram of the nitrated sample, the absorbance at 360 nm of two defined components allowed a rapid monitoring of the modified peptides (data not shown). Their ESI-MS analysis identified it as peptides 23−40 and 236−274 with a mass increase of 45 Da (theoretical masses of 2037.2 and 4175.8 Da, respectively), indicating the incorporation of one nitro group per tyrosine residue (Figure 4A,B). These species eluted almost 1.5 min later than the corresponding unmodified peptides and were absent in asample of untreated pACY1 (data not shown). The specific detection of the corresponding PTH-Tyr-NO_2 in these peptides, following Edman degradation analysis, definitely demonstrated that nitration of pACY1 resulted in the selective modification of Tyr30 and Tyr260. The overall results of the selective chemical modification experiments are summarized in Table 2.

pACY1 Cross-Linking. Chemical cross-linking of noncovalent protein complexes and the analysis of cross-linked species by SDS−PAGE have long been employed in biochemistry laboratories to determine the stoichiometry of constituent monomers (*54*). Cross-linking followed by mass spectrometric analysis has more recently been used to determine the stoichiometry of noncovalent protein complexes, and some studies have extended this approach to proteolytic digestion of the cross-linked species and subsequent identification of the linked peptides (*30*, *37*). Different reagents have been used for this purpose, depending on the nature of the amino acids occurring at the interface regions (*55*). On the basis of preliminary hypotheses resulting from sequence alignments between ACY1 and CBPG, we decided to use BS_3 to selectively cross-link spatially closed lysine residues.

Native pACY1 was incubated with different concentrations of the cross-linking agent under strictly controlled conditions, and the reaction mixtures were analyzed by SDS−PAGE. To limit the extent of modification, we chose a 1000:1 BS_3:pACY1 molar ratio as the optimal reaction condition (Figure 5). The 90 and 45 kDa Coomassie-stained bands, corresponding to the cross-linked dimer and eventually modified

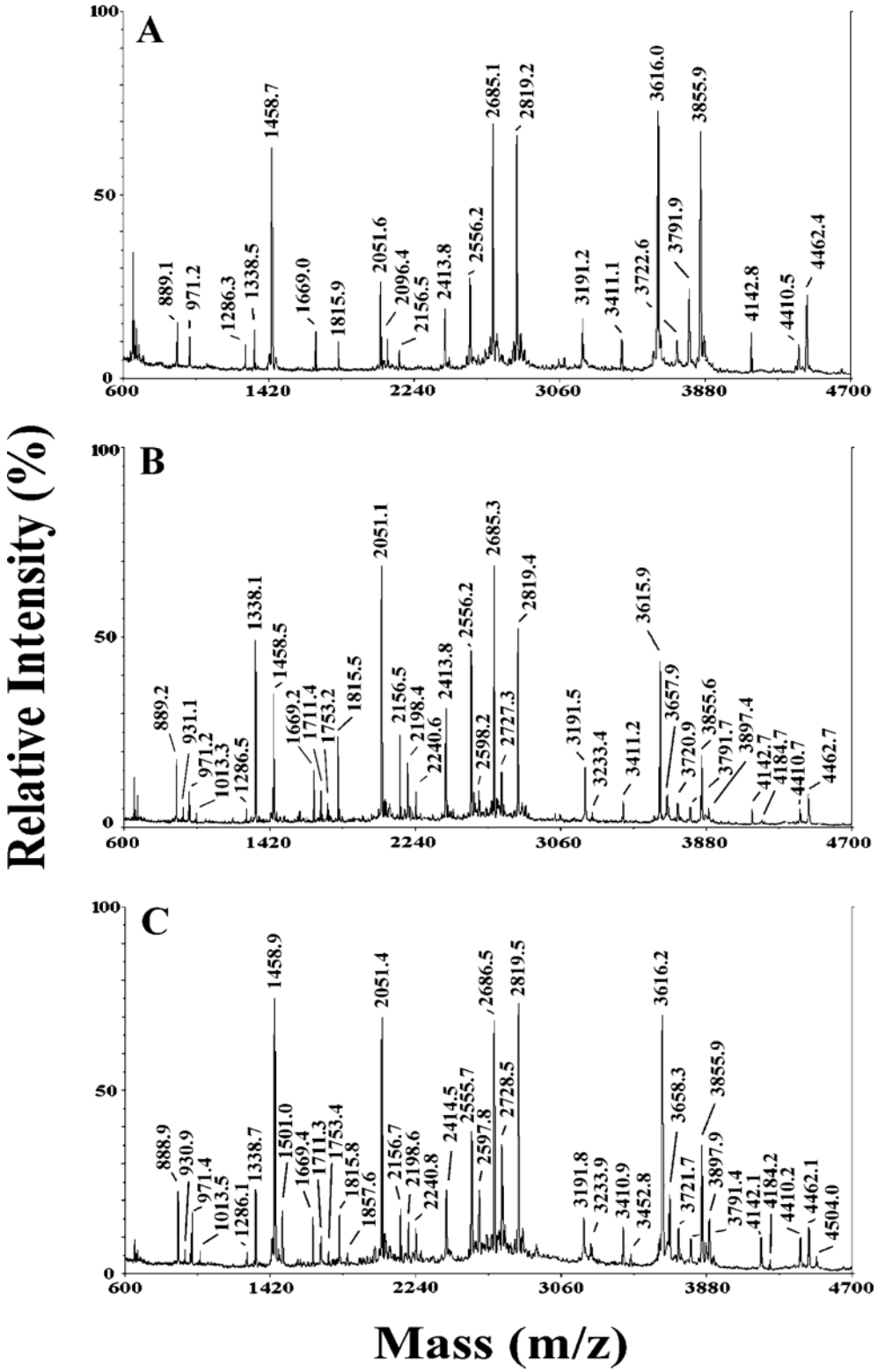

FIGURE 3: Selective acetylation of native pACY1 following reaction with acetic anhydride. The figure shows the MALDI-MS analysis of endoprotease GluC digest for the untreated pACY1 (A), acetylated pACY1 (acetic anhydride:pACY1 molar ratio of 2000:1) (B), and acetylated pACY1 (acetic anhydride:pACY1 molar ratio of 4000:1) (C). The assignments of individual signals to the corresponding peptides and theoretical mass values are reported in Table 3.

monomer, respectively, were digested *in situ* with trypsin. A sample of the nontreated pACY1 monomer was used as a blank. Aliquots from the extracted peptide mixtures were directly analyzed by MALDI-MS, and the spectra that were obtained are shown in Figure 6. Their comparison with the spectrum of the untreated pACY1 sample easily allowed the identification of the species originated from the reaction with BS_3. In fact, the spectrum of the BS_3-modified pACY1 monomer contained clear signals at *m*/*z* 716.7, 1129.4, 1340.6, 1376.6, and 1567.9; similarly, in the spectrum of the BS_3-cross-linked pACY1 dimer, peaks were present at *m*/*z* 716.8, 1340.5, and 1376.7. All these species were absent in the sample of untreated pACY1. On the basis of their molecular mass and trypsin specificity, these components

Table 3: MALDI-MS Analysis of Acetylated Peptides Obtained following the Reaction of pACY1 with Acetic Anhydride and Digestion with Endoprotease GluC

untreated pACY1 (measured *m/z*)	2000:1 Ac_2O:pACY1 ratio (measured *m/z*)	4000:1 Ac_2O:pACY1 ratio (measured *m/z*)	theoretical *m/z*	peptide
889.1	889.2	888.9	889.1	(156−162)
	931.1	930.9	931.1	(156−162)Ac
971.2	971.2	971.4	971.2	(123−130)
	1013.3	1013.5	1013.2	(123−130)Ac
1286.3	1286.5	1286.1	1286.4	(101−112)
1338.5	1338.1	1338.7	1338.5	(163−174)
1458.7	1458.5	1458.9	1458.7	(40−52)
		1501.0	1500.7	(40−52)Ac
1669.0	1669.2	1669.4	1669.0	(217−230)
	1711.4	1711.3	1711.0	(217−230)Ac
	1753.2	1753.4	1753.0	(217−230)Ac_2
1815.9	1815.5	1815.8	1816.0	(303−317)
		1857.6	1858.0	(303−317)Ac
2051.6	2051.1	2051.4	2051.3	(131−147)
		2096.4	2097.4	(320−337)
2156.5	2156.7	2156.7	2156.5	(212−230)
	2198.4	2198.6	2198.5	(212−230)Ac
	2240.6	2240.8	2240.5	(212−230)Ac_2
2413.8	2413.8	2414.5	2414.6	(317−337)
2556.2	2556.2	2555.7	2555.9	(190−211)
	2598.2	2597.8	2597.9	(190−211)Ac
2685.1	2685.3	2686.5	2686.9	(279−302)
	2727.3	2728.5	2728.9	(279−302)Ac
2819.2	2819.4	2819.5	2819.3	(380−406)
3191.2	3191.5	3191.8	3191.5	(256−285)
	3233.4	3233.9	3233.5	(256−285)Ac
3411.1	3411.2	3410.9	3410.8	(303−331)
		3452.8	3452.8	(303−331)Ac
3616.0	3615.9	3616.2	3616.0	(182−212)
	3657.9	3658.3	3658.0	(182−212)Ac
3722.6	3720.9	3721.7	3722.2	(347−379)
3791.9	3791.7	3791.4	3791.3	(7−39)
3855.9	3855.6	3855.9	3855.5	(53−88)
	3897.4	3897.9	3897.5	(53−88)Ac
4142.8	4142.7	4142.1	4142.6	(175−211)
	4184.7	4184.2	4184.6	(175−211)Ac
4410.5	4410.7	4410.2	4410.0	(340−379)
4462.4	4462.7	4462.1	4462.9	(1−39)
		4504.0	4504.9	(1−39)Ac

were assigned to peptides 232−235, 230−235, 197−208, 127−136, and 218−229, all presenting a mass increase of 156 or 312 Da (theoretical *m/z* values of 716.6, 1129.9, 1340.3, 1376.4, and 1567.7, respectively). These mass differences were associated with covalent adducts generated from the reaction of BS_3 with the ϵ-amino group of lysine residues, followed by the release of *N*-hydroxysuccinimide during enzymatic hydrolysis (*56*). This reaction and similar adducts have been recently reported for 3,3′-dithiobis-(succinimidyl)propionate, a disulfide-containing BS_3 homologue used for protein cross-linking (*37*). These results definitely demonstrated that Lys128, Lys202, Lys220, Lys231, and Lys233 were modified by BS_3 and are accessible residues on the molecular surface of the pACY1 dimer. Furthermore, the spectrum of the BS_3-cross-linked pACY1 dimer digest uniquely showed the occurrence of an intense signal at *m/z* 2082.5 that was associated with the cross-linked peptide (218−229)-CO-$(CH_2)_6$-CO-(230−233) (theoretical *m/z* value of 2082.3). The nature of this species was also proven by automated Edman degradation of the purified peptide, following chromatographic separation. The unique occurrence of this cross-linked component in the digest of the BS_3-cross-linked pACY1 dimer and not in the BS_3-modified pACY1 monomer demonstrated that this species was responsible for the covalent dimerization observed by SDS−PAGE (Figure 5). Lys220 and Lys231 were unambiguously identified as the cross-linked residues. These findings were verified by an independent experiment with endoprotease LysC. In this case, the spectrum of the BS_3-cross-linked pACY1 dimer uniquely showed the presence of a clear signal

Table 4: MALDI-MS Analysis of Acetylated Peptides following the Reaction of pACY1 with Acetic Anhydride and Digestion with 5% Formic Acid at 55 °C[a]

untreated pACY1 (measured *m/z*)	2000:1 Ac_2O:pACY1 ratio (measured *m/z*)	4000:1 Ac_2O:pACY1 ratio (measured *m/z*)	theoretical *m/z*	peptide
1214.6	1214.1	1214.7	1214.5	(113−122)
2018.2	2018.5	2018.8	2018.1	(303−319)
		2060.3	2060.1	(303−319)Ac
2107.4	2107.6	2107.1	2107.2	(94−112)
	2149.6	2149.3	2149.2	(94−112)Ac
2557.0	2557.3	2557.2	2557.0	(233−255)
2686.9	2686.7	2687.2	2686.9	(279−302)
	2728.5	2728.9	2728.9	(279−302)Ac
2814.4	2814.4	2814.7	2814.3	(231−255)
3027.5	3027.1	3027.6	3027.5	(320−346)
3133.3	3133.9	3133.4	3133.6	(146−173)
	3175.4	3175.8	3175.6	(146−173)Ac
3616.3	3615.8	3616.4	3616.0	(182−212)
	3658.3	3658.1	3658.0	(182−212)Ac
3940.4	3940.9	3940.5	3940.7	(113−145)
	3982.4	3982.7	3982.7	(113−145)Ac
4240.8	4240.4	4240.6	4240.9	(233−272)
4498.3	4498.5	4498.6	4498.2	(231−272)
6522.6	6522.4	6522.8	6522.5	(347−406)
10400.7	10400.5	10401.3	10400.8	(1−93)
	10442.7	10442.5	10442.8	(1−93)Ac
		10484.7	10484.8	(1−93)Ac_2
		10526.2[a]	10526.8	(1−93)Ac_3

[a] Traces.

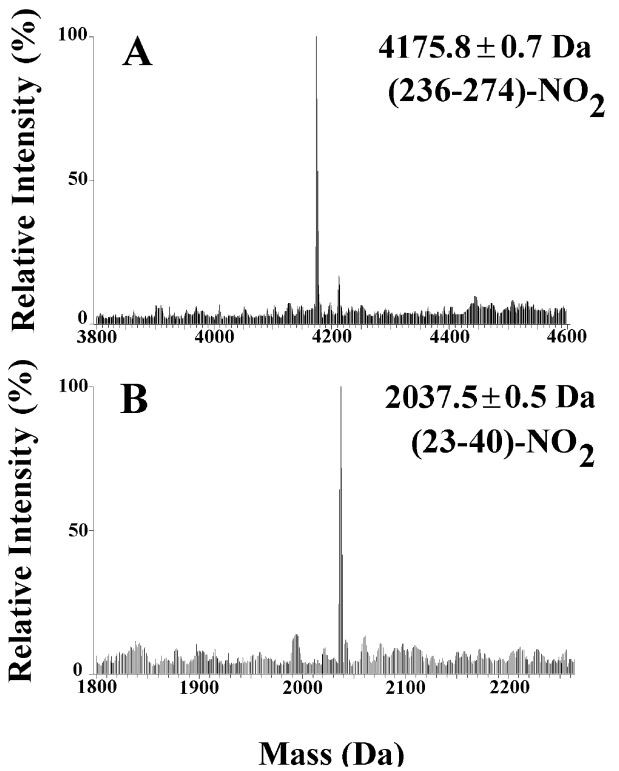

FIGURE 4: Selective nitration of native pACY1 following reaction with tetranitromethane. Panels show the transformed ESI-MS spectra of the modified peptides absorbing at 360 nm obtained from mononitrated pACY1 following digestion with trypsin.

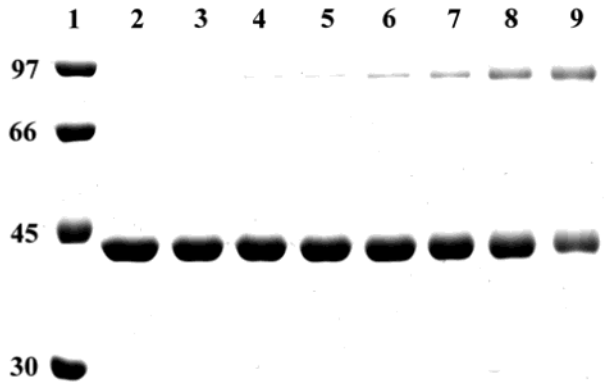

FIGURE 5: SDS−PAGE characterization of native pACY1 cross-linked with BS_3: lane 1, molecular mass markers (phosphorylase B, albumin, ovalbumin, and carbonic anhydrase); lane 2, untreated pACY1; and lanes 3−9, pACY1 incubated with 10-, 50-, 100-, 250-, 500-, 1000-, and 2500-fold molar excesses of BS_3, respectively. Proteins were visualized by Coomassie R250 staining. The 90 kDa BS_3-cross-linked pACY1 dimer and the 45 kDa BS_3-reacted pACY1 monomer (lane 8), together with a 45 kDa untreated pACY1 monomer (lane 1), were excised and digested *in situ* with trypsin before mass spectrometric characterization.

at *m*/*z* 3989.0 that was associated with peptide (218−233)-CO-$(CH_2)_6$-CO-(218−233) (theoretical *m*/*z* value of 3989.6), bearing Lys220 and Lys231 linked by the BS_3 spacer arm.

Molecular Modeling. Gel filtration chromatography experiments definitely demonstrated that porcine ACY1 occurs as a dimer in its native form. On this basis, the ACY1 dimer was modeled using as a template the crystal structure of carboxypeptidase G2, the only dimeric protein with a known structure that is significantly homologous in sequence to pACY1. However, the availability of a single template, its relatively low level of full-length sequence identity (~20%), and a different number of Zn^{2+} ions for the catalytic subunit (two for CBPG and one for pACY1) (*20*, *57*) made sequence alignment a rather crucial step in the modeling of the pACY1 dimer. In fact, standard modeling procedures based on automatic alignment on the CBPG structure failed to produce a dimeric model consistent with experimental data. Significant differences were observed at loop regions and at the dimer interface, where residue accessibility and the calculated distance between Lys220 and Lys231 mainly contrasted with limited proteolysis, chemical modification, and cross-linking-derived data.

To overcome these limitations, we performed a number of different alignments, using different pairwise alignments and information from multiple alignments and from predicted (for target) and observed (for template) secondary structure. In all cases, this approach allowed the identification of discrepancies in the putative site of Zn coordination between pACY1 and CBPG. In fact, either by using multialignment or by including CLUSTALW secondary structure gap penalties in the simple pairwise alignment, not all CBPG metal coordination residues were conserved. The crystal structure of CBPG shows that each subunit of the homodimer contains two zinc atoms. Zn1 is coordinated by the carboxylate oxygen atoms of Asp141 and Glu176, the N^{ϵ} atom of His385, and an oxygen atom from a water molecule. Similarly, Zn2 coordination involves the second oxygen of the Asp141 carboxylate, Glu200, the N^{ϵ} atom of His112, and the same water molecule binding Zn1. In both pairwise alignment with the secondary structure gap penalties and multialignment with hACY, His385 and Glu200 were not conserved. This result was confirmed by extending the multiple alignment to other aminoacylases. These results suggested a possible new type of metal coordination for pACY, involving Asp112, His79, and Glu147 (corresponding to CBPG residues Asp141, His112, and Glu176, respectively), in agreement with the hypothesis of Biagini and Puigserver (*21*).

The alignment that was used (see below) also exhibited a uniform distribution of conserved regions along the sequences, despite the relatively low level of homology between the target and template, with no insertion or deletion longer than three residues (Figure 7). The only exception is represented by a region of the dimerization domain that corresponds to the template 240−254 helix (boxed area in Figure 7), hereafter dubbed HD, where most alignments showed a less conserved profile, with large insertions and low degree of homology. On the basis of all the secondary structure predictions, the corresponding target helix 221−235 was elongated at the N-terminus of five residues in the final models, after deleting the template 234−242 segment. The seven C-terminal residues of the target sequence, for which no template information was available, were excluded from the final modeling.

To compare the different alignments in HD, we built 50 pACY(1−399) dimer models for each: (a) pairwise alignment with CLUSTALW secondary structure gap penalties and (b) manually edited version of (a) with modified alignment and surrounding insertions in the helical region of the dimerization domain (corresponding to template segment 240−254), to avoid a single long insertion (10 residues) following HD, and to provide accessibility to Lys220 and Lys231. The multialignment provided substantially the same result obtained in (a), with the exception of the HD, which, however, was optimized in (b). The criteria used to keep or reject the model obtained from each set were both the quality scores of models and the agreement with the experimental data. Analysis of resulting models confirmed both the predicted difficulties in hosting a large insertion in this region and the eventual wrong orientation of cross-linked Lys residues; therefore, only models from alignment (b) were retained for the subsequent analysis.

The best models (PROCHECK *G*-factors ranging from −0.01 to −0.20) underwent an energy minimization protocol with restraints on the secondary structure and on the Asp112, His79, and Glu147 atoms involved in Zn^{2+} coordination. The structure with the lowest *G*-factor following minimization was also used to build a model of the experimentally detected cross-linked dimer (see Experimental Procedures). Model validation of the minimized structures was performed with different methods. PROCHECK *G*-factors ranged from −0.29 to −0.25, ProsaII combined, pair, and surface *Z*-scores ranged from −8.07 to −7.63, from −5.74 to −5.23, and from −6.27 to −5.49, respectively. These values were quite satisfactory for a structure obtained from homology modeling with a single template with a relatively low level of sequence identity and a dimeric modular structure.

A good overall agreement with the template was suggested by the results of different model superpositions on the template, which also identified eventual intermolecular and interdomain rearrangements occurring in pACY1 models. In particular, after selection of conserved secondary structure

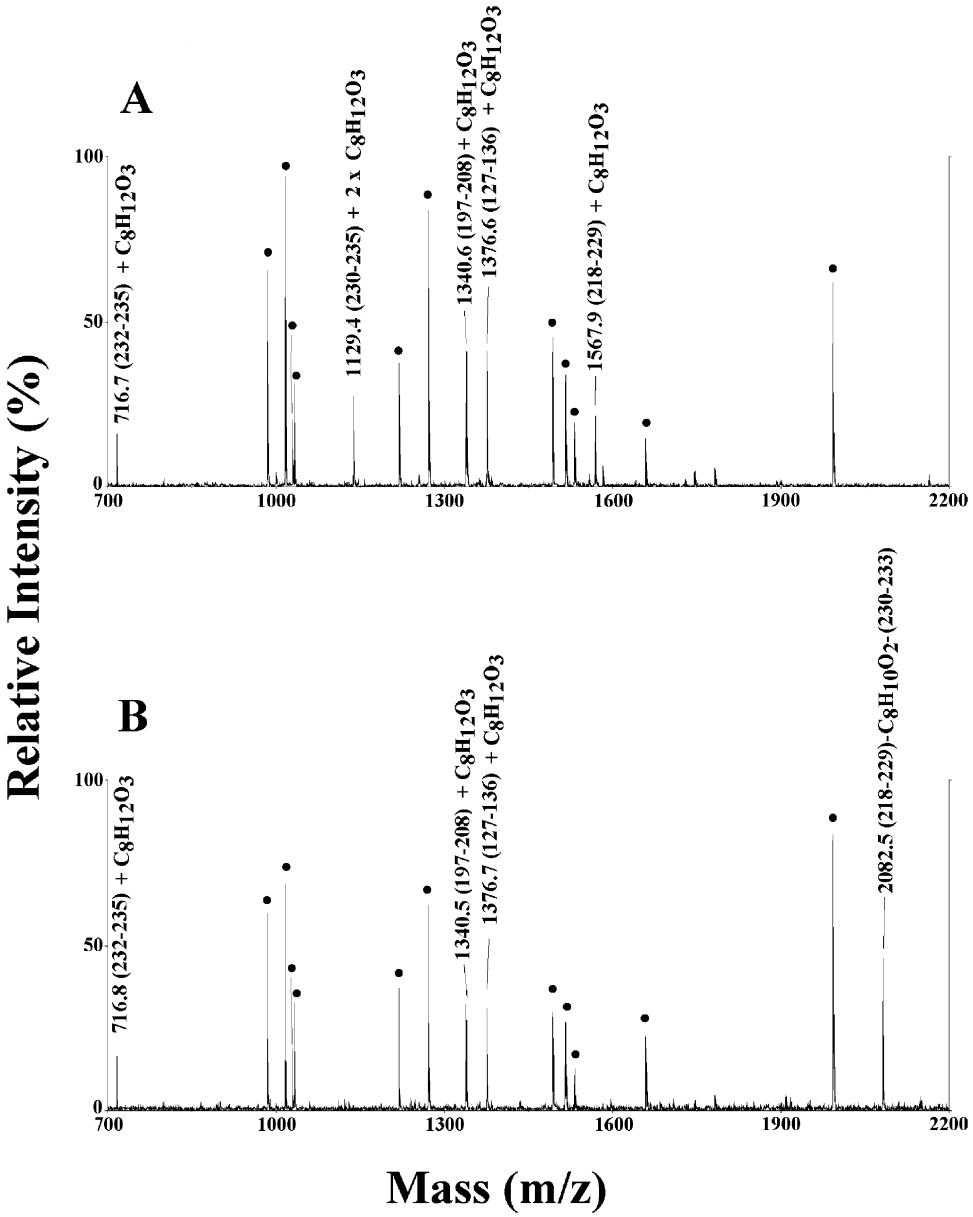

FIGURE 6: Mass spectrometric characterization of cross-linked pACY1. MALDI-MS analysis of the BS_3-reacted pACY1 monomer following digestion with trypsin (A). MALDI-MS analysis of the BS_3-cross-linked pACY1 dimer following digestion with trypsin (B). The panel shows the signals associated with BS_3-modified peptides identified with respect to an unreacted pACY1 sample used as a blank. Dots denote canonic peptides also observed in the pACY1 blank.

regions (corresponding to residues 1−18, 35−43, 154−163, 315−331, and 379−497 of the catalytic domain and residues 191−200, 266−275, 278−292, and 300−307 of the dimerization domain in the target sequence), five different superpositions were performed, by considering the backbone atoms of (a) all the selected ranges for the dimers, (b) all the selected ranges for a single monomer, (c) the selected ranges of dimerization domains for the dimer, (d) the selected ranges of dimerization domains for a single monomer, and (e) the selected ranges of catalytic domains for a single monomer. The corresponding ranges of calculated rms deviations for the minimized models were 1.3−2.0 (a), 0.9−1.0 (b), 0.6−0.8 (c), 0.5−0.7 (d), and 0.7−0.9 (e). These values, indicating a good overall fit of the target structures with the template, show that, even after model relaxation, the deviations were distributed along the whole protein sequence, rather than concentrated on either a single domain or interdomain or intermolecular interfaces.

The main structural elements of the resulting pACY1 dimeric model are shown in Figure 8A. Each monomer consists of two domains: the catalytic domain, a globular, compact subunit that comprises residues 1−188 and 311−399, forming a β-sheet sandwiched between α-helices and a second β-sheet located on the surface, and the dimerization domain, which comprises residues 189−310, folding into a β-sheet flanked on one side by two α-helices. Figure 8B shows the residues exposed on the model surface according to the limited proteolysis, chemical modification, and cross-linking experiments. A good agreement was observed between the calculated structure and the experimental data, with preferential hydrolysis and modification sites belonging to external loops, typically located at the end of structured regions, as easily detected by model visual inspection. These observations found a quantified response in the calculated SAS% for these residues (Table 5). The only low values, observed for Glu309 and Gln311 (both located on the surface

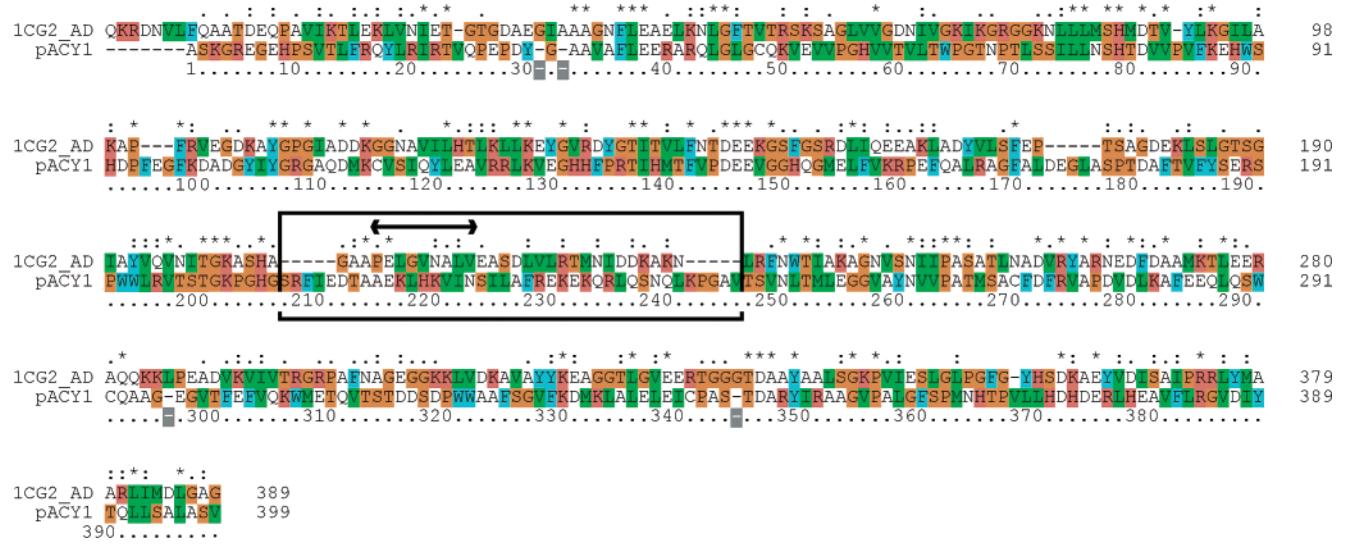

FIGURE 7: Final pACY1−CPBG sequence alignment used in the modeling procedure. Only residues present in the reference PDB structure for the template and included in the modeling for the target are shown. Residues are colored according to the standard CLUSTALX protein scheme (orange for G, P, S, and T, red for H, K, and R, blue for F, W, and Y, green for I, L, M, and V, and white for A, C, D, E, N, and Q). Residue conservation is encoded on the top row of the alignment. Asterisks denote positions which have a single, fully conserved residue. Dots denote the following "strong" groups as fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, and FYW. Double dots denote the following "weaker" groups as fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, and HFY. A box surrounds the HD region discussed in the text. A double-headed arrow shows the template residues omitted in the modeling procedure.

of a shallow cavity at the junction between the catalytic and dimerization domains of each monomer), were strongly dependent by even minor rearrangements in interdomain interfaces, thus suggesting either a slightly different relative orientation of the two domains or the presence of a rather flexible junction in this region. Furthermore, the congruency between the calculated structure and the experimental data was also evident at the dimerization domain, where the Lys residues involved in the dimer cross-linking occurred at a distance of 9.5 Å, perfectly compatible with the length of the BS_3 spacer arm (Figure 8C). Although close to the interacting regions of each monomer, these amino acids were still accessible to the cross-linking reagent and separated by a region that is able to easily accommodate the methylene groups of the spacer arm via positive hydrophobic interactions.

DISCUSSION

Knowledge of the tertiary and quaternary structure of proteins is a fundamental prerequisite for the correct understanding of their functional mechanism. This makes the successful prediction of a protein tertiary and quaternary structure a central issue in proteomics, in cases in which the experimental structural characterization is not yet available. However, although prediction and modeling techniques have greatly improved in the recent years (*58*), simple and effective tools are needed for experimentally testing tertiary and quaternary structural models. This is more evident for homomultimeric proteins for which information on the three-dimensional structure of each monomer, the number of interacting monomers, and the nature of the regions present at the molecular interface are needed to provide confidence in the prediction results. In these cases, automated procedures conventionally used to generate tertiary models of monomeric proteins sometimes are not effective, and different integrated methodologies are necessary to provide a successful prediction of the protein structure. For this reason, a multimethodological approach combining size exclusion chromatography, limited proteolysis, selective chemical modification, and cross-linking experiments, in conjunction with advanced mass spectrometry and modeling procedures, was employed to probe the solution structure of porcine ACY1. A similar approach has been successfully used to achieve a topological description of the molecular surface in monomeric proteins or the interacting regions in protein complexes (*25*−*38*).

As a first stage, the aggregation and redox state of the enzyme were investigated by size exclusion chromatography measurements and extensive alkylation experiments, demonstrating that ACY1 exists in solution as a dimer with all cysteines present in a reduced state. These findings definitely ruled out previous reports on the existence of pACY1 in the monomeric state and containing the Cys270−Cys292 disulfide (*24*). Similarly, on the basis of the high level of sequence identity between the murine and porcine enzymes, a possible occurrence of rat ACY1 as a trimer was considered poorly probable (*22*). Furthermore, a series of limited proteolysis experiments with different proteases in conjunction with mass spectrometric analysis of the reaction products allowed us to define the preferential cleavage sites on protein surface. Selective modification at specific lysines and tyrosines, followed by mass mapping experiments, identified the amino acids displaying the highest chemical reactivity, occurring in the most exposed regions on the molecular surface. The overall results (Table 2) showed a good agreement between the different approaches and indicated that the 26 hydrolysis and/or modification sites were distributed all along the pACY1 polypeptide chain. In parallel, cross-linking experiments on native pACY1 confirmed the dimeric nature of the enzyme and identified Lys220 and Lys231 as the intermonomer cross-linked residues. These findings strongly suggested that these amino

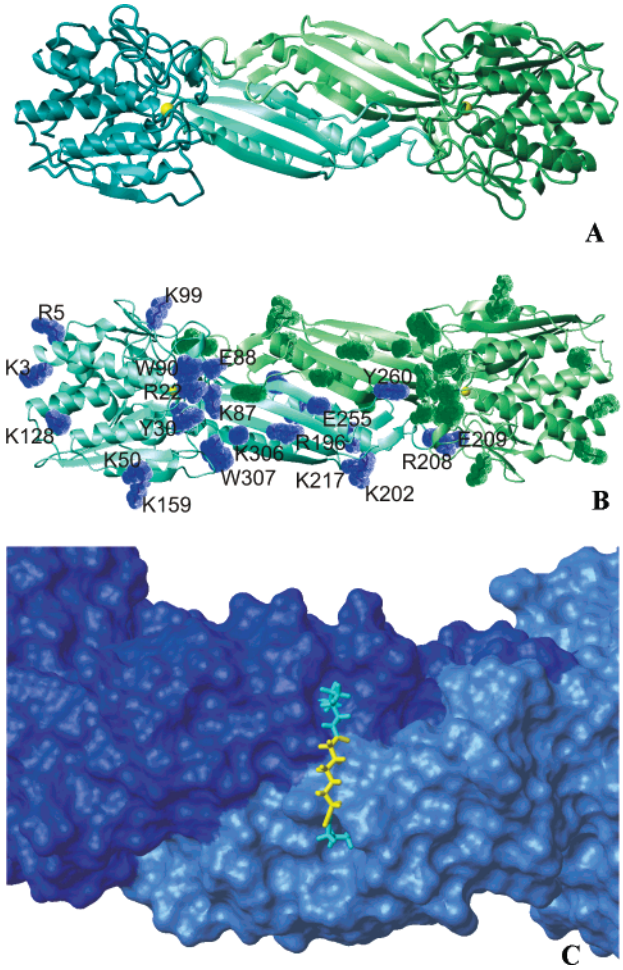

FIGURE 8: Different view of the three-dimensional model proposed for dimeric porcine ACY1. (A) Ribbon representation of the final pACY1 dimer model. The two monomers are cyan and green with their catalytic domains in darker shades. Zinc ions are yellow. (B) Experimentally determined exposed residues in the pACY1 dimer. A ribbon plot of the pACY1 dimer is shown. The proposed view allows visualization of most of the residues found as exposed on pACY1 (see Table 2); these amino acids are labeled on one monomer and are represented as stick bonds and dotted van der Waals surfaces. (C) Model of a cross-linked dimer of pACY1. The solvent accessible surfaces of the monomers are shown in two shades of blue. Lys220 (cyan) from monomer A (dark blue), Lys231 (cyan) from monomer B (light blue), and the OC-$(CH_2)_6$-CO spacer arm (yellow) are represented with stick bonds.

acids are located close to the dimer interface in the native protein, at a relative distance of lower than 11.4 Å.

Standard modeling procedures from the crystal structure of carboxypeptidase G2, based on automatic alignment, failed to produce a dimeric model consistent with these experimental results. Significant discrepancies were observed at the interface region and mainly contrasted with accessibility and cross-linking-derived data. Therefore, a series of optimized pACY1 three-dimensional models were calculated on the basis of different pairwise and multiple-sequence alignments, also using information derived from predicted and observed secondary structures. All models underwent energy minimization, and structure validation was performed with different methods. The refined model reported in Figure 8A appeared to be compatible with all experimental outputs. In fact, a visual inspection of the location on the dimer surface for the experimentally identified exposed residues and the analysis of their calculated SAS values showed a remarkably

Table 5: Calculated Solvent Accessibility of Selected Amino Acid Residues in the Refined Model of Dimeric pACY1[a]

accessibility of experimentally determined proteolysis/modification sites		accessibility of amino acids occurring at the dimerization interface	
residue	SAS%	residue	SAS%
Lys3	26.5	Glu174	11.9
Arg5	43.7	Arg190	6.9
Arg22	32.0	Trp194	14.0
Tyr30	46.3	Gly206	0.6
Lys50	33.4	Ser207	12.7
Lys87	44.3	Leu218	1.9
Glu88	40.1	His219	10.6
Trp90	24.8	Ile222	1.0
Lys99	39.5	Asn223	10.7
Lys128	28.9	Ile225	0.6
Lys159	36.8	Leu226	0.6
Arg196	36.0	Arg229	0.4
Lys202	42.9	Glu230	10.4
Arg208	57.2	Leu236	15.2
Phe209	37.5	Lys242	5.1
Lys217	36.6	Ala245	8.5
Lys220	23.1	Val246	12.6
Lys231	26.9	Ser248	0.9
Lys233	23.0	Asn250	2.7
Glu255	30.8	Leu251	1.1
Tyr260	34.0	Thr252	0.1
Lys282	43.8	Met253	13.3
Lys306	41.4	Leu254	1.1
Trp307	40.4	Gly256	2.4
Glu309	1.5	Gly257	7.2
Gln311	4.4	Val263	1.8
		Met267	0.8
		Cys270	5.8
		Asp272	0.1
		Arg274	2.2

[a] Accessibility is expressed as the percentage of maximum SAS (SAS%); this parameter is defined as the actual SAS value for a residue in the pACY1 dimer divided by the SAS value of the same amino acid in its current conformation calculated for a hypothetical isolated state times 100. Arg208, Phe209, and Tyr260 were located close to the border of the predicted dimer interface.

good overall agreement between model predictions and experimental results (Figure 8B). At the same time, this model accounted for a distance and a relative orientation of Lys220 and Lys231 in two different monomers that are correct for formation of the experimentally detected chemical cross-link. In fact, the model of the cross-linked dimer (Figure 8C) showed that only minimal rearrangements in the terminal atoms of both Lys residues were required to obtain an ideal arrangement and a fully relaxed structure for the adduct. This result was made even more critical by the lack of direct structural information from the template in this region, where the structural features of the model were mainly driven by statistical information derived from multiple alignments and secondary structure predictions. In particular, the analysis of the model structure in this region revealed that all residues predicted to occur inside the dimer interface (Table 5) were not recognized or labeled by proteases or chemical reagents. An even more remarkable feature of the model is the observation that many residues predicted to be close to the border of the interface (i.e., involved in intermonomer contacts but exhibiting SAS% values larger than 20; see Table 5) were recognized or labeled by proteases or chemical reagents. Moreover, previous biochemical investigations of the pACY1 sensitivity to thiol-directed inhibitors were easily interpreted on the basis of the model

proposed here (*59*). In fact, a specific residue, Cys115, occurred close to the active site region in the model and should be hypothesized to be a possible target of these reagents (data not shown). According to these results, we conclude that the structure reported in Figure 8 should provide a good three-dimensional description of dimeric pACY1 in solution.

A general overall agreement with the carboxypeptidase G2 molecular architecture was suggested by the results of different pACY1 superpositions with CBPG. Despite the relatively low level of sequence identity between pACY1 and CBPG, conserved regions were uniformly distributed along the polypeptide chains and a common number of secondary structure elements resulted from the comparison of the final pACY1 model and the CBPG structure. Insertions and deletions were small in size on average and, not considering the HD region, located in external loops of the catalytic domain, far from both the dimerization domain and the putative Zn-binding site. Significant differences were observed in the dimerization domain, where the helix stabilizing intermonomer interaction was longer than the corresponding helix in CBPG. In this region, five additional residues are located, without affecting the overall fold of the dimerization region. Moreover, additional residues with respect to CBPG were arranged in two five-residue loops at both ends of the HD helix, occupying rather empty regions, and potentially improving both intra- and intermolecular nonbonded interactions between the dimerization and catalytic domains. The predicted intermonomer contact surface in pACY1 appeared to be increased by the larger size of the HD region, with an average calculated increment of 1018 $Å^2$ in the different models (corresponding to 5.6% of the total monomer target SAS burial after dimerization) over carboxypeptidase G2.

Furthermore, the structure presented three predicted Zn-binding residues (His79, Asp112, and Glu147) in a spatial arrangement fully compatible with a tetrahedral coordination of one Zn ion, with the fourth coordination atom eventually provided by a solvent molecule, or by a bidentate coordination from either acidic residue. The geometrical properties of this metal binding site are similar to those already observed for other metalloenzymes (*60*). The above-mentioned residues are all present in the conserved sequence motifs [(S,G,A)-**H**-X-D-X-V, G-X-X-**D**, and X-E-**E**], common to animal or plant aminoacylases, bacterial succinyldiaminopimelate desuccinylases, bacterial and plant acetylornithine deacetylases, bacterial carboxypeptidase G2s, and yeast carboxypeptidase S (*21*). The sequences of these metallopeptidases are appreciably homologous with that of pACY1 (including the dimerization domain), and the enzymes have a dimeric nature (except yeast carboxypeptidase S) and zinc or cobalt as a cofactor. At the moment, the number of metal ions per catalytic subunit for these proteins has been determined only for aminoacylase 1 (one metal) (*57*) and carboxypeptidase G2 (two metals) (*20*). Sequence alignment studies revealed that the polypeptide region containing the three conserved amino acid motifs are significantly homologous to other peptidases whose three-dimensional structure has been determined so far; this is the case for *Aeromonas proteolytica* aminopeptidase and *Streptomyces griseus* aminopeptidase (*61*, *62*). Although all these enzymes are monomeric species and do not have the dimerization domain sequence, they present an active site containing two zinc ions per catalytic subunit. The amino acids involved in coordination of the second metal do not seem to be fully conserved in pACY1. On this basis, we believe that the number of the metal ions occurring in the pACY1 active site and the geometry of their coordination are important features, to be further investigated in the future by detailed X-ray crystallography studies. These data, together with the structural considerations reported in this paper, will definitely elucidate the structure−function relationships for this metalloprotease, with altered expression levels in different renal and small cell lung carcinomas. New investigations on possible biochemical mechanisms in which pACY1 and its substrates should be involved will determine if this protease has a specific role in neoplacy or its altered activity is simply associated with a generic loss of genetic material on chromosome 3p21, a region where several other genes involved in the regulation of intracellular protein degradation may be perturbed.

It is well-known that a major problem in predicting protein structure by homology modeling is that the selected sequence alignment may not be the best with respect to the correct equivalence of residues assessed by structural or functional criteria. This is even more evident for model of multimeric proteins where an additional energy contribution associated with quaternary structure is essential for structure stabilization and incorrect alignment of multimerization domains can totally disturb proper folding. In this case, it has been proposed (*63*) that a number of suboptimal alignments should be generated and examined before selecting the appropriate one for model building. In this case, an effective procedure that is able to rapidly filter the different possible alignments is strongly demanded. This goal still represents the major obstacle in large-scale automatic modeling, where manual inspection and/or careful analysis of all possible suboptimal sequence alignments is prohibitive.

The results presented in this paper are a good example of how fast and reliable techniques such as limited proteolysis, selective chemical modification, and cross-linking experiments can be used to highlight problems in automatic model building and to direct an optimized modeling procedure toward a refined model fully compatible with all experimental observations. The development of automatic algorithms, including information derived from this type of experimental data, will provide a higher degree of confidence for computer-generated structural models.

REFERENCES

1. Gade, W., and Brown, J. L. (1981) *Biochim. Biophys. Acta 662*, 86−93.
2. Heese, D., Loffler, H., and Rohm, K. H. (1990) *Biol. Chem. Hoppe-Seyler 369*, 559−566.
3. Jones, W., Scaloni, A., and Manning, J. M. (1994) *Methods Enzymol. 244*, 227−231.
4. Scaloni, A., Jones, W. M., Barra, D., Pospischil, M., Sassa, S., Popowicz, A., Manning, L., Schneewind, O., and Manning, J. M. (1992) *J. Biol. Chem. 267*, 3811−3818.
5. Jones, W. M., Manning, L. R., and Manning, J. M. (1986) *Biochem. Biophys. Res. Commun. 139*, 244−250.
6. Chongcharoen, K., and Sharma, K. K. (1998) *Biochem. Biophys. Res. Commun. 247*, 136−141.
7. Cook, R. M., Burke, B. J., Buchhagen, D. L., Minna, J. D., and Miller, Y. E. (1993) *J. Biol. Chem. 268*, 17010−17017.
8. Mitta, M., Ohnogi, H., Yamamoto, A., Kato, I., Sakiyama, F., and Tsunasawa, S. (1992) *J. Biochem. 112*, 737−742.

9. Miyagi, M., Sakiyama, F., Kato, I., and Tsunasawa, S. (1995) *J. Biochem. 118*, 771−779.
10. Erlandsson, R., Boldog, F., Persson, B., Zabarovsky, E. R., Allikmets, R. L., Sumegi, J., Klein, G., and Jornvall, H. (1991) *Oncogene 6*, 1293−1295.
11. Jones, W., Scaloni, A., Bossa, F., Popowicz, A., Schneewind, O., and Manning, J. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2194−2198.
12. Ginzinger, D. G., Shridhar, V., Baldini, A., Taggart, R. T., Miller, O. J., and Smith, D. I. (1992) *Am. J. Hum. Genet. 50*, 826−833.
13. Miller, Y. E., Minna, J. D., and Gazdar, A. F. (1989) *J. Clin. Invest. 83*, 2120−2124.
14. Balabanov, S., Zimmermann, U., Protzel, C., Scharf, C., Klebingat, K. J., and Walther, R. (2001) *Eur. J. Biochem. 268*, 5977−5980.
15. Scaloni, A., Jones, W. M., Pospischil, M., Schneewind, O., Popowicz, A., Bossa, F., Graziano, S., and Manning, J. M. (1992) *J. Lab. Clin. Med. 120*, 546−552.
16. Naylor, S. L., Johnson, B. E., Minna, J. D., and Sakaguchi, A. Y. (1987) *Nature 329*, 451−454.
17. McNail-Killary, A., Wolf, M. E., Giambernardi, T. A., and Naylor, S. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 10877−10881.
18. Fulop, V., Bocskei, Z., and Polgar, L. (1998) *Cell 94*, 161−170.
19. Feese, M., Scaloni, A., Jones, W. M., Manning, J. M., and Remington, S. (1993) *J. Mol. Biol. 233*, 546−549.
20. Rowsell, S., Pauptit, R. A., Tucker, A. D., Melton, R., Blow, D. M., and Brick, P. (1997) *Structure 5*, 337−347.
21. Biagini, A., and Puigserver, A. (2001) *Comp. Biochem. Physiol., Part B 128*, 469−481.
22. Giardina, T., Perrier, J., and Puigserver, A. (2000) *Eur. J. Biochem. 267*, 6249−6255.
23. Spormann, D. O., Heim, J., and Wolf, D. H. (1991) *Eur. J. Biochem. 197*, 399−405.
24. Palm, G. J., and Rohm, K. H. (1995) *J. Protein Chem. 14*, 233−240.
25. Zappacosta, F., Pessi, A., Bianchi, E., Venturini, S., Sollazzo, M., Tramontano, A., Marino, G., and Pucci, P. (1996) *Protein Sci. 5*, 802−813.
26. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) *Folding Des. 266*, 223−230.
27. Suckau, D., Kohl, J., Karwath, G., Schneider, K., Casaretto, M., Bitter-Suermann, D., and Przybylski, M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9848−9852.
28. Zhao, Y., Muir, T., Kent, S. B., Tischer, E., Scardina, J. M., and Chait, B. T. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 4020−4024.
29. Cohen, S. L., Ferrè-D'amare, A. R., Burley, S. K., and Chait, B. T. (1995) *Protein Sci. 4*, 1088−1099.
30. Scaloni, A., Miraglia, N., Orrù, S., Amodeo, P., Motta, A., Marino, G., and Pucci, P. (1998) *J. Mol. Biol. 277*, 945−958.
31. Scaloni, A., Monti, M., Acquaviva, R., Tell, G., Damante, G., Formisano, S., and Pucci, P. (1999) *Biochemistry 38*, 64−72.
32. Suckau, D., Mak, M., and Przybylski, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5630−5634.
33. Glocker, M. O., Borchers, C., Fiedler, W., Suckau, D., and Przybylski, M. (1994) *Bioconjugate Chem. 5*, 583−590.
34. Zappacosta, F., Ingallinella, P., Scaloni, A., Pessi, A., Bianchi, E., Sollazzo, M., Tramontano, A., Marino, G., and Pucci, P. (1997) *Protein Sci. 6*, 1901−1909.
35. Leite, J. F., and Cascio, M. (2002) *Biochemistry 41*, 6140−6148.
36. Ploug, M., Rahbek-Nielsen, H., Ellis, V., Roepstorff, P., and Dano, K. (1995) *Biochemistry 34*, 12524−12534.
37. Bennett, K. L., Kussmann, M., Bjork, P., Godzwon, M., Mikkelsen, M., Sorensen, P., and Roepstorff, P. (2000) *Protein Sci. 9*, 1503−1518.
38. McLafferty, F. W., Fridriksson, E., Horn, D., Lewis, M. A., and Zubarev, R. A. (1999) *Science 284*, 1289−1290.
39. Vilardo, P. G., Scaloni, A., Amodeo, P., Barsotti, C., Cecconi, I., Cappiello, M., Lopez Mendez, B., Rullo, R., Dal Monte, M., Del Corso, A., and Mura, U. (2001) *Biochemistry 40*, 11985−11994.
40. Allegrini, S., Scaloni, A., Ferrara, L., Pesi, R., Pinna, P., Camici, M., Eriksson, S., and Tozzi, M. G. (2001) *J. Biol. Chem. 276*, 33526−33532.
41. Peitsch, M. C. (1996) *Biochem. Soc. Trans. 24*, 274−279.
42. Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol. 234*, 779−815.
43. Rowsell, S., Pauptit, R. A., Tucker, A., Melton, R. G., Blow, D. M., and Brick, P. (1997) *Structure 5*, 337−347.
44. Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
45. Kabsch, W., and Sander, C. (1983) *Biopolymers 22*, 2577−2637.
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
47. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1999) *AMBER 6*, University of California, San Francisco.
48. Weiner, S. J., Kollman, P. A., Case, D. A., Chandra Singh, U., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. K. (1984) *J. Am. Chem. Soc. 106*, 765−784.
49. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986) *J. Comput. Chem. 7*, 230−252.
50. Sippl, M. J. (1993) *Proteins 17*, 355−362.
51. Lee, B., and Richard, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
52. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
53. Petersson, A. S., Steen, H., Kalume, D. E., Caidahl, K., and Roepstorff, P. (2001) *J. Mass Spectrom. 36*, 616−625.
54. Ji, T. H. (1983) *Methods Enzymol. 91*, 580−609.
55. Hermanson, G. T. (1996) *Bioconjugate techniques*, Academic Press, San Diego.
56. Young, M., Tang, N., Hempel, J., Oshiro, C., Taylor, E., Kuntz, I., Gibson, B., and Dollinger, G. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5802−5806.
57. Heese, D., Berger, S., and Rohm, K. H. (1990) *Eur. J. Biochem. 188*, 175−180.
58. Moult, J., Hubbard, T., Bryant, S. H., Fidelis, K., and Pedersen, J. T. (1999) *Proteins S1*, 2−6.
59. Heese, D., and Rohm, K. H. (1989) *Biol. Chem. Hoppe-Seyler 370*, 607−612.
60. Castagnetto, J. M., Hennessy, S. W., Roberts, V. A., Getzoff, E. D., Tainer, J. A., and Pique, M. E. (2002) *Nucleic Acids Res. 30*, 379−382.
61. Greenblatt, H. M., Almog, O., Maras, B., Spungin-Bialik, A., Barra, D., Blumberg, S., and Shoham, G. (1997) *J. Mol. Biol. 265*, 620−636.
62. Chevrier, B., Schalk, C., D'Orchymont, H., Rondeau, J. M., Moras, D., and Tarnus, C. (1994) *Structure 2*, 283−291.
63. Saqi, M. A., Bates, P. A., and Sternberg, M. J. (1992) *Protein Eng. 5*, 305−311.

BI0206715