3D Structure of *Sulfolobus solfataricus* Carboxypeptidase Developed by Molecular Modeling is Confirmed by Site-Directed Mutagenesis and Small Angle X-Ray Scattering

Emanuela Occhipinti,* Pier Luigi Martelli,† Francesco Spinozzi,‡ Federica Corsi,‡ Cristina Formantici,* Laura Molteni,* Heintz Amenitsch,§ Paolo Mariani,‡ Paolo Tortora,* and Rita Casadio†

*Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, I-20126 Milano, Italy; †Laboratory of Biocomputing, CIRB/Department of Biology, University of Bologna, I-40126 Bologna, Italy; ‡Istituto di Scienze Fisiche and INFM, Università di Ancona, I-60131 Ancona, Italy; and §Sincrotrone di Trieste S.p.A., I-34016 Basovizza, Trieste, Italy

ABSTRACT *Sulfolobus solfataricus* carboxypeptidase (CPSso) is a thermostable zinc-metalloenzyme with a M_r of 43,000. Taking into account the experimentally determined zinc content of one ion per subunit, we developed two alternative 3D models, starting from the available structures of *Thermoactinomyces vulgaris* carboxypeptidase (Model A) and *Pseudomonas* carboxypeptidase G2 (Model B). The former enzyme is monomeric and has one metal ion in the active site, while the latter is dimeric and has two bound zinc ions. The two models were computed by exploiting the structural alignment of the one zinc- with the two zinc-containing active sites of the two templates, and with a threading procedure. Both computed structures resembled the respective template, with only one bound zinc with tetrahedric coordination in the active site. With these models, two different quaternary structures can be modeled: one using Model A with a hexameric symmetry, the other from Model B with a tetrameric symmetry. Mutagenesis experiments directed toward the residues putatively involved in metal chelation in either of the models disproved Model A and supported Model B, in which the metal-binding site comprises His^{108}, Asp^{109}, and His^{168}. We also identified Glu^{142} as the acidic residue interacting with the water molecule occupying the fourth chelation site. Furthermore, the overall fold and the oligomeric structure of the molecule was validated by small angle x-ray scattering (SAXS). An ab initio original approach was used to reconstruct the shape of the CPSso in solution from the experimental curves. The results clearly support a tetrameric structure. The Monte Carlo method was then used to compare the crystallographic coordinates of the possible quaternary structures for CPSso with the SAXS profiles. The fitting procedure showed that only the model built using the *Pseudomonas* carboxypeptidase G2 structure as a template fitted the experimental data.

INTRODUCTION

In recent years, plenty of work was carried out in one of our laboratories to characterize a carboxypeptidase from the extreme thermophilic archaeon *Sulfolobus solfataricus* (CPSso) (Colombo et al., 1992; Villa et al., 1993; Colombo et al., 1995; Bec et al., 1996; Mombelli et al., 1996). Our previous investigations initially led to purification and characterization of the enzyme from its natural source (Colombo et al., 1992), then to cloning of the enzyme-encoding gene and its expression in *Escherichia coli* (Colombo et al., 1995). CPSso was shown to be a zinc-metalloprotease, with a molecular mass of ~43 kDa, deduced from the gene sequence and in good agreement with the value assessed in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Colombo et al., 1992). Gel-filtration data indicated that the enzyme has a tetrameric structure (Colombo et al., 1992). This, however, does not represent a conclusive evidence in support of a defined aggregation state. The recombinant protein was indistinguishable from the natural form in terms of catalytic activity and chemical-physical properties (Colombo et al., 1995). As expected, CPSso displayed a remarkable heat (Villa et al., 1993; Colombo et al., 1995) and pressure resistance (Bec et al., 1996; Mombelli et al., 1996), up to 85°C and at least 400 MPa, respectively. Nevertheless, the structural features responsible for these properties are not completely understood. Our previous investigations showed that salt bridges and the bound metal ion play a role in thermostability (Villa et al., 1993), whereas piezostability has been related to possible hydration of hydrophobic patches exposed to the solvent (Bec et al., 1996; Mombelli et al., 1996). We found that the enzyme also has singular catalytic properties. First, unlike most known carboxypeptidases (Zuhlke et al., 1976; Makinen et al., 1984; Skidgel and Erdös, 1998), it removes any amino acid from the C-terminus of short peptides, with the sole exception of proline, and it also hydrolyzes N-blocked amino acids, thus acting as an aminoacylase; second, despite its thermophilicity it maintains a significant fraction of its maximal activity even at room temperature; third, it also maintains a significant activity in organic solvents at high concentrations (Colombo et al., 1992). In particular, we found that CPSso is fairly stable and active in dimethylformamide up to 60% and dimethylsulfoxide up to 70%; as a result we could exploit the enzyme to achieve synthesis of N-blocked amino acid in organic media (manuscript in preparation). Our interest in this enzyme is also justified by the remarkable biotechnological potential of aminoacylases (Tewary, 1990).

Submitted October 18, 2002, and accepted for publication April 8, 2003.

Address reprint requests to Dr. Paolo Tortora, Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano, Italy. Tel.: +39-02-6448-3401; Fax: +39-02-6448-3565; E-mail: Paolo.Tortora@unimib.it.

0006-3495/03/08/1165/11 $2.00

Of course, possible future developments of our investigations would greatly benefit from the availability of the three-dimensional structure of CPSso. This would provide a better understanding of the catalytic mechanism, as well as of the determinants of enzyme specificity and stability. It also would pave the way to a strategy of protein engineering aimed at further improving enzyme stability in aqueous and organic media, thus providing better performance in processes of chemical synthesis. Unfortunately, crystallization trials have failed so far. On the other hand, structures of closely related proteins are not yet available. Actually, CPSso belongs to a family of metallopeptidases (M40) that also includes several hydrolytic enzymes of unknown structure from eukaryotic and eubacterial sources (Tortora and Vanoni, 1998; Barrett et al., 1998). They mainly catalyze the hydrolytic removal of small moieties such as succinate, β-alanine, acyl groups and carbamate from the α-amino group of amino acid residues, and display a sequence identity to CPSso in the range of 25%–40%. However, for none of these enzymes the structure is available to date. The enzyme of known structure whose sequence is more closely related to CPSso is *Pseudomonas* carboxypeptidase G2 (CPG2), a dimeric protein belonging to a different family (M20), but part of the same clan (MH) of CPSso (Barrett et al., 1998; Rowsell et al., 1997), which shares with the archaebacterial enzyme 21% of identical residues and 32% of conserved residues.

In the database of known protein structures, the carboxypeptidase from *Thermoactinomyces vulgaris* (Teplyakov et al., 1992) (CPTvu) is also available. This is a monomer, containing only one catalytic zinc. By taking advantage of the structural alignment of the one zinc- with the two zinc-containing active sites, and using a threading procedure we computed two possible three-dimensional structures of CPSso. This approach led initially to the development of two models. One has a structure resembling that of CPTvu with one bound zinc ion in the active site (Model A); the other model (Model B) envisages a structure resembling that of CPG2 but, unlike the template, has one instead of two bound zinc ions. With these two models, two different quaternary structures have been modeled: one using Model A with a hexameric symmetry, the other from Model B with a tetrameric symmetry (http://pqs.ebi.ac.uk).

Experiments of mutagenesis directed toward residues putatively involved in zinc chelation, substantially confirmed Model B and also provided suggestions to its further refinement. Likewise, they contributed to disproving Model A.

Finally, the molecule was investigated by small angle x-ray scattering (SAXS), a powerful technique providing information on shape, size and aggregation state of biological macromolecules in solution. In particular, as it is very sensitive to geometric shape, SAXS is an appropriate technique for studying protein conformational changes and aggregation processes that occur in solution (Trewhella, 1997; Kataoka et al., 1995; Pollack et al., 1999; Perez et al., 2001; Cinelli et al., 2001). Previous works also show that SAXS data can be conveniently used to confirm protein-computed structures (Casadio et al., 1999; Mariani et al., 2000). Thus, an ab initio original approach, based on the multiple expansion method (Svergun and Stuhrmann, 1991; Spinozzi et al., 1998), was used to reconstruct the shape of the CPSso in solution from the SAXS experimental curves. The results indicate that the enzyme has a tetrameric arrangement.

With the models at hand, two different quaternary structures can be modeled: one using Model A with a hexameric symmetry, the other from Model B with a tetrameric symmetry (http://pqs.ebi.ac.uk). The Monte Carlo method was then used to compare the crystallographic coordinates of the possible molecular structures for CPSso with the SAXS profiles. The fitting procedure shows that only the quaternary model built using the as a template, and modified to accommodate only one zinc ion in the catalytic site (Model B), agrees with the experimental data.

MATERIALS AND METHODS

Construction of expression vectors coding for histidine-tagged wild-type and mutated CPSso

The gene coding for wild-type CPSso was previously expressed in *E. coli* using the vector pT7-7 (Colombo et al., 1995). The gene was amplified by PCR using the following oligonucleotides, where restriction sites for subcloning are in italics and the sequence coding for $Gly(His)_6$ immediately downstream the start codon is shown in bold: 5′-CG*CCATGG***GTCATCACCATCACCATCAC**GATTTAGTTGATAAGTTAAAA-3′ (upstream); 5′-CG*GAATTC*TTATTTATTACTGAACTTTACTGCCAATAATGCGTG (downstream). The gene was excised with *Nco*I and *Eco*RI and inserted into the same restriction sites of the pET-11d plasmid. The plasmid was transformed into the *E. coli* expression strain BL21(DE3) [pLysE]. Genes coding for mutated CPSso enzymes were produced starting from the wild type and using the commercially available Quick Change mutagenesis kit (Stratagene, La Jolla, CA).

CPSso activity and protein assays

CPSso activity was assayed continuously using benzoyl-arginine as a substrate. The assay mixture contained 0.1 M potassium Mes pH 6.5, 0.1 mM benzoyl-arginine, in a total volume of 1.5 ml. After a 5-min preincubation at 60°C the enzyme was added, and the decrease in absorbance monitored at the same temperature and 239 nm. At this wavelength, the decrease of the molar absorption coefficient was 6240 M^{-1} cm^{-1}. One unit of enzyme activity is defined as the amount of enzyme that hydrolyzes 1 μmol of substrate/min under the standard assay conditions. Authentic and histidine-tagged CPSso enzyme did not show significant differences in terms of catalytic behavior. Protein content was determined either using the Coomassie protein assay reagent from Pierce (Rockford, IL) and bovine plasma immunoglobulin G as the standard protein, or on the basis of the extinction coefficient at 280 nm, as reported (Mach et al., 1992).

Affinity purification of histidine-tagged CPSso

Wild-type and mutated histidine-tagged CPSso were purified by Ni-chelate affinity chromatography. *E. coli* cells were grown under shaking at 37°C in 6 l of LB medium containing 0.1% glucose, 50 μg/ml ampicillin, 25 μg/ml

choramphenicol until A_{600} reached 0.8. Induction was then carried out for 3.5 h with 0.4 mM isopropyl β-D-thiogalactopyranoside. Cells were harvested (~20 g wet weight), washed with 50 mM Tris-HCl, pH 8.0, 0.1 M NaCl, 10 μM $ZnCl_2$, and resuspended in 3 volumes of 50 mM Tris-HCl, pH 8.0, 1 mM PMSF, 2 mM 2-mercaptoethanol, complete EDTA-free protease inhibitor (1 tablet/50ml; Roche, Mannheim, Germany), 11 units/ml bovine pancreas DNase (Sigma, St. Louis, MO). Cells were disrupted by sonication and centrifuged at 40000 $\times$ *g* for 30 min. After addition of 10% (by vol) glycerol and 10 mM imidazole, the supernatant was loaded at a flow rate of 1 ml/min onto a Ni-NTA Superflow (Qiagen, Hilden, Germany) column (bed volume 5 ml) preequilibrated with 50 mM potassium Mes, pH 6.0, 10 mM imidazole, 5 mM 2-mercaptoethanol, 10% (by vol) glycerol. The column was washed with 10 volumes of the same buffer and the enzyme eluted with a 60-ml continuous linear imidazole gradient, 10 mM to 1 M, in the same buffer. 4-ml fractions were collected. Fractions displaying the highest purity (at least 95%, as assessed in SDS-PAGE; Laemmli, 1970) were employed for activity and zinc content determinations, as well as for SAXS experiments. Before metal assay, suitable aliquots of the fractions were concentrated in Centricon YM-30 microconcentrators (Millipore, Bedford, MA) to ~20 mg/ml.

Determination of enzyme-bound zinc

For enzyme-bound zinc determinations, CPSso of the best purity (at least 95%, as assessed in SDS-PAGE) was first stripped of the metal by dialysis as previously reported (Colombo et al., 1992). This treatment completely abolished activity. Zinc was then reconstituted by addition of 1 mM $ZnCl_2$, which fully restored activity. Metal in excess was removed by passing the enzyme through a 0.25-ml Chelex 100 column (Sigma, St. Louis, MO) preequilibrated with 50 mM Tris-HCl, pH 7.2. Protein-containing fractions (0.1 ml volume) were then pooled. Protein content was accurately determined on the basis of wild-type and mutated CPSso extinction coefficients at 280 nm, as reported (Mach et al., 1992). To determine metal content, protein was first extensively digested by incubation with pronase (0.1 mg/mg CPSso) for 4 h at 37°C, then spectrophotometric zinc determination was performed essentially as reported (Goulding and Matthews, 1997), using 4-(2-pyridylazo)-resorcinol. The blank was a buffer passed through Chelex 100 and treated with pronase as described above. Likewise, all other buffers employed after removal of excess zinc from CPSso preparations were previously passed through a Chelex 100 column.

Modeling of CPSso

Pairwise and multiple sequence alignment

Pairwise sequence alignment was computed using LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html). All the sequences of two zinc- and one zinc-containing carboxypeptidases were searched in the SwissProt database (Release 40.7 of December 19, 2001). The resulting sequences were aligned with CLUSTALW (Thompson et al., 1994). The residues involved in the binding of the two zinc and one zinc atoms were deduced from this alignment.

Structural alignments

The atomic solved structures of carboxypeptidases were searched in the PDB database and were superimposed with the BRAGI program (ftp.gbf.de/pub/Bragi) to obtain pairwise structural alignments.

Secondary structure prediction

The secondary structure of CPSso was predicted with a program based on neural networks (Jacoboni et al., 2000) and with a consensus-based procedure (Cuff and Barton, 1999); the secondary structure of the atomic resolved carboxypeptidases respectively from *Thermoactinomyces vulgaris* (Teplyakov et al., 1992) (CPTvu, 1obr) and from *Pseudomonas spirullum* (Rowsell et al., 1997) (CPG2, 1cg2) was defined with the DSSP program (Kabsch and Sander, 1983). When necessary, solvent exposure of residues was evaluated with the DSSP program.

Model building and evaluation

Model building by comparison was performed with the MODELLER program (Šali and Blundell, 1993). For a given alignment, 10 model structures were built and evaluated with the PROCHECK suite of programs (Laskowski et al., 1993). Only the best evaluated model was retained after the analysis.

We also computed possible assemblies of the modeled monomers. Complexes were built using BRAGI according to crystal symmetries of selected templates (as deduced from the MSD-PQS database, http://pqs.ebi.ac.uk) by means of structural superimposition with the templates. Close contacts were removed with the DEBUMP program of the WHATIF package (Vriend, 1990).

SAXS experiments

Experiments were performed using the SAXS beamline at the ELETTRA synchrotron (Trieste, Italy). The wavelength of x-rays was 1.54 Å and the sample-to-detector distance was 2.5 m. The scattering vector Q range was 0.01−0.15 $Å^{-1}$. CPSso samples at a concentration of 6 mg/ml in 50 mM Tris/Mes buffer, pH 6.0, 6% (by vol) glycerol and 0.75 M NaCl were measured using 1-mm glass capillaries. To avoid radiation damage, the exposure time was 300 s/frame. The experimental intensities were corrected for background, buffer contributions, detector inhomogeneities, and sample transmission. Some basic equations and the methods used for SAXS data analysis are described below. In the frame of the so-called two-phase model, diluted protein solutions (on the order of 10^{-5} M) can be considered as consisting of randomly oriented, homogeneous particles (of constant electron density ρ), dispersed in a solvent with homogeneous electron density ρ_s (Guinier and Fournet, 1955). If the system is monodisperse (i.e., the conformational and aggregation state of the protein is completely defined), the SAXS intensity can be approximated as

$$I(Q) = c(\Delta\rho)^2 V^2 P(Q) S_M(Q), \tag{1}$$

where Q is the exchanged wave vector, equal to $4\pi \sin\theta/\lambda$ (2θ is the full scattering angle), c is the molar protein concentration, $\Delta\rho = \rho - \rho_s$ is the contrast and V is the protein volume. $P(Q)$ is the averaged squared form factor, which depends on the protein structure, while $S_M(Q)$ is the so-called measured structure factor, which depends on the interference between waves scattered by different particles and which can be neglected for diluted protein solutions. The form factor $P(Q)$ is connected via an isotropic Fourier transform to the distance distribution function $p(r)$, the probability of finding pairs of small volume elements at a distance r within the protein,

$$\begin{aligned} P(Q) &= \int_0^\infty p(r)\frac{\sin(Qr)}{Qr}dr \Leftrightarrow p(r) \\ &= \frac{2r}{\pi}\int_0^\infty P(Q)Q\sin(Qr)dQ. \end{aligned} \tag{2}$$

At small Q, the scattering intensity can be approximated by the Guinier law (Guinier and Fournet, 1955)

$$I(Q) \cong I(0)\exp(-R_g^2 Q^2/3), \tag{3}$$

where R_g is the particle radius of gyration, related to the average dimension of the scattering particle and connected with the distance distribution function by

$$R_g^2 = \frac{1}{2}\int_0^\infty r^2 p(r)dr, \tag{4}$$

and $I(0)$, defined by

$$I(0) = c(\Delta\rho)^2 V^2, \tag{5}$$

is the scattering intensity at zero angle. For globular proteins, the Guinier approximation is valid only for $QR_g \leq 1.3$ (Feigin and Svergun, 1987). Considering Eq. 3, by linear fitting the data in the so-called Guinier plot (i.e., ln $I(Q)$ vs. Q^2), R_g, and $I(0)$ can be easily determined.

Particle shape reconstruction methods

Due to the limitation on the accessible Q range and the loss of information incurred from averaging the scattered intensity over all particle orientations, the derivation of the $p(r)$ function from the experimental SAXS curve is a rather complicated problem, the solution of which might not be unique. So, to reconstruct the shape of the scattering particle, different procedures should be adopted, usually based on the comparison of the form factor of refined model shapes to the experimentally observed scattering intensity. Two different methods have been used in this work, and both are valid under the assumptions of the two-phase model.

The Monte Carlo form factor for globular protein

When the structure of a protein is known, the distance distribution functions $p(r)$ can be calculated from the crystallographic coordinates using the Monte Carlo methods (Mariani et al., 2000; Hansen, 1990; Henderson, 1996; Ashton et al., 1997; Svergun, 1997). In particular, any arbitrary but compact protein shape can be described by the position function $s(\mathbf{r})$, which gives the probability that the point $r \equiv (r, \omega_r)$ (where ω_r indicates the polar angles α_r and β_r) lies within the particle, as

$$s(\mathbf{r}) = \begin{cases} 1 & r \leq \mathcal{F}(\omega_r) \\ \exp\{-[r - \mathcal{F}(\omega_r)]^2/2\sigma^2\} & r > \mathcal{F}(\omega_r) \end{cases}. \tag{6}$$

Here $F(\omega_r)$ is the shape function, which represents the distance between the particle center and the border in the direction ω_r. To account for the effect of the chain mobility on the protein surface or/and to the presence of a hydration shell, the particle border is described with a half-Gaussian profile, characterized by the variance σ (Svergun et al., 1998). The function $F(\omega_r)$ can be easily evaluated from the envelope surface of the van der Waals spheres centered in the protein atomic coordinates and the $p(r)$ histogram, which is then directly calculated taking into account the distances between all pairs of N_R points that have been randomly generated and evaluated according to Eq. 6. The Monte Carlo form factor is then calculated by the Eq. 2, and the data fitting procedures are described below.

Protein shape reconstruction from the multipole expansion method

The in-solution particle shape can be also reconstructed from SAXS data by the multipole expansion method. In the original method, the shape function $F(\omega_r)$ is expanded in series of spherical harmonics $Y_{l,m}(\omega_r)$ (Svergun and Stuhrmann, 1991). According to the recent introduction of the group theory to explore the fitting parameter space and of the maximum entropy principle to avoid truncation effects (Spinozzi et al., 1998), the expression of the shape function can be written as

$$\mathcal{F}^S(\omega_r) = \exp\left\{\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{l,m} Y^S_{l,m}(\omega_r)\right\}, \tag{7}$$

where the symmetrized spherical harmonics $Y^S_{l,m}(\omega_r)$ (calculated up to the maximum rank L) are compatible with the symmetry group G and $\{a_{l,m}\}$ is the set of parameters to be determined in order to fit the experimental scattering curve. Therefore, the set $\{a_{l,m}\}$ contains all information about protein size and shape. Accordingly, the squared form factor $P(Q)$ can be written as an even power series of Q

$$P(Q) = \sum_n b_n Q^{2n}, \tag{8}$$

where the b_n are complex functions of the variational parameters $a_{l,m}$, through the following set of expressions

$$b_n = \sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{p=0}^{P} \varphi^{(p)}_{l,m}\varphi^{(n-l-p)*}_{l,m}, \tag{9}$$

$$\varphi^{(p)}_{l,m} = \frac{4\pi d_{l,p}}{V}\left[\frac{f^{(3+1+2p)}_{l,m}}{3+l+2p} + \sum_{k=0}^{1+l+2p}\binom{2+l+2p}{k} 2^{(k-1)/2}\Gamma([k+1]/2)\sigma^{k+1} f^{(2+1+2p-k)}_{l,m} + \delta_{l,0}(4\pi)^{1/2}2^{(1+2p)/2}\Gamma([3+2p]/2)\sigma^{3+2p}\right], \tag{10}$$

$$f^{(q)}_{l,m} = \sum_{l_1=0}^{L}\sum_{l_2=|l-l_1|}^{l+l_l}\left[\frac{(2l_1+1)(2l_2+1)}{4\pi(2l+1)}\right]^{1/2} C(l_1,l_2,l;0,0) \times \sum_{m_1=-l_1}^{l_1} C(l_1,l_2,l;m_1,m-m_1) f_{l_1,m_1} f^{(q-1)}_{l_2,m-m_1}, \tag{11}$$

$$f_{l,m} = \int d\omega_r \mathcal{F}^S(\omega_r) Y^*_{l,m}(\omega_r), \tag{12}$$

$$d_{l,p} = \frac{(-1)^p}{2^p p![2(l+p)+1]!!}. \tag{13}$$

The quantities $d_{l,p}$ are the expansion coefficients of the spherical Bessel functions in power series, $j_l(x) = \sum_p = {}_0{}^P d_{l,p} x^{l+2p}$, with $P \cong 35-40$, depending on the maximum x value, $\delta_{l,m}$ is the Kronecher delta function, $\Gamma(x)$ is the Euler's gamma function and $C(l_1, l_2, l; m_1, m - m_1)$ are the Clebsh-Gordan coefficients. The integrals in Eq. 12, which only depend on the set $\{a_{l,m}\}$, are calculated with the Gauss-Legendre quadrature method, using 32 points for both variables α_r and β_r. Note that, as previously discussed (Spinozzi et al., 1998), the set of parameters $\{a_{l,m}\}$ for a globular protein can be efficiently determined by shrinking their range of existence at $|a_{l,m}/a_{0,0}| < t$, where the bound parameter t is near 0.1–0.5. It is important to stress that for each point group G, only a subset of parameters $a_{l,m}$ can be different from zero. However, the number of those parameters, M, increases on progression from higher to lower symmetries. In the fitting procedure (see below), the $\{a_{l,m}\}$ space is explored using a sequence of decreasing symmetries, from the most symmetrical group, the spherical one, K_h, defined

by only one parameter $a_{0,0}$, to the completely asymmetrical C_1 group, which is defined by $M = (L + 1)^2$ parameters (Spinozzi et al., 1998).

Data fitting

The merit functional to be minimized is defined as

$$\chi^2 = \frac{1}{N_S - M - 1}\left\{\sum_{i=1}^{N_S}\left[\frac{I(Q_i) - \kappa P(Q_i) - B}{\delta_i}\right]^2 + \left[\frac{V - V_0}{\delta_V}\right]^2\right\}, \quad (14)$$

where $I(Q_i)$ is the experimental SAXS curve measured at the point $Q_i \equiv i\pi/D_{max}$; $P(Q_i)$ is the fitting form factor calculated at the same point by Eqs. 2 or 8; κ is a scaling factor corresponding to the scattering intensity at $Q = 0$ (see Eq. 1); B is a flat background; V is the reconstructed particle volume; δ_i and δ_V are the experimental uncertainties of the scattering curve at the point Q_i and of the known volume V_0, respectively. It should be in fact observed that the protein volume V_0, which can be estimated from the amino acid sequence (Jacrot and Zaccai, 1981) taking into account the protein aggregation state, can be included in the numerical procedure as a further constraint. In any case, the number of free parameters cannot exceed the information content (More, 1980). In a SAXS curve, the information content is estimated by the Shannon's theorem as $N_S = D_{max}(Q_{max} - Q_{min})/\pi$, where D_{max} is the maximum particle diameter and $(Q_{max} - Q_{min})$ is the experimental window range. Therefore, the useful points correspond to the N_s Shannon channels. It is worth stressing that when an increase in the number of parameter M does not determine an effective improvement in the data fit, an increased χ^2 results from Eq. 14. Hence, the minimum χ^2 corresponds to the solution in which good agreement with the experimental data is obtained using the smallest number of parameters (i.e., considering the simpler and symmetrical particle shape).

RESULTS

Determination of bound zinc in wild-type CPSso

Our previous investigations clearly established that CPSso is a zinc-metalloprotease (Colombo et al., 1992; Villa et al., 1993). Thus, an obvious prerequisite for the development of a correct 3D model for this enzyme is the determination of the molar metal/protein ratio. By using a spectrophotometric method that exploits the spectral changes resulting from the formation of a complex between zinc and 4-(2-pyridylazo)-resorcinol (see Materials and Methods), we consistently determined a molar ratio close to one between bound zinc and wild-type enzyme subunit (Table 1).

TABLE 1 Zinc content and enzyme activity of CPSso and some mutants

Enzyme	Specific activity (units/mg)	Molar ratio (zinc/subunit)
Wild-type CPSso	12.0	0.93
$His^{104}Ala$	11.5	n.d.*
$His^{108}Ala$	b.d.†	0.04
$Asp^{109}Leu$	b.d.	0.03
$His^{111}Ala$	12.3	n.d.
$Glu^{142}Ser$	b.d.	0.02
$His^{168}Ala$	b.d.	0.03
$His^{245}Ala$	11.8	n.d.

*Not determined.
†Below detection limit.
Metal and protein content of purified CPSso variants were determined as reported in Materials and Methods. Each measurement is the mean of at least three independent determinations. Standard deviations never exceeded 10%.

Modeling of CPSso

It is well known that building by homology can only be successfully applied when protein sequence identity is higher than 40% (Sternberg,1996). Below this limit, in the so-called ''twilight'' zone, it is questionable whether building by homology can be applied and other procedures are necessary. Here we take advantage of some experimental findings that characterize the relationship between structure and function, posing constraints to our alignment methods and basically developing an expert-driven threading procedure.

A search in the database of protein structures indicates that carboxypeptidases can be clustered on structural basis into two basic sets: one-domain monomers, containing one catalytic zinc ion, such as the enzyme from *Thermoactinomyces vulgaris* (Teplyakov et al., 1992) (CPTvu, 1obr), with hexameric symmetry (http://pqs.ebi.ac.uk); two-domain dimers, one catalytic and the other for dimerization, with a dimeric/tetrameric symmetry (http://pqs.ebi.ac.uk), containing two catalytic zinc ions per monomer, such as carboxypeptidase G2 from *Pseudomonas spirullum* (Rowsell et al., 1997) (CPG2, 1cg2).

Sequence alignment of CPSso with the database of structures reveals that the chain from *Sulfolobus solfataricus* has a low sequence identity with both templates. Carboxypeptidase G2 (CPG2) shares 21% of identical residues and 32% of conserved residues with CPSso. It is present in the database as a dimer; however its crystallographic symmetry is also consistent with a tetrameric structure (http://pqs.ebi.ac.uk). Furthermore, the catalytic site of CPG2 contains two zinc ions, whereas we found that CPSso has only one bound metal. The other template, which contains only one catalytic zinc ion in the monomer has a sequence identity with CPSso of 16% and is consistent with an hexameric structure (http://pqs.ebi.ac.uk).

The question is then which template should be used, given the substantially low sequence identity of the target with both templates. Our procedure started with the structural alignment of the two selected templates, to highlight which residues in the two zinc ion-containing monomer are conserved in the one zinc-containing protein. This procedure indicated that the catalytic domain of CPG2 (comprising residues 1–192 and 304–393) and CPTvu are similar (the root mean square deviation of the aligned structures is 0.25 nm). The structural alignment (shown in Fig. 1) indicates that residues coordinating the Zn ion in CPTvu are aligned

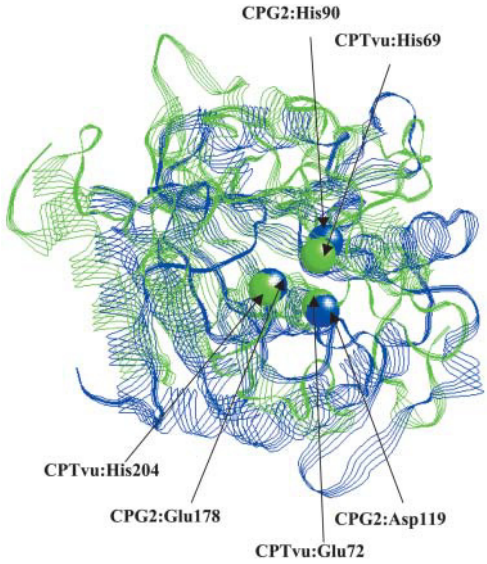

FIGURE 1 Structural superimposition of CPTvu with the catalytic domain of CPG2. CPTvu is in green and the catalytic domain of CPG2 (residues 1–192 and 304–393) is in blue. Green spheres highlight the Cα atoms of the residues coordinating to the zinc ion in CPTvu; blue spheres represent the Cα atoms of residues coordinating to the first zinc ion in CPG2.

with those coordinating the first Zn ion in CPG2. More specifically, His^{69}, Glu^{72}, and His^{204} of CPTvu correspond respectively to His^{90}, Asp^{119}, and Glu^{178} of CPG2. Furthermore, in both templates a water molecule is hydrogen bonded to a glutamate residue in position 154 and 277 in CPG2 and CPTvu, respectively. We used these data to further constrain manually the alignment of our target to either template, based on the assumption that if CPSso has one metal binding site per monomer, it is the one common to both templates.

First, we predicted the secondary structure of our target with a neural network-based method, implemented in house (Jacoboni et al., 2000) Then we aligned the predicted secondary structure of the target to that of either template while maintaining the constraint of conserving zinc ion-coordinating residues. The alignment was then used to run a program of building by homology. Based on this, we computed two models for CPSso that are shown in Fig. 2, *A* and *B*. In the two models the zinc ion is coordinated by different residues of the CPSso sequence: in Model A, the metal is coordinated by His^{108}, Asp^{109}, and His^{245}, and water is hydrogen-bonded to Glu^{327}; in Model B the metal is coordinated by His^{108}, Asp^{109}, and His^{168}, and water is bound to Glu^{142}. Furthermore, as expected from the building procedure, Model A resembles the CPTvu structure, whereas Model B is similar to the CPG2 structure. Experiments described below show that Model B, based on CPG2 but containing only one catalytic zinc, fits well with our experimental data, whereas Model A, based on the CPTvu structure, does not.

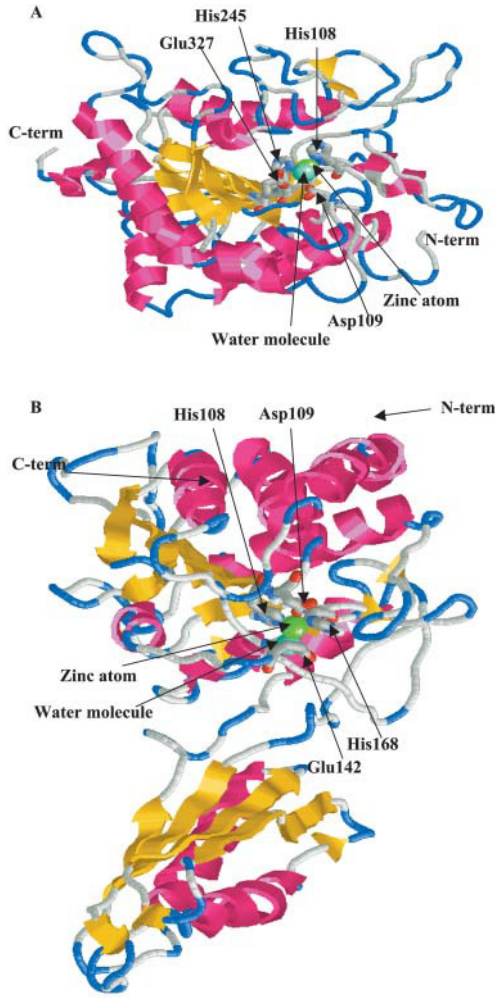

FIGURE 2 Models for CPSso. The model based on the structure of CPTvu is depicted in *A*, while the model based on the structure of CPG2 is depicted in *B*. α-helices, β-strands, and turns are in magenta, yellow, and blue respectively. The zinc ions and the water molecules are represented with green and cyan spheres, respectively. The residues coordinating to the zinc ion and the water molecules are represented by sticks.

Identification of residues putatively involved in zinc binding

According to the two structural models (Fig. 2, *A* and *B*), the metal is chelated either by the residues His^{245}, His^{108}, Asp^{109} (Model A), or by His^{168}, His^{108}, Asp^{109} (Model B).

However, according to our computations the models could also accommodate His^{104} or His^{111} in the metal binding site, due to their proximity to the active site. Therefore, all three histidine residues, i.e., His^{104}, His^{108}, and His^{111}, being close in the sequence, might be involved in metal chelation. So we produced mutants for each of three histidines, which showed that the replacement of His^{104}, as well as of His^{111} did not compromise enzyme function. This was instead the case when His^{108} was replaced, allowing us to identify it as a key residue in the catalytic device of both models. Also, the lack

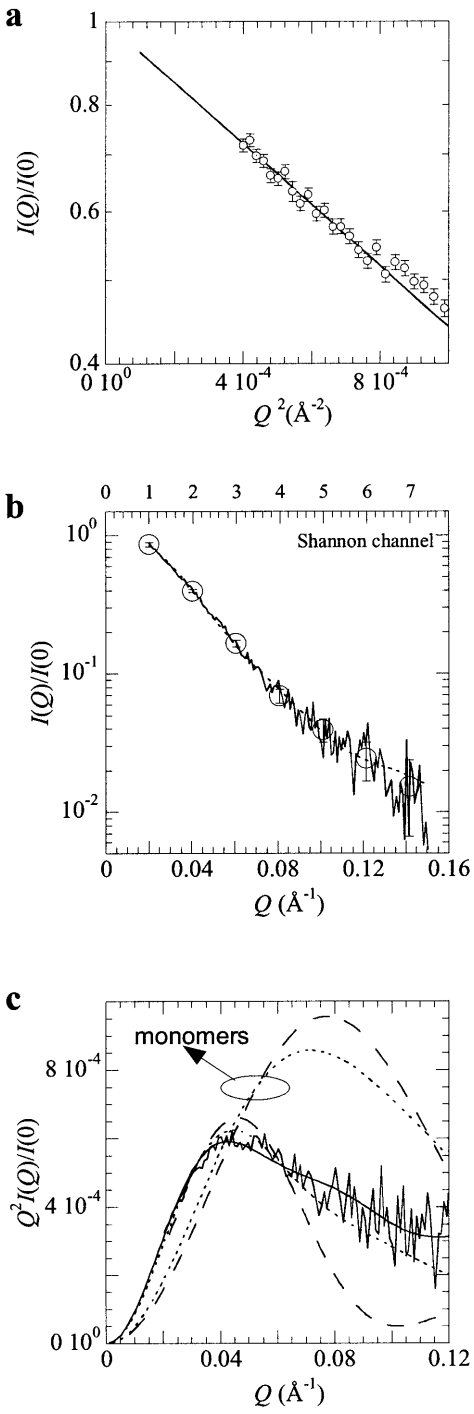

FIGURE 3 SAXS data. Experimental SAXS data have been normalized for the $I(0)$ value obtained by applying the Guinier law (see Eq. 3) and are shown in different forms. (*a*) Guinier plot (ln $I(Q)/I(0)$ vs. Q^2). The solid line corresponds to the best fit obtained by using the Guinier law (see Eq. 3). (*b*) semilogarithmic plot (ln $I(Q)/I(0)$ vs. Q). The dotted line corresponds to the fit obtained using the multipole expansion method. The corresponding best fitting parameters are reported in Table 2. The position of the Shannon channels ($Ns = 7$) is shown by open circles. (*c*) Kratky plot ($Q^2I(Q)/I(0)$ vs. Q). The best fit curves calculated by using the Monte Carlo method from the two computer-designed models are given. Model A (CPTvu-based) and B (CPG2-based), both in monomeric and in oligomeric form, are represented by the dashed and the dotted lines, respectively. Model B, tetrameric assembly, $\sigma = 1.5 \pm 0.6$ Å; Model A, hexameric assembly, $\sigma = 0.5 \pm 0.3$ Å. The solid line was obtained by the multipole expansion method.

of activity and bound metal in D109L (Table 1) was again consistent with both Models A and B. We found, however, that the replacement of His^{168}, but not of His^{245}, caused complete loss of activity and bound zinc (Table 1). This led us to discard the Model A and support Model B.

As to the fourth coordination site, it is well known that in metalloproteases it is occupied by a water molecule, which in turn interacts with an acidic residue via hydrogen bonding (Alberts et al., 1998). Based on Model B, we observed that the residue Glu^{142} could play this role, so we produced the mutant $Glu^{142}Ser$, which was actually nonfunctional. Thus, we suggest that the active site consists of a zinc ion chelated by the residues His^{108}, Asp^{109}, His^{168}, and a water molecule interacting with Glu^{142}.

SAXS data analysis

Experimental SAXS curve for CPSso is reported in Fig. 3, in the form of semilogarithmic, Kratky and Guinier plots. The presence of a well-defined peak in the Kratky plot clearly indicates that the protein has a globular form (Kataoka et al., 1993; Kataoka et al., 1995). Moreover, the linearity observed in the Guinier plot up to $Q^2 = 7 \times 10^{-4}$ Å^{-2} suggests sample monodispersity. Using the Guinier approximation (Eq. 3), a radius of gyration of 50 ± 2 Å has been then calculated, which is much larger than the one expected for CPSso in the monomeric form, i.e., $R_g^{sph} = 18.3$ Å. This has been calculated from the enzyme volume V_0, determined from the amino acid composition (Jacrot and Zaccai, 1981) and considering a spherical shape ($R_g^{sph} = (3/5)(3V_0/4\pi)^{2/3}$). This suggests the presence in solution of small oligomeric aggregates.

To assess the particle shape, the whole experimental curve was then analyzed by using the multipole expansion method described under Materials and Methods (see SAXS experiments). The maximum rank L was fixed to 6, the width σ of the hydration layer at the particle border was set to 2 Å and the bound parameter t to 0.3. An initial analysis was performed excluding the protein volume constraint in the χ^2 (Eq. 14). The best recovered shape (not shown here) had a volume of ~250,000 Å^3, a value 4–5 times larger than the monomer volume calculated from the amino acid sequence. Hence, multipole expansion analyses were arranged by fixing the nominal protein volume V_0 to 3, 4, 5, and 6 times the monomer volume, respectively, i.e., assuming that CPSso would form trimers, tetramers, pentamers, or hexamers. For all cases, the ratio δ_V/V_0 was fixed at 10%. As a result, the best fit was obtained by considering the presence of tetrameric aggregates: the fitting SAXS curve, the corresponding $p(r)$ and the reconstructed shape function $F^S(\omega_r)$ are shown in Figs. 3, 4, and 5, respectively, and the fitting parameters are reported in Table 2.

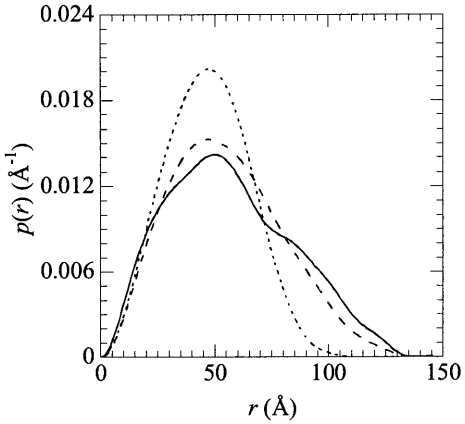

FIGURE 4 Distance distribution function. Comparison between the $p(r)$ functions obtained by fitting the SAXS experimental data with the Monte Carlo method from the two computer-designed models (*dashed line*, CPG2, tetrameric assembly, $\sigma = 1.5 \pm 0.6$ Å; *dotted line*, CPTvu, hexameric assembly, $\sigma = 0.5 \pm 0.3$ Å) and with the multipole expansion method (*solid line*).

Two points are of particular interest. First, the quality of the fit is very good, particularly if the small number of fitting parameters, $M = 4$, are taken into account. Actually, they are much lower than the Shannon's channel, $N_S = 7$. Second, the point group symmetry that better fits the experimental curve, D_{2h}, is compatible with the tetrameric structure. This feature is clearly evident on inspecting the four-lobated shape, which can be considered to be the most complete translation of the information collected in the SAXS curve.

Because of the intrinsically low symmetry of the present SAXS experiments, the determination of the monomer orientation inside the reconstructed shape has no unique solution, even though the derived point group symmetry for the scattering particle reduces the possible tetrameric states. Therefore, the tetrameric structure obtained by molecular modeling (as described above) was directly compared with the particle shape derived by the SAXS multipole approach. In Fig. 5, the computed tetrameric structure, assembled using the CPG2 structure as a template has then been superimposed on the particle shape obtained by SAXS multipole analysis. Using a "best-fill" approach (Cinelli et al. 2001), the six positional and orientational coordinates of the computed structure have been adjusted to minimize the quantity

$$rms = \left[V - \sum_{i=1}^{N_A} s(\mathbf{r}_i) V_i^{(vdW)} \right]^2, \qquad (15)$$

where N_A is the number of atoms, $s(\mathbf{r})$ the position function (Eq. 6), $V_i^{(vdW)}$ the van der Waals volume of the i-th atom, and $\mathbf{r}_i$ its position. Here $s(\mathbf{r}_i)$ plays the role of a filter function, which discards the i-th atom if it does not fall inside the shape. The agreement between the two structures is very good, in particular considering that they have been obtained using two completely different approaches, i.e., the modeling prediction and an ab initio analysis of SAXS experimental data.

As a further and conclusive support, the SAXS experimental curve has been fitted with the scattering intensity calculated from the crystallographic coordinates of the two computer-designed models. Since the experimental radius of gyration, as well as the particle volume obtained from the multipole expansion method, indicated the presence in solution of small oligomeric aggregates, both modeled monomers and their possible assemblies were assessed. The Monte Carlo approach previously described was used and in the fitting procedure only three parameters were considered: the width σ of the hydration layer at the particle border, and the two scaling factors, κ and B. The calculated scattering intensities are shown in Fig. 3 (*lower frame*): it can be clearly observed that a good agreement with the experi-

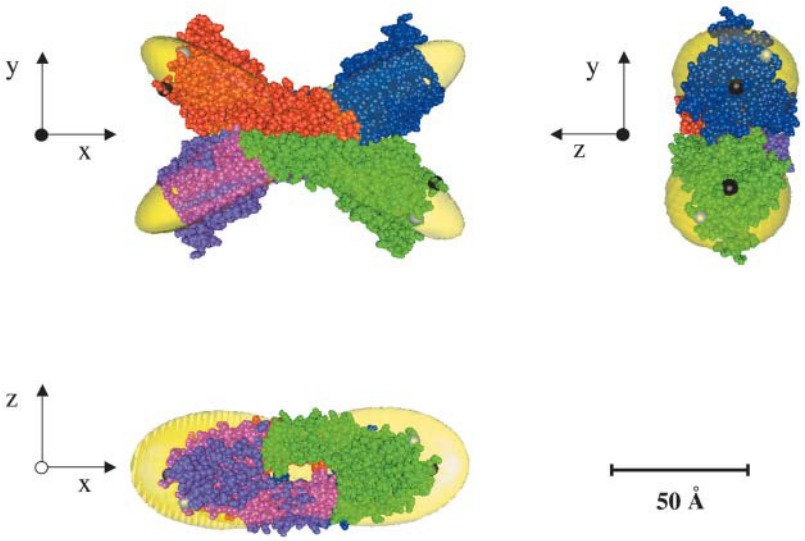

FIGURE 5 Reconstructed CPSso shape function. Scaled representation of the shape of the scattering particle, as obtained by multiple expansion analysis of the SAXS experimental data. Three different orientations are reported. To illustrate the agreement with the computer-designed model, the CPG2 tetrameric structure is superimposed to the reconstructed shape. The black and the white spheres represent the C- and the N-terminal ends, respectively, of the four polypeptide chains.

TABLE 2 SAXS data fitting parameters

Method	χ^2	κ (a.u.)	B (a.u.)	R_g (Å)	V_P (10^5 Å^3)	Fit parameters	
						Group symmetry	D_{2h}
						M	4
						N_S	7
Multipole expansion	1.23	2.01 ± 0.03	0	44.5 ± 0.1	2.02 ± 0.02	D_{max}	155 Å
						$a_{0,0}$	9.73 ± 0.03
						$a_{2,0}$	−2.95 ± 0.05
						$^R a_{2,2}$	2.24 ± 0.02
						$a_{4,0}$	−2.96 ± 0.03
Monte Carlo	2.62	1.82 ± 0.03	0	41.0 ± 0.6	2.0 ± 0.3	σ	1.5 ± 0.6 Å

Values of the parameters involved in the different procedures used to analyze the SAXS experimental data. Symbols and definitions as in the text.

mental data is obtained only with the tetrameric model based on the CPG2 structure (Model B).

DISCUSSION

In recent years, we subjected *S. solfataricus* carboxypeptidase (CPSso) to extensive investigations (Colombo et al., 1992; Villa et al., 1993; Colombo et al., 1995; Bec et al., 1996; Mombelli et al., 1996). This led us to discover its unusual features, notably a broad substrate specificity and a remarkable stability in organic solvents. We also found that, thanks to this latter property, it also catalyzes the synthesis of N-blocked amino acids in these media (manuscript in preparation). Of course, to understand the structural determinants of stability and specificity more completely, in addition to designing protein engineering experiments aimed at improving its performance as a biocatalyst, the 3D structure of the molecule is necessary. Unfortunately, we could not solve its structure by x-ray crystallography as crystallization trials have failed so far. Thus, the objective of this paper was to develop the 3D structure by a molecular modeling approach.

We first determined the metal content of CPSso. This is an indispensable prerequisite for the development of a reliable model. Our measurements consistently produced a molar ratio metal/subunit close to one. Taking this constraint into account, two models were developed, one based on the structure of CPTvu (Model A), the other on CPG2 (Model B), although sequence alignment revealed that the *Sulfolobus* enzyme has low sequence identity with both templates. On the other hand, the *Pseudomonas* enzyme belongs to the same protease clan (MH) as CPSso, but to a different family (M20) (Barrett et al., 1998) and is the protein of known structure whose sequence is more closely related to CPSso. However, CPG2 has two bound zinc ions per subunit (Rowsell et al., 1997), unlike CPTvu, which has only one zinc (Teplyakov et al., 1992). Thus, the CPG2-based model was developed with only one metal-binding site.

Site-directed mutagenesis allowed us to definitely discard Model A. Regarding Model B, this envisaged a zinc-binding site consisting of His^{108}, Asp^{109}, and His^{168}. The replacement of each of these three residues by noncatalytic ones abolished activity and bound zinc. In particular, the validation of the histidine residue involved in the active site was obtained by detecting a loss of activity and bound metal in mutant H108A, and not in mutants H104A and H111A (Table 1). This set of mutations was required since our computations were consistent with either His^{104} or His^{108} being involved in the active site, due to their proximity in the residue sequence. Thus, site-directed mutagenesis allowed us to confirm, and also refine Model B to some extent.

Based on further mutagenesis experiments we also identified Glu^{142} as the acidic residue interacting with the water molecule occupying the fourth chelation site (Alberts et al., 1998).

The active site defined by our model involves the occurrence of two consecutive residues chelating the metal, i.e., His^{108} and Asp^{109}. When the present model was developed, this was a novel and unexpected feature for zinc-binding sites of metalloenzymes. However, a similar pattern was recently found in *S. solfataricus* alcohol dehydrogenase, whose structure was solved by x-ray crystallography (Esposito et al., 2002). It was thus shown that the active site of this zinc-metalloenzyme consists of four residues bound to the metal, i.e., Cys^{38}, His^{68}, Glu^{69}, and Cys^{154}, with a fairly regular tetrahedral geometry. This shows that no steric hindrance prevents two consecutive residues, specifically a histidine followed by an acidic residue, from coordinating to the same metal ion.

Two recent articles report the characterization of *Pyrococcus horikoshi* OT3 carboxypeptidase/aminoacylase (Ishikawa et al., 2001) and *Pyrococcus furiosus* aminoacylase (Story et al., 2001). These two enzymes are closely related to each other, sharing as much as 45% of identical residues with CPSso. Furthermore, they also exhibit a molar ratio zinc/subunit close to one, which further supports the assignment of the same metal content to CPSso. On the whole, our results therefore substantiate the idea that, despite the distant evolutionary relationship between CPG2 and CPSso, the two enzymes have a very similar overall fold, but developed different metal-binding sites and that the structure we developed for the active site may be shared also by other aminoacylase/carboxipeptidases from other thermophiles.

Although site-directed mutagenesis allowed us to select Model B by determining the residues involved in the metal-binding site, this, obviously, could not validate the overall fold of the protein, or its quaternary structure. They were obtained by SAXS experiments that clearly showed a four-lobate conformation of CPSso. This nicely fits with the tetrameric structure of the CPG2-based model. Actually, gel-filtration data lead us to assign an apparent molecular mass of 170 kDa, in good agreement with a tetrameric structure (Colombo et al., 1992). This, however, could not be regarded as a conclusive result in support to a defined aggregation state for *Sulfolobus* enzyme. Furthermore, molecular modeling indicates symmetries compatible with both a dimeric (CPG2 is a dimer in the PDB database) and also a tetrameric aggregation state. So, by combining mutagenesis and SAXS experiments we provided independent and complementary support for the CPG2-based model that can also have a tetrameric symmetry.

The availability of a 3D model for CPSso makes it feasible various experimental prospects. For instance, in principle, it allows to mutagenize the enzyme for better understanding the structural features of its unique specificity. It is worth mentioning that the aforementioned *Pyrococcus horikoshi* OT3 carboxypeptidase/aminoacylase (Ishikawa et al., 2001) has a clearly different specificity with respect to the *Sulfolobus* enzyme, despite their close evolutionary relationship. In particular, *Pyrococcus* enzyme does not remove basic residues from N-blocked amino acids, which are very good substrates for CPSso. Furthermore, it freely cleaves dipeptides but it barely attacks tri- and tetrapeptides, whereas CPSso freely hydrolyzes peptides of up to five residues (data not shown). Although the availability of the 3D model alone should make it possible to identify the structural determinants involved in substrate recognition, comparison of the sequences of the two enzymes should also greatly facilitate this task. Likewise, by taking advantage of a similar approach, the heat resistance of CPSso might also be increased, as the *Pyrococcus* enzyme is significantly more thermostable than its counterpart from *Sulfolobus*. Thus, the model developed in this paper should enable us to produce new mutated forms of CPSso, with new catalytic properties and improved stability combined as desired. At the same time, protein engineering of CPSso should also allow us to further refine the model.

Finally, the methodological significance of this paper should also be stressed. In the absence of the crystallographic structure, the combination of computational prediction techniques, site-directed mutagenesis and in-solution SAXS, should allow us to determine the major properties of an enzyme protein in solution with greater confidence, for example the geometry of the catalytic site, the overall fold, as well as the quaternary structure. This is relevant to the growing impact of proteomics, with the expected high throughput sample production, which is likely to result in a growing yield of noncrystallizable proteins.

This work was partially supported by grants from the Italian Ministry for University and Research (in particular, PRIN 2001: ''Hydrolases from Thermophilic Microorganisms: Structural and Functional Aspects, Homologous and Heterologous Expression''), by a grant for a target project in Biotechnology and a project on Molecular Genetics, both of the Italian Centro Nazionale delle Ricerche (CNR), delivered to R.C. R.C. also acknowledges the EC grant Biowulf IST 1999-20232 for supporting the development of DNCBLAST, a parallelized version of PSI-BLAST for PC nets. P.L.M. is the recipient of a fellowship from the Italian Center for National Researches (CNR) devoted to a target project of Molecular Genetics (Law No. 449-1997).

REFERENCES

Alberts, I. L., K. Nadassy, and S. J. Wodak. 1998. Analysis of zinc binding sites in protein crystal structures. *Protein Sci.* 7:1700–1716.

Ashton, A. W., M. K. Boehm, J. R. Gallimore, M. B. Pepys, and S. J. Perkins. 1997. Pentameric and decameric structures in solution of serum amyloid P component by x-ray and neutron scattering and molecular modelling analyses. *J. Mol. Biol.* 272:408–422.

Barrett, A. J., N. D. Rawlings, and F. Woessner. 1998. Clan MH containing varied co-catalytic metallopeptidases. *In* Handbook of Proteolytic Enzymes, A. J. Barrett, N. D. Rawlings, and F. Woessner, editors. Academic Press Ltd. 1412–1416.

Bec, N., A. Villa, P. Tortora, V. V. Mozhaev, C. Balny, and R. Lange. 1996. Enhanced stability of carboxypeptidase from *Sulfolobus solfataricus* at high pressure. *Biotechnol. Lett.* 18:483–488.

Casadio, R., E. Polverini, P. Mariani, F. Spinozzi, F. Carsughi, A. Fontana, P. Polverino de Laureto, G. Matteucci, and C. M. Bergamini. 1999. The structural basis for the regulation of tissue transglutaminase by calcium ion. *Eur. J. Biochem.* 262:672–679.

Cinelli, S., F. Spinozzi, R. Itri, S. Finet, F. Carsughi, G. Onori, and P. Mariani. 2001. Structural characterization of the pH-denatured states of ferricytochrome-c by synchrotron small angle X-ray scattering. *Biophys. J.* 81:3522–3533.

Colombo, S., S. D'Auria, P. Fusi, L. Zecca, C. A. Raia, and P. Tortora. 1992. Purification and characterization of a thermostable carboxypeptidase from the extreme thermophilic archaebacterium *Sulfolobus solfataricus*. *Eur. J. Biochem.* 206:349–357.

Colombo, S., G. Toietta, L. Zecca, M. Vanoni, and P. Tortora. 1995. Molecular cloning, nucleotide sequence and expression of a carboxypeptidase-encoding gene from the archaebacterium *Sulfolobus solfataricus*. *J. Bacteriol.* 177:5561–5566.

Cuff, J. A., and G. J. Barton. 1999. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins.* 34:508–519.

Esposito, L., F. Sica, C. A. Raia, A. Giordano, M. Rossi, L. Mozzarella, and A. Zagari. 2002. Crystal structure of the alcohol dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* 318:463–477.

Feigin, L. A., and D. I. Svergun. 1987. Structure Analysis by Small-Angle X-ray and Neutron Scattering. Plenum Press, New York.

Goulding, C. W., and R. Matthews. 1997. Cobalamin-dependent methionine synthase from *Escherichia coli*: involvement of zinc in homocysteine activation. *Biochemistry.* 36:15749–15757.

Guinier, A., and G. Fournet. 1955. Small Angle Scattering Of X-Ray. Wiley, New York.

Hansen, S. 1990. Calculation of small-angle scattering profiles using Monte Carlo simulation. *J. Appl. Crystallogr.* 23:344–346.

Henderson, S. J. 1996. Monte Carlo modeling of small-angle scattering data from non-interacting homogeneous and heterogeneous particles in solution. *Biophys. J.* 70:1618–1627.

Ishikawa, K., H. Ispida, I. Matsui, Y. Kawarabayasi, and H. Kikuchi. 2001. Novel bifunctional hyperthermostable carboxypeptidase/aminoacylase from *Pyrococcus horikoshii* OT3. *Appl. Environ. Microbiol.* 67:673–679.

Jacoboni, I., P. L. Martelli, P. Fariselli, M. Compiani, and R. Casadio. 2000. Predictions of protein segments with the same aminoacid sequence and different secondary structure: a benchmark for predictive methods. *Proteins*. 41:535–544.

Jacrot, B., and G. Zaccai. 1981. Determination of molecular weight by neutron scattering. *Biopolymers*. 20:2413–2426.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22:2577–2637.

Kataoka, M., Y. Hagihara, K. Mihara, and Y. Goto. 1993. Molten globule of cytochrome c studied by the small angle X-ray scattering. *J. Mol. Biol.* 229:591–596.

Kataoka, M., I. Nishii, T. Fujisawa, T. Ueki, F. Tokunaga, and Y. Goto. 1995. Structural characterization of molten globule and native states of apomyoglobin by solution X-ray scattering. *J. Mol. Biol*. 249:215–228.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227:680–685.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.

Mach, H., C. R. Middaugh, and R. V. Lewis. 1992. Statistical determination in the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. *Anal. Biochem.* 200:74–80.

Makinen, M. W., G. B. Wells, and S. O. Kang. 1984. Structure and mechanism of carboxypeptidase A. *Adv. Inorg. Biochem.* 6:1–69.

Mariani, P., F. Carsughi, F. Spinozzi, S. Romanzetti, G. Meier, R. Casadio, and C. M. Bergamini. 2000. Ligand-induced conformational changes in tissue transglutaminase: Monte Carlo analysis of small-angle scattering data. *Biophys. J*. 78:3240–3251.

Mombelli, E., N. Bec, P. Tortora, C. Balny, and R. Lange. 1996. Pressure and temperature control of a thermophilic carboxypeptidase from *Sulfolobus solfataricus*. *Food Biotechnol.* 10:131–142.

More, P. B. 1980. Small-angle scattering content and error analysis. *J. Appl. Crystallogr.* 13:168–175.

Perez, J., P. Vachette, D. Russo, M. Desmadril, and D. Durand. 2001. Heat-induced unfolding of neocarzinostatin, a small all-β protein investigated by small-angle x-ray scattering. *J. Mol. Biol.* 308:721–743.

Pollack, L., M. W. Tate, N. C. Darnton, J. B. Knight, S. M. Gruner, W. A. Eaton, and R. H. Austin. 1999. Compactness of the denatured state of a fast-folding protein measured by submillisecond small angle x-ray scattering. *Proc. Natl. Acad. Sci. USA*. 96:10115–10117.

Rowsell, S., R. A. Pauptit, A. D. Tucker, R. G. Melton, D. M. Blow, and P. Brick. 1997. Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure*. 5:337–347.

Šali, A., and T. L. Blundell. 1993. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.

Skidgel, R. A., and E. G. Erdös. 1998. Cellular carboxypeptidases. *Immunol. Rev.* 161:129–141.

Spinozzi, F., F. Carsughi, and P. Mariani. 1998. Particle shape reconstruction by small-angle scattering: integration of group theory and maximum entropy to multiple expansion method. *J. Chem. Phys.* 109:10148–10158.

Sternberg, M. J. E. 1996. Protein structure prediction—principles and approaches. *In* Protein Structure Prediction—A Practical Approach. M. J. E. Sternberg, editor. Oxford University Press, Oxford. 1–30.

Story, S. V., A. M. Grunden, and M. W. W. Adams. 2001. Characterization of an aminoacylase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* 183:4259–4268.

Svergun, D. I. 1997. Restoring three-dimensional structure of biopolymers from solution scattering. *J. Appl. Crystallogr.* 30:792–797.

Svergun, D. I., and H. B. Stuhrmann. 1991. New developments in direct shape determination from small-angle scattering: I. Theory and model calculations. *Acta Crystallogr.* A47:736–744.

Svergun, D. I., S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA*. 95:2267–2272.

Teplyakov, A., K. Polyakov, G. Obmolova, B. Strokopytov, I. Kuranova, A. Osterman, N. Grishin, S. Smulevitch, O. Zagnitko, and O. Galperina. 1992. Crystal structure of carboxypeptidase T from *Thermoactinomyces vulgaris*. *Eur. J. Biochem.* 208:281–288.

Tewary, Y. B. 1990. Thermodynamics of industrially-important, enzyme-catalyzed reactions. *Appl. Biochem. Biotechnol.* 18:706–709.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.

Tortora, P., and M. Vanoni. 1998. Carboxypeptidase from *Sulfolobus solfataricus*. *In* Handbook of Proteolytic Enzymes, A. J. Barrett, N. D. Rawlings, and F. Woessner, editors. Academic Press Ltd. 1441–1443.

Trewhella, J. 1997. Insights into biomolecular function from small-angle scattering. *Curr. Opin. Struct. Biol.* 7:702–708.

Villa, A., L. Zecca, P. Fusi, S. Colombo, G. Tedeschi, and P. Tortora. 1993. Structural features responsible for kinetic thermal stability of a carboxypeptidase from the archaebacterium *Sulfolobus solfataricus*. *Biochem. J.* 295:827–831.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8:52–56.

Zuhlke, H., D. F. Steiner, A. Lernmark, and C. Lipsey. 1976. Carboxypeptidase B-like and trypsin-like activities in isolated rat pancreatic islets. Ciba Foundation Symposium. 41:183–195.