Crystallographic and Biochemical Investigations of Kumamolisin-As, a Serine-Carboxyl Peptidase with Collagenase Activity*

Alexander Wlodawer‡§, Mi Li‡¶, Alla Gustchina‡, Naoki Tsuruoka‖, Masako Ashida**,
Hiroyuki Minakata‡‡, Hiroshi Oyama§§, Kohei Oda§§, Tokuzo Nishino**, and Toru Nakayama**

*From the ‡Protein Structure Section, Macromolecular Crystallography Laboratory, NCI-Frederick, National Institutes of Health, Frederick, Maryland 21702, ¶Basic Research Program, Science Applications International Corp.-Frederick, NCI-Frederick, National Institute of Health, Frederick, Maryland 21702, the ‖Department of Biochemistry, School of Medicine, Kanazawa Medical University, Uchinada 1-1, Ishikawa 920-0293, Japan, the **Department of Biomolecular Engineering, Graduate School of Engineering, Tohoku University, 07 Aoba-yama, Sendai 980-8579, Japan, ‡‡Suntory Institute for Bioorganic Research, Mishima-gun, Osaka 618-8503, Japan, and the §§Department of Applied Biology, Faculty of Textile Science, Kyoto Institute of Technology, Sakyo-ku, Kyoto 606-8585, Japan*

Kumamolisin-As (previously called ScpA) is the first known example of a collagenase from the sedolisin family (MEROPS S53). This enzyme is active at low pH and in elevated temperatures. In this study that used x-ray crystallographic and biochemical methods, we investigated the structural basis of the preference of this enzyme for collagen and the importance of a glutamate residue in the unique catalytic triad (Ser^{278}-Glu^{78}-Asp^{82}) for enzymatic activity. Crystal structures of the uninhibited enzyme and its complex with a covalently bound inhibitor, *N*-acetyl-isoleucyl-prolyl-phenylalaninal, showed the occurrence of a narrow S2 pocket and a groove that encompasses the active site and is rich in negative charges. Limited endoproteolysis studies of bovine type-I collagen as well as kinetic studies using peptide libraries randomized at P1 and P1′, showed very strong preference for arginine at the P1 position, which correlated very well with the presence of a negatively charged residue in the S1 pocket of the enzyme. All of these features, together with those predicted through comparisons with fiddler crab collagenase, a serine peptidase, rationalize the enzyme's preference for collagen. A comparison of the Arrhenius plots of the activities of kumamolisin-As with either collagen or peptides as substrates suggests that collagen should be relaxed before proteolysis can occur. The E78H mutant, in which the catalytic triad was engineered to resemble that of subtilisin, showed only 0.01% activity of the wild-type enzyme, and its structure revealed that Ser^{278}, His^{78}, and Asp^{82} do not interact with each other; thus, the canonical catalytic triad is disrupted.

A novel peptidase, initially named ScpA (1) and now called kumamolisin-As (2), was recently identified by us in the culture filtrate of a thermoacidophilic soil bacterium *Alicyclobacillus sendaiensis* strain NTAP-1 (1). Specificity analyses using macromolecular substrates including globular and other fibrillar proteins showed that kumamolisin-As is highly specific for collagen (3, 4) and thus could be considered as a collagenase, although with some unusual properties. Most noticeably, this enzyme exhibits the maximum activity at acidic pH ~4.0. This is in striking contrast to all known collagenases, which are either zinc-dependent metallopeptidases (5) or chymotrypsin-like serine proteinases (6, 7), with an optimum pH for activity at neutral to alkaline regions.

A primary structure analysis of this novel "acid collagenase" revealed that it is a member of the sedolisin family, a recently established class of serine peptidases with a unique catalytic triad, Ser-Glu-Asp, in place of the Ser-His-Asp triad of classical serine peptidases (2, 4). Moreover, the enzyme was found to be very similar in its primary structure to kumamolisin, a well characterized member of the family (8–10), exhibiting 92.7% identity with its mature form. This high level of identity led to the change of the name from the initially used ScpA (1) to kumamolisin-As (2). Kumamolisin-As was the first member of the sedolisin family to be shown capable of degrading collagen, but further analysis of the substrate preferences of kumamolisin detected some collagenolytic properties, although not as pronounced (4).

Collagen is an insoluble structural protein that accounts for ~30% of the total weight of animal proteins (5). It is the predominant constituent of skin, tendons, and cartilage as well as the organic component of bones, teeth, and the cornea. Collagen is also found in the connective tissues of nearly all organs as insoluble fibers embedded in the extracellular matrix, where it serves to provide both their structure and strength (5).

Enzymatic degradation of collagen has attracted medical attention because it is closely related to the etiology of many human diseases (5). Considering the abundance of collagen in nature, microbial degradation of collagen should also be of biogeochemical significance in global cycling of nitrogen (11). Due to its rigidified fibrillar structure, collagen is not generally degraded by ordinary peptidases but can only efficiently be degraded by the collagen-specific enzymes named collagenases (5).

Kumamolisin-As was the first example of a collagenase from the sedolisin family, and analyses of its subsite specificity, its mode of collagen binding, and the role of its unique catalytic triad could be very interesting issues to be clarified in compar-

ison with the classical types of collagenases. We report here the results of such a study, conducted using crystallographic and biochemical approaches. In addition, we created and characterized the E78H mutant of the enzyme, in which the glutamate residue of its catalytic triad was replaced by a histidine in order to mimic the catalytic triad of the classical serine peptidases.

EXPERIMENTAL PROCEDURES
Materials

Collagen (type I, from bovine Achilles tendon) and high performance liquid chromatography (HPLC)[1] grade acetonitrile were purchased from Nacalai Tesque (Kyoto, Japan). All chemicals for peptide synthesis were obtained from PerkinElmer Life Science. An internally quenched fluorogenic substrate (IQF), NMA-MGPH*FFPK(DNP)DRDR ([2-(*N*-methylamino)benzoyl]-L-methionyl-glycyl-L-prolyl-L-histidyl-L-phenylalanyl-L-phenylalanyl-L-prolyl-*N*ε-(2,4-dinitrophenyl)-L-lysyl-D-arginyl-D-arginine amide) was a product of the Peptide Institute (Osaka, Japan). An inhibitor, *N*-acetyl-isoleucyl-prolyl-phenylalaninal (AcIPF), was synthesized as described previously (9, 12). Restriction enzymes and other DNA-modifying enzymes were purchased from TaKaRa Shuzo (Kyoto, Japan) or from Toyobo (Osaka, Japan). The plasmid pScpA, which is a derivative of pET15b (Novagen, Madison, WI), was constructed as described previously (4) and was used for the expression of the full-length kumamolisin-As gene. All other chemicals used were of analytical or sequencing grade, as appropriate.

Enzyme Assays

Two methods of assessing collagenase activity were employed in this study.

Method I—This assay system contained 100 μM sodium acetate, pH 4.0, 2 mg of collagen, and enzyme in a final volume of 0.5 ml. After preincubation at 60 °C, the reaction was started by the addition of the enzyme. Incubation was carried out at 60 °C with shaking (at 1000 rpm) for 10–60 min (depending on the amount of enzyme to be assayed) using a micromixer model E-36 (TAITEC Co., Saitama, Japan) that maintained the homogeneous distribution of the collagen powder in the reaction mixture during incubation. The blank reference mixture did not contain the enzyme. The reaction was stopped by adding 1.0 μl of 0.2 M HCl and chilling the mixture on ice for 15 min followed by centrifugation. Supernatant (100 μl) was mixed with 400 μl of Ninhydrin Color Reagent Solution (Nacalai Tesque) and was heated at 97 °C for 10 min, followed by chilling the mixture on ice. 2-Propanol (1.0 ml) was then added to the mixture, and the increase in absorbance of the supernatant at 570 nm (ΔA_{570}) was determined. The ΔA_{570} of the reaction mixture, in which collagen (2 mg) was completely degraded by the addition of an excess amount of the collagenase under these assay conditions, was also determined and used for unit calculations.

Method II—For assaying the enzymatic hydrolysis of the IQF substrate, NMA-MGPH*FFPK(DNP)DRDR (where an asterisk indicates the scissile site), the standard assay mixture contained varying amounts of the substrate, 50 mM sodium acetate buffer (pH 4.0), and the enzyme in a final volume of 300 μl. The stock enzyme solution used for Method II contained 0.1% (w/v) Tween 80. The assay mixture without the enzyme was brought to 40 °C, and the reaction was started by the addition of the enzyme. After incubation for 10 min, the reaction was stopped by the addition of 300 μl of 1 M Tris-HCl, pH 9.0, followed by chilling the mixture on ice. Fluorescence intensity changes of the reaction mixture (excitation, 340 nm; emission, 440 nm) were determined with a Shimadzu fluorescence spectrophotometer RF-5000. The fluorescence intensity change where the substrate was completely degraded by the addition of an excess amount of the collagenase under these assay conditions was also determined and was used for unit calculations. Protein was determined by the method of Bradford (13) using a kit (Bio-Rad) with bovine serum albumin as the standard. Kinetic parameters and their S.E. values were determined by nonlinear regression analysis (14) using the initial velocity data obtained by means of assay method II.

Construction of Peptide Libraries

To analyze the P1 and P1′ specificities of kumamolisin-As, peptide libraries were designed on the basis of the amino acid sequence of a nonapeptide, Met-Gly-Pro-Arg*Gly-Phe-Pro-Gly-Ser, where an asterisk indicates the scissile site. A peptide of such a sequence was previously identified as a good substrate for kumamolisin-As (4). The P1′ library was an equimolar mixture of peptide components, Met-Gly-Pro-Arg*Xaa-Phe-Pro-Gly-Ser, with Xaa being all of the different amino acids other than Cys. The individual relative retention times of the peptide components and their cleavage products were predicted from their amino acid sequences by Rekker's method as described by Sasagawa and Teller (15). Five different sets of peptides (termed P1′a, P1′b, P1′c, P1′d, and P1′e; see below), each consisting of three or four peptides, were synthesized for the P1′ library using an ABI 433A automated peptide synthesizer; the combination of amino acids (Xaa) at P1′ of peptides for each set was determined on the basis of the predicted retention times so that all of the substrate peptide and the cleavage products could be separated from one another by a single HPLC assay (see below). Thus, amino acid residues at the P1′ position of each set were as follows: Trp, His, Gly, and Arg for P1′a; Phe, Leu, Pro, and Ala for P1′b; Tyr, Thr, and Glu for P1′c; Ile, Val, Gln, and Asp for P1′d; and Met, Ser, Lys, and Asn for P1′e.

For the evaluation of the P1 specificity of the enzyme, the P1 library, Met-Gly-Pro-Xaa*Phe-Phe-Pro-Gly-Ser, was also constructed in a manner similar to the one described above. Three different sets of peptides were synthesized for the P1 library and were termed P1a (Xaa; His, Asn, Tyr, and Trp), P1b (Xaa; Arg, Glu, Ala, Asp, Val, Leu, and Phe), and P1c (Xaa; Lys, Thr, Gln, Gly, Ser, Met, Pro, and Ile).

Primary structures of the peptide libraries were verified by automated Edman degradation. The individual substrates and cleavage products separated on HPLC were identified by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry using a Bruker REFLEX III spectrometer.

Specificity Assays at the P1 and P1′ Sites by HPLC

The reaction mixture consisted of 100 mM sodium acetate, pH 4.0, a set of synthetic peptides (typically 250 μM for each component), and 0.185 nM of enzyme in a final volume of 200 μl. The mixture without the enzyme was preincubated at 60 °C, and the reaction was started by the addition of the enzyme. At intervals, a 200-μl aliquot of solution was withdrawn and mixed with 50 μl of 1 M potassium phosphate, pH 7.0; the resulting mixture was chilled on ice. The peptides were analyzed by a reversed-phase HPLC using an automated Gilson 305 system equipped with a Shimadzu SPD-10A *VP* UV-visible detector: column, YMC-Pack ODS-A A-303 (4.6 × 250 mm; YMC Co., Kyoto, Japan); flow rate, 0.7 ml/min; solvent A, 0.1% (v/v) trifluoroacetic acid; solvent B, 0.1% (v/v) trifluoroacetic acid in 60% (v/v) acetonitrile. After injection (50 μl) onto a column that was equilibrated with 35% solvent B, the column was initially developed isocratically for 5 min, followed by linear gradients from 35% solvent B to 48% solvent B in 20 min and from 48% solvent B to 100% solvent B in 1 min. The column was then washed isocratically with 100% solvent B for 5 min, followed by a linear gradient from 100% B to 35% B in 1 min. This gradient profile ensured base-line separation of the oligopeptide components. The chromatograms were obtained with detection at 215 nm, and the amounts of peptides were determined from peak integrals by using a Shimadzu Chromatopak CR8A data processor. Three independent digestions were carried out to check the reproducibility of the assays. MALDI-TOF mass spectrometry analysis showed that cleavage of the peptides took place only at the predicted site, indicated above by an asterisk.

The initial velocity (*v*) (*i.e.* change of the individual substrate concentration at zero time) was calculated from the initial linear part of the decrease in the corresponding substrate peak areas after separation of the components by reversed-phase HPLC. Quantitative evaluation of the specificity profiles of the enzyme was based on the equation that describes the enzymatic reaction when competing substrates are present (16, 17),

$$(v_i/[S_i])/(v_j/[S_j]) = (k_{\text{cat},i}/K_{m,i})/(k_{\text{cat},j}/K_{m,j}) \quad\quad \text{(Eq. 1)}$$

where *i* and *j* denote the individual substrates, and k_{cat} and K_m are the catalytic constant and the Michaelis constant, respectively. Individual concentrations of the substrates, $[S_i]$, were determined by amino acid analyses.

Mutagenesis, Protein Expression, and Purification

We constructed an active site mutant of kumamolisin-As in which Glu[78] was replaced by His. The plasmid, pScpA/E78H, was prepared by *in vitro* mutagenesis of the plasmid pScpA (4) using a QuikChange™ mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's guidelines. The mutation was verified by DNA sequencing on

---

both strands using a Dye-Terminator Cycle Sequencing Kit (Beckman Coulter, Fullerton, CA) with a CEQ 2000 DNA analysis system (Beckman Coulter). Each of the plasmids, pScpA and pScpA/E78H, was used to transform *Escherichia coli* BL21 (DE3), and the transformant cells were grown at 37 °C in an LB medium (1 liter) containing 50 μg/ml ampicillin until the optical turbidity of the culture reached 0.6 at 600 nm. Expression of the wild-type and mutated kumamolisin-As genes was attained by adding isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.8 mM, followed by further fermentation for 3 h. The cells were harvested by centrifugation (5,000 × g, at 4 °C for 10 min), suspended in an appropriate volume of a 0.05 M sodium acetate buffer, pH 4.0, and disrupted at 4 °C by ultrasonication at 10 kHz. The cell debris was removed by centrifugation (18,000 × g, at 4 °C for 10 min), and the resultant supernatant (pH 4.0) was incubated at 55 °C for 5 h, followed by centrifugation. Almost all of the endogenous *E. coli* proteins were removed by centrifugation, and the resultant supernatant contained the expressed product of >96% homogeneity. The supernatant was concentrated by ultrafiltration with a YM-10 membrane using an Amicon 8200 unit and dialyzed at 4 °C against 0.02 M potassium phosphate buffer, pH 7.5 (buffer A). The concentrate was then loaded on an ΔKTA system equipped with a Mono Q HR10/10 column (Amersham Biosciences) equilibrated with buffer A. After loading the enzyme solution onto the column, followed by an extensive washing of the column with buffer A, the enzyme was eluted with a linear gradient of 0.15-0.5 M NaCl in buffer A in 75 min at a flow rate of 0.5 ml/min. In order to avoid any possible contamination of the mutant preparations with the wild-type activity, chromatographic purification of the E78H mutant was completed first, followed by that of the wild-type enzyme.

Limited Endoproteolysis

Collagen (3 mg) was reacted with 0.065 pM kumamolisin-As at 60 °C for 30 min in 0.5 ml of a 0.05 M sodium acetate buffer, pH 4.0. The mixture was then analyzed for proteolytic cleavage of collagen by SDS-PAGE (18). Protein bands in the gel were transferred to the polyvinylidene difluoride membrane by electroblotting, and the membrane was stained with Coomassie Brilliant Blue R250. Stained portions corresponding to the 70- and 30-kDa fragments of the membrane were excised using dissecting scissors and subjected to automated Edman degradation to determine the *N*-terminal amino acid sequences of the fragments.

Crystallization

Crystals of kumamolisin-As complexed with AcIPF, an inhibitor specifically designed for kumamolisin (10), as well as of uninhibited native kumamolisin-As and its E78H mutant were obtained by the vapor diffusion method. The complex was prepared by mixing 100 μl of 2.3 mg/ml kumamolisin-As in 10 mM sodium acetate buffer at pH 5.0 with 2.5 μl of 10 mg/ml AcIPF in Me₂SO. The samples of native kumamolisin, both complexed and uninhibited, were mixed with the well solution containing 27% polyethylene glycol 8000, 0.18 M ammonium sulfate, 10 mM dithiothreitol in 25 mM sodium acetate buffer at pH 4.0 or 4.2 at a 1:1 ratio, with the total volume of 4 μl. Crystals of the E78H mutant were grown in a mother liquor containing 0.2 M ammonium sulfate, 30% polyethylene glycol 8000 in deionized water.

X-ray Data Collection and Structure Refinement

X-ray diffraction data for the wild-type and E78H mutant apoenzymes were collected on a MAR345 detector mounted on a Rigaku RU200 rotating anode x-ray generator, operated at 50 kV and 100 mA. Data for the inhibitor complex were collected on beamline X9B, NSLS, Brookhaven National Laboratory, using an ADSC Quantum4 CCD detector. The reflections were integrated and merged using the HKL2000 suite (19), with the results summarized in Table I. All structures were refined using the program SHELXL (20). After each round of refinement, the models were compared with the respective electron density maps and modified using the interactive graphics display program O (21). The default SHELXL restraints were used for the geometrical (22) and displacement parameters; temperature factors were refined isotropically, due to the limited resolution of data. Water oxygen atoms were refined with unit occupancies, although some of the sites are probably only partially occupied. The refinement results are also presented in Table I. The coordinates and structure factors have been deposited in the Protein Data Bank (accession codes 1sn7, 1siu, and 1sio for the wild-type apoenzyme, E78H mutant, and the inhibitor complex, respectively).

RESULTS

Specificity Analysis of Kumamolisin-As—We have previously shown that digestion of collagen with kumamolisin-As at a substrate/enzyme ratio of 10^3:1 (mol/mol) yielded more than 50 peptides (4). Analysis of the primary structure of some of these peptides suggested that the enzyme may preferably act on the -Pro-Xaa*Gly-Yaa-Zaa- sequence of the α_1 and α_2 chains of collagen. When collagen was incubated with a lower amount of the enzyme (*i.e.* substrate to enzyme ratio, 2×10^5:1 mol/mol), incubation resulted in specific cleavage of collagen at a single site to produce two degradation products, one with a molecular mass of 70 kDa and another with a molecular mass of 30 kDa, as analyzed by SDS-PAGE. After electroblotting the protein bands of the fragments to the membranes, N-terminal amino acid sequences were analyzed by automated Edman degradation. The N-terminal sequence, NH₂-Gly-Leu-Hyp-Gly-Glu-Arg-Gly-Arg-Hyp-, could only be unambiguously determined for the 70-kDa fragment. A comparison of this sequence with the published amino acid sequence of the α_1 chain of bovine type I collagen (23, 24) revealed that it corresponded to the sequence starting from position 127 of the α_1 peptide. Since the N terminus of the α_1 chain of bovine type I collagen is blocked (24), these results showed that the specific cleavage site of collagen under these conditions was -Arg^{126}-*Gly^{127}- of the α_1 collagen chain.

The specificity at the P1′ site was then analyzed using the peptide libraries, Met-Gly-Pro-Arg*Xaa-Phe-Pro-Gly-Ser, in which the individual components differ only at the P1′ site. Each set of peptides (see "Experimental Procedures") was digested with kumamolisin-As, and the reaction was monitored by separation of the substrates and cleavage products by reversed-phase HPLC. The calculated specificity profile at the P1′ site is presented in Fig. 1*A*, indicating some preference for aromatic or bulky aliphatic amino acids. Unexpectedly, Gly was one of the least preferred amino acids at this position, although that residue is present at the P1′ position of the preferred cleavage site in collagen. We then analyzed the P1 specificity profile using the peptide library, Met-Gly-Pro-Xaa*Phe-Phe-Pro-Gly-Ser, where Phe was located at the P1′ site. We found that the P1 site had very high specificity for Arg (Fig. 1*B*), in excellent agreement with the results obtained by the limited proteolysis of collagen.

On the basis of the preliminary results of the specificity studies, we designed an IQF substrate, NMA-MGPH*FFPK-(DNP)DRDR, which could be utilized for a highly sensitive fluorometric assay of kumamolisin-As. This substrate was developed based on the addition of the fluorescent tag, *N*-methyl-anthranilic acid (NMA), and the quenching tag, 2,4-dinitrophenol (DNP), to a peptide where His and Phe were located at the respective P1 and P1′ positions. The design was based on our preliminary analysis of the specificity of kumamolisin-As that appeared to indicate that both histidine and arginine were equally good P1 substituents. Although later reinterpretation of the data summarized in Fig. 1*B* has shown this not to be the case, P1 His appears to be sufficient to provide a basis for the design of an acceptable substrate. Two D-arginine residues were added to the C terminus to enhance the solubility of the peptide in water. MALDI-TOF mass spectrometry analysis showed that cleavage of this peptide by kumamolisin-As took place only at the site indicated by an *asterisk*. The k_{cat} and K_m values of kumamolisin-As for the IQF substrate at pH 4.0 and at 40 °C were 395 ± 7 s^{-1} and 1.0 ± 0.2 μM, respectively. The k_{cat} and calculated k_{cat}/K_m values, 390 s$^{-1}\cdot\mu$M^{-1}, were even greater than the value (351 s^{-1} and 1.6 s$^{-1}\cdot\mu$M^{-1}, respectively) obtained at 60 °C for the parent peptide for the design of the IQF substrate, Met-Gly-Pro-Arg*Gly-Phe-Pro-Gly-Ser (4).
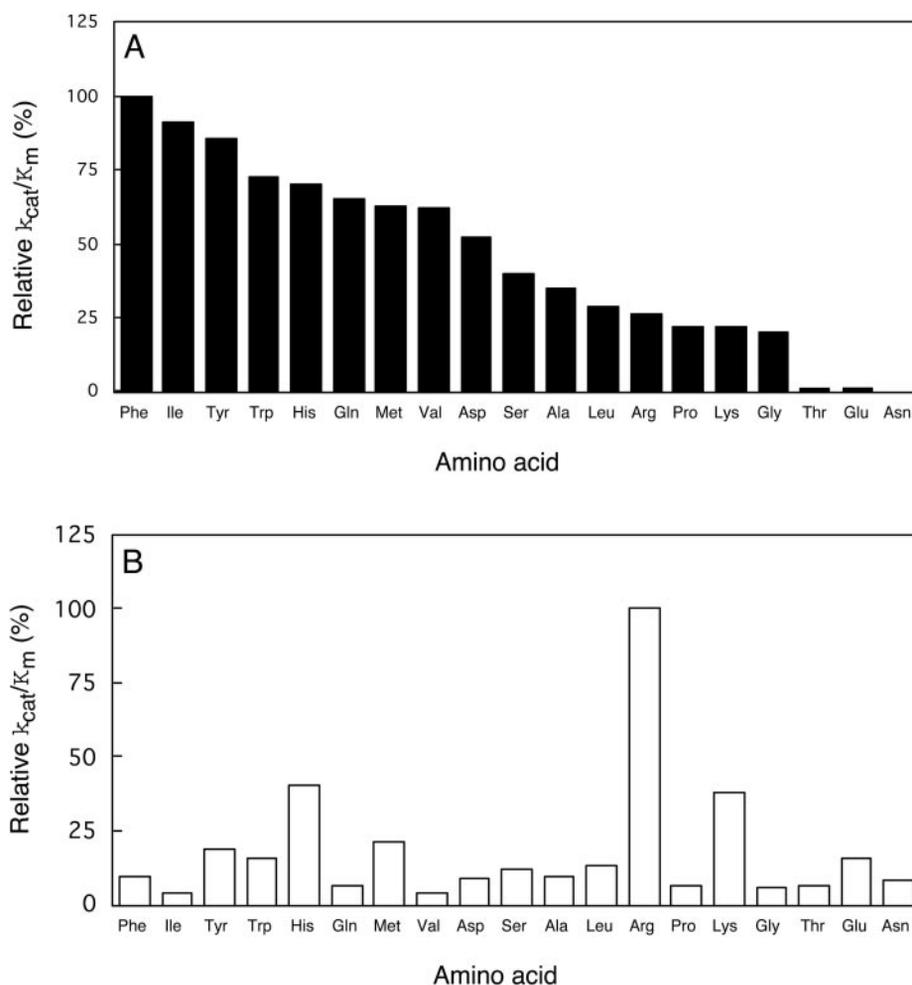
FIG. 1. **Graphical representation of specificity profiles of kumamolisin-As at P1′ and P1 sites.** The specificity profile for the P1′ site (*A*) was determined using the P1′ library, Met-Gly-Pro-Arg*Xaa-Phe-Pro-Gly-Ser, and that for the P1 site (*B*) was determined using the P1 library, Met-Gly-Pro-Xaa*Phe-Phe-Pro-Gly-Ser. For experimental details, see "Experimental Procedures."
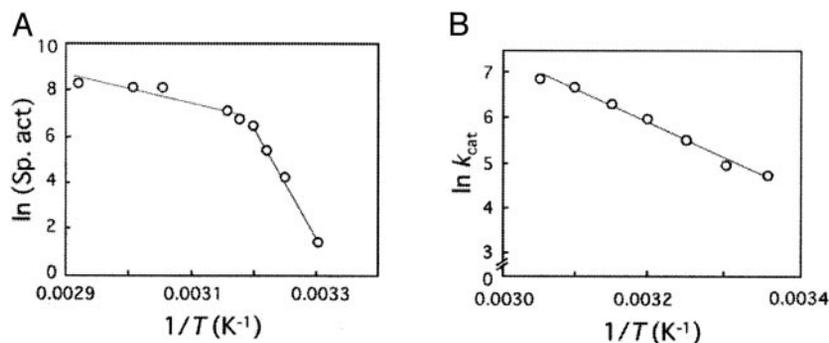
FIG. 2. **Effect of temperature on the collagenase activity (*A*) and the NMA-MGPH*FFPK(DNP)DRDR-hydrolyzing activity of kumamolisin-As (*B*).** *A*, the enzyme was assayed for collagenase activity at different temperatures by assay method I. Natural logarithm of specific activities (*Sp. act.*; μg of collagen solubilization·min^{-1}·mg of enzyme^{-1}) was plotted against the reciprocal of absolute temperature. *B*, the enzyme was assayed for hydrolysis of NMA-MGPH*FFPK(DNP)DRDR at different temperatures by assay method II. Natural logarithm of k_{cat} values was plotted against the reciprocal of absolute temperature.

Temperature Activity Profiles—Fig. 2 shows the Arrhenius plots of collagenase activity and the NMA-MGPH*FFPK-(DNP)DRDR hydrolyzing activity of kumamolisin-As. Arrhenius plots of collagenase activity showed a biphasic profile with a breakpoint at around 40 °C, which corresponds to the denaturing temperature of collagen. An activation energy, 396 kJ·mol^{-1}, that was calculated from plots below 40 °C was greater than the value (51 kJ·mol^{-1}) calculated from plots above 40 °C. On the other hand, the plots of the NMA-MGPH*FFPK(DNP)DRDR hydrolyzing activity showed a monophasic profile with a calculated activation energy of 61 kJ·mol^{-1}.

Crystal Structure of Kumamolisin-As—The preparations of kumamolisin-As used for the enzymatic studies were also successfully crystallized. Three crystal structures of this enzyme have now been solved and refined with data extending to the resolution of 1.8–2.3 Å. Wild-type apoenzyme and uninhibited E78H mutant crystallized in an isomorphous manner in space group P1 with a single molecule in the asymmetric unit. A complex of the wild-type enzyme with an inhibitor AcIPF crystallized in space group P2$_1$ with three molecules in the asymmetric unit; nevertheless, this crystal form yielded highest resolution diffraction data using a synchrotron x-ray source. Both crystal forms were distinct from the two reported crystal

FIG. 3. *Ribbon diagram* **showing the tracing of the main chain of kumamolisin-As.** Helices and strands are shown as *ribbons*, whereas loops are drawn as *thin lines*. The active site Ser[278] is marked in *orange*, and the covalently bound inhibitor AcIPF is shown in *magenta*.
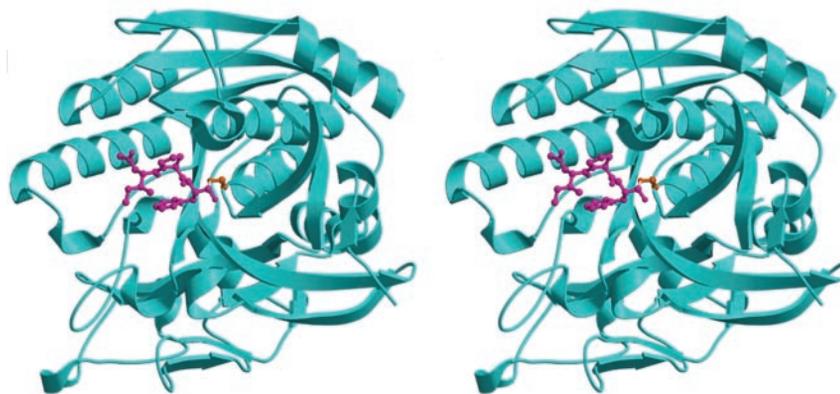
forms of kumamolisin (both in space group P2₁ but with very different unit cell parameters (10)) or from any of the crystal forms reported for sedolisin (previously known as PCP or PSCP) (25). The packing of the molecules in the two crystal forms of kumamolisin-As is quite different, and the residues involved in creating intermolecular contacts vary between them.

As expected, the structures of all variants of kumamolisin-As are extremely similar to both sedolisin and kumamolisin, with the agreement with the latter being particularly close (see below). All three enzymes are members of the family of serine-carboxyl peptidases (MEROPS S53-sedolisins) (2), with a general protein fold resembling that of serine proteases from the subtilisin family, although sedolisins are considerably larger than subtilisins. Therefore, practically all of the secondary structure elements found in subtilisins are also present in sedolisins, although, for obvious reasons, the opposite is not true.

The crystal structure of the native kumamolisin-As comprises residues 4–357, missing three residues at its N terminus and seven residues at the C terminus; the rest of the chain could be traced without any interruptions (Fig. 3). The location of the last C-terminal residue visible in the electron density is the same as in the structures of kumamolisin. That part of the sequence of the latter enzyme was subject to several reinterpretations (9, 10), most recently in a new NCBI entry, gi: 25137473. The latest reinterpretation of the primary structure of kumamolisin provides a good match with the corresponding part of kumamolisin-As, and it is thus not surprising that the structures of the two enzymes are also in agreement in this area.

Three *cis* peptide bonds are present in kumamolisin-As. Two of them involve prolines, Pro[181] and Pro[251], with both residues involved in creating sharp turns of the polypeptide chain. Their conformation is conserved not only in kumamolisin but also in sedolisin and, for Pro[181], even in subtilisin, attesting to the importance of such conformation for the maintenance of the polypeptide fold for this protein superfamily. The third, rare *cis* peptide not adjacent to a proline is found between Ile[330] and Tyr[331], in a turn stabilized by strong hydrogen bonds extending from Ser[167] Oγ to the carbonyl oxygen of Ile[330] and from the main-chain amide of Tyr[331] to the Oδ1 of Asn[322]. Interestingly, Tyr[331] is conserved in sedolisin, yet that residue is found in the latter enzyme in the common *trans* configuration. However, the preceding residue in sedolisin is a glycine, and this part of the peptide chain is extended, whereas it makes two rather sharp turns in kumamolisin-As. The *cis* conformation of Tyr[331] is fully conserved in kumamolisin as well.

The Ca²⁺ binding site, found in all sedolisin-like enzymes (2), is also present in kumamolisin-As. This ion exhibits almost perfect octahedral coordination by the carboxyl oxygens of Asp[316] and Asp[338], main chain carbonyls of residues 317, 334, and 338, and the completely buried Wat[501]. The latter water also accepts hydrogen bonds from the amide nitrogen of residue 317 and from the side chain Oγ of Ser[341], with the angle between these groups being ~100°. The temperature factors of groups involved in the creation of this site, including Wat[501] and the Ca²⁺ ion are comparatively low (average B ~22 Å² in the native structure, even lower in the inhibitor complex), attesting to the stability of this structural feature of sedolisins.

The Active Site and the Inhibitor Binding Site—The active site triad is formed in kumamolisin-As by Ser[278], Glu[78], and Asp[82]. The side chains of these residues are connected by short hydrogen bonds into an extended catalytic machinery that includes two additional residues, Glu[32] and Trp[129], also present in kumamolisin (10). In order to define the substrate binding site, we solved the structure of kumamolisin-As complexed with the inhibitor AcIPF, previously used in the structural studies of sedolisin (12) and kumamolisin (10). As in these two other structures, this inhibitor is covalently bound to Ser[278], defining this residue as the catalytic nucleophile (Fig. 4). Asp[164] and an amide nitrogen of the catalytic serine form an oxyanion hole, which accommodates the hemiacetal oxygen atom of the inhibitor. Three main-chain hydrogen bonds are formed between the inhibitor and the strand 128–132 of the kumamolisin As: one between the amide nitrogen of Phe at the P1 position and carbonyl oxygen of Ser[128] and two between Ile at the P3 position and Gly[130], respectively. The S1 binding pocket is formed by Ala[161], Gly[163], Gly[130], Thr[277], and Asp[179], the carboxylate of the latter hydrogen bonded to the carboxylate of Asp[169] (the distance between the oxygen atoms was 2.5 Å). The presence of negatively charged residues in the S1 pocket provides the structural basis for the recognition of an Arg-Gly sequence in collagen that is specifically cleaved by kumamolisin-As (see above). On the other hand, such a short hydrogen bond between two aspartates indicates that one of these residues has to be protonated, which may explain the enhanced enzymatic activity at low pH. The proline ring in the S2 binding pocket is tightly packed between the aromatic side chain of Trp[129], and the hydrophobic fragment of Glu[78]; therefore, small hydrophobic residues seem to be good candidates for the P2 position in the substrates of kumamolisin-As. The side chain of isoleucine at the P3 position interacts with the P1 Phe ring of the inhibitor and with residues Gly[131]-Pro[132] of the enzyme.

A superposition of the crystal structures of subtilisin complexed with a protein inhibitor eglin (26) and fiddler crab collagenase complexed with ecotin (27) on kumamolisin-As complexed with AcIPF allows modeling of a longer peptide bound in the active site of the latter enzyme (Fig. 5). We can

FIG. 4. **Electron density $2F_o - F_c$ map, contoured at 1σ level, for the inhibitor AcIPF covalently bound to Ser[278] in kumamolisin-As.**
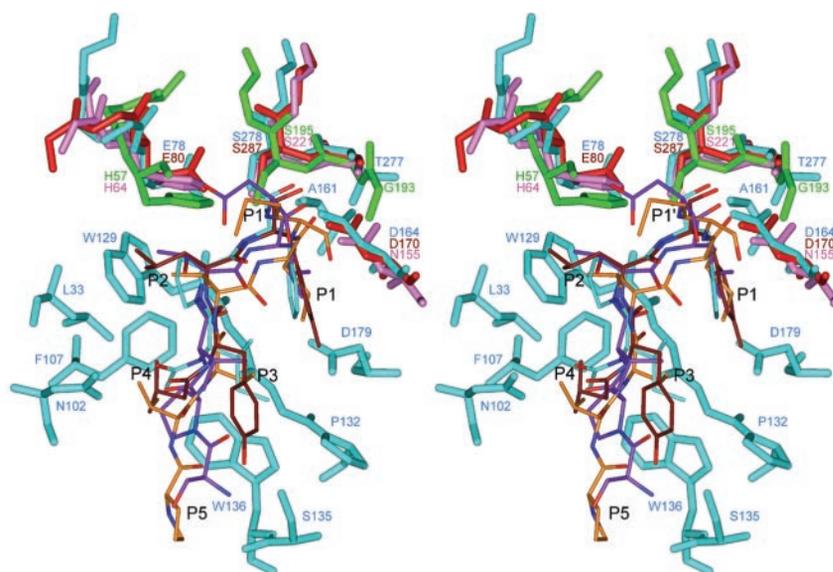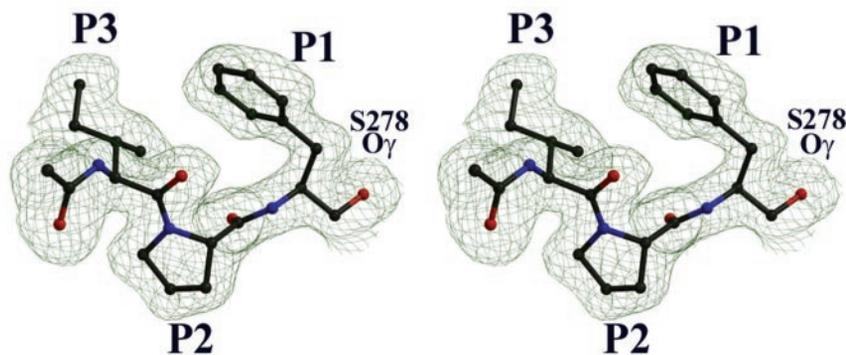
FIG. 5. **A superposition of the active sites of four proteases.** Kumamolisin-As with bound AcIPF is shown in *blue*; sedolisin and bound tyrostatin are colored *red* and *brown*, respectively; subtilisin is shown in *lilac*, and its inhibitor eglin is shown in *violet*. Fiddler crab collagenase is *colored green*, whereas its inhibitor ecotin is *orange*. The Cα atoms of kumamolisin-As, sedolisin, and subtilisin were superimposed with the program ALIGN (34). Superposition of kumamolisin-As with fiddler crab collagenase is described under "Results."

postulate that the S4 pocket occupies very well defined area between the rings of Trp[129] and Phe[107], with additional interactions with the P4 residue of the substrate also provided by the side chains of Leu[33] and Asn[102]. The side chain of the P5 residue points into the solvent, and the closest enzyme residues are Ser[135] and Trp[136]. The S1′ pocket seems to have enough space to accommodate residues larger than Gly, which is found at the P1′ position of the preferred cleavage site in collagen. Therefore, the presence of a Gly residue in this position should be attributed to the properties of the collagen molecule rather than to the substrate specificity of kumamolisin-As.

Conserved Solvent Molecules—Since the structures reported here have been solved at medium to high resolution, a considerable number of solvent positions could be identified (Table I). The number of assigned solvent molecules approaches that of the amino acids present in the asymmetric unit, at least in the two structures solved at the resolution of 2 Å or higher. Many solvent positions are conserved in all three structures, despite the presence of a mutation in one of them and the bound inhibitor molecules in another one. The number of conserved solvent molecules (within 1 Å of each other) is 104 for the two isomorphous structures (native *versus* E78H mutant), whereas the number of common solvents for the native kumamolisin-As and molecule A of the complex is 122. The number of conserved solvent molecules that are present in all three structures (defined as above) is 77. The common water molecules have identical numbers in the deposited coordinates but increased by 1000 for each molecule of the complex.

At least nine completely or partially buried solvent molecules can be identified in equivalent locations in both the native and mutant kumamolisin-As, as well as in all three molecules of the inhibitor complex. These waters have been numbered 501–509 in the first two structures and 1501–1509, 2501–2509, and 3501–3509 in the three molecules present in the asymmetric unit of the inhibitor complex. Wat[501] was discussed above in the context of the calcium-binding site, since it is providing the sixth oxygen ligand of the complexed metal ion. Wat[502] and Wat[505] interact with each other, being buried behind the active site residue Asp[82], in an area inaccessible to solvent (or the substrate). Wat[505] interacts with the Oγ atoms of both Ser[126] and Ser[128]. Two other pairs of buried waters are Wat[503] and Wat[504] as well as Wat[506] and Wat[507]. All four of these waters are located in the same vicinity, and each of them interacts with 2–3 polar groups belonging to the main chain of the enzyme. The final pair consists of Wat[508] and Wat[509]. All of these water molecules must be very important for maintaining the fold of this class of proteins, since they are also present in virtually the same locations in not only kumamolisin but also sedolisin. Interestingly, the only solvent molecule from that group that has no direct counterpart in sedolisin is Wat[505], substituted in the latter by an Oδ1 atom of Asn[131] that makes virtually the same interactions. This substitution emphasizes the fact that either a polar atom belonging to an amino acid or a water molecule might play the same role in stabilizing local structure of a protein.

Comparison of Kumamolisin-As and Kumamolisin—Kumamolisin-As is very closely related in sequence and in structure

TABLE I

Details of X-ray data collection and structure refinement.

Parameters	Values		
	Wild type	E78H	Wild type/AcIPF
Space group	P1	P1	P2$_1$
Unit cell dimensions (Å)			
a	41.81	42.16	49.37
b	44.95	45.04	238.73
c	49.11	49.43	49.25
α	113.9	114.82	90.0
β	106.1	106.36	113.7
γ	102.3	102.03	90.0
Resolution (Å)	2.0	2.3	1.8
Measured reflections	28,815	23,703	328,423
R_{merge} (%)	3.5 (12.5)[a]	4.3 (11.8)	5.3 (31.1)
$I/\sigma(I)$	25.8 (6.7)	17 (6.1)	23 (2.9)
Completeness (%)	90 (61.6)	94.9 (80.6)	97 (76)
Refinement			
R, no σ cut-off (%)	15.0	15.9	17.3
R_{free} (%)	25.5	28.7	24.3
Reflections used in refinement	16,732	11,509	84,178
Reflections used for R_{free}	886	607	4,457
Root mean square bond lengths (Å)	0.010	0.008	0.011
Root mean square angle distances (Å)	0.033	0.030	0.032
Protein atoms	2526	2527	7583
Inhibitor atoms			87
Other ligand atoms[b]	1	1	18
Water sites	286	179	935
Protein Data Bank accession code	1sn7	1siu	1sio

[a] Values in the highest resolution shell are shown in parentheses.

[b] One Ca^{2+} ion is present in all molecules; one sulfate ion is bound to each enzyme molecule in the inhibitor complex.

to kumamolisin. Only 57 residues (10.1%) differ between the complete enzyme sequences that include the prosegments, whereas 27 residues differ in the mature enzymes (92.7% identity). The only insertion in kumamolisin-As is a single residue in the propeptide. Even for those residues that do differ, substitutions are mostly conservative. Some of these differences are found on the surface of the molecules and thus do not seem to influence the overall structures of the enzyme. The side chains of A40T, P52S, S65T, S73N, L120H, A203E, Q204R, A229S, H233R, A271T, A298P, A311P, D312E, Q329R, and V349I (the first residue in each position refers to kumamolisin-As and the second to kumamolisin) are located primarily in surface loops, and the differences in their identity do not seem to affect the global structure of the protein. Three other residues, T137A, S138P, and A140S, are part of a single turn. An interesting feature in the latter case is that whereas Thr[137] makes a hydrogen bond to an amide nitrogen of Ala[140] in kumamolisin-As, the presence of Pro[138] in kumamolisin (instead of Ser in kumamolisin-As) provides stability to the turn even in the absence of any extra hydrogen bonding.

Several buried residues also differ between these enzymes, and a few of the differences are not compensated. One of the buried residues is Phe[95] in kumamolisin-As (Ile in kumamolisin). The side chain of this residue packs against the conserved Ile[83] in either protein, whereas other neighboring side chains are not shifted, thus creating a small cavity in kumamolisin. On the other hand, a small cavity is present in kumamolisin-As next to the buried Ala[216]. The corresponding Ser[216] in kumamolisin is hydrogen-bonded to the carbonyl oxygen of Trp[208]. There is no rearrangement of the residues around this site, and thus both a cavity and an unsatisfied polar interaction are present in the former enzyme. In the case of another buried residue, an extra methyl group present in Leu[253] in kumamolisin-As (compared with Val in kumamolisin) leads only to a small rearrangement of the neighboring side chains, and neither enzyme has any visible cavity in this area. Interestingly enough, one cavity of significant size (15 Å^3 in kumamolisin, as estimated with the program VOIDOO (28)) is found in an area surrounded by the side chains of Leu[188], Leu/Val[253], Phe[284],

Leu[287], Leu[302], and Leu[351]. This cavity is surrounded by non-polar atoms only and appears to be empty.

The root mean square deviation between the Cα coordinates of the uninhibited wild-type kumamolisin-As and molecule A of the native kumamolisin (Protein Data Bank code 1gt9) is 0.33 Å for 327 atom pairs. This value can be compared with the 0.08 Å for 309 atom pairs found by superimposing the two molecules of kumamolisin present in the asymmetric unit.

The root mean square deviation between molecules A and B of the inhibited kumamolisin-As is 0.27 Å for 335 atom pairs, with the most significant differences between them found at the N termini, where the two molecules diverge by as much as 6 Å. The first three residues in both structures are exposed to the solvent and have different conformations. Similar overall comparisons for molecules A and C yield the value of 0.27 Å for 331 pairs, and comparisons for B and C yield 0.31 Å for 339 pairs. By comparison, molecules A of the inhibitor complexes of kumamolisin-As and kumamolisin (Protein Data Bank code 1gtl) are only 0.37 Å for 331 target pairs, whereas a comparison of the former with another inhibited kumamolisin complex studied at much higher resolution (Protein Data Bank code 1gt) is 0.35 Å for 335 pairs. Clearly, the overall structure of these molecules is extremely similar.

Comparison of Kumamolisin-As and Fiddler Crab Collagenase—The only collagenolytic serine proteases described to date belong to the chymotrypsin family (7). Since kumamolisin-As is the first subtilisin-like serine protease shown to have collagenase activity, we compared it with the chymotrypsin-like enzymes in order to analyze the structural basis for the recognition of a collagen molecule. The structure of fiddler crab collagenase complexed with the dimeric serine protease inhibitor ecotin (27) was selected for this comparison, since it had been suggested that the protease binding loop of ecotin adopts a conformation mimicking that of the cleaved strand of collagen (27). Since the global folds of these two enzymes that belong to different families of serine proteases have no similarity at all, it is not possible to obtain meaningful superimpositions of the overall structures by using Cα traces. However, since both enzymes recognize and cleave collagen molecules in preference
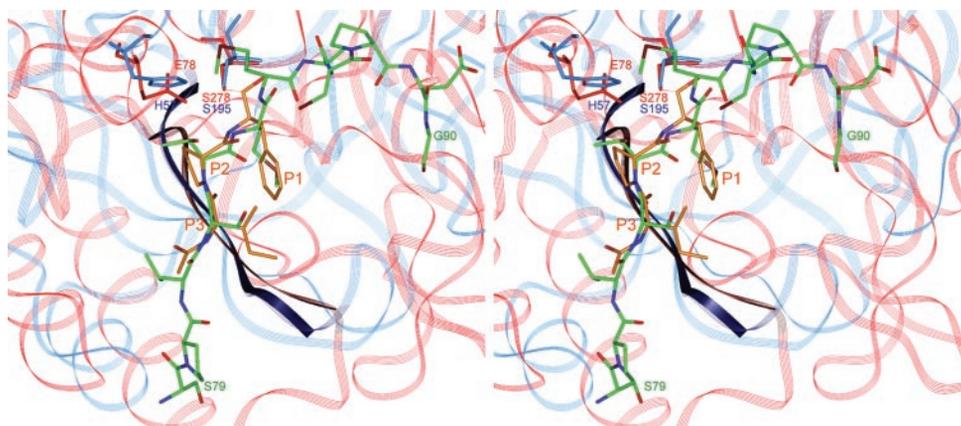
FIG. 6. **Superposition of kumamolisin-As and fiddler crab collagenase.** The two structures were superimposed as described under "Results." The main chain tracing of kumamolisin-As and fiddler crab collagenase is shown as *red* and *blue ribbons*, respectively, and the superimposed side chains of the active sites residues are shown as *sticks* in corresponding *colors*. The strands 127–132 in kumamolisin-As and 213–218 in collagenase are emphasized in *black* (see under "Results"). The inhibitor AcIPF is shown in *gold*, and a fragment of ecotin is shown in *green*.

to any other substrates, we searched for the catalytically important residues, which could be used as guide points in the superposition of their structures. The most successful trial was based on the superposition of three structural elements: the catalytic Ser[278] and Ser[195], general base residues Glu[78] and His[57], and the oxyanion holes (Fig. 5), which led to the remarkably accurate alignment of the residues at the P1–P3 positions in ecotin and of the inhibitor bound to kumamolisin-As (Fig. 6). It was also found upon this superposition that the strand 213–218 in the latter enzyme, which is involved in extensive hydrogen-bonded interactions with ecotin, is also well aligned with the strand 127–132 in kumamolisin-As (Fig. 6). The conformations of both strands, as well as of the ligands bound to the enzymes, are very similar, suggesting that corresponding types of interactions may also be maintained in kumamolisin. The protease binding loops of ecotin fit perfectly well into the large groove in kumamolisin-As, and the surface of this groove is negatively charged similarly to collagenase (Fig. 7). Both types of structural features, namely a large groove with extended negative charge and an extensive hydrogen-bonded network, were considered to be important for specific recognition of collagen.

Biochemical Characterization of the E78H Mutant—The E78H mutant of kumamolisin-As was expressed as a soluble protein and purified from the crude extract of transformant *E. coli* cells by a two-step purification procedure: an acid treatment of the extract at 55 °C followed by chromatography on Mono Q. In the crude extract of transformant cells, the 57-kDa precursor of the E78H mutant could be very slowly converted during the acid treatment into a mature form, which was indistinguishable in size from the 37-kDa mature form of the wild type enzyme. The expressed E78H mutant exhibited only very low enzyme activity. Highly sensitive fluorometric enzyme assay using NMA-MGPH*FFPK(DNP)DRDR allowed us to determine the kinetic parameters of the mutant for the substrate at pH 4.0 and 40 °C. The K_m value (0.7 ± 0.2 μM) was essentially identical to that of the wild-type enzyme, whereas the k_{cat} and k_{cat}/K_m values (0.033 ± 0.006 s^{-1} and 0.047 μM$^{-1} \cdot$s^{-1}, respectively) were 0.008–0.012% of the values of the wild type.

The Structure of the E78H Mutant—Crystals of the E78H mutant of kumamolisin-As could only be grown for the uninhibited form of the enzyme; cocrystallization of the mutant with the inhibitor was not successful. These crystals are fully isomorphous with those of the uninhibited wild-type enzyme, and the structures are very similar (root mean square deviation of 0.275 Å for 346 Cα pairs). However, the conformation of the

catalytic Ser[278] differs significantly between these proteins (Fig. 8). Whereas the torsion angle χ1 is $-78°$ for the wild-type apoenzyme and the inhibitor complex, it is $74°$ for the mutant. The Oγ atom of Ser[278] interacts only with two water molecules. One of them is a highly conserved water (Wat[570]) that is also bound to the main chain carbonyl of Gly[275], and the other is Wat[648], mediating an interaction with the carboxylate group of Asp[164], the residue that forms the oxyanion hole. The conformation of His[78] differs very significantly from that of Glu[78], with the imidazole ring held in place by the interactions with the indole moiety of Trp[129] as well as by hydrogen bonds with Oγ of Ser[128] and its main chain carbonyl. His[78] makes a short contact with the P2 Pro of AcIPF of the inhibited structure (closest approach $\sim$2 Å), possibly explaining the failure to observe a mutant-inhibitor complex. A switch of the catalytic histidine to a position that prevented formation of a proper triad was previously described for a D102N mutant of rat trypsin with very low activity (29). Although the conformation of Asp[82] is virtually the same in all three structures of kumamolisin-As, substitution of glutamic acid by histidine in the mutant does not allow formation of a hydrogen bond between residues 78 and 82 analogous to the bond seen in both the wild-type apoenzyme and in the inhibitor complex between the carboxylates of Asp[82] and Glu[78]. The closest distance between any atoms of the imidazole ring of His[78] and the carboxylate of Asp[82] is 3.65 Å, and this contact involves a carbon rather than nitrogen of the ring. Thus, the catalytic triad of the mutant simply does not exist, since the three residues do not interact with each other at all. Interestingly, however, the extension of the catalytic triad, first described in kumamolisin (10) and consisting of Glu[32] and Trp[129] is still in place, and a chain of hydrogen bonds between these two side chains and Asp[82] is present.

DISCUSSION

Kumamolisin-As can efficiently degrade an insoluble, fibrillar form of collagen but acts only poorly on albumin and casein. The biological significance of this collagenolytic activity could be further substantiated by the fact that the producer bacterium, *A. sendaiensis* NTAP-1, cannot grow on peptone (a partial hydrolyzate of casein) as a sole nitrogen source but can grow on collagen and gelatin (4). It is likely that this soil bacterium plays a saprophytic role through degradation of collagenous materials (such as those from carcasses) in cooperation with other saprophytes under thermoacidophilic conditions that emerges during microbial proliferation and should
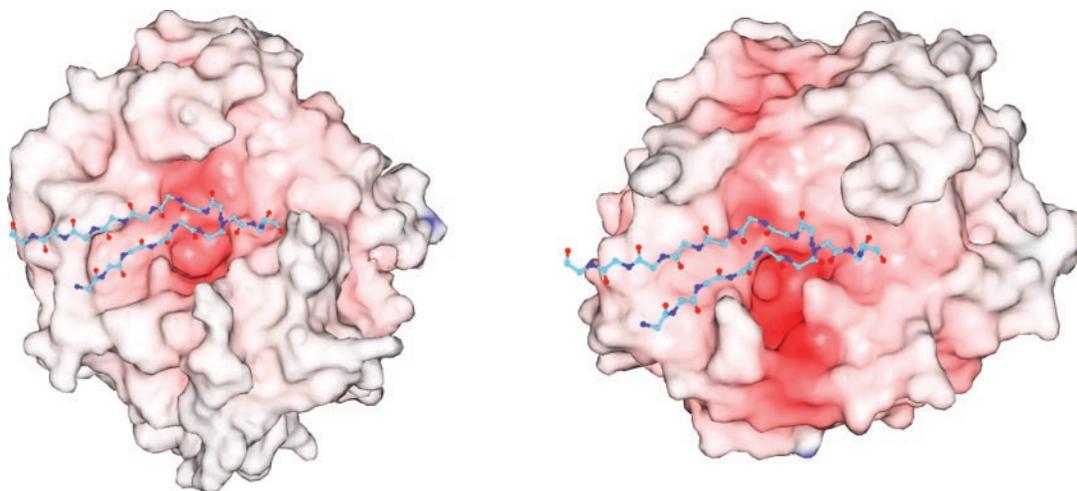
FIG. 7. **Charge distribution on the surface of kumamolisin-As (*left*) and fiddler crab collagenase (*right*).** The coordinates shown in *ball-and-stick* are for ecotin bound to the latter enzyme and modeled in the structure of the former one. This figure was created with the program GRASP (35).

FIG. 8. **Comparison of the active sites of the wild-type kumamolisin-As (native and inhibited) and the E78H mutant.** The coordinates of molecule A of the inhibitor complex (*orange*) were superimposed on those of the E78H mutant (*blue*) and the native, uninhibited kumamolisin-As (*green*). The inhibitor is *colored* in *magenta*. Wat^{570} is conserved in all three structures, whereas Wat^{648} is present only in the mutant.
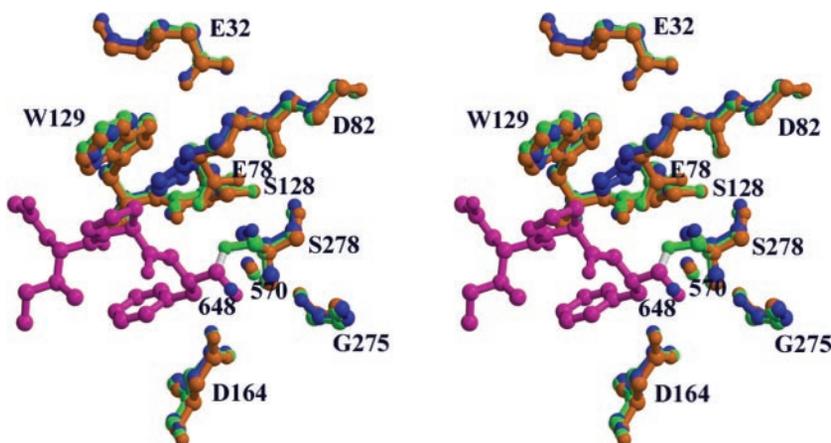
play an important role in global cycling of nitrogen. Thus, kumamolisin-As, together with kumamolisin, have been identified as novel collagenases, which are highly thermostable and show the highest activity at acidic pH 3.9 (4). These enzymes are the first examples of collagenases from the sedolisin family, quite distinct from the previously described collagenase families that include zinc-dependent metallopeptidases and serine peptidases with trypsin/chymotrypsin fold (6, 7). It must be stressed, however, that the use of the term "collagenase" might still be subject of some controversy, since kumamolisin-As does not make a specific three-quarters/one-quarter cut of the triple-helical, native collagen, primarily between -Gly^{775}-*Ile^{776}-, characteristic for the previously described collagenases. However, kumamolisin-As does prefer collagen to any other protein substrate and preferentially cleaves the α_1 collagen chain at the single specific cleavage site, -Arg^{126}-*Gly^{127}-. We would also like to emphasize the differences in the assay conditions under which a single cut of collagen molecule was observed for kumamolisin-As (pH 4, 60 °C) and for other collagenases (pH 8, 25 °C). Since the structure of the collagen triple helix is highly dependent on both parameters (see below), it may differ under these two assay conditions, which, in turn, may impact the specificities for the cleavage sites.

The results of the present structural and biochemical studies provide important clues that should help in understanding why this member of the sedolisin family can specifically and efficiently degrade collagen. The observed specific, collagenolytic action of kumamolisin-As could be explained in terms of (i) the presence of a substrate-binding groove on the surface of the enzyme, (ii) substrate specificities at the P1 and P2 sites, and (iii) relaxation of collagen under the reaction conditions of low pH and high temperature. Each of these aspects is further discussed below.

The crystal structure of kumamolisin-As revealed the presence of a long groove that encompasses the active site. Three catalytic residues (Ser^{278}, Glu^{78}, and Asp^{82}) as well as Asp^{164}, which participates in the formation of an oxyanion hole, are located at the bottom of the groove. Docking studies using ecotin, which mimics the collagen α chain (27), predicted that this groove should provide sufficient space for binding of a relaxed collagen molecule. Moreover, many negative charges are clustered on the surface of the groove, facilitating the binding of positively charged collagen (pI > 9). A substrate-binding groove with a similar electrostatic nature has also been identified in the fiddler crab collagenase (7) (Fig. 7). When the crystal structure of kumamolisin-As with the ecotin modeled in its active was superimposed on sedolisin (30), we found that the loop 226–229 of the latter enzyme, which lies over the substrate-binding groove, is longer and has a different conformation than a corresponding loop in kumamolisin-As. Since that loop collides with ecotin (and, by extrapolation, collagen), this observation may explain the very low collagenase activity of sedolisin (4).

Limited endoproteolysis as well as subsite specificity studies revealed that kumamolisin-As shows the highest P1 preference for Arg and, to a lesser extent, for His. This observation could be consistently explained by the fact that Asp^{179} is located at the bottom of the S1 site; the P1 preference for Arg and His

likely arises from electrostatic interactions of their positively charged side chains with the carboxylate of Asp[179]. The other important structural feature revealed by the crystal structure of kumamolisin-As is that the S2 subsite is so narrow that it appears to be able to accommodate only a small amino acid residue, such as glycine, proline, or alanine. Although we did not experimentally examine the P2 preference of kumamolisin-As, it was shown that kumamolisin, in which the S2 subsite is very similar (10), indeed exhibits a P2 preference for small amino acids (31). Such a structure of the S2 pocket is expected to cause significant restraints on the enzyme's preference for protein substrates. It is important to note that collagen contains a significant number of arginine residues (*e.g.* 4–5 mol/mol%), many of which are immediately preceded in the collagen α chains by a small amino acid. For example, nine -Pro-Arg-sequences as well as more than 15 related sequences (*e.g.* -Gly-Arg- and -Ala-Arg-) could be identified in the primary structure of the α_1 (I) chain of bovine collagen (24), and these sequences should be effectively accommodated in the -S2-S1-subsites of kumamolisin-As.

Because the N termini of peptides derived from bovine type-I collagen upon digestion with kumamolisin-As were all shown to be glycines (4), the possible importance of this amino acid at the P1′ position could also be postulated. However, the present analysis of kumamolisin-As revealed relatively broad preference for the amino acids at the P1′ site of peptide substrates, where the aromatic or bulky aliphatic amino acids were preferred. Crystal structure of kumamolisin-As also indicates that the S1′ site has enough space to accommodate a larger size of the residue than Gly and therefore appeared to be less important for the specificity for collagen.

Arrhenius plots of the collagenase activity of kumamolisin-As show a biphasic profile with a break point at around 40 °C. Since the plots of the IQF peptide-hydrolyzing activity show a monophasic profile within the same temperature range, it is highly unlikely that this break arose from the conformational changes of the enzyme. The break point corresponds to the starting denaturing temperature of the triple helical structure of collagen and can most likely be ascribed to partial denaturation of the collagen substrate. The activation energy calculated from plots above 40 °C for collagenase activity was significantly lower than the value below this temperature; therefore, it is strongly suggested that the substrate should be at least partially denatured before proteolysis can occur. All collagenases cleave insoluble forms of collagen molecules that aggregate into collagen fibrils. The relaxation of fibrils as well as the "helix-to-coil" transition of collagen molecules should at least in part occur under the thermoacidophilic conditions (pH 4.0 and 60 °C) of the enzyme assays. Thus, the enzyme probably acted on such relaxed (or partially unwound) portions of the substrate to cleave the -Pro-Arg*Gly- and related sequences in the collagen α chains.

For comparison, the mode of collagenolytic actions of Zn^{2+}-dependent collagenases of mammals (matrix metalloproteinases, MEROPS family M10) and bacteria (MEROPS family M9) needs to be mentioned. The crystal structure of a matrix metalloproteinase from porcine synovia showed that the mature form of the enzyme consists of an N-terminal catalytic domain and a C-terminal hemopexin-like domain with a characteristic four-bladed β-propeller structure. It was proposed that the hemopexin-like domain first interacts with a collagen molecule and relaxes it to make it possible to undergo subsequent cleavage by the catalytic domain (32). The importance of relaxation of a collagen molecule during its enzymatic degradation was also suggested for a clostridial collagenase, an exo-type M9 collagenase (33). Although the crystal structure of the clostrid-

ial collagenase is not yet determined, it was suggested that a C-terminal domain of the enzyme should play an important role in the binding and relaxation of collagen. These studies have clearly shown that the collagen molecule has to be relaxed before it could be enzymatically degraded. In the case of M9/M10 collagenases, all of which are enzymes of mesophilic organisms, such relaxation could be attained at nondenaturing temperatures of collagen by the action of the collagen-binding domain. In the case of kumamolisin-As, however, relaxation of the substrate can be easily attained under the growth conditions of the thermoacidophilic bacterium producing this enzyme; for such reason, the enzyme might have evolved as a collagenase without such a collagen-binding domain.

Sedolisins, including kumamolisin-As, have been defined as serine peptidases with a unique catalytic triad (Ser-Glu-Asp) instead of the canonical Ser-His-Asp triad as well as an aspartic acid residue in the oxyanion hole, and these structural features have been postulated to be among factors that cause acidophilic activities of these enzymes. Thus, another important objective of this study was to examine the effect of substitution of Glu[78] of kumamolisin-As by histidine on the enzymatic activity and the hydrogen bond network around catalytic triad of this sedolisin family enzyme. Previous mutational studies with kumamolisin showed that replacement of the glutamic acid residue corresponding to Glu[78] of kumamolisin-As by alanine caused a complete loss of its proteolytic activity (10). In this study, that residue was mutated in order to mimic the canonical catalytic triad of a serine protease in a member of the sedolisin family. We have shown that the mutant has only very low activity (~0.008% of k_{cat} for the wild-type enzyme) with a K_m value that was essentially identical to the value of wild type. Crystal structure of the uninhibited form of the E78H mutant showed that the conformation of the residues forming the new "catalytic triad" was such that the expected hydrogen bonding pattern was not present at all. By changing only the torsion angles of the side chains of Ser[278] and His[78] (without adjustment of any main-chain parameters), it is possible to create a strong hydrogen bond between these two residues, but it is not possible to adjust His[78] in any way that would result in that residue also making a hydrogen bond to Asp[82]. This observation must be related to the much higher level of flexibility of glutamate (with three side-chain torsion angles) compared with histidine (with only two). Thus, it is unlikely that a normal catalytic triad could be recreated in this mutant without the (unlikely) considerable rearrangement of the main chain in the vicinity of the active site. These observations strongly suggested that His[78] of the mutant could not function as a general base in a triad, albeit very weak enzyme activity of the mutant could arise from the possible formation of a catalytic dyad consisting of His[78] and Ser[278]. This, in turn, suggested that the glutamate residue at position 78 of kumamolisin-As should play an indispensable role in the maintenance of the structural integrity of the active site of the enzyme. Finally, we should also note that the preservation of the low pH activity maximum may be due to the presence of Asp[164] in the oxyanion hole. A double mutant E78H/D164N might exhibit activity at higher pH, but this point can be answered only by further experimentation, now in progress.

REFERENCES

1. Tsuruoka, N., Isono, Y., Shida, O., Hemmi, H., Nakayama, T., and Nishino, T. (2003) *Int. J. Syst. Evol. Microbiol.* **53,** 1081–1084
2. Wlodawer, A., Li, M., Gustchina, A., Oyama, H., Dunn, B. M., and Oda, K. (2003) *Acta Biochim. Polon.* **50,** 81–102
3. Nakayama, T., Tsuruoka, N., Akai, M., and Nishino, T. (2000) *J. Biosci.*

Bioeng. **89,** 612–614

4. Tsuruoka, N., Nakayama, T., Ashida, M., Hemmi, H., Nakao, M., Minakata, H., Oyama, H., Oda, K., and Nishino, T. (2003) *Appl. Environ. Microbiol.* **69,** 162–169

5. Harrington, D. J. (1996) *Infect. Immun.* **64,** 1885–1891

6. Okamoto, M., Yonejima, Y., Tsujimoto, Y., Suzuki, Y., and Watanabe, K. (2001) *Appl. Microbiol. Biotechnol.* **57,** 103–108

7. Tsu, C. A., Perona, J. J., Fletterick, R. J., and Craik, C. S. (1997) *Biochemistry* **36,** 5393–5401

8. Murao, S., Ohkuni, K., Nagao, M., Hirayama, K., Fukuhara, K., Oda, K., Oyama, H., and Shin, T. (1993) *J. Biol. Chem.* **268,** 349–355

9. Oyama, H., Hamada, T., Ogasawara, S., Uchida, K., Murao, S., Beyer, B. B., Dunn, B. M., and Oda, K. (2002) *J. Biochem.* (*Tokyo*) **131,** 757–765

10. Comellas-Bigler, M., Fuentes-Prior, P., Maskos, K., Huber, R., Oyama, H., Uchida, K., Dunn, B. M., Oda, K., and Bode, W. (2002) *Structure* **10,** 865–876

11. Madigan, M. T., Martinko, J. M., and Parker, J. (2000) *Biology of Microorganisms*, Prentice-Hall, Saddle River, NJ

12. Wlodawer, A., Li, M., Gustchina, A., Dauter, Z., Uchida, K., Oyama, H., Goldfarb, N. E., Dunn, B. M., and Oda, K. (2001) *Biochemistry* **40,** 15602–15611

13. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254

14. Leatherbarrow, R. J. (1991) *Trends Biochem. Sci.* **15,** 455–458

15. Sasagawa, T., and Teller, D. C. (1987) in *Handbook of HPLC for the Separation of Amino Acids, Peptides and Proteins* (Hancock, W. S., ed) pp. 53–65, CRC Press, Inc., Boca Raton, FL

16. Antal, J., Pal, G., Asboth, B., Buzas, Z., Patthy, A., and Graf, L. (2001) *Anal. Biochem.* **288,** 156–167

17. Cornish-Bowden, A. (1995) *Fundamentals of Enzyme Kinetics*, pp. 105–111, Portland Press, London

18. Laemmli, U. K. (1970) *Nature* **227,** 680–685

19. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326

20. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol.* **277,** 319–343

21. Jones, T. A., and Kjeldgaard, M. (1997) *Methods Enzymol.* **277,** 173–208

22. Engh, R., and Huber, R. (1991) *Acta Crystallogr. A* **47,** 392–400

23. Glanville, R. W., Breitkreutz, D., Meitinger, M., and Fietzek, P. P. (1983) *Biochem. J.* **215,** 183–189

24. Miller, E. J. (1984) in *Extracellular Matrix Biochemistry* (Piez, K. A., and Reddi, A. H., eds) pp. 41–81, Elsevier Science Publishing, New York

25. Dauter, Z., Li, M., and Wlodawer, A. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 239–249

26. Bode, W., Papamokos, E., and Musil, D. (1987) *Eur. J. Biochem.* **166,** 673–692

27. Perona, J. J., Tsu, C. A., Craik, C. S., and Fletterick, R. J. (1997) *Biochemistry* **36,** 5381–5392

28. Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 178–185

29. Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer-Moore, J., Xuong, N. H., Hamlin, R., Rutter, W. J., and Craik, C. S. (1987) *Science* **237,** 905–909

30. Wlodawer, A., Li, M., Dauter, Z., Gustchina, A., Uchida, K., Oyama, H., Dunn, B. M., and Oda, K. (2001) *Nat. Struct. Biol.* **8,** 442–446

31. Oda, K., Ogasawara, S., Oyama, H., and Dunn, B. M. (2000) *J. Biochem.* (*Tokyo*) **128,** 499–507

32. Li, J., Brick, P., O'Hare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., and Cawston, T. E. (1995) *Structure* **3,** 541–549

33. Matsushita, O., Jung, C. M., Minami, J., Katayama, S., Nishi, N., and Okabe, A. (1998) *J. Biol. Chem.* **273,** 3643–3648

34. Cohen, G. E. (1997) *J. Appl. Crystallogr.* **30,** 1160–1161

35. Nicholls, A. (1992) *GRASP: Graphical Representation and Analysis of Surface Properties*, Columbia University, New York, NY