Role of the S′ Subsites in Serine Protease Catalysis. Active-Site Mapping of Rat Chymotrypsin, Rat Trypsin, α-Lytic Protease, and Cercarial Protease from *Schistosoma mansoni*†

Volker Schellenberger,‡,§ Christoph W. Turck,‖ and William J. Rutter*,‡

Department of Biochemistry and Biophysics, Hormone Research Institute (Box 0534), and Department of Medicine, University of California, San Francisco, California 94143, and Howard Hughes Medical Institute, Box 0724, San Francisco, California 94143

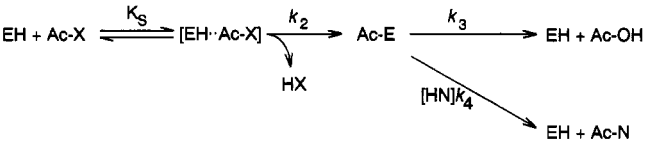
Received November 9, 1993; Revised Manuscript Received February 4, 1994®

ABSTRACT: The S′ subsite specificity of four homologous serine proteases, rat chymotrypsin, rat trypsin, α-lytic protease, and cercarial protease from *Schistosoma mansoni*, was studied by measuring acyl-transfer reactions to 100 pentapeptide nucleophiles. Peptides of the general structures H-Xaa-Ala-Ala-Ala-Ala-NH_2, H-Ala-Xaa-Ala-Ala-Ala-NH_2, and H-Ala-Ala-Xaa-Ala-Ala-NH_2 were synthesized, where Xaa is D-Ala, Cit, and all natural amino acids except Cys. The variable residues of these nucleophiles occupy the P'_1, P'_2, and P'_3 positions in acyl-transfer reactions. The P'_1 and P'_2 residues were found to influence the efficiency of the nucleophiles by more than 2 orders of magnitude, whereas the S'_3 subsite shows a lower specificity in all four enzymes. We synthesized consensus peptides of the general structure H-aa_1-aa_2-aa_3-Ala-Ala-NH_2, in which two or three positions were occupied by amino acids that showed the highest specificity in the first series of nucleophiles. Peptides with optimal amino acid residues in the P'_2 and P'_3 positions show a very high efficiency in chymotrypsin- and trypsin-catalyzed reactions. Otherwise, large specific side chains in the P'_1 and P'_3 positions of the nucleophiles show less than additive binding contributions due to steric hindrance. Comparison of chymotrypsin-catalyzed acyl-transfer reactions to nucleophiles of the structures H-Xaa-Leu-Arg-Ala-Ala-NH_2 and H-Xaa-Ala-Ala-Ala-Ala-NH_2 reveals a significantly different P'_1 specificity for both series which confirms steric hindrance between large P'_1 and P'_3 residues. α-Lytic protease and cercarial protease exhibit very little specificity toward the P'_3 side chain, but the efficiency of nucleophiles with D-Ala or Gly in P'_3 position was significantly reduced compared to that of nucleophiles with other P'_3 side chains. These results suggest significant enzyme–nucleophile contacts beyond the S'_3 subsite. Acyl transfer to alanine peptides of varying length revealed the existence of four S′ subsites for α-lytic protease and at least five S′ subsites for cercarial protease.

The active sites of endoproteinases can be divided into subsites which bind residues of a peptide substrate on either side of the scissile bond [subsite nomenclature of Schechter and Berger (1967)]. The number of subsites of a protease can be deduced from the specificity of the enzyme for substrate residues in the various positions. For many proteases the S subsites, which bind residues N-terminal of the cleavage site, have been well characterized using series of chromogenic or fluorogenic substrates and various synthetic inhibitors. Much less information is available about the S′ subsites of most enzymes, which bind the substrate on the C-terminal side of the scissile bond.

Serine proteases are able to catalyze acyl-transfer reactions according to Scheme 1. Free enzyme EH and acyl donor Ac-X form the Michaelis complex [EH··Ac-X] and subsequently the acyl enzyme Ac-E. The acyl enzyme can be attacked by water, leading to the hydrolysis product Ac-OH, or alternatively by an added peptide nucleophile NH. In the latter case a peptide product, Ac-N, is formed. Acyl transfer to a peptide nucleophile is the reversal of the acylation of the enzyme by a peptide substrate. Consequently, both processes show the same specificity (Fersht et al., 1973). Recently, we showed that *mixtures* of competing nucleophiles can be employed in a single acyl-transfer reaction. This allows simultaneous calculation of relative rate constants and a very efficient analysis of the S′ specificity of serine proteases (Schellenberger et al., 1993b).

Scheme 1



Chymotrypsin and trypsin are among the best studied enzymes (Fersht, 1985). They show a high level of homology on the level of primary and tertiary structure. The role of the S subsites of chymotrypsin and trypsin in substrate binding and catalysis has been extensively investigated [reviewed in Schellenberger et al. (1991c)]. Although the importance of S′–P′ interactions in chymotrypsin-catalyzed hydrolysis has been known for a long time (Fersht et al., 1973), little information is available about the S′ subsite specificity of chymotrypsin and trypsin. In the studies published so far, amino acid or dipeptide amides have been used to investigate the S'_1 specificity of chymotrypsin and trypsin (Fersht et al.,

† This work was supported by NIH Grants DK21344 and NCRR RR01614 and a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft to V.S.

‡ Department of Biochemistry and Biophysics and Hormone Research Institute, University of California.

§ Present address: Genencor International, Inc., 180 Kimball Way, South San Francisco, CA 94080.

‖ Department of Medicine, University of California, and Howard Hughes Medical Institute.

® Abstract published in *Advance ACS Abstracts,* March 15, 1994.

0006-2960/94/0433-4251$04.50/0

1973; Hanisch et al., 1987; Petkov & Stoineva, 1984; Schellenberger et al., 1990). However, in a recent analysis we observed that chymotrypsin shows a significantly altered S'_1 specificity with *longer* peptide nucleophiles, which can interact with the enzyme beyond the S'_1 subsite (Schellenberger et al., 1993b).

In order to further elucidate the role of the S'_2 and S'_3 subsites in chymotrypsin and trypsin catalysis, we investigated a large series of pentapeptide nucleophiles with various residues in the P'_2 and P'_3 positions. The resulting data give a detailed picture of the S' specificity of both enzymes. Furthermore, the analysis of consensus peptides revealed the interplay of the S'_1 to S'_3 subsites. We also determined the S'_1–S'_3 subsite specificities of two homologous serine proteases, α-lytic protease and cercarial protease from *Schistosoma mansoni*. The resulting specificity profiles suggest the presence of S'–P' contacts beyond the S'_3 subsite in these enzymes. To get more information about the number of S' subsites in both enzymes, we measured acyl-transfer reactions with series of nucleophiles, H-Ala$_n$-NH_2 and H-Ala$_n$-OH with n between 1 and 5. The resulting data confirm the presence of an S'_4 subsite in α-lytic protease and of S'_4 and S'_5 subsites in cercarial protease.

MATERIALS AND METHODS

Rat trypsin and rat chymotrypsin were expressed in yeast, purified, and activated as described previously (Hedstrom et al., 1992; Schellenberger et al., 1993b). α-Lytic protease was a gift of D. A. Agard (UCSF), and cercarial protease was a gift of J. H. McKerrow (UCSF). Ac-Phe-OEt,[1] Bz-Arg-OEt, alanine, and peptides of the structure H-Ala$_n$-OH with n between 2 and 5 were purchased from Sigma Chemical Co. H-Ala-Ala-NH_2 and Ac-Ala-Ala-Ala-OMe were purchased from Bachem (Torrance, CA).

Preparation of Peptide Mixtures. The peptides H-Ala-Ala-Ala-NH_2, H-Ala-Ala-Ala-Ala-NH_2, and H-Ala-Ala-Ala-Ala-Ala-NH_2 were synthesized with an ABI Model 431 peptide synthesizer (Applied Biosystems, Foster City, CA). The resin was aliquoted, and subsequent amino acids were added with an Advanced Chemtech MPS 350 multiple peptide synthesizer (Louisville, KY) using double coupling and a 12-fold excess of the appropriate Fmoc amino acid (Bachem, Torrance, CA). Deprotection, characterization by HPLC, and mass spectrometry were performed as described previously (Schellenberger et al., 1993b). Peptides were dissolved in water at 100 μM concentrations (individual peptides) and stored at −20 °C. H-Ala-Ala-Ile-Ala-Ala-NH_2 was used as a reference nucleophile because it could be separated from all other peptides by HPLC. In our initial investigations of the S'_1 subsite specificity H-Ser-Ala-Ala-Ala-Ala-NH_2 served as a reference. In this stury we repeated several of these experiments with H-Ala-Ala-Ile-Ala-Ala-NH_2 as a reference, but no significant change in the calculated rate constants was observed.

Acyl Transfer. Enzyme-catalyzed acyl transfer was performed essentially as described previously (Schellenberger et al., 1993b). Approximately 40 μM peptide nucleophiles, 10 mM acyl donor, and buffer (0.1 M borate buffer, pH 9.0, and 0.2 M KCl) were mixed and equilibrated to 25 °C, and the reaction was started by the addition of enzyme. The acyl donor esters were added from 400 mM stock solutions, leading to the presence of 2.5% acetonitrile in reactions with chymotrypsin and cercarial protease and of 2.5% DMSO in the reactions with α-lytic protease. Enzyme concentrations were chosen which resulted in 30–80% conversion of the acyl donor ester during 1 h of reaction time. Typically, 6–10 samples were taken from the reaction mixtures after different times. The nucleophiles were derivatized with dansyl chloride and quantified by HPLC. For each reaction a set of peak areas was obtained, one value for each nucleophile and each sample. From these data relative rate constants were calculated using a computer program that takes into account the pairwise dependence of all nucleophile concentrations (Schellenberger et al., 1993a). All experiments were performed in triplicate. Many nucleophiles were included in more than one mixture, which further improved the statistical quality of the data. All nucleophiles were completely soluble under the assay conditions. Relative rate constants were identical within the limit of error when the nucleophile concentrations were varied in the range of 20–80 μM, confirming the absence of significant interactions between the nucleophiles in solution.

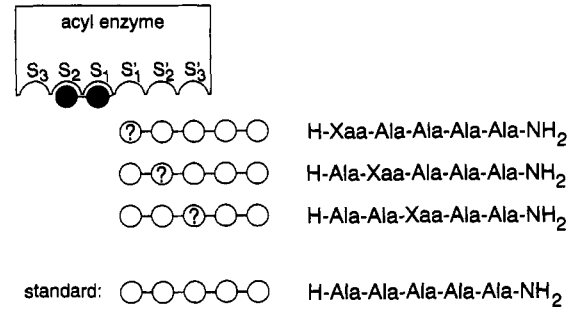


FIGURE 1: Cartoon explaining the structure of the nucleophiles used to study the S'_1–S'_3 specificity. The enzyme is shown with its S_2 and S_1 subsites occupied by the acyl part of the acyl donor ester (filled circles). The nucleophiles are shown in alignment with the S' subsites they are binding to during acyl-transfer. Question marks indicate the variable positions of the nucleophiles.

RESULTS

Mapping the S'_1–S'_3 Subsites. A series of peptides of the general structure H-Xaa-Ala-Ala-Ala-Ala-NH_2, H-Ala-Xaa-Ala-Ala-Ala-NH_2, and H-Ala-Ala-Xaa-Ala-Ala-NH_2, where Xaa is D-Ala, Cit, and all natural amino acids except Cys, were prepared by multiple peptide synthesis. These peptides were used as nucleophiles in acyl-transfer reactions with Ac-Phe-OEt (chymotrypsin and cercarial protease), Bz-Arg-OEt (trypsin), and Ac-Ala-Ala-Ala-OMe (α-lytic protease) as acyl donor esters. Under these conditions the variable residues of the nucleophiles occupied the P'_1, P'_2, or P'_3 position, respectively, as illustrated in Figure 1. Acyl-transfer reactions were measured by our recently reported method (Schellenberger et al., 1993b). The decrease of individual nucleophile concentrations during acyl transfer to mixtures of nucleophiles was analyzed by HPLC. From the resulting data we calculated second-order rate constants k_4 (see Scheme 1) for each nucleophile in the mixture using multiple substrate kinetics (Schellenberger et al., 1993a). Eight mixtures were prepared from the series of 61 nucleophiles such that each nucleophile was included in at least one mixture and all nucleophiles could be quantitated by HPLC. H-Ala-Ala-Ile-Ala-Ala-NH_2 was included in all peptide mixtures as a reference compound.

Second-order rate constants could be obtained with high precision for most of the peptides. A small number of nucleophiles showed a very low acyl-transfer efficiency, so that the concentration of these peptides did not significantly

[1] Abbreviations: Ac, acetyl; Boc, *tert*-butoxycarbonyl; BPTI, bovine pancreatic trypsin inhibitor; Bz, benzoyl; Cit, citrulline; -OEt, ethyl ester; -OMe, methyl ester; OMTKY, turkey ovomucoid third domain.

Table 1: Second-Order Rate Constants for Acyl Transfer to Nucleophiles of the Structure H-Xaa-Ala-Ala-Ala-Ala-NH_2, H-Ala-Xaa-Ala-Ala-Ala-NH_2, and H-Ala-Ala-Xaa-Ala-Ala-NH_2[a]

nucleophile			k_{rel}			
P'_1	P'_2	P'_3	chymotrypsin	trypsin	α-lytic protease	cercarial protease
D-Ala	Ala	Ala	<0.05	<0.07	<0.07	<0.05
Pro	Ala	Ala	<0.05	<0.07	<0.07	<0.05
Gly	Ala	Ala	0.27 ± 0.02	0.25 ± 0.04	0.31 ± 0.01	0.23 ± 0.02
Val	Ala	Ala	1.00 ± 0.01	0.93 ± 0.16	1.68 ± 0.10	0.65 ± 0.01
Ile	Ala	Ala	1.04 ± 0.03	5.57 ± 0.48	1.63 ± 0.06	0.61 ± 0.03
Leu	Ala	Ala	1.07 ± 0.01	3.70 ± 0.25	2.14 ± 0.09	1.64 ± 0.06
Met	Ala	Ala	1.42 ± 0.04	9.29 ± 0.91	2.55 ± 0.09	2.31 ± 0.08
Phe	Ala	Ala	0.49 ± 0.02	2.02 ± 0.21	1.18 ± 0.05	0.63 ± 0.03
Tyr	Ala	Ala	0.40 ± 0.04	1.59 ± 0.26	1.11 ± 0.03	0.38 ± 0.04
Trp	Ala	Ala	0.40 ± 0.02	2.96 ± 0.51	1.64 ± 0.02	0.58 ± 0.02
His	Ala	Ala	1.03 ± 0.04	1.15 ± 0.10	0.48 ± 0.05	0.50 ± 0.03
Ser	Ala	Ala	1.35 ± 0.07	0.73 ± 0.09	1.40 ± 0.03	1.11 ± 0.03
Thr	Ala	Ala	1.09 ± 0.03	0.48 ± 0.07	0.34 ± 0.06	0.62 ± 0.02
Asn	Ala	Ala	0.29 ± 0.03	0.51 ± 0.07	0.23 ± 0.11	0.24 ± 0.16
Gln	Ala	Ala	0.41 ± 0.03	1.09 ± 0.16	0.32 ± 0.07	0.70 ± 0.02
Asp	Ala	Ala	<0.05	<0.07	<0.07	0.13 ± 0.09
Glu	Ala	Ala	<0.05	0.32 ± 0.08	<0.07	0.18 ● 0.01
Lys	Ala	Ala	3.99 ± 0.47	1.29 ± 0.34	1.61 ± 0.05	0.34 ± 0.05
Arg	Ala	Ala	10.68 ± 0.36	1.67 ± 0.18	1.52 ± 0.15	0.28 ± 0.03
Cit	Ala	Ala	1.64 ± 0.03	3.90 ± 0.28	1.16 ± 0.03	0.75 ± 0.03
Ala	D-Ala	Ala	<0.05	<0.07	<0.07	<0.05
Ala	Pro	Ala	<0.05	<0.07	<0.07	<0.05
Ala	Gly	Ala	<0.05	0.19 ± 0.09	0.27 ± 0.04	0.58 ± 0.03
Ala	Val	Ala	2.77 ± 0.03	0.66 ± 0.03	1.19 ± 0.04	0.09 ± 0.03
Ala	Ile	Ala	4.82 ± 0.08	0.46 ± 0.08	2.25 ± 0.11	<0.05
Ala	Leu	Ala	13.76 ± 1.37	0.27 ● 0.11	15.61 ± 2.58	0.19 ± 0.03
Ala	Met	Ala	7.79 ± 0.22	0.54 ± 0.08	5.81 ± 0.76	0.11 ± 0.03
Ala	Phe	Ala	4.47 ± 0.07	1.06 ± 0.07	1.49 ± 0.09	<0.05
Ala	Tyr	Ala	1.93 ± 0.16	0.52 ± 0.20	1.61 ± 0.08	0.08 ± 0.02
Ala	Trp	Ala	2.11 ± 0.02	1.78 ± 0.02	1.53 ± 0.05	0.23 ± 0.02
Ala	His	Ala	1.48 ± 0.04	0.19 ± 0.12	0.22 ± 0.03	<0.05
Ala	Ser	Ala	0.49 ± 0.03	0.46 ± 0.07	0.31 ± 0.04	0.36 ± 0.03
Ala	Thr	Ala	0.60 ± 0.04	0.21 ± 0.07	0.33 ± 0.06	0.50 ± 0.04
Ala	Asn	Ala	0.86 ± 0.06	0.19 ± 0.09	0.09 ± 0.04	0.09 ± 0.03
Ala	Gln	Ala	<0.05	0.19 ± 0.08	0.43 ± 0.02	<0.05
Ala	Asp	Ala	<0.05	<0.07	0.09 ± 0.04	<0.05
Ala	Glu	Ala	<0.05	0.12 ± 0.14	0.11 ± 0.07	<0.05
Ala	Lys	Ala	<0.05	2.94 ± 0.06	1.03 ± 0.04	<0.05
Ala	Arg	Ala	0.58 ± 0.07	nd[b]	0.51 ± 0.01	<0.05
Ala	Cit	Ala	1.84 ± 0.04	0.33 ± 0.17	1.95 ± 0.14	<0.05
Ala	Ala	D-Ala	1.13 ± 0.02	1.31 ± 0.05	0.36 ± 0.06	<0.05
Ala	Ala	Pro	0.52 ± 0.08	0.88 ± 0.03	1.42 ± 0.05	1.96 ± 0.11
Ala	Ala	Gly	1.05 ± 0.12	1.01 ± 0.02	0.33 ± 0.06	0.12 ± 0.08
Ala	Ala	Val	1.74 ± 0.38	1.60 ± 0.24	1.28 ± 0.03	0.73 ± 0.15
Ala	Ala	Ile	2.45 ± 0.13	1.89 ± 0.03	1.33 ± 0.02	1.09 ± 0.02
Ala	Ala	Leu	1.91 ± 0.05	1.81 ± 0.03	1.46 ± 0.04	1.43 ± 0.03
Ala	Ala	Met	2.16 ± 0.10	1.54 ± 0.03	1.28 ± 0.05	1.22 ± 0.03
Ala	Ala	Phe	3.00 ± 0.06	1.83 ± 0.11	1.18 ± 0.03	0.93 ± 0.03
Ala	Ala	Tyr	2.58 ± 0.09	1.56 ± 0.10	1.20 ± 0.03	1.05 ± 0.05
Ala	Ala	Trp	4.44 ± 0.17	3.35 ± 0.13	1.39 ± 0.06	1.47 ± 0.05
Ala	Ala	His	2.50 ± 0.14	1.66 ± 0.07	1.22 ± 0.06	0.54 ± 0.03
Ala	Ala	Ser	1.77 ± 0.04	1.01 ± 0.05	1.13 ± 0.04	0.55 ± 0.05
Ala	Ala	Thr	2.45 ± 0.05	1.25 ± 0.08	1.06 ± 0.01	0.77 ± 0.01
Ala	Ala	Asn	1.60 ± 0.09	1.05 ± 0.03	0.82 ± 0.04	0.51 ± 0.03
Ala	Ala	Gln	1.97 ± 0.01	1.15 ± 0.10	1.12 ± 0.03	0.81 ± 0.01
Ala	Ala	Asp	0.63 ± 0.09	1.13 ± 0.04	0.84 ± 0.03	0.74 ± 0.02
Ala	Ala	Glu	0.60 ± 0.12	0.62 ± 0.06	0.85 ± 0.05	1.13 ± 0.01
Ala	Ala	Lys	3.43 ± 0.11	1.33 ± 0.04	1.33 ± 0.05	0.63 ± 0.03
Ala	Ala	Arg	15.31 ± 1.90	nd[b]	1.60 ± 0.10	0.86 ± 0.01
Ala	Ala	Cit	1.85 ± 0.07	1.38 ± 0.04	1.38 ± 0.04	1.13 ± 0.01

[a] Acyl donors were Bz-Arg-OEt for trypsin, Ac-Ala-Ala-Ala-OMe for α-lytic protease, and Ac-Phe-OEt for chymotrypsin and cercarial protease. Rate constants are relative to the constant of H-Ala-Ala-Ala-Ala-Ala-NH_2. Data are mean ± SD calculcated from triplicate experiments. [b] These nucleophiles were not analyzed, as they were rapidly hydrolyzed by trypsin.

decrease during acyl transfer. H-Ala-Arg-Ala-Ala-Ala-NH_2 and H-Ala-Ala-Arg-Ala-Ala-NH_2 could not be studied with trypsin because both peptides were rapidly hydrolyzed by the enzyme. For all other nucleophiles less than 10% hydrolysis occurred under reaction conditions in the absence of acyl donor ester. Thus, the observed decreases in nucleophile concentrations were due to acyl transfer. Except for mixtures with Lys- or Tyr-containing peptides which give dansylated acyl-transfer products, no new peaks were observed during acyl-transfer reactions.

Second-order rate constants for the series of nucleophiles are summarized in Table 1. It is obvious that acyl transfer provides a detailed picture of the S′ subsite specificity of all four enzymes. The data reveal that no changes to the

Table 2: Second-Order Rate Constants for Acyl Transfer to Nucleophiles of the General Structure H-aa_1-aa_2-aa_3-Ala-Ala-NH_2[a]

chymotrypsin		trypsin	
nucleophile	k_{rel}	nucleophile	k_{rel}
H-Arg-Leu-Ala-Ala-Ala-NH_2	35 ± 2	H-Met-Lys-Ala-Ala-Ala-NH_2	21.8 ± 2.7
H-Arg-Ala-Arg-Ala-Ala-NH_2	33 ± 6	H-Met-Ala-Trp-Ala-Ala-NH_2	11.1 ± 1.1
H-Ala-Leu-Arg-Ala-Ala-NH_2	205 ± 6	H-Ala-Lys-Trp-Ala-Ala-NH_2	11.6 ± 1.4
H-Arg-Leu-Arg-Ala-Ala-NH_2	57 ● 9	H-Met-Lys-Trp-Ala-Ala-NH_2	27.2 ± 3.2

[a] Rate constants are given relative to the constant of H-Ala-Ala-Ala-Ala-Ala-NH_2 and are mean ± SD calculated from triplicate experiments.

Table 3: Second-Order Rate Constants for Chymotrypsin-Catalyzed Acyl Transfer to Nucleophiles of the Structure H-Xaa-Leu-Arg-Ala-Ala-NH_2[a]

P'_1 residue	k_{rel}	P'_1 residue	k_{rel}
D-Ala	<0.05	His	16.8 ± 2.3
Pro	0.22 ± 0.05	Ser	291 ± 16
Gly	286 ± 38	Thr	272 ± 54
Ala	205 ± 6	Asn	8.2 ± 1.7
Val	54 ± 7	Gln	8.7 ± 1.5
Ile	34 ± 4	Asp	2.7 ± 0.4
Leu	26 ± 3	Glu	1.4 ± 0.2
Met	46 ± 6	Lys	40 ± 5
Phe	17.6 ± 1.9	Arg	47 ± 4
Tyr	17.0 ± 2.4	Cit	34 ± 4
Trp	12.2 ± 1.4		

[a] Rate constants are relative to the constant of H-Ala-Ala-Ala-Ala-Ala-NH_2. Data are mean ± SD calculcated from triplicate experiments.

nucleophile backbones are tolerated in the P'_1 and P'_2 positions, as nucleophiles with D-Ala or Pro in these positions show no reactivity. The enzymes show different preferences for amino acid side chains in the various positions. The S'_3 subsites of the studied enzymes are less specific than the S'_1 and S'_2 subsites.

Nucleophilic Efficiency of Consensus Peptides H-aa_1-aa_2-aa_3-Ala-Ala-NH_2. In order to investigate the interdependence of the S'_1–S'_3 subsites in chymotrypsin and trypsin, we synthesized consensus peptides in which two or three positions were occupied by the amino acids showing the highest specificity in the first series of peptides. For both enzymes we prepared a nucleophile mixture containing H-Ala-Ala-Ile-Ala-Ala-NH_2 and H-Ala-Ala-Ala-Ala-Ala-NH_2 as reference compounds and all seven possible peptides with one, two, or three specific residues in the P'_1–P'_3 positions. These mixtures were readily resolved by HPLC. The efficiency of acyl transfer of all nucleophiles in these mixtures was determined from the same experiments, thus ensuring identical reaction conditions.

Second-order rate constants for the peptides are compiled in Table 2. The data reveal that the specificity of the enzymes in any position is influenced by the occupancy of the other positions. This is especially obvious for chymotrypsin where H-Arg-Leu-Arg-Ala-Ala-NH_2 shows a lower efficiency than H-Ala-Leu-Arg-Ala-Ala-NH_2.

Chymotrypsin S'_1 Specificity with Nucleophiles of the Structure H-Xaa-Leu-Arg-Ala-Ala-NH_2. The data in Table 2 indicate that the occupancy of the P'_2 and P'_3 positions by Leu and Arg leads to a significant change in the S'_1 specificity of chymotrypsin. In order to further analyze this effect, we synthesized a series of peptides of the structure H-Xaa-Leu-Arg-Ala-Ala-NH_2. Three mixtures were prepared from these peptides, and H-Thr-Leu-Arg-Ala-Ala-NH_2 and H-Ala-Ala-Ala-Ala-Ala-NH_2 were added as reference compounds to all mixtures. All resulting data are shown in Table 3. The dramatic difference in the S'_1 specificity of chymotrypsin as compared to that of trypsin is obvious from the data. Instead of positively charged residues, the S'_1 subsite now shows a preference for amino acids with small hydrophilic side chains.

Efficiency of Nucleophiles of Varying Length. There is a marked difference in the S'_3 specificities of α-lytic protease and cercarial protease compared to the S'_3 specificities of chymotrypsin and trypsin. Both α-lytic protease and cercarial protease show almost no specificity for the *side chain* of the P'_3 residue, but D-Ala and Gly in P'_3 lead to very low efficiencies as nucleophiles in the acyl-transfer reactions. This finding is in marked contrast to the S'_3 specificity of chymotrypsin and trypsin and indicates enzyme–nucleophile interactions beyond the S'_3 subsite. In order to further investigate such distant interactions, we performed acyl-transfer reactions using nucleophiles of the structure H-Ala_n-NH_2 and H-Ala_n-OH with n between 1 and 5. All H-Ala_n-NH_2 peptides showed almost identical retention times under our standard HPLC conditions. To allow analysis by multiple acyl transfer, we prepared five nucleophile mixtures. All these mixtures contained one nucleophile of each alanine peptide series and H-Ala-Ala-Ile-Ala-Ala-NH_2, H-Leu-Ala-Ala-Ala-Ala-NH_2, and H-Tyr-Ala-Ala-Ala-Ala-NH_2 as internal references. Relative rate constants that were obtained for α-lytic protease, cercarial protease, and chymotrypsin are given in Table 4. The data reveal significant differences in the number of Ala residues that are required for efficient acyl transfer by the three enzymes.

DISCUSSION

Our data demonstrate that the acyl-transfer reaction to nucleophile mixtures is a powerful method for the analysis of structure–function relationships in the S′ subsites of serine proteases. The method is able to detect even small differences in the efficiencies of nucleophiles, thus providing information about low-energy enzyme–substrate contacts which are currently not predictable on the basis of the enzyme structure alone.

Structural Basis of the S′ Specificity of Chymotrypsin and Trypsin. Complexes between proteases and peptide inhibitors can provide structural information about enzyme–substrate interactions during catalysis (Bode & Huber, 1991). Comparison of the superimposed structures of complexes between chymotrypsin and turkey ovomucoid third domain (Fujinaga et al., 1987) and between trypsin and bovine pancreatic trypsin inhibitor (Ruhlmann et al., 1973) shows that the P_2–P'_2 residues of both inhibitors are bound with virtually identical backbone conformations to both enzymes. The inhibitor backbones form many close contacts with conserved residues of the enzymes. The most important contact in the S′ subsites is a hydrogen bond between O(Phe-41) and HN(P'_2), which is found in both complexes. This explains why no acyl transfer to H-Ala-Pro-Ala-Ala-Ala-NH_2 was observed with both enzymes, since this peptide is unable to act as a hydrogen bond donor in P'_2.

Crystal structures of the protease–inhibitor complexes also provide some information about the orientation of the P′ side

Table 4: Second-Order Rate Constants for the Acyl Transfer to Ala-Peptides of Varying Length[a]

nucleophile	k_{rel} chymotrypsin	k_{rel} α-lytic protease	k_{rel} cercarial protease
H-Ala-NH_2	0.23 ± 0.01	<0.07	<0.05
H-Ala-Ala-NH_2	0.95 ± 0.14	0.18 ± 0.08	<0.05
H-Ala-Ala-Ala-NH_2	0.85 ± 0.03	0.25 ± 0.07	<0.05
H-Ala-Ala-Ala-Ala-NH_2	1.08 ± 0.08	1.08 ± 0.04	0.27 ± 0.04
H-Ala-Ala-Ala-Ala-Ala-NH_2	1.00	1.00	1.00
H-Ala-OH	<0.05	<0.07	<0.05
H-Ala-Ala-OH	0.24 ± 0.04	0.18 ± 0.09	<0.05
H-Ala-Ala-Ala-OH	0.62 ± 0.10	0.13 ± 0.05	<0.05
H-Ala-Ala-Ala-Ala-OH	0.65 ± 0.02	1.16 ± 0.04	0.28 ± 0.03
H-Ala-Ala-Ala-Ala-Ala-OH	0.81 ± 0.02	1.03 ± 0.04	0.95 ± 0.01

[a] Rate constants are relative to the constant of H-Ala-Ala-Ala-Ala-Ala-NH_2. Data are mean ± SD calculcated from triplicate experiments.

chains. Due to the extended backbone conformation, the P'_1 and P'_3 side chains point in one direction whereas the P'_2 side chain points in the opposite direction. Large amino acid residues in P'_1 and P'_3 can probably form contacts with the same region on the enzyme surface; this explains the similarity in the S'_1 and S'_3 specificities of both chymotrypsin and trypsin. Most likely, large side chains in P'_1 and P'_3 compete for contacts with the same region of the enzyme surface. This explains the dramatically changed S'_1 specificity of chymotrypsin that was observed with nucleophiles of the structure H-Xaa-Leu-Arg-Ala-Ala-NH_2 (see Table 3). In these nucleophiles the P'_3 Arg seems to occupy a part of the S'_1 subsite leading to a specificity for small hydrophilic residues in P'_1. The interference of large P'_1 and P'_3 side chains is much stronger for chymotrypsin where two positively charged residues in the consensus peptides have to bind in close proximity. Arg residues in P'_1 and P'_3 probably interact with Asp-64 and Asp-35 of chymotrypsin.

The crystal structures suggest that P'_2 side chains bind to a region on the enzyme surface that is lined by His-40 and Tyr-151 in trypsin and by His-40 and Leu-143 in chymotrypsin. As the P'_2 and P'_3 side chains point in opposite directions, residues with large and specific side chains can be accommodated in both positions. In fact, positive cooperativity is observed for specific P'_2 and P'_3 residues which is probably entropy driven. Introduction of a specific side chain in the P'_2 or P'_3 position leads to an increased number of enzyme nucleophile contacts and consequently to a loss of conformational freedom of the bound nucleophile. The accompanying entropy loss has to be paid for by the P'_2–S'_2 or P'_3–S'_3 side-chain contacts. If both the P'_2 and P'_3 positions of a nucleophile are occupied by specific amino acid residues, then the entropic costs of determining the nucleophile conformation have to be paid only once.

A model which explains the interdependence of the S′ subsite interactions is given in Figure 2. It shows a peptide substrate in an extended conformation with the amino acid side chains represented by dotted ovals. The C-terminal part of the substrate can be held in place by either P'_2–S'_2 or P'_3–S'_3 side-chain interactions. If the S'_3 subsite is occupied by a large residue as in H-Xaa-Leu-Arg-Ala-Ala-NH_2, then the space available for the P'_1 side chain is considerably reduced and consequently small amino acids are preferred.

Structure of the S′ Subsites of α-Lytic Protease. The importance of the S′ subsites in α-lytic protease-catalyzed hydrolysis of peptides is known from studies of substrates of the structure Ac-Pro-Ala-Pro-Ala-X, where X represents various amino acid- and peptide-derived leaving groups (Bauer et al., 1981). k_{cat}/K_M values for these substrates increased 10-fold for X = Ala-NH_2 and 44-fold for X = Ala-Ala-NH_2 as compared with that of X = NH_2. Further information

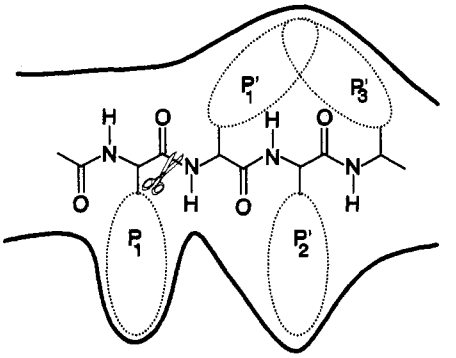


FIGURE 2: Model of the S_1–S'_3 binding sites of chymotrypsin and trypsin.

about substrate binding in the S′ subsites can be obtained from the recently determined crystal structure of a complex between α-lytic protease and an analogue of Boc-Ala-Ala-Pro-Val-Ala-Ala, where Val is replaced with an analogous phosphonate ester and the subsequent Ala is replaced with lactate (Bone et al., 1991). Two hydrogen bonds were observed between the enzyme and the P'_2 residue of the inhibitor. One hydrogen bond is formed between the P'_2 amide proton and the carbonyl oxygen of Leu-41 of the enzyme. This hydrogen bond seems to be conserved in all proteases of the trypsin family as was also observed in the trypsin–BPTI complex (Bone et al., 1991) and in a chymotrypsin–OMTKY complex (Fujinaga et al., 1987). The other hydrogen bond is observed between the carbonyl oxygen of the P'_2 residue and the amide of Leu-41 in α-lytic protease. This bond cannot be formed by trypsin or chymotrypsin due to large differences in the orientation of residue 41 in these enzymes.

The different hydrogen bond interactions in the S′ subsites provide an explanation for some of the results given in Table 4. In chymotrypsin only one hydrogen bond is observed in the S′ subsites between NH(P'_2) and O(Phe-41). H-Ala-NH_2 is able to satisfy this bond, and consequently, chymotrypsin catalyzes acyl transfer to this nucleophile. Nucleophiles which cannot donate this hydrogen bond, like H-Ala-OH or H-Ala-OMe, are very inefficient in chymotrypsin-catalyzed acyl-transfer reactions (Schellenberger & Jakubke, 1986). α-Lytic protease requires the formation of a second hydrogen bond between O(P'_2) and NH(Leu-41) of the enzyme which cannot be satisfied by H-Ala-NH_2, and as a result no acyl transfer to this nucleophile is observed.

The crystal structure also reveals some information about possible interactions of P'_1 and P'_2 side chains with α-lytic protease. The P'_1 side chain of the inhibitor points into a channel on the enzyme surface which is lined by His-57 and Ser-41. The P'_2 side chain points into a cleft that is formed by the side chains of Leu-41 and Thr-142. The hydrophobic

nature of this part of the enzyme surface readily explains the preference of α-lytic protease for hydrophobic P'_2 side chains.

Evidence for Contacts beyond S'_3 in α-Lytic Protease and Cercarial Protease. Table 1 reveals very low specificities of α-lytic protease and cercarial protease for the P'_3 residue. This suggests that few contacts between the P'_3 side chain and the enzyme occur during catalysis. If we assume that no contacts are formed between the enzyme and the nucleophile beyond P'_3, then the nucleophile should be able to rotate in P'_3, and one should not observe stereospecificity.

The low efficiency of H-Ala-Ala-Gly-Ala-Ala-NH_2 in acyl transfer observed for both enzymes also indicates contacts beyond S'_3. H-Ala-Ala-Gly-Ala-Ala-NH_2 is structurally more flexible than nucleophiles with other P'_3 residues. If this nucleophile is bound to the enzyme on both sides of the P'_3 residue during catalysis, then a peptide with P'_3 Gly will lose more entropy than nucleophiles with other P'_3 residues.

Further evidence for extended nucleophile-binding sites in α-lytic protease and cercarial protease is provided by the data shown in Table 4. For chymotrypsin, all peptide amides with two or more Ala residues show nearly identical efficiencies. This enzyme seems to form productive contacts only with the P'_1 and P'_2 residues of the nucleophiles. α-Lytic protease reaches its maximum catalytic efficiency only if the nucleophile can bind to the S'_1–S'_4 subsites. The data in Table 4 indicate an even larger number of S′ subsites for cercarial protease. This enzyme needs at least four Ala residues to exhibit some acyl-transfer efficiency, but a substantially higher rate was observed for H-Ala-Ala-Ala-Ala-Ala-NH_2, which can satisfy contacts with the S'_1–S'_5 subsites.

The studies of acyl-transfer reactions with nucleophiles of the structure H-Ala_n-OH confirm the large number of S′ subsites for α-lytic protease and cercarial protease. The reduced efficiency of these nucleophiles in chymotrypsin-catalyzed acyl transfer is caused by electrostatic repulsion between the negatively charged C-termini of these nucleophiles and negative charges in the S′ subsites of the enzyme (Schellenberger et al., 1991b). It has been shown that nucleophiles of the structure H-Gly_n-OH with n between 2 and 5 show essentially identical efficiencies in chymotrypsin-catalyzed acyl-transfer reactions at high salt concentrations when electrostatic interactions are effectively shielded (Schellenberger et al., 1991a).

Currently there is no information available about the residues which form the extended S′ subsites of α-lytic protease and cercarial protease. Crystal structures of serine protease–inhibitor complexes provide no information about subsites beyond S'_3. The relatively low interaction energies of contacts in these subsites would make computer modeling very speculative.

High Conformational Freedom of Gly-Containing Nucleophiles. The data in Table 1 reveal that in most reactions the efficiency of the Gly-containing peptides is about 3–4-fold reduced as compared to that of H-Ala-Ala-Ala-Ala-Ala-NH_2. This reflects the higher structural flexibility of Gly peptides in solution. If a peptide nucleophile is bound to the enzyme on both sides of an amino acid residue, then a peptide with Gly in that position will lose more entropy than a nucleophile with a different amino acid. Consequently, H-Ala-Ala-Gly-Ala-Ala-NH_2 and H-Ala-Ala-Ala-Ala-Ala-NH_2 show identical reactivities with chymotrypsin and trypsin, which lack nucleophile contact beyond S'_3, but both nucleophiles differ in their reactivity with α-lytic protease and cercarial protease, which bind the nucleophile beyond P'_3.

Comparison of the data in Table 1 shows an unusual high reactivity of H-Ala-Gly-Ala-Ala-Ala-NH_2 in cercarial protease-catalyzed reactions. In fact, Gly is the most specific residue after Ala in P'_2 for this enzyme. If we consider that the reactivity of H-Ala-Gly-Ala-Ala-Ala-NH_2 is reduced due to its high mobility in solution, then it becomes obvious that the S'_2 subsite of cercarial protease prefers Gly over all other amino acid residues.

S′ Specificity of Cercarial Protease. Cercarial protease is able to cleave protein substrates like laminin and fibronectin. It is a type-specific collagenase with activity against basement membrane collagens IV and VIII but not interstitial collagens (McKerrow et al., 1985). There is no information available about the exact cleavage sites of the enzyme in collagen. In order to fulfill its function, cercarial protease has to degrade various protein substrates of predominantly fibrillar structure. In general such proteins are very protease-resistant due to their coiled coil conformation. We may speculate that the extended binding site of cercarial protease, which allows the enzyme to interact with a long region of its substrate, provides the additional binding energy necessary for the cleavage of protease-resistant substrates.

The pronounced specificity of the P'_2 subsite of cercarial protease for amino acids with small side chains may also be rationalized by considering the extremely high number of Gly residues in collagen.

Our data, which provide detailed information about the S'_1–S'_3 specificity of cercarial protease, should help in the refinement of three-dimensional models of the enzyme and in the design of more potent inhibitors. Peptide analogues based on fluoromethyl ketones (Imperiali & Abeles, 1987) or phosphonate esters (Sampson & Bartlett, 1991) are capable of interacting with both the S and S′ subsites of serine proteases and should allow a more efficient and specific inhibition of this enzyme.

ACKNOWLEDGMENT

Rat trypsin was a gift of Lizbeth Hedstrom (Brandeis University). α-Lytic protease and cercarial protease were kindly provided by D. A. Agard and J. H. McKerrow (UCSF). Mass spectra of the peptide nucleophiles were provided by the UCSF Mass Spectrometry Facility (A. L. Burlingame, Director). We wish to thank Jeffrey Johnson (UCSF) for reading and correcting the manuscript.

REFERENCES

Bauer, C.-A., Brayer, G., Sielecki, A., & James, M. (1981) *Eur. J. Biochem. 120*, 289–294.

Bode, W., & Huber, R. (1991) *Curr. Opin. Struct. Biol. 1*, 45–52.

Bone, R., Sampson, N., Bartlett, P., & Agard, D. (1991) *Biochemistry 30*, 2263–2272.

Fersht, A. (1985) *Enzyme structure and mechanism*, W. H. Freeman and Co., New York.

Fersht, A. R., Blow, D. M., & Fastrez, J. (1973) *Biochemistry 12*, 2035–2041.

Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M., Jr., & James, M. N. G. (1987) *J. Mol. Biol. 195*, 397–418.

Hanisch, U.-K., Könnecke, A., Schellenberger, V., & Jakubke, H.-D. (1987) *Biocatalysis 1*, 129–136.

Hedstrom, L., Szilagyi, L., & Rutter, W. J. (1992) *Science 255*, 1249–1253.

Imperiali, B., & Abeles, R. H. (1987) *Biochemistry 26*, 4474–4477.

McKerrow, J., Jones, P., Sage, H., & Pino-Heiss, S. (1985) *Biochem. J. 231*, 47–51.

Petkov, D. D., & Stoineva, I. (1984) *Biochem. Biophys. Res. Commun. 118*, 317–323.

Ruhlmann, A., Kukla, D., Schwager, P., Bartels, K., & Huber, R. (1973) *J. Mol. Biol. 77*, 417–436.
Sampson, N., & Bartlett, P. (1991) *Biochemistry 30*, 2255–2263.
Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.
Schellenberger, V., & Jakubke, H.-D. (1986) *Biochim. Biophys. Acta 869*, 54–60.
Schellenberger, V., Schellenberger, U., Mitin, Y. V., & Jakubke, H.-D. (1990) *Eur. J. Biochem. 187*, 163–167.
Schellenberger, V., Jakubke, H.-D., & Kasche, V. (1991a) *Biochim. Biophys. Acta 1078*, 8–11.
Schellenberger, V., Kosk, M., Jakubke, H.-D., & Aaviksaar, A. (1991b) *Biochim. Biophys. Acta 1078*, 1–7.
Schellenberger, V., Braune, K., Hoffmann, H.-J., & Jakubke, H.-D. (1991c) *Eur. J. Biochem. 199*, 623–636.
Schellenberger, V., Siegel, R., & Rutter, W. (1993a) *Biochemistry 32*, 4344–4348.
Schellenberger, V., Turck, C., Hedstrom, L., & Rutter, W. (1993b) *Biochemistry 32*, 4349–4353.