

Characterization of a Serine Protease Homologous to House Dust Mite Group 3 Allergens from the Scabies Mite *Sarcoptes scabiei*[§]

Received for publication, September 2, 2009, and in revised form, October 5, 2009. Published, JBC Papers in Press, October 7, 2009, DOI 10.1074/jbc.M109.061911

Simone A. Beckham^{‡§1}, Sarah E. Boyd^{¶1,2}, Simone Reynolds^{‡||}, Charlene Willis[‡], Masego Johnstone[‡], Angela Mika[‡], Pavla Simerská^{**}, Lakshmi C. Wijeyewickrema[§], A. Ian Smith[§], David J. Kemp[‡], Robert N. Pike^{§3}, and Katja Fischer^{‡3,4}

From the [‡]Infectious Diseases and Immunology Division, Queensland Institute for Medical Research, Herston, Queensland 4029, Australia, the [¶]Clayton School of Information Technology, Faculty of Information Technology, and the [§]Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Faculty of Medicine, Monash University, Clayton, Victoria 3800, Australia, and the ^{**}School of Molecular and Microbial Sciences and the ^{||}School of Veterinary Sciences, University of Queensland, St. Lucia, Queensland 4072, Australia

The scabies mite, *Sarcoptes scabiei* var. *hominis*, infests human skin, causing allergic reactions and facilitating bacterial infection by *Streptococcus* sp., with serious consequences such as rheumatic fever and rheumatic heart disease. To identify a possible drug target or vaccine candidate protein, we searched for homologues of the group 3 allergen of house dust mites, which we subsequently identified in a cDNA library. The native protein, designated Sar s 3, was shown to be present in the mite gut and excreted in fecal pellets into mite burrows within the upper epidermis. The substrate specificity of proteolytically active recombinant rSar s 3 was elucidated by screening a bacteriophage library. A preference for substrates containing a RS(G/A) sequence at the P1-P2' positions was revealed. A series of peptides synthesized as internally quenched fluorescent substrates validated the phage display data and high performance liquid chromatography/mass spectrometry analysis of the preferred cleaved substrate and confirmed the predicted cleavage site. Searches of the human proteome using sequence data from the phage display allowed the *in silico* prediction of putative physiological substrates. Among these were numerous epidermal proteins, with filaggrin being a particularly likely candidate substrate. We showed that recombinant rSar s 3 cleaves human filaggrin *in vitro* and obtained immunohistological evidence that the filaggrin protein is ingested by the mite. This is the first report elucidating the substrate specificity of Sar s 3 and its potential role in scabies mite biology.

Scabies is a transmissible parasitic skin infestation caused by the mite *Sarcoptes scabiei*. It is common worldwide and spreads

rapidly under crowded conditions, such as those found in institutional settings, e.g. prisons and long term care facilities. Scabies is a major public health problem in socially disadvantaged communities, such as those found in indigenous populations and in developing countries (1, 2). Infestation occurs when the adult female mite burrows in the skin. Pruritic scabies lesions are often accompanied by bacterial infections (3), particularly Group A streptococci. These secondary infections cause significant sequelae (cellulitis, septicemia, and post-streptococcal glomerulonephritis), and the increased community streptococcal burden contributes to extreme levels of acute rheumatic fever and rheumatic heart disease (4). Recently, it has been reported that treatment efficacy for scabies is decreasing (5, 6), indicating that the development of novel control strategies is necessary.

Mite gut proteases involved in host protein digestion offer an avenue for interfering with parasite establishment. In scabies mite-related but non-parasitic astigmatid house dust mites, serine and cysteine proteases have been implicated in skin protein digestion (7–11). Group 3 allergens, typified by Der p 3 from *Dermatophagoides pteronyssinus*, are abundant serine proteases of the S1 family (12). They are thought to be secreted into the gut of the mite and shown to be excreted in fecal pellets, which can become aerosols and, hence, are implicated among elicitors of the allergic response in asthma (11, 13). Interference with scabies mite homologs of house dust mite gut allergens may lead to inhibition of mite digestion, thereby presenting a novel approach to protective intervention. We have previously described a family of at least 33 scabies mite proteins homologous to the serine protease group 3 major allergens of house dust mites (14). Surprisingly, all but one of these molecules is predicted to be catalytically inactive due to mutations of the catalytic amino acids in the active site (scabies mite-inactivated protease paralogues). The sequence exhibiting the closest match to the group 3 allergens of house dust mites (Yv7016G03) displayed a complete catalytic triad and was designated Sar s 3 according to the nomenclature used for house dust mite allergens. Here we have expressed and purified an active recombinant form of the enzyme and used it to characterize the enzyme activity and substrate specificity. The specificity information was used to examine the role(s) of Sar

^{*} This work was supported by a Australian National Health and Medical Research Council fellowship and Program Grant 496600 (to D. J. K.).

[§] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Tables 1–4.

¹ Current address: Centre for Cancer Research, Monash Institute of Medical Research, Monash University, Clayton, Victoria, 3168, Australia.

² Supported by an Australian postdoctoral fellowship and Australian Research Council Discovery Grant DP0771183.

³ These contributors were joint senior authors.

⁴ To whom correspondence should be addressed: P. O. Royal Brisbane Hospital QLD 4029, Australia. Tel.: 61-7-33620417; Fax: 61-7-33620104; E-mail: Katja.Fischer@qimr.edu.au.

s 3. In addition, the site of expression of Sar s 3 within the mite body was localized. Together, these data have provided significant insight into the role of the enzyme in scabies biology.

EXPERIMENTAL PROCEDURES

Localization of Sar s 3, IgG, and Filaggrin by Immunohistochemistry—The production of antibodies in mice, tissue preparation of human mite-infested skin samples, and the immunolocalization of Sar s 3 were accomplished as previously outlined (15). Antibodies to recombinant Sar s 3 were raised in C57BL/6 female mice. Pre-bleeds from mice used for antibody production served as negative controls. Anti-human IgG (Silenus, Melbourne, Australia) was used as a positive control to identify the gut.

For the localization of human filaggrin, formalin-fixed and paraffin-embedded skin biopsies from human *S. scabiei* patients underwent an antigen retrieval procedure in a solution containing 1% (w/v) Revealit-Ag in 200 mM Tris, pH 7.5 (Immuno-Solution, Australia), at 85 °C for 20 min. The sections were cooled and washed with distilled water for 10 min followed by three 5-min washes in phosphate-buffered saline at pH 7.2. The sections were blocked with 10% (v/v) donkey serum in 1% (w/v) bovine serum albumin in phosphate-buffered saline. Endogenous peroxidase activity was blocked with 3% (v/v) H₂O₂ in blocking buffer. The sections were incubated overnight with a polyclonal antibody against filaggrin (Covance, Jomar Biosciences, Australia) at a 1:200 dilution. After washing, the sections were incubated with Dako-EnVision (DakoCytomation) anti-rabbit-horseradish peroxidase conjugate for 45 min. For the detection of human IgG, the sections were incubated with anti-human IgG-horseradish peroxidase polyclonal antibody (Abcam, Sapphire Biosciences, Australia) at a 1:500 dilution. All slides were washed in phosphate-buffered saline, and the Vector VIP peroxidase substrate kit (Vector Laboratories) was used for staining following the manufacturer's recommendations.

Ethical approval for the production of antibodies in mice was obtained from the Queensland Institute of Medical Research Animal Ethics Committee. Ethical approval for the use of shed skin crusts from the bedding of a patient with recurrent crusted scabies was obtained from the Human Research Ethics Committee of the Northern Territory Department of Health and Community Services and the Menzies School of Health Research.

Expression of rSar s 3 in Escherichia coli and Purification of the Enzyme—The sequence encoding mature rSar s 3 (Yv7016G03 (14)) was amplified from the scabies mite EST cDNA clone 7016G03 using specific primers (5'-ACCGGT-CGACATTGTCGGCGGTCGTTTAGCTAAGCC-3') and (5'-ACCGCTGCAGTTAATTATTTCTAAGGATATTTT-TGATCCATTGG-3') (Sigma) with the restriction sites SalI and PstI required for directional cloning into the pQE9 vector (Qiagen). The PCR product was digested at these sites, ligated into the pQE9 vector, transformed into XL1-blue *E. coli* cells (Stratagene), and expressed as an N-terminal hexahistidine fusion protein. Transformants were confirmed by sequencing with BigDye 3.1 (Applied Biosystems) using primers T3 and T7. One clone was selected, and the recombinant protein was

expressed and purified under denaturing conditions on nickel-immobilized metal affinity chromatography (Qiagen) as described by the manufacturer.

Expression of rSar s 3 in Pichia pastoris and Purification of the Enzyme—To direct secretion of the expressed protein into the medium, all constructs placed the predicted N terminus flush with the Kex2 signal peptide cleavage site in the transfer vector pPICZαA (Invitrogen). A stop codon at the 3' end of the Sar s 3 sequence was introduced to ensure that the recombinant protein did not include any of the C-terminal tags of the vector. The plasmid was propagated in *E. coli* strain TOP10F', linearized with SacI, and electroporated into *P. pastoris* KM71H (Invitrogen). Media for colony growth and biomass or induction cultures were made according to the manufacturer's instructions. Cells were initially plated onto supported nitrocellulose on agar plates containing 100 μg/ml zeocin (Invitrogen). After incubation for 24 h at 29 °C, filters were moved onto yeast extract peptone dextrose medium with sorbitol containing 100, 500, 750, or 2000 μg/ml zeocin for selection of transformants with increased numbers of Sar s 3 gene copies. To ensure pure clonal isolates, single colonies were plated twice on yeast extract/peptone/dextrose/zeocin agar plates. Transformants were assessed to select for the highest expressing clone following the manufacturer's instructions. Biomass cultures were grown for 24 h, and induction cultures were grown for 60–70 h with shaking at 29 °C. rSar s 3 expression was detected by immunoblotting using mouse antiserum generated against the rSar s 3 produced in *E. coli*. Correct processing of the recombinant protein at the Kex2 cleavage site was confirmed after purification by N-terminal sequencing in the Peptide Biology Laboratory, Monash University, Australia (data not shown).

Induction cultures were centrifuged at 12,000 × g for 45 min at 4 °C, and the supernatants were stored at –80 °C until required. The supernatant was thawed and prepared for hydrophobic interaction chromatography by the gradual addition of NH₂(SO)₄ to a final concentration of 1.5 M in 25 mM sodium acetate and adjusted to pH 5.0. The preparation was centrifuged for 30 min at 20,000 × g. The supernatant was filtered through a 0.45-μm filter, and the filtrate was loaded onto a 5-ml HiTrap phenyl-Sepharose column (GE Healthcare) pre-equilibrated with 25 mM sodium acetate, 1.5 M NH₂(SO)₄, pH 5.0. Unbound material was removed by washing with equilibration buffer. Protein was eluted using a linear gradient of NH₂(SO)₄ from 1.5 to 0 M in 25 mM sodium acetate, pH 5.0, over 80 ml. The rSar s 3 eluted between 0.6 and 0.8 M NH₂(SO)₄. Peak fractions were pooled and subjected to dialysis against 10 mM MES,⁵ pH 5.5, loaded onto a 5-ml HiTrap SP-Sepharose FF column (GE Healthcare), and washed using 50 ml of 10 mM MES, pH 5.5, to remove unbound material. Protein was eluted using a linear gradient of 0–0.5 M NaCl over 80 ml, with rSar s 3 eluting between 0.1 and 0.2 M NaCl.

Purification was monitored by analysis on 10% SDS-PAGE gels and subsequent visualization with Coomassie Brilliant Blue

⁵ The abbreviations used are: MES, 2(*n*-morpholino)ethanesulfonic acid; Abz, ortho-aminobenzoic acid; CAPS, cyclohexylamino propanesulfonic acid; PoPS, prediction of protease specificity; WT, wild type; CE, cornified envelope.

R-250 dye, silver staining, and immunoblotting analysis. Protease activity in fractions containing rSar s 3 was also determined using standard assay conditions.

Determination of Kinetic Constants for Cleavage of Peptide Substrates—Enzyme assays were performed in standard assay buffer (100 mM Tris, 10 mM CaCl₂, and 0.05% (v/v) polyethylene glycol 8000, pH 8.2) with commercially available methylcoumarin substrates (Sigma) at 37 °C in a final volume of 100 μ l in 96-well plates. Fluorescence was continuously measured on a POLARstar fluorescent microtiter plate reader (BMG Labtech, Melbourne, Australia) at 30-s intervals for 30 min at excitation and emission wavelengths of 370 and 460 nm, respectively.

Initial velocity values at different substrate concentrations were determined for rSar s 3 wild type (WT) in assay buffer containing 4% (v/v) DMSO and fitted using nonlinear regression on Graphpad Prism 4, according to the equation $v = V_{\max}[S]/K_m + [S]$.

All values determined were derived from linear increases in fluorescence intensities, and all errors were below 10%. As no specific inhibitor for active site titration has been identified, k_{cat} values were calculated on the basis of the molarity of total rSar s 3.

pH-dependent Activity and Stability of rSar s 3—The pH dependence of rSar s3 (50 nM) was determined at 37 °C in 20 mM sodium acetate, 20 mM MES, 20 mM Tris, 20 mM CAPS, 0.05% (w/v) polyethylene glycol 8000 adjusted to the pH range 4–11.5. The enzyme activity was started by the addition of 10 μ M Gly-Gly-Arg-7-amino-4-methylcoumarin as a substrate. After 45 min the assay conditions were alkalized with 1.5 M Tris, 10 mM CAPS, pH 12.0, to normalize the pH and reveal any partially masked fluorescence. The protease activity was measured as total fluorescence at excitation and emission wavelengths of 370 and 460 nm, respectively. The alkalization step is necessary as the useful fluorescence properties of methylcoumarin peptides reside in the anion, which is not formed at acidic pH values (16). All values were corrected by control curves measured for the pH-dependent changes in substrate and product-only fluorescence.

rSar s 3 Protease Activity in the Presence of Salts and Protease Inhibitors—The effects of salts and inhibitors were tested by measuring their effects on cleavage of 10 μ M Gly-Gly-Arg-7-amino-4-methylcoumarin by 30 nM rSar s 3 at 37 °C ($n = 3 \pm$ S.D.). The effect of NaCl, phenylmethylsulfonyl fluoride, EDTA, α 1-antitrypsin, and antithrombin III was measured in the standard assay buffer. The effects of CaCl₂ were measured using concentrations ranging from 0 to 100 mM. The effect of ZnCl₂ was measured in the standard assay buffer at pH 6.5 due to the reduced solubility of the substance at higher pH values. Phenylmethylsulfonyl fluoride was dissolved in isopropanol at a final concentration of 10% (v/v) in the assay. All measurements were corrected using chemical control experiments. To visualize binding of rSar s 3 to the nonspecific inhibitor α ₂-macroglobulin (Sigma), the enzyme and inhibitor were incubated at 37 °C for 30 min. The formation of covalent complexes was monitored without prior denaturing and after harsh denaturing conditions (the sample was boiled for 5 min in 1% SDS, 10 mM dithiothreitol) on 4–12% gradient SDS-PAGE gels. Protein bands were stained using Coomassie Blue R-250 or transferred

to nitrocellulose for Western blotting using the anti-Sar s 3 antibodies mentioned above.

Determination of Protease Substrate Specificity Using a Bacteriophage Display Library—The substrate specificity for rSar s 3 spanning the P5-P4' positions was determined using a T7Select1–1b phage display library with an arginine residue fixed at the fifth position generated as previously described (17, 18). Briefly, incubations of rSar s 3 WT (2 μ M) with the libraries loaded onto the nickel-chelate matrices were conducted in 50 mM phosphate, 1 mM MgSO₄, 5.6 mM NaCl, pH 8.2, at 37 °C overnight. The selection process was repeated six times, and phage-displayed peptide sequences were determined as previously described.

The Use of Fluorescent-quenched Substrates to Validate Substrate Specificity Data—To investigate whether sequences obtained from the phage display analysis were optimal peptide substrates for Sar s 3 and to further investigate the preferences identified, five fluorescence-quenched peptide substrates were synthesized (H-Abz-AMWRSGKLY-3-NO₂, H-Abz-AMWRAGKLY-3-NO₂, H-Abz-AMWRSVKLY-3-NO₂, H-Abz-AMWRAVKLY-3-NO₂, H-Abz-AGARSGGVY-3-NO₂; GL Biochem Ltd, Shanghai, China). Kinetic parameters were determined as for the methylcoumarin substrates; however, excitation and emission wavelengths of 305 and 420 nm, respectively, were used.

Confirmation of the Cleavage Site in a Peptide Substrate by Liquid Chromatography-Mass Spectrometry—The position of cleavage in the substrate Abz-AGARSGGVY-NO₂ by rSar s 3 was confirmed by incubating 10 mM substrate with 1 mg/ml rSar s 3 in 100 mM Tris, 10 mM CaCl₂, 0.05% (v/v) polyethylene glycol 8000, pH 8.2. The reaction was performed at room temperature for 5 h. Analytical reverse phase high performance liquid chromatography was performed using a 0–60% linear gradient from 0.1% (v/v) aqueous trifluoroacetic acid to 90% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid over 30 min on a Vydac (Hesperia, CA) analytical C18 column (218TP54; 5 μ m; 4.6 $\times$ 250 mm) with a 1 ml/min flow rate and detection at 214 nm. Electrospray ionization mass spectrometry was performed on a PerkinElmer Life Sciences Sciex API3000 using the Analyst 1.4 (Applied Biosystems/MDS Sciex, Toronto, ON, Canada) software. Samples (1–10 μ l) were injected into acetonitrile, H₂O mobile phases containing 0.1% (v/v) acetic acid and run in positive ion mode.

Searching the Human Proteome Using Prediction of Protease Specificity (PoPS) Program—The specificity model for Sar s 3 in PoPS was generated from the $\Delta\sigma$ phage display data detailed in [supplemental Table 2](#). The values for each subsite were scaled to the range –5.0 to +5.0, as required by PoPS (19). The PoPS model also allows weights to be set for each subsite, which are multiplied by the values of the matrix during the prediction of cleavage sites (18). The weights for the Sar s 3 model were set to be the scale factors derived from the scaling of each subsite, so that the PoPS model was mathematically equivalent to the values in [supplemental Table 2](#). The human proteome was then searched using the PoPS algorithm as described previously (19). Hypothetical and predicted proteins were not considered in the analysis.

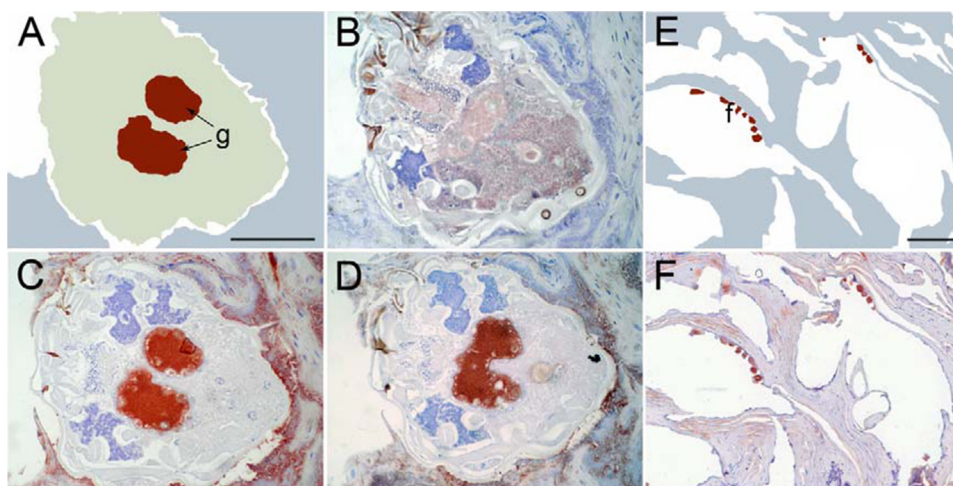


FIGURE 1. Immunohistological colocalization of Sar s 3 and host IgG indicates that Sar s 3 is located in the mite digestive system. A schematic diagram (A) outlines the features visible in three serial histological sections of a mite in the epidermis. The epidermis is shown in blue, the burrow is in white, the body of the mite is in gray, and the mite digestive system is in red. B–D represent the matching serial histological sections through the same mite. Red staining indicates binding of antibody to protein. B was probed with pre-immune mouse serum as a negative control remained unstained, whereas C was probed with anti-human IgG, a marker for mite gut (20) and D was probed with anti-Sar s 3-stained equivalent regions. E shows a schematic diagram of mite burrows at a lower magnification, and F represents the matching section probed with anti-Sar s 3 with positively stained feces. g, gut; f, feces; scale bars equal 100 μ m.

In Vitro Cleavage of Filaggrin with rSar s 3—1 μ g of a recombinant filaggrin fragment (0.82 μ M) corresponding to the C-terminal amino acids 3838–4062 of human filaggrin (Abcam) was incubated with 0, 2, and 20 μ g of rSar s 3 (0, 2.3, and 23 μ M) in 100 mM Tris, 10 mM CaCl_2 , 0.05% (v/v) polyethylene glycol 8000, pH 8.2. The reaction was performed at 30 $^\circ\text{C}$ for 10 min and then stopped by the addition of SDS-PAGE loading buffer and heat inactivation at 95 $^\circ\text{C}$. Filaggrin degradation was analyzed by immunoblotting using a mouse monoclonal antibody against recombinant human filaggrin (Affinity Bio Reagents, Golden, CO).

RESULTS

Localization of Sar s 3 to the Mite Digestive System and Fecal Pellets in Human Skin—Recombinant rSar s 3 and a series of three scabies mite-inactivated protease paralogues were expressed as insoluble products using an *E. coli* system, and the resolubilized rSar s 3 was used to raise polyclonal antibodies in mice. Testing of the antibodies using immunoblotting analysis revealed that the antibodies against rSar s 3 were specific to the enzyme and did not cross-react with scabies mite-inactivated protease paralogues from the scabies mite (data not shown).

The localization of Sar s 3 protein within and outside the scabies mite was shown by using the polyclonal antibodies for immunohistochemical staining of serial sections of human skin infested with mites (Fig. 1). A section probed with preimmune serum showed only minor background staining in the digestive system of the mites (Fig. 1A). In an adjacent section, human IgG, which for the purpose of this experiment was used as a positive control, was localized in the gut (Fig. 1B) as previously documented (20). In an adjacent section the anti-Sar s 3 antibody also yielded strong staining in the digestive system of the mite (Fig. 1C). In addition to its location in the digestive tract of the mite, Sar s 3 was localized to external acellular masses,

which we conclude are scybala (fecal pellets; Fig. 1D). The scybala were negative for staining with pre-immune serum (data not shown), consistent with this conclusion. From these results we are able to conclude that Sar s 3 is located to the digestive tract of the Scabies mite and is excreted in the fecal matter of the parasite. The enzyme is, therefore, highly likely to be involved in digestive functions within the gut of the mite and would be predicted to gain access to skin proteins in the epidermis via its fecal deposition.

Expression and Purification of rSar s 3—The rSar s 3 protein expressed in *P. pastoris* was soluble, in contrast to the material previously expressed in *E. coli*. Three initial constructs were trialed; WT Sar s 3 including the propeptide, enzyme with the propeptide modified by

site-directed mutagenesis of the Lys residue at position 7 of the pro-enzyme to an Asp residue (21), and the enzyme without a propeptide (Fig. 2A). Only the wild type construct with the complete propeptide without mutation proved to be proteolytically active and was investigated further.

Recombinant rSar s 3 enzymes (WT and glycosylation null mutants) were purified using hydrophobic interaction and ion exchange chromatography, generating on average 80 μ g of purified protein per liter of biomass culture. Two bands were observed for the WT enzyme by SDS-PAGE analysis (Fig. 2B). Both were cross-reactive with antisera against rSar s 3 produced in *E. coli*, and when treated with peptide *N*-glycosidase F, a single band was produced (data not shown), suggesting glycosylation on both predicted *N*-linked glycosylation sites (Asn-21 and Asn-101). Both bands from rSar s 3 WT had the same *N*-terminal sequence, which matched the predicted mature *N*-terminal sequence of the protease (Fig. 2B). The putative *N*-linked glycosylation sites were altered by site-directed mutagenesis (Fig. 2A) either individually or in combination (rSar s 3, N21Q, N101Q, and N21Q/N101Q mutants), and the proteins were produced in *P. pastoris*. The rSar s 3 N21Q and N101Q mutant enzymes were expressed and purified, but the N21Q/N101Q mutant was produced in lesser quantities and was unstable; it was, therefore, deemed unsuitable for further analysis.

Activity of rSar s 3 against Di- and Tripeptide Coumarin Substrates—Recombinant rSar s 3 WT exhibited trypsin-like protease activity of variable magnitude against 8 of the 10 fluorogenic substrates tested (Fig. 2C). This activity was abolished by the Lys to Asp mutation at position 7 in the propeptide of the enzyme, indicating that cleavage to reveal the *N*-terminal Ile residue is important for activity of the enzyme.

The activity of rSar s 3 WT as well as the N21Q and N101Q mutants was determined using various fluorogenic peptide

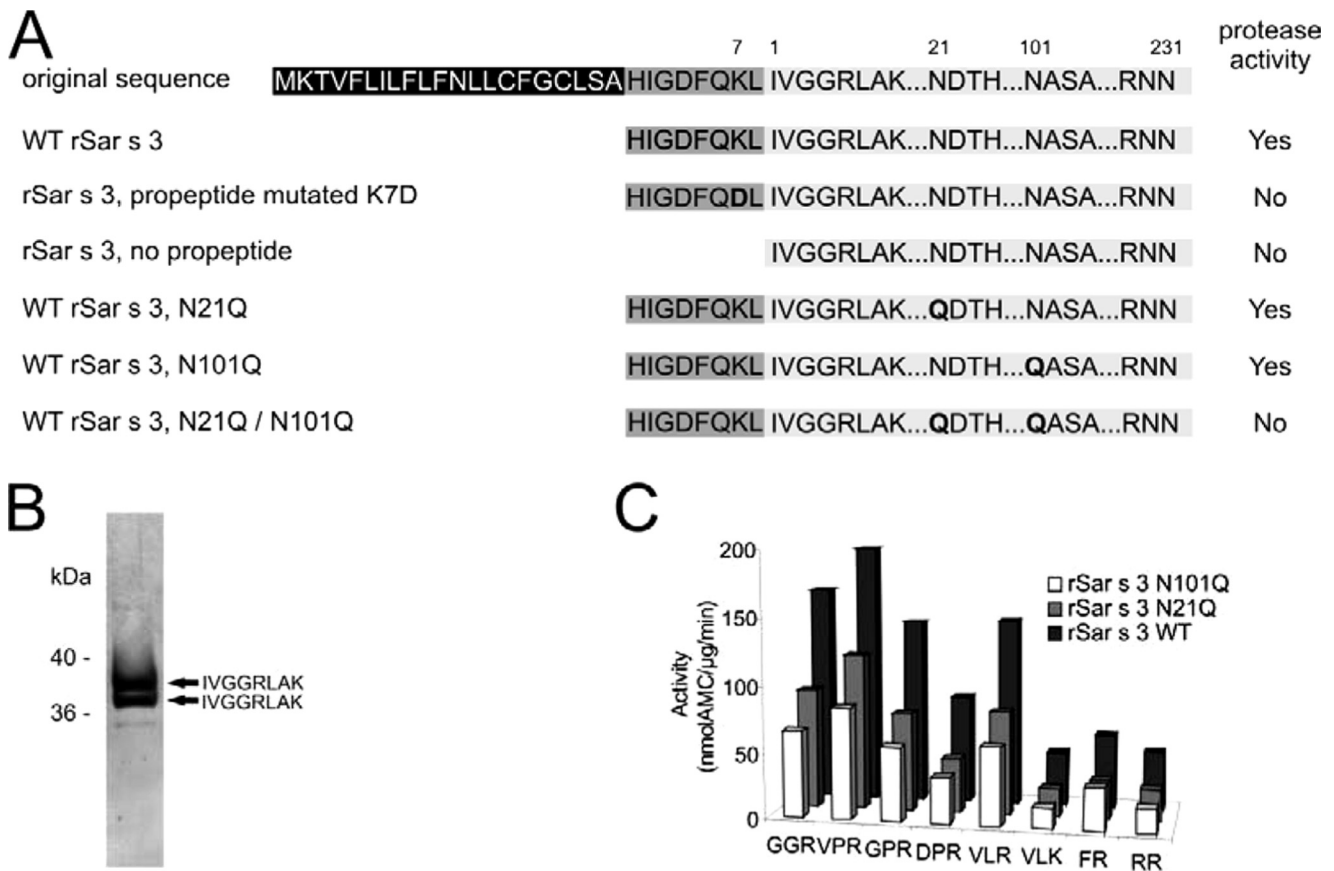


FIGURE 2. Alignment of *Sar s 3* constructs and analysis of r*Sar s 3* purity and substrate specificity. *A*, shown is amino acid alignment of the *Sar s 3* constructs used, including wild type *Sar s 3* and mutants with changes to the propeptide and predicted glycosylation sites. *B*, purified recombinant r*Sar s 3* WT was subjected to SDS-PAGE and visualized by silver staining. The N-terminal sequences of the bands are shown next to the protein bands. *C*, shown is specific activity of r*Sar s 3* WT, r*Sar s 3* N21Q, and r*Sar s 3* N101Q for cleavage of different methylcoumarin peptide substrates at a concentration of 50 μ M. The values are shown in nmol of 7-amino-4-methylcoumarin (AMC)/ μ g/min.

substrates (Fig. 2C). Negligible activity was observed against the substrate used to detect chymotrypsin-like activity (*N*-succinyl-Ala-Ala-Pro-Phe- and *N*-succinyl-Ala-Ala-Phe-7-amino-4-methylcoumarin), whereas substrates with Arg residues at the P₁ position were favored over substrates with Lys residues at P₁ (as shown by the direct comparison between D-Val-Leu-Arg-7- and Boc-Val-Leu-Lys-7-amido-4-methylcoumarin substrates). These results indicate that the *Sar s 3* enzyme is trypsin-like in its preference. The r*Sar s 3* WT, N21Q, and N101Q enzyme forms exhibited similar activity profiles against the different peptide substrates, indicating that the site of glycosylation does not alter the substrate specificity nor does it substantially alter the catalytic activity. Interestingly, the activity of r*Sar s 3* WT was optimal at markedly alkaline pH values (Fig. 3A). The enzyme had only 40% of its highest activity at pH 7.5 and less than 30% of its highest activity in the acidic pH range.

Classical effectors of proteases were used to characterize *Sar s 3* activity. At concentrations of NaCl found in mammalian physiological fluids, the *Sar s 3* activity was inhibited by 20% (Fig. 3B). Sensitivity of proteases to NaCl has been shown for other serine proteases, such as thrombin (22). Calcium, often necessary as activator and/or stabilizer of serine proteases, did not stimulate but, rather, inhibited r*Sar s 3* enzyme activity (Fig. 3C). ZnCl₂ did not affect *Sar s 3* activity at physiological concentrations in the micromolar range, whereas excess concen-

trations revealed inhibition up to 90% (Fig. 3D). In comparison, caspase-3 and metalloproteases such as aminopeptidases involved in neuropeptide production are highly sensitive to Zn²⁺ (23, 24). Phenylmethylsulfonyl fluoride inhibited r*Sar s 3* only at higher concentrations (equal to or above 10 mM), whereas 1 mM EDTA did not exhibit any effect. Phenylmethylsulfonyl fluoride is a classical serine protease inhibitor, whereas EDTA targets metal and calcium-activated proteases. The commercially available human serpins, α_1 -antitrypsin and antithrombin, were poor inhibitors of r*Sar s 3*. The residual protease activities were 96.4 and 91.8% at initial inhibitor/enzyme ratios of 100 and 10, respectively, for antitrypsin and antithrombin.

The covalent complex formed between α_2 -macroglobulin and endopeptidases can be useful to characterize novel proteases. Proteinase entrapment by α_2 -macroglobulin is irreversible; however, the complexed enzyme retains substantial activity against small peptides but is almost completely inactive against proteins larger than 25 kDa (25). We investigated the *Sar s 3*- α_2 -macroglobulin interaction by incubating *Sar s 3* and α_2 -macroglobulin followed by SDS-PAGE and Western analysis. Under non-reducing conditions, *Sar s 3* co-migrated with α_2 -macroglobulin, demonstrating the formation of α_2 -macroglobulin-proteinase complexes (Fig. 3E, panel I and II). An obvious gel shift occurred when the same samples were boiled

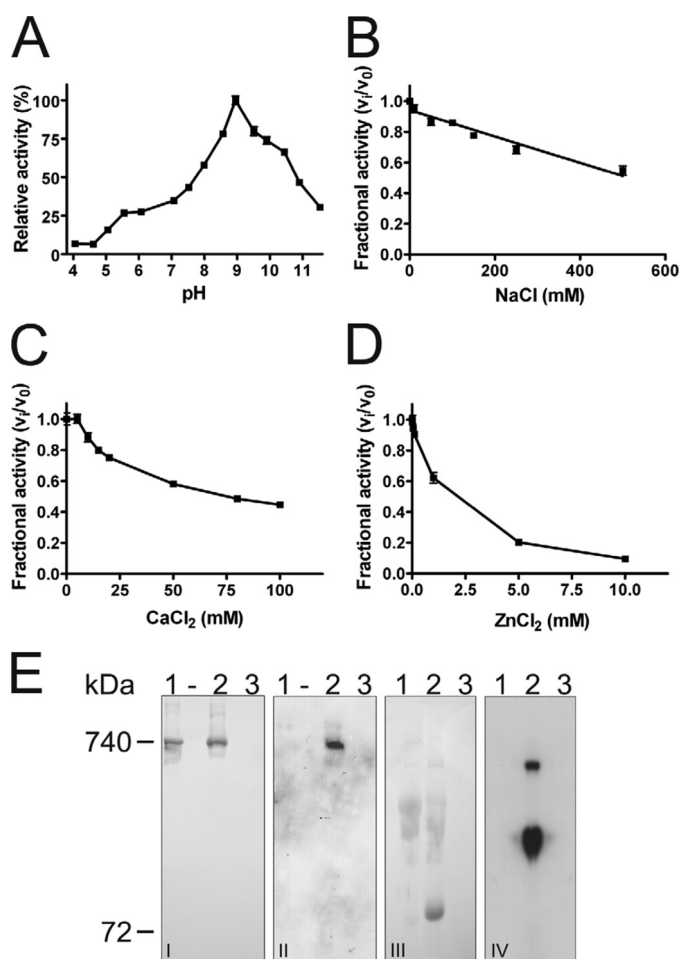


FIGURE 3. Effect of pH, salts, and α_2 -macroglobulin on rSar s 3 activity. A, shown is pH dependence of rSar s 3 WT protease activity against Gly-Gly-Arg-7-amino-4-methylcoumarin. The results shown are the means $\pm$ S.D. ($n = 3$) from 3 independent purifications and measurements. The curve was corrected by using chemical control experiments, i.e. pH-dependent changes in substrate and product-only fluorescence, and normalized to the maximum of the mean activity. rSar s 3 exhibited optimal activity at pH 9.0. B–D, shown is the effect of NaCl (B), CaCl₂ (C), measured at pH 8.2, and ZnCl₂ (D), measured at pH 6.5, on rSar s 3 at a 30 nM concentration (shown are means $\pm$ S.D., $n = 3$). E, Sar s 3 forms a covalent linkage with α_2 -macroglobulin. The formation of covalent links between α_2 -macroglobulin and rSar s 3 was monitored by analysis on 4–12% gradient SDS-PAGE gels. Before loading, samples were either not treated (panels I and II) or boiled for 5 min in 1% SDS, 10 mM dithiothreitol (panels III and IV). Within each panel, lane 1 was loaded with 5 μ g α_2 -macroglobulin, lane 2 was loaded with 5 μ g of α_2 -macroglobulin and 5 μ g of rSar s 3 after 30 min of incubation at 37 $^{\circ}$ C, and lane 3 was loaded with 5 μ g of rSar s 3. Panels I and III were visualized by staining with Coomassie Blue R-250, and panels II and IV were visualized after Western blot analysis using specific anti-Sar s 3 antibodies. Free rSar s 3, which runs below 40 kDa (see Fig. 2B), was not visible on the region of gel shown.

for 5 min in 1% SDS and 10 mM dithiothreitol, which is characteristic for endopeptidases complexed with α_2 -macroglobulin (26). This test ensured that the observed binding was specific (Fig. 3E, panel III and IV). In addition, as described by Hall and Roberts (26), the denaturing and reducing treatment caused the α_2 -macroglobulin to be partially converted into multiple low molecular weight bands, of which some still bound Sars3. Low molecular weight protein, presumably corresponding to unbound Sar s 3, was visible at the very bottom of both Western blots (not shown). The characteristics of the enzyme revealed in this study indicate that the enzyme is a somewhat atypical serine protease.

Determination of Substrate Specificity Using Phage Display—The sequences of 100 phage-displayed peptides were determined from the sublibrary generated after six cycles of screening with phage display library (supplemental Table 1). Analysis of the sequences revealed that the most frequently occurring sequences were RAMRSGKL (5 times), SRNVRSRAG (4 times), and RARARSRQA (3 times). Numerous sequences were found twice, e.g. YARFRSGGI, YSVARSGY, LTRRRSAVV, RARMRSVV, and TKLRRAVIS. Interestingly, it was immediately apparent that a number of these repeated sequences contained a similar motif; that is, Arg followed by Ser or Ala followed by Gly or Ala. The motif did not necessarily occur in the middle of the sequence, as a number of sequences were repeated with the potential motif occurring at the beginning of the sequence: e.g. RSGERKAVK, RSGIRRVVT, and RSGMRLKLM. The remaining 70 sequences occurred once only. The potential motif was present in 58 of the 100 sequences (33 of 70 unique sequences), and elements of the motif were present in almost all sequences. Because earlier data (Fig. 2C) indicated that the enzyme preferred to cleave after Arg residues, these data indicate that the enzyme has a preference for a small hydrophilic residue in P1' (Ser) and a small residue at P2' (Ala/Gly). The data also suggest that the enzyme substrate specificity is most strikingly dictated by prime side residues.

Analysis of the phage display data aligned with the Arg residue fixed at the fifth position was then carried out to reveal a measure of the number of standard deviations ($\Delta\sigma$) away from an expected "normal" to identify over-representation of particular amino acids at any given substrate position from P5–P4' (27) (see supplemental Table 2 for detailed values). This analysis revealed rSar s 3 WT to exhibit a preference for residues at P2 in the following order: Ala > Trp > Met ($\Delta\sigma$ value for the most preferred residue (Ala) = 4.9 (Fig. 4)). For the P3 position, the preference was for Arg > Gly > Ala > Met = Lys ($\Delta\sigma$ value for Arg = 4.9), whereas at the P4 position the preference was for Ala > Ser ($\Delta\sigma$ value for Ala = 4.9), and at the P5 position, the preference was Arg > Tyr ($\Delta\sigma$ value for Arg = 7.6). These data indicate that P5 was the most discriminating of the non-prime side positions followed by P4. It should be noted, however, that it is likely that the values for P5–P3 in particular are most likely strongly influenced by the reasonably frequent occurrence of the previously noted motif (RAG) at the first, second, and third positions of the sequences (11 non-redundant sequences of 82).

Generally, as might be expected from the previously discussed motif, the $\Delta\sigma$ values were higher for the prime sites of the substrates, in particular at the P1' and P2' positions. At the P1' position, the $\Delta\sigma$ value for the Ser residue was 8.3 followed by Ala with a $\Delta\sigma$ value of 4.4. At the P2' position, the $\Delta\sigma$ value for Gly was 8.5 and that for Ala was 6.7. The high values for these residues reinforced the manual detection of the motif identified previously and yielded a quantitative assessment of the importance of the different amino acids at the P1' and P2' positions, with Ser residues most important at P1' and Gly residues marginally more important than Ala residues at the P2' position. At the P3' position Gly was the most preferred residue, with a $\Delta\sigma$ value of 2.6 and Val at 2.2; thus, this position was apparently the least selective of all analyzed. Valine was the most preferred residue at the P4' position,

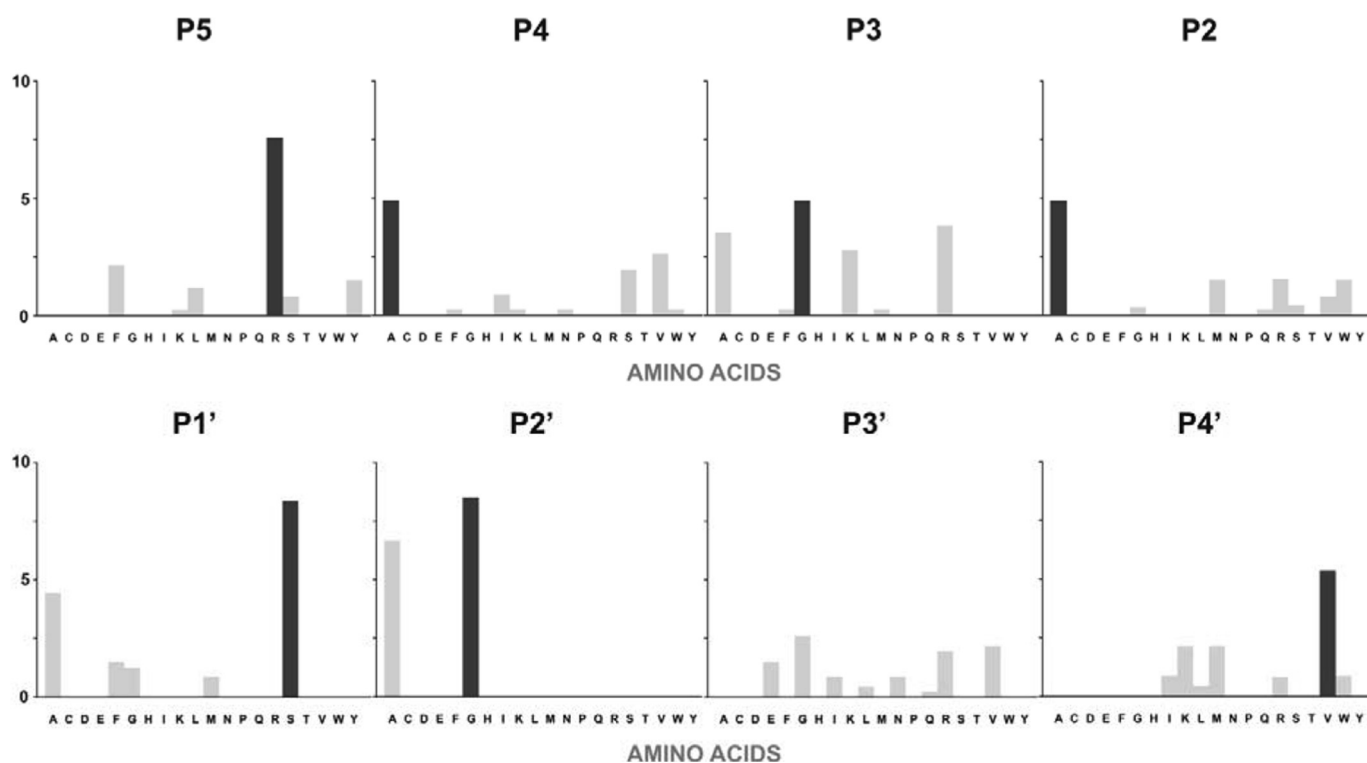


FIGURE 4. The substrate specificity of rSar s 3 as revealed by phage display analysis. The $\Delta\sigma$ values are shown for each amino acid at the P5-P4' positions. Negative values were omitted for clarity, and the highest preference at each position (for values close to or above five) is indicated with the dark-shaded bar.

with a $\Delta\sigma$ value of 5.8. The consensus sequence from this analysis was RAGARSGVV.

Validation of Phage Display Data Using Internally Quenched Fluorescent Peptides—To investigate whether sequences obtained from the phage display analysis were optimal peptide substrates for Sar s 3 and to further investigate the preferences identified, a fluorescence-quenched peptide substrate based on the most frequently occurring peptide, YAMWRSGKL (5/100 sequences)m was synthesized (H-Abz-AMWRSGKLY-3-NO₂). The peptide exhibits the signature motif, RSG, at the P1-P2' positions. The Ser and Gly residues at the P1' and P2' positions of the peptide were also selectively altered to Ala and Val, respectively, to investigate their importance (H-Abz-AMWRAGKLY-3-NO₂ and H-Abz-AMWRSVKLY-3-NO₂). In addition, a substrate containing changes at both the P1' and P2' positions was synthesized (H-Abz-AMWRAVKLY-3-NO₂). Finally, a peptide representing the consensus sequence for all sequences was synthesized (H-Abz-AGARSGGVY-3-NO₂).

On the basis of the k_{cat} and k_{cat}/K_m values obtained, the most favored substrate was that based on the consensus sequence AGARSGGVY (Table 1). The substrate, AMWRAGKLY, representing the most commonly found sequence in the phage display data, was the second most favored sequence closely followed by AMWRSGKLY. AMWRSVKLY was second least favored, and AMWRAVKLY was the least favored. With respect to the importance of the Ser residue at the P1' position, there was a ~ 1.2 -fold difference between the k_{cat}/K_m values for AMWRAGKLY and AMWRSGKLY compared with >8 -fold difference in k_{cat}/K_m values when comparing AMWRAGKLY and AMWRAVKLY, indicating the S2' pocket is more impor-

TABLE 1

Validation of phage display data for rSar s 3

Determination of the kinetics of cleavage of representative substrate sequences by the protease is presented. The substrates were synthesized as fluorescence-quenched peptides.

Peptide substrate	K_m μM	K_{cat} s^{-1}	k_{cat}/K_m $\text{M}^{-1} \text{s}^{-1}$
Abz-AGARSGGVY-3-NO ₂	72.79	0.118095	1622.41
Abz-AMWRAGKLY-3-NO ₂	49.67	0.037143	747.7926
Abz-AMWRSGKLY-3-NO ₂	42.99	0.02619	609.2225
Abz-AMWRSVKLY-3-NO ₂	45.32	0.009262	204.3668
Abz-AMWRAVKLY-3-NO ₂	37.61	0.003214	85.46359

tant in dictating substrate specificity than the S1' pocket in the context of these peptide substrates.

The cleavage of the peptides at the Arg residue in the substrate Abz-AGARSGGVY-NO₂ by Sar s 3 was confirmed by carrying out liquid chromatography-mass spectrometry analysis on the peptide before and after cleavage by the enzyme. After incubation with Sar s 3 for 5 h, the peptide substrate was cleaved to yield products of $[M + H]^+$ m/z 493.4 (calculated 492.3) and $[M + H]^+$ m/z 525.1 (calculated 527.2). Thus, the analytical liquid chromatography-mass spectrometry analysis confirmed that cleavage of the substrate occurred after the Arg residue to yield the predicted products Abz-AGAR and SGGVY-NO₂. A control reaction using commercial trypsin instead of Sar s 3 resulted in identical products (data not shown).

Prediction of Novel Human Substrates for Sar s 3 Using PoPS—The specificity data from the phage display analysis was used to generate a model of Sar s 3 specificity that was employed to scan the human proteome using the PoPS program, as detailed previously (28). Hypothetical and predicted proteins were removed, and then the top 100 predicted hits were considered.

Within these predicted substrates, a number of intracellular proteins were ranked highly for cleavage, but it is less likely that Sar s 3 would encounter these proteins in the extracellular environment in which it is likely to be acting. However, within the top 100 scoring hits were four skin-specific proteins, keratin 1, filaggrin, desmoplakin (isoforms I and II), and envoplakin (supplemental Table 3).

Although keratin 1 displayed only one RSGYRSGGG sequence, filaggrin was of particular interest, with multiple predicted cleavage sites. All of these contained the consensus motif for cleavage by Sar s 3 (RS(A/G)) at the P1-P2' positions (supplemental Table 4), as well as very high PoPS scores. Profilaggrin contains an abundance of more than 90 potential Sar s 3 cleavage sites including 28 RSG and 12 RSA sequences, suggesting that it is a likely target of Sar s 3.

In Vitro Digestion of Human Filaggrin and Keratin with rSar s 3—To determine whether Sar s 3 could directly degrade the human skin protein, filaggrin, a recombinant fragment of human filaggrin containing six putative cleavage sites, was treated with the enzyme, and the products were examined by immunoblotting with anti-filaggrin antibody. After a 10-min incubation time, 1 μ g of filaggrin was partially digested by 2 μ g of Sar s 3 and completely digested by 20 μ g of the enzyme (Fig. 5), representing an estimated specific activity of 0.135 nmol/min/mg of protein. It must be noted at this point that we have

yet to establish an active site titration method for rSar s 3, and therefore, this activity against the protein substrate is likely to underestimate the specific activity of the enzyme for this substrate. We also incubated the protein with rSar s 3 and examined the cleavage using a protein stain after SDS-PAGE. The results showed that 5 μ g of the enzyme was able to essentially completely degrade the protein after 2 h with no intermediate bands present, indicating that at these enzyme concentrations the filaggrin fragment used was being digested at all potential sites within the protein (results not shown). Incubation with lower concentrations of enzyme did not yield any digestion of the protein. It should be noted that the available recombinant fragment of filaggrin does not contain sites predicted to be most optimal for cleavage by rSar s 3; thus, cleavage of this fragment might be considered to be considerably less than optimal, perhaps explaining the requirement for a large amount of enzyme to efficiently cleave the substrate. The cleavage of human keratin in a skin lysate was examined using 5 μ g of enzyme over 2 h, but no cleavage of this protein could be seen (results not shown).

Localization of Human Filaggrin to the Mite Digestive System—Using commercial antibodies against human filaggrin along with antibodies against human IgG as a positive control for gut localization, the presence of filaggrin in the mite gut was demonstrated (Fig. 6).

DISCUSSION

The scabies mite is a parasite that lives in the epidermal layer of human skin. The mite is thought to continually feed on constituents of the uppermost layer of the epidermis known as the stratum corneum or cornified envelope (CE). The CE is formed by flattened dead-cell remnants to create a physical barrier against the environment. At the molecular level, the CE is formed by proteins, including filaggrin, loricrin, trichohyalin, involucrin, small proline-rich proteins, and keratin intermediate filaments. These components are highly cross-linked by transglutaminases. In addition, a complex series of insoluble lipids help produce a complete barrier. The function of the CE is to exclude foreign substances and harmful organisms, such as viruses, bacteria, and fungi, and to prevent the loss of vital fluids. This system is continuously regenerated by differentiating keratinocytes in a highly organized process (for review, see Refs. 29 and 30). When the female scabies mite burrows into the

superficial skin layers, it moves by mechanically disrupting the CE (31). Proteases would be required to digest the ingested skin proteins and perhaps also might play a role in degrading skin proteins outside of the mite (32). Scybala (feces) are left behind as they travel through the epidermis, creating linear lesions clinically recognized as burrows (33). We have shown that a serine protease from the scabies mite, Sar s 3, can be found in the digestive tract of the mite and also in fecal matter deposited in the skin by the parasite;

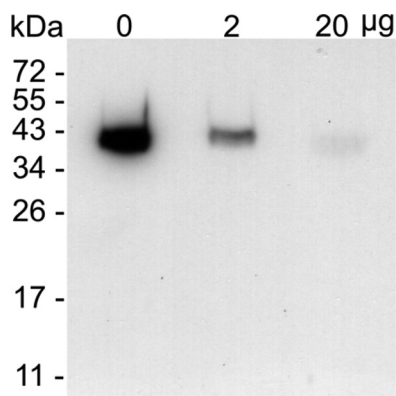


FIGURE 5. *In vitro* digestion of filaggrin with rSar s 3. A tagged recombinant filaggrin fragment corresponding to amino acids 3838–4062 of human filaggrin was incubated for 10 min with 0, 2, and 20 μ g of rSar s 3, expressed and purified from *P. pastoris*. The digests were analyzed by immunoblotting using the anti-filaggrin antibody.

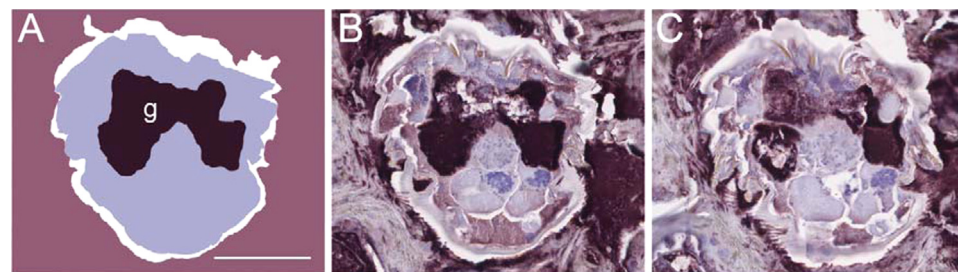


FIGURE 6. Immunohistological colocalization of filaggrin and host IgG indicates that human filaggrin is ingested into the mite gut. A schematic diagram (A) outlines the features visible in two serial histological sections of a mite in the epidermis. The epidermis is shown in light purple, the burrow is in white, the body of the mite is in blue, and the mite digestive system is in dark purple. g, gut; the scale bar equals 100 μ m. B and C represent the matching serial sections through the same mite. Dark purple staining indicates binding of antibody to protein. B was probed with anti-human IgG, and C was probed with anti-filaggrin antibodies.

this led us to question whether the enzyme was involved in digesting skin proteins either within the mite or externally.

We successfully expressed a recombinant form of the Sar s 3 serine protease in a yeast expression system. The enzyme was purified and shown to be an active, mature form with the propeptide cleaved at the position required to liberate the N-terminal Ile residue required to complete the oxyanion hole of the enzyme. The data indicate that the propeptide of the enzyme is vital for formation of active enzyme and that processing after the Leu residue at position 8 of the propeptide is necessary to produce mature protease. The Lys residue at position 7 is clearly required for the cleavage by the yeast enzyme responsible for generating mature and active rSar s 3. The enzymes responsible for similarly processing the enzyme within the mite are currently unknown, although the cleavage is unlikely to be an autocatalytic event given that Sar s 3 appears to be trypsin-like and, therefore, unlikely to cleave after the Leu residue at position 8. Our current hypothesis is, therefore, that the Sar s 3 enzyme is cleaved by an unknown Scabies mite enzyme in the digestive tract to form active, mature enzyme, therefore, allowing Sar s 3 to be active in both the digestive tract of the mite and in the feces deposited in the skin.

The rSar s 3 was not inhibited by a number of host protease inhibitors, although α_2 -macroglobulin was capable of inhibiting the enzyme. The enzyme had a highly alkaline pH optimum, although the physiological implications of this are unclear given that the pH of the mite gut is unknown. The enzyme also displayed a strong specificity for residues at the P1-P2' positions of substrates, with the motif of RSG/A being strongly recognized at these positions, suggesting that the S1' and S2' subsites of the enzyme might be somewhat occluded and, thus, unable to accept larger amino acids at these positions in substrates. Attempts to crystallize the enzyme thus far have been unsuccessful, and models could not be built based on known structures due to poor sequence homology. Thus, further insights into the reasons for the high specificity at these positions await further structural insights. Nevertheless, the information regarding the substrate specificity of the enzyme, which was confirmed using peptide substrates, allowed us to examine the human proteome for potential substrates, revealing that several skin proteins contained potential cleavage sites for the mite enzyme.

We have shown here that rSar s 3 was able to digest a recombinant fragment of the human skin protein, filaggrin (filament aggregating protein), which is a key component of the CE. We predict that the full-length filaggrin will be significantly more sensitive to Sar s 3 digestion as the protein outside of the recombinant fragment used contains highly optimal predicted sites for cleavage by the protease. Filaggrin is formed from profilaggrin, a highly phosphorylated, histidine-rich, 500-kDa polypeptide consisting of 10–12 tandemly arranged 35-kDa filaggrin units. Candi *et al.* (29) termed these units “KIFglue,” the matrix that embeds, aligns, and aggregates keratin intermediate filaments and subsequently maintains the epidermal texture. If there is a reduction or complete absence of filaggrin, as occurs in heritable skin disorders such as atopic dermatitis or ichthyosis vulgaris (34), the skin barrier may be defective and permeable to external factors such as allergens or pathogens. The

mechanical disruption of the filaggrin matrix by the scabies mite and subsequent digestion of the protein in the mite digestive tract would be predicted to have similar effects.

The excretion of the protease in the mite feces would be expected to allow the protease access to the proteins found in the epidermal layer. As proteins in house dust mite feces have been found to gain considerable access to lung tissue of allergy sufferers (35), it is similarly likely that the protease in the mite fecal matter might gain such access to skin. The actions of this protease would be anticipated to cause considerable disruption to the proper assembly and structure of the epidermal layer. The precise cleavage of proteins such as filaggrin has been shown to be essential to proper formation of this vital protective layer (36). Disruption of the structure of this protein, which has been shown here to be a substrate of the Sar s 3 protease, would be expected to substantially disrupt the CE, and indeed such disruption of the epidermis is commonly observed in sufferers of scabies mite infestations. The disruption of the epidermis would also be expected to render patients susceptible to secondary infections, which is in fact the case (4). Thus, the Sar s 3 protease has emerged as an excellent candidate for the development of inhibitors to combat the allergic symptoms engendered in patients infested with scabies mites as well as the associated secondary infections. The unique specificity motif of the protease identified in this study together with some atypical enzymatic properties might be of considerable utility in the development of such inhibitors.

Acknowledgments—We thank Kate Mounsey, Deborah Holt, Shelley Walton, and Bart Currie from the Tropical and Emerging Infectious Diseases Division of the Menzies School of Health Research in Darwin, Australia for providing material for localization studies and Mei-Fong Ho of the Queensland Institute for Medical Research for technical assistance.

REFERENCES

- Walton, S. F., Dougall, A., Pizzutto, S., Holt, D., Taplin, D., Arlian, L. G., Morgan, M., Currie, B. J., and Kemp, D. J. (2004) *Int. J. Parasitol.* **34**, 839–849
- Walton, S. F., and Currie, B. J. (2007) *Clin. Microbiol. Rev.* **20**, 268–279
- Brook, I. (1995) *J. Clin. Microbiol.* **33**, 2139–2140
- Currie, B. J., and Carapetis, J. R. (2000) *Australas. J. Dermatol.* **41**, 139–145
- Currie, B. J., Harumal, P., McKinnon, M., and Walton, S. F. (2004) *Clin. Infect. Dis.* **39**, e8–12
- Mounsey, K. E., Holt, D. C., McCarthy, J., Currie, B. J., and Walton, S. F. (2008) *Future Microbiol.* **3**, 57–66
- Chua, K. Y., Stewart, G. A., Thomas, W. R., Simpson, R. J., Dilworth, R. J., Plozza, T. M., and Turner, K. J. (1988) *J. Exp. Med.* **167**, 175–182
- Stewart, G. A., Ward, L. D., Simpson, R. J., and Thompson, P. J. (1992) *Immunology* **75**, 29–35
- Stewart, G. A., Kollinger, M. R., King, C. M., and Thompson, P. J. (1994) *Allergy* **49**, 553–560
- Smith, W. A., Chua, K. Y., Kuo, M. C., Rogers, B. L., and Thomas, W. R. (1994) *Clin. Exp. Allergy* **24**, 220–228
- Thomas, W. R., and Smith, W. (1999) *Clin. Exp. Allergy* **29**, 1583–1587
- Rawlings, N. D., and Barrett, A. J. (1994) *Methods Enzymol.* **244**, 19–61
- Platts-Mills, T. A. (1998) *J. Allergy Clin. Immunol.* **101**, S364–S366
- Holt, D. C., Fischer, K., Allen, G. E., Wilson, D., Wilson, P., Slade, R., Currie, B. J., Walton, S. F., and Kemp, D. J. (2003) *J. Invest. Dermatol.* **121**, 1419–1424
- Willis, C., Fischer, K., Walton, S. F., Currie, B. J., and Kemp, D. J. (2006)

- Am. J. Trop. Med. Hyg.* **75**, 683–687
16. John, R. A. (2002) in *Enzyme Assays* (Eisenthal, R., and Danson, M. J., eds) pp 75–76, Oxford University Press, Oxford
 17. Kerr, F. K., O'Brien, G., Quinsey, N. S., Whisstock, J. C., Boyd, S., de la Banda, M. G., Kaiserman, D., Matthews, A. Y., Bird, P. I., and Pike, R. N. (2005) *J. Biol. Chem.* **280**, 39510–39514
 18. Cwirla, S. E., Peters, E. A., Barrett, R. W., and Dower, W. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6378–6382
 19. Boyd, S. E., Pike, R. N., Rudy, G. B., Whisstock, J. C., and Garcia de la Banda, M. (2005) *J. Bioinform. Comput. Biol.* **3**, 551–585
 20. Rapp, C. M., Morgan, M. S., and Arlian, L. G. (2006) *J. Med. Entomol.* **43**, 539–542
 21. Hanquier, J., Sorlet, Y., Desplancq, D., Baroche, L., Ebtinger, M., Lefèvre, J. F., Pattus, F., Hersherberger, C. L., and Vertès, A. A. (2003) *Appl. Environ. Microbiol.* **69**, 1108–1113
 22. Wells, C. M., and Di Cera, E. (1992) *Biochemistry* **31**, 11721–11730
 23. Hwang, S. R., and Hook, V. (2008) *FEBS Lett.* **582**, 2527–2531
 24. Perry, D. K., Smyth, M. J., Stennicke, H. R., Salvesen, G. S., Duriez, P., Poirier, G. G., and Hannun, Y. A. (1997) *J. Biol. Chem.* **272**, 18530–18533
 25. Beynon, R., and Bond, J. S. (eds) (2001) *Proteolytic Enzymes*, pp. 109–110, Oxford University Press, Oxford
 26. Hall, P. K., and Roberts, R. C. (1978) *Biochem. J.* **173**, 27–38
 27. Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) *Protein Sci.* **3**, 1197–1205
 28. Scott, F. L., Fuchs, G. J., Boyd, S. E., Denault, J. B., Hawkins, C. J., Dequiedt, F., and Salvesen, G. S. (2008) *J. Biol. Chem.* **283**, 19499–19510
 29. Candi, E., Schmidt, R., and Melino, G. (2005) *Nat. Rev. Mol. Cell Biol.* **6**, 328–340
 30. McGrath, J. A., and Uitto, J. (2008) *Trends Mol. Med.* **14**, 20–27
 31. Van Neste, D. (1985) *Dermatologica* **171**, 343–348
 32. Fimiani, M., Mazzatenta, C., Alessandrini, C., Paccagnini, E., and Andreassi, L. (1997) *J. Submicrosc. Cytol. Pathol.* **29**, 105–113
 33. Prins, C., Stucki, L., French, L., Saurat, J. H., and Braun, R. P. (2004) *Dermatology* **208**, 241–243
 34. Nomura, T., Akiyama, M., Sandilands, A., Nemoto-Hasebe, I., Sakai, K., Nagasaki, A., Ota, M., Hata, H., Evans, A. T., Palmer, C. N., Shimizu, H., and McLean, W. H. (2008) *J. Invest. Dermatol.* **128**, 1436–1441
 35. Wan, H., Winton, H. L., Soeller, C., Tovey, E. R., Gruenert, D. C., Thompson, P. J., Stewart, G. A., Taylor, G. W., Garrod, D. R., Cannell, M. B., and Robinson, C. (1999) *J. Clin. Invest.* **104**, 123–133
 36. List, K., Szabo, R., Wertz, P. W., Segre, J., Haudenschild, C. C., Kim, S. Y., and Bugge, T. H. (2003) *J. Cell Biol.* **163**, 901–910

SUPPLEMENTARY DATA

Supplementary Table 1: *Amino acid sequences derived from phage display analysis for the specificity of the Sar s 3 protease. The number of times the sequence appeared is indicated in the right hand column.*

P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	Repeated
R	A	M	W	R	S	G	K	L	5
S	R	N	V	R	S	R	A	G	4
R	A	R	A	R	S	R	Q	A	3
Y	S	V	A	R	S	G	Y	V	2
T	K	L	R	R	A	Y	I	S	2
R	S	G	E	R	K	A	V	K	2
R	S	G	I	R	R	V	V	T	2
W	R	S	A	R	L	T	R	P	2
R	A	R	M	R	S	A	V	V	2
R	S	G	M	R	L	K	L	M	2
L	T	R	R	R	S	A	V	V	2
Y	A	R	F	R	S	G	G	I	2
L	Q	A	R	R	G	Y	V	V	1
S	Y	K	R	R	D	W	S	V	1
L	L	G	R	R	G	W	L	R	1
A	R	M	R	R	A	V	G	T	1
S	V	T	A	R	S	V	V	L	1
R	A	G	Y	R	T	V	V	G	1
R	S	V	A	R	M	T	L	R	1
R	S	A	G	R	F	S	T	V	1
V	R	A	R	R	G	S	N	V	1
V	R	A	R	R	S	G	N	V	1
F	P	A	G	R	S	S	N	R	1
H	V	L	S	R	S	S	G	S	1
F	V	Q	P	R	A	R	D	S	1
G	W	R	S	R	G	R	V	W	1
R	S	A	T	R	G	R	F	L	1
R	V	A	Q	R	L	R	R	R	1
R	A	G	Y	R	L	R	L	Q	1
R	A	G	G	R	G	P	R	K	1
S	I	T	A	R	S	P	R	Q	1
Y	N	R	A	R	S	P	A	S	1
R	A	A	L	R	R	N	F	K	1
L	W	E	V	R	F	M	T	D	1
R	S	V	W	R	R	M	P	R	1
A	F	E	P	R	A	L	A	L	1
V	F	L	R	R	A	L	I	R	1
R	S	G	V	R	F	L	S	T	1
S	K	G	R	R	G	L	R	W	1
R	A	G	G	R	T	L	S	K	1
L	G	F	L	R	V	L	G	R	1

L	K	W	R	R	A	I	R	L	1
I	V	K	G	R	A	G	A	V	1
L	A	K	N	R	A	G	G	M	1
L	S	L	A	R	A	G	E	V	1
P	L	A	R	R	A	G	Y	V	1
R	A	A	E	R	A	G	Q	W	1
S	V	G	P	R	A	G	E	V	1
R	A	A	L	R	F	G	E	G	1
R	V	G	A	R	G	G	G	N	1
L	A	R	A	R	L	G	I	I	1
K	L	K	V	R	M	G	Q	V	1
R	A	A	T	R	P	G	N	A	1
F	R	G	W	R	S	G	G	M	1
G	I	A	S	R	S	G	G	R	1
L	W	R	Q	R	S	G	R	K	1
R	P	S	M	R	S	G	R	A	1
T	V	W	L	R	S	G	I	M	1
T	V	R	S	R	S	G	W	Y	1
Y	A	R	V	R	S	G	R	R	1
F	I	S	A	R	V	G	R	L	1
F	P	R	A	R	W	G	R	I	1
R	S	K	M	R	A	F	T	M	1
R	S	G	S	R	V	D	K	V	1
L	L	L	S	R	A	A	L	S	1
W	Q	R	S	R	A	A	L	A	1
K	L	R	V	R	F	A	V	V	1
G	L	M	W	R	M	A	R	L	1
R	V	G	A	R	M	A	V	K	1
R	A	Y	M	R	R	A	A	V	1
F	F	R	Q	R	S	A	L	V	1
H	V	K	V	R	S	A	L	E	1
I	N	K	S	R	S	A	L	M	1
K	P	R	N	R	S	A	G	R	1
R	T	F	A	R	S	A	S	L	1
R	A	R	M	R	S	A	V	V	1
R	S	G	A	R	S	A	E	G	1
S	L	R	W	R	S	A	R	T	1
S	T	F	F	R	S	A	G	W	1
S	P	R	A	R	S	A	M	F	1
Y	N	T	G	R	S	A	G	I	1
S	I	R	S	R	T	A	E	L	1

Supplementary Table 2: $\Delta\sigma$ values for amino acids at each position in phage display data for Sar s 3. The highest values at particular positions are shown in the dark shaded boxes. Values close to or at zero were omitted for clarity.

Amino acid	1	2	3	4	5	6	7	8	9
A	-1.5	4.9	3.5	4.9		4.4	6.7	-0.1	-0.6
C									
D						-1.0	-1.0	-1.0	-1.0
E			-0.4	-0.4				1.5	-1.0
F	2.1	0.2	0.2	-0.4		1.5	-1.0	-0.4	-1.0
G	-1.0	-1.9	4.9	0.4		1.3	8.5	2.6	-0.6
H	-0.4								
I	-0.4	0.9		-1.0			-1.0	0.9	0.9
K	0.2	0.2	2.8			-1.0	-1.0	-0.4	2.1
L	1.2	-0.3	-1.1	-1.4		-1.1	-0.7	0.4	0.4
M			0.2	1.5		0.9	-0.4	-1.0	2.1
N		0.2	-1.0	-0.4			-1.0	0.9	-1.0
P	-1.9	-0.1		-1.0		-1.9	-1.0	-1.9	-1.9
Q		-0.4	-1.0	0.2				0.2	-0.4
R	7.6	-0.7	3.8	1.6	27.5	-1.4	-0.3	1.9	0.8
S	0.8	1.9	-1.8	0.4		8.3	-1.4	-1.4	-1.1
T	-1.0	-1.0	-1.0	-1.5		-1.0	-1.5	-1.0	-0.6
V	-1.0	2.6	-1.0	0.8		-1.0	-0.6	2.2	5.3
W	-0.4	0.2	-0.4	1.5		-1.0	-0.4	-1.0	0.9
Y	1.5	-1.0	-1.0	-0.4			-0.4	-0.4	-1.0

Supplementary Table 3: *Proteins in the Homo sapiens proteome predicted to be cleaved by Sar s 3. In silico* determination by PoPS, based on phage display substrate specificity data. Skin related proteins are in bold. Proteins shaded in grey are predicted to be structurally inaccessible for cleavage.

NCBI ID	Protein Name	PoPS Score
NP_005608 NP_001020242 NP_001020241	Ribosomal protein S14	73.6
NP_002217 NP_660142	Jagged 2 isoforms a and b precursor	72.6
NP_005294	G protein-coupled receptor 40	72.5
NP_004515 NP_001026973	Lethal giant larvae homolog 2 isoforms a and c	72.3
NP_997253	G protein-coupled receptor 153	71.9
NP_667338	Limbin	71.9
NP_057460	Ankyrin repeat and FYVE domain containing 1 isoform 1	71.5
NP_620686	A disintegrin-like and metalloprotease (repolysin type) with thrombospondin type 1 motif, 15 preproprotein	71.5
NP_057153	Mitochondria-associated granulocyte macrophage CSF signaling molecule	71.2
NP_076950	Nucleolar protein GU2	70.8
NP_001385	Dual specificity phosphatase 4 isoform 1	70.8
NP_003898	Eukaryotic translation initiation factor 2B, subunit 5 epsilon, 82kDa	70.6
NP_002198	Integrin, alpha 9 precursor	70.4
NP_006112	Keratin 1	70.3
NP_039256	Neuregulin 1 isoform GGF2	70.3
NP_071407	Cadherin related 23 isoform 1 precursor	70.2
NP_115828	Junctophilin 4	70.2
NP_001002909	G patch domain containing 8	69.8
NP_612456	Zinc finger protein HIT-39	69.8
NP_976324	RNA binding motif protein 12B	69.7
NP_079513	Sclerostin precursor	69.7
NP_057204 NP_899196	Ring finger protein 12	69.7
NP_258260	FCH and double SH3 domains 1	69.5
NP_932070	BH3 interacting domain death agonist isoform 1	69.5
NP_059128	prostate tumor overexpressed gene 1	69.4
NP_001026997 NP_001027003	BolA-like protein 2 isoforms a and b	69.4
NP_001034271	BolA-like protein 2B	69.4
NP_036433	Membrane associated guanylate kinase, WW and PDZ domain containing 2	69.3
NP_000921	Plasminogen activator, tissue type isoform 1 preproprotein	69.3
NP_004874 NP_053584 NP_053585 NP_053586	Neuregulin 2 isoforms 1, 2, 3 and 4	69.2
NP_000398	Glycoprotein Ib, beta polypeptide precursor	69.1
NP_579874	Death effector domain-containing DNA binding protein 2	69.1
NP_068752	Invasion inhibitory protein 45	69.1

NP_006172	Netrin 2-like	69
NP_057013 NP_660200 NP_660197 NP_660198 NP_660199	Lipoyltransferase 1	69
NP_057537 NP_909119	NCK interacting protein with SH3 domain isoforms 1 and 2	68.8
NP_055077	Leprecan-like 2	68.5
NP_056031	Histidine acid phosphatase domain containing 1	68.5
NP_001073921	G protein-coupled receptor 108 isoform 1	68.4
NP_056067	EH domain binding protein 1	68.4
NP_002007	Filaggrin	68.3
NP_112486	Queuine tRNA-ribosyltransferase 1 (tRNA-guanine transglycosylase)	68.3
NP_710159	HtrA serine peptidase 4	68.3
NP_002324	Low density lipoprotein receptor-related protein 3	68.3
NP_588615	Zinc finger protein 618	68.2
NP_067051	Serine arginine-rich pre-mRNA splicing factor SR-A1	68.2
NP_004549	Neurturin precursor	68
NP_065743	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetyl-galactosaminyltransferase-like 1	68
NP_619635	Glutamate receptor, ionotropic, N-methyl-D-aspartate 3B	67.9
NP_001073277	Chromosome 1 open reading frame 37 duplicate copy	67.9
NP_057496	Insulin-like growth factor 2 antisense	67.9
NP_006535	SEC10 protein	67.8
NP_067638 NP_671737	Transient receptor potential cation channel, subfamily V, member 4 isoforms a and b	67.8
NP_619649	Necdin-like 2	67.8
NP_061747 NP_115268	Protocadherin gamma subfamily B, 3 isoforms 1 and 2 precursor	67.7
NP_055159	Four jointed box 1	67.7
NP_004658	Hect domain and RLD 2	67.5
NP_573439 NP_573438 NP_005695	Protein tyrosine phosphatase, receptor type, U isoform 1 and 3 precursor	67.4
NP_733752	Transcription factor-like protein 4 isoform gamma	67.4
NP_061197 NP_542435 NP_542437 NP_542436	Transient receptor potential cation channel, subfamily V, member 1	67.3
NP_004783	Peptidyl-prolyl isomerase G (cyclophilin G)	67.3
NP_067054 NP_001074012	Chromosome 19 open reading frame 29	67.3
NP_620305	SON DNA-binding protein isoform F	67.2
NP_036608	Ventral anterior homeobox 2	67.2
NP_037489 NP_001015508	Purine-rich element binding protein G isoform A and B	67.1
NP_065948	Hyperpolarization activated cyclic nucleotide-gated potassium channel 3	67.1

NP_076438 NP_001035187 NP_001035188 NP_001035186	Motile sperm domain containing 3 isoform a and b	67
NP_758454	Methylmalonic aciduria type A	67
NP_001013597	Diacylglycerol O-acyltransferase 2-like 3	67
NP_002964	Ataxin 2	66.9
NP_055093	Heat shock 70kDa protein 4-like	66.7
NP_115618	Plexin A1	66.6
NP_036286	Eukaryotic translation initiation factor 2C, 2	66.6
NP_114429	Coiled-coil domain containing 8	66.6
NP_690007 NP_071907 NP_690008	Hypoxia-inducible factor-3 alpha isoform a, b and c	66.5
NP_000038	Arylsulfatase E precursor	66.5
NP_001923	Palmitoylated membrane protein 3	66.5
NP_000983	Ribosomal protein L29	66.5
NP_000937	DNA primase small subunit, 49kDa	66.4
NP_115587	Lysyl oxidase-like 4 precursor	66.4
NP_055707	RNA-binding motif protein 16	66.4
NP_055390	Schwannomin interacting protein 1	66.4
NP_997309	Forkhead box I2	66.4
NP_005617	Splicing factor, arginine/serine-rich 4	66.4
NP_005080	U2 small nuclear RNA auxiliary factor 1-like 2	66.4
NP_705843	Ral-GDS related protein Rgr	66.3
NP_073562	Fanconi anemia, complementation group F	66.3
NP_055575	TatD DNase domain containing 2	66.3
NP_036427	Calcineurin binding protein 1	66.3
NP_296375	Coiled-coil domain containing 120	66.3
NP_004406 NP_001008844	Desmoplakin isoforms I and II	66.2
NP_699165 NP_878315	Scavenger receptor class F, member 2 isoforms 1 and 2	66.2
NP_000456	BRCA1 associated RING domain 1	66.2
NP_001285	Cleft lip and palate associated transmembrane protein 1	66.2
NP_001979	Envoplakin	66.2
NP_003924	BAI1-associated protein 3	66.1
NP_005327 NP_976221	High density lipoprotein binding protein	66.1
NP_001073893	Jumonji domain containing 3	66.1

Supplementary Table 4: *Sites within filaggrin, predicted to be cleaved by Sar s 3*. Using the substrate specificity data generated by the phage display, PoPS was used to predict the most likely cleavage sites within filaggrin (NP_002007.1). 'R' in bold represents the P1 position of the cleavage site.

Location	Site	Score
R3379-S3380	RSG G RSGRS	68.3
R3375-S3376	SARD R SGGR	68.1
R461-S462	RSG E RSGRS	67.7
R782-S783	SARD R SGER	66.9
R2407-S2408	RSAG R SGRS	66.2
R3703-S3704	RSAG R SGRS	66.2
R2727-S2728	SARG R SGET	66.1
R3051-S3052	SARG R SGET	66.1
R786-S787	RSG E RSRRS	58.9
R1110-S1111	WSG G RSGRS	58.5