The *Streptococcus equi* prophage-encoded protein SEQ2045 is a hyaluronan-specific hyaluronate lyase that is produced during equine infection

Anna-Marie Lindsay,¹ Meng Zhang,¹ Zoe Mitchell,² Matthew T. G. Holden,³ Andrew S. Waller,² Iain C. Sutcliffe¹ and Gary W. Black¹

¹School of Applied Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
²Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk CB8 7UU, UK
³The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

**Correspondence**
Gary W. Black
gary.black@northumbria.ac.uk

INTRODUCTION

*Streptococcus equi* is a Lancefield group C streptococcus. It is principally a pathogen of equines and causes the widespread and highly contagious disease ‘strangles’, which manifests as a nasopharyngeal infection that infiltrates the lymph nodes of the head and neck (Harrington et al., 2002; Waller & Jolley, 2007). Bacterial multiplication causes the formation of abscesses which rupture and drain through sinus tracts to the upper respiratory mucosa or through the skin (Timoney, 2004; Waller & Jolley, 2007). Regrettably, the development of an effective vaccine has been slow (Waller & Jolley, 2007). A better understanding of *S. equi* pathogenesis is essential if improved vaccines are to be developed. Likely candidates for novel vaccine targets are the cell surface and extracellular virulence factors produced by *S. equi*.

The closely related human pathogen, *Streptococcus pyogenes*, produces many virulence factors in common with *S. equi*, for example, superantigens (Proft et al., 2003) and a hyaluronic acid (HA) capsule (Bisno et al., 2003; Harrington et al., 2002). Interestingly, several of these are encoded by genes carried by temperate bacteriophages that have integrated into the bacterial chromosome as prophages. One such prophage, SF370.1, from *S. pyogenes* SF370, encodes the pyrogenic mitogen SpeC and DNase virulence factors (Canchaya et al., 2002). A hyaluronate lyase (Hyl), termed HylP1, is also encoded by SF370.1. Typically, Hyls have been shown to be secreted and are capable of degrading HA and chondroitins (Baker & Pritchard, 2000; Hynes & Walton, 2000; Nukui et al., 2003; Pritchard et al., 1994; Shain et al., 1996), which are major components of mammalian connective tissue, thus possibly facilitating the invasion of the bacteria and their toxins (Hynes & Walton, 2000). However, HylP1 is thought not to be involved in the degradation of connective tissue, as it is unlikely to be secreted, since it lacks a signal peptide, and it only degrades HA and does not process chondroitins (Smith et al., 2005). Additionally, HylP1 is sequence-distinct from the Hyls that degrade chondroitins and has been shown to contain a triple-stranded β-helix, a structural fold only found in bacteriophage tail-fibre proteins (Smith et al., 2005). It has therefore been postulated that the role of HylP1 is the penetration of the HA capsule of streptococci during phage infection (Hynes et al., 1995), facilitating bacterial lysogenization.

In an attempt to understand the role of prophage-encoded Hyls further, we have analysed SEQ2045, the closest HylP1 homologue in *S. equi* 4047, as this organism also produces a HA capsule (Harrington et al., 2002). We have shown that SEQ2045 is biochemically very similar to HylP1 and that antibodies to SEQ2045 are produced by infected horses.

Abbreviations: CDS, coding sequence; HA, hyaluronic acid; Hyl, hyaluronate lyase.

A supplementary figure illustrating the purification of N-terminally hexahistidine-tagged SEQ2045 is available with the online version of this paper.
METHODS

Cloning and mutagenesis. The strain of S. equi used, 4047, was isolated in 1990 by the Animal Health Trust from an abscess from a New Forest pony with strangles (Kelly et al., 2006) and is virulent in horses (Hamilton et al., 2006). This strain is the subject of the S. equi genome-sequencing project at the Sanger Institute (http://www.sanger.ac.uk/Projects/S_equi/). Genomic DNA from S. equi 4047 was isolated using a DNeasy extraction kit (Qiagen). Primer pair 5′-CATATgTCAAAAgAaGgATgCTACgC3′/5′-gATCCTATTTTgTTg-ATTgAgTTgTTTTTACC3′ was used to amplify ORF seq2045 from this genomic DNA using the KOD hot-start polymerase kit (Merck Biosciences). The amplified product was ligated into pCR-Blunt (Invitrogen) then removed from pCR-Blunt on a NdeI–XhoI fragment and ligated into pET-28a (Merck Biosciences) to produce pSEQ2045. Plasmids pD171A and pY183A, which encode the SEQ2045 catalytic mutants D171A and Y183A respectively, were generated from pSEQ2045 using the QuikChange site-directed mutagenesis kit (Agilent Technologies) and mutagenic primer pairs 5′-gAgACgCgACAACA3′ and 5′-gAgTTTTTTTAAC3′, respectively.

Protein expression. N-terminally hexahistidine-tagged SEQ2045 and SEQ2045 catalytic mutants, D171A and Y183A, were produced in and purified from Escherichia coli BL21(DE3) (Merck Biosciences) carrying pSEQ2045, pD171A and pY183A, respectively. The protocol described by Charnock et al. (2001) was followed, except that cultures were grown at 30 °C post-induction, and concentration and exchange of the purified proteins into 18.2 M 2 M NaCl was achieved using 10 kDa cut-off concentrator units (Viva Science). The purity of SEQ2045, D171A and Y183A was judged by SDS-PAGE and Coomassie blue staining (Laemmli, 1970). MS analysis of a tryptic digest of purified SEQ2045 was used to confirm the identity of the protein according to the methods and criteria of Zhang et al. (2007).

Biochemical assays. Enzyme assays were performed as described by Smith et al. (2005) with reaction conditions of 2 mg substrate ml^{-1}, 8 mM CaCl2, 50 mM CH3COONa (pH 7.5) unless otherwise stated. The range of HA concentrations used for determining Kcat and Km values was 0.25 to 2 mg ml^{-1}. Goodness-of-fit statistical analysis of the linear trend lines of the resulting double-reciprocal plots produced R2 values of ≥0.9042. The following buffers were used to determine the pH optimum: CH3COONa (pH 5.0 and 5.5), CH3COONa (pH 6.0, 6.5, 7.0 and 7.5) and glycine (pH 8.0). The effect of temperature on activity was determined by performing assays at 27, 37, 47, 57, 67 and 77 °C. The temperature at which the rate of the reaction was reduced to 50% was determined by incubating assays at 27.7–87.5 °C for 20 min and then assaying at 37 °C. The effects of divalent cations on activity were determined using the standard reaction conditions, except that 8 mM CaCl2 was replaced with BaCl2, CoCl2, CuCl2, MgCl2, MnCl2, NiCl2, SrCl2 or ZnCl2. Inhibition studies with ascobic acid were performed using the standard reaction conditions, except that ascobic acid was added to final concentrations of 2–20 mM. High-performance anion-exchange chromatography was performed according to the method of Lauder et al. (2000).

Western blot analysis. Western blotting was performed as described by Hamilton et al. (2000) with 7 µg purified SEQ2045 per lane. The preimmune and strangles convalescent sera were from naturally infected ponies and were prepared at the Animal Health Trust. Control sera (Hyperimmune-RE and donor herd serum) were from Veterinary Immunogensics and Sigma, respectively. The secondary antibody, anti-horse IgG (whole molecule)–alkaline phosphatase conjugate (Sigma) was produced in rabbit.

RESULTS AND DISCUSSION

SEQ2045 from S. equi 4047 is similar to HylP1 from S. pyogenes SF370

A BLAST search of the S. equi 4047 genome (http://www.sanger.ac.uk/Projects/S_equi/) with HylP1 from S. pyogenes SF370 revealed a single homologous protein-coding sequence (CDS), seq2045 (E probability $7.6 \times 10^{-128}$). Proteins SEQ2045 and HylP1 share 75% amino acid identity (Fig. 1). This high identity score suggested that SEQ2045, like HylP1, contains a triple-stranded β-helix, a structural fold only found in phage tail-fibre proteins (Smith et al., 2005). SEQ2045 also contains ten internal repeats of the collagen-like motif Gly-X-Y within the N-terminal region of the protein (Fig. 1). This collagen-like motif has been found in many streptococcal proteins, including several Hyls (Ferretti et al., 2001; Stern & Stern, 1992), but not HylP1. Consistent with this, analysis of the genomic context of SEQ2045 indicated it to be carried within a phage designated ϕSeq.4 in the genome annotation. Thus we conclude that SEQ2045 is part of a phage tail.

A TBLASTX comparison of ϕSeq.4 and prophage SF370.1 from S. pyogenes shows that these prophages have mosaic similarity, with the regions encoding the Hyls being the most conserved (Fig. 2). There is also partial similarity elsewhere, but it is low level (Fig. 2). Most of the prophage CDSs are divergent and the orientation of the CDSs on both phages is conserved in terms of strand bias (Fig. 2), as in most other S. pyogenes prophages. The only region in the phages that demonstrates a marked difference in gene orientation is the region carrying the cargo genes at the left end of the alignment (Fig. 2). The S. equi prophage is notable as it also encodes proteins SEQ2036 and SEQ2037 (Fig. 2), which are 100% identical to the mitogens SeeH and Seel from S. equi CF32 (Artiushin et al., 2002).

Bioinformatic analyses predicted that SEQ2045 is a non-secreted protein (data not shown). SEQ2045 and the SEQ2045 catalytic mutants, D171A and Y183A, were obtained in large quantities (yields ~40 mg l^{-1}) as 42 kDa N-terminally hexahistidine-tagged proteins that were readily purified to homogeneity (for example see Supplementary Fig. S1, available with the online version of this paper). Three non-redundant SEQ2045 peptides were identified by MS, confirming the identity of the purified protein (data not shown). The temperature-activity optimum was 37 °C (Fig. 3A) and the temperature at which the rate of the reaction was reduced by 50% was 44 °C (Fig. 3B). The pH-activity optimum for SEQ2045 against HA was ~7 (Fig. 3C). SEQ2045 was shown to have no requirement for any specific divalent ions tested and was not inhibited by ascobic acid at concentrations up to 20 mM (data not shown). Kinetic analysis of SEQ2045 on HA substrate yielded a $K_{cat}$ of 7.61 ± 0.63 s^{-1} and a $K_{m}$ of 0.29 ± 0.05 mg ml^{-1} (equivalent to 0.71 ± 0.13 mM, when the concentration of HA is expressed as the concentration of its disaccharide unit based on its molecular mass;
The \( k_{\text{cat}} \) values show notable similarity to the \( k_{\text{cat}} \) determined for HylP1, 7.23 s\(^{-1}\), and another Hyl, HylP, from bacteriophage H4489A from S. pyogenes, 4.9 s\(^{-1}\) (Baker et al., 2002; Smith et al., 2005). These turnover rates are in stark contrast to a sequence-distinct Hyl from Streptococcus agalactiae, which has a \( k_{\text{cat}} \) of around 8600 s\(^{-1}\) on HA (Pritchard et al., 1994). However, with regard to the binding capacity of SEQ2045 to HA, the calculated \( K_m \) value is about fivefold lower than that of HylP1 (3.69 ± 0.06 mM) and about eightfold higher than that of a sequence-distinct Hyl from Streptococcus pneumoniae (0.09 ± 0.03 mM) (Nukui et al., 2003).

High-performance anion-exchange chromatography analysis of SEQ2045 digestion products of HA showed that end-products were predominantly the \( A_4,5 \) unsaturated

Fig. 3D). The \( k_{\text{cat}} \) values show notable similarity to the \( k_{\text{cat}} \) determined for HylP1, 7.23 s\(^{-1}\), and another Hyl, HylP, from bacteriophage H4489A from S. pyogenes, 4.9 s\(^{-1}\) (Baker et al., 2002; Smith et al., 2005). These turnover rates are in stark contrast to a sequence-distinct Hyl from Streptococcus agalactiae, which has a \( k_{\text{cat}} \) of around 8600 s\(^{-1}\) on HA (Pritchard et al., 1994). However, with regard to the binding capacity of SEQ2045 to HA, the calculated \( K_m \) value is about fivefold lower than that of HylP1 (3.69 ± 0.06 mM) and about eightfold higher than that of a sequence-distinct Hyl from Streptococcus pneumoniae (0.09 ± 0.03 mM) (Nukui et al., 2003).

High-performance anion-exchange chromatography analysis of SEQ2045 digestion products of HA showed that end-products were predominantly the \( A_4,5 \) unsaturated

---

**Fig. 1.** Box-shaded CLUSTAL W (1.83) sequence alignment of SEQ2045 from S. equi 4047 and HylP1 from S. pyogenes SF370. Black background, identical residues; grey background, similar residues; white background, different residues. Identity, 75%. The arrows delineate the ten internal repeats of the collagen-like motif Gly-X-Y.

**Fig. 2.** Pairwise comparison of prophage SF370.1 from S. pyogenes SF370 and prophage \( \Phi \)Seq.4 from S. equi 4047 displayed using the Artemis Comparison Tool (ACT) (Carver et al., 2005). The red bars linking each sequence represent similarity matches identified by tBLASTx analysis (Altschul et al., 1990). The darker the red, the higher the similarity. The \( \Phi \)Seq.4 CDSs are from DNA bases 2064760 to 2105593 of the S. equi 4047 genome (http://www.sanger.ac.uk/Projects/S_equi/), encoding SEQ2036 through to SEQ2089. The SF370.1 CDSs are as specified in Canchaya et al. (2002). CDSs likely to belong to the same functional module are marked with the same colour. CDSs encoding the mitogens SeeH (SEQ2036) and Seel (SEQ2037) (Artiushin et al., 2002), and the Hyl (SEQ2045), are indicated with arrows.
hyalurono-hexasaccharide and the Δ4,5 unsaturated hyalurono-octasaccharide, with small amounts of Δ4,5 unsaturated hyalurono-tetrasaccharide, Δ4,5 unsaturated hyalurono-decasaccharide and longer oligosaccharides (Fig. 4). Early time points showed that a wide range of Δ4,5 unsaturated oligosaccharides are produced, indicating that SEQ2045 has an endo-acting mode of digestion. The wide size range of oligosaccharides released is the result of the random internal cuts of HA by SEQ2045.

Unlike the sequence-distinct Hyls, SEQ2045 showed no catalytic activity against chondroitin 4-sulphate, chondroitin 6-sulphate and dermatan sulphate (data not shown) and was not inhibited by ascorbic acid, confirming that SEQ2045, like HylP1 from *S. pyogenes* SF370, is functionally very different from the sequence-distinct Hyls.

The SEQ2045 catalytic mutants, D171A and Y183A, like the analogous HylP1 catalytic mutants, D137A and Y149A (Smith et al., 2005), had little activity against HA (8 and 13 % respectively compared to SEQ2045; data not shown), verifying the role of Asp171 and Tyr183 in catalysis and confirming SEQ2045 to be a lyase, rather than a hydrolase (Smith et al., 2005).

**Fig. 3.** Graphs used in the determination of the biochemical and biophysical properties of SEQ2045. (A) Plot of temperature against rate of reaction for temperature-activity optimum determination; (B) plot of pre-incubation temperature against rate of reaction at 37 °C for determining the temperature at which the rate of the reaction was reduced to 50 % (i.e. thermal midpoint); (C) plot of pH against rate of reaction for the pH-activity optimum determination; (D) double-reciprocal plot (the inset is the corresponding non-reciprocal plot). Means ± SD (n=3) (all panels).

The only predicted hyaluronate lyase activity of *S. equi* 4047 is encoded by *seq2045*

A pairwise DNA sequence comparison of the region of *S. equi* 4047 genome that encodes the sole recognizable nonprophage-encoded Hyl against the highly similar region (97 % identity) of the close relative *Streptococcus zooepidemicus* H70 (subject of the *S. zooepidemicus* genome-sequencing project at the Sanger Institute, http://www.sanger.ac.uk/Projects/S_zooepidemicus/) revealed an internal deletion of TCTC within the *S. equi* genome sequence (data not shown). This frame-shift results in a pseudogene, *seq1480*, that encodes a C-terminally truncated product of only 217 amino acids, ~20 % of the size of the *S. zooepidemicus* H70 putative nonprophage-encoded Hyl. The truncated protein lacks the catalytic residues and substrate-binding sites of the full-length *S. zooepidemicus* enzyme and will therefore not be active against HA. It is unclear if the adjacent pseudogene, *seq1479*, which encodes the C-terminal portion of the Hyl, is likely to be expressed, but as the protein encoded by this ORF lacks a signal peptide it is not predicted be an extracellular Hyl (data not shown). Interestingly, truncated Hyl sequences have also
been reported in a significant proportion of *Streptococcus suis* strains and these strains lacked detectable Hyl activity (King et al., 2004).

Additionally, bioinformatic analysis of the three other prophage sequences present in *S. equi* 4047 (φSeq.1–φSeq.3) revealed that they do not encode intact Hyls, only other collagen-like motif-containing proteins (data not shown). Consequently, it is highly likely that the only predicted Hyl activity of *S. equi* strain 4047 is encoded by *seq2045*.

**SEQ2045 is produced during *S. equi* infection of horses**

A post facto serological investigation of protein expression in vivo was undertaken using serum taken from a horse following natural *S. equi* infection (convalescent serum). Western blot analysis clearly showed that equine infection with *S. equi* initiated an immunogenic response against SEQ2045 that was not observed with control sera (commercial horse sera from donor herds; data not shown). This indicated that specific antibodies are raised against SEQ2045 during the course of *S. equi* infection. Subsequently, results obtained with three pre- and post-infection matched pony serum samples (Fig. 5) confirmed that antibodies are made to SEQ2045 specifically in response to *S. equi* infection. These data are in excellent agreement with the findings of Artiushin et al. (2002), which showed that the prophage-encoded superantigens SeeH and SeeI (Fig. 2) are recognized by convalescent sera from horses that have recovered from *S. equi* infection.

**Conclusions**

SEQ2045 is a Hyl that is incapable of degrading HA further than a Δ4,5 unsaturated hyalurono-tetrasaccharide and has no activity against chondroitin or dermatan sulphates. As such, it is unlikely that the principal role of the enzyme is the dissemination of *S. equi* through equine tissue and its primary role is most likely to be in phage penetration through the host capsule. However, it is clearly expressed during infection of horses, as herein demonstrated serologically, and it is notable that phage particles from lysates of *S. equi* have been demonstrated to possess *in vitro* Hyl activity (Timoney et al., 1991). Thus phage induction in vivo may itself represent an additional mechanism leading to host connective tissue damage. It has also been noted that *in vitro* secreted Hyl activity was absent from culture supernatants of four out of ten *S. equi* clinical isolates (Sting et al., 1990). Thus the extent to which *S. equi* is able to degrade host HA may vary from strain to strain. In some strains, as exemplified by strain 4047, the only source of Hyl activity is likely to be prophage-encoded, whereas in other strains both prophage- and nonprophage-encoded Hyls may be present.

In light of the results presented here it may therefore be necessary to re-evaluate the role of prophage-encoded Hyls from *S. pyogenes*, since it is likely that they are expressed during human infections, as phage mobilization has been...
demonstrated both in vitro and in a mouse infection model (Broudy & Fischetti, 2003) and antibody response to S. pyogenes phage Hyls has been reported in patients recovering from acute glomerulonephritis (Halperin et al., 1987). For both S. pyogenes and S. equi, phage induction in vivo may thus lead to the production and release of both prophage-encoded virulence factors, e.g. DNase (Broudy et al., 2002) and superantigens, and phage structural proteins with the capacity to damage connective tissues (i.e. HylPI and SEQ2045).

We have recently reported the importance of studying S. equi infections of the horse in the natural host (Hamilton et al., 2006). Experimental infections of the horse with S. equi may therefore be a useful model for studying the role of Hyls in human S. pyogenes infections. Moreover, as the study of in vivo interactions between phages and pathogenic bacteria remains an important new frontier for phage biologists (Chibani-Chennoufi et al., 2004), S. equi and the horse may provide an important system for studying the contribution of phages to bacterial virulence in the natural host.

REFERENCES


