Published in final edited form as:

Ticks Tick Borne Dis. 2011 December ; 2(4): 219–224. doi:10.1016/j.ttbdis.2011.09.004.

Distinctive amino acid composition profiles in salivary proteins of the tick *Ixodes scapularis*

Austin L. Hughes* and Robert Friedman

Dept. of Biological Sciences, University of South Carolina, Columbia SC 29208, USA

Abstract

Bioinformatic analysis of the amino acid composition of proteins of the tick *Ixodes scapularis* showed that, in comparison to other secreted proteins, salivary proteins in general have higher frequencies of polar residues and lower frequencies of the non-polar residues leucine and valine. Computer prediction of linear B-cell epitopes showed that polar residues were associated with the presence of high-quality epitopes and that tick salivary proteins included significantly more proteins with predicted high-quality epitopes than did other secreted proteins. The results provided no evidence that salivary proteins as a whole have evolved characteristics minimizing their antigenicity to the vertebrate host. Certain salivary proteins may indeed have evolved low antigenicity, but the *I. scapularis* sialome include at least some apparently antigenic proteins that might be tested experimentally to determine whether they would be suitable candidates for anti-tick vaccines.

Keywords

B-cell epitope prediction; *Ixodes scapularis*; Salivary proteins; Secreted proteins; Sialome; Vaccine

Introduction

The saliva of ticks contains numerous compounds with diverse biological functions, including disabling of the vertebrate host's clotting mechanisms, suppressing host inflammatory responses, and minimizing the host's immune recognition of the tick (Valenzuela, 2004; Wang and Nuttall, 1999). Many of the salivary constituents are proteins, and recent years have seen increased interest in the set of salivary proteins or "sialome" of ticks and other blood-feeding arthropods (Ribeiro and Francischetti, 2003). Salivary proteins are of interest both because they represent an untapped source of potential pharmacological agents (Hovius et al., 2008) and because they provide candidate components for anti-tick vaccines (Mulenga et al., 2000; Xu et al., 2005). Moreover, components of the tick saliva have been shown to play important roles in the transmission of pathogens for which the tick serves as a vector. For example, the salivary protein Salp15 of ticks in the genus *Ixodes* protects *Borrelia burgdorferi*, the spirochete causing Lyme disease, from complement-mediated killing (Schuijt et al., 2008).

© 2011 Elsevier GmbH. All rights reserved.

*Corresponding author: Dept. of Biological Sciences, Coker Life Sciences Building, 715 Sumter St., University of South Carolina, Columbia SC 29208, USA. Tel.: +1 803 777-9186; fax: +1 803 777-4002. austin@biol.sc.edu (A. Hughes) .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Empirical studies have provided evidence of antigenicity of certain tick salivary proteins (Das et al., 2001; Mulenga et al., 2000; Narasimhan et al., 2007; Schuijt et al., 2011). However, attempts to develop vaccines based on tick salivary proteins have so far been unsuccessful, and it has been hypothesized that this result may at least in part reflect relatively low overall immunogenicity of salivary proteins to the vertebrate host (Mulenga et al., 2000). Indeed, it has been proposed that tick salivary proteins may have evolved characteristics producing low antigenicity as a mechanism to prevent a host immune response against tick saliva (Tellam et al., 1992). In addition, as noted by Mulenga et al. (2000), the fact that a protein induces an antibody response, does not necessarily mean that this will lead to a protective response; and some tick proteins that induce immune responses may have characteristics that prevent the development of protective immunity on the part of the host.

The low immunogenicity of tick salivary proteins led to the idea of a vaccine based on so-called “concealed antigens”; i.e., tick proteins that are not normally exposed to the host immune system during tick infestation and therefore are unlikely to have evolved immune-evasive characteristics (Willadsen, 1987). One concealed antigen, the gut protein Bm86, has been used as the basis of a vaccine against the cattle tick *Boophilus microplus* (Willadsen et al., 1989). On the other hand, there are reasons why it may still be desirable to develop anti-tick vaccines based on salivary antigens or based on a combination of salivary and concealed antigens (Mulenga et al., 2000). Since vaccines based on concealed antigens may not prevent feeding, they may not be effective in preventing the spread of tick-borne diseases.

Because of extensive previous experimental work, the draft genomic sequence of *Ixodes scapularis*, the main vector of *B. burgdorferi* in the eastern United States, includes a well-annotated set of salivary proteins (Ribeiro et al., 2006; Valenzuela et al., 2002; Xu et al., 2005). On the hypothesis that salivary proteins have evolved features that minimize their antigenicity to the vertebrate host, one would predict that these proteins should have distinctive characteristics setting them apart not only from most other proteins, but even from non-salivary secreted proteins, since the latter are not subject to selective pressures arising from host immune recognition. Here, we use bioinformatic methods to explore the biochemical characteristics of these proteins in order to identify characteristics that distinguish salivary proteins from other predicted secreted proteins and other proteins of *I. scapularis*. By relating features of amino acid composition to computational prediction of the occurrence of B-cell epitopes, we test the hypothesis that at least some salivary proteins have evolved unique sequence features that reduce antigenicity to the vertebrate host. Furthermore, we use the same analyses to predict which *Ixodes* salivary proteins are likely to be antigenic and therefore might be viable candidates for future empirical investigation into their potential as vaccine candidates.

Materials and methods

Sequence data

I. scapularis protein sequences (N=20,486) with mapped “supercontig” positions were obtained from VectorBase (version IscaW1, CpipJ1.2; Lawson et al., 2009). We filtered out predicted proteins lacking a predicted start site. To predict the proteins which are secreted from the cell, we employed 2 algorithms to find a signal peptide sequence and a transmembrane (TM) helical domain. The predicted secreted protein set were those proteins with a predicted signal peptide, but no predicted TM domain. We used the SignalP software (Bendtsen et al., 2004) to predict the presence of a signal peptide. We scored a signal peptide as present when both of the following conditions were met: (i) the “D-score” (or discrimination score) used by the neural network method of SignalP indicated presence of a

signal peptide; and (ii) the probability of a signal peptide from the Hidden Markov Model procedure in Signal P was at least 95%. After the predicted signal peptide was cleaved in silico from the sequence, the TM domains were predicted by the TMHMM software (version 2.0) at the web server <http://www.cbs.dtu.dk/services/TMHMM> (Krogh et al., 2001).

Secreted salivary proteins were identified by BLASTX search (in all possible reading frames) against the putative secreted salivary proteins in the EST database of Ribeiro et al. (2006). This procedure was implemented in a Perl-based bioinformatic pipeline, and the BLASTX searches were conducted between the VectorBase protein sequences and the “contiguous sequences” of the EST database. The output was parsed, so the VectorBase gene names were associated by sequence identity to a “contiguous sequence” name. In some cases, the contiguous sequence contained homology to multiple sequences, so we enforced a rule whereby we chose only the VectorBase gene with the best match to the contiguous sequence.

We assembled the protein families of *I. scapularis* by BLASTCLUST (Altschul et al., 1997), using default values except for a 40/60 criterion where 40% is the minimum amino acid similarity between any 2 matching sequences across at least 60% of their sequence lengths. Once all pairs of homologous sequences were identified, the procedure clustered these into families by a single-linkage method.

Prior to analysis of protein sequences, we cleaved in silico the signal peptide from all proteins for which a signal peptide was predicted. In a few cases, the software predicted an unrealistically long signal peptide. We excluded from analyses all proteins for which the predicted signal peptide was greater than 50 amino acid residues in length (N=78, none of them salivary proteins). For the predicted secreted proteins (N=993), the median length of the predicted signal peptide was 22.00 residues [23.31±0.18 (mean ± S.E.)]. Computation of the isoelectric point (pI) and the molecular mass (average mass) of proteins was performed by an available Perl script (AG Evdokimov; <http://www.xtals.org>).

Statistical methods

Based on protein family prediction, we analyzed amino acid composition in 10,947 proteins encoded by single-member families (singletons) in *I. scapularis*. We analyzed single-member families for these analyses in order to avoid statistical problems of pseudo-replication based on analysis of evolutionarily related proteins. Of these proteins, 226 were classified as salivary proteins, 922 were classified as other secreted proteins, and 9799 were classified as other proteins. Of the 226 salivary proteins, 71 (31.4%) were classified as secreted proteins by the criteria of possessing a predicted signal peptide, but not a TM domain, whereas the rest lacked these criteria. The latter may have been misclassified due to inaccuracy in the prediction of signal peptides, or they may represent proteins secreted by a non-classical pathway; i.e., a pathway not involving a signal peptide (Nickel, 2003).

Using the singleton data set, we computed the proportion of each of the 20 amino acids in each protein and applied principal components analysis (PCA) to the correlation matrix in order to reduce the dimensionality of these 20 variables. In order to provide a physical interpretation of principal components (PCs), we examined correlations between the PC loadings on individual amino acids with known chemical properties of the amino acids obtained from the ProtScale tool (<http://expasy.org/tools/protscale.html>) of the ExPASy Proteomics Server (Gasteiger et al., 2003). Amino acid properties with significant correlations included volume (Grantham, 1974), polarity (Grantham, 1974), and P_{α} , the tendency to form α -helices (Deléage and Roux, 1987). Average values of each of these quantities were computed for each individual protein.

We used the BCPREDS software (EL-Manzalawy et al., 2008) to predict linear B-cell epitopes in proteins. We searched for epitopes of 20 amino acids in length in a sliding window along the protein sequence. In these analyses, we excluded proteins less than 20 amino acids in length. The software assigns a score (0-1.0) reflecting the quality of the prediction. Salivary and other secreted proteins were classified in terms of predicted antigenicity as low (no predicted B-cell epitope with score ≥ 0.75), intermediate, and high (at least one predicted epitope with score ≥ 0.99). A predicted epitope with a score of 0.99 or better was defined as a “high-quality epitope”. Because some variables analyzed were not normally distributed, non-parametric statistical methods were used to test for differences among functional groups of proteins (Hollander and Wolfe, 1973). In preliminary analyses, parametric tests on means were also used; however, since the results were substantially the same, only the results of non-parametric tests are reported here.

Results

General protein characteristics

We analyzed amino acid sequence properties of 10,947 proteins encoded by single-member families (singletons) in *I. scapularis* (Supplementary Table S1). Similar proportions of salivary proteins (86/226 or 38.1%), other secreted proteins (336/922 or 36.4%), and other proteins (3755/9799 or 38.3%) were acidic. Median molecular mass differed significantly among salivary (18.7 kDa), other secreted (12.6 kDa), and other (21.3 kDa) proteins (Kruskal-Wallis test; $P < 0.001$). The distribution of molecular masses among salivary proteins was highly positively skewed (skewness = 2.68). This positive skewness resulted from the fact that several of the proteins in the salivary set had very high molecular masses, including 4 proteins greater than 200 kDa in mass. The largest of these (ISCW00929; molecular mass = 250.95 kDa) was shown by BLAST search to be homologous to a secreted protein of the human epididymus (Li et al., 2010).

Principal components analysis

We applied principal components analysis to the correlation matrix of 20 variables representing the proportion of the 20 common amino acids in the 10,947 singleton proteins. The first 3 principal components (PC1–PC3) accounted for 30.7% of the overall variance; the respective percentages of the variance accounted for by the first 3 principal components were 13.8% (PC1), 9.8% (PC2), and 7.1% (PC3). The coefficients (loadings) of PC1–PC3 on the proportions of the different amino acids represented complex patterns of contrasts, and these patterns showed correlations with known chemical properties of the amino acids (Fig. 1). PC1 coefficients were significantly negatively correlated with amino acid residue volume ($r = -0.574$; $P = 0.008$; Fig. 1A). Similarly, the PC1 scores for the 10,947 proteins were significantly negatively correlated with the average volume of the amino acid residues in the protein ($r = -0.786$; $P < 0.001$). One reason that the correlation between PC1 and residue volume was not stronger was the fact that one relatively large residue, arginine, showed a relatively low PC1 coefficient, along with the 4 small residues glycine, alanine, proline, and serine (Fig. 1A). Therefore, we used the proportion of these 5 residues (GAPSR) as an approximate proxy for PC1. For the 10,947 proteins, GAPSR was positively correlated with PC1 score ($r = 0.862$; $P < 0.001$).

PC2 coefficients were positively correlated with amino acid residue polarity ($r = 0.783$; $P < 0.001$; Fig. 1B). Examination of the PC2 coefficients suggested that this PC represented mainly a contrast between the highly polar residues aspartic acid, glutamic acid, glutamine, and lysine with the other residues. Therefore, we used the proportion of these 4 residues (DEKQ) as a proxy for PC2. For the 10,947 proteins, DEKQ was positively correlated with PC2 score ($r = 0.904$; $P < 0.001$).

PC3 coefficients were positively correlated with P_{α} , the tendency to form α -helices ($r=0.525$; $P=0.018$; Fig. 1C). Examination of the PC3 coefficients suggested that this PC represented mainly a contrast between leucine and valine, on the one hand, and all other residues. Therefore, we used the proportion of these 2 residues (LV) as a proxy for PC3. For the 10,947 proteins, DEKQ was positively correlated with PC2 score ($r=0.696$; $P<0.001$).

We used canonical correlation to further explore the relationship between the 2 sets of variables measured on each of the 10,947 individual proteins: (i) scores on PC1, PC2, and PC3; and (ii) values of GAPSRS, DEKQ, and LV. The canonical correlation was 0.948 ($P<0.001$). For each set of variables, the variance extracted was 100%, implying that the canonical variables accounted for all of the variance in each set.

Median scores for PC1, PC2, and PC3 differed significantly among salivary proteins, other secreted proteins, and other proteins (Kruskal-Wallis test; $P<0.001$ in each case; Fig. 2A). In individual comparisons, median PC1 score for salivary proteins (-0.225) was significantly different from median PC1 score for other secreted proteins (0.132 ; Dunn's test; $P<0.01$; Fig. 2A). Likewise, median PC2 and PC3 scores (-0.027 and -0.560 , respectively) were significantly different from median PC2 and PC3 scores for other secreted proteins (-0.314 ; Dunn's test; $P<0.05$; and -0.026 ; Dunn's test; $P<0.001$, respectively; Fig. 2A).

In order to interpret these differences in terms of individual amino acids, we likewise compared median GAPSRS, DEKQ, and LV among the 3 groups of proteins (Fig. 2B). Median values of the latter 3 quantities differed significantly among salivary proteins, other secreted proteins, and other proteins (Kruskal-Wallis test; $P<0.001$ in each case; Fig. 2B). In individual comparisons, median GAPSRS for salivary proteins (0.323) did not differ significantly from that for other secreted proteins (0.331 ; Dunn's test; n.s.; Fig. 2B). On the other hand, median DEKQ for salivary proteins (0.185) was significantly different from that for other secreted proteins (0.201 ; Dunn's test; $P<0.01$); and median LV for salivary proteins (0.148) was significantly different from that for other secreted proteins (0.160 ; Dunn's test; $P<0.001$; Fig. 2B).

B-cell epitope predictions

We used the BCPREDS software to predict linear B-cell epitopes in singleton proteins classified as salivary proteins or other secreted proteins. Salivary and other secreted proteins were classified in terms of predicted antigenicity as low (no predicted B-cell epitope with score ≥ 0.75); intermediate; and high (at least one predicted epitope with score ≥ 0.99). The distribution of these 3 categories differed significantly between salivary and other secreted proteins ($P<0.001$; Fig. 3). Salivary proteins included a higher percentage (68.6%) of proteins with high antigenicity than did other secreted proteins (54.6%; Fig. 3). Salivary proteins also included a lower percentage (0.4%) of proteins with low antigenicity than did other secreted proteins (9.5%; Fig. 3).

We used binary logistic regression to test whether GAPSRS, DEKQ, and LV predicted the occurrence of at least one epitope with score ≥ 0.99 (high-quality epitope or HQEp). There was a highly significant logistic regression with significant positive coefficients for GAPSRS and DEKQ and a significant negative coefficient for LV (Table 1). Thus, high proportions of GAPSRS and DEKQ and low proportions of LV were associated with the occurrence of one or more HQEp (Table 1).

A plot of LV vs. DEKQ for the salivary proteins revealed certain extreme values (Fig. 4; Table 2). For example, ISCW010560, a highly hydrophobic putative protein, had very low DEKQ and high LV, with consequent low predicted antigenicity (Fig. 4; Table 2). By contrast, the highly acidic protein ISCW008184 ($pI=4.3$) showed very high DEKQ and low

LV and had high predicted antigenicity (Fig. 4). BLAST search showed ISCW008184 to be homologous to calreticulin, a calcium-binding protein found in vertebrates and other animals (Rokeach et al., 1991). ISCW008141, with high DEKQ and low LV, showed high sequence identity to Salp25D (Table 2), identified experimentally to be antigenic by Das et al. (2001). ISCW000240, likewise with high DEKQ and low LV, showed a lower sequence identity to Salp20, which was also experimentally identified as antigenic (Das et al., 2001; Table 2). In spite of the rather low sequence overall sequence identity in the latter case (70.9%), there was one 20-amino-acid window with 95% sequence identity between ISCW000240 and Salp20 which received the highest possible BCPREDS score (1.000; Table 2).

Discussion

We analyzed the amino acid composition of 10,947 single-member protein families (singletons) of *I. scapularis* using principal components analysis. The first principal component corresponded mainly to the frequencies of small amino acid residues (G, A, P, and S) and of R (GAPSR). Secreted salivary proteins showed intermediate frequencies of the latter amino acids, with a median GAPSR lower than that of other secreted proteins, though not significantly so (Fig. 2B). The second principal component corresponded to high frequencies of the polar residues D, E, K, and Q (DEKQ). Secreted salivary proteins showed significantly elevated median DEKQ compared to other secreted proteins (Fig. 2B). The third principal component corresponded mainly to frequencies of the non-polar residues L and V (LV). Secreted salivary proteins showed significantly lower median LV than other secreted proteins (Fig. 2B).

LV was found to be strongly negatively associated with the occurrence of computationally predicted linear B-cell epitopes in *I. scapularis* secreted proteins, while DEKQ was positively associated with the occurrence of predicted B-cell epitopes. Because of the relatively high DEKQ and low LV in secreted salivary proteins, salivary secreted proteins included a significantly higher proportion of proteins with high-quality epitopes than did other secreted proteins, and the frequency of proteins of low antigenicity was substantially lower among salivary proteins than among other secreted proteins. These results did not support the hypothesis that tick salivary proteins as a group have evolved low antigenicity to the vertebrate host.

The latter conclusion is consistent with results of empirical studies showing the antigenicity of certain tick salivary proteins (Das et al., 2001; Mulenga et al., 2000; Schuijt et al., 2011). For a variety of reasons, many of the proteins identified as antigenic by empirical studies were not included in our data set of salivary proteins. In several cases, no complete predicted mRNA in the *I. scapularis* genome assembly corresponded to the previously identified protein, suggesting that substantial gene prediction work remains in the case of *I. scapularis* salivary proteins. Nonetheless, one of the predicted salivary proteins in our data set (ISCW008141) showed high DEKQ, and low LV appeared to correspond to a protein (Salp26A) previously identified experimentally as antigenic (Table 2). Another sequence in our data set (ISCW000240) represented the best available match in the *I. scapularis* genome with another previously identified antigenic protein (Salp20), but in this case, the sequence similarity was lower (Table 2). However, the most highly antigenic region showed high similarity between ISCW000240 and Salp20 (Table 2). In the latter case, the exact relationship between the previously reported mRNA sequence and the predicted gene in the *I. scapularis* genome remains unclear.

Although our analyses did not reveal an overall tendency toward reduced antigenicity in salivary proteins, certain salivary proteins may have evolved reduced antigenicity. The short highly hydrophobic salivary proteins such as ISCW010560 and ISCW002972 may be

possible examples (Table 2). Since neither of these proteins has any known homologs, it is an intriguing hypothesis that such proteins may have the reduction of the antigenicity of saliva as their major or only function. Obviously, quantitative data on the expression of these and other salivary proteins would be required to test this hypothesis. In addition to the antigenicity of proteins themselves, other mechanisms may reduce the antigenicity of tick saliva, including direct inhibition of CD4+ T cell immune responses (Anguita et al., 2002).

On the other hand, the results identified a number of salivary proteins with predicted high antigenicity, particularly ISCW008184 (calreticulin) and ISCW017739. These may be examples of proteins subject to functional constraints that prevent their evolving low antigenicity characteristics. Interestingly, human calreticulin is known to be an autoantigen, being a target of antibodies in autoimmune disease (Rokeach et al., 1991). Such proteins, along with those identified in empirical studies (Das et al., 2001; Mulenga et al., 2000; Schuijt et al., 2011), may merit investigation as possible vaccine components. Further research will be needed to identify the best candidates, including experimental confirmation of both B-cell and T-cell antigenicity of the most promising candidates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health.

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
- Anguita J, Ramamoorthi N, Hovius JW, Das S, Thomas V, Persinski R, Conze D, Askenase PW, Rincón M, Kantor FS, Fikrig E. Salp15, an *Ixodes scapularis* salivary protein, inhibits CD4+ T cell activation. *Immunity.* 2002; 16:849–859. [PubMed: 12121666]
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 2004; 340:783–795. [PubMed: 15223320]
- Das S, Banerjee G, DePonte K, Marcantonio N, Kantor FS, Fikrig E. Salp25D, an *Ixodes scapularis* antioxidant, is 1 of 14 immunodominant antigens in engorged tick salivary glands. *J. Infect. Dis.* 2001; 184:1056–1084. [PubMed: 11574922]
- Deléage G, Roux B. An algorithm for protein secondary structure prediction based on class prediction. *Protein Engineering.* 1987; 1:289–294. [PubMed: 3508279]
- Dong XN, Chen Y, Chen YH. Surface simulation synthesis: a new strategy to spy alpha-helix structure. *Vaccine.* 2007; 25:6569–6571. [PubMed: 17055133]
- EL-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognition.* 2008; 21:243–255.
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985; 39:783–791.
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003; 31:3484–3488.
- Grantham R. Amino acid difference formula to help explain protein evolution. *Science.* 1974; 185:862–864. [PubMed: 4843792]
- Hollander, M.; Wolfe, DA. *Nonparametric Statistical Methods.* John Wiley & Sons; New York: 1999.
- Hovius JW, Levi M, Fikrig E. Salivating for knowledge: potential pharmacological agents in tick saliva. *PLoS Med.* 2008; 5:e43. [PubMed: 18271624]

- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* 2001; 305:567–580. [PubMed: 11152613]
- Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, Hammond M, Hill CA, Konopinski N, Lobo NF, MacCallum RM, Madey G, Megy K, Meyer J, Redmond S, Severson DW, Stinson EO, Topalis P, Birney E, Gelbart WM, Kafatos FC, Louis C, Collins FH. VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res.* 2009; 37:D583–D587. [PubMed: 19028744]
- Li J, Liu F, Wang H, Liu X, Liu J, Li N, Wan F, Wang W, Zhang C, Jin S, Liu J, Zhu P, Liu Y. Systematic mapping and functional analysis of a family of human epididymal secretory sperm-located proteins. *Mol. Cell. Proteomics.* 2010; 9:2517–2528. [PubMed: 20736409]
- Liu W-M, Chou K-C. Prediction of protein secondary structure content. *Protein Engineering.* 1999; 12:1051–1050. [PubMed: 10611398]
- Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* 1999; 285:2177–2198. [PubMed: 9925793]
- Mulenga A, Sugimoto C, Onuma M. Issues in tick vaccine development: identification and characterization of potential candidate vaccine antigens. *Microbes Infect.* 2000; 2:1351–1361.
- Narasimhan S, DePonte K, Marcantonio N, Liang X, Royce TE, Nelson KF, Booth CJ, Koski B, Anderson JF, Kantor F, Fikrig E. Immunity against *Ixodes scapularis* salivary proteins expressed within 24 hours of attachment thwarts tick feeding and impairs *Borrelia* transmission. *PLoS ONE.* 2007; 2(5):e451. [PubMed: 17505544]
- Nickel W. The mystery of nonclassical protein secretion: a current view on cargo proteins and potential export routes. *Eur. J. Biochem.* 2003; 270:2109–2119. [PubMed: 12752430]
- Ribeiro JM, Francischetti IM. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu. Rev. Entomol.* 2003; 48:73–88. [PubMed: 12194906]
- Ribeiro JM, Alarcon-Chaidez F, Francischetti IB, Mans BJ, Mather TN, Valenzuela JG, Wikel SK. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem. Mol. Biol.* 2006; 36:111–129. [PubMed: 16431279]
- Rokeach LA, Haselby JA, Meilof JF, Smreenk RJ, Unnasch TR, Greene BM, Hoch SO. Characterization of the autoantigen calreticulin. *J. Immunol.* 1991; 147:3031–3039. [PubMed: 1919005]
- Schuijt TJ, Hovius JW, van Burgel ND, Ramamoorthi N, Fikrig E, van Dam AP. The tick salivary protein Salp15 inhibits the killing of serum-sensitive *Borrelia burgdorferi* sensu lato isolates. *Infect. Immun.* 2008; 76:2888–2894. [PubMed: 18426890]
- Schuijt TJ, Narasimhan S, Daffre S, DePonte K, Hovius JW, van't Veer C, van der Poll T, Bakhtiari K, Meijers JC, Boder ET, van Dam AP, Fikrig E. Identification and characterization of *Ixodes scapularis* antigens that elicit tick immunity using yeast surface display. *PLoS ONE.* 2011; 6(1):e15926. [PubMed: 21246036]
- Tellam, RL.; Smith, D.; Kemp, DH.; Willadsen, P. Vaccination against ticks. In: Wong, WK., editor. *Animal Parasite Control Utilising Biotechnology.* CRC Press; Boca Raton FL: 1992. p. 303-311.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Diggins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997; 25:4876–4882. [PubMed: 9396791]
- Valenzuela JG. Exploring tick saliva: from biochemistry to “sialomes” and functional genomics. *Parasitology.* 2004; 129:S83–S94. [PubMed: 15938506]
- Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Mather TN, Ribeiro JM. Exploring the sialome of the tick, *Ixodes scapularis*. *J. Exp. Biol.* 2002; 205:2843–2864. [PubMed: 12177149]
- Wang H, Nuttall PA. Immunoglobulin-binding proteins in ticks: a new target for vaccine development against a blood-feeding parasite. *Cell. Mol. Life Sci.* 1999; 56:286–295. [PubMed: 11212356]
- Willadsen P. Immunological approaches to the control of ticks. *Int. J. Parasitol.* 1987; 17:671–677. [PubMed: 3294678]
- Willadsen P, Riding GA, McKenna RV, Kemp DH, Tellam RL, Nielsen JN, Lahnstein J, Cobon GS, Gough JM. Immunologic control of a parasite arthropod. Identification of a protective antigen from *Boophilus microplus*. *J. Immunol.* 1989; 143:1346–1351. [PubMed: 2745982]

Xu Y, Bruno JF, Luft BJ. Identification of novel tick salivary gland proteins for vaccine development. *Biochem. Biophys. Res. Comm.* 2005; 326:901–904. [PubMed: 15607754]

Fig. 1.

(A) Plot of PC1 scores of the 20 amino acids vs. amino acid volume ($r=-0.786$; $P<0.001$); (B) Plot of PC2 scores of the 20 amino acids vs. amino acid polarity ($r=0.783$; $P<0.001$); (C) Plot of PC3 scores of the 20 amino acids vs. P_a ($r=0.525$; $P=0.018$).

Fig. 2.

(A) Median scores of PC1, PC2, and PC3 in singleton salivary proteins, other secreted proteins, and other proteins (Kruskal-Wallis tests of differences among categories: $P < 0.001$ in each case); (B) Median GAPS, DEKQ, and LV in singleton salivary proteins, other secreted proteins, and other proteins (Kruskal-Wallis tests of differences among categories: $P < 0.001$ in each case).

Fig. 3. Percentages of proteins of high, intermediate (int.), and low predicted antigenicity in (A) salivary and (B) other secreted proteins. Numbers in each category are shown. Test of homogeneity: $\chi^2=16.7$; 2 d.f.; $P<0.001$).

Fig. 4. Plot of LV vs. DEKQ for salivary proteins. Proteins are characterized as having low (◆), intermediate (+), or high (○) predicted antigenicity. Individual proteins with unusually high LV and low DEKQ and with unusually high DEKQ and low LV are indicated.

Table 1

Binary logistic regression of the occurrence of high-quality epitope against variables describing amino acid composition of salivary and other secreted proteins.

Predictor	Coefficient	<i>P</i>
Constant	−1.74	=0.009
GAPSR	6.32	<0.001
DEKQ	5.30	<0.001
LV	−6.63	<0.001

Table 2

Examples of salivary proteins with unusually high and low DEKQ relative to LV.

	Protein	LV	DEKQ	Max. BCPREDS score	Putative homology outside Acari (BLAST search)
Low DEKQ	ISCW010560	0.289	0.044	0.229	None found
	ISCW002972	0.261	0.116	0.679	None found
	ISCW018681	0.218	0.115	0.717	Androgen-induced protein 1
	ISCW020648	0.200	0.126	0.241	Ribonuclease
High DEKQ	ISCW008184	0.086	0.411	1.000	Calreticulin
	ISCW017739	0.084	0.361	0.996	None found
	ISCW010106	0.043	0.199	1.000	Papilin
	ISCW013070	0.039	0.250	1.000	Kunitz domain-containing
	ISCW008141	0.167	0.296	0.901	None found ^a
	ISCW000240	0.099	0.237	1.000	None found ^b

^a97.1% amino acid sequence identity to Salp26A (AF209919), identified as antigenic by Das et al. (2001). Percent amino acid identity in 20 amino acid window of highest BCPREDS score: 95%.

^b70.9% amino acid sequence identity to Salp20 (AF209917), identified as antigenic by Das et al. (2001). Percent amino acid identity in 20 amino acid window of highest BCPREDS score: 95%.