Virus and Bacterial Membrane Proteins

Clifford J Hawkins BSc PhD DSc

Two viruses and seven bacteria have been chosen as examples to illustrate the structures of membrane proteins.

Influenza

All the pathogenic subtypes and mutations of the influenza virus so far identified have an abundance of particular prolyl peptides (**XPY,** where **P** is proline and one or both of the adjacent amino acids, **X** and **Y**, are hydrophilic amino acids – love water) in their hemagglutinin and neuraminidase surface membrane proteins. (see Figure 1: the hemagglutinin and neuraminidase proteins are the hair-like structures on the surfaces of the viruses) In addition, neuraminidase proteins - essential for host cell invasion and proliferation and in the former, for causing an immune response that damages the host's organs - in each variant of the influenza virus for which the structures are known.

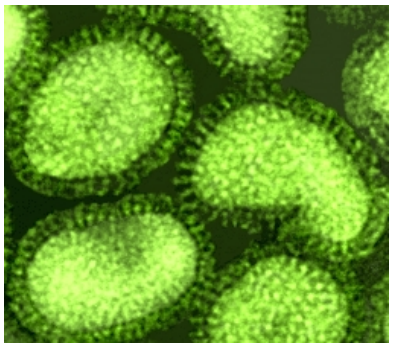

Figure 1: influenza virus

For example, **the sequence of N1 neuraminidase** in the Japan/China H5N1 avian influenza (Mase, M., Eto,M., Tanimura,N., Imai,K., Tsukamoto,K., Horimoto,T., Kawaoka,Y., Yamaguchi,S. "Isolation of a genotypically unique H5N1 influenza virus from duck meat imported into Japan from China" **Virology 339** (1), 101-109 (2005)) has multiple potential sites for ginger enzyme hydrolysis (**bolded**): these sites are adjacent to P-3, 48, 93, 120, 154, 167, 169, 198, 246, 272, 283, 302, 326, 328, 340, 377, 410, 420, 431, and 458, and the blue highlighted prolines are conserved over the three H5N1, the H1N1 swine flu of 1918 and 2009, and H9N7 bird flu, with the pink highlighted prolines are not conserved in the H9N7 bird flu:

1 M**NP**NQKITTI GSICMVIGIV SLMLQIGNII SIWVSHSIQT GNQHQA**EPC**N QSIITYENNT
61 WVNQTYVNIS NTNFLTEKAV NLVTLAGNSS L**CP**ISGWAVY SKDNGIRIGS KGDVFVIR**EP**
121FISCSHLECR TFFLTQGALL NDKHSNGTVK DR**SPH**RTLMS **CP**VGEA**PSPY**NSRFESVAWS
181ASACHDGTSW LTIGISG**PD**N GAVAVLKYDG IITDTIKSWR NNILRTQESE CACVNGSCFT
241VMTDG**PS**NGQ ASYKIFRIEK GKVVKSAELNA**PN**YHYEECSC**YPD**AGEITCVCRDNWHGSN
301**RP**WVSFNQNL EYRIGYICSG VFGD**NPRPN**DGTGSCG**P**V**SPK**GAYGIKGFSFRYGNGVWIG
361RTKSTNSRSG FEMIW**DPN**GW TGTDSNFSVK QDIVAITDWSGYSGSFVQ**HPE**LTGLDCI**RP**
421**C**FWVELIRG**R PK**ESTIWTSG SSISFCGVNS DTVGWSW**PD**G AEL**P**FTIDK

Each letter in the sequence represents an amino acid. The important hydrophilic amino acids adjacent to prolines are: asparagine, **N**; glutamate, **E**; cysteine, **C**; serine, **S**; histidine, **H**; aspartate, **D**; arginine, **R**; lysine, **K;** tyrosine, **Y**.

Cleavage of the neuraminidase proteins by the ginger enzyme will prevent the virus invading the host cells and prevent proliferation of the virus in the host.

The identified epitopes for hemagglutinin are not linear peptides but are what is called "conformational" epitopes in which the participating amino acids are brought together via the three-dimensional structure of the protein. Seven of the H1N1 2009 (Swine flu) active prolyl peptides are included in the conformational epitopes and six of these are conserved prolines in H1 and H5. Numerous other non-proline amino acids in the epitopes are not conserved.(Deem, M.W., Pan, K. "The epitope regions of H1-subtype influenza A, with application to vaccine efficacy" **Protein Eng., Design & Selection**, 1-4 (2009,July 3))

EPITOPES (CONFORMATIONAL)

A: RQLSSFERFPKSWPNHDKGTWGD
B: VSCPHAGAFKDKGKE LVL GIHH
C: DTVLENVVTH AFAMER AGSSSHTQPKNTLPFQNI
D: AYIVDLLVKKGNSYPLSSSDQSLYQNADTYVFV SKKFKPVDERNYY
E: VNLEKHNLLGKCNIAG LGNPETFEATGLR

Underlined letters are three or more amino acid linear peptides.
Yellow highlighted letters are part of an active proline peptide group which is cleaved by the ginger enzyme.

The inhibitors currently on the market function differently to the ginger enzyme: they are specifically designed to 'plug' the active site of neuraminidase where the neuraminidase opens the virus' surface membrane so the virus can enter the host's cells. They rely on the amino acids that surround the active site (glutamate, arginine and aspartate) to stabilize the binding of the plug in the hole. (see Figures 2). The above N1 variant in H5N1 bird flu does not have arginine, R, in positions 92 and 371 and subsequently the binding and efficacy of the 'plug' could be significantly reduced.

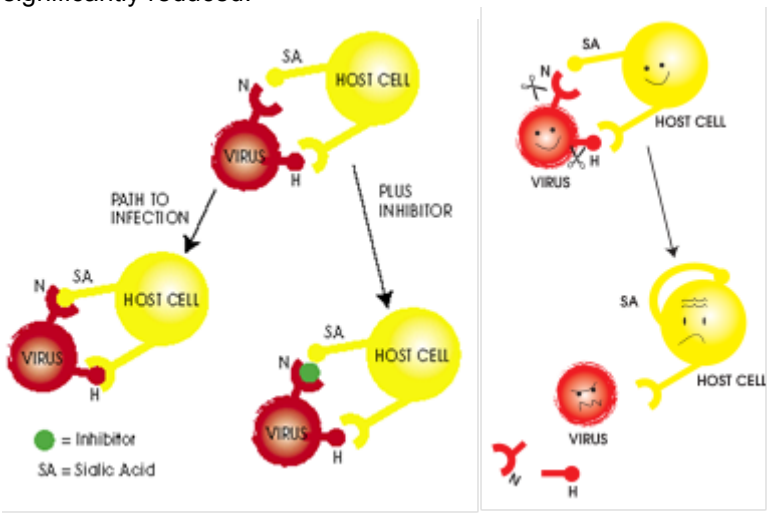

Figure 2: The influenza virus invades a host cell by initially binding to receptors on the host cell, one of which binds to the hemagglutinin (H) on the virus, and one with sialic acid (SA) which binds to the neuraminidase (N) protein on the virus. Current inhibitors such as Relenza are designed to mimic the sialic acid and to bind to the neuraminidase blocking the link to the host cell. The Biohawk ginger product acts (on the right) like a pair of scissors and specifically cuts off the hemagglutinin and neuraminidase proteins from the surface of the virus completely preventing the virus infecting the cell and replicating itself.

Importantly, although the various H5 structures show significant mutations, the potential sites for ginger enzyme hydrolysis are largely conserved and number at least 15. In the H5-hemagglutinins recently found in the Vietnam (Nguyen T.D., Hanh T.H., Puthavathana P., Long H.T., Buranathai C., Lim W., Webster R.G., Hoffmann E. "Lethality to Ferrets of H5N1 Influenza Viruses Isolated from Humans and Poultry in 2004" **J. Virol. 79,** 2191-2198 (2005)) and Japan/China variants, and previously identified in Singapore H5N1 avian influenza (Ha, Y., Stevens, D.J., Skehel, J.J., Wiley, D.C. "Structure of Avian H5 Haemagglutinin Complexed with LSTA" **Proc. Nat. Acad. Sci. USA, 98,** 11181 (2001)), the prolines are conserved. In addition there is an extra proline, P-233, in the Singapore H5N1, which has the hydrophilic arginine and lysine adjacent to it, but in the Vietnam and Japan/China variants, proline is replaced by serine. The hydrophilic amino acids adjacent to the prolines are also conserved, except for those adjacent to P-101 with asparagine (N-100) for the Vietnam and Singapore variants and serine (S-100) for the Japan/China protein, P-108 with the non-hydrophilic glycine for the Vietnam and Japan/China variants and the hydrophilic glutamate adjacent to the proline for the Singapore variant - giving an additional site for hydrolysis, P-134 with lysine for the Vietnam and Japan/China proteins and with arginine for the Singapore protein, and for P-337, which has glutamine following it for all three but with serine before the proline for the Vietnam and Japan/China variants and valine for the Singapore protein. The target prolines for ginger enzyme hydrolysis are: P-65, 81, 90, 101, 108 (Singapore only), 134, 174, 197, 210, 227, 233 (Singapore only) 251, 266, 297, 312, 319, 337, 506.

This gives the ginger enzyme excellent opportunity to hydrolyze the H5 at multiple sites. The conservation of the prolyl residues in the hemagglutinin structures is suggestive of these having a specific role in the function of the protein. Hydrolysis by Biohawk's ginger at these specific sites in the neuraminidase and hemagglutinin proteins, would cleave them from the virus inhibiting its ability to invade host cells and to proliferate. Further, hydrolysis of the hemagglutinin structure would prevent this viral protein from stimulating damaging cytokine fluxes. Independent studies have confirmed the ginger enzyme inhibits H5N1 bird flu (Selleck, P "Efficacy of Zingibain in inactivating H5N1 Avian Influenza Virus" Report July 2007- A/chicken/Vietnam/8/2004 **H5N1**)

Papilloma Virus

All capsid proteins of papilloma viruses are proline-rich with a high degree of conservation of the proline peptides. The known structures of the proteins associated with common warts and with anogenital papilloma infections are as follows. The proline peptides are highlighted:

PV L1A Sequences-warts common

```
CPV1        . MWRPSDNKLYVPPPAPVSKVLTTDAYVTRTKIFYHASSSRLLAVGNPYFPIRK.....ANKTIVPKVSGFQF          67
RHPV1R   MSMWRPSDSKVYLPP…PVSKVVSTDEYVSRTSIYYHAGSSRLLAVGHPYYAVKK..GNNKVSVPKVSGLQY          68
HPV29    MALWRSSDNLVYLPP.TPVSKVISTDD….YVTRTNIYYYAGSSRLLTVGHPHYSIPKSSGNKVDVPKVSAFQY          70
HPV2a    MALWRPNESKVYLPP.TPVSKVI..STDVYVTRTNVYYHGGSSRLLTVGHPYYSIKK....SNNKVAVPKVSGYQY          69
HPV27    MALWRPNESKVYLPP.TPVSKVI..STDVYVTRTNVYYHGGSSRLLTVGHPYYSIKK..GSNNRLAVPKVSGYQY          70
HPV57    MAMWRPNESKVYLPP.TPVSKVL.STDVYVTRTNVYYHGGSSRLLTVGHPYYSIKK SGNNKVSVPKVSGYQY          70
HPV26    MALWRTSDSKVYLPP.TPVSRVVNTDE…YVTRTGIYYYAGSSRLLTLGHPYFSIPK....TGQKAEIPKVSAYQY          69

CPV1     RVFKIVL..PDPNKFALPDTSIFDSTSQRLVWACI…GLEVGRGQPLGVGYCGHPCLNKFDDVENSASYAVNPGQDNR  141
RHPV1R   RVFRVRLPDPNKFGLPDANFYDPNTQRLVWACLGVEVGRGQPLGVGTSGHPLLNKLDDTENGPKVAGGQGADNR  142
HPV29    RVFRVRLPDPNKFGLPDARIYNPEAERLVWACTGVEVGRGQPLGVGLSGHPLYNKLNDTENSNIAHAENGQDSR  144
HPV2a    RVFHVKLPDPNKFGLPDADLYDPDTQRLLWACVGVEVGRGQPLGVGVSGHPYYNRLDDTENAHTPDT..ADDGR  141
HPV27    RVFHVKLPDPNKFGLPDADLYDPDTQRLLWACVGVEVGRGQPLGVGVSGHPYYNRQDDTENAHTLDS..AEDGR  142
HPV57    RVFHVKLPDPNKFGLPDANLYDPDTQRLLWACVGVEVGRGQPLGVGISGHPYYNKQDDTENSHNPDA..ADDGR  142
HPV26    RVFRVHLPDPNKFGLPDPQLYNPDTERLVWACVGVEVGRGQPLGIGLSGHPLFNKLDDTENSHLATVNADTDNR  143

CPV1     VNVAMDYKQTQLCLVGCAPPLGEHWGKGKQCSGVSVQDGDCPPLELVTSVIQDGDMVDTGFGAMDFAELQSNKS215
RHPV1R   ECVSMDYKQTQLCMLGCKPPVGEHWGKGNPC..TTGAAGDCPALELVNSVIQDGDMVDTGYGAMDFNALQANKS  214
HPV29    DNIAVDYKQTQLCILGCTPPMGEHWGKGTVCARTSSAAGDCPPLELMTTHIEDGDMVDTGYGAMDFAALQVNKS  218
HPV2a    ENISMDYKQTQLFILGCKPPIGEHWSKGTTC.NGSSAAGDCPPLQFTNTTIEDGDMVETGFGALDFATLQSNKS  214
HPV27    ENISMDYKQTQLFILGCKPSIGEHWSKGTTC.NGSSAAGDCPPLQFTNSTIEDGDMVETGFGALDFATLQSNRS  215
HPV57    EYISMDYKQTQLFILGCKPPIGEHWSKGTTC.SGSSAVGDCPPLQFTNTTIEDGDMVETGFGALDFAALQSNKS  215
HPV26    DNVSVDNKQTQLCIIGCTPPLGEHWGIGTICKNTQTQRGDCPPLELISSIIEDGDMIDTGFGAMDFTALQATKS  217

CPV1     DVPLDICTSTCKYPDYLQMAADPYGDRLFFYLRKEQMFARHFFNRAGTVGEQIPDELFVKGTT...SRATVSSN  286
RHPV1R   DVPIDICTSVCKYPDYLKMASDPYGDSLFFYLRREQMFVRHLFNRAGTMGDSVPDDLYIKGSG...SNVKLASH  285
HPV29    DVPLDICQSTCKYPDYLGMAADPYGDSMFFFLRREQLFARHFFNRAGVVGDKIPDSLYLKGNN...GRETPGSA  289
HPV2a    DVPLDICTNTCKYPDYLKMAAEPYGDSMFFSLRREQMFTRHFFNLGGKMGDTIPDELYIKSTS....VPTPGSH  284
HPV27    DVPLDICTNVCKYPDYLKMAAEPYGDSMFFSLRREQMFTRHFFNRAGKMGDTIPDELYIKSTT....ISDPGSH  285
HPV57    DVPLDICTNICKYPDYLKMAADPYGDSMFFSLRREQMFTRHFFNRGGSMGDALPDELYVKSST....VQTPGSY  285
HPV26    DVPIDISQSTCKYPDYLKMSADTYGNSMFFFLRREQLFARHFYNKAGAVGDAIPTTLYIKGAES..GREPPTSS  289

CPV1     IYFNTPSGSLVSSEAQLFNKPYWLHKAQGHNNGICWGNTLFVTVVDTTRSTNMTVCASTTSSP...SATYTASE  357
RHPV1R   VFYPTPSGSMVTSDAQLFNKPYWLQKAQGHNNGICWGNQVFLTVVDTTRSTNMTLCASTASTV...TTPYNNES  356
HPV29    IYSPTPSGSMVTSEAQIFNKPYWLQQAQGHNNGICWANQVFLTVVDTTRSTNMSLCATTESQP...LTTYDATK  360
HPV2a    VYTSTPSGSMVSSEQQLFNKPYWLRRAQGHNNGMCWGNRVFLTVVDTTRSTNVSLCATEASDT.....NYKATN  353
HPV27    VYTSTPSGSMVSSEQQLFNKPYWLRRAQGHNNGMCWGNRIFLTVVDTTRSTNVSLCAAEVSDN....TNYKATN  355
HPV57    VYTSTPSGSMVSSEQQLFNKPYWLRRAQGHNNGMCWGNRIFLTVVDTTRSTNVSLCATVTTET.....NYKASN  354
HPV26    IYSATPSGSMVTSDAQLFNKPYWLQRAQGHNNGICWGNQLFVTCVDTTRSTNLTISTLSAASA...STPFKPSD  360

CPV1     YKQYMRHVEEFDLQFIFQLCTIKLTAELMAYIHTMNPTVLEEWNFGLSPPPNGTLEDTYRYVQSQAITCQK.P.  429
RHPV1R   FKEYLRHVEEFDLQFIFQLCKVTLNTEVMAYIHSMDASILEDWNFGLQPPPSGSLQDTYRFVTSAAITCQK.P.  428
HPV29    IKEYLRHGEEYDLQFIFQLCKVTLTPEIMAYLHTMNSALLEDWNFGLTLPPSTSLEDTYRFVTSSAITCQK.D.  432
HPV2a    FKEYLRHMEEYDLQFIFQLCKITLTPEIMAYIHNMDPQLLEDWNFGVPPPPSASLQDTYRYLQSQAITCQK.P.  425
HPV27    FKEYLRHMEEYDLQFIFQLCKITLTPEIMAYIHNMDPQLLEDWNFGVPPPPSASLQDTYRYLQSQAITCQK.P.  427
HPV57    YKEYLRHMEEYDLQFIFQLCKITLTPEIMAYIHNMDARLLEDWNFGVPPPPSASLQDTYRYLQSQAITCQK.P.  426
HPV26    YKQFIRHGEEYELQFIFQLCKITLTTDVMAYIHLMNASILEDWNFGLTLPPTASLEDAYRFIKNSATTCQR.N.  432

CPV1     T.PDKEKQDPYAGLSFWEVNLKEKFSSELEQYPLGRKFLLQTGVQSTSLARAG...TKRAA.....STST.ATP  493
RHPV1R   A.PPKEKEDPLAKYTFWEVDLKEKFSADLDQFPLGRKFLLQAGMRARPTLRAP....KRTAS...STSSS.SPR  493
HPV29    L.APTEKQDPYAKLNFWDVDLKDRFTLDLSQFPLGRKFLLQIGARRRSVVPSR....KRRT.....TTTAPTPA  496
HPV2a    T.PPKTPTDPYASLTFWDVDLSESFSMDLDQFPLGRKFLLQRGAM.PTVSRKR....AAVS.....GTTP.PTS  487
HPV27    T.PPKTPTDPYANMTFWDVDLRESFSMDLDQFPLGRKFLLQRGTT.............P.....TVSR.KRT  479
HPV57    T.PPKTPTDPYATMTFWDVDLSESFSMDLDQFPLGRKFLLQRGAT..PTVSRK....RAAA.....TAAA.PTA  487
HPV26    A.PPVPKEDPFQKFKFWDVDLKEKFSIDLDQFPLGRKFMLQAGIQRRPKLGTK....RPLS.....STS..SST  494

CPV1     TR.KKVKRK.....  501
RHPV1R   KR.KRTKR.....  500
HPV29    KR.KRSKK.....  503
HPV2a    KR.KRVRR.....  494
HPV27    AV.GRGH.....  485
HPV57    KR.KKVRR.....  494
HPV26    KR.KKRKLTK....  503
HPV52    KK.KVKR.....  503
HPV58    KR.KKVKK.....  498
HPV67    RK.KVKR.....  500
```

HPV L1A Sequences-Anogenital

```
HPV32    MSVWRPSDNKVYLPP.PPVSKVVSTDEYVQRTNYFYHASSSRLLAVGHPYYTIKK....TPNRTSIPKVSGLQY          69
HPV11R   ..  MWRPSDSTVYVPPPNPVSKVVATDAYVKRTNIFYHASSSRLLAVGHPYYSIKK....VNKTVVPKVSGYQY          67
HPV6bR   ..  MWRPSDSTVYVPPPNPVSKVVATDAYVTRTNIFYHASSSRLLAVGHPYFSIKR....ANKTVVPKVSGYQY          67
HPV18R   MALWRPSDNTVYLPP.PSVARVVNTDD…YVTRTSIFYHAGSSRLLTVGNPYFRVPAGGGNKQDIPKVSAYQY          70
HPV16R   MSLWLPSEATVYLPP.VPVSKVVSTDEYVARTNIYYHAGTSRLLAVGHPYFPIKK..PNNNKILVPKVSGLQY          70
HPV31    MSLWRPSEATVYLPP.VPVSKVVSTDEYVTRTNIYYHAGSARLLTVGHPYYSIPKSDNPKKIVVPKVSGLQY          71
HPV33    MSVWRPSEATVYLPP.VPVSKVVSTDEYVSRTSIYYYAGSSRLLAVGHPYFSIKNPTNAKKLLVPKVSGLQY          71

HPV32    RVFRVRLPDPNKFTLPETNLYNPETQRMVWACVGLEVGRGQPLGVGLSGHPLLNRLDDTENGPRYAAGPGTDNR  143
HPV11R   RVFKVVLPDPNKFALPDSSLFDPTTQRLVWACT..GLEVGRGQPLGVGVSGHPLLNKYDDVENSGGYGGNPGQDNR  141
HPV6bR   RVFKVVLPDPNKFALPDSSLFDPTTQRLVWACT.GLEVGRGQPLGVGVSGHPFLNKYDDVENSGS.GGNPGQDNR  140
HPV18R   RVFRVQLPDPNKFGLPDTSIYNPETQRLVWACAGVEIGRGQPLGVGLSGHPFYNKLDDTESSHAATSNVSEDVR  144
HPV16R   RVFRIHLPDPNKFGFPDTSFYNPDTQRLVWACVGVEVGRGQPLGVGISGHPLLNKLDDTENASAYAANAGVDNR  144
HPV31    RVFRVRLPDPNKFGFPDTSFYNPETQRLVWACVGLEVGRGQPLGVGISGHPLLNKFDDTENSNRYAGGPGTDNR  145
HPV33    RVFRVRLPDPNKFGFPDTSFYNPDTQRLVWACVGLEIGRGQPLGVGISGHPLLNKFDDTETGNKYPGQPGADNR  145

HPV32    ENVSMDCKQTQLCLVGCKPAIGEHWGKGAACSA..QSNGDCPPLELQNSVIQDGDMADVGFGAMDFSALQTSKA  215
HPV11R   VNVGMDYKQTQLCMVGCAPPLGEHWGKGTQCSNTSVQNGDCPPLELITSVIQDGDMVDTGFGAMNFADLQTNKS  215
HPV6bR   VNVGMDYKQTQLCMVGCAPPLGEHWGKGKQCTNTPVQAGDCPPLELITSVIQDGDMVDTGFGAMNFADLQTNKS  214
```

```
HPV18R   DNVSVDYKQTQLCILGCAPAIGEHWAKGTACKSRPLSQGDCPPLELKNTVLEDGDMVDTGYGAMDFSTLQDTKC  218
HPV16R   ECISMDYKQTQLCLIGCKPPIGEHWGKGSPCTNVAVNPGDCPPLELINTVIQDGDMVDTGFGAMDFTTLQANKS  218
HPV31    ECISMDYKQTQLCLLGCKPPIGEHWGKGSPCSNNAITPGDCPPLELKNSVIQDGDMVDTGFGAMDFTALQDTKS  219
HPV33    ECLSMDYKQTQLCLLGCKPPTGEHWGKGVACTN.AAPANDCPPLELINTIIEDGDMVDTGFGCMDFKTLQANKS  218

HPV32    EVPLDIMNSISKYPDYLKMSAEAYGDNMFFFLRREQMFVRHLFNRAGTLGEPVPEDMYIKASNGASGRNNLASS  289
HPV11R   DVPLDICGTVCKYPDYLQMAADPYGDRLFFYLRKEQMFARHFFNRAGTVGEPVPDDLLVKGGN...NRSSVASS  286
HPV6bR   DVPIDICGTTCKYPDYLQMAADPYGDRLFFFLRKEQMFARHFFNRAGEVGEPVPDTLIIKGSG...NRTSVGSS  285
HPV18R   EVPLDICQSICKYPDYLQMSADPYGDSMFFCLRREQLFARHFWNRAGTMGDTVPQSLYIKGTG...MRASPGSC  289
HPV16R   EVPLDICTSICKYPDYIKMVSEPYGDSLFFYLRREQMFVRHLFNRAGTVGENVPDDLYIKGSG...STANLASS  289
HPV31    NVPLDICNSICKYPDYLKMVAEPYGDTLFFYLRREQMFVRHFFNRSGTVGESVPTDLYIKGSG...STATLANS  290
HPV33    DVPIDICGSTCKYPDYLKMTSEPYGDSLFFFLRREQMFVRHFFNRAGTLGEAVPDDLYIKGSG...TTASIQSS  289

HPV32    IYYPTPSGSMVTSDAQIFNKPYWLQQAQGHNNGICWGNQVFLTVVDTTRSTNMTVCATVTTED.....TYKSTN  358
HPV11R   IYVHTPSGSLVSSEAQLFNKPYWLQKAQGHNNGICWGNHLFVTVVDTTRSTNMTLCASVSKSA.....TYTNSD  355
HPV6bR   IYVNTPSGSLVSSEAQLFNKPYWLQKAQGHNNGICWGNQLFVTVVDTTRSTNMTLCASVTTSS.....TYTNSD  354
HPV18R   VYSPSPSGSIVTSDSQLFNKPYWLHKAQGHNNGVCWHNQLFVTVVDTTRSTNLTICASTQSPV...PGQYDATK  360
HPV16R   NYFPTPSGSMVTSDAQIFNKPYWLQRAQGHNNGICWGNQLFVTVVDTTRSTNMSLCAAISTSE....TTYKNTN  359
HPV31    TYFPTPSGSMVTSDAQIFNKPYWMQRAQGHNNGICWGNQLFVTVVDTTRSTNMSVCAAIANSD....TTFKSSN  360
HPV33    AFFPTPSGSMVTSESQLFNKPYWLQRAQGHNNGICWGNQVFVTVVDTTRSTNMTLCTQVTSDS.....TYKNEN  358

HPV32    FKEYLRHAEEYDIQFIFQLCKITLSVEVMSYIHTMNPDILDDWNVGVAPPPSGTLEDSYRFVQSQAIRCQA.K.  430
HPV11R   YKEYMRHVEEFDLQFIFQLCSITLSAEVMAYIHTMNPSVLEDWNFGLSPPPNGTLEDTYRYVQSQAITCQK.P.  427
HPV6bR   YKEYMRHVEEYDLQFIFQLCSITLSAEVMAYIHTMNPSVLEDWNFGLSPPPNGTLEDTYRYVQSQAITCQK.P.  426
HPV18R   FKQYSRHVEEYDLQFIFQLCTITLTADVMSYIHSMNSSILEDWNFGVPPPPTTSLVDTYRFVQSVAITCQK.D.  432
HPV16R   FKEYLRHGEEYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPGGTLEDTYRFVTSQAIACQK.H.  431
HPV31    FKEYLRHGEEFDLQFIFQLCKITLSADIMTYIHSMNPAILEDWNFGLTTPPSGSLEDTYRFVTSQAITCQK.T.  432
HPV33    FKEYIRHVEEYDLQFVFQLCKVTLTAEVMTYIHAMNPDILEDWQFGLTPPPSASLQDTYRFVTSQAITCQK.T.  430

HPV32    V.TAPEKKDPFSDYSFWEVNLSEKFSSDLDQFPLGRKFLLQAGLRARPKLTAV....KRTASS.SQKSSS.PAK  497
HPV11R   T.PEKEKQDPYKDMSFWEVNLKEKFSSELDQFPLGRKFLLQSGYRGRTSARTG...IKRPA....VSKPS.TAP  492
HPV6bR   T.PEKEKPDPYKNLSFWEVNLKEKFSSELDQYPLGRKFLLQSGYRGRSSIRTG...VKRPA....VSKAS.AAP  491
HPV18R   A.APAENKDPYDKLKFWNVDLKEKFSLDLDQYPLGRKFLVQAGLRRKPTIGPR....KRSAP..SATTSS.KPA  498
HPV16R   T.PPAPKEDPLKKYTFWEVNLKEKFSADLDQFPLGRKFLLQAGLKAKPKFTLG....KRKATP.TTSSTS.TTA  498
HPV31    A.PQKPKEDPFKDYVFWEVNLKEKFSADLDQFPLGRKFLLQAGYRARPKFKAG....KRSAP...SASTT.TPA  497
HPV33    V.PPKEKEDPLGKYTFWEVDLKEKFSADLDQFPLGRKFLLQAGLK.AKPKLKR....AAPT.....STRT.SSA  492

HPV32    RR.KTRK.....  503
HPV11R   KR.KRTKTKK....  501
HPV6bR   KR.KRAKTKR....  500
HPV18R   KR.VRVRARK....  507
HPV16R   KR.KKRKL......  505
HPV31    KR.KKTKK......  504
HPV33    KR.KKVKK......  499
```

The multiple highlighted proline peptides (and the other prolines which lack adjacent hydrophilic amino acids) render these proteins inert to normal hydrolysis. These intact proline-rich peptides in which the prolines are adjacent to hydrophilic amino acids are ideal epitopes for triggering a severe immune response and for binding to host cells. The Biohawk ginger with its high level of enzyme has the potential to cleave these peptides, resulting in breakdown of the viral capsid and structural proteins, prevention of host cell invasion and avoidance of immune system stimulation and the associated tissue damaging cytokine flux. The multiple exposed sites for the ginger enzyme cleavage on one side of the protein are coloured orange in the following model of the HVP16-L1 protein (Figure 3).

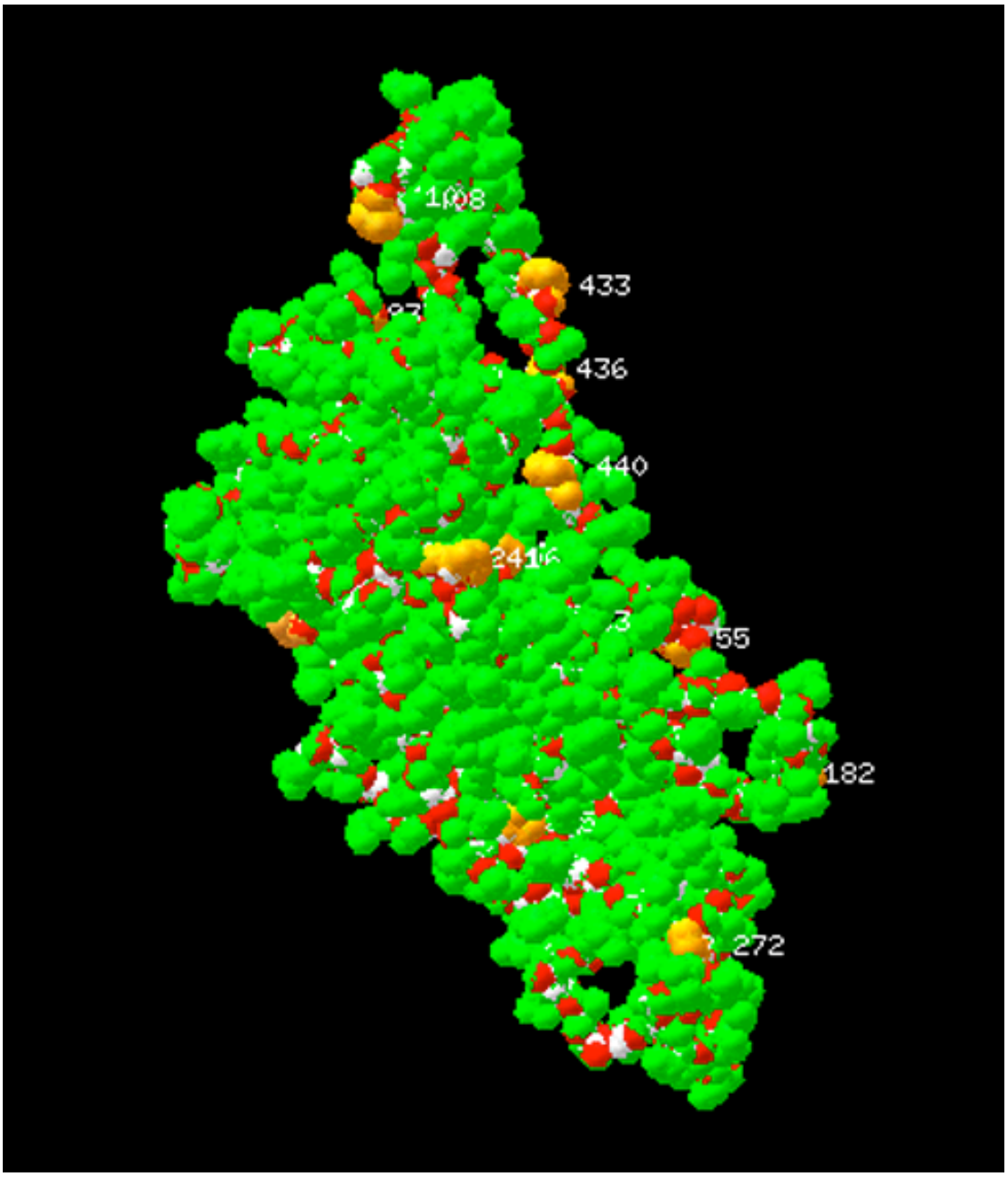

Figure 3. A model of the HPV16-L1 capsid protein responsible for generating a number of forms of squamous cell carcinomas. In the present picture green represents the surface groups and the backbone has been shown in red and white. The suitable prolines for ginger enzyme cleavage are shown in orange-yellow. The numbers represents the proline number in the HPV16-L1 protein.

Papilloma virus is of particular interest to Biohawk because it is associated with disease in humans and other species, for example equine sarcoid, bovine eye cancer, common warts, corns, some squamous cell carcinomas, and cervical cancer. Case studies have shown a very favourable response to treatment with the Biohawk creams (for example, Skin Rejuvenator). For example, the following common warts (Figure 4) were treated with a ginger cream once at 3 pm on one day and the next morning the growths had fallen off the hand.

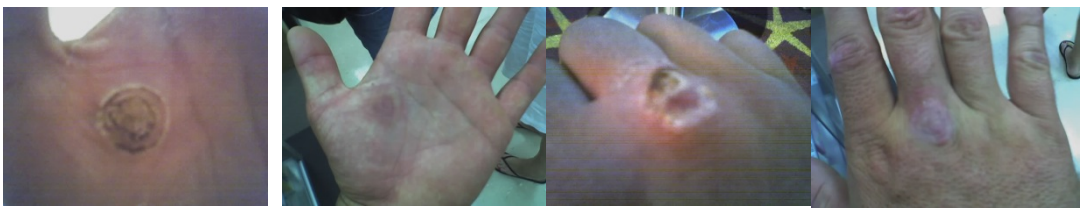

Figure 4: Common wart treated with an active ginger cream

Bacteria

Some bacteria have a carbohydrate coating on the membrane mostly attached to the membrane proteins through the amino acids, serine (highlighted pink) and threonine (highlighted grey). This coating makes it more difficult for antibiotics to penetrate the membrane and makes it more difficult for the ginger enzyme to digest the proline-rich proteins. Biohawk's Pine Crush was developed to remove the carbohydrate coating. In other types of bacteria the proline-rich proteins project out of the membrane and although they may have serine and threonine, the carbohydrate coating does not prevent the ginger enzyme from digesting the proline-rich membrane proteins. Some examples of membrane proteins for common types of bacteria are given below.

Whooping Cough: Bordetella pertussis

***Bordetella pertussis*, the causative agent of whooping cough,** is an aerobic coccobacillus capsulate of the genus *Bordetella*. The acellular pertussis vaccine components, pertussis toxoid (PT), pertussis filamentous haemagglutinin (FHA) and pertactin (PRN) are extracted from phase I *Bordetella pertussis,* and are then purified and stabilised. The structures of FHA and PRN are given below.

FHA protein [Bordetella pertussis]

1 MNTNLYRLVF SHVRGMLVPV SEHCTVGNTF CGRTRGQARS GARATSLSVA PNALAWALML
61 ACTGLPLVTH AQGLVPQGQT QVLQGGNKVP VVNIADPNSG GVSHNKFQQF NVANPGVVFN
121 NGLTDGVSRI GGALTKNPNL TRQASAILAE VTDTSPSRLA GTLEVYGKGA DLIIANPNGI
181 SVNGLSTLNA SNLTLTTGRP SVNGGRIGLD VQQGTVTIER GGVNATGLGY FDVVARLVKL
241 QGAVSSKQGK PLADIAVVAG ANRYDHATRR ATPIAAGARG AAAGAYAIDG TAAGAMYGKH
301 ITLVSSDSGL GVRQLGSLSS PSAITVSSQG EIALGDATVQ RGPLSLKGAG VVSAGKLASG
361 GGAVNVAGGG AVKIASASSV GNLAVQGGGK VQATLLNAGG TLLVSGRQAV QLGAASSRQA
421 LSVNAGGALK ADKLSATRRV DVDGKQAVAL GSASSNALSV RAGGALKAGK LSATGRLDVD
481 GKQAVTLGSV ASDGALSVSA GGNLRANELV SSAQLEVRGQ REVALDDASS ARGMTVVAAG
541 ALAARNLQSK GAIGVQGGEA VSVANANSDA ELRVRGRGQV DLHDLSAARG ADISGEGRVN
601 IGRARSDSDV KVSAHGALSI DSMTALGAIG VQAGGSVSAK DMRSRGAVTV SGGGAVNLGD
661 VQSDGQVRAT SAGAMTVRDV AAAADLALQA GDALQAGFLK SAGAMTVNGR DAVRLDGAHA
721 GGQLRVSSDG QAALGSLAAK GELTVSAARA ATVAELKSLD NISVTGGERV SVQSVNSASR
781 VAISAHGALD VGKVSAKSGI GLEGWGAVGA DSLGSDGAIS VSGRDAVRVD HARSLADISL 841
GAEGGATLGA VEAAGSIDVR GGSTVAANSL HANRDVRVSG KDAVRVTAAT SGGGLHVSSG
901 RQLDLGAVQA RGALALDGGA GVALQSAKAS GTLHVQGGEH LDLGTLAAVG AVDVNGTGDV
961 RVAKLVSDAG ADLQAGRSMT LGIVDTTGDL QARAQQKLEL GSVKSDGGLQ AAAGGALSLA
1021 AAEVAGALEL SGQGVTVDRA SASRARIDST GSVGIGALKA GAVEAASPRR ARRALRQDFF
1081 TPGSVVVRAQ GNVTVGRGDP HQGVLAQGDI IMDAKGGTLL LRNDALTENG TVTISADSAV
1141 LEHSTIESKI SQSVLAAKGD KGKPAVSVKV AKKLFLNGTL RAVNDNNETM SGRQIDVVDG
1201 RPQITDAVTG EARKDESVVS DAALVADGGP IVVEAGELVS HAGGIGNGRN KENGASVTVR
1261 TTGNLVNKGY ISAGKQGVLE VGGALTNEFL VGSDGTQRIE AQRIENRGTF QSQAPAGTAG
1321 ALVVKAAEAI VHDGVMATKG EMQIAGKGGG SPTVTAGAKA TTSANKLSVD VASWDNAGSL
1381 DIKKGGAQVT VAGRYAEHGE VSIQGDYTVS ADAIALAAQV TQRGGAANLT SRHDTRFSNK
1441 IRLMGPLQVN AGGPVSNTGN LKVREGVTVT AASFDNETGA EVMAKSATLT TSGAARNAGK
1501 MQVKEAATIV AASVSNPGTF TAGKDITVTS RGGFDNEGKM ESNKDIVIKT EQFSNGRVLD
1561 AKHDLTVTAS GQADNRGSLK AGHDFTVQAQ RIDNSGTMAA GHDATLKAPH LRNTGQVVAG
1621 HDIHIINSAK LENTGRVDAR NDIALDVADF TNTGSLYAEH DATLTLAQGT QRDLVVDQDH
1681 ILPVAEGTLR VKAKSLTTEI ETGNPGSLIA EVQENIDNKQ AIVVGKDLTL SSAHGNVANE
1741 ANALLWAAGE LTVKAQNITN KRAALIEAGG NARLTAAVAL LNKLGRIRAG EDMHLDAPRI
1801 ENTAKLSGEV QRKGVQDVGG GEHGRWSGIGYVNYWLRAGNGKKAGTIAAPWYGGDLTAEQ
1861 SLIEVGKDLY LNAGARKDEH RHLLNEGVIQ AGGHGHIGGD VDNRSVVRTV SAMEYFKTPL
1921 PVSLTALDNR AGLSPATWNF QSTYELLDYL LDQNRYEYIW GLYPTYTEWS VNTLKNLDLG
1981 YQAKPAPTAP PMPKAPELDL RGHTLESAEG RKIFGEYKKL QGEYEKAKMA VQAVEAYGEA
2041 TRRVHDQLGQ RYGKALGGMD AETKEVDGII QEFAADLRTV YAKQADQATI DAETDKVAQR
2101 YKSQIDAVRL QAIQPGRVTL AKALSAALGA DWRALGHSQL MQRWKDFKAG KRGAEIAFYP
2161 KEQTVLAAGA GLTLSNGAIH NGENAAQNRG RPEGLKIGAH SATSVSGSFD ALRDVGLEKR
2221 LDIDDALAAV LVNPHIFTRI GAAQTSLADG AAGPALARQA RQAPETDGMV DARGLGSADA
2281 LASLASLDAA QGLEVSGRRN AQVADAGLAG PSAVAAPAVG AADVGVEPVT GDQVDQPVVA
2341 VGLEQPVATV RVAPPAVALP RPLFETRIKF IDQSKFYGSR YFFEQIGYKP DRAARVAGDN
2401 YFDTTLVREQ VRRALGGYES RLPVRGVALV AKLMDSAGTV GKALGLKVGV APTAQQLKQA
2461 DRDFVWYVDT VIDGQKVLAP RLYLTEATRQ GITDQYAGGG ALIASGGDVT VNTDGHDVSS
2521 VNGLIQGRSV KVDAGKGKVV VADSKGAGGG IEADDEVDVS GRDIGIEGGK LRGKDVRLKA

2581 DTVKVATSMR YDDKGRLAAR GDGALDAQGG QLHIEAKRLE TAGATLKGGK VKLDVDDVKL
2641 GGVYEAGSSY ENKSSTPLGS LFAILSSTTE TNQSAHANHY GTRIEAGTLE GKMQNLEIEG
2701 GSVDAAHTDL SVARDARFKA AADFAHAEHE KDVRQLSLGA KVGAGGYEAG FSLGSESGLE
2761 AHAGRGMTAG AEVKVGYRAS HEQSSETEKS YRNANLNFGG GSVEAGNVLD IGGADINRNR
2821 YGGAAKGNAG TEEALRMRAK KVESTKYVSE QTSQSSGWSV EVASTASARS SLLTAATRLG
2881 DSVAQNVEDG REIRGELMAA QVAAEATQLV TADTAAVALS AGISADFDSS HSRSTSQNTQ
2941 YLGGNLSIEA TEGDATLVGA KFGGGDQVSL KAAKSVNLMA AESTFESYSE SHNFHASADA
3001 NLGANAVQGA VGLGLTAGMG TSHQITNETG KTYAGTSVDA ANVSIDAGKD LNLSGSRVRG
3061 KHVVLDVEGD INATSKQDER NYNSSGGGWD ASAGVAIQNR TLVAPVGSAG FNFNTEHDNS
3121 RLTNDGAAGV VASDGLTGHV KGDANLTGAT IADLSGKGNL KVDGAVNAQN LKDYRDKDGG
3181 SGGLNVGISS TTLAPTVGVA FGRVAGEDYQ AEQRATIDVG QTKDPARLQV GGGVKGTLNQ
3241 DAAQATVVQR NKHWAGGGSE FSVAGKSLKK KNQVRPVETP TPDVVDGPPSRPTTPPASPQ
3301 PIRATVEVSS PPPVSVATVE VVPRPKVETG SAASASAGGA QVVPVTPPKV EVAKVEVVPR
3361 PKVETAQPLP PRPVVAEKVT TPAVQPQLAK VETVQPVKPE TTKPLPKPLP VAKVTKAPPP
3421 VVETAQPLPP VKPQKATPGP VAEVGKATVT TVQVQSAPPK PAPVAKQPAP APKPKPKPKP
3481 KAERPKPGKT TPLSGRHVVQ QQVQVLQRQA SDINNTKSLP GGKLPKPVTV KLTDENGKPQ
3541 TYTINRREDL MKLNGKVLST KTTLGLEQTF RLRVEDIGGK NYRVFYETNK

pertactin outer membrane protein [Bordetella pertussis]

1 MNMSLSRIVK AAPLRRTTLA MALGALGAAP AAHADWNNQS IVKTGERQHG IHIQGSDPGG
61 VRTASGTTIK VSGRQAQGIL LENPAAELQF RNGSVTSSGQ LSDDGIRRFL GTVTVKAGKL
121 VADHATLANV GDTWDDDGIA LYVAGEQAQA SIADSTLQGA GGVQIERGAN VTVQRSAIVD
181 GGLHIGALQS LQPEDLPPSR VVLRDTNVTA VPASGAPAAV SVLGASELTL DGGHITGGRA
241 AGVAAMQGAV VHLQRATIRR GDAPAGGAVPGGAVPGGAVP GGFGPGGFGPVLDGWYGVDV
301 SGSSVELAQS IVEAPELGAA IRVGRGARVT VSGGSLSAPH GNVIETGGAR RFAPQAAPLS
361 ITLQAGAHAQ GKALLYRVLP EPVKLTLTGG ADAQGDIVAT ELPSIPGTSI GPLDVALASQ
421 ARWTGATRAV DSLSIDNATW VMTDNSNVGA LRLASDGSVD FQQPAEAGRF KVLTVNTLAG
481 SGLFRMNVFA DLGLSDKLVV MQDASGQHRL WVRNSGSEPA SANTLLLVQT PLGSAATFTL
541 ANKDGKVDIG TYRYRLAANG NGQWSLVGAK APPAPKPAPQ PGPQPPQPPQ PQPEAPAPQP
601 PAGRELSAAA NAAVNTGGVG LASTLWYAES NALSKRLGEL RLNPDAGGAW GRGFAQRQQL
661 DNRAGRRFDQ KVAGFELGAD HAVAVAGGRWHLGGLAGYTRGDRGFTGDGGGHTDSVHVGG
721 YATYIADSGF YLDATLRASR LENDFKVAGS DGYAVKGKYR THGVGASLEA GRRFTHADGW
781 FLEPQAELAV FRAGGGAYRA ANGLRVRDEG GSSVLGRLGL EVGKRIELAG GRQVQPYIKA
841 SVLQEFDGAG TVHTNGIAHR TELRGTRAEL GLGMAAALGR GHSLYASYEY SKGPKLAMPW
901 TFHAGYRYSW

Homology in sequences of above proteins

Pertactin: PKPAP- - - QPGPQPPQPPQPQPEAPAPQP
PKPAP- - - QP–P– P- - - P –P –P- A - -P- P

FHA protein: PKPAPVAKQPAPAPKPKPK PKPKAERPKP

Staphylococcus aureus

Cap5P [Staphylococcus aureus]

1 MCLNFREDNV MKKIMVIFGT RPEAIKMAPL VKEIDHNGNF EANIVITAQH RDMLDSVLSI
61 FDIQADHDLN IMQDQQTLAG LTANALAKLD SIINEEQPDM ILVHGDTTTT FVGSLAAFYH
121 QIPVGHVEAG LRTHQKYSPF PEELNRVMVS NIAELNFAPT VIAAKNLLFE NKDKERIFIT
181 GNTVIDALST TVQNDFVSTI INKHKGKKVI LLTAHRRENI GEPMHQIFKA VRDLADEYKD
241 VVFIYPMHRN PKVRAIAEKY LSGRNRIELI EPLDAIEFHN FTNQSYLVLT DSGGIQEEAP
301 TFGKPVLVLR NHTERPEGVE AGTSRVIGTD YDNIVRNVKQ LIEDDEAYQR MSQANNPYGD
361 GQASRRICEA IEYYFGLRTD KPDEFVPLRH K

Protein A signal fusion protein.

1 MKKKNIYSIR KLGVGIASVT LGTLLISGGV TPAANAAQHD EAVDNKFNKE QQNAFYEILH
61 LPNLNEEQRN AFIQSLKDDP SQSANLLAEA KKLNDAQAPK VDNKFNKEQQ NAFYEILHLP
121 NLNEEQRNAF IQSLKDDPSQ SANLLAEAKK LNDAQAPKVD ANSSSVPGDP LESTCRHASL
181 ALAVVLQRRD WENPGVTQLN RLAAHPPFAS WRNSEEARTD RPSQQLRSLNGEWRFRCNGW
241 R

Biofilm-associated surface protein [Staphylococcus aureus]

1 MGNKQGFLPN KLNKYSIRKF TVGTASLLVG TTLFFGIGSE AQAAELDTIT KEDVKSQDKG
61 EALDIKNIKESEKDVTTEDDNNAEVQNSAQTVDKSENSNDTAVESTNDSV KTDETKETSE
121 NKSAQDDDNIKEDSNTQEESTNTSSQSSEVPQTKKDTNETSETAIDEDASTKEQNNKDND
181 TAQDDDNIKE DSNTQEESTN TSSQSSEVPQ TKKEQPDKSS NSIKEPDKQQ EEVAKEEKAI
241 TEIADKNKEL ELKNNKTDKN EESELESNLS SSENKKDTVE SFLNSQLSDS ETKKIMENAN
301 IDYDKATDEE INTEILRASL IEMANNKKKT ETLATPQRTM FRAMATPTAL RAAVNQDEEL
361 QKSLGYTDNY TFASMLFDPG KLDSDDALNS NIIPFDLHSY MSGANSGNRY KIDLKLDPII
421 AEHVTKISAN PSGSNKPVEF VRNKDENGNL TDTWEVNFIR ANDGLFGGAE ILSQYTAKNG
481 KIELDDTVGN IISNAGNLSN NKLNHQVFVR DSRENKIVRT SESSGYFLTK ADDDLVNLEN
541 NVSTENNNAF KASSGSATYN ENVGEFGGIL IDQQIMKNGI FSYSKTKANQ WAYNYQIDKD
601 LLPYIEGVEL HQYDYKGLNG FDKNYDAKNK VADLTIDEVG NGTITSDNLN KLIEFNNALP
661 ETVGVRVVLK LNKSVNNILT KDAKYDSEGN LIRETTKQKE DFTFAGYLTD SKGALINNTL
721 GTSTLALQDY DKDGLLDRYE RQLSLSDAEN EDTDGDGKND GDEVVNYKTS PLVGKPQAAD
781 ITTEDTVVSG SVPLKEGAAT QTAKVINAEG TTVGTATVNS DGTFSVSIPN SPEGTYTIAI
841 DSPNYDNDEV NTFEIVDNSK LPAPSINPVD DNDQQIVVNG TSGSTVTVTD SNNNVLGTVT
901 IPADDTSAAI NVDTPLEAGT VLTSTASKDG KTSDVSDQIT VTDATAPDAP TLDEVNTDAT
961 QVTGQAEPNS TVKLTFPDGT TATGTADDQG NYTIDIPSNV DLNGGEELQV TATDKDGNTS
1021 EPSSANVTDT TAPDAPTVND VTSDATQVTG QAEPNSTVKL TFPDGTTATGTADDQGNYTI
1081 DIPSNVDLNG GEELQVTATD KDGNTSEPSS ANVTDTTAPD APTVNDVTSDATQVTGQAEP
1141 NSTVKLTFPD GTTATGTADD QGNYTIDIPS NVDLNGGEEL QVTATDKDGNTSEPSSANVT
1201 DTTAPDAPTV NDVTSDATQV TGQAEPNSTV KLTFPDGTTA TGTADDQGNYTIDIPSNVDL
1261 NGGEELQVTA TDKDGNTSEP SSANVTDTTA PDAPTVNDVT SDATQVTGQAEPNSTVKLTF
1321 PDGTTATGTA DDQGNYTIDI PSNVDLNGGE ELQVTATDKD GNTSEPSSANVTDTTAPDAP
1381 TVNDVTSDAT QVTGQAEPNS TVKLTFPDGT TATGTADDQG NYTIDIPSNVDLNGGEELQV
1441 TATDKDGNTS EPSSANVTDT TAPDAPTVND VTSDATQVTG QAEPNSTVKLTFPDGTTATG
1501 TADDQGNYTI DIPSNVDLNG GEELQVTATD KDGNTSEPSS ANVTDTTAPDAPTVNDVTSD
1561 ATQVTGQAEP NSTVKLTFPD GTTATGTADD QGNYTIDIPS NGDLNGGEELQVTATDKDGN
1621 TSEPSSANVT DTTASDAPTV NDVTSDATQV TGQAEPNSTV KLTFPDGTTATGTADDQGNY
1681 TIDIPSNVDL NGGEELQVTA TDKDGNTSEP SSANVTDTTA PDAPTVNDVTSDATQVTGQA
1741 EPNSTVKLTF PDGTTATGTA DDQGNYTIDI PSNVDLNGGE ELQVTATDKDGNTSEPKLTN
1801 VTDTTASDAP TVNDVTSDAS QVTGQAEPNS TVKLTFPDGT TATGTADDQGNYTIDIPSNV
1861 DLNGGEELQV TATDKDGNTS EPSSANVTDT TAPDAPTVND VTSDATQVTGQAEPNSTVKL
1921 TFPDGTTATG TADDQGNYTI DIPSNGDLNG GEKLQVTATD KDGNTSEPSSANVTDTTAPD
1981 APTVNDVTSD ATQVTGQAEP NSTVKLTFPD GTTATGTADD QGNYTIDIPSNVDLNGGEEL
2041 QVTATDKDGN TSEPSSANVT DTTAPDAPTV NDVTSDATQV TGQAEPNSTVKLTFPDGTTA
2101 TGTADDQGNY TIDIPSNVDL NGGEELQVTA TDKDGNTSES TNTTIIDSDDNSDNGNNSGA
2161 GDTSDSDDNSGNGDNSGAGD NSDSDDNSDNGNNSGAGDNSDSDDNSDNEDNSSSNKDSIN
2221 QDSNVNSNDS KHDKQNELPE TGEKEVRNGT LFGTLFAGLG SLLLFTKRRR KENDKK

Streptococcus agalactiae

Protein immunoglobulin-a-binding beta antigen

1 AIKQQTIFDI DNAKTEVEID NLVHDAFSKM NATVAKFQKG LETNTPETPD TPKIPELPQA
61 PDTPQAPDTP HVPESPKAPE APRVPESPKT PDTPHVPESP KAPEAPRVPE SPKTPDTPHV
121 PESPKAPEAP RVPESPKTPD TPHVPESPKA PEAPRVPESP KTPDTPHVPE SPKAPEAPRV
181 PESPKTPEAP HVPESPKTPE APKIPEPPKT PDVPKLPDVP KLPDVPKLPD APKLPDGLNK
241 VGQAVFTSTD GNTKVTVFD KPTDADKLHL KEVTTKELA

Penicillin binding protein 1a

1 MITIKKESVI KLLKYAFGII MGFIILAIVI GGLLFAYYVS RSPKLTDQAL KSVNSSLVYD
61 GNNKLIADLG SEKRESVSAD SIPLNLVNAI TSIEDKRFFK HRGVDIYRIL GAAWHNLVSS
121 NTQGGSTLDQ QLIKLAYFST NKSDQTLKRK SQEVWLALQM ERKYTKEEIL TFYINKVYMG
181 NGNYGMRTTA KSYFGKDLKE LSIAQLALLA GIPQAPTQYD PYKNPESAQT RRNTVLQQMY
241 QDKNISKKEY DQAVATPVTD GLKELKQKST YPKYMDNYLK QVISEVKQKT GKDIFTAGLK
301 VYTNINTDAQ KQLYDIYNSD TYIAYPNNEL QIASTIMDAT NGKVIAQLGG RHQNENISFG
361 TNQSVLTDRD WGSTMKPISA YAPAIDSGVY NSTGQSLNDS VYYWPGTSTQ LYDWDRQYMG
421 WMSMQTAIQQ SRNVPAVRAL EAAGLDEAKS FLEKLGIYYP EMNYSNAISS NNSSSDAKYG
481 ASSEKMAAAY SAFANGGTYY KPQYVNKIEF SDGTNDTYAA SGSRAMKETT AYMMTDMLKT
541 VLTFGTGTKA AIPGVAQAGK TGTSNYTEDE LAKIEATTGI YNSAVGTMAP DENFVGYTSK
601 YTMAIWTGYK NRLTPLYGSQ LDIATEVYRA MMSYLTGGYS ADWTMPEGLY RSGSYLYING
661 TTTTGTYSSS VYKNIYQNSG QSSQSSSSTS SEKQKEDKNT ANDANSSSPQ VETPNNGNAT

721 TPNNSNQTVP GTGHGNGNGN NNTVPNGN

Pi-2a ancillary protein 2
1 MKKIRKSLGL LLCCFLGLVQ LAFFSVASVN ADTPNQLTIT QIGLQPNTTE EGISYRLWTV
61 TDNLKVDLLS QMTDSELNQK YKSILTSPTD TNGQTKIALP NGSYFGRAYK ADQSVSTIVP
121 FYIELPDDKL SNQLQINPKR KVETGRLKLI KYTKEGKIKK RLSGVIFVLY DNQNQPVRFK
181 NGRFTTDQDG ITSLVTDDKG EIEVEGLLPG KYIFREVKAL TGYRISMKDA VVAVVANKTQ
241 EVEVENEKET PPPTNPKPSQ PLFPQSFLPK TGMIIGGGLT ILGCIILGIL FIFLRKTKNS
301 KSERNDTV

Streptococcus pneumoniae

PspA [Streptococcus pneumoniae]
1 MNKKKMILTS LASVAILGAG FVASSPTFVR AEEAPVANQS KAEKDYDAAV KKSEAAKKDY
61 ETAKKKAEDA QKKYDEDQKK TEAKAEKERK ASEKIAEATK EVQQAYLAYL QASNESQRKE
121 ADKKIKEATQ RKDEAEAAFA TIRTTIVVPE PSELAETKKK AEEATKEAEV AKKKSEEAAK
181 EVEVEKNKIL EQDAENEKKI DVLQNKVADL EKGIAPYQNE VAELNKEIAR LQSDLKDAEE
241 NNVEDYIKEG LEQAITNKKA ELATTQQNID KTQKDLEDAE LELEKVLATL DPEGKTQDEL
301 DKEAAEAELN EKVEALQNQV AELEEELSKL EDNLKDAETN NVEDYIKEGL EEAIATKKAE
361 LEKTQKELDA ALNELGPDGD EEETPAPAPQPEKPAEEPEN PAPAPKPEKSADQQAEEDYA
421 RRSEEEYNRLTQQQPPKAEKPAPAPQPEQPAPAPKIGWKQENGMWYFYNTDGSMATGWLQ
481NNGSWYYLNSNGAMATGWLQYNGSWYYLNANGAMATGWLQYNGSWYYLNANGAMATGWLQ
541YNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWAKVHGSWYYLNANGSMATGWVK 601
DGETWYYLEA SGSMKANQWF QVSDKWYYVN GLGSLSVNTT VDGYKVNANG EWV

cbpA [Streptococcus pneumoniae GA19998] gram pos signal peptide
1 MFASKSERKV HYSIRKFSIG VASVVVASLF LGGVVHAEEV RRGNNLTVTS SGDEVESHYQ
61 SILEKVRKSL EKDRHTQNVD LIKKLQDIKR TYLYNLKEKP EAELTSKTKK ELDAAFEKFK
121 KEPELTKKLA EAEKKAKDQK EEDHRNYPTN TYKTIELEIA EAEVGVAKAE LELVQAQVQI
181 PQDTEKINAA KAKVEAAKSN VKKLEKIKSD IEKTYLYKLD NSTKETPKSR VRRNSPQVGD
241 SRELKETIDK AKETLSTYMV TRLTKLDPSV FWFADLLMDA KKVVEEYKTK LEDASDKKSV
301 EDLRKEAEGK IESLIVTHQN REKENQPAPQ PGGQAGGSMV VPPVTQTPPS TSQSPGQKAT
361 EAEKKKLQDL IRQFQEALNK LDDETKTVPD GAKLTGEAGK AYNETRTYAK EVVDKSKKLL
421 SQTAVTMDEL AMQLTKLNDA MSKLKEAKAK LVPEVKPQPE NPEPKPQPEG EKPSVPDINQ
481 EKEKAKLAIA TYMSKILDDI KKHHLKKEKH HQIVALIKDL DKLKKQALSE IDNVNTKVEI
541 ENTVHKVFAD MDTVVTKFQK GLIQNTPQVP EAPKSPEVPK VSDTPKAPDT PQVPEAPKSP 601
EVPKVPEAPK APDTPQVPEA PKSPEVPKVP DTPKAPDTPQ VPEAPKAPDT PQIPEAPAPE
661 TPAPAPEAPKTGWKQENGMWYFYNTDGSMATGWLEYNGSWYYLNANGAMATGWLEYNGSW
721 YYLNTNGAMETGWLEYNGSWYYLNTNGAMETGWLEYNGSWYYLNTNGAME TGWLEYNGSW
781 YYLNTNGAMETGWLEYNGSWYYLNTNGAMETGWLEYNGSWYYLNANGSMA TGWLKDGDTW
841 YYLEASGAMK ESQWFKVSDK WYYVNGSGAL AVNTTVGGYR VNANGKWVN

Pseudomonas aeruginosa

ExoU [Pseudomonas aeruginosa PA103].
1 MHIQSLGATA SSLNQEPVET PSQAAHKSAS LRQEPSGQGL GVALKSTPGI LSGKLPESVS
61 DVRFSSPQGQ GESRTLTDSA GPRQITLRQF ENGVTELQLS RPPLTSLVLS GGGAKGAAYP
121 GAMLALEEKG MLDGIRSMSG SSAGGITAAL LASGMSPAAF KTLSDKMDLI SLLDSSNKKL
181 KLFQHISSEI GASLKKGLGN KIGGFSELLL NVLPRIDSRA EPLERLLRDE TRKAVLGQIA
241 THPEVARQPT VAAIASRLQS GSGVTFGDLD RLSAYIPQIK TLNITGTAMF EGRPQLVVFN
301 ASHTPDLEVA QAAHISGSFP GVFQKVSLSD QPYQAGVEWT EFQDGGVMIN VPVPEMIDKN
361 FDSGPLRRND NLILEFEGEA GEVAPDRGTR GGALKGWVVG VPALQAREML QLEGLEELRE
421 QTVVVPLKSE RGDFSGMLGG TLNFTMPDEI KAHLQERLQE RVGEHLEKRL QASERHTFAS 481
LDEALLALDD SMLTSVAQQN PEITDGAVAF RQKARDAFTE LTVAIVSANG LAGRLKLDEA
541 MRSALQRLDA LADTPERLAW LAAELNHADN VDHQQLLDAM RGQTVQSPVL AAALAEAQRR 601
KVAVIAENIR KEVIFPSLYR PGQPDSNVAL LRRAEEQLRH ATSPAEINQA LNDIVDNYSA
661 RGFLRFGKPL SSTTVEMAKA WRNKEFT

exotoxin A, partial [Pseudomonas aeruginosa].
1 ALLERNYPTG AEFLGDGGDV SFSTRGTQNW TVERLLQAHR QLEERGYVFV GYHGTFLEAA
61 QSIVFGGVRA RSQDLDAIWR GFYIAGDPAL AYGYAQDQEP DARGRIRIGA LLRVYVPRSS
121 LPGFYRTGLT LAAPEAAGEV ERLIGHPLPL RLDAITGPEE EGGRLETILG WPLAERTVVI
181 PSAIPTDPRN VGGDLDPSSI PDKEQAISAL PDYASQPGKP PREDLK

toxA gene product [Pseudomonas aeruginosa PAO1]
1 MHLTPHWIPL VASLGLLAGG SFASAAEEAF DLWNECAKAC VLDLKDGVRS SRMSVDPAIA
61 DTNGQGVLHY SMVLEGGNDA LKLAIDNALS ITSDGLTIRL EGGVEPNKPV RYSYTRQARG
121 SWSLNWLVPI GHEKPSNIKV FIHELNAGNQ LSHMSPIYTI EMGDELLAKL ARDATFFVRA
181 HESNEMQPTL AISHAGVSVV MAQAQPRREK RWSEWASGKV LCLLDPLDGV YNYLAQQRCN
241 LDDTWEGKIY RVLAGNPAKH DLDIKPTVIS HRLHFPEGGS LAALTAHQAC HLPLETFTRH
301 RQPRGWEQLE QCGYPVQRLV ALYLAARLSW NQVDQVIRNA LASPGSGGDL GEAIREQPEQ
361 ARLALTLAAA ESERFVRQGT GNDEAGAASA DVVSLTCPVA AGECAGPADS GDALLERNYP
421 TGAEFLGDGG DISFSTRGTQ NWTVERLLQA HRQLEERGYV FVGYHGTFLE AAQSIVFGGV 481
RARSQDLDAI WRGFYIAGDP ALAYGYAQDQ EPDARGRIRN GALLRVYVPR SSLPGFYRTG
541 LTLAAPEAAG EVERLIGHPL PLRLDAITGP EEEGGRLETI LGWPLAERTV VIPSAIPTDP
601 RNVGGDLDPS SIPDKEQAIS ALPDYASQPG KPPREDLK

Escherichia coli

Surface protein [E. coli]
1 MTTPNPLAKT KGAGTTFWMY TGKGDAFANP LSDTDWLRLA MVKDLQPGEM TADAEDDTYL
61 DDEDADWKTT TQGQKSVGDT SATLAWRPGD SGQKKLVQLF DSGEVCAFRI KYPNGTVDVF
121RGWLSSLGKT IASKDVMTRT VKISGVGRPY LAEEGXETVG VTGLTVAPAS ASVKAGATTT
181LTFTVKPDGA SDKAISVHSS DPQTASVTLS GLVATVKGVK QGSVSIVGMT SDGEFVAVAA
241VTVSAP

Plasmodium falciparum - malaria

PfMSA2
1 SIRRSMAESK SPTGTGASGS AGSGDGASGS AGSGDGASGS AGSGDGAVAS ARNGANPGAD
61 AEGSSSTPAT TTTTTTTTTT TTTNDAEAST STSSENPNHN NAETNPKGNG EVQEPNQANK
121 ETQNNSNVQQ DSQTKSNVPP TQDADTKSPT AQPEQAENSA PTAEQTESPE LQSAPEN

P proline P where ginger enzymes digest protein
T threonine that can bind to carbohydrate
S serine that can bind to carbohydrate