Supplementary File: gp114-l15-16 and gp34-35-36 amino acid percent identity and amino acid conservation and secondary structure alignment.

Percent amino acid identity was performed using ClustalW default parameters within Geneious (Geneious version 6.1.4, http://www.geneious.com/) (Kearse et al., 2012). For gpl15 / gp35, the respective proteins were divided into N-terminal, middle, and C-terminal for identity determinations. Amino acid conservation and secondary structure prediction alignment was performed using PRALINE Multiple Sequence Alignment set at default parameters (http://www.ibi.vu.nl/programs/pralinewww/) (Simossis & Heringa, 2005). The color key for the amino acid conservation output is as shown. Alignment comparison of predicted secondary structure used DSSP (Kabsch & Sander, 1983) and PSIPRED (Jones, 1999) and was performed separately for each c2-type (gpl14-l15-l16) and bIL67-type (gp34-35-36) group of phage proteins; color assignments are red for helix and blue for sheet.

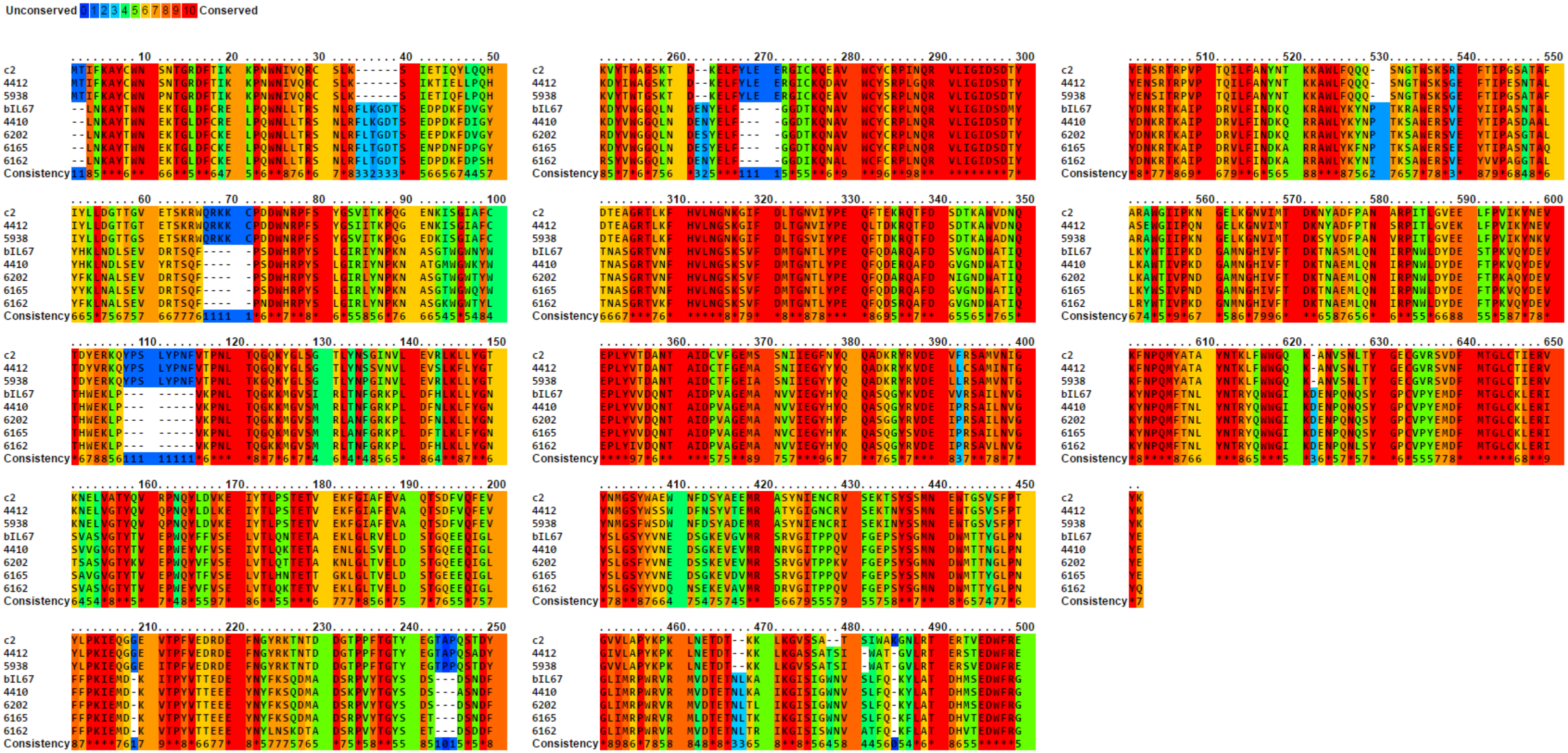
Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Molec Biol* **292**,195-202.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-637.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S. & other authors (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Simossis, V.A. & Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* **33** (Web Server issue), W289–W294.

gpl14 – gp34 amino acid alignment / % identity

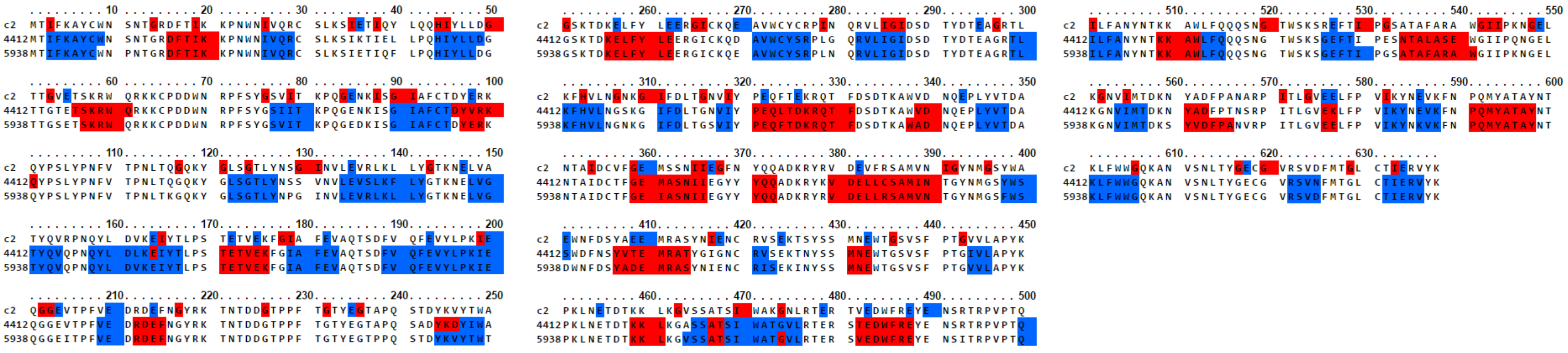


Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

```
           ........10.........20.........30.........40.........50
c2         MTIFKAYCWN SNTGRDFTIK KPNWNIVQRC SLK------S IETIQYLQQH
4412       MTIFKAYCWN SNTGRDFTIK KPNWNIVQRC SLK------S IKTIELLPQH
5938       MTIFKAYCWN PNTGRDFTIK KPNWNIVQRC SLK------S IETIQFLPQH
bIL67      --LNKAYTWN EKTGLDFCRE LPQWNLLTRS NLRFLKGDTS EDPDKFDVGY
4410       --LNKAYTWN EKTGLDFCRE LPQWNLLTRS NLRFLKGDTS EDPDKFDIGY
6202       --LNKAYTWN EKTGLDFCKE LPQWNLLTRS NLRFLTGDTS EEPDKFDVGY
6165       --LNKAYTWN EKTGLDFCKE LPQWNLLTRS NLRFLTGDTS ENPDNFDPGY
6162       --LNKAYTWN EKTGLDFCKE LPQWNLLTRS NLRFLTGDTS EDPDKFDPSH
Consistency1185***6** 66**5**647 5*6**876*6 7*8332333* 5665674457

           ........60.........70.........80.........90........100
c2         IYLLDGTTGV ETSKRWQRKK CPDDWNRPFS YGSVITKPQG ENKISGIAFC
4412       IYLLDGTTGT ETSKRWQRKK CPDDWNRPFS YGSIITKPQG ENKISGIAFC
5938       IYLLDGTTGS ETSKRWQRKK CPDDWNRPFS YGSVITKPQG EDKISGIAFC
bIL67      YHKLNDLSEV DRTSQF---- -PSDWHRPYS LGIRIYNPKN ASGTWGWNYW
4410       YHKLNDLSEV YRTSQF---- -PSDWHRPYS LGIRIYNPKN ATGMWGWKYW
6202       YFKLNALSEV DRTSQF---- -PSDWHRPYS LGIRLYNPKN ASGTWGWTYW
6165       YYKLNALSEV DRTSQF---- -PSDWHRPYS LGIRLYNPKN ASGTWGWQYW
6162       YFKLNALSEV DRTSQF---- -PNDWHRPYS LGIRLYNPKN ASGKWGWTYL
Consistency665*756757 6677761111 1*6**7**8* 6*55856*76 66545*5484

           .......110........120........130........140........150
c2         TDYERKQYPS LYPNFVTPNL TQGQKYGLSG TLYNSGINVL EVRLKLLYGT
4412       TDYVRKQYPS LYPNFVTPNL TQGQKYGLSG TLYNSSVNVL EVSLKFLYGT
5938       TDYERKQYPS LYPNFVTPNL TKGQKYGLSG TLYNPGINVL EVRLKLLYGT
bIL67      THWEKLP--- -----VKPNL TQGKKMGVSI RLTNFGRKPL DFHLKLLYGN
4410       THWEKLP--- -----VKPNL TQGKKMGVSM RLTNFGRKPL DFNLKLFYGN
6202       THWEKLP--- -----VKPNL TQGKKMGVSM RLANFGRKPL DFNLKLLYGN
6165       THWEKLP--- -----VKPNL TQGQKMGVSM RLANFGRKPL DFTLKLFYGN
6162       THWEKLP--- -----VKPNL TQGKKMGVSM RLTNFGRKPL DFHLKLLYGN
Consistency*678856111 11111*6*** *8*7*6*7*4 6*4*48565* 864**87**6

           .......160........170........180........190........200
c2         KNELVATYQV RPNQYLDVKE IYTLPSTETV EKFGIAFEVA QTSDFVQFEV
4412       KNELVGTYQV QPNQYLDLKE IYTLPSTETV EKFGIAFEVA QTSDFVQFEV
5938       KNELVGTYQV QPNQYLDVKE IYTLPSTETV EKFGIAFEVA QTSDFVQFEV
bIL67      SVASVGTYTV EPWQYFFVSE LVTLQNTETA EKLGLRVELD STGQEEQIGL
4410       SVVGVGTYTV EPWEYVFVSE IVTLQKTETA ENLGLSVELD STGQEEQIGL
6202       TSASVGTYKV EPWQYVFVSE LVTLQTTETA KNLGLTVELD STGQEEQIGL
6165       SAVGVGTYTV EPWQYTFVSE LVTLHNTETT GKLGLTVELD STGEEEQIGL
6162       SVASVGTYTV EPWEYVFVSE LVTLQKTETV EKLGLTVELD STGQEEQIGL
Consistency6454*8**5* 7*48*5597* 86**55***6 777*856*75 7*7655*757

           .......210........220........230........240........250
c2         YLPKIEQGGE VTPFVEDRDE FNGYRKTNTD DGTPPFTGTY EGTAPQSTDY
4412       YLPKIEQGGE VTPFVEDRDE FNGYRKTNTD DGTPPFTGTY EGTAPQSADY
5938       YLPKIEQGGE ITPFVEDRDE FNGYRKTNTD DGTPPFTGTY EGTPPQSTDY
bIL67      FFPKIEMD-K ITPYVTTEDE YNYFKSQDMA DSRPVYTGYS DS---DSNDF
4410       FFPKIEMD-K VTPYVTTEEE YNYFKSQDMA DSKPVYTGYS DS---ASNDF
6202       FFPKIEMD-K VTPYVTTEEE YNYFKSQDMA DSRPVYTGYS DS---DSNDF
6165       FFPKIEMD-K VTPYVTTEEE YNYFKSQDMA DSRPVYTGYS ET---DSNDF
6162       FFPKIEMD-K VTPYVTTEEE YNYLNSKDTA DSRPVYTGYS ET---DSDDF
Consistency87****7617 9**8*6677* 8*57775765 *75*58**55 851015*5*8

           .......260........270........280........290........300
c2         KVYTWAGSKT D--KELFYLE ERGICKQEAV WCYCRPINQR VLIGIDSDTY
4412       KDYIWAGSKT D--KELFYLE ERGICKQDAV WCYSRPLGQR VLIGIDSDTY
5938       KVYTWTGSKT D--KELFYLE ERGICKQEAV WCYSRPLNQR VLIGIDSDTY
bIL67      KDYVWGGQLN DENYELF--- -GGDTKQNAV WCYCRPLNQR VLIGIDSDMY
4410       KDYVWGGQLN DENYELF--- -GGDTKQNAV WCYCRPLNQR VLIGIDSDTY
6202       RDYVWGGQLN DESYELF--- -GGDTKQNAV WCYCRPLNQR VLIGIDSDTY
6165       KDYVWGGQLN DESYELF--- -GGDTKQNAV WCYCRPLNQR VLIGIDSDTY
6162       RSYVWGGQLN DENYELF--- -GGDIKQNAL WCFCRPLNQR VLIGIDSDIY
Consistency85*7*6*756 *325***111 15*55**6*9 **96**98** ********7*

           .......310........320........330........340........350
c2         DTEAGRTLKF HVLNGNKGIF DLTGNVIYPE QFTEKRQTFD SDTKAWVDNQ
4412       DTEAGRTLKF HVLNGSKGIF DLTGNVIYPE QLTDKRQTFD SDTKAWVDNQ
5938       DTEAGRTLKF HVLNGNKGIF DLTGSVIYPE QFTDKRQTFD SDTKAWADNQ
bIL67      TNASGRTVNF HVLNGSKSVF DMTGNTLYPE QFQDARQAFD SVGNDWATIQ
4410       TNASGRTVNF HVLNGSKSVF DMTGNTLYPE QFQDERQAFD GVGNDWATIQ
6202       TNASGRTVNF HVLNGSKSVF DMTGNTLYPE QFQDARQAFD NIGNDWATIQ
6165       TNASGRTVNF HVLNGSKSVF DMTGNTLYPE QFQDDRQAFD GVGNDWATIQ
6162       TNASGRTVKF HVLNGSKSVF DMTGNTLYPE QFQDSRQAFD GVGNDWATIQ
Consistency6667***76* *****8*79* *8**878*** *8695**7** 65565*765*

           .......360........370........380........390........400
c2         EPLYVTDANT AIDCVFGEMS SNIIEGFNYQ QADKRYRVDE VFRSAMVNIG
4412       EPLYVTDANT AIDCTFGEMA SNIIEGYYYQ QADKRYKVDE LLCSAMINTG
5938       EPLYVTDANT AIDCTFGEIA SNIIEGYYYQ QADKRYRVDE LLRSAMVNTG
bIL67      EPLYVVDQNT AIDPVAGEMA NVVIEGYHYQ QASQGYKVDE VVRSAILNVG
4410       EPLYVVDQNT AIDPVAGEMA NVVIEGYHYQ QASQGYRVDE IPRSAILNVG
6202       EPLYVVDQNT AIDPVAGEMA NVVIEGYHYP QASGGYRVDE IPRSAILNVG
6165       EPLYVVDQNT AIDPVAGEMA NVCIEGYHYK QASQGYSVDE IPRSAILNVG
6162       EPLYIVDQNT AIDPVAGEMA NVVIEGYHYQ QASQGYRVDE IPRSAVLNVG
Consistency****97*6** ***575**89 757***96*7 **765*7*** 837**78*7*

           .......410........420........430........440........450
c2         YNMGSYWAEW NFDSYAEEMR ASYNIENCRV SEKTSYSSMN EWTGSVSFPT
4412       YNMGSYWSSW DFNSYVTEMR ATYGIGNCRV SEKTNYSSMN EWTGSVSFPT
5938       YNMGSFWSDW NFDSYADEMR ASYNIENCRI SEKINYSSMN EWTGSVSFPT
bIL67      YSLGSYYVNE DSGKEVGVMR SRVGITPPQV FGEPSYSGMN DWMTTYGLPN
4410       YSLGSYYVNE DSGKEVEVMR NRVGITPPQV FGEPSYSSMN DWMTTNGLPN
6202       YSLGSFYVNE DSSKEVEVMR SRVGVTPPKV FGEPSYSGMN DWMTTNGLPN
6165       YSLGSYYVNE DSGKEVDVMR SRVGITPPQV FGEPSYSSMN DWMTTYGLPN
6162       YSLGSYYVDQ NSEKEVAVMR DRVGITPPQV FGEPSYSGMN DWMTTYGLPN
Consistency*78**87664 75475745** 5667955579 55758**7** 8*657477*6

           .......460........470........480........490........500
c2         GVVLAPYKPK LNETDT--KK LKGVSSA--T SIWAKGNLRT ERTVEDWFRE
4412       GIVLAPYKPK LNETDT--KK LKGASSATSI -WAT-GVLRT ERSTEDWFRE
5938       GVVLAPYKPK LNETDT--KK LKGVSSATSI -WAT-GVLRT ERSVEDWFRE
bIL67      GLIMRPWRVR MVDTETNLKA IKGISIGWNV SLFQ-KYLAT DHMSEDWFRG
4410       GLIMRPWRVR MVDTETNLKA IKGISIGWNV SLFQ-KYLAT DHMSEDWFRG
6202       GLIMRPWRVR MVDTETNLTL IKGISIGWNV SLFQ-KYLAT DHMSEDWFRG
6165       GLIMRPWRVR MLDTETNLTK IKGISIGWNV SLFQ-KFLAT DHVTEDWFRG
6162       GLIMRPWRVR MVDTETNLTR IKGISISWNV ATFQ-KFLAT DHVTEDWFRG
Consistency*8986*7858 848*8*3365 8**8*56458 4456054*6* 8655*****5

           .......510........520........530........540........550
c2         YENSRTRPVP TQILFANYNT KKAWLFQQQ- SNGTWSKSRE FTIPGSATAF
4412       YENSRTRPVP TQILFANYNT KKAWLFQQQ- SNGTWSKSGE FTIPESNTAL
5938       YENSITRPVP TQILFANYNT KKAWLFQQQ- SNGTWSKSGE FTIPGSATAF
bIL67      YDNKRTKAIP DRVLFINDKQ KRAWLYKYNP TKRAWERSVE YIIPASNTAL
4410       YDNKRTKAIP DRVLFINDKQ RRAWLYKYNP TKSAWERSVE YTIPASDAAL
6202       YDNKRTKAIP DRVLFINDKQ KRAWLYKYNP TKSAWERSVE YTIPASGTAL
6165       YDNKRTKAIP DRVLFINDKA RRAWLYKFNP TKSAWERSEE YTIPASNTAQ
6162       YDNKRTKAIP DRVLFINDKA RRAWLYKYNT TKSAWERSVE YVVPAGGTAL
Consistency*8*77*869* 679**6*565 88***87562 7657*78*3* 879*6848*6

           .......560........570........580........590........600
c2         ARAWGIIPKN GELKGNVIMT DKNYADFPAN ARPITLGVEE LFPVIKYNEV
4412       ASEWGIIPQN GELKGNVIMT DKNYADFPTN SRPITLGVEK LFPVIKYNEV
5938       ARAWGIIPKN GELKGNVIMT DKSYVDFPAN VRPITLGVEE LFPVIKYNKV
bIL67      LKYWTIIPKD GAMNGHIVFT DKTNASMLQN IRPNWLDYDE STPKVQYDEV
4410       LKAWTIVPKD GAMNGHIVFT DKTNAEMLQN IRPNWLDYDE FTPKVQYDEV
6202       LKAWTIVPND GAMNGHIVFT DKTNAEMLQN IRPNWLDYDE FTPKAQYDEV
6165       LKYWSIVPND GAMNGHIVFT DKTNAEMLQN IRPNWLDYDE FTPKVQYDEV
6162       LRYWTIVPKD GNMNGHIVFT DKTNAEMLQN IRPNWLDYDE FTPKVQYDEV
Consistency674*5*9*67 *586*7996* **6587656* 6**55*6688 55*587*78*

           .......610........620........630........640........650
c2         KFNPQMYATA YNTKLFWWGQ K-ANVSNLTY GECGVRSVDF MTGLCTIERV
4412       KFNPQMYATA YNTKLFWWGQ K-ANVSNLTY GECGVRSVNF MTGLCTIERV
5938       KFNPQMYATA YNTKLFWWGQ K-ANVSNLTY GECGVRSVDF MTGLCTIERV
bIL67      KYNPQMFTNL YNTRYQWWGI KDENPQNQSY GPCVPYEMDF MTGLCKLERI
4410       KYNPQMFTNL YNTRYQWWGI KDENPQNQSY GPCVPYEMDF MTGLCKLERI
6202       KYNPQMFTNL YNTRYQWWGI KDENPQNQSY GPCVPYEMDF MTGLCKLERI
6165       KYNPQMFTNL YNTRYQWWGI KDENPQNQSY GPCVPYEMDF MTGLCKLERI
6162       KYNPQMFTNL YNTRYQWWGI KDENPQNLSY GPCVPYEMDF MTGLCKLERI
Consistency*8****8766 ***865***5 *36*57*57* *6*555778* *****68**9

           ..
c2         YK
4412       YK
5938       YK
bIL67      YE
4410       YE
6202       YE
6165       YE
6162       YQ
Consistency*7
```

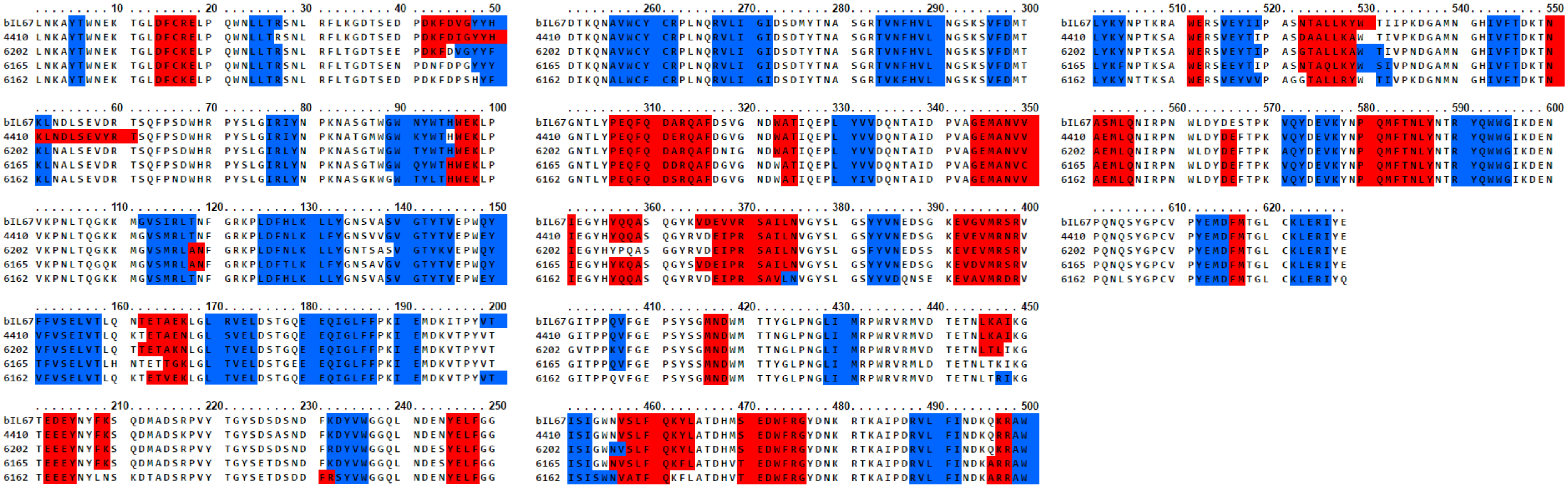
	c2	D4412	M5938	bIL67	D4410	M6162	M6165	M6202
c2		90	94	42	42	41	41	41
D4412	90		91	42	41	40	41	41
M5938	94	91		41	41	41	41	40
bIL67	42	42	41		94	89	91	93
D4410	42	41	41	94		89	92	93
M6162	41	40	41	89	89		89	89
M6165	41	41	41	91	92	89		92
M6202	41	40	40	93	93	89	92	

gpl14 secondary structure prediction



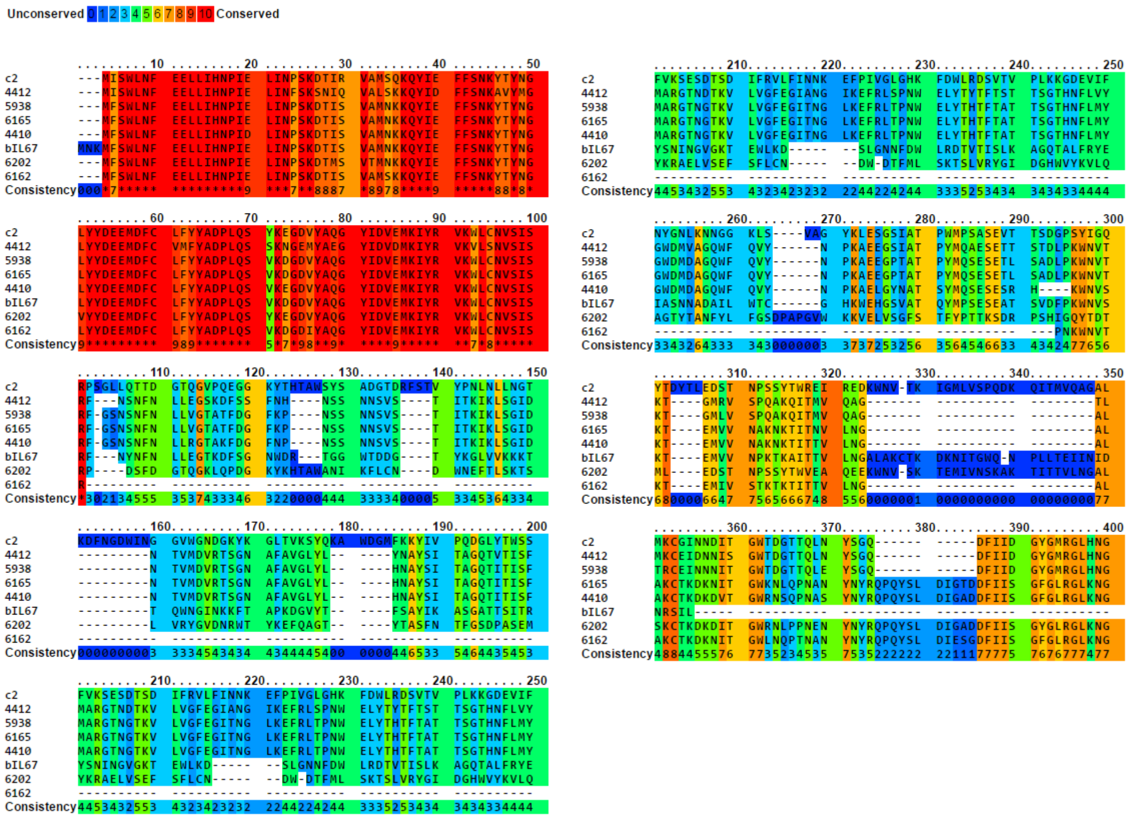
```
    .........10.........20.........30.........40.........50
c2   MTIFKAYCWN SNTGRDFTIK KPNWNIVQRC SLKSIETIQY LQQHIYLLDG
4412 MTIFKAYCWN SNTGRDFTIK KPNWNIVQRC SLKSIKTIEL LPQHIYLLDG
5938 MTIFKAYCWN PNTGRDFTIK KPNWNIVQRC SLKSIETIQF LPQHIYLLDG

    .........60.........70.........80.........90.........100
c2   TTGVETSKRW QRKKCPDDWN RPFSYGSVIT KPQGENKISG IAFCTDYERK
4412 TTGTETSKRW QRKKCPDDWN RPFSYGSIIT KPQGENKISG IAFCTDYVRK
5938 TTGSETSKRW QRKKCPDDWN RPFSYGSVIT KPQGEDKISG IAFCTDYERK

    .........110.........120.........130.........140.........150
c2   QYPSLYPNFV TPNLTQGQKY GLSGTLYNSG INVLEVRLKL LYGTKNELVA
4412 QYPSLYPNFV TPNLTQGQKY GLSGTLYNSS VNVLEVSLKF LYGTKNELVG
5938 QYPSLYPNFV TPNLTKGQKY GLSGTLYNPG INVLEVRLKL LYGTKNELVG

    .........160.........170.........180.........190.........200
c2   TYQVRPNQYL DVKEIYTLPS TETVEKFGIA FEVAQTSDFV QFEVYLPKIE
4412 TYQVQPNQYL DLKEIYTLPS TETVEKFGIA FEVAQTSDFV QFEVYLPKIE
5938 TYQVQPNQYL DVKEIYTLPS TETVEKFGIA FEVAQTSDFV QFEVYLPKIE

    .........210.........220.........230.........240.........250
c2   QGGEVTPFVE DRDEFNGYRK TNTDDGTPPF TGTYEGTAPQ STDYKVYTWA
4412 QGGEVTPFVE DRDEFNGYRK TNTDDGTPPF TGTYEGTAPQ SADYKDYIWA
5938 QGGEITPFVE DRDEFNGYRK TNTDDGTPPF TGTYEGTPPQ STDYKVYTWT

    .........260.........270.........280.........290.........300
c2   GSKTDKELFY LEERGICKQE AVWCYCRPIN QRVLIGIDSD TYDTEAGRTL
4412 GSKTDKELFY LEERGICKQD AVWCYSRPLG QRVLIGIDSD TYDTEAGRTL
5938 GSKTDKELFY LEERGICKQE AVWCYSRPLN QRVLIGIDSD TYDTEAGRTL

    .........310.........320.........330.........340.........350
c2   KFHVLNGNKG IFDLTGNVIY PEQFTEKRQT FDSDTKAWVD NQEPLYVTDA
4412 KFHVLNGSKG IFDLTGNVIY PEQLTDKRQT FDSDTKAWVD NQEPLYVTDA
5938 KFHVLNGNKG IFDLTGSVIY PEQFTDKRQT FDSDTKAWAD NQEPLYVTDA

    .........360.........370.........380.........390.........400
c2   NTAIDCVFGE MSSNIIEGFN YQQADKRYRV DEVFRSAMVN IGYNMGSYWA
4412 NTAIDCTFGE MASNIIEGYY YQQADKRYKV DELLCSAMIN TGYNMGSYWS
5938 NTAIDCTFGE IASNIIEGYY YQQADKRYRV DELLRSAMVN TGYNMGSFWS

    .........410.........420.........430.........440.........450
c2   EWNFDSYAEE MRASYNIENC RVSEKTSYSS MNEWTGSVSF PTGVVLAPYK
4412 SWDFNSYVTE MRATYGIGNC RVSEKTNYSS MNEWTGSVSF PTGIVLAPYK
5938 DWNFDSYADE MRASYNIENC RISEKINYSS MNEWTGSVSF PTGVVLAPYK

    .........460.........470.........480.........490.........500
c2   PKLNETDTKK LKGVSSATSI WAKGNLRTER TVEDWFREYE NSRTRPVPTQ
4412 PKLNETDTKK LKGASSATSI WATGVLRTER STEDWFREYE NSRTRPVPTQ
5938 PKLNETDTKK LKGVSSATSI WATGVLRTER SVEDWFREYE NSITRPVPTQ

    .........510.........520.........530.........540.........550
c2   ILFANYNTKK AWLFQQQSNG TWSKSREFTI PGSATAFARA WGIIPKNGEL
4412 ILFANYNTKK AWLFQQQSNG TWSKSGEFTI PESNTALASE WGIIPQNGEL
5938 ILFANYNTKK AWLFQQQSNG TWSKSGEFTI PGSATAFARA WGIIPKNGEL

    .........560.........570.........580.........590.........600
c2   KGNVIMTDKN YADFPANARP ITLGVEELFP VIKYNEVKFN PQMYATAYNT
4412 KGNVIMTDKN YADFPTNSRP ITLGVEKLFP VIKYNEVKFN PQMYATAYNT
5938 KGNVIMTDKS YVDFPANVRP ITLGVEELFP VIKYNKVKFN PQMYATAYNT

    .........610.........620.........630........
c2   KLFWWGQKAN VSNLTYGECG VRSVDFMTGL CTIERVYK
4412 KLFWWGQKAN VSNLTYGECG VRSVNFMTGL CTIERVYK
5938 KLFWWGQKAN VSNLTYGECG VRSVDFMTGL CTIERVYK
```

gp34 secondary structure prediction



```
     .........10.........20.........30.........40.........50
bIL67 LNKAYTWNEK TGLDFCRELP QWNLLTRSNL RFLKGDTSED PDKFDVGYYH
4410  LNKAYTWNEK TGLDFCRELP QWNLLTRSNL RFLKGDTSED PDKFDIGYYH
6202  LNKAYTWNEK TGLDFCKELP QWNLLTRSNL RFLTGDTSEE PDKFDVGYYF
6165  LNKAYTWNEK TGLDFCKELP QWNLLTRSNL RFLTGDTSEN PDNFDPGYYY
6162  LNKAYTWNEK TGLDFCKELP QWNLLTRSNL RFLTGDTSED PDKFDPSHYF

     .........60.........70.........80.........90.........100
bIL67 KLNDLSEVDR TSQFPSDWHR PYSLGIRIYN PKNASGTWGW NYWTHWEKLP
4410  KLNDLSEVYR TSQFPSDWHR PYSLGIRIYN PKNATGMWGW KYWTHWEKLP
6202  KLNALSEVDR TSQFPSDWHR PYSLGIRLYN PKNASGTWGW TYWTHWEKLP
6165  KLNALSEVDR TSQFPSDWHR PYSLGIRLYN PKNASGTWGW QYWTHWEKLP
6162  KLNALSEVDR TSQFPNDWHR PYSLGIRLYN PKNASGKWGW TYLTHWEKLP

     .........110.........120.........130.........140.........150
bIL67 VKPNLTQGKK MGVSIRLTNF GRKPLDFHLK LLYGNSVASV GTYTVEPWQY
4410  VKPNLTQGKK MGVSMRLTNF GRKPLDFNLK LFYGNSVVGV GTYTVEPWEY
6202  VKPNLTQGKK MGVSMRLANF GRKPLDFNLK LLYGNTSASV GTYKVEPWQY
6165  VKPNLTQGQK MGVSMRLANF GRKPLDFTLK LFYGNSAVGV GTYTVEPWQY
6162  VKPNLTQGKK MGVSMRLTNF GRKPLDFHLK LLYGNSVASV GTYTVEPWEY

     .........160.........170.........180.........190.........200
bIL67 FFVSELVTLQ NTETAEKLGL RVELDSTGQE EQIGLFFPKI EMDKITPYVT
4410  VFVSEIVTLQ KTETAENLGL SVELDSTGQE EQIGLFFPKI EMDKVTPYVT
6202  VFVSELVTLQ TTETAKNLGL TVELDSTGQE EQIGLFFPKI EMDKVTPYVT
6165  TFVSELVTLH NTETTGKLGL TVELDSTGEE EQIGLFFPKI EMDKVTPYVT
6162  VFVSELVTLQ KTETVEKLGL TVELDSTGQE EQIGLFFPKI EMDKVTPYVT

     .........210.........220.........230.........240.........250
bIL67 TEDEYNYFKS QDMADSRPVY TGYSDSDSND FKDYVWGGQL NDENYELFGG
4410  TEEEYNYFKS QDMADSKPVY TGYSDSASND FKDYVWGGQL NDENYELFGG
6202  TEEEYNYFKS QDMADSRPVY TGYSDSDSND FRDYVWGGQL NDESYELFGG
6165  TEEEYNYFKS QDMADSRPVY TGYSETDSND FKDYVWGGQL NDESYELFGG
6162  TEEEYNYLNS KDTADSRPVY TGYSETDSDD FRSYVWGGQL NDENYELFGG

     .........260.........270.........280.........290.........300
bIL67 DTKQNAVWCY CRPLNQRVLI GIDSDMYTNA SGRTVNFHVL NGSKSVFDMT
4410  DTKQNAVWCY CRPLNQRVLI GIDSDTYTNA SGRTVNFHVL NGSKSVFDMT
6202  DTKQNAVWCY CRPLNQRVLI GIDSDTYTNA SGRTVNFHVL NGSKSVFDMT
6165  DTKQNAVWCY CRPLNQRVLI GIDSDTYTNA SGRTVNFHVL NGSKSVFDMT
6162  DIKQNALWCF CRPLNQRVLI GIDSDIYTNA SGRTVKFHVL NGSKSVFDMT

     .........310.........320.........330.........340.........350
bIL67 GNTLYPEQFQ DARQAFDSVG NDWATIQEPL YVVDQNTAID PVAGEMANVV
4410  GNTLYPEQFQ DERQAFDGVG NDWATIQEPL YVVDQNTAID PVAGEMANVV
6202  GNTLYPEQFQ DARQAFDNIG NDWATIQEPL YVVDQNTAID PVAGEMANVV
6165  GNTLYPEQFQ DDRQAFDGVG NDWATIQEPL YVVDQNTAID PVAGEMANVC
6162  GNTLYPEQFQ DSRQAFDGVG NDWATIQEPL YIVDQNTAID PVAGEMANVV

     .........360.........370.........380.........390.........400
bIL67 IEGYHYQQAS QGYKVDEVVR SAILNVGYSL GSYYVNEDSG KEVGVMRSRV
4410  IEGYHYQQAS QGYRVDEIPR SAILNVGYSL GSYYVNEDSG KEVEVMRNRV
6202  IEGYHYPQAS GGYRVDEIPR SAILNVGYSL GSFYVNEDSS KEVEVMRSRV
6165  IEGYHYKQAS QGYSVDEIPR SAILNVGYSL GSYYVNEDSG KEVDVMRSRV
6162  IEGYHYQQAS QGYRVDEIPR SAVLNVGYSL GSYYVDQNSE KEVAVMRDRV

     .........410.........420.........430.........440.........450
bIL67 GITPPQVFGE PSYSGMNDWM TTYGLPNGLI MRPWRVRMVD TETNLKAIKG
4410  GITPPQVFGE PSYSSMNDWM TTNGLPNGLI MRPWRVRMVD TETNLKAIKG
6202  GVTPPKVFGE PSYSGMNDWM TTNGLPNGLI MRPWRVRMVD TETNLTLIKG
6165  GITPPQVFGE PSYSSMNDWM TTYGLPNGLI MRPWRVRMLD TETNLTKIKG
6162  GITPPQVFGE PSYSGMNDWM TTYGLPNGLI MRPWRVRMVD TETNLTRIKG

     .........460.........470.........480.........490.........500
bIL67 ISIGWNVSLF QKYLATDHMS EDWFRGYDNK RTKAIPDRVL FINDKQKRAW
4410  ISIGWNVSLF QKYLATDHMS EDWFRGYDNK RTKAIPDRVL FINDKQRRAW
6202  ISIGWNVSLF QKYLATDHMS EDWFRGYDNK RTKAIPDRVL FINDKQKRAW
6165  ISIGWNVSLF QKFLATDHVT EDWFRGYDNK RTKAIPDRVL FINDKARRAW
6162  ISISWNVATF QKFLATDHVT EDWFRGYDNK RTKAIPDRVL FINDKARRAW

     .........510.........520.........530.........540.........550
bIL67 LYKYNPTKRA WERSVEYIIP ASNTALLKYW TIIPKDGAMN GHIVFTDKTN
4410  LYKYNPTKSA WERSVEYTIP ASDAALLKAW TIVPKDGAMN GHIVFTDKTN
6202  LYKYNPTKSA WERSVEYTIP ASGTALLKAW TIVPNDGAMN GHIVFTDKTN
6165  LYKFNPTKSA WERSEEYTIP ASNTAQLKYW SIVPNDGAMN GHIVFTDKTN
6162  LYKYNTTKSA WERSVEYVVP AGGTALLRYW TIVPKDGNMN GHIVFTDKTN

     .........560.........570.........580.........590.........600
bIL67 ASMLQNIRPN WLDYDESTPK VQYDEVKYNP QMFTNLYNTR YQWWGIKDEN
4410  AEMLQNIRPN WLDYDEFTPK VQYDEVKYNP QMFTNLYNTR YQWWGIKDEN
6202  AEMLQNIRPN WLDYDEFTPK AQYDEVKYNP QMFTNLYNTR YQWWGIKDEN
6165  AEMLQNIRPN WLDYDEFTPK VQYDEVKYNP QMFTNLYNTR YQWWGIKDEN
6162  AEMLQNIRPN WLDYDEFTPK VQYDEVKYNP QMFTNLYNTR YQWWGIKDEN

     .........610.........620........
bIL67 PQNQSYGPCV PYEMDFMTGL CKLERIYE
4410  PQNQSYGPCV PYEMDFMTGL CKLERIYE
6202  PQNQSYGPCV PYEMDFMTGL CKLERIYE
6165  PQNQSYGPCV PYEMDFMTGL CKLERIYE
6162  PQNLSYGPCV PYEMDFMTGL CKLERIYQ
```

gpl15 – gp35 amino acid alignment / % identity



Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

```
           .........10 .........20 .........30 .........40 .........50
c2         ---MISWLNF  EELLIHNPIE  LINPSKDTIR  VAMSQKQYIE  FFSNKYTYNG
4412       ---MISWLNF  EELLIHNPIE  LINFSKSNIQ  VALSKKQYID  FFSNKAVYMG
5938       ---MFSWLNF  EELLIHNPIE  LINPSKDTIS  VAMNKKQYIE  FFSNKYTYNG
6165       ---MFSWLNF  EELLIHNPIE  LINPSKDTIS  VAMNKKQYIE  FFSNKYTYNG
4410       ---MFSWLNF  EELLIHNPID  LINPSKDTIS  VAMNKKQYIE  FFSNKYTYNG
bIL67      MNKMFSWLNF  EELLIHNPIE  LINPSKDTIS  VAMNKKQYIE  FFSNKYTYNG
6202       ---MFSWLNF  EELLIHNPIE  LINPSKDTMS  VTMNKKQYIE  FFSNKYTYNG
6162       ---MFSWLNF  EELLIHNPIE  LINPSKDTIS  VAMSKKQYIE  FFSNKYTYNG
Consistency000*7*****  *********9  ***7**8887  *8978****9  *****88*8*

           .........60 .........70 .........80 .........90 .........100
c2         LYYDEEMDFC  LFYYADPLQS  YKEGDVYAQG  YIDVEMKIYR  VKWLCNVSIS
4412       LYYDEEMDFC  VMFYADPLQS  SKNGEMYAEG  YIDVDMKIYR  VKVLSNVSIS
5938       LYYDEEMDFC  LFYYADPLQS  VKDGDVYAQG  YIDVEMKIYR  VKWLCNVSIS
6165       LYYDEEMDFC  LFYYADPLQS  VKDGDVYAQG  YIDVEMKIYR  VKWLCNVSIS
4410       LYYDEEMDFC  LFYYADPLQS  VKEGDVYAQG  YIDVEMKIYR  VKWLCNVSIS
bIL67      LYYDEEMDFC  LFYYADPLQS  VKDGDVYAQG  YIDVEMKIYR  VKWLCNVSIS
6202       VYYDEEMDFC  LFYYADPLQS  YKEGDVYAQG  YIDVEMKIYR  VKWLCNVSIS
6162       LYYDEEMDFC  LFYYADPLQS  VKDGDIYAQG  YIDVEMKIYR  VKWLCNVSIS
Consistency9*********  989*******  5*7*98**9*  ****9*****  **7*8*****

           .........110.........120.........130.........140.........150
c2         RPSGLLQTTD  GTQGVPQEGG  KYTHTAWSYS  ADGTDRFSTV  YPNLNLLNGT
4412       RF---NSNFN  LLEGSKDFSS  FNH----NSS  NNSVS----T  ITKIKLSGID
5938       RF-GSNSNFN  LLVGTATFDG  FKP----NSS  NNSVS----T  ITKIKLSGID
6165       RF-GSNSNFN  LLVGTATFDG  FKP----NSS  NNSVS----T  ITKIKLSGID
4410       RF-GSNSNFN  LLRGTAKFDG  FNP----NSS  NNSVS----T  ITKIKLSGID
bIL67      RF---NYNFN  LLEGTKDFSG  NWDR---TGG  WTDDG----T  YKGLVVKKKT
6202       RP----DSFD  GTQGKLQPDG  KYKHTAWANI  KFLCN----D  WNEFTLSKTS
6162       R---------  ----------  ----------  ----------  ----------
Consistency*302134555  3537433346  3220000444  3333400005  3345364334

           .........160.........170.........180.........190.........200
c2         KDFNGDWING  GVWGNDGKYK  GLTVKSYQKA  WDGMFKKYIV  PQDGLYTWSS
4412       ---------N  TVMDVRTSGN  AFAVGLYL--  ----YNAYSI  TAGQTVTISF
5938       ---------N  TVMDVRTSGN  AFAVGLYL--  ----HNAYSI  TAGQTITISF
6165       ---------N  TVMDVRTSGN  AFAVGLYL--  ----HNAYSI  TAGQTITISF
4410       ---------N  TVMDVRTSGN  AFAVGLYL--  ----HNAYSI  TAGQTITISF
bIL67      ---------T  QWNGINKKFT  APKDGVYT--  ----FSAYIK  ASGATTSITR
6202       ---------L  VRYGVDNRWT  YKEFQAGT--  ----YTASFN  TFGSDPASEM
6162       ----------  ----------  ----------  ----------  ----------
Consistency0000000003  3334543434  4344445400  0000446533  5464435453

           .........210.........220.........230.........240.........250
c2         FVKSESDTSD  IFRVLFINNK  EFPIVGLGHK  FDWLRDSVTV  PLKKGDEVIF
4412       MARGTNDTKV  LVGFEGIANG  IKEFRLSPNW  ELYTYTFTST  TSGTHNFLVY
5938       MARGTNGTKV  LVGFEGITNG  LKEFRLTPNW  ELYTHTFTAT  TSGTHNFLMY
6165       MARGTNGTKV  LVGFEGITNG  LKEFRLTPNW  ELYTHTFTAT  TSGTHNFLMY
4410       MARGTNGTKV  LVGFEGITNG  LKEFRLTPNW  ELYTHTFTAT  TSGTHNFLMY
bIL67      YSNINGVGKT  EWLKD-----  --SLGNNFDW  LRDTVTISLK  AGQTALFRYE
6202       YKRAELVSEF  SFLCN-----  --DW-DTFML  SKTSLVRYGI  DGHWVYKVLQ
6162       ----------  ----------  ----------  ----------  ----------
Consistency4453432553  4323423232  2244224244  3335253434  3434334444

           .........210.........220.........230.........240.........250
c2         FVKSESDTSD  IFRVLFINNK  EFPIVGLGHK  FDWLRDSVTV  PLKKGDEVIF
4412       MARGTNDTKV  LVGFEGIANG  IKEFRLSPNW  ELYTYTFTST  TSGTHNFLVY
5938       MARGTNGTKV  LVGFEGITNG  LKEFRLTPNW  ELYTHTFTAT  TSGTHNFLMY
6165       MARGTNGTKV  LVGFEGITNG  LKEFRLTPNW  ELYTHTFTAT  TSGTHNFLMY
4410       MARGTNGTKV  LVGFEGITNG  LKEFRLTPNW  ELYTHTFTAT  TSGTHNFLMY
bIL67      YSNINGVGKT  EWLKD-----  --SLGNNFDW  LRDTVTISLK  AGQTALFRYE
6202       YKRAELVSEF  SFLCN-----  --DW-DTFML  SKTSLVRYGI  DGHWVYKVLQ
6162       ----------  ----------  ----------  ----------  ----------
Consistency4453432553  4323423232  2244224244  3335253434  3434334444

           .........260.........270.........280.........290.........300
c2         NYGNLKNNGG  KLS----VAG  YKLESGSIAT  PWMPSASEVT  TSDGPSYIGQ
4412       GWDMVAGQWF  QVY------N  PKAEEGSIAT  PYMQAESETT  STDLPKWNVT
5938       GWDMDAGQWF  QVY------N  PKAEEGPTAT  PYMQSESETL  SADLPKWNVT
6165       GWDMDAGQWF  QVY------N  PKAEEGPTAT  PYMQSESETL  SADLPKWNVT
4410       GWDMDAGQWF  QVY------N  PKAELGYNAT  SYMQSESESR  H----KWNVS
bIL67      IASNNADAIL  WTC------G  HKWEHGSVAT  QYMPSESEAT  SVDFPKWNVS
6202       AGTYTANFYL  FGSDPAPGVW  KKVELVSGFS  TFYPTTKSDR  PSHIGQYTDT
6162       ----------  ----------  ----------  ----------  ---PNKWNVT
Consistency3343264333  3430000003  3737253256  3564546633  4342477656

           .........310.........320.........330.........340.........350
c2         YTDYTLEDST  NPSSYTWREI  REDKWNV-TK  IGMLVSPQDK  QITMVQAGAL
4412       KT----GMRV  SPQAKQITMV  QAG-------  ----------  --------TL
5938       KT----GMLV  SPQAKQITMV  QAG-------  ----------  --------AL
6165       KT----EMVV  NAKNKTITNV  LNG-------  ----------  --------AL
4410       KT----EMVV  NAKNKTITTV  LNG-------  ----------  --------AL
bIL67      KT----EMVV  NPKTKAITTV  LNGALAKCTK  DKNITGWQ-N  PLLTEIINID
6202       ML----EDST  NPSSYTWVEA  QEEKWNV-SK  TEMIVNSKAK  TITTVLNGAL
6162       KT----EMIV  STKTKTITTV  LNG-------  ----------  --------AL
Consistency6800006647  7565666748  5560000001  0000000000  0000000077

           .........360.........370.........380.........390.........400
c2         MKCGINNDIT  GWTDGTTQLN  YSGQ------  -----DFIID  GYGMRGLHNG
4412       MKCEIDNNIS  GWTDGTTQLN  YSGQ------  -----DFIID  GYGMRGLHNG
5938       TRCEINNNIT  GWTDGTTQLE  YSGQ------  -----DFIID  GYGMRGLHNG
6165       AKCTKDKNIT  GWKNLQPNAN  YNYRQPQYSL  DIGTDDFIIS  GFGLRGLKNG
4410       AKCTKDKDVT  GWRNSQPNAS  YNYRQPQYSL  DIGADDFIIS  GFGLRGLKNG
bIL67      NRSIL-----  ----------  ----------  ----------  ----------
6202       SKCTKDKDIT  GWRNLPPNEN  YNYRQPQYSL  DIGADDFIIS  GYGLRGLKNG
6162       AKCTKDKNIT  GWLNQPTNAN  YNYRQPQYSL  DIESGDFIIS  GFGLRGLKNG
Consistency4884455576  7735234535  7535222222  2211177775  7676777477
```

gpl15 / gp35 N-terminal

	c2	D4412	M5938	bIL67	D4410	M6162	M6165	M6202
c2		79	94	94	94	94	94	93
D4412	79		78	78	77	79	78	75
M5938	94	78		100	98	98	100	95
bIL67	94	78	100		98	98	100	95
D4410	94	77	98	98		96	98	95
M6162	94	79	98	98	96		98	93
M6165	94	78	100	100	98	98		95
M6202	93	75	95	95	95	93	95	

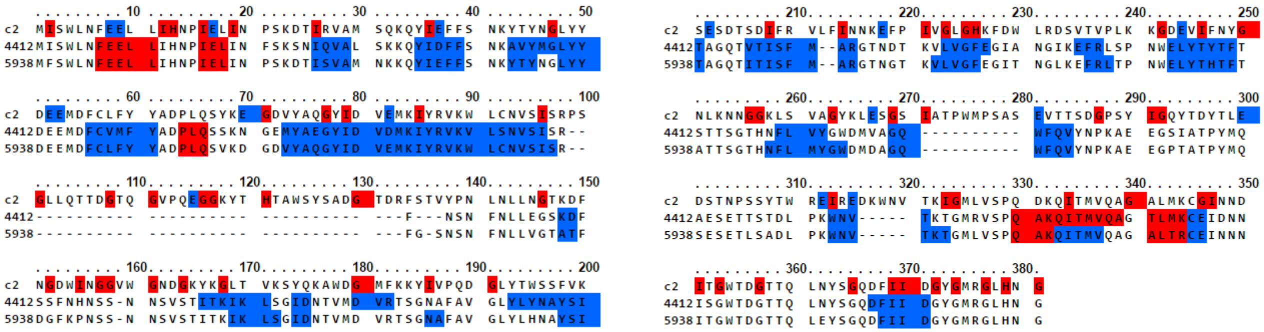
gpl15 / gp35 Middle

	c2	D4412	M5938	bIL67	D4410	M6162	M6165	M6202
c2		22	25	0	24	-	25	0
D4412	22		85	28	84	-	85	0
M5938	20	85			94	-	100	0
bIL67	45	28	25		24	-	25	0
D4410	19	84	94	24		-	94	0
M6162	-	-	-	-	-		-	-
M6165	-	85	100	25	94	-		0
M6202	8	0	0	0	0	-	0	

gpl15 / gp35 C-terminal

	c2	D4412	M5938	bIL67	D4410	M6162	M6165	M6202
c2		88	89	45	42	47	45	46
D4412	88	89	89	46	41	49	46	46
M5938	89	89		45	41	47	45	45
bIL67	45	46	45		86	82	86	84
D4410	42	41	41	86		80	90	87
M6162	47	49	47	82	80		84	82
M6165	45	46	45	86	90	84		84
M6202	46	46	45	84	87	82	84	

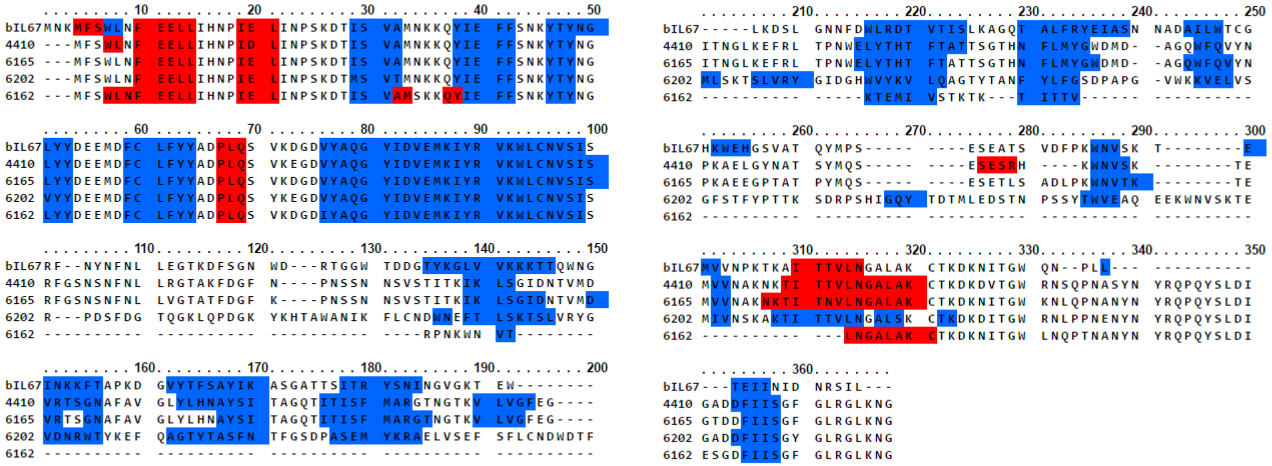
gpl15 secondary structure prediction



```
    .........10 .........20 .........30 .........40 .........50
c2  MISWLNFEEL LIHNPIELIN PSKDTIRVAM SQKQYIEFFS NKYTYNGLYY
4412MISWLNFEEL LIHNPIELIN FSKSNIQVAL SKKQYIDFFS NKAVYMGLYY
5938MFSWLNFEEL LIHNPIELIN PSKDTISVAM NKKQYIEFFS NKYTYNGLYY

    .........60 .........70 .........80 .........90 .........100
c2  DEEMDFCLFY YADPLQSYKE GDVYAQGYID VEMKIYRVKW LCNVSISRPS
4412DEEMDFCVMF YADPLQSSKN GEMYAEGYID VDMKIYRVKV LSNVSISR--
5938DEEMDFCLFY YADPLQSVKD GDVYAQGYID VEMKIYRVKW LCNVSISR--

    .........110.........120.........130.........140.........150
c2  GLLQTTDGTQ GVPQEGGKYT HTAWSYSADG TDRFSTVYPN LNLLNGTKDF
4412---------- ---------- ---------- ---F---NSN FNLLEGSKDF
5938---------- ---------- ---------- ---FG-SNSN FNLLVGTATF

    .........160.........170.........180.........190.........200
c2  NGDWINGGVW GNDGKYKGLT VKSYQKAWDG MFKKYIVPQD GLYTWSSFVK
4412SSFNHNSS-N NSVSTITKIK LSGIDNTVMD VRTSGNAFAV GLYLYNAYSI
5938DGFKPNSS-N NSVSTITKIK LSGIDNTVMD VRTSGNAFAV GLYLHNAYSI
```

```
    .........210.........220.........230.........240.........250
c2  SESDTSDIFR VLFINNKEFP IVGLGHKFDW LRDSVTVPLK KGDEVIFNYG
4412TAGQTVTISF M--ARGTNDT KVLVGFEGIA NGIKEFRLSP NWELYTYTFT
5938TAGQTITISF M--ARGTNGT KVLVGFEGIT NGLKEFRLTP NWELYTHTFT

    .........260.........270.........280.........290.........300
c2  NLKNNGGKLS VAGYKLESGS IATPWMPSAS EVTTSDGPSY IGQYTDYTLE
4412STTSGTHNFL VYGWDMVAGQ ---------- WFQVYNPKAE EGSIATPYMQ
5938ATTSGTHNFL MYGWDMDAGQ ---------- WFQVYNPKAE EGPTATPYMQ

    .........310.........320.........330.........340.........350
c2  DSTNPSSYTW REIREDKWNV TKIGMLVSPQ DKQITMVQAG ALMKCGINND
4412AESETTSTDL PKWNV----- TKTGMRVSPQ AKQITMVQAG TLMKCEIDNN
5938SESETLSADL PKWNV----- TKTGMLVSPQ AKQITMVQAG ALTRCEINNN

    .........360.........370.........380.
c2  ITGWTDGTTQ LNYSGQDFII DGYGMRGLHN G
4412ISGWTDGTTQ LNYSGQDFII DGYGMRGLHN G
5938ITGWTDGTTQ LEYSGQDFII DGYGMRGLHN G
```

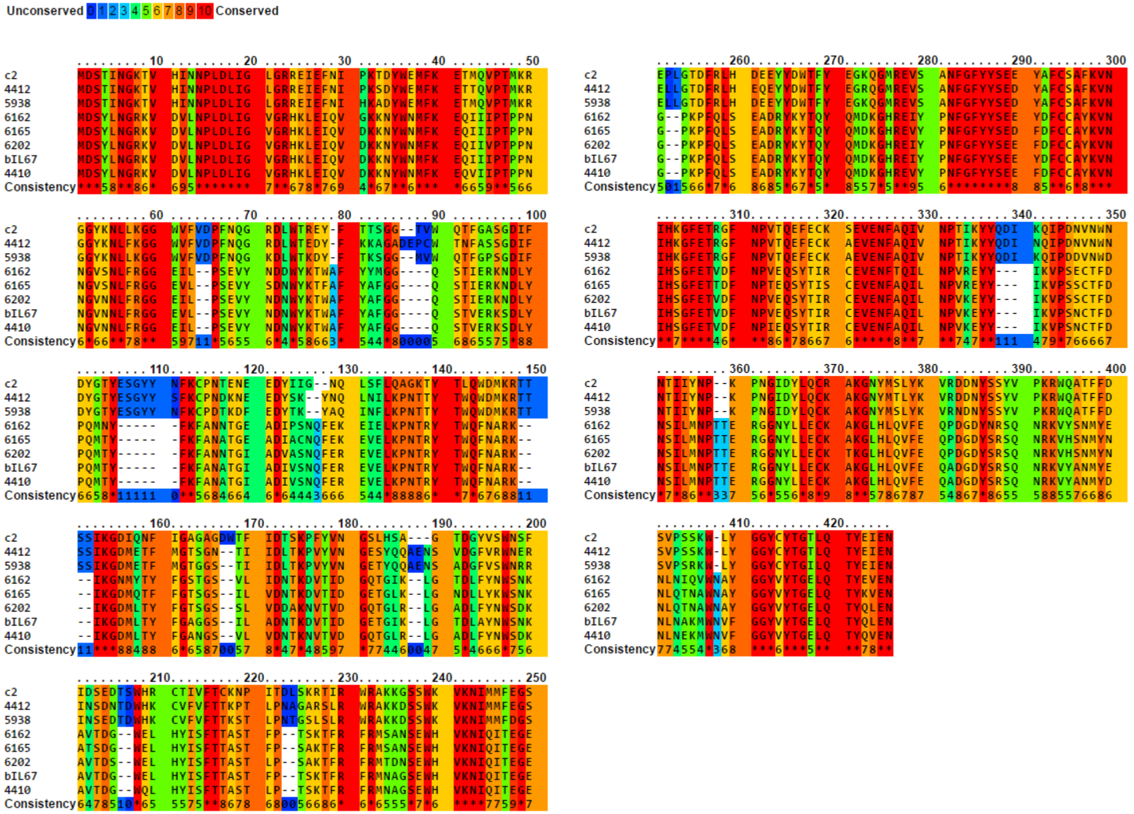
gp35 secondary structure prediction



```
     .........10 .........20 .........30 .........40 .........50
bIL67MNKMFSWLNF EELLIHNPIE LINPSKDTIS VAMNKKQYIE FFSNKYTYNG
4410 ---MFSWLNF EELLIHNPID LINPSKDTIS VAMNKKQYIE FFSNKYTYNG
6165 ---MFSWLNF EELLIHNPIE LINPSKDTIS VAMNKKQYIE FFSNKYTYNG
6202 ---MFSWLNF EELLIHNPIE LINPSKDTMS VTMNKKQYIE FFSNKYTYNG
6162 ---MFSWLNF EELLIHNPIE LINPSKDTIS VAMSKKQYIE FFSNKYTYNG

     .........60 .........70 .........80 .........90 .........100
bIL67LYYDEEMDFC LFYYADPLQS VKDGDVYAQG YIDVEMKIYR VKWLCNVSIS
4410 LYYDEEMDFC LFYYADPLQS VKEGDVYAQG YIDVEMKIYR VKWLCNVSIS
6165 LYYDEEMDFC LFYYADPLQS VKDGDVYAQG YIDVEMKIYR VKWLCNVSIS
6202 VYYDEEMDFC LFYYADPLQS YKEGDVYAQG YIDVEMKIYR VKWLCNVSIS
6162 LYYDEEMDFC LFYYADPLQS VKDGDIYAQG YIDVEMKIYR VKWLCNVSIS

     .........110.........120.........130.........140.........150
bIL67RF--NYNFNL LEGTKDFSGN WD---RTGGW TDDGTYKGLV VKKKTTQWNG
4410 RFGSNSNFNL LRGTAKFDGF N----PNSSN NSVSTITKIK LSGIDNTVMD
6165 RFGSNSNFNL LVGTATFDGF K----PNSSN NSVSTITKIK LSGIDNTVMD
6202 R---PDSFDG TQGKLQPDGK YKHTAWANIK FLCNDWNEFT LSKTSLVRYG
6162 ---------- ---------- ---------- ----RPNKWN VT--------

     .........160.........170.........180.........190.........200
bIL67INKKFTAPKD GVYTFSAYIK ASGATTSITR YSNINGVGKT EW--------
4410 VRTSGNAFAV GLYLHNAYSI TAGQTITISF MARGTNGTKV LVGFEG----
6165 VRTSGNAFAV GLYLHNAYSI TAGQTITISF MARGTNGTKV LVGFEG----
6202 VDNRWTYKEF QAGTYTASFN TFGSDPASEM YKRAELVSEF SFLCNDWDTF
6162 ---------- ---------- ---------- ---------- ----------
```

```
     .........210.........220.........230.........240.........250
bIL67-----LKDSL GNNFDWLRDT VTISLKAGQT ALFRYEIASN NADAILWTCG
4410 ITNGLKEFRL TPNWELYTHT FTATTSGTHN FLMYGWDMD- -AGQWFQVYN
6165 ITNGLKEFRL TPNWELYTHT FTATTSGTHN FLMYGWDMD- -AGQWFQVYN
6202 MLSKTSLVRY GIDGHWVYKV LQAGTYTANF YLFGSDPAPG -VWKKVELVS
6162 ---------- -----KTEMI VSTKTK---T ITTV------ ----------

     .........260.........270.........280.........290.........300
bIL67HKWEHGSVAT QYMPS----- ----ESEATS VDFPKWNVSK T--------E
4410 PKAELGYNAT SYMQS----- ----ESESRH ----KWNVSK --------TE
6165 PKAEEGPTAT PYMQS----- ----ESETLS ADLPKWNVTK --------TE
6202 GFSTFYPTTK SDRPSHIGQY TDTMLEDSTN PSSYTWVEAQ EEKWNVSKTE
6162 ---------- ---------- ---------- ---------- ----------

     .........310.........320.........330.........340.........350
bIL67MVVNPKTKAI TTVLNGALAK CTKDKNITGW QN--PLL--- ----------
4410 MVVNAKNKTI TTVLNGALAK CTKDKDVTGW RNSQPNASYN YRQPQYSLDI
6165 MVVNAKNKTI TNVLNGALAK CTKDKNITGW KNLQPNANYN YRQPQYSLDI
6202 MIVNSKAKTI TTVLNGALSK CTKDKDITGW RNLPPNENYN YRQPQYSLDI
6162 ---------- ---LNGALAK CTKDKNITGW LNQPTNANYN YRQPQYSLDI

     .........360........
bIL67---TEIINID NRSIL---
4410 GADDFIISGF GLRGLKNG
6165 GTDDFIISGF GLRGLKNG
6202 GADDFIISGY GLRGLKNG
6162 ESGDFIISGF GLRGLKNG
```

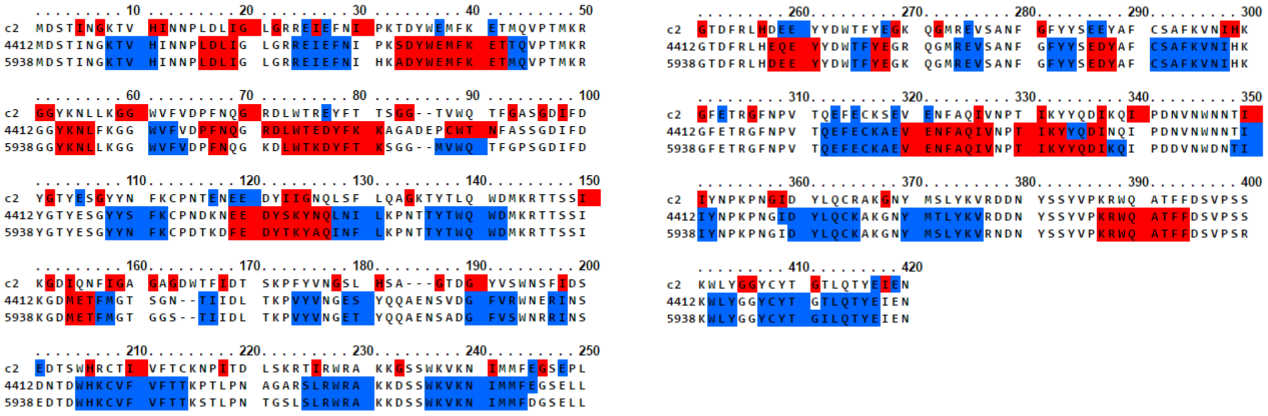
gpl16 – gp36 amino acid alignment / % identity



```
            .........10 .........20 .........30 .........40 .........50
c2          MDSTINGKTV  HINNPLDLIG  LGRREIEFNI  PKTDYWEMFK  ETMQVPTMKR
4412        MDSTINGKTV  HINNPLDLIG  LGRREIEFNI  PKSDYWEMFK  ETTQVPTMKR
5938        MDSTINGKTV  HINNPLDLIG  LGRREIEFNI  HKADYWEMFK  ETMQVPTMKR
6162        MDSYLNGRKV  DVLNPLDLIG  VGRHKLEIQV  GKKNYWNMFK  EQIIIPTPPN
6165        MDSYLNGRKV  DVLNPLDLIG  VGRHKLEIQV  DKKNYWNMFK  EQIIIPTPPN
6202        MDSYLNGRKV  DVLNPLDLIG  VGRHKLEIQV  DKKNYWNMFK  EQIIVPTPPN
bIL67       MDSYLNGRKV  DVLNPLDLIG  VGRHKLEIQV  DKKNYWNMFK  EQIIIPTPPN
4410        MDSYLNGRKV  DVLNPLDLIG  VGRHKLEIQV  DKKNYWNMFK  EQVIIPTPPN
Consistency ***58**86*  695*******  7**678*769  4*67**6***  *6659**566

            .........60 .........70 .........80 .........90 .........100
c2          GGYKNLLKGG  WVFVDPFNQG  RDLWTREY-F  TTSGG--TVW  QTFGASGDIF
4412        GGYKNLFKGG  WVFVDPFNQG  RDLWTEDY-F  KKAGADEPCW  TNFASSGDIF
5938        GGYKNLLKGG  WVFVDPFNQG  KDLWTKDY-F  TKSGG--MVW  QTFGPSGDIF
6162        NGVSNLFRGG  EIL--PSEVY  NDDWYKTWAF  YYMGG----Q  STIERKNDLY
6165        NGVSNLFRGG  EVL--PSEVY  SDNWYKTFAF  YAFGG----Q  STIERKNDLY
6202        NGVNNLFRGG  EIL--PSEVY  NDNWYKTWAF  YAFGG----Q  STIERKNDLY
bIL67       NGVNNLFRGG  EVL--PSEVY  NDNWYKTWAF  YAFGG----Q  STVERKSDLY
4410        NGVNNLFRGG  EIL--PSEVY  NDNWYKTWAF  YAFGG----Q  STVERKSDLY
Consistency 6*66**78**  59711*5655  6*4*58663*  544*800005  6865575*88

            .........110.........120.........130.........140.........150
c2          DYGTYESGYY  NFKCPNTENE  EDYIIG--NQ  LSFLQAGKTY  TLQWDMKRTT
4412        DYGTYESGYY  SFKCPNDKNE  EDYSK--YNQ  LNILKPNTTY  TWQWDMKRTT
5938        DYGTYESGYY  NFKCPDTKDF  EDYTK--YAQ  INFLKPNTTY  TWQWDMKRTT
6162        PQMNY-----  -FKFANNTGE  ADIPSNQFEK  EIELKPNTRY  TWQFNARK--
6165        PQMTY-----  -FKFANATGE  ADIACNQFEK  EVELKPNTRY  TWQFNARK--
6202        PQMTY-----  -FKFANNTGI  ADVASNQFER  EVELKPNTRY  TWQFNARK--
bIL67       PQMTY-----  -FKFANATGI  ADIVSNQFER  EVELKPNTRY  TWQFNARK--
4410        PQMTY-----  -FKFANATGI  ADIVSNQFER  EVELKPNTRY  TWQFNARK--
Consistency 6658*11111  0**5684664  6*64443666  544*88886*  *7*676881 1

            .........160.........170.........180.........190.........200
c2          SSIKGDIQNF  IGAGAGDWTF  IDTSKPFYVN  GSLHSA---G  TDGYVSWNSF
4412        SSIKGDMETF  MGTSGN--TI  IDLTKPVYVN  GESYQQAENS  VDGFVRWNER
5938        SSIKGDMETF  MGTGGS--TI  IDLTKPVYVN  GETYQQAENS  ADGFVSWNRR
6162        --IKGNMYTY  FGSTGS--VL  IDNTKDVTID  GQTGIK--LG  TDLFYNWSNK
6165        --IKGDMQTF  FGTSGS--IL  VDNTKDVTID  GETGLK--LG  NDLLYKWSNK
6202        --IKGDMLTY  FGTSGS--SL  VDDAKNVTVD  GQTGLR--LG  ADLFYNWSDK
bIL67       --IKGDMLTY  FGAGGS--IL  ADNTKDVTID  GETGIK--LG  TDLAYNWSNK
4410        --IKGDMLTY  FGANGS--VL  VDNTKNVTVD  GQTGLR--LG  ADLFYNWSDK
Consistency 11***88488  6*65870057  8*47*48597  *774460047  5*4666*756

            .........210.........220.........230.........240.........250
c2          IDSEDTSWHR  CTIVFTCKNP  ITDLSKRTIR  WRAKKGSSWK  VKNIMMFEGS
4412        INSDNTDWHK  CVFVFTTKPT  LPNAGARSLR  WRAKKDSSWK  VKNIMMFEGS
5938        INSEDTDWHK  CVFVFTTKST  LPNTGSLSLR  WRAKKDSSWK  VKNIMMFDGS
6162        AVTDG--WEL  HYISFTTAST  FP--TSKTFR  FRMSANSEWH  VKNIQITEGE
6165        ATSDG--WEL  HYISFTTAST  FP--SAKTFR  FRMSANSEWH  VKNIQITEGE
6202        AVTDS--WEL  HYISFTTAST  LP--SAKTFR  FRMTDNSEWH  VKNIQITEGE
bIL67       AVTDG--WEL  HYISFTTAST  FP--TSKTFR  FRMNAGSEWH  VKNIQITEGE
4410        AVTDG--WQL  HYISFTTAST  LP--TSKTFR  FRMNAGSEWH  VKNIQITEGE
Consistency 6478510*65  5575**8678  6800566686*  6*6555*7*6  ****7759*7

            .........260.........270.........280.........290.........300
c2          EPLGTDFRLH  DEEYYDWTFY  EGKQGMREVS  ANFGFYYSEE  YAFCSAFKVN
4412        ELLGTDFRLH  EQEYYDWTFY  EGRQGMREVS  ANFGFYYSED  YAFCSAFKVN
5938        ELLGTDFRLH  DEEYYDWTFY  EGKQGMREVS  ANFGFYYSED  YAFCSAFKVN
6162        G--PKPFQLS  EADRYKYTQY  QMDKGHREIY  PNFGFYYSEE  YDFCCAYKVN
6165        G--PKPFQLS  EADRYKYTQY  QMDKGHREVY  PNFGFYYSED  FDFCCAYKVN
6202        G--PKPFQLS  EADRYKYTQY  QMDKGHREIY  PNFGFYYSEE  FDFCCAYKVN
bIL67       G--PKPFQLS  EADRYKYTQY  QMDKGHREIY  PNFGFYYSEE  YDFCCAYKVN
4410        G--PKPFQLS  EADRYKYTQY  QMDKGHREVY  PNFGFYYSEE  FDFCCAYKVN
Consistency 501566*7*6  8685*67*5*  8557*5**95  6********8  85**6*8***

            .........310.........320.........330.........340.........350
c2          IHKGFETRGF  NPVTQEFECK  SEVENFAQIV  NPTIKYYQDI  KQIPDNVNWN
4412        IHKGFETRGF  NPVTQEFECK  AEVENFAQIV  NPTIKYYQDI  NQIPDNVNWN
5938        IHKGFETRGF  NPVTQEFECK  AEVENFAQIV  NPTIKYYQDI  KQIPDDVNWD
6162        IHSGFETVDF  NPVEQSYTIR  CEVENFTQIL  NPVREYY---  IKVPSECTFD
6165        IHSGFETTDF  NPTEQSYTIS  CEVENFAQIL  NPVREYY---  IKVPSSCTFD
6202        IHSGFETVDF  NPVEQSYTIR  CEVENFAQIL  NPVKEYY---  IKVPSSCTFD
bIL67       IHSGFETVDF  NPVEQSYTIR  CEVENFAQIL  NPVKEYY---  IKVPSNCTFD
4410        IHSGFETVDF  NPIEQSYTIR  CEVENFAQIL  NPVKEYY---  IKVPSNCTFD
Consistency **7****46*  **86*78667  6*****8**7  **747**111  479*766667

            .........360.........370.........380.........390.........400
c2          NTIIYNP--K  PNGIDYLQCR  AKGNYMSLYK  VRDDNYSSYV  PKRWQATFFD
4412        NTIIYNP--K  PNGIDYLQCK  AKGNYMTLYK  VRDDNYSSYV  PKRWQATFFD
5938        NTIIYNP--K  PNGIDYLQCK  AKGNYMSLYK  VRNDNYSSYV  PKRWQATFFD
6162        NSILMNPTTE  RGGNYLLECK  AKGLHLQVFE  QPDGDYNRSQ  NRKVYSNMYE
6165        NSILMNPTTE  RGGNYLLECK  AKGLHLQVFE  QPDGDYSRSQ  NRKVHSNMYN
6202        NSILMNPTTE  RGGNYLLECK  TKGLHLQVFE  QPDGDYSRSQ  NRKVHSNMYN
bIL67       NSILMNPTTE  RGGNYLLECK  AKGLHLQVFE  QADGDYSRSQ  NRKVYANMYE
4410        NSILMNPTTE  RGGNYLLECK  AKGLHLQVFE  QADGDYSRSQ  NRKVYANMYD
Consistency *7*86**337  56*556*8*9  8**5786787  54867*8655  5885576686

            .........410.........420.......
c2          SVPSSKW-LY  GGYCYTGTLQ  TYEIEN
4412        SVPSSKW-LY  GGYCYTGTLQ  TYEIEN
5938        SVPSRKW-LY  GGYCYTGILQ  TYEIEN
6162        NLNIQVWNAY  GGYVYTGELQ  TYEVEN
6165        NLQTNAWNAY  GGYVYTGELQ  TYKVEN
6202        NLQTNAWNAY  GGYVYTGELQ  TYQLEN
bIL67       NLNAKMWNVF  GGYVYTGELQ  TYQLEN
4410        NLNEKMWNVF  GGYVYTGELQ  TYQVEN
Consistency 774554*368  ***6***5**  **78**
```

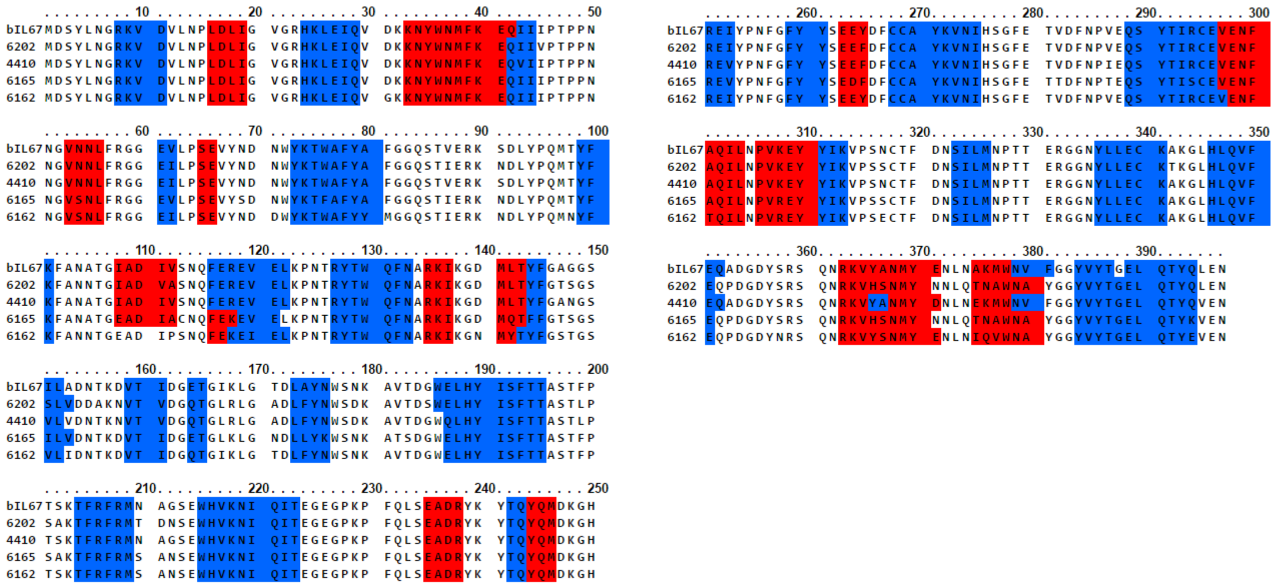
	c2	D4412	M5938	bIL67	D4410	M6162	M6165	M6202
c2		79	81	38	37	37	37	37
D4412	79		88	39	39	38	40	39
M5938	81	88		40	39	39	40	39
bIL67	38	39	40		94	90	88	89
D4410	37	40	40	94		89	88	91
M6162	37	38	39	90	89		89	88
M6165	37	40	40	88	88	89		90
M6202	37	39	39	89	91	88	90	

gpl16 secondary structure prediction



```
    .........10.........20.........30.........40.........50
c2   MDSTINGKTV HINNPLDLIG LGRREIEFNI PKTDYWEMFK ETMQVPTMKR
4412 MDSTINGKTV HINNPLDLIG LGRREIEFNI PKSDYWEMFK ETTQVPTMKR
5938 MDSTINGKTV HINNPLDLIG LGRREIEFNI HKADYWEMFK ETMQVPTMKR

    .........60.........70.........80.........90.........100
c2   GGYKNLLKGG WVFVDPFNQG RDLWTREYFT TSGG--TVWQ TFGASGDIFD
4412 GGYKNLFKGG WVFVDPFNQG RDLWTEDYFK KAGADEPCWT NFASSGDIFD
5938 GGYKNLLKGG WVFVDPFNQG KDLWTKDYFT KSGG--MVWQ TFGPSGDIFD

    .........110.........120.........130.........140.........150
c2   YGTYESGYYN FKCPNTENEE DYIIGNQLSF LQAGKTYTLQ WDMKRTTSSI
4412 YGTYESGYYS FKCPNDKNEE DYSKYNQLNI LKPNTTYTWQ WDMKRTTSSI
5938 YGTYESGYYN FKCPDTKDFE DYTKYAQINF LKPNTTYTWQ WDMKRTTSSI

    .........160.........170.........180.........190.........200
c2   KGDIQNFIGA GAGDWTFIDT SKPFYVNGSL HSA---GTDG YVSWNSFIDS
4412 KGDMETFMGT SGN--TIIDL TKPVYVNGES YQQAENSVDG FVRWNERINS
5938 KGDMETFMGT GGS--TIIDL TKPVYVNGET YQQAENSADG FVSWNRRINS

    .........210.........220.........230.........240.........250
c2   EDTSWHRCTI VFTCKNPITD LSKRTIRWRA KKGSSWKVKN IMMFEGSEPL
4412 DNTDWHKCVF VFTTKPTLPN AGARSLRWRA KKDSSWKVKN IMMFEGSELL
5938 EDTDWHKCVF VFTTKSTLPN TGSLSLRWRA KKDSSWKVKN IMMFDGSELL

    .........260.........270.........280.........290.........300
c2   GTDFRLHDEE YYDWTFYEGK QGMREVSANF GFYYSEEYAF CSAFKVNIHK
4412 GTDFRLHEQE YYDWTFYEGR QGMREVSANF GFYYSEDYAF CSAFKVNIHK
5938 GTDFRLHDEE YYDWTFYEGK QGMREVSANF GFYYSEDYAF CSAFKVNIHK

    .........310.........320.........330.........340.........350
c2   GFETRGFNPV TQEFECKSEV ENFAQIVNPT IKYYQDIKQI PDNVNWNNTI
4412 GFETRGFNPV TQEFECKAEV ENFAQIVNPT IKYYQDINQI PDNVNWNNTI
5938 GFETRGFNPV TQEFECKAEV ENFAQIVNPT IKYYQDIKQI PDDVNWDNTI

    .........360.........370.........380.........390.........400
c2   IYNPKPNGID YLQCRAKGNY MSLYKVRDDN YSSYVPKRWQ ATFFDSVPSS
4412 IYNPKPNGID YLQCKAKGNY MTLYKVRDDN YSSYVPKRWQ ATFFDSVPSS
5938 IYNPKPNGID YLQCKAKGNY MSLYKVRNDN YSSYVPKRWQ ATFFDSVPSR

    .........410.........420
c2   KWLYGGYCYT GTLQTYEIEN
4412 KWLYGGYCYT GTLQTYEIEN
5938 KWLYGGYCYT GILQTYEIEN
```

gp36 secondary structure prediction



```
     .........10.........20.........30.........40.........50
bIL67 MDSYLNGRKV DVLNPLDLIG VGRHKLEIQV DKKNYWNMFK EQIIIPTPPN
6202  MDSYLNGRKV DVLNPLDLIG VGRHKLEIQV DKKNYWNMFK EQIIVPTPPN
4410  MDSYLNGRKV DVLNPLDLIG VGRHKLEIQV DKKNYWNMFK EQVIIPTPPN
6165  MDSYLNGRKV DVLNPLDLIG VGRHKLEIQV DKKNYWNMFK EQIIIPTPPN
6162  MDSYLNGRKV DVLNPLDLIG VGRHKLEIQV GKKNYWNMFK EQIIIPTPPN

     .........60.........70.........80.........90.........100
bIL67 NGVNNLFRGG EVLPSEVYND NWYKTWAFYA FGGQSTVERK SDLYPQMTYF
6202  NGVNNLFRGG EILPSEVYND NWYKTWAFYA FGGQSTIERK NDLYPQMTYF
4410  NGVNNLFRGG EILPSEVYND NWYKTWAFYA FGGQSTVERK SDLYPQMTYF
6165  NGVSNLFRGG EVLPSEVYSD NWYKTFAFYA FGGQSTIERK NDLYPQMTYF
6162  NGVSNLFRGG EILPSEVYND DWYKTWAFYY MGGQSTIERK NDLYPQMNYF

     .........110.........120.........130.........140.........150
bIL67 KFANATGIAD IVSNQFEREV ELKPNTRYTW QFNARKIKGD MLTYFGAGGS
6202  KFANNTGIAD VASNQFEREV ELKPNTRYTW QFNARKIKGD MLTYFGTSGS
4410  KFANATGIAD IVSNQFEREV ELKPNTRYTW QFNARKIKGD MLTYFGANGS
6165  KFANATGEAD IACNQFEKEV ELKPNTRYTW QFNARKIKGD MQTFFGTSGS
6162  KFANNTGEAD IPSNQFEKEI ELKPNTRYTW QFNARKIKGN MYTYFGSTGS

     .........160.........170.........180.........190.........200
bIL67 ILADNTKDVT IDGETGIKLG TDLAYNWSNK AVTDGWELHY ISFTTASTFP
6202  SLVDDAKNVT VDGQTGLRLG ADLFYNWSDK AVTDSWELHY ISFTTASTLP
4410  VLVDNTKNVT VDGQTGLRLG ADLFYNWSDK AVTDGWQLHY ISFTTASTLP
6165  ILVDNTKDVT IDGETGLKLG NDLLYKWSNK ATSDGWELHY ISFTTASTFP
6162  VLIDNTKDVT IDGQTGIKLG TDLFYNWSNK AVTDGWELHY ISFTTASTFP

     .........210.........220.........230.........240.........250
bIL67 TSKTFRFRMN AGSEWHVKNI QITEGEGPKP FQLSEADRYK YTQYQMDKGH
6202  SAKTFRFRMT DNSEWHVKNI QITEGEGPKP FQLSEADRYK YTQYQMDKGH
4410  TSKTFRFRMN AGSEWHVKNI QITEGEGPKP FQLSEADRYK YTQYQMDKGH
6165  SAKTFRFRMS ANSEWHVKNI QITEGEGPKP FQLSEADRYK YTQYQMDKGH
6162  TSKTFRFRMS ANSEWHVKNI QITEGEGPKP FQLSEADRYK YTQYQMDKGH

     .........260.........270.........280.........290.........300
bIL67 REIYPNFGFY YSEEYDFCCA YKVNIHSGFE TVDFNPVEQS YTIRCEVENF
6202  REIYPNFGFY YSEEFDFCCA YKVNIHSGFE TVDFNPVEQS YTIRCEVENF
4410  REVYPNFGFY YSEEFDFCCA YKVNIHSGFE TVDFNPIEQS YTIRCEVENF
6165  REVYPNFGFY YSEDFDFCCA YKVNIHSGFE TTDFNPTEQS YTISCEVENF
6162  REIYPNFGFY YSEEYDFCCA YKVNIHSGFE TVDFNPVEQS YTIRCEVENF

     .........310.........320.........330.........340.........350
bIL67 AQILNPVKEY YIKVPSNCTF DNSILMNPTT ERGGNYLLEC KAKGLHLQVF
6202  AQILNPVKEY YIKVPSSCTF DNSILMNPTT ERGGNYLLEC KTKGLHLQVF
4410  AQILNPVKEY YIKVPSNCTF DNSILMNPTT ERGGNYLLEC KAKGLHLQVF
6165  AQILNPVREY YIKVPSSCTF DNSILMNPTT ERGGNYLLEC KAKGLHLQVF
6162  TQILNPVREY YIKVPSECTF DNSILMNPTT ERGGNYLLEC KAKGLHLQVF

     .........360.........370.........380.........390.......
bIL67 EQADGDYSRS QNRKVYANMY ENLNAKMWNV FGGYVVYTGEL QTYQLEN
6202  EQPDGDYSRS QNRKVHSNMY NNLQTNAWNA YGGYVVYTGEL QTYQLEN
4410  EQADGDYSRS QNRKVYANMY DNLNEKMWNV FGGYVVYTGEL QTYQVEN
6165  EQPDGDYSRS QNRKVHSNMY NNLQTNAWNA YGGYVVYTGEL QTYKVEN
6162  EQPDGDYNRS QNRKVYSNMY ENLNIQVWNA YGGYVVYTGEL QTYEVEN
```