

Sequences

Created Saturday 19 October 2013

>1LQT_A

RPYYIAIVGSGPSAFFAAASLLKAADTTEDLDMAVDMLEMLPTPWGLVRSGVAPDHPKIKSISKQFEKTAE
DPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGE-DLPGSIAAVDFVGWYNAHPHFEQVSP
DLSGARAVVIGNGNVALDVARILLTDPDVLARTDIADHALES LRPRGIQEVVIVGRRGPLQAAFTTLELREL
ADLDGVDVVIDPAELDGITDEDAAVGKVCQKNIKVLRGYADREPRPGHRRMVFRFLTSPIEIKGK---RKV
ERIVLGRNELVSDGSGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQSGTIPNVGGRI-----NG
SPNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEHADQVADWLAARQPKLV TSAH
WQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG---

>466500

R---VAIIGSGPAGFYAAEALLKRTDTAVD---VDMFDRLPFPGLVRGGVAPDHQRIKAVTRVFASTAA
RPTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGE-GIERCTPATVFGWYNGHPDYRHARF
DLSVRRAAVVGNGNVAVDVARILLRTRAELERTDIAAHAEALRESQVREVYLLGRRGPAQAAFS PAELREL
GTLEAADPVVDPADLAGLSDPAPGG-----NLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAG-AV
AGLRLEKNRLERRPDGTVVARGTGETEVLEVGLVLPVAVGYAADRIPGVPFDEKARVIANEDGRVVDPVLRSP
LPGEYVVGWARS GPQGLIGEHRRA SAHVVAH MVAD-----AAGLATRPLPPRDAIDDLRARGVRPVSVFD
WARLDEVEVARGARRGAPRDKLVDVAAMLAI-LGEGE

>466500

MRVAIIGSGPAGFYAAEALLKRTDTAVDVMFDRLPFPGLVRGGVAPDHQRIKAVTRVFASTAARPTFR
FLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGEGIERCTPATVFGWYNGHPDYRHARFDLSV
RRAAVVGNGNVAVDVARILLRTRAELERTDIAAHAEALRESQVREVYLLGRRGPAQAAFS PAELRELGT
LEAADPVVDPADLAGLSDPAPGGNLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAGAVAGLRLEKN
RLERRPDGTVVARGTGETEVLEVGLVLPVAVGYAADRIPGVPFDEKARVIANEDGRVVDPVLRSPLPGEYV
VGWARS GPQGLIGEHRRA SAHVVAH MVADAAGLATRPLPPRDAIDDLRARGVRPVSVFDWARLDEVEVA
RGARRGAPRDKLVDVAAMLAILGEGE

>1LQT_A

MRPYIAIVGSGPSAFFAAASLLKAADTTEDLDMAVDMLEMLPTPWGLVRSGVAPDHPKIKSISKQFEKT
AEDPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGEDLPGSIAAVDFVGWYNAHPHFEQ
VSPDLSGARAVVIGNGNVALDVARILLTDPDVLARTDIADHALES LRPRGIQEVVIVGRRGPLQAAFTTL
ELRELADLDGVDVVIDPAELDGITDEDAAVGKVCQKNIKVLRGYADREPRPGHRRMVFRFLTSPIEIKG
KRKVERIVLGRNELVSDGSGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQSGTIPNVGGRING
SPNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEHADQVADWLAARQPKLVTS
AHWQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG

>1LQT_B

MRPYIAIVGSGPSAFFAAASLLKAADTTEDLDMAVDMLEMLPTPWGLVRSGVAPDHPKIKSISKQFEKT
AEDPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGEDLPGSIAAVDFVGWYNAHPHFEQ
VSPDLSGARAVVIGNGNVALDVARILLTDPDVLARTDIADHALES LRPRGIQEVVIVGRRGPLQAAFTTL
ELRELADLDGVDVVIDPAELDGITDEDAAVGKVCQKNIKVLRGYADREPRPGHRRMVFRFLTSPIEIKG
KRKVERIVLGRNELVSDGSGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQSGTIPNVGGRING
SPNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEHADQVADWLAARQPKLVTS
AHWQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG

Alignment

Created Saturday 19 October 2013

Adjusted Rosetta Alignment, finalAln.fa

>466500

```
R---VAIIGSGPAGFYAAEALLKRTDTAVD----VDMFDRLPFPGLVRGGVAPDHQRIKAVTRVFASTAAR
PTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGE-GIERCTPATVFVGWYNGHPDYRHARFD
LSVRRAAVVGNGNVAVDVARILLRTAEELERTDIAAHALEALRESQVREVYLLGRRGPAQAASFPAELRELG
TLEAADPVVDPADLAGLSDPAPGG-----NLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAG-AVA
GLRLEKNRLERRPDGTVVARGTGETEVLEVGLVLPVAVGYAADRIPGVFPFDEKARVIANEDGRVVDPVLRSP
PGEYVVGWARSQGLIGEHRRAHVAHVMVAD-----AAGLATRPLPPRDAIDDLRARGVRPVSVFDW
ARLDEVEVARGARRGAPRDKLVDVAAMLAILGE
```

>1LQT_A

```
RPYYIAIVGSGPSAFFAAASLLKAADTTEDLMAVDMLEMLPTPWGLVRSGVAPDHPKIKSISKQFEKTAED
PRFRFFGNVVVGEHVQPGELSERVDAVIYAVGAQSDRMLNIPGE-DLPGSIAAVDFVGWYNAHPHFQVSPD
LSGARAVVIGNGNVAVDVARILLTDPDVLARTDIADHALESRLPRGIQEVVIVGRRGPLQAAFTTLELRELA
DLDGVDVVIDPAELDGITDEDAAVGKVCQNIKVLRGYADRE-RPGHRRMVFRFLTSPIEIKGK---RKVE
RIVLGRNELVSDGSGRVAAKDTGEREELPAQLVRSVGYRGVPTPGLPFDDQSGTIPNVGGRI-N---GSPN
--EYVVGWIKRGPTGVIGTNKKDAQDQDVTDLIKNLGNAKE--GAECKSFDHADQVADWLAARQPKLVTSAHW
QVIDAFERAAGEPHGRPRVKLASLAELLRI-GLG-
```

Changes made here:

>466500

VVDPVLRSPLPGEYVVGW

I-N---GSPN--EYVVGW

>1LQT_A

From here:

>466500

VVDPVLRSPLPGEYVVGW

I-----NGSPNEYVVGW

>1LQT_A

Working rosetta alignment

Took 1LQT_A, manually aligned using old profile-profile alignment

>466500

```
R---VAIIGSGPAGFYAAEALLKRTDTAVD----VDMFDRLPFPGLVRGGVAPDHQRIKAVTRVFASTAAR
PTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGE-GIERCTPATVFVGWYNGHPDYRHARFD
LSVRRAAVVGNGNVAVDVARILLRTAEELERTDIAAHALEALRESQVREVYLLGRRGPAQAASFPAELRELG
TLEAADPVVDPADLAGLSDPAPGG-----NLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAG-AVA
GLRLEKNRLERRPDGTVVARGTGETEVLEVGLVLPVAVGYAADRIPGVFPFDEKARVIANEDGRVVDPVLRSP
PGEYVVGWARSQGLIGEHRRAHVAHVMVAD-----AAGLATRPLPPRDAIDDLRARGVRPVSVFDW
ARLDEVEVARGARRGAPRDKLVDVAAMLAILGE
```

>1LQT_A

```
RPYYIAIVGSGPSAFFAAASLLKAADTTEDLMAVDMLEMLPTPWGLVRSGVAPDHPKIKSISKQFEKTAED
PRFRFFGNVVVGEHVQPGELSERVDAVIYAVGAQSDRMLNIPGE-DLPGSIAAVDFVGWYNAHPHFQVSPD
LSGARAVVIGNGNVAVDVARILLTDPDVLARTDIADHALESRLPRGIQEVVIVGRRGPLQAAFTTLELRELA
DLDGVDVVIDPAELDGITDEDAAVGKVCQNIKVLRGYADRE-RPGHRRMVFRFLTSPIEIKGK---RKVE
RIVLGRNELVSDGSGRVAAKDTGEREELPAQLVRSVGYRGVPTPGLPFDDQSGTIPNVGGRI-----NGS
PNEYVVGWIKRGPTGVIGTNKKDAQDQDVTDLIKNLGNAKE--GAECKSFDHADQVADWLAARQPKLVTSAHW
QVIDAFERAAGEPHGRPRVKLASLAELLRI-GLG-
```

Without End Seq Matching too much

>466500

```
R---VAIIGSGPAGFYAAEALLKRTDTAVD----VDMFDRLPFPGLVRGGVAPDHQRIKAVTRVFASTAAR
PTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGE-GIERCTPATVFVGWYNGHPDYRHARFD
LSVRRAAVVGNGNVAVDVARILLRTAEELERTDIAAHALEALRESQVREVYLLGRRGPAQAASFPAELRELG
TLEAADPVVDPADLAGLSDPAPGG-----NLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAG-AVA
GLRLEKNRLERRPDGTVVARGTGETEVLEVGLVLPVAVGYAADRIPGVFPFDEKARVIANEDGRVVDPVLRSP
PGEYVVGWARSQGLIGEHRRAHVAHVMVAD-----AAGLATRPLPPRDAIDDLRARGVRPVSVFDW
ARLDEVEVARGARRGAPRDKLVDVAAMLAI-LGE
```

>1LQT_A

```
RPYYIAIVGSGPSAFFAAASLLKAADTTEDLMAVDMLEMLPTPWGLVRSGVAPDHPKIKSISKQFEKTAED
```

PRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGE-DLPGSIAAVDFVGWYNAHPHFQVSPD
 LSGARAVVIGNGNVALDVARILLTDPDLARTDIADHALESRLPRGIQEVVIVGRRGPLQAAFTTLELREL
 DLDGVDVVIDPAELDGITDEDAAVGKVCQNIKVLRGYADRE-RPGHRRMVFRFLTSPIEIKGK---RKVE
 RIVLGRNELVSDGSGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQSGTIPNVGGRI-----NGS
 PNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFDHADQVADWLAARQPKLVTSAHWQV
 IDAFERAAGEPHGRPRVKLASLAELLRIGLG---

>466500

--MRVAIIGSGPAGFYAAEALLKRTDT----AVDVMDFDRLPTPFGVLVRGGVAPDHQRIK
 AVTRVFASTAARPTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGEGIER
 CTPATVFVGWYNGHPDYRHARFDSLVRRAAVVGNNGNAVVDVARILLRTAEELERTDIAAH
 ALEALRESQVREVYLLGRRGPAQAASFPAELRELGTLEAADPVVDPADLAGLSDPAPGGN
 LEILRS-----FAARRPRPDARRLHLRFLVSPTEVLADAAGAVAGLRLEKNRLERRPDG
 TVVARGTGETEVLEVGLVLPVAVGYAADRIPGVPFDEKARVIANEDGRVVDPVLRSPPLGE
 YVVGWARSQPQGLIGEHRRAAHVVAHMAVADAA----GLATRPLPPRDAIDVLRARGVR
 PVSFVDWARLDEVEVARGARRGAPRDKLVDVAAMLAILGE

>1LQT_A

RPYYIAIVGSGPSAFFAAASLLKAADTTEDLMAVDMLEMLPTPWGLVRSVAPDHPKIK
 SISKQFEKTAEDPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGEDLPG
 SIAAVDFVGWYNAHPHFQVSPDLSGARAVVIGNGNVALDVARILLTDPDLARTDIADH
 ALESRLPRGIQEVVIVGRRGPLQAAFTTLELRELADLDGVDVVIDPAELDGITDEDAAV
 GKVCQNIKVLRGYADRE-RPGHRRMVFRFLTSPIEIKGKRK--VERIVLGRNELVSDGSG
 RVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQSGTIPNVGGRI-----SPNE
 YVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFDHADQVADWLAARQPK
 LVTSAHWQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG--

Profile-Profile-without MET

>466500

R---VAIIGSGPAGFYAAEALLKRTDTAVD----VDMDFDRLPTPFGVLVRGGVAPDHQRIKAVTRVFASTAA
 RPTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGE-GIERCTPATVFVGWYNGHPDYRHARF
 DLSVRRRAAVVGNNGNAVVDVARILLRTAEELERTDIAAHALEALRESQVREVYLLGRRGPAQAASFPAELREL
 GTLEAADPVVDPADLAGLSDPAPGG-----NLEILRSFAARRPRPDARRLHLRFLVSPTEVL
 ADAAG-AVAGLRLEKNRLERRPDGTVVARGTGETEVLEVGLVLPVAVGYAADRIPGVPFDEKARVIANED
 GRVVDPVLRSPPLGEYVVGWARSQPQGLIGEHRRAAHVVAHMAVAD-----AAGLATRPLPPRDAIDVLR
 ARGVRPVSFVDWARLDEVEVARGARRGAPRDKLVDVAAMLAI-LGE

>1LQT_A

RPYYIAIVGSGPSAFFAAASLLKAADTTEDLMAVDMLEMLPTPWGLVRSVAPDHPKIKSISKQFEKTAE
 DPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGE-DLPGSIAAVDFVGWYNAHPHFQVSP
 DLSGARAVVIGNGNVALDVARILLTDPDLARTDIADHALESRLPRGIQEVVIVGRRGPLQAAFTTLELREL
 ADLDGVDVVIDPAELDGITDEDAAVGKVCQNIKVLRGYADREPRPGHRRMVFRFLTSPIEIK
 GK---RKVERIVLGRNELVSDGSGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQSGTIPNVG
 GRI-----NGSPNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEDHADQVADWLA
 ARQPKLVTSAHWQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG---

Profile-Profile

>466500

MR---VAIIGSGPAGFYAAEALLKRTDTAVD----VDMDFDRLPTPFGVLVRGGVAPDHQRIKAVTRVFASTAA
 RPTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGE-GIERCTPATVFVGWYNGHPDYRHARF
 DLSVRRRAAVVGNNGNAVVDVARILLRTAEELERTDIAAHALEALRESQVREVYLLGRRGPAQAASFPAELREL
 GTLEAADPVVDPADLAGLSDPAPGG-----NLEILRSFAARRPRPDARRLHLRFLVSPTEVL
 ADAAG-AVAGLRLEKNRLERRPDGTVVARGTGETEVLEVGLVLPVAVGYAADRIPGVPFDEKARVIANED
 GRVVDPVLRSPPLGEYVVGWARSQPQGLIGEHRRAAHVVAHMAVAD-----AAGLATRPLPPRDAIDVLR
 ARGVRPVSFVDWARLDEVEVARGARRGAPRDKLVDVAAMLAI-LGE

>1LQT_A

MRPYYIAIVGSGPSAFFAAASLLKAADTTEDLMAVDMLEMLPTPWGLVRSVAPDHPKIKSISKQFEKTAE
 DPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGE-DLPGSIAAVDFVGWYNAHPHFQVSP
 DLSGARAVVIGNGNVALDVARILLTDPDLARTDIADHALESRLPRGIQEVVIVGRRGPLQAAFTTLELREL
 ADLDGVDVVIDPAELDGITDEDAAVGKVCQNIKVLRGYADREPRPGHRRMVFRFLTSPIEIK
 GK---RKVERIVLGRNELVSDGSGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQSGTIPNVG
 GRI-----NGSPNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEDHADQVADWLA
 ARQPKLVTSAHWQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG---

PHYRE

>Undefined

MRVAIIGSGPAGFYAAEALLK---
 RTDTAVDVMDFDRLPTPFGVLVRGGVAPDHQRIKAVTRVFASTAARPTFRFLGNVRLGRDVTVDLRRHYHQIVYA
 TGSESDRRLGIPGEGIERCTPATVFVGWYNGHPDYRHARFDSLVRRAAVVGNNGNAVVDVARILLRTAEELERTDIAAH
 ALEALRESQVREVYLLGRRGPA

QAAFPSPAELELGTLEAADPVVDPADLAGLS-----
 DPAPGGNLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAGAVAGLRLEKNRLERRPDGT
 VVARGTGETEVLEVLVPAVGAAADRIPIGVFDEKARVIANEDGRVVDVPLRSPLPGEYVVGWARS GPQGLIGEHRRASAHVVAH
 MVADAAGLATRPLP
 PR-----DAIDVLRARGVRPVSVFDWARLDEVEVARGARRGAPRDKLVDVAAMLAILGE
 >c1lqtB_
 YYIAIVGSGPSAFFAAASLLKAADTTEDLDMVDMLEMLPTPWGLVRSVGVAPDHPKISISKQFEKTAEDPRFRFFGNVVVGEHVQPG
 ELSERYDAVIYA
 VGAQSDRMLNIPGEDLPGSIAAVDFVWYNAHPHFEQVSPDLSGARAVVIGNGNVALDVARILLTDPDLARTDIADHALESRLPRGI
 QEVVIVGRRGPL
 QAAFTTLELRELADLDGVDVVIDPAELDGITDEDAAGVGVCKQNIKVLRGYADREPRPGHRRMVFRFLTSPIEIKGK--
 RKVERIVLGRNELVSDGSGR
 VAAKDTGEREELPAQLVVRVSVGYRGVPTPLPFDDQSGTIPNVGG-----
 RINGSPNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAEC
 KSFPDHADQVADWLAARQPKLVSAHWQVIDAFERAAGEPHGRPRVKLASLAELLRI-----

MAFFT

>466500
 MR---VAIIGSGPAGFYAAEALLKRTDTAVD----VDMFDRLPPTFGLVRGGVAPDHQRI
 KAVTRVFASTAARPTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGEGIE
 RCTPATVVFVGWYNGHPDYRHARFDLSVRRAAVVGNGNVAVDVARILLRTRAELERTDIAA
 HALEALRESQVREVYLLGRRGPAQAAFPSPAELELGTLEAADPVVDPADLAGLSDPAPGG
 -----NLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAGAVAGLRLEKNRLERRP
 DGTVVARGTGETEVLEVLVPAVGAAADRIPIGVFDEKARVIANEDGRVVDVPLRSPLP
 GEYVVGWARS GPQGLIGEHRRASAHVVAH MV-----ADAAGLATRPLPRDAIDVLR
 RGVRPVSVFDWARLDEVEVARGARRGAPRDKLVDVAAMLAILGE
 >1LQT_A
 MRPYIAIVGSGPSAFFAAASLLKAADTTEDLDMVDMLEMLPTPWGLVRSVGVAPDHPKI
 KSISKQFEKTAEDPRFRFFGNVVVGEHVQPGELSERVDAVIYAVGAQSDRMLNIPGEDLP
 GSIAAVDFVWYNAHPHFEQVSPDLSGARAVVIGNGNVALDVARILLTDPDLARTDIAD
 HALESRLPRGIQEVVIVGRRGPLQAAFTTLELRELADLDGVDVVIDPAELDGITDEDAAG
 VGVCKQNIKVLRGYADREPRPGHRRMVFRFLTSPIEI--KGKRVKVERIVLGRNELVSDG
 SGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPLPFDDQSGTIPNVGGRI-----NGSP
 NEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEDHADQVADWLA
 RQPKLVSAHWQVIDAFERAAGEPHGRPRVKLASLAELLRI-GLG-

R-bio3D

Created Sunday 20 October 2013

> seq <-

"MRVAIIGSGPAGFYAAEALLKRTDTAVD VDMFDRLPPTFGLVRGGVAPDHQRIKAVTRVFASTAARPTFRFLGNVRLGRDVTVDLRRHYHQIVYATGSESDRRLGIPGEGIERCTPATVVFVGWYNGHPDYRHARFDLSVRRAAVVGNGNVAVDVARILLRTRAELERTDIAAH
 ALEALRESQVREVYLLGRRGPAQAAFPSPAELELGTLEAADPVVDPADLAGLSDPAPGGNLEILRSFAARRPRPDARRLHLRFLVSPTEVLADAAGAVAGLRLEKNRLERRPDGTVVARGTGETEVLEVLVPAVGAAADRIPIGVFDEKARVIANEDGRVVDVPLRSPLPGEY
 VVGWARS GPQGLIGEHRRASAHVVAH MVADAAGLATRPLPRDAIDVLRARGVRPVSVFDWARLDEVEVARGARRGAPRDKLVD
 VAAMLAILGE"

seq <- as.character(unlist(strsplit(seq, "")))

> seq

[1] "M" "R" "V" "A" "I" "I" "G" "S" "G" "P" "A" "G" "F" "Y" "A" "A" "E" "A"
 [19] "L" "L" "K" "R" "T" "D" "T" "A" "V" "D" "V" "D" "M" "F" "D" "R" "L" "P"
 [37] "T" "P" "F" "G" "L" "V" "R" "G" "G" "V" "A" "P" "D" "H" "Q" "R" "I" "K"
 [55] "A" "V" "T" "R" "V" "F" "A" "S" "T" "A" "A" "R" "P" "T" "F" "R" "F" "L"
 [73] "G" "N" "V" "R" "L" "G" "R" "D" "V" "T" "V" "D" "D" "L" "R" "R" "H" "Y"
 [91] "H" "Q" "I" "V" "Y" "A" "T" "G" "S" "E" "S" "D" "R" "R" "L" "G" "I" "P"
 [109] "G" "E" "G" "I" "E" "R" "C" "T" "P" "A" "T" "V" "F" "V" "G" "W" "Y" "N"
 [127] "G" "H" "P" "D" "Y" "R" "H" "A" "R" "F" "D" "L" "S" "V" "R" "R" "A" "A"
 [145] "V" "V" "G" "N" "G" "N" "V" "A" "V" "D" "V" "A" "R" "I" "L" "L" "R" "T"
 [163] "R" "A" "E" "L" "E" "R" "T" "D" "I" "A" "A" "H" "A" "L" "E" "A" "L" "R"
 [181] "E" "S" "Q" "V" "R" "E" "V" "Y" "L" "L" "G" "R" "R" "G" "P" "A" "Q" "A"
 [199] "A" "F" "S" "P" "A" "E" "L" "R" "E" "L" "G" "T" "L" "E" "A" "A" "D" "P"
 [217] "V" "V" "D" "P" "A" "D" "L" "A" "G" "L" "S" "D" "P" "A" "P" "G" "G" "N"
 [235] "L" "E" "I" "L" "R" "S" "F" "A" "A" "R" "R" "P" "R" "P" "D" "A" "R" "R"
 [253] "L" "H" "L" "R" "F" "L" "V" "S" "P" "T" "E" "V" "L" "A" "D" "A" "A" "G"
 [271] "A" "V" "A" "G" "L" "R" "L" "E" "K" "N" "R" "L" "E" "R" "R" "P" "D" "G"
 [289] "T" "V" "V" "A" "R" "G" "T" "G" "E" "T" "E" "V" "L" "E" "V" "G" "L" "V"
 [307] "L" "P" "A" "V" "G" "Y" "A" "A" "D" "R" "I" "P" "G" "V" "P" "F" "D" "E"
 [325] "K" "A" "R" "V" "I" "A" "N" "E" "D" "G" "R" "V" "V" "D" "P" "V" "L" "R"
 [343] "S" "P" "L" "P" "G" "E" "Y" "V" "V" "G" "W" "A" "R" "S" "G" "P" "Q" "G"

```

[361] "L" "I" "G" "E" "H" "R" "A" "S" "A" "H" "V" "V" "A" "H" "M" "V" "A"
[379] "D" "A" "A" "G" "L" "A" "T" "R" "P" "L" "P" "P" "R" "D" "A" "I" "D" "D"
[397] "V" "L" "R" "A" "R" "G" "V" "R" "P" "V" "S" "F" "V" "D" "W" "A" "R" "L"
[415] "D" "E" "V" "E" "V" "A" "R" "G" "A" "R" "R" "G" "A" "P" "R" "D" "K" "L"
[433] "V" "D" "V" "A" "A" "M" "L" "A" "I" "L" "G" "E" "G" "E"
score <-
"9899989587999999998469999947997779988827861669899781089999999998499199918486772899999844999999628898877
997998898848789999998589655578877828839998997589999999703900103665306778776426753699995789788487748887640
3667530027445224445034444567776420255677754799944785599628998288999998984379987110279955999899987347778
888887464678877789963816887777999898028853752889999999999999741410688788987889999986399751699999999999
827122899246299999998741489"
score <- as.character(unlist(strsplit(score, "")))
seqscore <- as.data.frame(cbind(seq, score))
> seqscore <- cbind(as.character(seq[2:length(seq)]), as.numeric(score[2:length(score)]))
> targetType[targetType$Type == "AlphaHelix",]
Type Start.Res Start.Position End.Res End.Position Interval
2 AlphaHelix PRO 9 THR 24 15
4 AlphaHelix LEU 40 GLY 44 4
6 AlphaHelix VAL 58 ALA 64 6
8 AlphaHelix VAL 82 HIS 88 6
11 AlphaHelix ALA 117 ASN 125 8
13 AlphaHelix ASN 149 LEU 159 10
14 AlphaHelix ARG 162 GLU 166 4
15 AlphaHelix ALA 172 ARG 179 7
18 AlphaHelix PRO 201 GLU 206 5
22 AlphaHelix ASP 227 ALA 242 15
31 AlphaHelix TRP 352 SER 355 3
32 AlphaHelix GLY 362 ASP 378 16
33 AlphaHelix PRO 389 ARG 400 11
35 AlphaHelix PHE 407 ARG 423 16
36 AlphaHelix VAL 434 ILE 440 6
>
> ahelix <- targetType[targetType$Type == "AlphaHelix",]
seqscore[(ahelix$Start.Position:ahelix$End.Position),]
> for(x in 1:length(ahelix)){
print(seqscore[ahelix$Start.Position[x]:ahelix$End.Position[x],])
print(seqscore[ahelix$Start.Position[x]:ahelix$End.Position[x],])
}
for(x in 1:nrow(ahelix)){
print(mean(as.numeric(seqscore[ahelix$Start.Position[x]:ahelix$End.Position[x],][,2])))
}
for(x in 1:nrow(targetType)){
targetType$Interval[x] <- (mean(as.numeric(seqscore[targetType$Start.Position[x]:targetType$End.Position[x],][,2])))
}
plotmatrix(with(targetType, data.frame(Type, Interval)))

```

Threading

Created Monday 18 November 2013

Threaded target prediction reveals FDPNDO secondary structure is comprised of 4 sheets

BA contains 6 strands with mixed parallel/anti parallel (

BA1; 3-5,

BA2; 29-32,

BA3; 69-72,

BA4; 92-94,

BA5; 348-349,

BA6 405-406)

BB contains 2 antiparallel strands (

BB1; 100-101,

BB2; 311-312)

BC contains 5 parallel strands (

BC1; 141-145,

BC2; 185-189,

BC3; 213-216,

BC4; 250-255,

BC5; 304-307)

BD contains 3 antiparallel stands (
 BD1; 258-264,
 BD2; 273-283,
 BD3; 289-301)
 4 beta alpha beta units exist involving: BA1XBA2;
 2 beta hairpins
 3 beta bulges
 16 strands
 22 helices
 18 helix-helix interac
 25 beta turns
 1 gamma turn

8-Thread pdb from alignment

Created Saturday 19 October 2013

python

**/media/DATA/Rosetta/rosetta_tools/scripts/thread_pdb_from_alignment.py --
 template=1LQT_A --target=466500 --chain=A --align_format=fasta finalAln.fa
 1LQT_A.pdb 466500_threaded.pdb**

dashte@dashte:~/Documents/HOMOLOGY/2-Model\$ **python**
/media/DATA/Rosetta/rosetta_tools/scripts/thread_pdb_from_alignment.py --template=1LQT_A --target=466500 --chain=A --
align_format=fasta finalAln.fa 1LQT_A.pdb 466500_threaded.pdb

117 117
 260 260
 282 284
 352 352
 354 356
 361 362
 401 402
 463 463
 2 4
 31 34
 117 117
 241 247
 285 285
 395 400

WARNING: gap of size -2 in target sequence.

We cannot completely thread this protein in an automatic way,
 manual inspection and adjustment of loop files will be required.

WARNING: gap of size -3 in target sequence.

We cannot completely thread this protein in an automatic way,
 manual inspection and adjustment of loop files will be required.

WARNING: gap of size -6 in target sequence.

We cannot completely thread this protein in an automatic way,
 manual inspection and adjustment of loop files will be required.

WARNING: gap of size -5 in target sequence.

We cannot completely thread this protein in an automatic way,
 manual inspection and adjustment of loop files will be required.

dashte@dashte:~/Documents/HOMOLOGY/2-Model\$ **python**
/media/DATA/Rosetta/rosetta_tools/scripts/thread_pdb_from_alignment.py --template=1LQT_A --target=466500 --
align_format=fasta finalAln.fa 1LQT_A.pdb 466500_threaded.pdb

118 118
 283 285
 353 358
 3 5
 32 35
 118 118
 242 248
 286 286
 396 401
 464 464

WARNING: gap of size -2 in target sequence.

We cannot completely thread this protein in an automatic way,

manual inspection and adjustment of loop files will be required.

WARNING: gap of size -3 in target sequence.

We cannot completely thread this protein in an automatic way,
manual inspection and adjustment of loop files will be required.

WARNING: gap of size -6 in target sequence.

We cannot completely thread this protein in an automatic way,
manual inspection and adjustment of loop files will be required.

WARNING: gap of size -5 in target sequence.

We cannot completely thread this protein in an automatic way,
manual inspection and adjustment of loop files will be required.

ARG MET

Residue mismatch between alignment and PDB, check that PDB sequence and alignment sequence are identical

11-Loops File

Created Sunday 20 October 2013

Att3 - only assign loops at weak regions

LOOP 2 6 0 0 0
 LOOP 27 34 0 0 0
 LOOP 64 80 0 0 0
 LOOP 113 117 0 0 0
 LOOP 132 136 0 0 0
 LOOP 161 165 0 0 0
 LOOP 180 183 0 0 0
 LOOP 201 203 0 0 0
 LOOP 229 234 0 0 0
 LOOP 245 250 0 0 0
 LOOP 267 271 0 0 0
 LOOP 284 293 0 0 0
 LOOP 315 318 0 0 0
 LOOP 332 342 0 0 0
 LOOP 383 391 0 0 0
 LOOP 423 425 0 0 0

Attempt #2 - shorten loops

LOOP 2 4 0 0 0
 LOOP 26 29 0 0 0
 LOOP 60 63 0 0 0
 LOOP 78 81 0 0 0
 LOOP 112 117 0 0 0
 LOOP 132 136 0 0 0
 LOOP 161 165 0 0 0
 LOOP 181 183 0 0 0
 LOOP 229 234 0 0 0
 LOOP 245 250 0 0 0
 LOOP 267 271 0 0 0
 LOOP 284 287 0 0 0
 LOOP 315 318 0 0 0
 LOOP 337 342 0 0 0
 LOOP 345 347 0 0 0
 LOOP 379 391 0 0 0
 LOOP 407 410 0 0 0

Attempt#1

2	4	0	0	0
27	30	0	0	0
33	34	0	0	0
64	67	0	0	0
79	80	0	0	0
113	117	0	0	0
132	136	0	0	0
161	165	0	0	0
181	183	0	0	0
201	203	0	0	0
229	231	0	0	0
233	234	0	0	0

```

245      246      0      0      0
267      271      0      0      0
284      287      0      0      0
315      317      0      0      0
332      333      0      0      0
338      342      0      0      0
383      391      0      0      0
423      425      0      0      0

```

Loop regions were identified using Phyre2 homology prediction server alignment of Predicted Secondary Structure against Template known and predicted secondary structure along with I-TASSER Predicted Secondary Structure prediction.

Mean SS confidence score was analyzed through R as 6.77, to which we rounded down to 6. 73.9% residues had confidence scores above

Regions where Confidence score

LOOP 117 117

282 284

354 356

361 362

401 402

2 4

31 34

241 247

285 285

395 400

TABLE 1 | Explanation of information contained in the loops file (Step 11).

Column 1	LOOP	The loops file identity tag
Column 2	< integer> ^a	Loop start residue number. <i>Note:</i> the starting structure must have real coordinates for all residues outside the loop definition, plus the first and last residue of each loop region
Column 3	< integer>	Loop end residue number
Column 4	< integer>	Cut point residue number, must be greater than the first residue of the loop and less than the end residue of the loop. Default (0)—let loop rebuild protocol choose cut point
Column 5	< float>	Skip rate. Default (0)—never skip
Column 6	< boolean>	Extend loop. Default (0)—false

^aThe < > indicates areas where the user is to specify the integer, float or boolean (0 for false or 1 for true).

12-modeling.options

Created Tuesday 05 November 2013

```

-loops:input_pdb 466500_threaded.pdb #input threaded pdb file
-loops:fa_input #input will be in all-atom mode
-loops:loop_file 466500_.loops #generated loop definitions
-loops:frag_size 9 3 1
-loops:frag_files aat000_09_05.200_v1_3 aat000_03_05.200_v1_3 none
-loops:remodel quick_ccd
-loops:refine refine_kic
-loops:extended true
-loops:idealize_after_loop_close
-loops:relax fastrelax
-loops:fast
-ex1
-ex2

```

13-running CM job

Created Tuesday 05 November 2013

```

/home/swajid/Rosetta/rosetta_2013wk42_bundle/main/source/bin/loopmodel.default.linuxgccrelease @modeling.options -database
/home/swajid/Rosetta/rosetta_2013wk42_bundle/main/database -nstruct 5
/media/DATA/Rosetta/rosetta_2013wk40_bundle/main/source/bin/loopmodel.default.linuxgccrelease @modeling.options -database
media/DATA/Rosetta/rosetta_2013wk40_bundle/main/database -nstruct 100
/media/DATA/Rosetta/rosetta_2013wk40_bundle/main/source/bin/loopmodel.default.linuxgccrelease \
-database media/DATA/Rosetta/rosetta_2013wk40_bundle/main/database \
-nstruct 100 \
-loops:input_pdb $HOME/Documents/HOMOLOGY/Nov5/2-Model/466500_threaded.pdb \
-loops:fa_input \
-loops:loop_file $HOME/Documents/HOMOLOGY/Nov5/2-Model/466500_.loops \
-loops:frag_size 9 3 1 \
-loops:frag_files $HOME/Documents/HOMOLOGY/Nov5/2-Model/aat000_09_05.200_v1_3
$HOME/Documents/HOMOLOGY/Nov5/2-Model/aat000_03_05.200_v1_3 none \
-loops:remodel quick_ccd \

```



```

-loops:refine refine_kic \
-loops:extended true \
-loops:idealize_after_loop_close \
-loops:relax fastrelax \
-loops:fast \
-ex1 \
-ex2 \
/media/DATA/Rosetta/rosetta_2013wk40_bundle/main/source/bin/loopmodel.default.linuxgccrelease@modeling.options - database
rosetta_database -nstruct 100
/media/DATA/Rosetta/rosetta_2013wk40_bundle/main/source/bin/loopmodel.default.linuxgccrelease \
-database media/DATA/Rosetta/rosetta_2013wk40_bundle/main/database \
-nstruct 100 \
-loops:input_pdb 466500_threaded.pdb \
-loops:fa_input \
-loops:loop_file 466500_.loops \
-loops:frag_size 9 3 1 \
-loops:frag_files aat000_09_05.200_v1_3 aat000_03_05.200_v1_3 none \
-loops:remodel quick_ccd \
-loops:refine refine_kic \
-loops:extended true \
-loops:idealize_after_loop_close \
-loops:relax fastrelax \
-loops:fast \
-ex1 \
-ex2 \
rosetta_source/bin/loopmodel.default.linuxgccrelease@modeling.options - database rosetta_database -nstruct 100

```