

Figure 2. Initial trimer design with Rosetta. (A) gp41 trimer (cold colors) shown in the context of whole Env (salmon). (B) disulphide bonds that were engineered SOS in between gp41 and gp120. (C) The c-terminus of alpha helical HR2 domain (S546 through Y588) interacts with the first helix in gp41 sequence (T530 through R544). RosettaDesign predicted mutation sites to strengthen HR2 – adjacent gp41 N-terminus interaction: N656 from one monomer (light purple) to M535 of HR2 from another monomer (green), and polar K655 which faces hydrophobic core. (D) Triple mutant M535W, N656Y and K655Y had the lowest Rosetta score. (E) Y656 and W535 pi-stacking interaction.

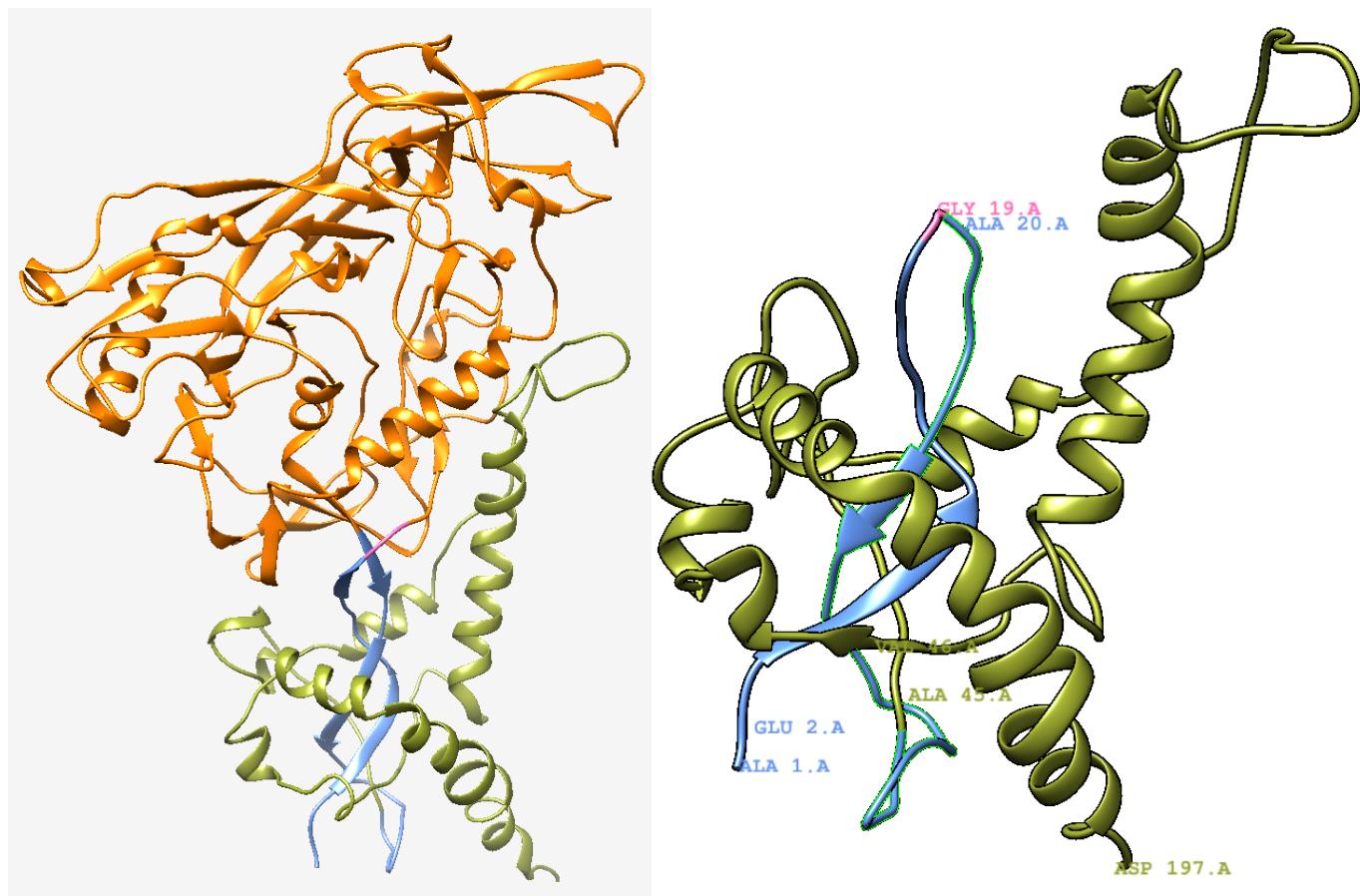


Figure 3. Gp140 (left) and gp140_trunc (right). Gp140 has a pg120 subunit (orange and blue) and gp41 (green). The gp120 portion that penetrates gp41's cavity is shown in blue, together with the linker region.

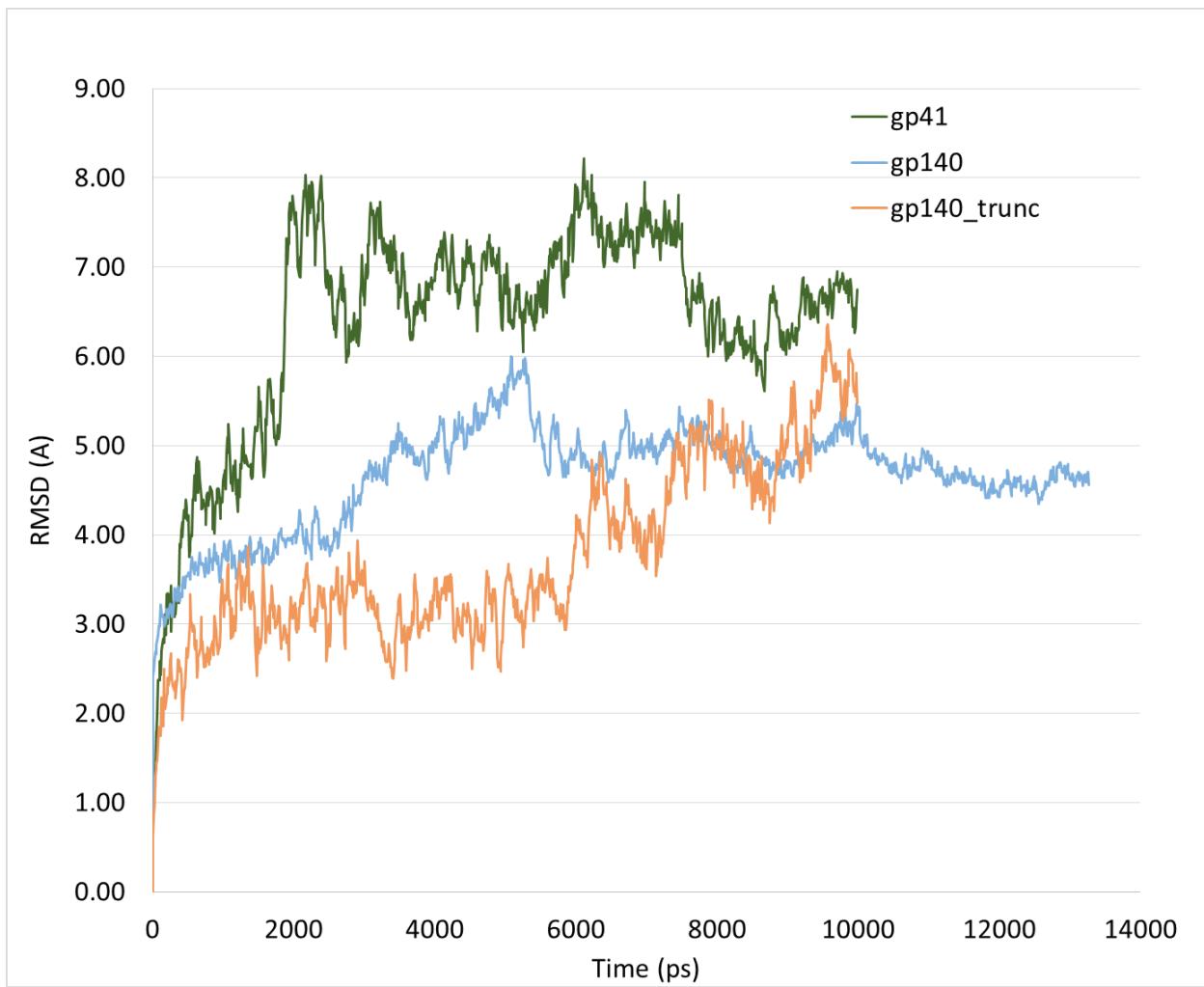


Figure 4. RMSD analysis of three constructs.

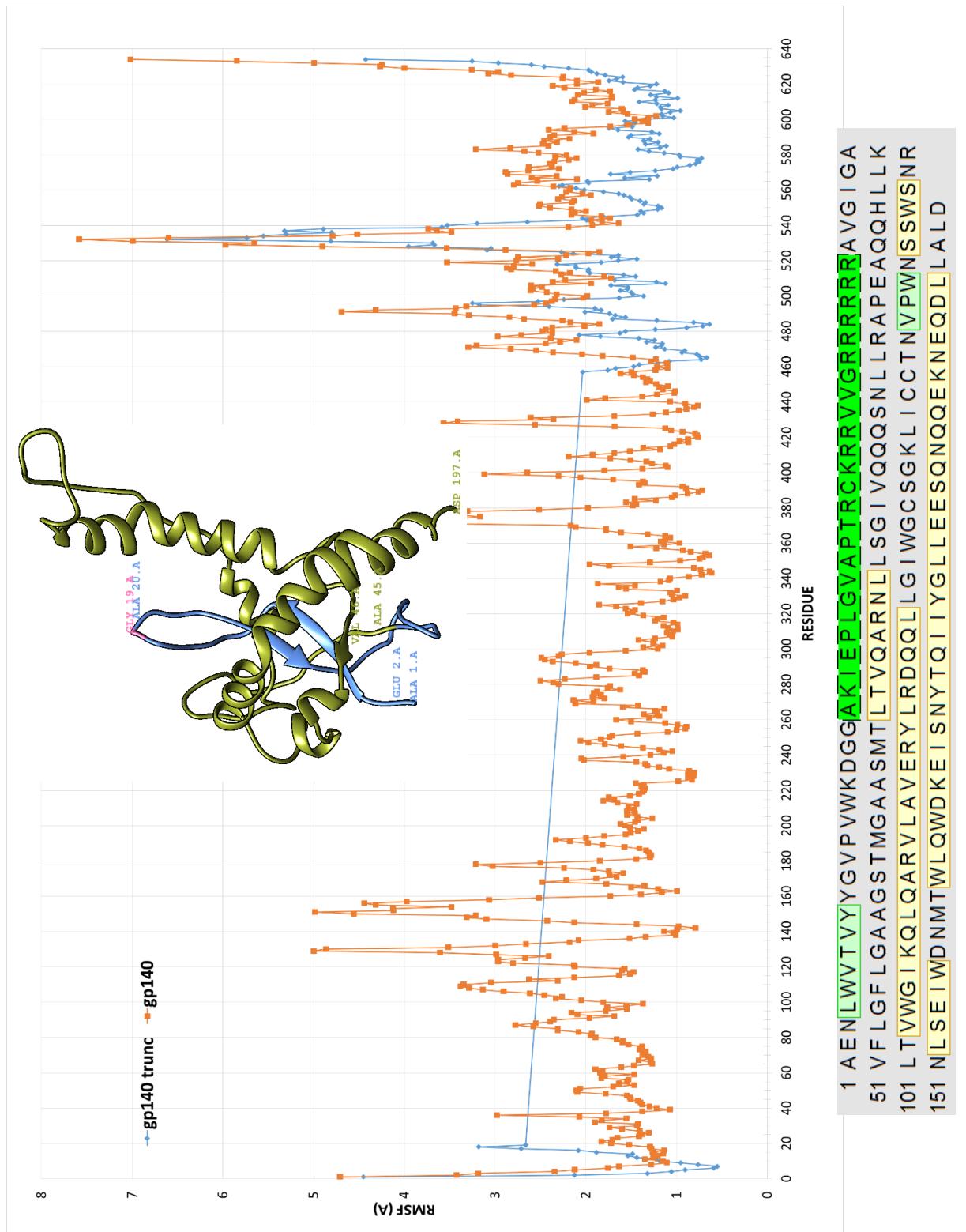


Figure 5. Aligned RMSF values for gp140 and gp140_trunc. Gp140_trunc is the same as gp41, except for a 44 aa sequence shown in green that was inserted, and originally belongs to gp120. The stabilizing insert is shown in blue in the cartoon.

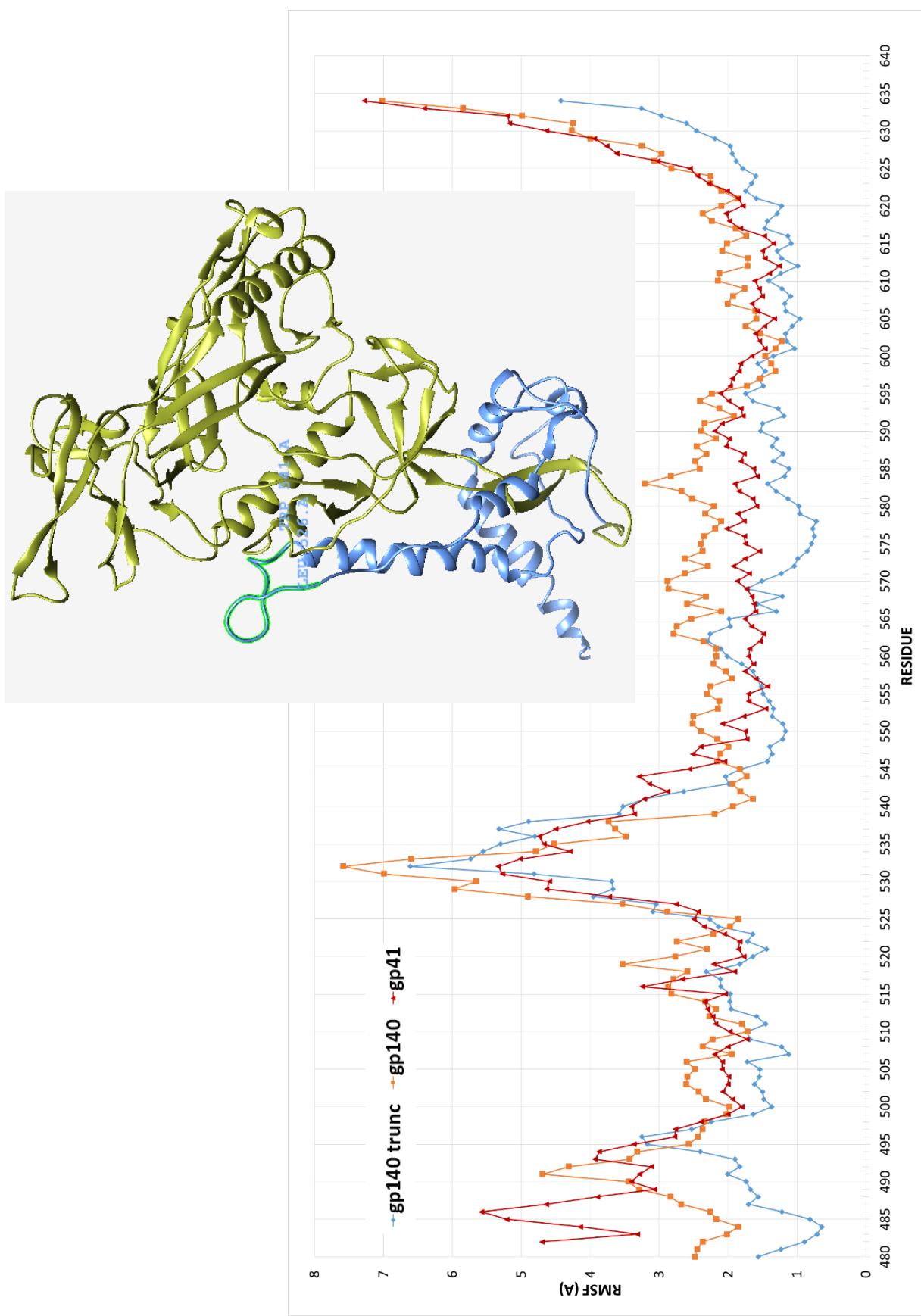


Figure 6. Drastic gp140 truncation leaving only 24 gp120 residues to stabilize gp41 produces the most stable structure. RMSF alignment of the corresponding gp41 sequence for gp140 (full gp120+gp41), full gp41, and truncated gp140. Highly disordered region L526-W541 shown in green.

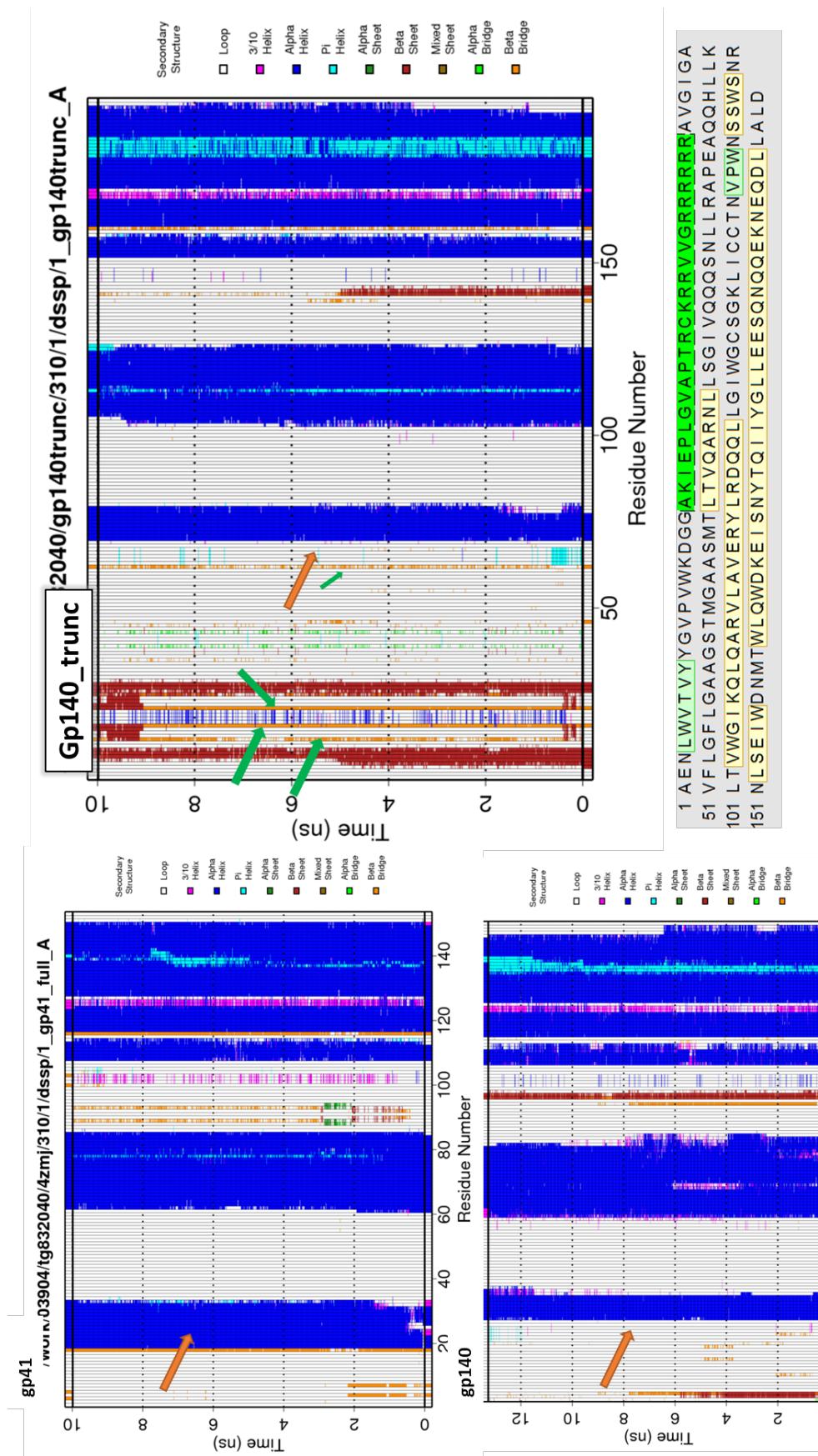


Figure 7. Secondary structure analysis indicates loss of alpha helical structure at 529-536 region of a helix that is adjacent to another protomer's HR1 (orange arrows), and is the site of M535W trimer-stabilizing mutation. Green arrows indicate sites of non-native salt bridge interactions in gp140_trunc.

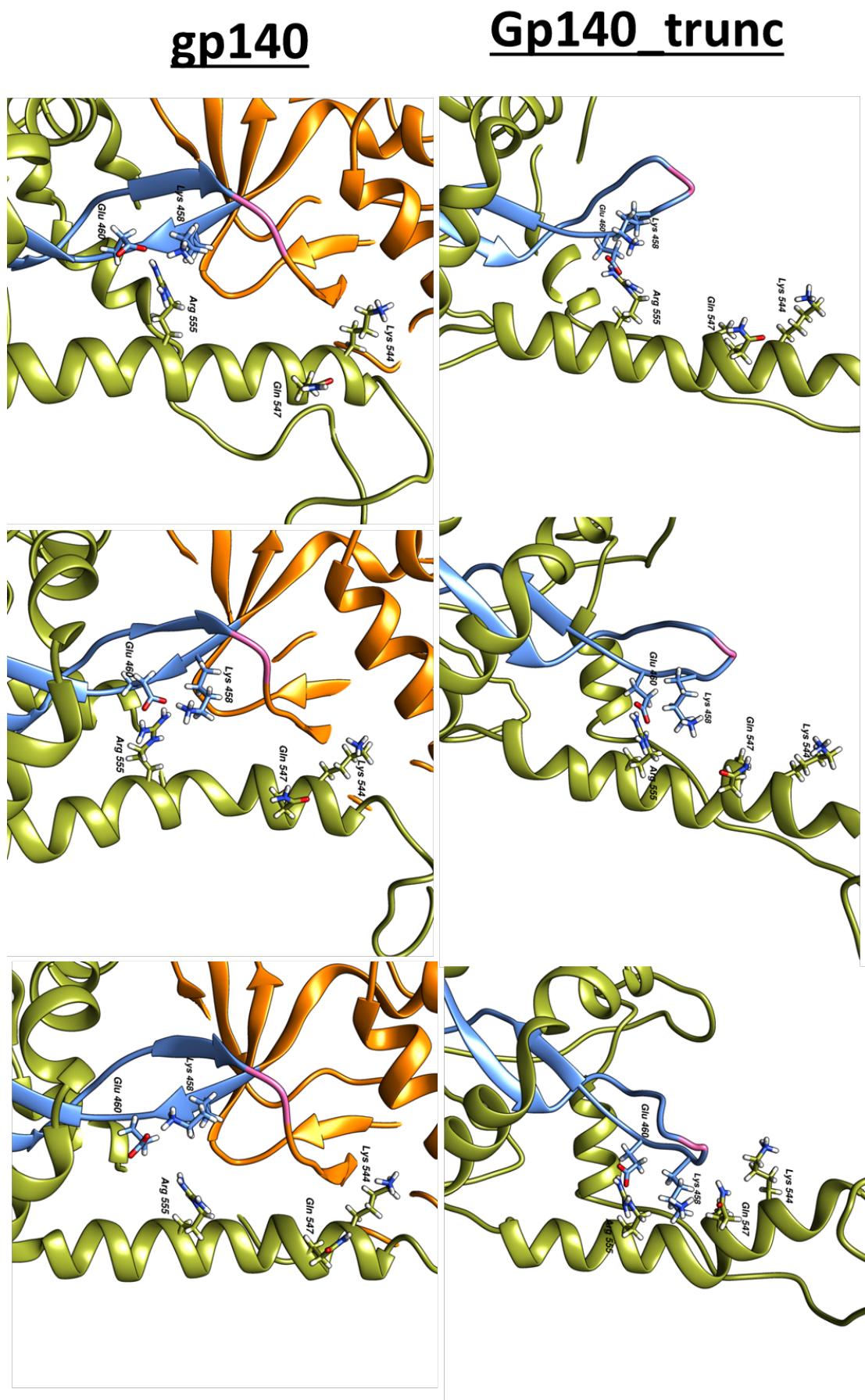


Figure 8. Transient and permanent salt bridges at E460-R555 and K458-Q547 break the HR1 alpha helix (HR1 helix is green). Residues from gp120 insert are blue.

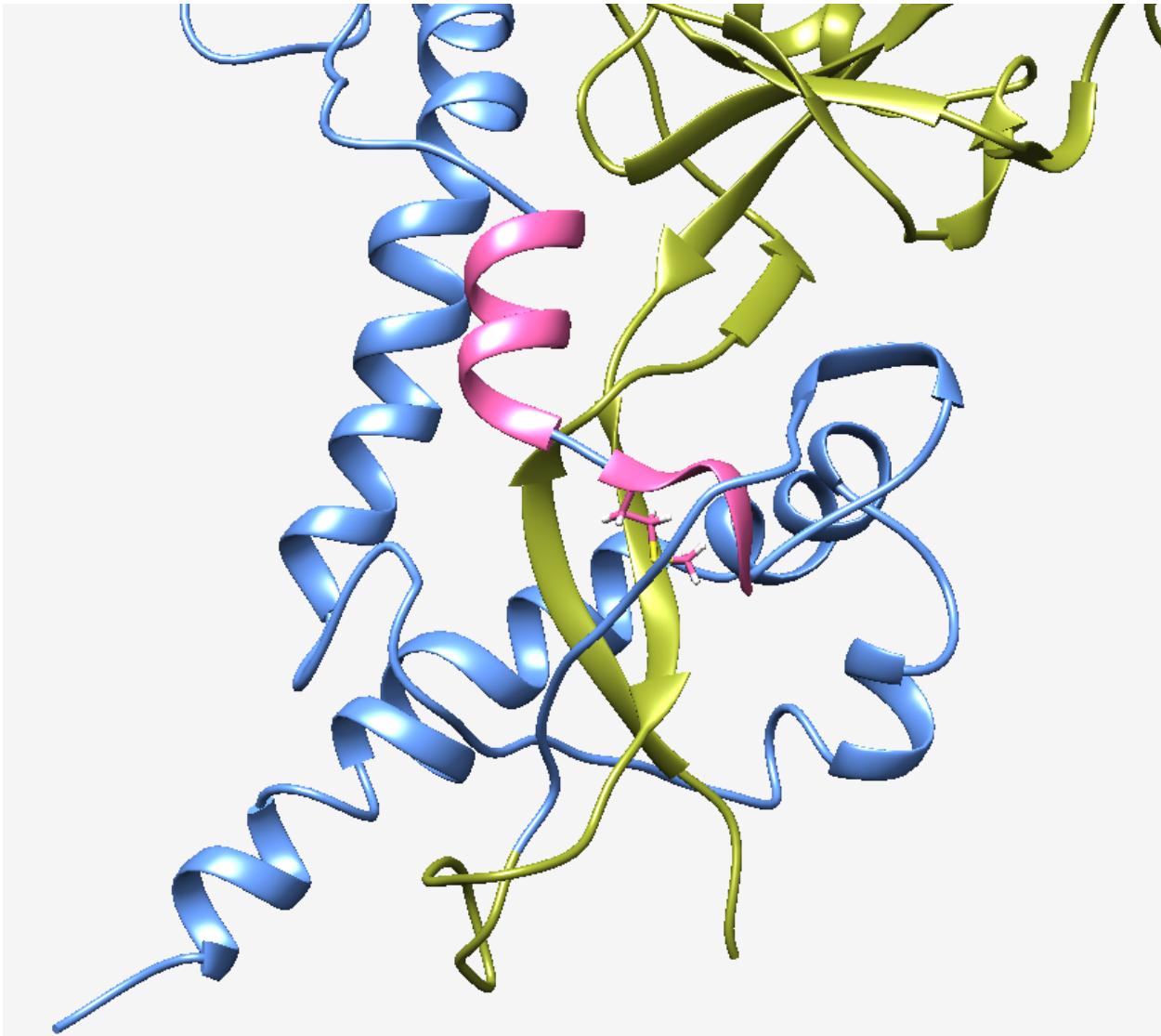


Figure 9. MD simulation proves that AASM(535)TLTVQARNL sequence (pink) does not have a stable helical secondary structure. Contrary to what the Xtal structure 4ZMJ is showing, the single helix breaks into two helices via T536. Therefore stabilizing M535 is of paramount importance. Thr536 is a blue linker, M535 is shown with side chain.

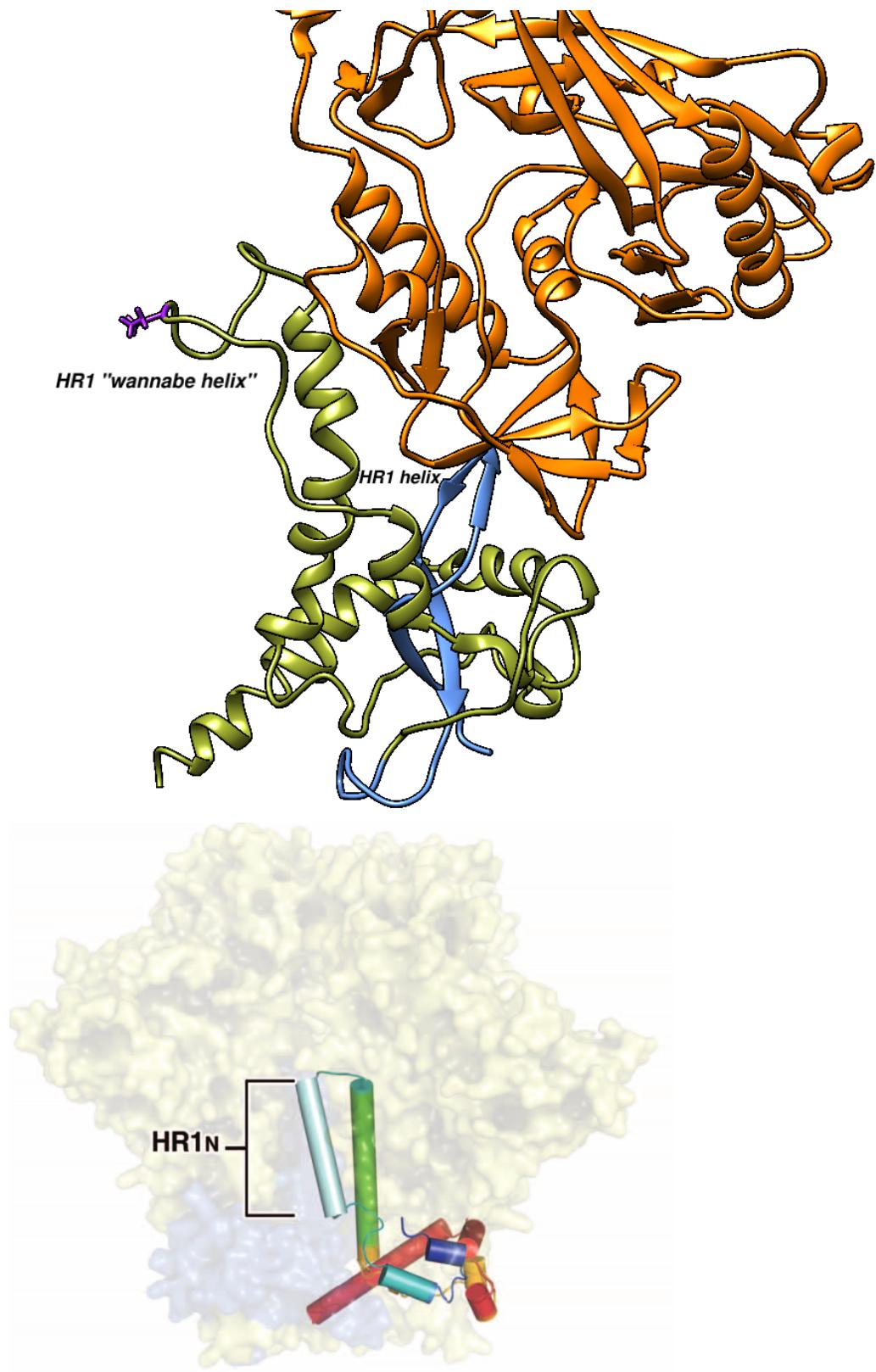


Figure 10. HR1's N-terminus wants to be a helix, but I559P prevents that.

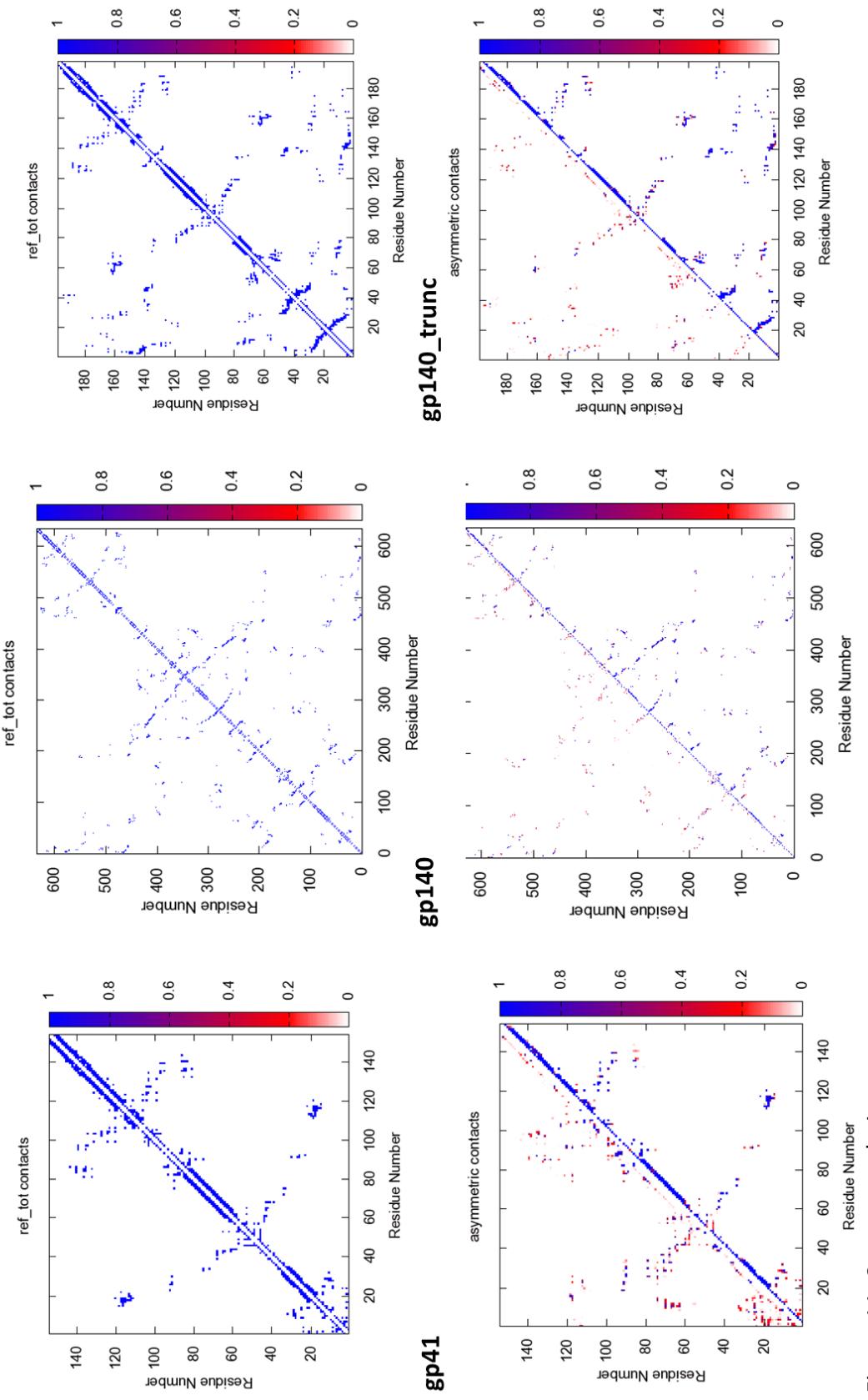
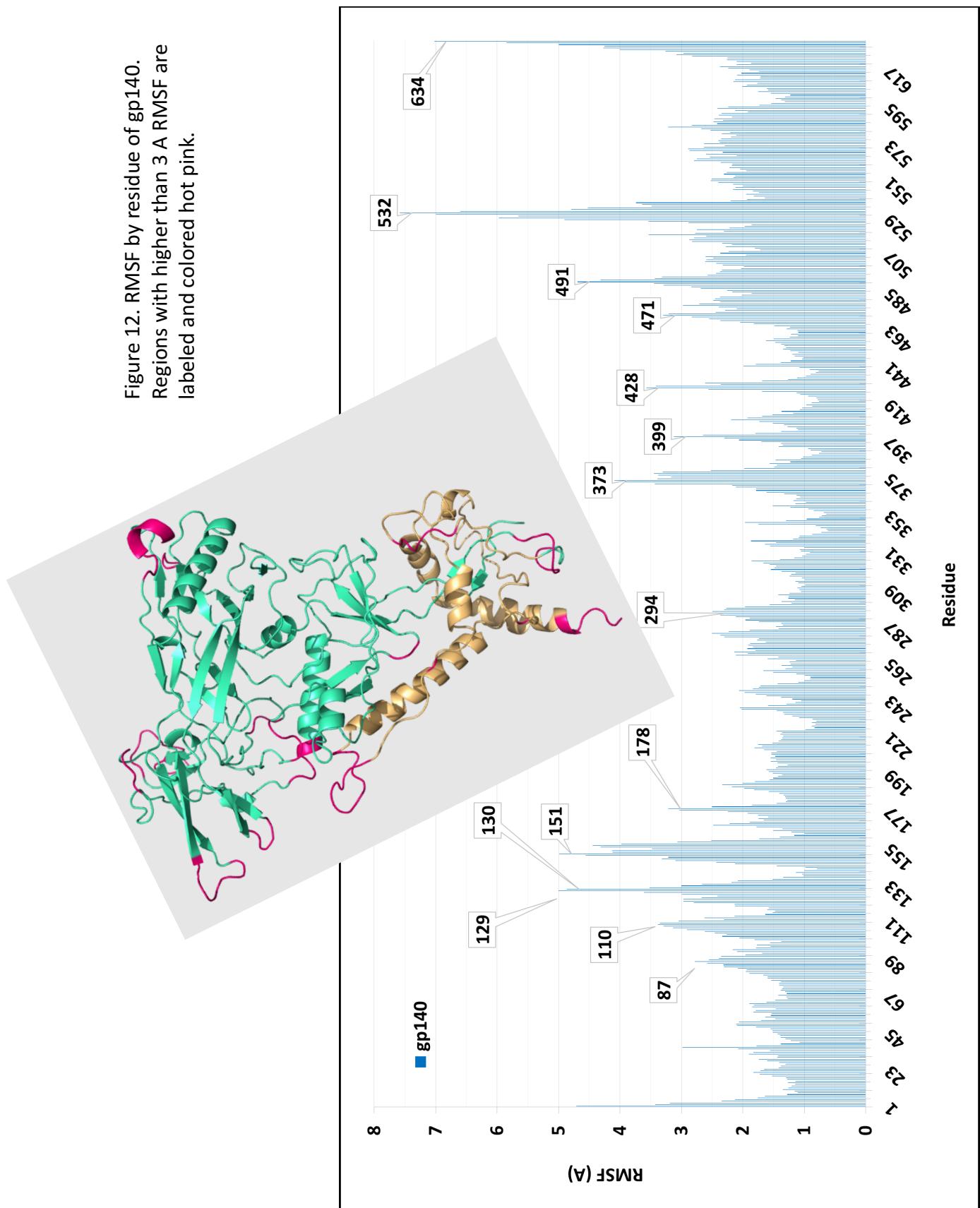
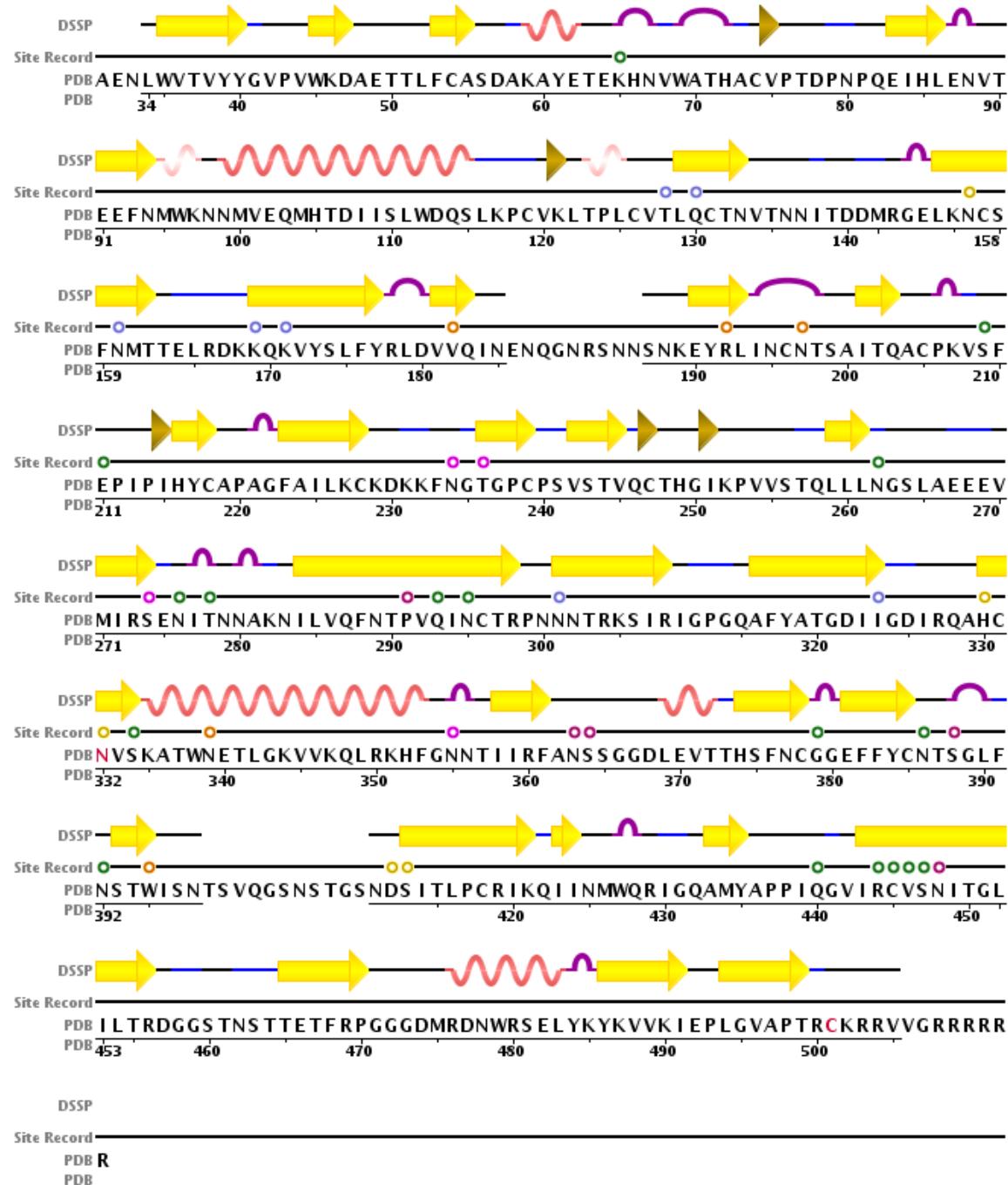


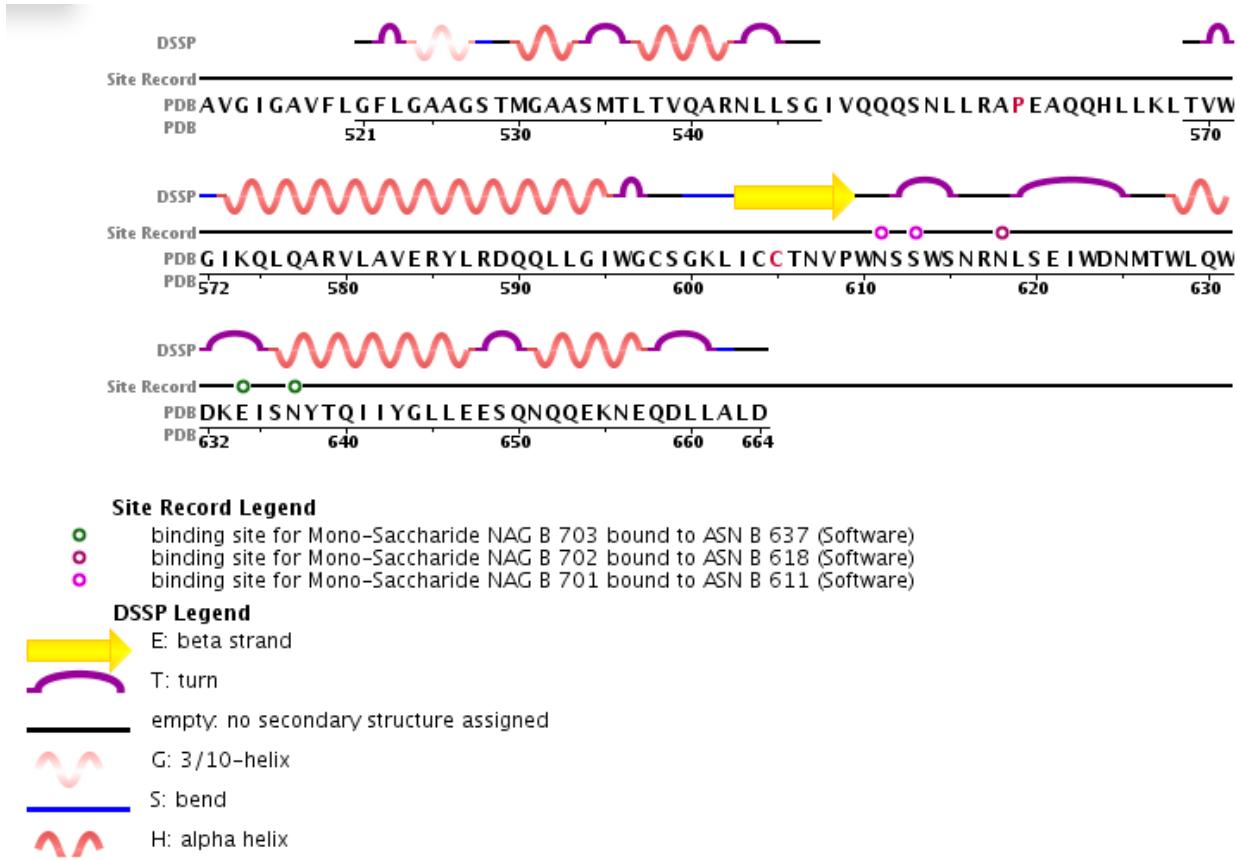
Figure 11. Contact analysis.

Figure 12. RMSF by residue of gp140.
Regions with higher than 3 Å RMSF are labeled and colored hot pink.





4zmj.pdb (#0) chain G 31 AENLWVTVYYGVPWKDAETTLFCASDAKAYETEKHNVWATHACVPTDPN
 4zmj.pdb (#0) chain G 81 PQEIHLENVTEEFNMWKNNMVEQMHTD IISLWDQSLKPCVKLTPLCVTLQ
 4zmj.pdb (#0) chain G 131 CTNVTNNITDDMRGEELKNCSFNMTTEL RDKKQKVYSLFYRLDVVQINENQ
 4zmj.pdb (#0) chain G 181 GNRSNNSNKEYRLINCNTSAITQACPKVSFEPIPIHYCAPAGFAILKCKD
 4zmj.pdb (#0) chain G 231 KKFNGTGPCPSVSTVQCTHGIKPVVSTQLLNLNGSLAEEEVMIRSENITNN
 4zmj.pdb (#0) chain G 281 AKNILVQFNTPVQINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHC
 4zmj.pdb (#0) chain G 331 NVSKATWNETLGKVVQLRKHFGNNTIIRFANSGGDLEVTTHSFNCGG
 4zmj.pdb (#0) chain G 381 FFYCNTSGLFNSTWISNTSVQGSNSTGSNDSTLPCRICKQIINMWQRIGQ
 4zmj.pdb (#0) chain G 431 AMYAPPIQGVIRCVSNIITGLILTRDGGSTNSTTETFRPGGGDMRDNWRS
 4zmj.pdb (#0) chain G 481 LYKYKVVKIEPLGVAPTRCKRRVVGRRRRR



```

4zmj_gp41model...principal chain 1 | AVG I GAVFL GFL G AAGS T MGA AS MTL TVQ ARNL LSG I VQQQS NLL RAP EAQQH LLKLT VW
4zmj_gp41model...principal chain 51 Q QH L L KLT VWG I KQLQARVL AVERYL RDQQL LG I WGCS GKL IC CTN VPWN
4zmj_gp41model...principal chain 101 S SWS NRNL SE I WDNMTWL QWD KE I SNYTQI I YGLLEES QNQQE KNE QDLL
4zmj_gp41model...principal chain 151 A LD

```

>P1;4zmj_trimer_restored_seq

AENLWVTVYYGVPWKDAETTLFCASDAKAYETEKHNVWATHACVPTDPNPQEIHLENVT
EEFNWKNNMVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCS
FNMTTELRDKKQKVYSLFYRLDVVQINENQGNRSNNSNKEYRLINCNTSAITQACPKVSF
EPIPIHYCAPAGFAILKCKDKKFNGTGPCSVSTVQCTHGIKPVVSTQLLLNGSLAEEEV
MIRSENITNNAKNILVQFNTPVQINCTRPNNTTRKSIRIGPGQAFYATGDIIGDIRQAHC
NVSKATWNELGKVVQLRKHFGNNTIIRFANSSGGDLEVTHSFNCGEFFYCNTSGLF
NSTWISNTSVQGSNSTGSNDSITLPCRIKQIINMWQRIGQAMYAPPIQGVIRCNSNITGL
ILTRDGGSTNSTTETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTRCKRRVGAVGIG
AVFLGFLGAAGSTMGAASMTLTQARNLLSGIVQQQSNLLRAPEAQHLLKLTWVGKQL
QARVLAVERYLRDQQQLLGIVGCGKLICTNVPWNSSWSNRNLSEIWDNMTWLQWDKEIS
NYTQIIYGLLEESQNQQEKNEQDLLALD*

>P1;4zmj_trimer.pdb

---LWVTVYYGVPWKDAETTLFCASDAKAYETEKHNVWATHACVPTDPNPQEIHLENVT
EEFNWKNNMVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCS
FNMTTELRDKKQKVYSLFYRLDVVQI-----NSNKEYRLINCNTSAITQACPKVSF
EPIPIHYCAPAGFAILKCKDKKFNGTGPCSVSTVQCTHGIKPVVSTQLLLNGSLAEEEV
MIRSENITNNAKNILVQFNTPVQINCTRPNNTTRKSIRIGPGQAFYATGDIIGDIRQAHC
NVSKATWNELGKVVQLRKHFGNNTIIRFANSSGGDLEVTHSFNCGEFFYCNTSGLF
NSTWISN-----NDSITLPCRIKQIINMWQRIGQAMYAPPIQGVIRCNSNITGL
ILTRDGGSTNSTTETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTRCKRRV-----
----GFLGAAGSTMGAASMTLTQARNLLSG-----TVWVGKQL
QARVLAVERYLRDQQQLLGIVGCGKLICTNVPWNSSWSNRNLSEIWDNMTWLQWDKEIS
NYTQIIYGLLEESQNQQEKNEQDLLALD

```

>BG505_full_sequence
structureX:pm.pdb:1      :A:1884  :F:::-1.00:-1.00

AENLWVTVYYGVPWKDAETTLFCASDAKAYETEKHNVWATHACVPTDPNPQEIHLENVTEEFNMWKNNMVEQMH
TDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCSFNMTTELRDKKQKVYSLFYRLDVQINENQ
GNRSNSNKEYRLINCNTSAITQACPVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCPSVSTVQCTHGIKPVV
STQLLLNGSLAEEEVMIRSENITNNAKNILVQFNTPVQINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAH
NVSKATWNETLGKVVQQLRKHFGNNTIIRFANSSGGDEVTTHSFNCGEFFYCNTSGLFNSTWIANTSQGSNS
TGSNDISITLPCRIKQIINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDGGSTNSTTETFRPGGGDMRDNWRE
LYKYKVVKIEPLGVAPTRCKRRVVG/AVGIGAVFLGFLGAAGSTMGAASMTLTVQARNLLSGIVQQQSNLLRAPE
AQQHLLKLTWVGIKQQLQARVLVERYLRDQQLGIWGCSKGKLICTNVPWNSSWSNRNLSEIWDNMTWLQWDKEI
SNYTQIIYGLLEESQNQQEKNEQDLLALD/AENLWVTVYYGVPWKDAETTLFCASDAKAYETEKHNVWATHACV
PTDPNPQEIHLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCS
FNMTTELRDKKQKVYSLFYRLDVQINENQGNRSNSNKEYRLINCNTSAITQACPVSFEPIPIHYCAPAGFAI
LKCKDKKFNGTGPCPSVSTVQCTHGIKPVVSTQLLLNGSLAEEEVMIRSENITNNAKNILVQFNTPVQINCTRPN
NNTRKSIRIGPGQAFYATGDIIGDIRQAHCVSKATWNETLGKVVQQLRKHFGNNTIIRFANSSGGDEVTTHSF
NCGEFFYCNTSGLFNSTWIANTSQGSNSTGSNDISITLPCRIKQIINMWQRIGQAMYAPPIQGVIRCVSNITGL
ILTRDGGSTNSTTETFRPGGGDMRDNWRESELYKYKVVKIEPLGVAPTRCKRRVVG/AVGIGAVFLGFLGAAGSTM
GAASMTLTVQARNLLSGIVQQQSNLLRAPEAQHLLKLTWVGIKQQLQARVLVERYLRDQQLGIWGCSKGKLIIC
TNVPWNSSWSNRNLSEIWDNMTWLQWDKEISNYTQIIYGLLEESQNQQEKNEQDLLALD/AENLWVTVYYGVPW
KDAETTLFCASDAKAYETEKHNVWATHACVPTDPNPQEIHLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCV
KLTPLCVTLQCTNVTNNITDDMRGELKNCSFNMTTELRDKKQKVYSLFYRLDVQINENQGNRSNSNKEYRLIN
CNTSAITQACPVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCPSVSTVQCTHGIKPVVSTQLLLNGSLAEEEV
MIRSENITNNAKNILVQFNTPVQINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCVSKATWNETLGKVV
KQLRKHFGNNTIIRFANSSGGDEVTTHSFNCGEFFYCNTSGLFNSTWIANTSQGSNSTGSNDISITLPCRIKQ
IINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDGGSTNSTTETFRPGGGDMRDNWRESELYKYKVVKIEPLGVA
PTRCKRRVVG/AVGIGAVFLGFLGAAGSTMGAASMTLTVQARNLLSGIVQQQSNLLRAPEAQHLLKLTWVGIKQ
LQARVLVERYLRDQQLGIWGCSKGKLICTNVPWNSSWSNRNLSEIWDNMTWLQWDKEISNYTQIIYGLLEESQ
NQQEKNEQDLLALD*

```

```

>4zmj_template

structureX:6chains1seq.pdb:1      :A:+1719:F:::-1.00:-1.00

---LWVTVYYGVPWKDAETTLFCASDAKAYETEKHNVWATHACVPTDPNPQEIHLENVTEEFNMWKNNMVEQMH
TDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCSFNMTTELRDKKQKVYSLFYRLDVQIN---
-----SNKEYRLINCNTSAITQACPVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCPSVSTVQCTHGIKPVV
STQLLLNGSLAEEEVMIRSENITNNAKNILVQFNTPVQINCTRPNNNTRKSIRIGPGQAFYATDGIIGDIRQAHC
NVSKATWNETLGKVVQKQLRKHFGNNTIIRFANSSGGDEVTTHSFNCGEFFYCNTSGLFNSTWISN-----
---NDSITLPCRIKQIINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDGGSNSTTETFRPGGGDMRDNRSE
LYKYKVVKIEPLGVAPTRCKRRV--/-----GFLGAAGSTMGAASMTLTVQARNLLSG-----
-----TVWGIKQLQARVLAVERYLRDQQLLGIWGCSGKLI CCTNVPWNSSWSNRNLSEIWDNMTWLQWDKEI
SNYTQIIYGLLEESQNQQEKNEQDLLALD/---LWVTVYYGVPWKDAETTLFCASDAKAYETEKHNVWATHACV
PTDPNPQEIHLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCS
FNMTTELRDKKQKVYSLFYRLDVQIN-----SNKEYRLINCNTSAITQACPVSFEPIPIHYCAPAGFAI
LKCKDKKFNGTGPCPSVSTVQCTHGIKPVVSTQLLNGSLAEEEVMIRSENITNNAKNILVQFNTPVQINCTRPN
NNTRKSIRIGPGQAFYATDGIIGDIRQAHCNVSKATWNETLGKVVQKQLRKHFGNNTIIRFANSSGGDEVTTHSF
NCGEFFYCNTSGLFNSTWISN-----NDSITLPCRIKQIINMWQRIGQAMYAPPIQGVIRCVSNITGL
ILTRDGGSNSTTETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTRCKRRV--/-----GFLGAAGSTM
GAASMTLTVQARNLLSG-----TVWGIKQLQARVLAVERYLRDQQLLGIWGCSGKLI CC
TNVPWNSSWSNRNLSEIWDNMTWLQWDKEISNYTQIIYGLLEESQNQQEKNEQDLLALD/---LWVTVYYGVPW
KDAETTLFCASDAKAYETEKHNVWATHACVPTDPNPQEIHLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCV
KLTPLCVTLQCTNVTNNITDDMRGELKNCSFNMTTELRDKKQKVYSLFYRLDVQIN-----SNKEYRLIN
CNTSAITQACPVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCPSVSTVQCTHGIKPVVSTQLLNGSLAEEEV
MIRSENITNNAKNILVQFNTPVQINCTRPNNNTRKSIRIGPGQAFYATDGIIGDIRQAHCNVSKATWNETLGKVV
KQLRKHFGNNTIIRFANSSGGDEVTTHSFNCGEFFYCNTSGLFNSTWISN-----NDSITLPCRIKQ
IINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDGGSNSTTETFRPGGGDMRDNRSELYKYKVVKIEPLGVA
PTRCKRRV--/-----GFLGAAGSTMGAASMTLTVQARNLLSG-----TVWGIKQ
LQARVLAVERYLRDQQLLGIWGCSGKLI CCTNVPWNSSWSNRNLSEIWDNMTWLQWDKEISNYTQIIYGLLEESQ
NQQEKNEQDLLALD*

```

Advanced Modeling Options

- Model implicit ligand binding
- Model glycosylation
- Option 1: Add flexible sugars to a semi-rigid protein that lacks glycosylation

Number of models

Glycosylation input file No file selected.

Flexible residues at glycosylation sites

Number of optimization steps

- Option 2: Sample multiple protein conformations with rigid sugars (see Sampling Options below)
- Add restraints between specific atoms
- Alter residue contact energies

Sampling Options

Generate a landscape that will be thermodynamically well sampled on the user's computer (using MODELLER)

Sample most probable conformations consistent with input crystal structure(s)

Number of simulations

MD temperature

Sample intermediate probability conformations consistent with input crystal structure(s)

Sample low probability conformations consistent with input crystal structure(s)

Sampling Options

Generate a landscape that will be thermodynamically well sampled on the user's computer (using MODELLER)

Sample most probable conformations consistent with input crystal structure(s)

Sample intermediate probability conformations consistent with input crystal structure(s)

Sample low probability conformations consistent with input crystal structure(s)

Number of simulations

MD temperature scanned or fixed value

Increase chain rigidity to maintain secondary structure at high temperature

Include chemical frustration (destabilize buried charged residues)

Z-score of the residue charge density, residues above this value cause chem. frust