Appendix A: The method in detail used to build Rhodopsin

This appendix contains a step-by-step explanation of how the validation structure for Rhodopsin was built. There were to be three final structures all built with slightly different parameters, and only at the last step was one of the three chosen as the best structure. This was done to test new methods and validate old ones as well.

1.0 TransMembrane Prediction

1.1 Setting up TMPred

Starting with the file: rhodopsin.fta, I start MembStruk and press button 1 on the MembStruk window in order to access the TM prediction GUI. I then press button 1 on the TMPred window and enter in textbox i of the Main TM window the following: /ul/sehall/ManualEx/rhod1/ and press the OK button (1).

Now I enter into textbox ii “Rhod1” and press the OK button (2). The next step is to press the TM predictions button (3).

1.2 Running Blast and ClustalW

First in the Predicting TM Regions window button 1 was pressed and the file “rhodopsin.fta” was chosen since it contains the fasta formatted Bovine rhodopsin sequence. No new files are produced from this button. Next, button 2 in the Predicting TM Regions window was pressed and then button 1 in the Blast Search window was pressed. This took about 5 minutes to complete. The Blast Search window is now closed since we obtained enough sequences to use for TM prediction.

The next step is to press button 3 in the Predicting TM Regions window. At this point I encountered an *error* and the clustal program didn’t work but remained frozen at the menu. The error was in my fasta formatted file; I did not have a line declaring the

name of the sequence "> rhodopsin" at the beginning of the file. This was easily fixed by placing this line in the beginning of the blastseq.txt file in order to fix the file and get clustal to run. Clustalw ran for 10 minutes.

At this point, I looked at the sequence_identities.txt file to decide which sequences should be kept in order to have an even distribution of sequences with differing homologies to the original sequence. The sequence_identities.txt file contains the following:

100 rhodopsin
89 gi|32363333|sp|Q8HY69|OPSD_SMI
86 gi|129205|sp|P22328|OPSD_CHICK
85 gi|1709476|sp|P52202|OPSD_ALLM
83 gi|2499369|sp|Q90245|OPSD_AMBT
83 gi|1709479|sp|P51470|OPSD_RANC
83 gi|400680|sp|P31355|OPSD_RANPI
83 gi|3024261|sp|P56515|OPSD_BUFM
83 gi|3024302|sp|P56516|OPSD_RANT
82 gi|3914251|sp|Q90215|OPSF_ANGA
81 gi|1171916|sp|P41591|OPSD_ANOC
80 gi|266704|sp|P29403|OPSD_XENLA
82 gi|21264487|sp|P35359|OPSD_BRA
79 gi|129208|sp|P22671|OPSD_LAMJA
81 gi|3024260|sp|P56514|OPSD_BUFB
79 gi|2499377|sp|Q98980|OPSD_PETM
79 gi|2499370|sp|Q90214|OPSD_ANGA
79 gi|10720156|sp|O93459|OPSD_SCY
79 gi|10720168|sp|Q9YH00|OPSD_LIT
78 gi|3024252|sp|O13227|OPSB_CONC
78 gi|10720172|sp|Q9YH04|OPSD_DIP
79 gi|1709477|sp|P51488|OPSD_CYPC
78 gi|417421|sp|P32309|OPSD_CARAU
78 gi|10720170|sp|Q9YH02|OPSD_SPA
78 gi|10720173|sp|Q9YH05|OPSD_DIP
78
gi|10720167|sp|Q9YGZ9|OPSD_MUG
78 gi|10720171|sp|Q9YH03|OPSD_SAR
79 gi|10720155|sp|O93441|OPSD_GAL
77 gi|10720164|sp|Q9YGZ6|OPSD_LIZ
77 gi|17368857|sp|Q9DGG4|OPSD_TET
77 gi|10720165|sp|Q9YGZ7|OPSD_LIZ
77 gi|10720162|sp|Q9YGZ4|OPSD_DIC
77 gi|2499375|sp|P79863|OPSD_RAJE
78 gi|10720158|sp|Q9YGZ0|OPSD_SAR
79 gi|1171917|sp|P41590|OPSD_ASTF
77 gi|3024301|sp|O42604|OPSD_ZEUF
77 gi|10720166|sp|Q9YGZ8|OPSD_CHE
77 gi|2499373|sp|P87369|OPSD_ORYL
76 gi|10720169|sp|Q9YH01|OPSD_MUL
77 gi|10720163|sp|Q9YGZ5|OPSD_SOL
76 gi|2499374|sp|P79848|OPSD_POER
77 gi|10720159|sp|Q9YGZ1|OPSD_ATH
77 gi|10720161|sp|Q9YGZ3|OPSD_SAL
76 gi|10720160|sp|Q9YGZ2|OPSD_GOB
75 gi|10720157|sp|Q9YGY9|OPSD_ZOS
75 gi|2499372|sp|P79756|OPSD_GAMA
75 gi|548431|sp|P35403|OPSD_POMMI
75 gi|3024287|sp|P79812|OPSD_NEOS
74 gi|3024303|sp|P79898|OPSD_SARD
74 gi|3024284|sp|P79807|OPSD_MYRV
74 gi|3024283|sp|P79798|OPSD_MYRB
74 gi|3024285|sp|P79808|OPSD_NEOA
74 gi|3024304|sp|P79901|OPSD_SARM
73 gi|3024308|sp|P79914|OPSD_SARX
72 gi|3024305|sp|P79902|OPSD_SARP
71 gi|3024307|sp|P79911|OPSD_SART
71 gi|3024306|sp|P79903|OPSD_SARS
71 gi|3024286|sp|P79809|OPSD_NEOA
70 gi|45382767|ref|NP_990821.1|
72 gi|417422|sp|P32311|OPSG_CARAU
71 gi|18858777|ref|NP_571329.1|
68 gi|1709472|sp|P51471|OPSB_ANOC
68 gi|548430|sp|P35357|OPSB_GECGE
70 gi|417423|sp|P32312|OPSH_CARAU
78 gi|3024275|sp|O42268|OPSD_ICTP
67 gi|1709481|sp|P51474|OPSI_ASTF
66 gi|21263839|sp|Q9W6A5|OPSG_BRA
78 gi|3024257|sp|O42294|OPSD_ABYK
77 gi|3024279|sp|O42330|OPSD_COTI
78 gi|3024299|sp|O42451|OPSD_PROJ
78 gi|3024258|sp|O42300|OPSD_BATM
77 gi|3024276|sp|O42307|OPSD_COTB
76 gi|3024277|sp|O42327|OPSD_COMD
77 gi|3024259|sp|O42301|OPSD_BATN
76 gi|3024278|sp|O42328|OPSD_COTG
77 gi|3024282|sp|O42452|OPSD_PARK
77 gi|3024280|sp|O42427|OPSD_LIMB
76 gi|3024281|sp|O42431|OPSD_LIMP
75 gi|3024289|sp|Q90373|OPSD_COTK
62 gi|3024309|sp|P87366|OPSG_ORYL
73 gi|3024300|sp|O42466|OPSD_TAUB
49 gi|45382921|ref|NP_990848.1|
48 gi|1709473|sp|P51472|OPSB_ASTF
47 gi|417419|sp|P32310|OPSB_CARAU
48 gi|3024263|sp|P87365|OPSB_ORYL
46 gi|18858357|ref|NP_571267.1|
43 gi|1709484|sp|P51473|OPSV_XENL
43 gi|1709482|sp|P51475|OPSP_CHIC
43 gi|45382143|ref|NP_990769.1|
43 gi|41017454|sp|O57605|OPSU_MEL
42 gi|1709483|sp|P51476|OPSP_COLL
41 gi|3024251|sp|O13092|OPSB_SAIB
41 gi|6680776|ref|NP_031564.1|
44 gi|2499382|sp|Q90309|OPSU_CARA
41 gi|4502387|ref|NP_001699.1|
43 gi|18859533|ref|NP_571394.1|
40 gi|25742620|ref|NP_112277.1|
40 gi|27807027|ref|NP_776992.1|
43 gi|3024320|sp|P87368|OPSV_ORYL
39 gi|3024317|sp|O42490|OPSP_PETM
39 gi|4503965|ref|NP_000504.1|
39 gi|2499381|sp|Q95170|OPSR_CAPH
39 gi|13634105|sp|O18913|OPSR_FEL
37 gi|548434|sp|P35358|OPSG_GECGE
38 gi|3024319|sp|P87367|OPSR_ORYL
39 gi|3915786|sp|O18910|OPSG_RABI
38 gi|9910526|ref|NP_064445.1|
38 gi|16758314|ref|NP_446000.1|
36 gi|464315|sp|P34989|OPSL_CALJA
38 gi|18859311|ref|NP_571250.1|
37 gi|12644063|sp|O35478|OPSG_SCI
38 gi|11386981|sp|Q9R024|OPSG_CAV
37 gi|6679975|ref|NP_032132.1|
38 gi|417424|sp|P32313|OPSR_CARAU
38 gi|548435|sp|P22332|OPSR_ASTFA
38 gi|1171918|sp|P41592|OPSR_ANOC
36 gi|3024311|sp|O12948|OPSR_XENL
37 gi|45382135|ref|NP_990771.1|
35 gi|129213|sp|P22330|OPSG_ASTFA
35 gi|129216|sp|P22331|OPSH_ASTFA
36 gi|3024298|sp|O42266|OPSP_ICTP
43 gi|3024313|sp|O18912|OPSR_HORS
42 gi|3024297|sp|O18911|OPSG_ODOV
42 gi|3024315|sp|O18914|OPSR_CANF
32 gi|3024291|sp|O13018|OPSO_SALS

The lines that are underlined are the ones that I want to keep. I generally keep 4 sequences per 10% of sequence homology, however the distribution of diversity is up to

the user as long as it is recorded. Generally, I tend to write this information down in a file called Notes.txt. Currently the only thing in this file is:

```
Running TMPred and using 4 per 10% for diversity selection
in phase 1 of TM prediction.
```

Now the file blastseq.txt is edited to only contain the 21 underlined sequences and the original file is saved under the name: blastseq.txt-1st. Now that it is edited for diversity, we run clustalw again pressing button 3 in the Predicting TM Regions window. Now we are ready to run the TM prediction program.

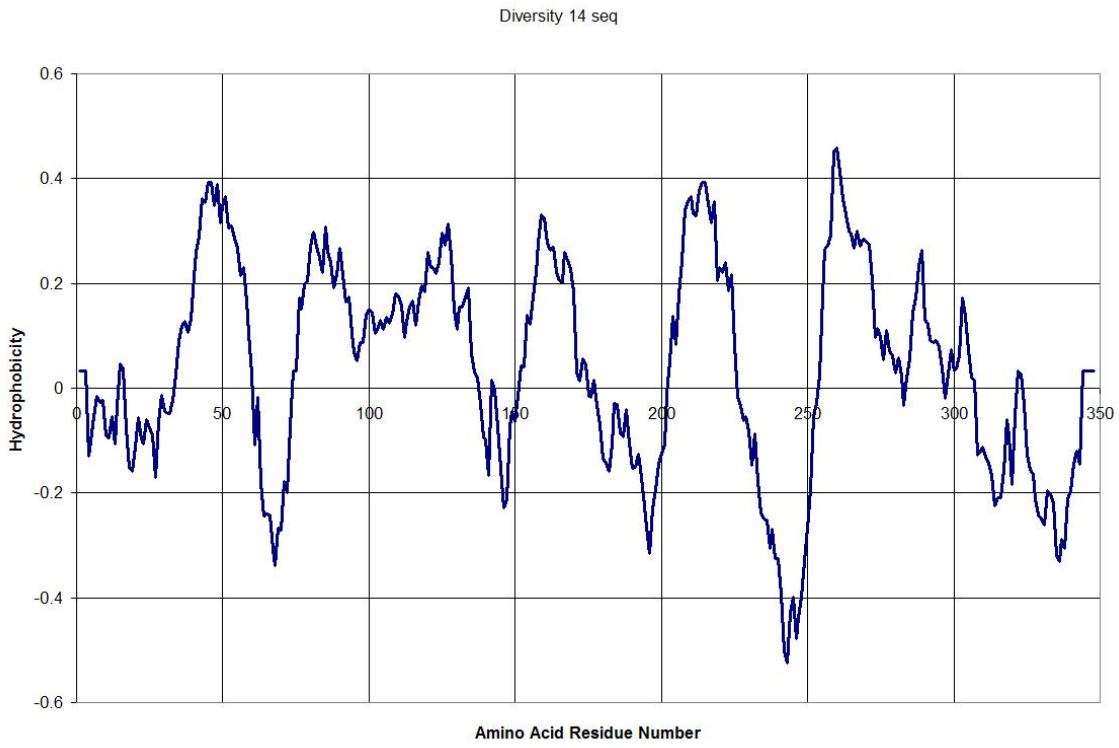
1.3 Running TM2ndS for the 1st time up to the successful Nth time

The first thing to do is press button 4 of the coarse grain prediction window. This was done several time to find a satisfactory sequence alignment. The first try the program tells me that I do not have a seven helix structure. This means that my sequences chosen in the blastseq.txt are either not diverse enough or too diverse. At this point, I will have to go back to that file and edit it.

I edit the blastseq.txt file to contain only the following sequences:

Rhodopsin
gi|32363333|sp|Q8HY69|OPSD_SMI
gi|129205|sp|P22328|OPSD_CHICK
gi|1709476|sp|P52202|OPSD_ALLM
gi|10720168|sp|Q9YH00|OPSD_LIT
gi|3024252|sp|O13227|OPSB_CONC
gi|3024304|sp|P79901|OPSD_SARM
gi|1709472|sp|P51471|OPSB_ANOC
gi|548430|sp|P35357|OPSB_GECGE
gi|1709481|sp|P51474|OPSI_ASTF
gi|45382921|ref|NP_990848.1|
gi|1709473|sp|P51472|OPSB_ASTF
gi|1709484|sp|P51473|OPSV_XENL
gi|13634105|sp|O18913|OPSR_FEL
gi|3024319|sp|P87367|OPSR_ORYL
gi|129213|sp|P22330|OPSG_ASTFA

Then I ran clustalW again pressing button 3 of the Predicting TM Regions window. After rerunning the TMPred program several times and not getting more than 6 helices in fact, the more sequences that were eliminated the less helices were found, I must add more sequences. Now it is best to plot out the data.txt file against the sequence.txt to see what we are looking at. I eliminated all gaps and got this graph:



Graph 1 – Hydrophobicity plot of the 14 sequence alignment done for rhodopsin.

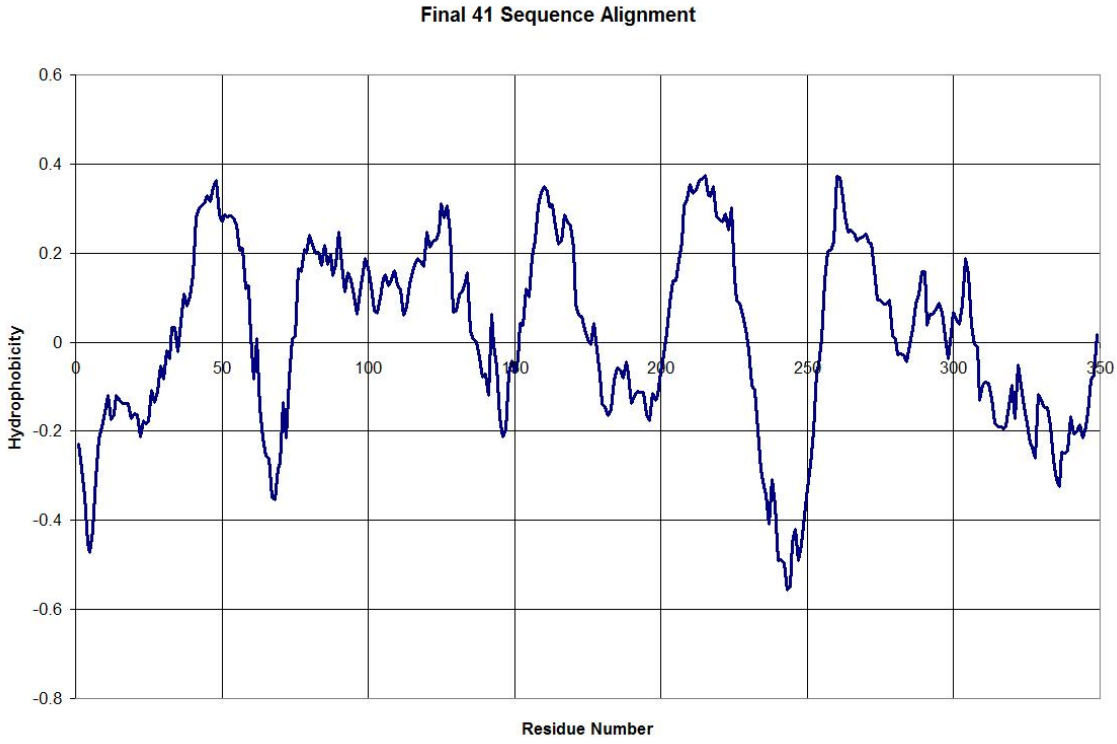
As can be seen in the Diversity 14 seq. graph we have 6 nice looking helical regions, but helix 7 (around 300) is split into two peaks and cannot be found by the program. Using the default method, I went back to the original blast and used all 125 sequences in the interactive TMPred. This doesn't yield 7 helices, so now I try several different possibilities using a script to pull out the sequences I want and placing them into another file.

The blastseq.txt file is enriched to 41 total sequences and then the interactive version of TMPred is run again with no success. However, right after that TMPred v2.0i is run and 7 helices are found. Now, I run the interactive version to get the data files and then run v2.0i again to overwrite the failed predictions file created by running the interactive version. Often the sign of a successful alignment is a large (15+) amount of gaps before and after the original sequence with not too many in between. Now that we

have all the files needed to run a fine TM prediction we are ready for the next step. Here is the final list of sequences used for the alignment:

100 rhodopsin	40 gi|25742620|ref|NP_112277.1|	38 gi|11386981|sp|Q9R024|OPSG_CAV
49 gi|45382921|ref|NP_990848.1|	40 gi|27807027|ref|NP_776992.1|	38 gi|417424|sp|P32313|OPSR_CARAU
48 gi|1709473|sp|P51472|OPSB_ASTF	43 gi|3024320|sp|P87368|OPSV_ORYL	38 gi|548435|sp|P22332|OPSR_ASTFA
47 gi|417419|sp|P32310|OPSB_CARAU	39 gi|3024317|sp|O42490|OPSP_PETM	38 gi|1171918|sp|P41592|OPSR_ANOC
48 gi|3024263|sp|P87365|OPSB_ORYL	39 gi|2499381|sp|Q95170|OPSR_CAPH	36 gi|3024311|sp|O12948|OPSR_XENL
46 gi|18858357|ref|NP_571267.1|	39 gi|13634105|sp|O18913|OPSR_FEL	37 gi|45382135|ref|NP_990771.1|
43 gi|1709484|sp|P51473|OPSV_XENL	37 gi|548434|sp|P35358|OPSG_GECGE	35 gi|129213|sp|P22330|OPSG_ASTFA
43 gi|1709482|sp|P51475|OPSP_CHIC	38 gi|3024319|sp|P87367|OPSR_ORYL	35 gi|129216|sp|P22331|OPSH_ASTFA
43 gi|45382143|ref|NP_990769.1|	39 gi|3915786|sp|O18910|OPSG_RABI	36 gi|3024298|sp|O42266|OPSP_ICTP
42 gi|1709483|sp|P51476|OPSP_COLL	38 gi|9910526|ref|NP_064445.1|	43 gi|3024313|sp|O18912|OPSR_HORS
41 gi|3024251|sp|O13092|OPSB_SAIB	38 gi|16758314|ref|NP_446000.1|	42 gi|3024297|sp|O18911|OPSG_ODOV
44 gi|2499382|sp|Q90309|OPSU_CARA	36 gi|464315|sp|P34989|OPSL_CALJA	42 gi|3024315|sp|O18914|OPSR_CANF
41 gi|4502387|ref|NP_001699.1|	38 gi|18859311|ref|NP_571250.1|	32 gi|3024291|sp|O13018|OPSO_SALS
43 gi|18859533|ref|NP_571394.1|	37 gi|12644063|sp|O35478|OPSG_SCI	

Below is the data vs. sequence graph of the successful alignment.



Graph 2 – Hydrophobic plot of the final 41 sequences that led to a successful TM prediction.

At this point our coarse grain TM prediction looks like this compared with the crystal TM regions:

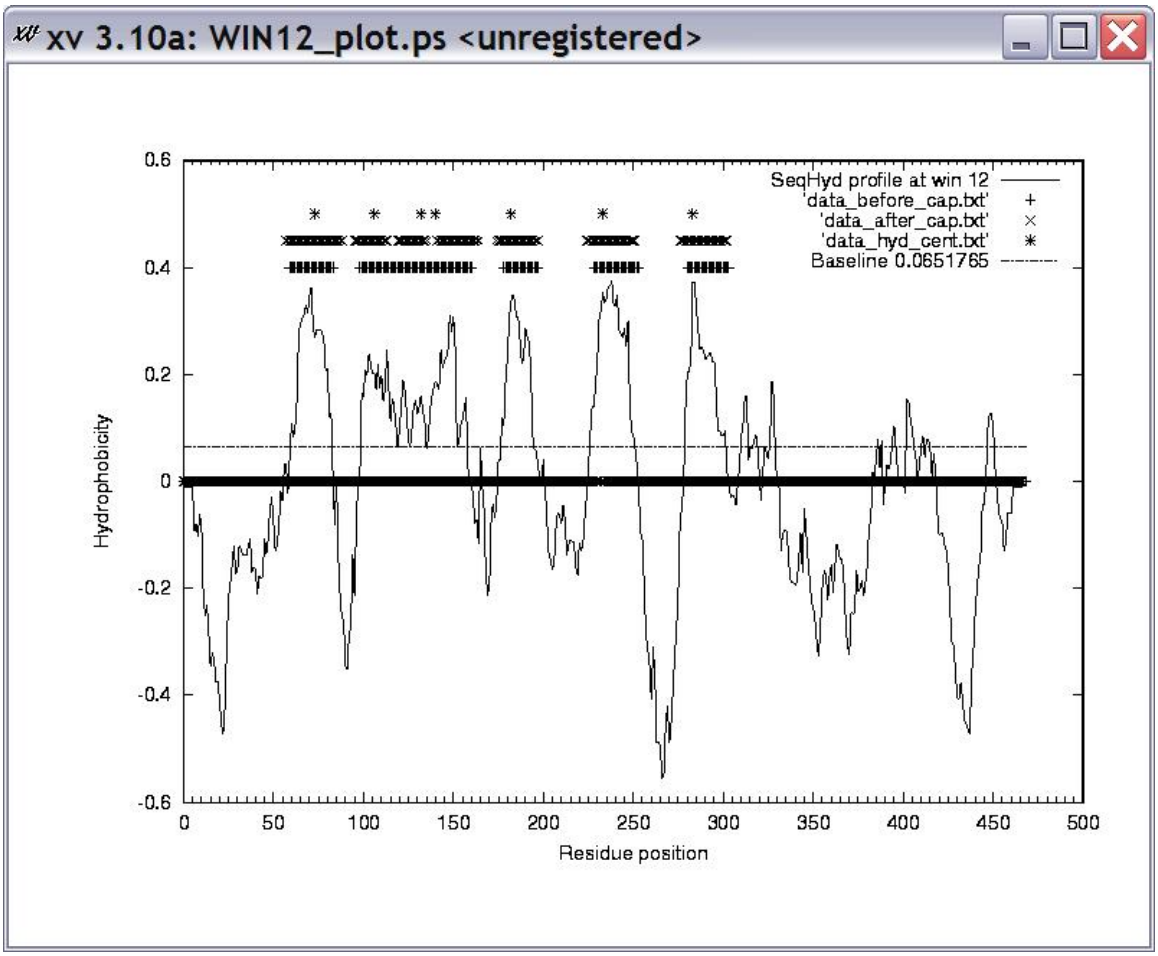
```
Crystal TM1:     WQFSMLAAYMFLLIMLGFPINFLTLYVTVQ
Pred. TM1:     PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH

Crystal TM2:    PLNYILLNLAVADLFMVFGGFTTTLYTSLH
```

```
Pred. TM2:     PLNYILLNLAVADLFMVFGGFTT

Pred. TM3:     YTSLHGYFVFGPTGCNL

Crystal TM3:   PTGCNLEGFFATLGGEIALWSLVVLAIERYVVV
Pred. TM4:             FFATLGGEIALWSLVVLAIER

Cyrstal TM4:   NHAIMGVAFTWVMALACAAPPLV
Pred. TM5:      HAIMGVAFTWVMALACAAPPLVG

Crystal TM5:   NESFVIYMFVVHFIIPLIVIFFCYGQ
Pred. TM6:      ESFVIYMFVVHFIIPLIVIFFCYGQLVF

Crystal TM6:   EKEVTRMVIIMVIAFLICWLPYAGVAFYIFT
Pred. TM7:          RMVIIMVIAFLICWLPYAGVAFYIFTH

Crystal TM7:   IFMTIPAFFAKTSAVYNPVI
```

The main problem lies with the split prediction of TM3 that eliminates helix 7. Ignoring this for the time being, the predictions are relatively good.

1.4 Fine Prediction on 1st Round

We take our previous predictions and now press button 2 of the Fine Grain Predictions Window. Now, we are done with the basic default method of running the TMPred, with the only exception of not running the iterative TMPred on the predictions. However, seeing the graph below of window 12 it should be obvious that helix 7 was not correctly identified. The next section will deal with what was done.



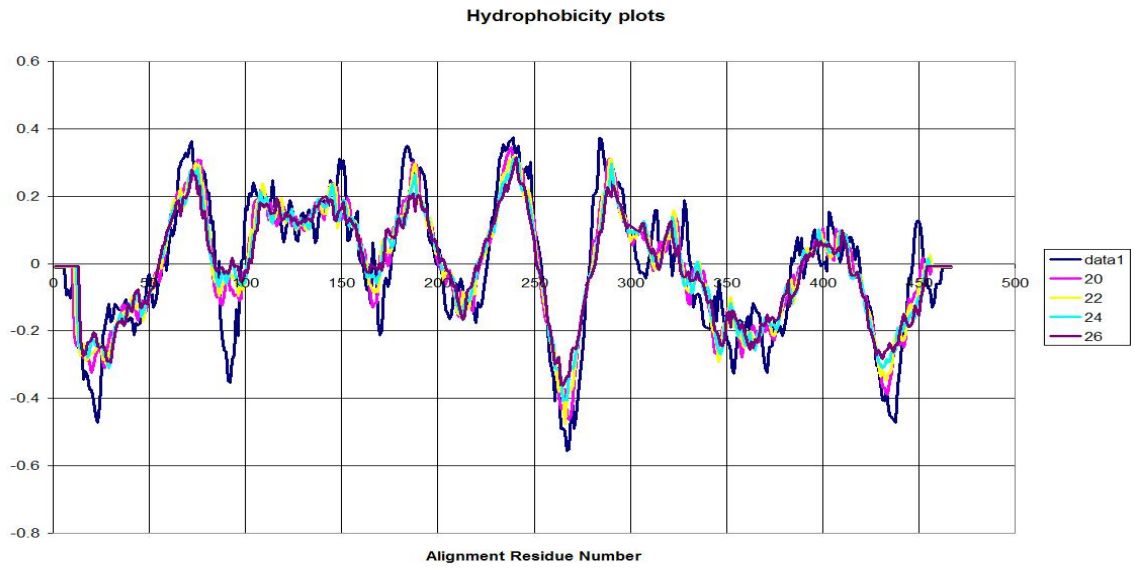
Graph 3 – Plot of Hydrophobicity after 1st Fine grain TMPred.

1.4.1 Fixing the loss of a TM region

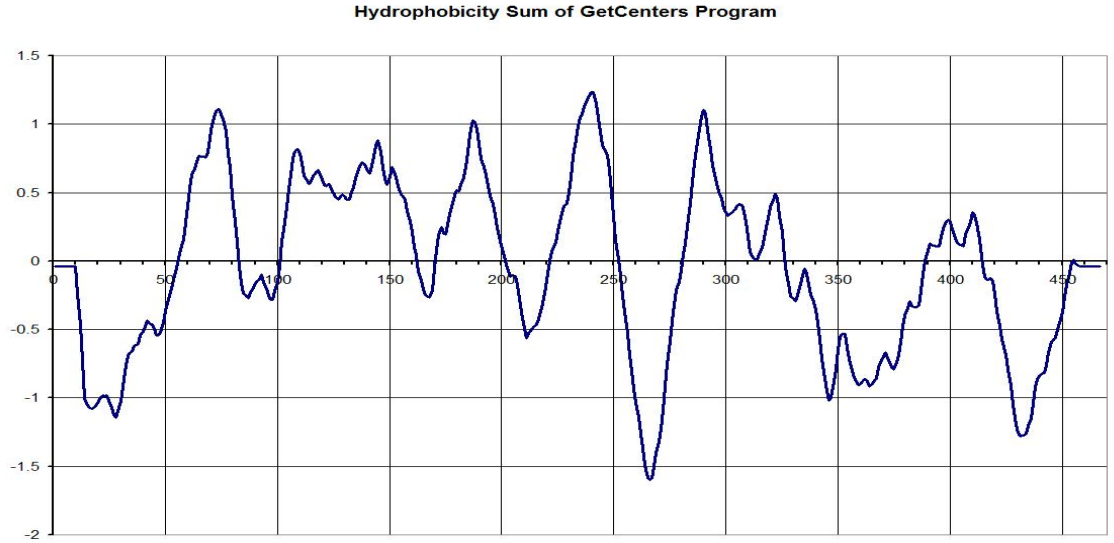
It is always a good idea to examine the graph of the data located in the sequence.txt file against in the data1.txt file. I did this in Excel, but any graphing program will do as long as you can determine the residue number according to the alignment that defines the start and end of a region. In particular, I will look closely at regions 2-4 to break them into just two regions and look at where helix 7 should be. I have also found it helpful to graph the data found in data#_nobasechange.txt where the # represents the windows (always even numbered) chosen by the get_centers program. In my case the get_centers program gave the following output:

```
…
for the data of window size 30The flags are :0000111100
15 13 20 18 18 17 18 15 16 13
19 11 11 13 14 14 13 13 18 19
4 15 2 2 17 17 16 17 3 15
8 9 6 7 4 5 4 1 1 1
10 10 11 16 13 14 14 13 12 18
15 14 13 12 14 16 16 17 16 15
8 9 11 11 13 14 14 15 16 17
The number of good windows is 4 with first and last indices of 4 7
18 17 18 15
For helix 1 the hydrophobic center index is 17
14 14 13 13
For helix 2 the hydrophobic center index is 13.5
17 17 16 17
For helix 3 the hydrophobic center index is 16.75
4 5 4 1
For helix 4 the hydrophobic center index is 3.5
13 14 14 13
For helix 5 the hydrophobic center index is 13.5
14 16 16 17
For helix 6 the hydrophobic center index is 15.75
13 14 14 15
For helix 7 the hydrophobic center index is 14
```

The line to look for is: "`The flags are :0000111100`". This tells the user that the windows 20, 22, 24, and 26 were used for the get centers program and offer the most reliable peaks in their TM regions. (The 0's are for not used and the 1's are for used, and it starts at 12 going up by 2 until it reaches 30.) So I will plot these four files together (I add up the values at every residue) with window 12 in data1.txt.



Graph 4 – Plot of all the data from the 1st Fine TM Prediction.



Graph 5 – Plot of the sum of the data contained in windows 20, 22, 24, and 26.

Now we can see from the sum plot that there is really only two hydrophobic regions from 100 to 160. We also know from looking at the sequence alignment, that after 385 there is nothing but gaps, so that last region at 380 to 415 is nothing but unimportant information. Since we need a region 7, we look to the end of region 6 (since often 6 and 7 tend to merge into one large region). This suggests that the two peaks located around 307 and 320 might be part of region 7. I would not choose the peak at

307 since it is too close to the end on region 6. Instead, I would choose the peak at 320 making the region start and end where the peak ends at 314 to 331 (taking the whole peak with no baseline since its area is small). This just leaves the problem of region 2 and region 3.

With the correcting of regions 2-4 into just regions 2 and 3, it is always best to go with the ends defined by the program. It this case, region 2 starts at 96 and region 4 ends at 165. Using this, we assume that region 2 begins at 96 and region 3 ends at 165. Now we just need to define the loop connecting the two regions. I tend to assign ends of regions at minima and let the fine TMPred program decide, so at 132 we have a significant minima for region 3. For region 2, the major dip found at 130 ends at 127 which only gives us a loop of 4 residues, so we move to the next dip at 122 which gives a loop of 9 residues and use that value. So region 2 is now 96 to 122 and region 3 is now 132 to 165.

With these new region definitions, we must edit the files that TMPred fine grain uses with the new assignments. These new files will be placed in a different directory and the Notes.txt file will reflect this. However, this program will still run automatically and will not input our choices for the regions. So we will use the interactive version of TM2ndS.

1.4.2 Running TMPred1_interactive

First I copied the temppirfile to the fixed directory and then ran the TMPred1_interactive program. I entered “N” to all questions except to:

```
“Would you like to have the option of graphing the
hydrophobic profile at each window size (the profiles will
be saves into data12_basechange.txt for window 12 with no
base modification for example; whichever profile you want
```

```
to graph needs to be renamed data.txt before using the
graph plot button)?(Pres Y or N)”
```

Since I want to see the data files for graphing and peak analysis. Also, I entered the base value of 0.0651765 (My local TM base). At the end of the window analysis, it tells me that it is not a seven helix protein, but no matter I already have the region information I want to use from the plot analysis.

1.4.3 TMPred2auto_mod

This is the program that was needed in order to get the final files from a Fine Grain prediction. The following regions were entered:

Region 1: 58 to 89
Region 2: 96 to 122
Region 3: 132 to 165
Region 4: 175 to 198
Region 5: 225 to 252
Region 6: 277 to 303
Region 7: 314 to 331
After step 7 I got the following output:

```
The sorted 0 interval is 58 89 with area 4.3366950338
The sorted 1 interval is 96 122 with area 3.0904441113
The sorted 2 interval is 132 165 with area 3.6924660828
The sorted 3 interval is 175 198 with area 4.1091795279
The sorted 4 interval is 225 252 with area 6.0229329658
The sorted 5 interval is 277 303 with area 4.2289971813
The sorted 6 interval is 314 331 with area 0.8024689609
THe break values57 88 W K
THe break values95 121 N L
THe break values131 164 G P
THe break values174 197 H W
THe break values224 251 S T
THe break values276 302 M Q
THe break values313 330 I I
```

The area values for each section is great being larger than 3, with the exception of region 7 that we know had problems before but is still 0.8 which is not bad. Our final default fine predictions are:

```
Crystal TM1:    WQFSMLAAYMFLLIMLGFPINFLTLYVTVQ
Fine TM1:      PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH
```

```
Crystal TM2:   PLNYILLNLAVADLFMVFGGFTTTLYTSLH
Fine TM2:      PLNYILLNLAVADLFMVFGGFTTTLYTSLH

Crystal TM3:   PTGCNLEGFFATLGGEIALWSLVVLAIERYVVV
Fine TM3:       TGCNLEGFFATLGGEIALWSLVVLAIERYVVVCK

Crystal TM4:   NHAIMGVAFTWVMALACAAPPLV
Fine TM4:      NHAIMGVAFTWVMALACAAPPLVG

Crystal TM5:   NESFVIYMFVVHFIIPLIVIFFCYGQ
Fine TM5:       ESFVIYMFVVHFIIPLIVIFFCYGQLVF

Crystal TM6:   EKEVTRMVIIMVIAFLICWLPYAGVAFYIFT
Fine TM6:           RMVIIMVIAFLICWLPYAGVAFYIFTH

Crystal TM7:   IFMTIPAFFAKTSAVYNPVI
Fine TM7:       FMTIPAFFAKTSAVYNPVIYIMMNK
```

1.5 Using the get_centers Program (Rhod1)

Running this program was fast since I only had to look at the contents of the predictions.txt file to get most of the information needed. The final output from get_centers was saved to the file {Prefix}-out-getcenters.txt. The program creates the file HPMCenters.txt. The output from this program is below:

```
The flags are :0011111010
15 13 20 18 18 17 18 15 16 13
20 12 12 14 15 15 14 22 19 20
18 19 16 17 14 15 14 11 10 11
10 10 11 16 13 14 14 13 12 18
15 14 13 12 14 16 16 17 16 15
8 9 11 11 13 14 14 15 16 17
17 2 15 13 12 12 11 10 9 8
The number of good windows is 5 with first and last indices of 2 6
20 18 18 17 18
For helix 1 the hydrophobic center index is 18.2
12 14 15 15 14
For helix 2 the hydrophobic center index is 14
16 17 14 15 14
For helix 3 the hydrophobic center index is 15.2
11 16 13 14 14
For helix 4 the hydrophobic center index is 13.6
13 12 14 16 16
For helix 5 the hydrophobic center index is 14.2
11 11 13 14 14
For helix 6 the hydrophobic center index is 12.6
15 13 12 12 11
```

```
For helix 7 the hydrophobic center index is 12.6
```

From this output we can see that we will want to use windows 16, 18, 20, 22, 24, and 28 in the BiPeak analysis. The centers obtained from this program are used for the structure (Rhod1).

1.6 (Test New Method - Pass) Default BiPeak Analysis Structure (Rhod2)

This new example will keep track of the progress of the BiPeak structure as it is run through the default method. First, I created a new directory called rhod2/ and placed all the data* files, sequence.txt, predictions.txt, HPMcenter.txt, and the hel? files into this new directory. I also created a new Notes.txt file to contain the details of how this structure is created. Then, I ran the BiPeakanalysis.exe program.

The files entered were:

data16_nobasechange.txt
data18_nobasechange.txt
data20_nobasechange.txt
data22_nobasechange.txt
data24_nobasechange.txt
data28_nobasechange.txt
END

Region 1: 58 to 89

```
Only 1 peak exists at residue  18 with x-intercept   18.3530326
  with a height of   1.730387
```

Region 2: 96 to 122

```
Only 1 peak exists at residue  14 with x-intercept   14.1311628
  with a height of   1.18803
```

Region 3: 132 to 165

```
The 1st peak is at residue  15 we have x-intercept   15.7454296
  with a height of   1.265608

The 2nd peak is at residue  13 we have x-intercept   13.6894375
  with a height of   1.214614
```

Region 4: 175 to 198

```
Only 1 peak exists at residue  14 with x-intercept   14.606748
  with a height of   1.576924
```

Region 5: 225 to 252

```
Only 1 peak exists at residue  14 with x-intercept   14.2975302
  with a height of   1.882872
```

Region 6: 277 to 303

```
Only 1 peak exists at residue  13 with x-intercept   13.9026367
  with a height of   1.621149
```

Region 7: 314 to 331

```
Only 1 peak exists at residue  9 with x-intercept   9.52075161
  with a height of   0.6416883
```

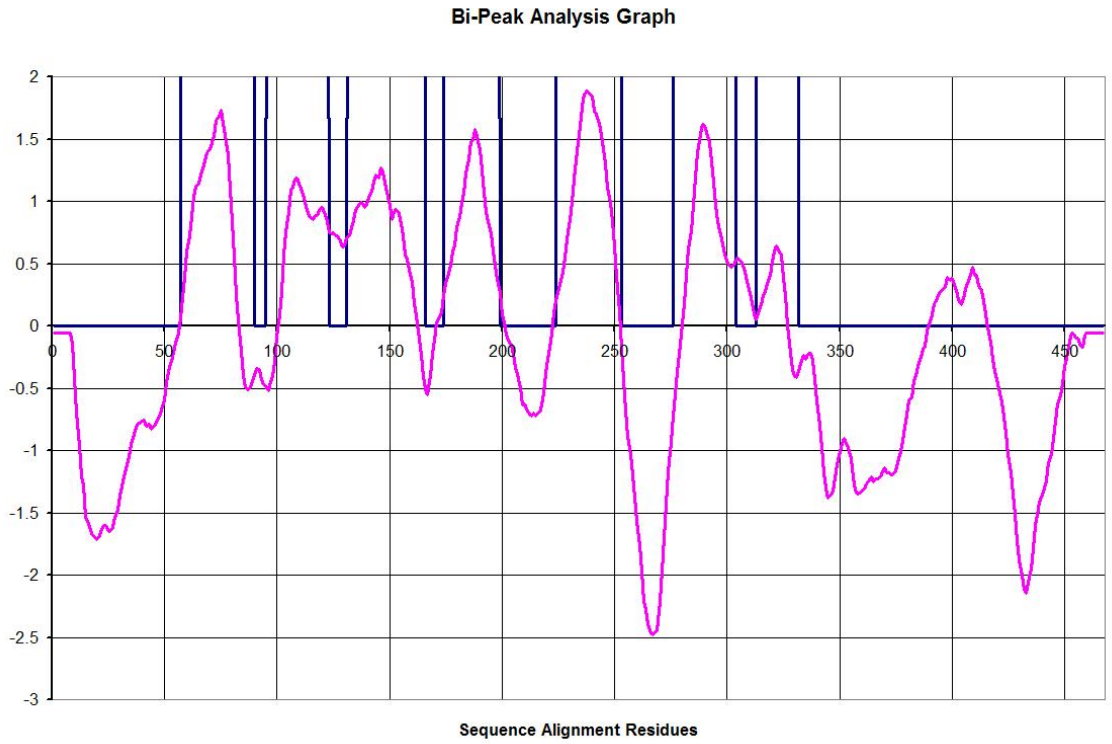
This output is saved in Rhod2-bipeak.txt.

This was easy to analyze since only helix 3 had two peaks to choose from. Now our TM prediction for helix 3 gives it 33 residues so the middle would be 16.5 and the original get_centers prediction is 15.2, so we will pick the first peak to use for translation.

So our final hydrophobic centers are:

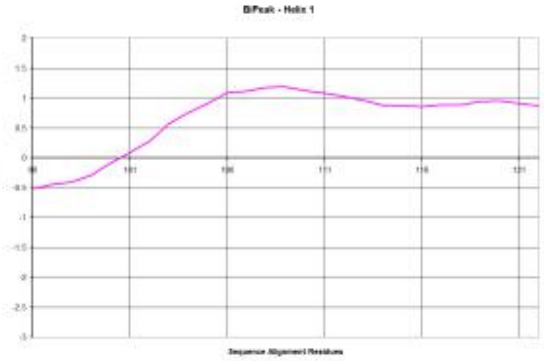
```
HELIX1  18.3530326
HELIX2  14.1311628
HELIX3  15.7454296
HELIX4  14.606748
HELIX5  14.2975302
HELIX6  13.9026367
HELIX7  9.52075161
```

Which I place in the HPMCenter.txt file. Below are the plot files that are looked at by the BiPeak program in making it's peak assignment.

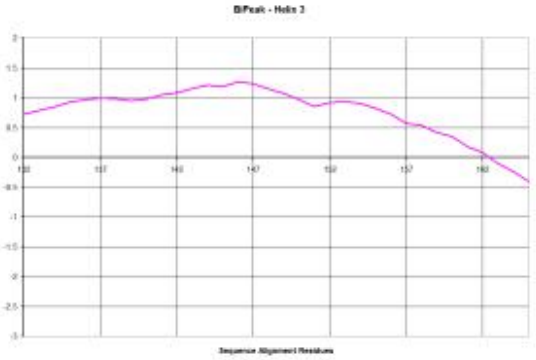


Graph 6 – The plot of hydrophobic values after BiPeak has obtained them from the data files.

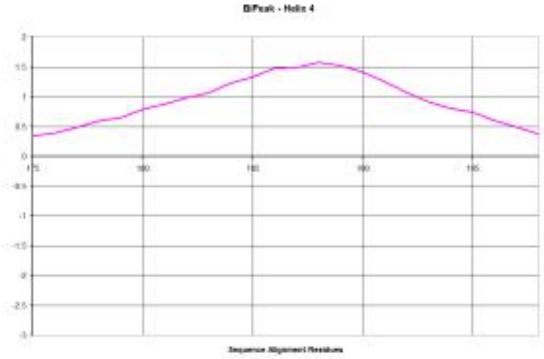
Graph 8 – Helix 1 BiPeak plot



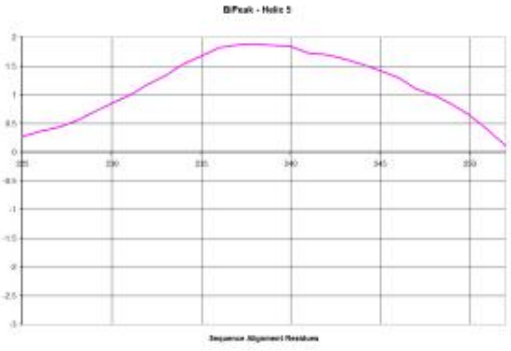
Graph 7 – Helix 2 BiPeak Plot



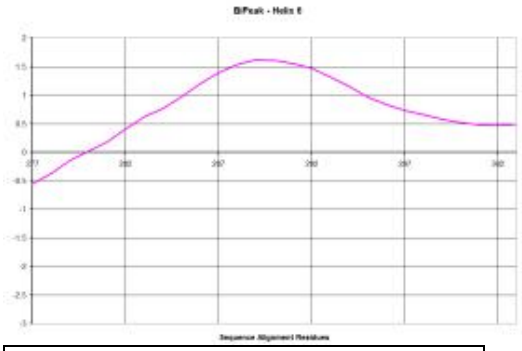
Graph 9 – Helix 3 BiPeak Plot



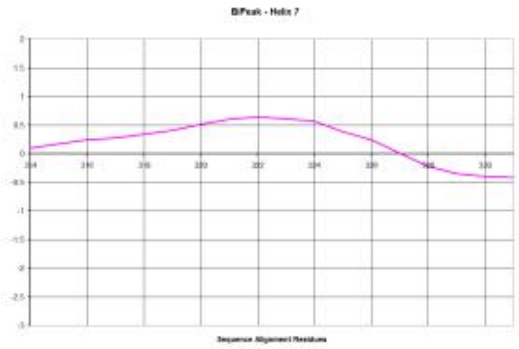
Graph 10 – Helix 4 BiPeak Plot



Graph 11 – Helix 5 BiPeak Plot



Graph 12 – Helix 6 BiPeak Plot



Graph 13 – Helix 7 BiPeak Plot

1.7 Creation of the Template

Rhod1 - Button 1 of the Create Template window was pressed and with no problems the Rhod1-Final.pdb file was generated. The template was created from the fine grain non-GUI predictions since the iterative method did not provide better results and is still in an experimental stage.

Rhod2 - Again the create template script ran with no problems and the final file (Rhod2-Final.pdb) is identical to the Rhod1-Final.pdb file produced in example 1 since the TM predictions are the same. The only difference between these two examples is the hydrophobic centers prediction, which will make a difference in the final file produced in Chapter 3 of the manual.

1.8 (Test New Method - Failed) Iterative TM Prediction

Rhod1 - Using the iterative TM prediction program from the Special Tools window, I ran it on the final fixed TM regions. Since I used the non-GUI version of TM2ndS, I had to create the output_TM_core.txt file myself. Using the rules described in step 2 of section 1.6.1 of the manual, I created the following:

```
AYMFLLIMLGFPINFLTLYV
ILLNLAVADLFMVFGGFTTT
EGFFATLGGEIALWSLVVLA
IMGVAFTWVMALACAAPPLV
IYMFVVHFIIPLIVIFFCYG
```

```
VIIMVIAFLICWLPYAGVAF
TIPAFFAKTSAVYNPVIYIM
```

Then I ran the program using the swissprot database since that is where I got the original 40 sequences that were used. I also used filtering entering “T” for every region. Below is the final TM Predictions obtained from this method:

```
Crystal TM1:          WQFSMLAAYMFLLIMLGFPINFLTLYVTVQ
Fine It.1 TM1:      EPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH

Crystal TM2:        PLNYILLNLAVADLFMVFGGFTTTLYTSLH
Fine It.1 TM2:      ---YILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTG

Crystal TM3:        PTGCNLEGFFATLGGEIALWSLVVLAIERYVVV
Fine It.1 TM3:       -------FFATLGGEIALWSLVVLAIERYV--

Crystal TM4:        NHAIMGVAFTWVMALACAAPPLV
Fine It.1 TM4:      NHAIMGVAFTWVMALACAAPPLV+

Crystal TM5:        NESFVIYMFVVHFIIPLIVIFFCYGQ
Fine It.1 TM5:       ESFVIYMFVVHFIIPLIVIFFCYGQLV+

Crystal TM6:        EKEVTRMVIIMVIAFLICWLPYAGVAFYIFT
Fine It.1 TM6:           RMVIIMVIAFLICWLPYAGVAFYIFTH

Crystal TM7:          IFMTIPAFFAKTSAVYNPVI
Fine It.1 TM7:      GP+FMTIPAFFAKTSAVYNPV-
```

Shaded residues and minuses are those that are worse than the original final prediction before using the iterative method. Plusses are where the iterative method got better results. Only TM regions 4, 5, and 7 improved and, since we wouldn’t know the correct answer we would have to take the entire result which is slightly worse than before the iterative method. So using the normal hydrophobic centers on our final predictions the iterative method did not improve the results, but does give us a reasonable alternative to the current method for considering new possibilities.

Rhod2 – It was run for Rhod2 with the following:

```
AYMFLLIMLGFPINFLTLYV
ILLNLAVADLFMVFGGFTTT
EGFFATLGGEIALWSLVVLA
MGVAFTWVMALACAAPPLVG
IYMFVVHFIIPLIVIFFCYG
IIMVIAFLICWLPYAGVAFY
FMTIPAFFAKTSAVYNPVIY
```

You will notice it has a few small changes from the one used in Example 1, since the hydrophobic centers are different. I then ran the program using the swissprot database, since that is where I got the original 40 sequences that were used. I also used filtering entering "T" for every region. Below is the final TM Predictions obtained from this method:

```
Crystal TM1:          WQFSMLAAYMFLLIMLGFPINFLTLYVTVQ
Fine It.2 TM1:      EPWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH

Crystal TM2:        PLNYILLNLAVADLFMVFGGFTTTLYTSLH
Fine It.2 TM2:      ---YILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTG

Crystal TM3:        PTGCNLEGFFATLGGEIALWSLVVLAIERYVVV
Fine It.2 TM3:      --------FFATLGGEIALWSLVVLAIERYV--

Cyrstal TM4:        NHAIMGVAFTWVMALACAAPPLV
Fine It.2 TM4:      NHAIMGVAFTWVMALACAAPPLV+

Crystal TM5:        NESFVIYMFVVHFIIPLIVIFFCYGQ
Fine It.2 TM5:       ESFVIYMFVVHFIIPLIVIFFCYGQLV+

Crystal TM6:        EKEVTRMVIIMVIAFLICWLPYAGVAFYIFT
Fine It.2 TM6:           RMVIIMVIAFLICWLPYAGVAFYIFTH

Crystal TM7:              IFMTIPAFFAKTSAVYNPVI
Fine It.2 TM7:      GSDFGP+FMTIPAFFAKTSAVYNP—
```

Regions 1, 2, 3, 4, 5, and 6 showed no difference from the predictions found in Example 1. Only region 7, showed any difference and it actually got worse. So once again the iterative method does not increase the overall TM prediction's accuracy.

2.0 TM Predictions file (Rhod1 and Rhod2)

First, a new directory was created (MembStruk) to contain the new structure files that are going to be built. Then a new Notes.txt file created for this directory, and the files hel1-7 and blastseq.txt copied from /rhod1/fixed. Now that everything was here the TM Predictions program was run into the file: Rhod1-TMs.doc. The contents of this file are below:

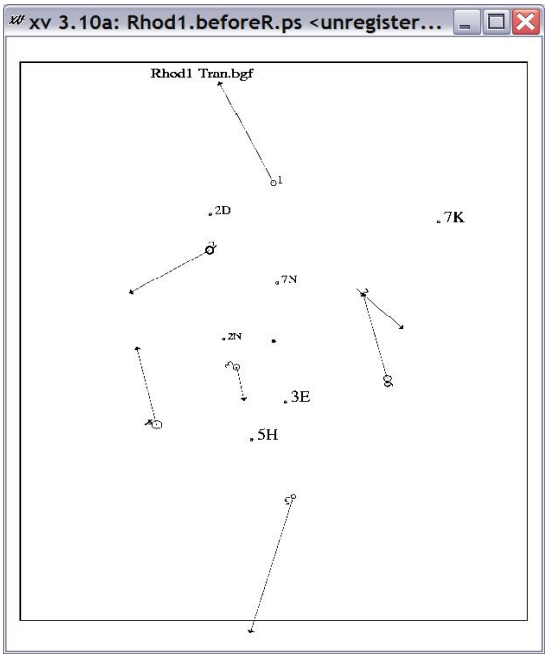
```
Listed below are the predictions used in the development
of the Transmembrane Regions utilized in TMPred and in
MembStruk

> rhodopsin

NT       1 MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAE 33 (33)

TM 1    34 PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQH 65 (32)

LP 1    66 KKLRT 70 (5)

TM 2    71 PLNYILLNLAVADLFMVFGGFTTTLYTSLH 100 (30)

LP 2   101 GYFVFGP 107 (7)

TM 3   108 TGCNLEGFFATLGGEIALWSLVVLAIERYVVVCK 141 (34)

LP 3   142 PMSNFRFGE 150 (9)

TM 4   151 NHAIMGVAFTWVMALACAAPPLVG 174 (24)

LP 4   175 WSRYIPEGMQCSCGIDYYTPHEETNN 200 (26)

TM 5   201 ESFVIYMFVVHFIIPLIVIFFCYGQLVF 228 (28)

LP 5   229 TVKEAAAQQQESATTQKAEKEVT 251 (23)

TM 6   252 RMVIIMVIAFLICWLPYAGVAFYIFTH 278 (27)

LP 6   279 QGSDFGPI 286 (8)

TM 7   287 FMTIPAFFAKTSAVYNPVIYIMMNK 311 (25)
```

```
CT      312 QFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA 348 (37)
```

From looking at the TM predictions we can see that we need to keep an eye on helix 1 and 2 for potential problems since the loop between them is small.

3.0 Translation and Rotation of the Template

Rhod1 - The template file was taken and run in the hcenterTR.script. The following were the output analysis files:

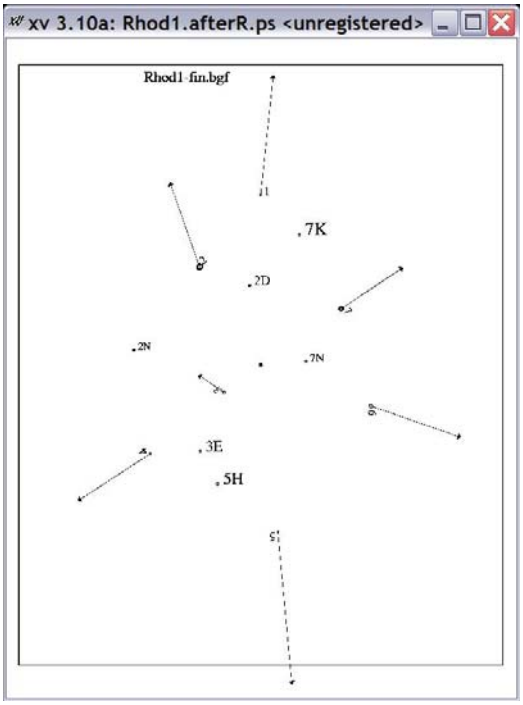


Graph 14 - Before hcenterTR Rhod1

```
Rhod1-Trans.bgf
Program Input
         Helical  Helical    HP    Mid.  Face
          Start     End     Cnt.   Res.  Deg.
         -------  -------  ------  ----  ----
HELIX  1       1      268  -18.20   15    226
HELIX  2     269      504  -14.00   15    166
HELIX  3     505      761  -15.20   15     46
HELIX  4     762      930  -13.60   15    214
HELIX  5     931     1170  -14.20   15    230
HELIX  6    1171     1394  -12.60   15    179
HELIX  7    1395     1596  -12.60   15    144

# Least Squares Plane Equation: Ax + By + Cz + D = 0
# A =    -25.4837026 B =     -6.9360587 C =      1.0000000 D =     -0.4648921
# RMS of helix CMs projected onto plane:    0.2479

                          Proj.   Proj.   Proj.   Plane   Plane   Proj.
         Helical  Plane    HPM     HPM     HPM     CM      CM      CM
          Bend    Tilt    Angle   Magn.    Fit    Dist.   Angle    Fit
         ------- ------- ------- ------- ------- ------- ------- -------
HELIX  1    9.8   32.8     25.6      9.9    0.0    14.1     0.0  -0.2136
```

```
HELIX  2   12.5   32.7     89.2      7.4    0.0     9.5    31.8   0.3038
HELIX  3    9.1   23.5     63.0      3.1    0.0     3.7   128.0  -0.2332
HELIX  4    7.3    3.1   -116.0      7.1    0.0    11.8   128.9  -0.3678
HELIX  5    6.5   24.1    -22.3     12.7    0.0    13.9   186.5  -0.1811
HELIX  6    4.6   11.7    124.6      7.9    0.0     9.6   249.7  -0.2640
HELIX  7    5.0   12.8    -79.5      5.0    0.0     8.1   305.2   0.0007
         ------- ------- ------- ------- ------- -------
RMS          8.3   22.7     83.2      8.1    0.0    10.6

RMS height of TMR:  15.9
RMS Radius of TMR:  10.6
Area of TMR Barrel:   5674.4
```

For a complete understanding of the output analysis file, see the MembComp manual. The important things to note are that (as seen in the graph), the hydrophobic moments are not pointing outward (the arrows) and the charged residues are not all contained inside the barrel (the 1 letter amino acids with a number in front for the helix where their C-alpha carbons attach). This is acceptable since this is the unrotated analysis and there should be problems.



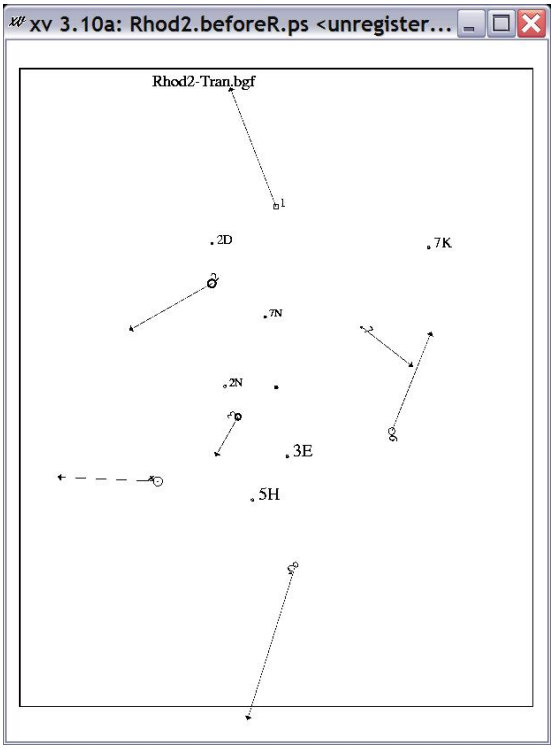
Graph 15 - After hcenterTR for Rhod1

```
Rhod1-fin.bgf
Program Input
         Helical  Helical    HP    Mid.  Face
          Start     End     Cnt.   Res.  Deg.
         -------  -------  ------  ----  ----
HELIX  1       1      268  -18.20   15    218
HELIX  2     269      504  -14.00   15    161
HELIX  3     505      761  -15.20   15     42
```

```
HELIX  4     762      930  -13.60   15    215
HELIX  5     931     1170  -14.20   15    231
HELIX  6    1171     1394  -12.60   15    177
HELIX  7    1395     1596  -12.60   15    147

# Least Squares Plane Equation: Ax + By + Cz + D = 0
# A =      4.2815970 B =     -0.7749614 C =      1.0000000 D =     -2.9495431
# RMS of helix CMs projected onto plane:    0.1405

                        Proj.   Proj.   Proj.   Plane   Plane   Proj.
         Helical  Plane   HPM     HPM     HPM     CM      CM      CM
          Bend    Tilt   Angle   Magn.    Fit    Dist.   Angle   Fit
        ------- ------- ------- ------- ------- ------- ------- -------
HELIX  1    7.8   30.3     -6.0     10.0    0.0    14.1     0.0  -0.0669
HELIX  2   10.6   36.2    -12.7      7.4    0.0     9.6    31.9   0.2246
HELIX  3   10.6   18.5    -70.6      2.6    0.0     3.8   126.9  -0.1298
HELIX  4    7.2    3.2     -5.7      7.1    0.0    11.7   128.9  -0.0795
HELIX  5    7.6   24.7     -0.6     12.6    0.0    13.9   185.9  -0.0613
HELIX  6    3.7   13.8      1.3      7.9    0.0     9.7   249.5  -0.1628
HELIX  7    5.4   17.7     -1.2      6.1    0.0     8.1   304.9   0.1726
        ------- ------- ------- ------- ------- -------
RMS         7.9   23.0     27.3      8.2    0.0    10.7

RMS height of TMR:  15.9
RMS Radius of TMR:  10.7
Area of TMR Barrel:   5667.5
```

This is an excellent final analysis of the protein. All the helical hydrophobic moments are pointing outward and all the charged residues are pointing inside the barrel of the protein. While it is not a major concern if there are problems at this stage in the structure, it is good to see that this is looking so well.

Rhod2 - The following is the output:



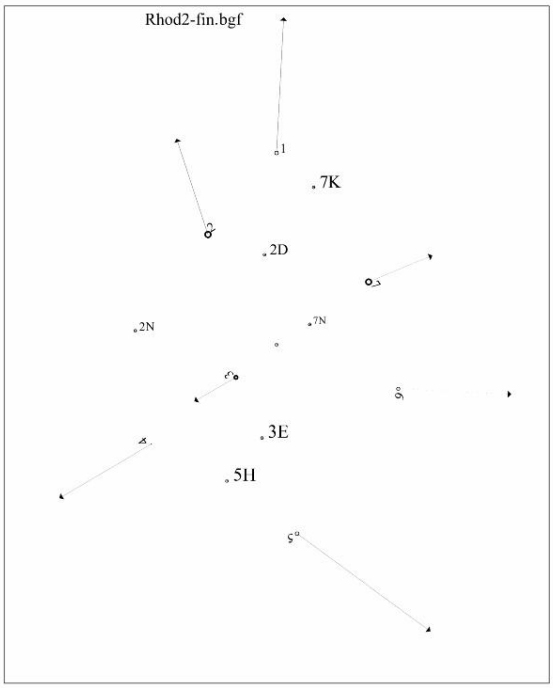
Graph 16 - Before hcenterTR for Rhod2

	Helical Bend	Plane Tilt	Proj. HPM Angle	Proj. HPM Magn.	Proj. HPM Fit	Plane CM Dist.	Plane CM Angle	Proj. CM Fit
HELIX 1	8.8	32.0	21.2	9.9	0.0	14.1	0.0	-0.2013
HELIX 2	11.7	23.3	87.7	7.4	0.0	9.5	31.8	0.3039
HELIX 3	8.4	21.9	22.5	3.5	0.0	3.8	127.8	0.2312
HELIX 4	7.1	4.7	-41.0	7.8	0.0	11.8	128.5	-0.3563
HELIX 5	5.8	23.8	-23.6	12.7	0.0	13.9	186.3	-0.1874
HELIX 6	7.8	9.6	89.0	8.3	0.0	9.7	249.3	-0.2626
HELIX 7	5.6	15.2	-73.1	5.0	0.0	8.2	305.3	0.0071
RMS	8.1	20.6	58.7	8.3	0.0	10.7		

Centered Comparision Table

	HPM Angle	HPM Magn.	Helical Bend	P. Face Deg.	Bisector Angle	
HELIX 1	10.9	2.0	2.6	219	-11.47	(2- 7)
HELIX 2	93.4	4.4	2.0	167	32.08	(1- 3)
HELIX 3	-6.8	2.8	2.7	40	-47.62	(2- 4)
HELIX 4	-63.9	3.6	2.1	213	28.54	(3- 5)
HELIX 5	-20.6	3.8	0.0	229	2.59	(4- 6)
HELIX 6	133.4	3.3	4.5	181	-3.49	(5- 7)
HELIX 7	-105.6	3.1	2.1	145	-0.63	(6- 1)
RMS	77.8	3.4	2.6		24.68	

This again shows that the file needs to be rotated more than the defaults from TM2ndS.

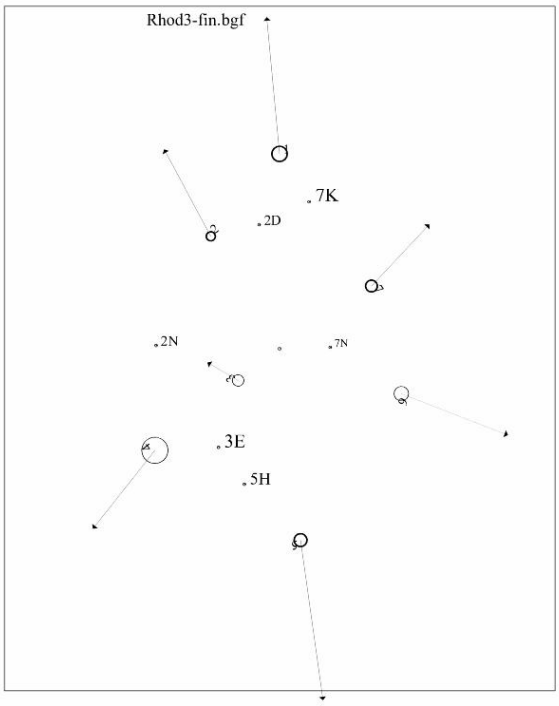


Graph 17 - After hcenterTR for Rhod2

```
# Least Squares Plane Equation: Ax + By + Cz + D = 0
# A =     21.0587805 B =     -6.9368154 C =      1.0000000 D =    -12.5336152
# RMS of helix CMs projected onto plane:    0.1490

                          Proj.    Proj.    Proj.    Plane    Plane    Proj.
         Helical  Plane     HPM      HPM      HPM      CM       CM       CM
          Bend    Tilt    Angle    Magn.    Fit     Dist.    Angle    Fit
        ------- ------- ------- ------- ------- ------- ------- -------
HELIX  1    7.3   29.2     -3.0      9.9    0.0    14.1     0.0  -0.1448
HELIX  2   11.7   36.5    -13.8      7.4    0.0     9.5    31.9   0.2053
HELIX  3    7.3   21.1     -8.9      3.5    0.0     3.8   128.8   0.1336
HELIX  4    7.2    0.0     -7.9      7.8    0.0    11.7   128.4  -0.0095
HELIX  5    7.3   24.8     47.5     12.1    0.0    13.9   186.2  -0.1384
HELIX  6    6.6   16.0     17.1      8.3    0.0     9.5   250.1  -0.1244
HELIX  7    6.2   15.0    -11.8      5.1    0.0     8.2   304.3   0.1992
        ------- ------- ------- ------- ------- -------
RMS         7.8   23.1     20.8      8.2    0.0    10.6

Centered Comparision Table
          HPM     HPM   Helical P. Face Bisector
         Angle   Magn.   Bend    Deg.    Angle
        ------- ------- ------- ------- --------
HELIX  1   -10.3   2.0      2.6    219     -11.90 ( 2- 7)
HELIX  2     5.7   4.4      2.0    162      32.27 ( 1- 4)
HELIX  3   -29.1   2.8      2.7     46      28.54 ( 4- 5)
HELIX  4   -22.9   3.6      2.1    214     -48.04 ( 2- 3)
HELIX  5     3.0   3.8      0.0    224       3.21 ( 3- 6)
HELIX  6    44.4   3.3      4.5    178      -4.86 ( 5- 7)
HELIX  7   -32.5   3.1      2.1    148       0.78 ( 6- 1)
        ------- ------- -------         --------
RMS         25.5   3.4      2.6             24.90
```

This is not as good as Rhod1, the hydrophobic moments for 5 and 6 are not exactly pointing at zero (180 degrees from the center point). However, there is the possibility of salt bridges: 2D-7K and 3E-5H.

Rhod3 - This structure is to contain the best possible structure from methods and analysis. The TM predictions are the same as examples 1 and 2 since they tend to be better on all helices except helix 7 than the predictions found in the Rhodopsin paper (Trabanino et. al. 2004). The output from hcenterTR is:



Graph 18 - After hcenterTR for Rhod3

		Helical Bend	Plane Tilt	Proj. HPM Angle	Proj. HPM Magn.	Plane HPM Fit	Plane CM Dist.	Proj. CM Angle	CM Fit
HELIX	1	10.1	37.9	5.4	9.9	0.0	14.1	0.0	0.5588
HELIX	2	7.4	38.7	-3.2	7.2	0.0	9.6	31.5	0.3254
HELIX	3	11.9	19.6	-67.6	2.6	0.0	3.8	127.4	-0.4199
HELIX	4	2.5	0.0	12.5	7.3	0.0	11.7	129.2	-0.9531
HELIX	5	8.1	14.7	1.7	11.7	0.0	14.0	186.3	0.4693
HELIX	6	7.8	15.6	-1.0	8.3	0.0	9.4	249.8	-0.5281
HELIX	7	8.4	18.9	12.6	6.1	0.0	8.1	304.2	0.4299
RMS		8.5	24.3	26.5	8.0	0.0	10.6		

```
Centered Comparision Table
          HPM     HPM   Helical P. Face Bisector
         Angle   Magn.   Bend    Deg.    Angle
        ------- ------- ------- ------- --------
HELIX  1   -16.4   2.0      2.6    222     -12.14 ( 2- 7)
HELIX  2    33.1   3.8      2.0    159      32.22 ( 1- 3)
HELIX  3   -41.6   2.1      0.0     43     -47.10 ( 2- 4)
HELIX  4    47.5   2.8      3.1    219      27.69 ( 3- 5)
HELIX  5    21.7   3.9      2.4    228       3.19 ( 4- 6)
HELIX  6    39.3   3.3      4.5    175      -4.52 ( 5- 7)
HELIX  7    -8.9   4.0      1.9    151       0.65 ( 6- 1)
        ------- ------- -------         --------
RMS         32.6   3.2      2.7             24.50
```

This is the best that I have seen a first rotation do to a structure. The hydrophobic moments are all pointing outward and there are 2 salt bridges: 2D-7K and 3E-5H.

4.0 Running Fixhelix

Rhod1 - Taking the final rotational file and running it through the fixhelix script we find no obvious problems at this moment. All the helices seem to be relatively fine with no obviously large distortions or bends in the structure. Below is the comparison of the fixhelix final structure to the crystal structure:

CRMS/MC-RMS to TM region of 1B crystal: 2.76 / 2.60
CRMS of helix 1 to 1B crystal: 1.08 ~20 degrees
CRMS of helix 2 to 1B crystal: 2.28 ~10 degrees
CRMS of helix 3 to 1B crystal: 1.08 ~10 degrees
CRMS of helix 4 to 1B crystal: 1.17 ~55 degrees
CRMS of helix 5 to 1B crystal: 1.65 ~10 degrees
CRMS of helix 6 to 1B crystal: 1.24 ~60 degrees
CRMS of helix 7 to 1B crystal: 1.65 ~10 degrees
Average 1.45

This structure was the best in terms of rotations of the 3 examples, and had the best CRMS values for the individual helices.

Rhod2 - Taking the final rotational file and running it through the fixhelix script, the only thing to watch would be helix 7, but it doesn't look like an issue from viewing the structure. The comparison to the crystal structure is below:

CRMS/MC-RMS to TM region of 1B crystal: 3.49 / 3.33

CRMS of helix 1 to 1B crystal:	2.47 ~40 degrees – little too bent
CRMS of helix 2 to 1B crystal:	2.20 ~10 degrees
CRMS of helix 3 to 1B crystal:	1.07 ~120 degrees
CRMS of helix 4 to 1B crystal:	0.61 ~5 degrees
CRMS of helix 5 to 1B crystal:	2.18 ~20 degrees
CRMS of helix 6 to 1B crystal:	2.28 ~90 degrees
CRMS of helix 7 to 1B crystal:	2.93 ~ 5 degrees
Average	1.96

This structure seems to be the worst, but it is still a fine structure at this point in the development stages. Generally, you will not be able to see problems in the structure unless they are 3 angstroms CRMS or worse to the crystal structure. So this structure's worst helix (7 at 2.93 CRMS) would be borderline if it could show up under personal scrutiny.

Rhod3 - Taking the final rotational file and running it through the fixhelix script we find no major problems. The comparison to the crystal structure is below:

CRMS/MC-RMS to TM region of 1B crystal: 3.01 / 2.83

CRMS of helix 1 to 1B crystal:	0.81 ~15 degrees
CRMS of helix 2 to 1B crystal:	2.23 ~15 degrees
CRMS of helix 3 to 1B crystal:	0.80 ~5 degrees
CRMS of helix 4 to 1B crystal:	1.18 ~45 degrees
CRMS of helix 5 to 1B crystal:	2.50 ~10 degrees
CRMS of helix 6 to 1B crystal:	1.12 ~90 degrees
CRMS of helix 7 to 1B crystal:	2.12 ~5 degrees
Average	1.54

This structure is looking fine at this stage in the game.

5.0 Rotation on the Fixhelix bundle

Rhod1 – Here, the after-rotation postscript file looks fine with a possible saltbridge between helix 2 and helix 7. The only point of concern is that the rotation of helix 3 is pointing more towards helix 2 than between helix 2 and 4. Usually helix 3 points between helix 2 and 4 or less often between 4 and 5. However, since the next step

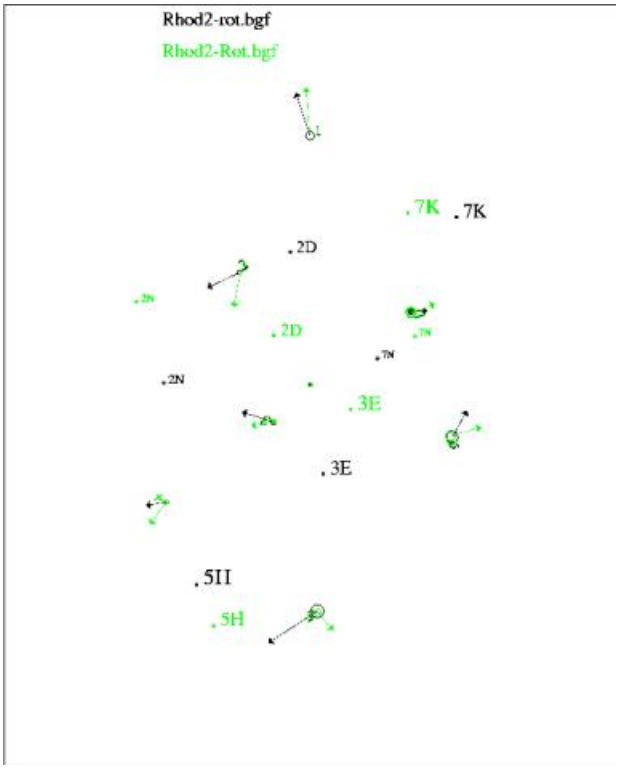
will focus on helix 3 first and rotates according to energy, this is not something that should be fixed at this time. After the next step, if helix 3 still look like it might be in the wrong position, then we can focus on helix 3 in the MembComp program and rotate it manually.

CRMS/MC-RMS to TM region of 1B crystal: 2.94 / 2.73

	Before	After
Rotation of helix 1 to 1B crystal:	~20	~20
Rotation of helix 2 to 1B crystal:	~10	~10
Rotation of helix 3 to 1B crystal:	~10	~90
Rotation of helix 4 to 1B crystal:	~55	~15
Rotation of helix 5 to 1B crystal:	~10	~10
Rotation of helix 6 to 1B crystal:	~60	~15
Rotation of helix 7 to 1B crystal:	~10	~10
Average	~25	~24

We see that the rotation has improved the overall structure, but in comparison to the crystal rotations, helix 3 got significantly worse. We were able to notice the change for the worse in helix 3 from the postscript file, and with no other knowledge would attempt to correct it in the next 2 steps.

Rhod2 - Looking at the before and after versions of the postscript files (graph 22), we can see several problems: 1) Helix 2's hydrophobic vector is pointing down instead of out from the center of the protein, though it did bring the 2D residue more inside the protein, 2) Helix 5 has had its His moved outward instead of inside, 3) and helix 7's Lys is still not inside the bundle where it would make a saltbridge with 2D. At this point, I would consider rotating helix 7 a bit, but this can also wait till after the next step.



Graph 19 - Rhod2 before (black) and after rotation (green)

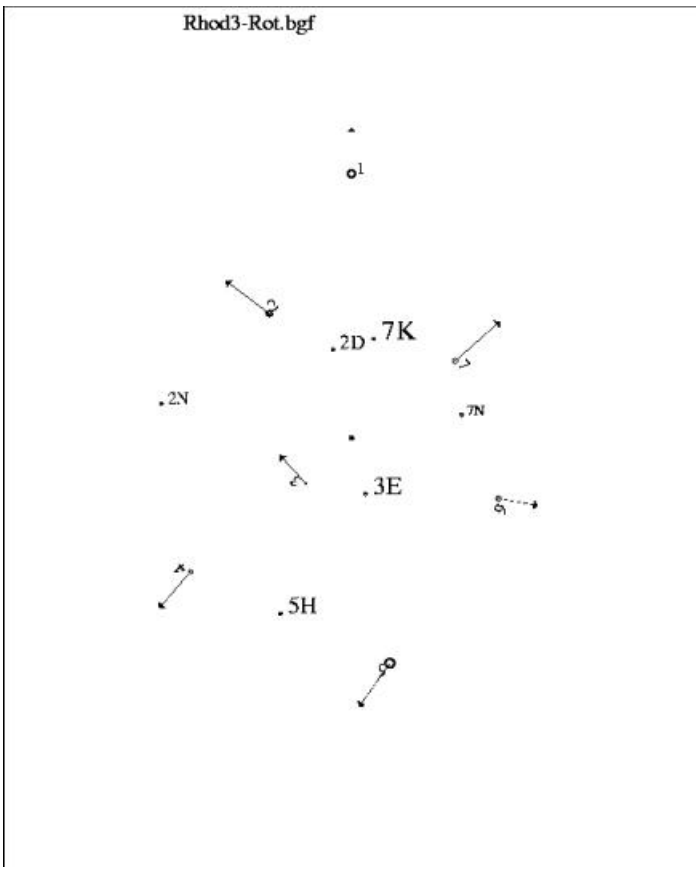
CRMS/MC-RMS to TM region of 1B crystal: 3.99 / 3.73

	Before	After
Rotation of helix 1 to 1B crystal:	~40	~30
Rotation of helix 2 to 1B crystal:	~10	~35
Rotation of helix 3 to 1B crystal:	~120	~90
Rotation of helix 4 to 1B crystal:	~5	~15
Rotation of helix 5 to 1B crystal:	~20	~75
Rotation of helix 6 to 1B crystal:	~90	~10
Rotation of helix 7 to 1B crystal:	~ 5	~15
Average	~41	~38

In the comparision of example 2 to the crystal structure, we can see that overall the structure improved. However, the two helices to be conserned about are helix 5 which we noted and helix 3 which is always a concern.

Rhod3 - In this example everything looks great. This is the kind of postscript file that you want to see after rotating the fixhelix bundle (see graph 23). Every helix is pointing away from the center of the protein and we have all but one polar residue inside

the bundle. The 2N is not a concern since the other helix 2 residue (2D) is making a saltbridge with the Lys in helix 7.



Graph 20 - Rhod3 after rotation of the Fixhelix bundle

CRMS/MC-RMS to TM region of 1B crystal: 3.00 / 2.80

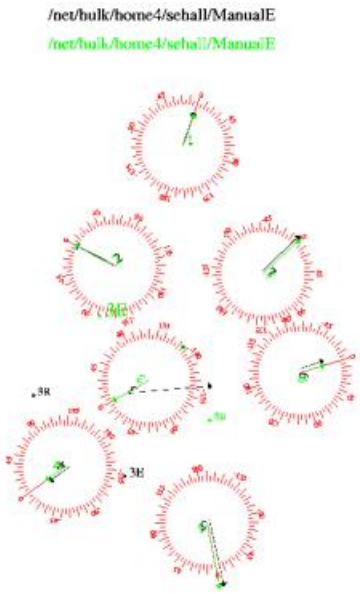
	Before	After
Rotation of helix 1 to 1B crystal:	~15	~20
Rotation of helix 2 to 1B crystal:	~15	~10
Rotation of helix 3 to 1B crystal:	~5	~30
Rotation of helix 4 to 1B crystal:	~45	~20
Rotation of helix 5 to 1B crystal:	~10	~5
Rotation of helix 6 to 1B crystal:	~90	~15
Rotation of helix 7 to 1B crystal:	~5	~15
Average	~26	~16

The comparison of this structure to the crystal structure shows that the rotation leveled the playing field, giving an average error across all helices. Helic 6 shows the

best improvement in terms of rotation. While this structure might not have the best CRMS, it does have the best overall rotation average error.

6.0 Running Rotmin

Rhod1 - Before we run rotmin for those using the beta program MembComp, we can analyze the top and bottom halves of Helix 3 to get a better idea of how it should be rotated. This analysis gives us the graph below, and from that we rotate Helix 3 by +50 degrees and name this new structure: Rhod1-H2anal.bgf.



Graph 21 - Helix 3 Analysis for Rhod1

The rotmin structure w/o H3 analysis had the following rotations: H1(15.0), H2(25.0), H3(-15.0), H4(-10.0), H5(5.0), H6(0), H7(-5.0). Comparing the two structures to the crystal using MembComp, we find that the hydrophobic centers are ~0.8 ang. rms from the crystal and the rotational differences are:

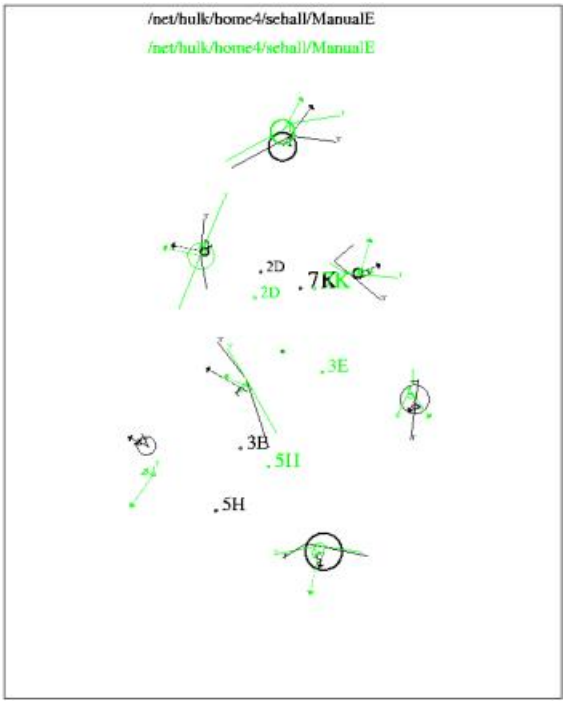
```
Hydrophobic Centers fit to plane: 0.8029
CRMS/MC-RMS to TM region of 1B crystal: 3.11 / 2.89 for No H3
CRMS/MC-RMS to TM region of 1B crystal: 2.89 / 2.74 for H3 Anal. Rotmin

             Plane of Intersection Difference (Rhod1-H3anal-rotmin)
```

Negative numbers represent the model smaller than the crystal
Negative degrees is counter-clockwise from the crystal

	Helical Bend	Plane Tilt	Proj. HPM Angle	Proj. HPM Magn.	Proj. HPM Fit	Plane CM Dist.	Plane CM Angle	Proj. CM Fit
HELIX 1	-3.0	-1.6	6.4	-0.8	-1.1	1.0	0.0	0.8557
HELIX 2	-2.6	13.2	-4.0	0.3	3.0	-0.1	2.5	-0.9525
HELIX 3	-1.3	-6.9	3.0	-1.2	1.8	-0.3	1.1	-0.0467
HELIX 4	-6.2	1.9	71.4	1.3	0.3	0.9	8.9	-0.1366
HELIX 5	-9.8	0.3	5.0	2.2	-1.6	-0.3	-1.1	-0.3870
HELIX 6	-15.1	-6.4	57.5	1.6	-0.1	-0.6	3.4	-0.0833
HELIX 7	-11.0	0.1	53.7	0.8	0.3	0.2	-0.6	0.1901
AVERAGE	7.0	4.3	28.7	1.2	1.2	0.5	2.5	0.3788
No H3 Ave.	7.7	3.2	56.0	2.0	1.1	0.4	2.9	0.4278

Comparision Plane Difference

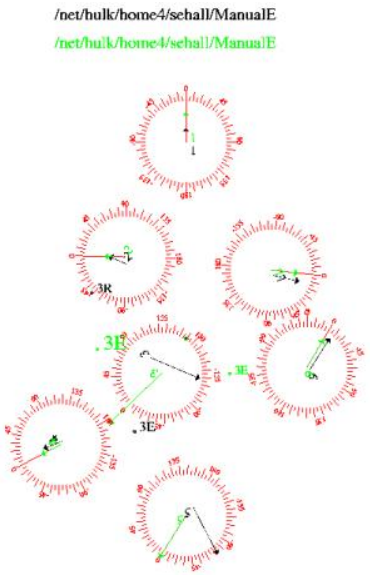
	HPM Angle	HPM Magn.	Helical Bend	P. Face Deg.	Bisector Angle
HELIX 1	-0.3	-1.2	-1.5	19	0.96 (2-7)
HELIX 2	-8.7	-1.9	-6.5	20	2.01 (1-4)
HELIX 3	29.1	0.3	1.0	6	2.86 (4-5)
HELIX 4	64.4	0.1	-13.9	-5	-7.16 (2-3)
HELIX 5	-47.0	-1.2	8.7	-1	3.28 (3-6)
HELIX 6	35.6	-0.2	-15.5	-7	-4.17 (5-7)
HELIX 7	56.6	0.8	-8.8	0	2.23 (6-1)
AVERAGE	34.5	0.8	8.0	8	
No H3 Ave.	53.9	1.0	8.8	10	



Graph 22 - Comparison of 1B Crystal (Black) to Example 1 (Green) after H3 Analysis and Rotmin

It can easily be seen that the H3 analysis helped this structure in getting the rotations almost 20 degrees better on average, with only helix 3's Glutamic acid not being in the general position.

Rhod2 - This structure was analyzed in MembComp (see example 3, or the MembComp manual) and had it's helix 3 rotated +60.0 degrees (from analysis of the graph below) and was saved in the H3anal directory under the name: Rhod2-H3anal.bgf. This example was run under rotmin before and after the H3 analysis rotation.

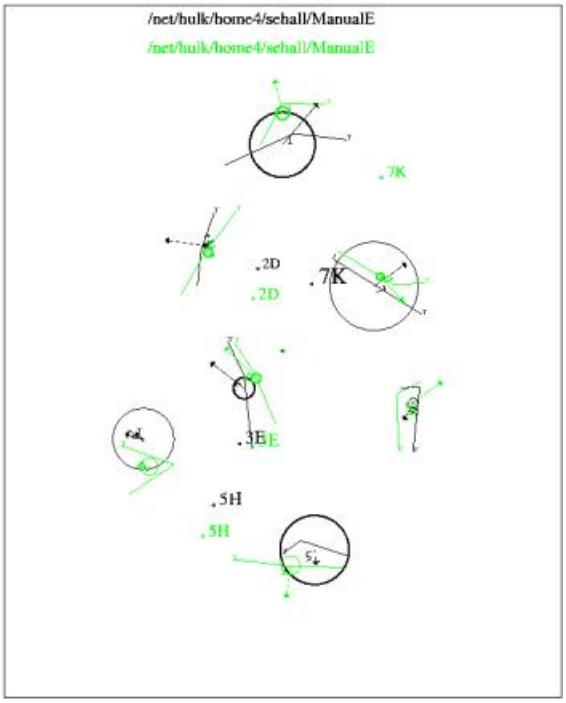


Graph 23 - Analysis of Helix 3 for Rhod2

We now compare the two structures to the crystal structure using MembComp. The first thing that we note is that this structure's hydrophobic centers are a fit of only ~1.9 ang. to the crystal (however this drops to ~1.3 ang. if we ignore helix 7 whose center at residue 9 is generally to low). ***This shows that the bipeak analysis must have a more detailed search of the windows to use for each individual helix than the ones used for***

get_centers to produce better centers than the current method. Below is the comparision to the crystal of the helical rotations to both structures:

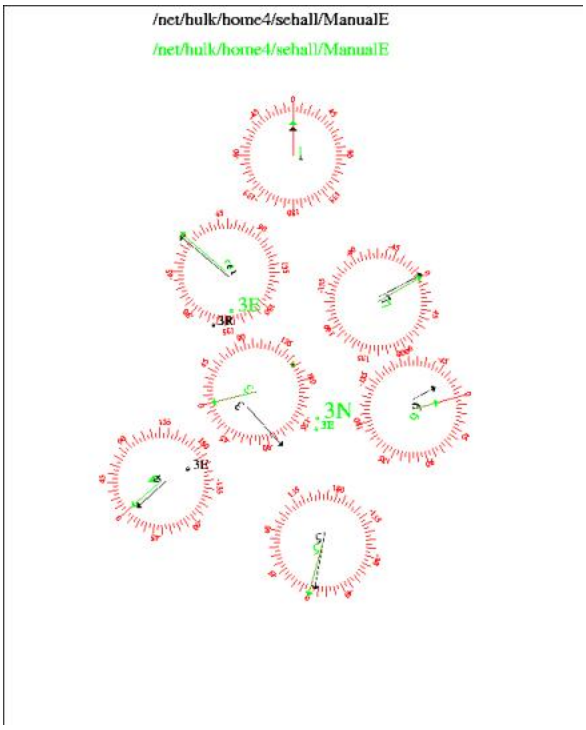
```
HPM centers fit to plane: 1.9768 (1.3 using 12.6 for Helix 7)
CRMS/MC-RMS to TM region of 1B crystal: 3.87 / 3.67 for No H3 rotmin
CRMS/MC-RMS to TM region of 1B crystal: 3.52 / 3.44 for H3 Anal rotmin

                    Plane of Intersection Difference (H3-anal-rotmin)
Negative numbers represent the model smaller than the crystal
Negative degrees is counter-clockwise from the crystal
                         Proj.   Proj.   Proj.   Plane   Plane   Proj.
         Helical  Plane    HPM     HPM     HPM     CM      CM      CM
          Bend    Tilt    Angle   Magn.    Fit    Dist.   Angle    Fit
         ------- ------- ------- ------- ------- ------- ------- -------
HELIX  1   11.4    -8.7    53.5   -1.4      1.5     2.2     0.0   0.4708
HELIX  2   -0.6     6.2  -201.3   -2.4      1.2    -0.4     1.0   0.3079
HELIX  3   -5.5    -4.8    -8.7    0.2      0.1    -1.1     1.5   0.3378
HELIX  4   23.6    17.0    30.0   -0.9     -0.8     0.9     8.6  -0.6331
HELIX  5  -12.9     8.3    -8.2    1.3      1.1     0.9    -6.9  -0.6840
HELIX  6   -8.3    -3.0   153.3    1.4     -0.8    -0.2    -0.9  -0.2897
HELIX  7  -15.1   -18.6   -84.7   -0.4     -0.8     0.7     2.0   0.2619
         ------- ------- ------- ------- ------- ------- ------- -------
AVERAGE    11.1     9.5    77.1    1.1      0.9     0.9     3.0   0.4265
No H3 Ave.  8.0     6.8    60.5    1.0      1.0     0.8     2.7   0.2228

               Comparision Plane Difference
           HPM     HPM   Helical P. Face Bisector
          Angle   Magn.    Bend    Deg.    Angle
         ------- ------- ------- ------- --------
HELIX  1   44.9   -1.4      5.6     20      1.47 (2-7)
HELIX  2  -94.9   -1.5     -7.5      7      3.28 (1-4)
HELIX  3  -30.8    0.1     -6.5     10     -0.68 (4-5)
HELIX  4  -31.1   -1.0     19.1    -27     -7.31 (2-3)
HELIX  5   -8.8   -0.4      8.5    -10       7.2 (3-6)
HELIX  6  115.3    1.4    -20.2     -5     -1.58 (5-7)
HELIX  7  -95.2    0.6    -14.9      8     -2.37 (6-1)
         ------- ------- ------- -------
AVERAGE    60.1    0.9     11.8     12
No H3 ave. 62.1    1.0      8.3      8
```



Graph 24 - Comparison of 1B crystal (black) to Example 2 after H3 anal. and rotmin

The comparison to the crystal structure shows us that the helix 3 analysis was needed and improved the overall structure. If you just look at rotations alone on the plane of intersection, the "No H3" structure looks better. However, with similar structures the Comparison table is more reliable. As you can see from Graph 26, only helix 7's Lys. is considerably out of position.

Rhod3 - This structure was run with rotmin and it also was rotated according to the Helix3 Analysis method and then run in rotmin again. The structure without any Helix3 analysis had the following rotations: H1 (-30.0), H2 (-5.0), H3 (-25.0), H4 (-5.0), H5 (-15.0), H6 (0), H7 (-5.0). However we also use MembComp to analyze the hydrophobic moment of the top and bottom of the helix. *This method is explained in detail in the MembComp Manual.* Using this method, we produce the following graph:



Graph 25 - Analysis of Helix 3 for Rhod3

Helix 3 being in the middle of the protein has two distinct hydrophobic areas, one side on the top half points between helices 2 and 4, while the opposite bottom half points between helices 4 and 5. Often these two vectors are ~90 degrees of each other when plotted on the plane of intersection. From the graph above, we can see that the lower half of helix 3 is pointing towards helix 5 and should be moved to be between 5 and 4. The more reliable indicator (*this is based on experience with other models*) tends to be the top half pointing between 2 and 4, so we should rotate helix 3 by +25 degrees. We call this structure: Rhod3-H3analy.bgf. Now we run this structure through the rotmin script and compare it to the crystal structure using the MembComp program. This comparision shows us that the hydrophobic centers are ~0.5 ang. rms from the crystal structure and the rotational differences are:

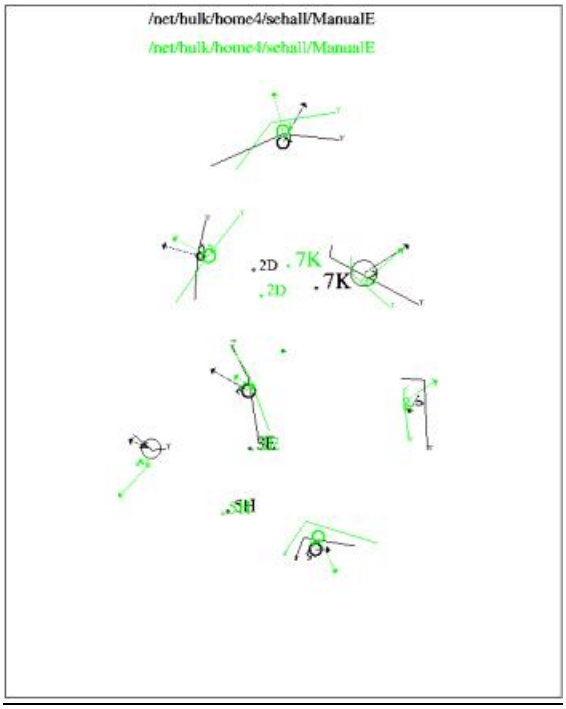
```
Hydrophobic Centers fit to plane: 0.5393
CRMS/MC-RMS to TM region of 1B crystal: 3.16 / 2.91 for No H3
CRMS/MC-RMS to TM region of 1B crystal: 2.98 / 2.82 for H3 Anal. Rotmin
```

Plane of Intersection Difference (H3 Analysis Rotmin)
Negative numbers represent the model smaller than the crystal
Negative degrees is counter-clockwise from the crystal

	Helical Bend	Plane Tilt	Proj. HPM Angle	Proj. HPM Magn.	Proj. HPM Fit	Plane CM Dist.	Plane CM Angle	Proj. CM Fit
HELIX 1	8.6	2.1	42.8	-0.4	0.5	0.9	0.0	0.4395
HELIX 2	-0.9	7.5	-7.8	0.0	1.7	-0.3	-3.0	0.4797
HELIX 3	-1.9	-3.9	-1.7	-1.6	1.6	-0.4	-0.3	0.1650
HELIX 4	-8.3	-3.8	64.4	1.4	-0.7	0.9	3.8	-0.1184
HELIX 5	6.5	10.7	-63.3	1.8	-2.4	-0.9	1.5	0.4062
HELIX 6	-11.6	-8.4	157.4	1.2	1.2	-0.7	-3.3	-0.1890
HELIX 7	-19.9	-15.5	11.5	-0.2	0.1	-0.4	-3.4	-0.2236
AVERAGE	8.2	7.4	49.8	0.9	1.2	0.6	2.2	0.2888
No H3 Ave.	9.2	8.2	67.0	0.9	1.0	0.6	3.1	0.2537

Comparision Plane Difference

	HPM Angle	HPM Magn.	Helical Bend	P. Face Deg.	Bisector Angle
HELIX 1	44.6	-0.4	5.5	21	-3.18 (2-7)
HELIX 2	-17.6	-0.8	1.4	22	4.9 (1-4)
HELIX 3	50.6	0.0	-3.4	9	2.95 (4-5)
HELIX 4	60.9	0.8	-16.1	2	-5.45 (2-3)
HELIX 5	-50.3	-1.5	24.3	-11	-3.28 (3-6)
HELIX 6	127.5	-0.1	-12.7	-8	2.33 (5-7)
HELIX 7	15.0	0.1	-17.7	6	1.73 (6-1)
AVERAGE	52.4	0.5	11.6	11	
No He Ave.	56.8	0.6	8.9	10	



Graph 26 - Comparison of 1B Crystal (Black) to Example 3 (Green) after H3 Analysis and Rotmin

This is by far the best structure of the three examples. If you were to judge only on CRMS you would pick Example 1, but as can plainly be seen in Graph 29, this example has the best fit to the crystal structure.

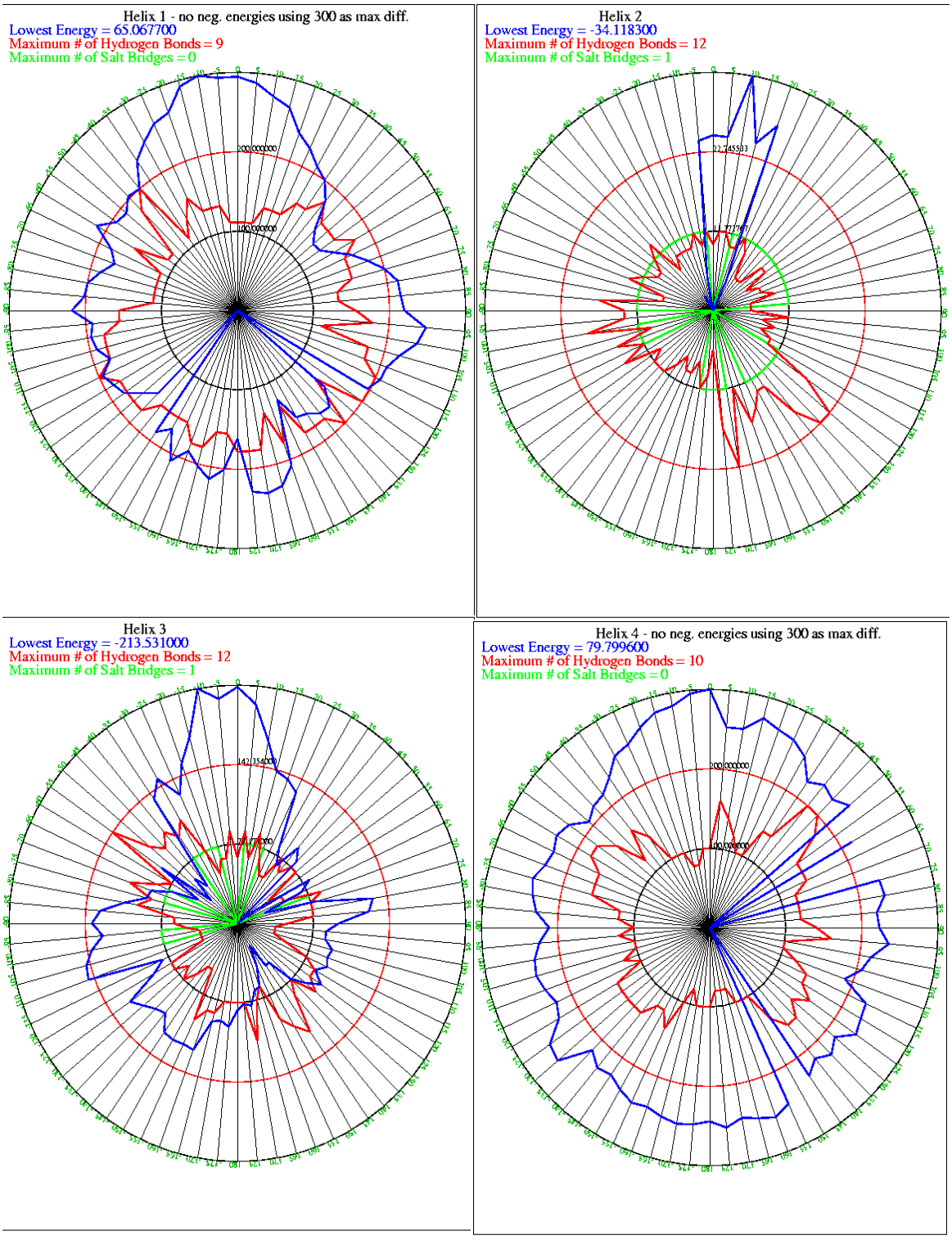
7.0 RBMD for Crystal Rhodopsin and Rhod[1-3]

All of the structures were run in RBMD before they had the helical scans done (see Section 8.0). There were no large changes in any of these examples, since RBMD is more of a fine tuning to the structures.

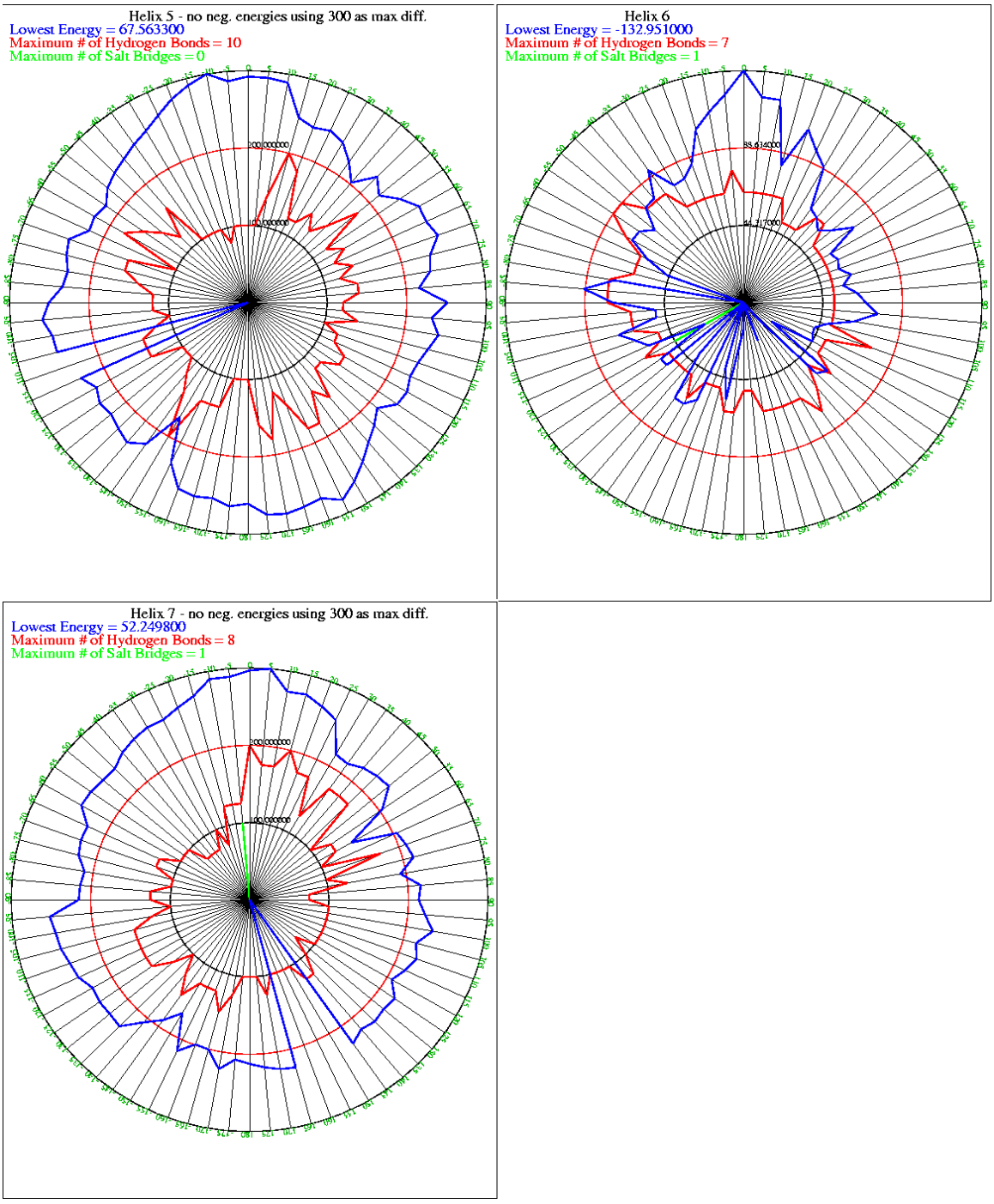
8.0 Full Rotational Scan

8.1 Crystal Rhodopsin

We took the crystal pdb structure and cut off the loops before running the bundle through RBMD. The TM regions were obtained from the pdb file itself called 1HZX chain b. We then took the bundle and ran each helix through the Individual Helix Scan program, obtaining the following *.ps files.

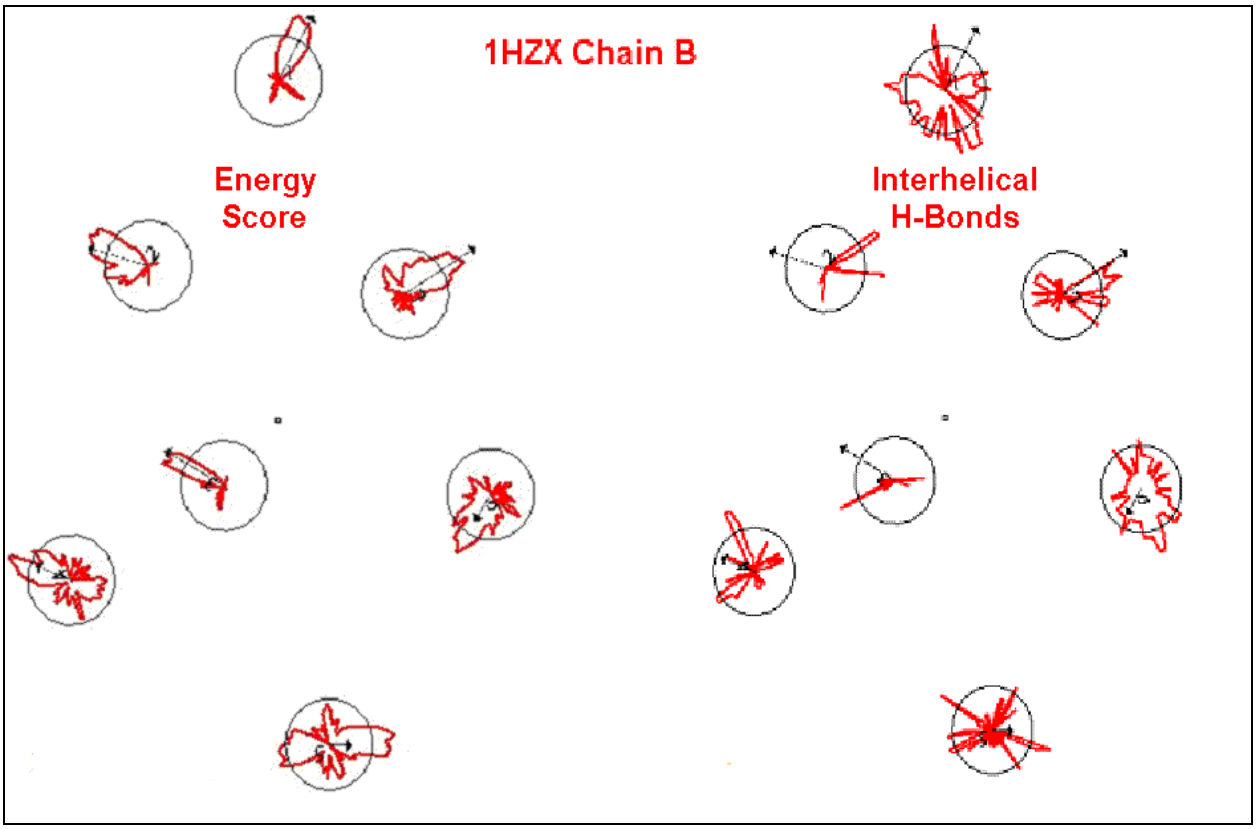


Graph 27 - Graphs of Individual Rotation Scans for Helices 1-4 of Crystal 1HZX



Graph 28 - Graphs of Individual Rotational Scans of Helices 5-7 of Crystal 1HZX

It is interesting to note that helices 2 and 3 have a range of rotations that still maintain the saltbridge. Also, helix 6, while it's energy points at helix 5 (see next graph), makes a saltbridge when it points towards the lipid.



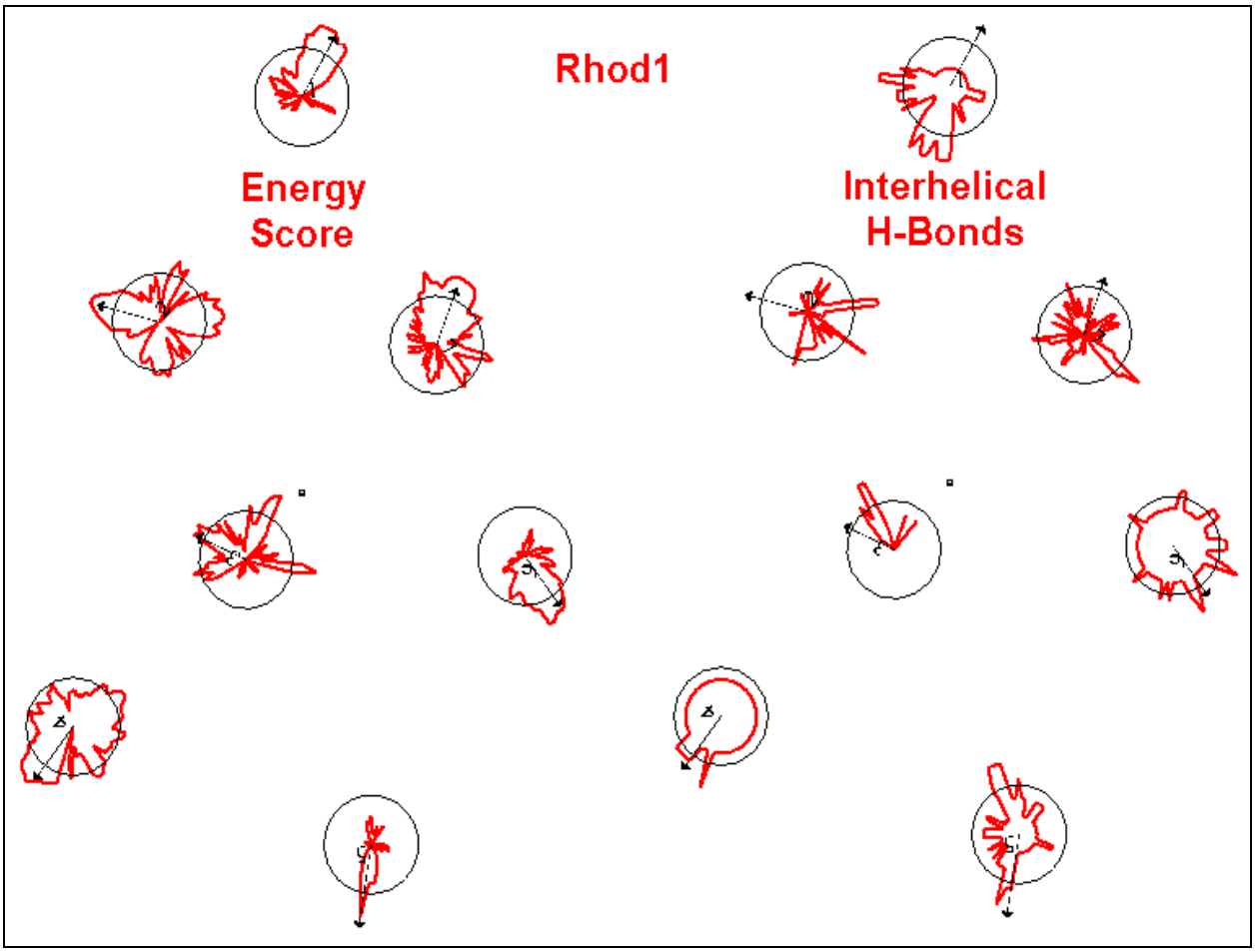
Graph 29 - Individual Rotation Scans for 1HZX on a MembComp Graph (using 100 energy scale for the top 1/3 energies, and showing the top 1/3 of Interhelical H-bonds).

From the graph above, the input rotations for the Combination program are: H1 0, 95; H2 0; H3 0; H4 0, 175, -135; H5 0, 170, -95, 65; H6 0, -105; H7 0, 100. This will produce 96 combinations.

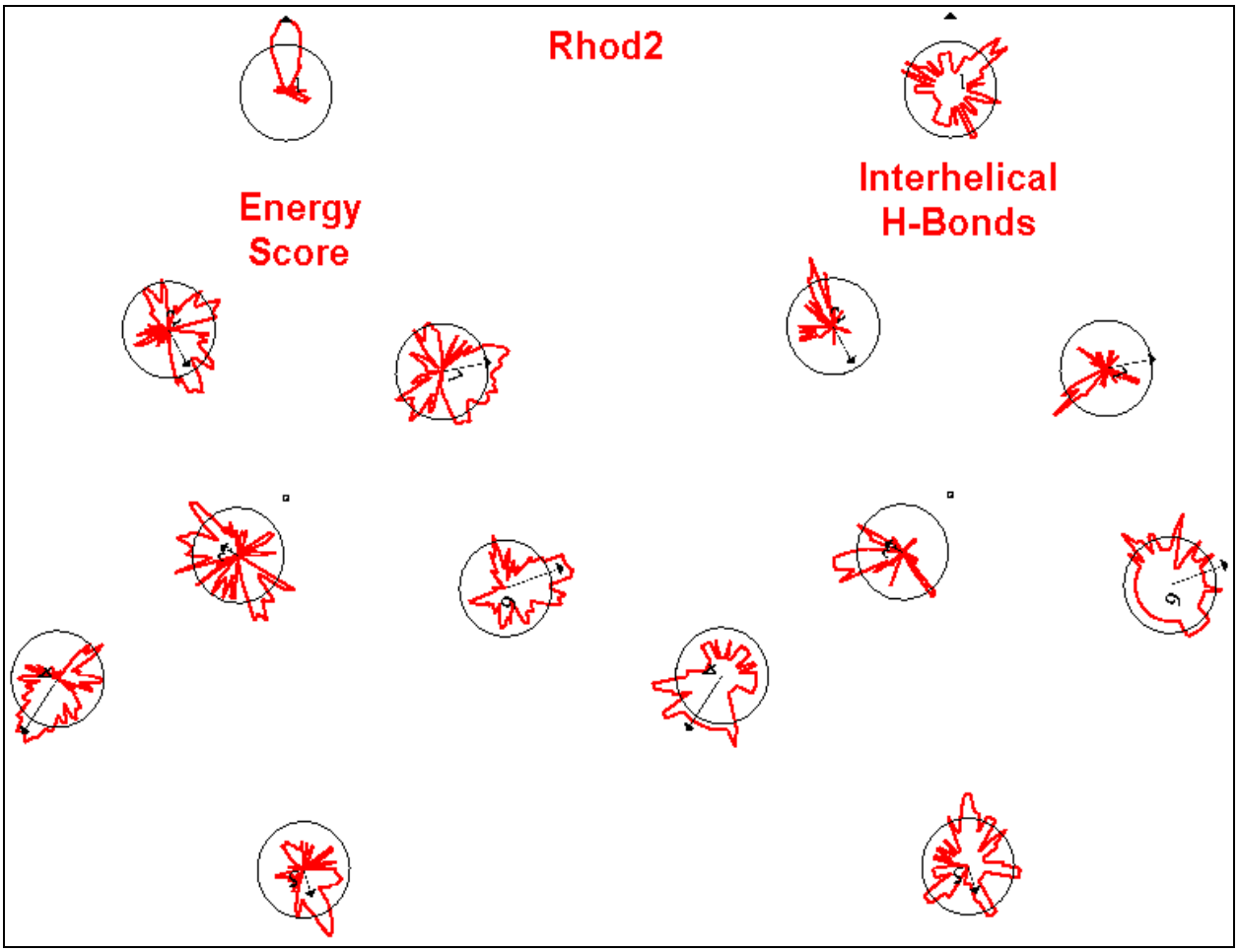
8.2 Individual Helical Scans for Rhod[1-3]

Below are the graphs from the scans of examples 1 through 3 and the combinations chosen from the information provided in these graphs. For Rhod1 the following combinations of rotations were chosen: H1 0, -90*, -135*, 85*; H2 0, 95, 160, -105; H3 95, 165, 0, -40; H4 0, -160, -70, 115; H5 0, 170*, -95*; H6 0, -130*; H7 0, 85*, 120*, 170*. Since that results in 4608 combinations, the rotations that have stars were not done (these were peaks that were 50 kcal/mol worse in energy then the best peak) leaving 64 possible combinations.

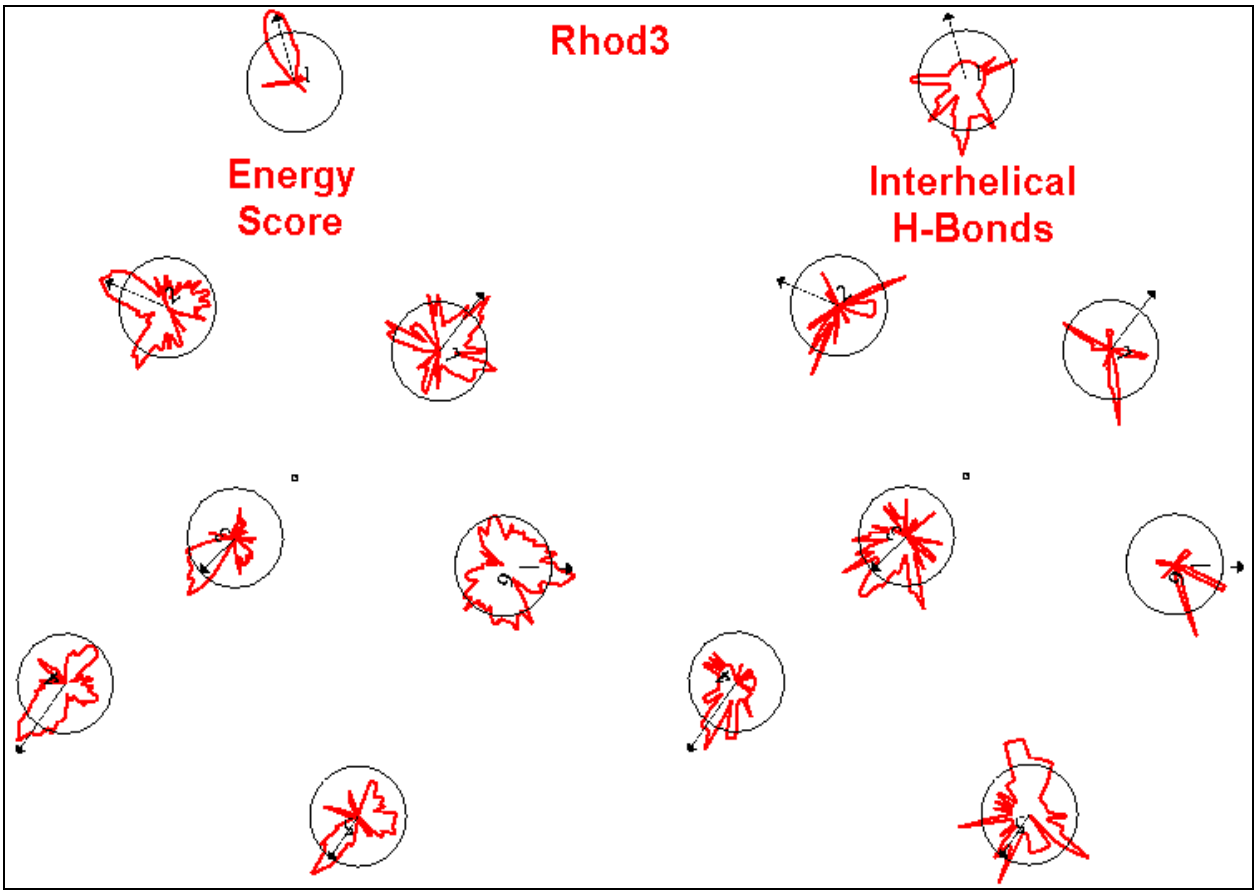
The possible rotations chosen from the graphs for Rhod2 are: H1 0, 110*; H2 0, -90, 180, 95*; H3 0, -155, -90*, 120*; H4 0, -160; H5 0, 160*, 90*; H6 0, -85, 115*; H7 0, 80, -95, 135 which give a total of 2304 combinations. These were reduced to those within 50 kcal/mol of the largest peak per helix which are not marked with a star for 96 combinations. The final structure Rhod3 had the following possible rotations: H1 0, -80*; H2 0, -85, 140; H3 0, -70*; H4 0, 180, 95*, -100*; H5 0, -145*, 70*, 125*; H6 0, -110, 100; H7 0, 70, -110 which is 1728 possible combinations, so again the peaks worse by 50 kcal/mol or more are not used and are marked with a star.



Graph 30 - Individual Rotation Scans for Rhod1 on a MembComp Graph (using 100 energy scale for the top 1/3 energies, and showing the top 1/3 of Interhelical H-bonds).



Graph 31 - Individual Rotation Scans for Rhod2 on a MembComp Graph (using 100 energy scale for the top 1/3 energies, and showing the top 1/3 of Interhelical H-bonds).



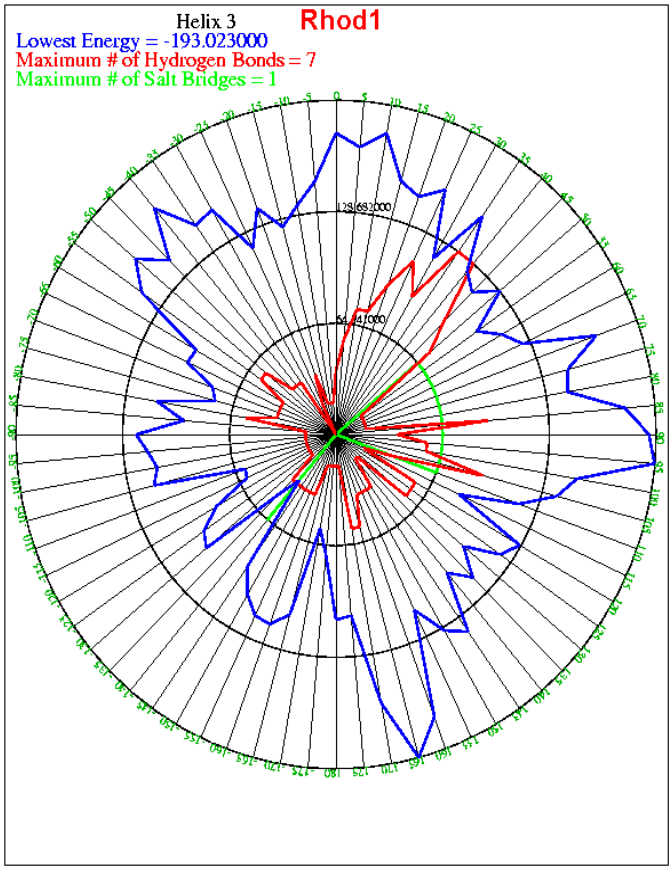
Graph 32 - Individual Rotation Scans for Rhod3 on a MembComp Graph (using 100 energy scale for the top 1/3 energies, and showing the top 1/3 of Interhelical H-bonds).

9.0 Helical Rotation Combination Generation

With the different possible rotations found through rotational scanning, all three MembStruk structures and the RBMD crystal were rotated and optimized for each possible combination chosen. The energies for each structure and their salt-bridges and hydrogen bonds were counted.

Rhod1 has it's lowest energy at the 0-0-0-0-0-0-0 position, but the first structure with a saltbridge and the lowest energy is found at 0-0-95-0-0-0-0. The standard default structure was used to build a final looped structure, but the EC2 loop would not close properly since the cystine bridge forces the EC2 loop to lay on the top of the structure.

This led to using the structure 0-0-90-0-0-0-0 since it contained the cystine in the correct place and had a low energy on the rotational scan.



Graph 33 - Helical scan of TM 3 on the Rhod1 Structure

This structure with helix 3 rotated +90 degrees has the lowest CRMS at: 2.586665 to 1U19. This lowest CRMS structure to the crystal was found by correctly assigning the cystine to place a closed EC2 loop. This shows the importance of forming the correct disulfide link to the EC2 loop for binding. The best structures compared to the 1HZX crystal are:

RMS is 2.495327 for Rhod1.0-0-95-0-0-0-0.CD.pdb to RhodC.0-0-0-0-0-0-0.CD.pdb
RMS is 3.549581 for Rhod2.0-0-0-0-0-0-0.CD.pdb to RhodC.0-0-0-0-0-0-0.CD.pdb
RMS is 2.691424 for Rhod3.0-0-0-0-0-100-0.CD.pdb to RhodC.0-0-0-0-0-0-0.CD.pdb

This shows that while Rhod1 correctly identified the helix 6 position, Rhod3 correctly identified the helix 3 position. **The methods may be improved by utilizing the Rhod1 build for everything but helix 3, and the Rhod3 build would be used for helix 3.**

10.0 Building of Loops and the Closing of the EC2 Loop

The structures Rhod1-0-0-0-0-0-0-0 and Rhod1-0-0-90-0-0-0-0 were used to build the final looped MembStruk structure. The loops for the Rhod1-0-0-0-0-0-0-0 structure were added using Whatif (Vriend 1990 reference found in manual). Once the loops were added, the EC_LOOP_SIM program (Trabanino 2004) was used to anneal the EC2 loop into a closed position. The final structure from this program resulted in a loop that was bent in the middle and spread out over the top of the protein instead of closing inside it. This was caused by the disulfide bond being formed away from the center of the protein due to the rotation of helix 3. This could indicate that the disulfide bond is formed during the binding of cis-retinal.

The structure Rhod1-0-0-90-0-0-0-0 was then run through Modeler (modeler references in manual) using a stepwise building of each individual loop followed by a minimization and sidechain optimization using Scwrl (Bower 1997). This produced a final open structure with the EC2 loop not closed on the protein. Next the EC_LOOP_SIM was used to close the loop. These two final structures, Open and Closed Rhod1, are the final structures used in docking.

11.0 Summary of Final Structure Properties

The building of the MembStruk 4.10 structures demonstrated the importance of the rotation of helix to the proper closing of the EC2 loop. The default MembStruk structure built had helix 3 rotated so that the EC2 loop could not close properly and

allowed Lys 296 to form a salt bridge with GLU 122. Additionally, While the final rotated structure involved a positive rotation of 90 degrees to allow the closing of the EC2 loop. This rotation of helix 3 also allowed the residue Lys 296 to be able to get involved in a Schiff's base bond with cis-retinal and Glu 122 to form a potential salt bridge with His 210. This suggests that the binding of cis-retinal might involve the rotation of helix 3 in forming the correct disulfide bond.