Chapter 5: Structure and Binding Site Analysis for CCR5 and CXCR4

1.0 Abstract

G-Protein Coupled Receptors (GPCRs) form a major target class for therapeutic drug development (1) and as such, their 3D protein structures are vital for future drug development. In particular, the chemokine receptor family is of particular importance for its potential for treatment of immunological pathologies. CCR5 and CXCR4 have been indicated as the most important HIV-1 coreceptors (2). This paper utilizes the MembStruk methods to develop 3D protein structures for CXCR4 and CCR5, and the HierDock protocol to define the binding site for both these receptors. A current drug patent (3) describes a ligand (Amd54) that targets both CCR5 and CXCR4, this ligand was used in HierDock to scan the receptors to locate the common binding site. Five other ligands were used for the scanning of CCR5 and five different ligands used for CXCR4. In both receptors, the predicted binding sites correlate well with the binding and mutational studies. This validates the MembStruk and HierDock protocols as well as providing new insights to the structural features of CCR5 and CXCR4.

2.0 Introduction

Integral membrane proteins are coded on 20-30% of genes (4) in humans and other organisms. These proteins take part in processes such as ion translocation, electron transfer, and transduction of extracellular signals. The G-protein-coupled receptor (GPCR) superfamily is one of the most important classes of transmembrane (TM) proteins being involved in cell communication processes and in mediating such senses as vision, smell, taste, and pain. Specifically, chemokines in particular are involved in cell growth and HIV infection (5). They are also involved in a variety of diseases related to

inflammatory cell localization: asthma, multiple sclerosis, atherosclerosis, arthritis, organ transplant rejection, and cancer (6).

Thus, chemokines have become important targets for drug development. One of the main challenges in drug design is antagonist cross-reactivity with other GPCRs and a reduced affinity in the animal models when compared to human (7). A specific example found for CCR5 are the inhibitors of Shering-Plough, where reactivity to muscarinic receptors was found (8), plus the antagonist SHC C of Shering Plough was found to have poor rodent receptor affinity (7).

The use of structural information becomes vitally important in understanding the cross-reactivity of drugs to different GPCRs. Unfortunately there is very little structure information on GPCRs although these proteins are important drug targets. In fact, there is only one experimental 3D structure for a single GPCR, bovine rhodopsin (9-10). The sequence identity to rhodopsin is low for most GPCRs of interest (17 % for dopamine, 14 % for serotonin) making the use of homology modeling for obtaining reliable structures not a valid option (11).

The MembStruk method provides a way to construct 3D structures of GPCRs without the use of homology modeling (12-13). This paper utilizes the MembStruk method to construct the 3D structures for CCR5 and CXCR4. Then the HierDock protocol was used to define the binding site through use of ligand scanning on the MembStruk structures (14-15). Amd54 (3) and five other compounds for each receptor was run through the binding site scanning methods from HierDock with good correlation to mutational studies.

3.0 Methods and Results

MembStruk version 4.3 was used in the building of the CCR5 and CXCR4 structures. HierDock version 2.5BS was then used for scanning the receptor for binding sites. The exact methodology for MembStruk and HierDock can be found in chapter 1, section 3 of this thesis. The deviations from the standard defaults are described here. Both structures were built with an open conformation to their EC-II loop.

3.1 Transmembrane (TM) Prediction for CCR5 and CXCR4

CCR5: The TM predictions previously published for CCR5 were used (16). These TM predictions come from a set of TM predictions made for CCR1 from MembStruk 4.1 then aligned to CCR5. This was done since the initial development of an aligned sequence set included the human CCR5 sequence as part of the alignment, so rerunning the TM2NDS did not produce any significant difference from the published predictions.

CXCR4: The prediction of the TM regions for CXCR4 involved a NCBI Blast search (17-18) on the SwissProt database and filtering out those hits less than a 200 bit score. ClustalW (19) was then used to align the sequences. This alignment was then used to filter out large groups of similar homologies by limiting the amount of sequences to fall within any 10 percentile to 2-4 when there exists enough. The final set of 21 sequences used for alignment is listed below:

```
CXCR4_HUMAN                    100% gi|18858505|ref|NP_571957.1|    65%
gi|17902281|gb|AAL47855.1|AF45  91% gi|29476914|gb|AAH50172.1|      65%
gi|1542889|emb|CAB02202.1|      91% gi|3327018|emb|CAA04493.1|      60%
gi|6753460|ref|NP_034041.1|     90% gi|3551197|dbj|BAA32797.1|      59%
gi|1666647|emb|CAA67893.1|      90% gi|18858507|ref|NP_571909.1|    59%
gi|28976130|gb|AAO47588.1|      91% gi|21928446|dbj|BAC05813.1|     95%
gi|9954428|gb|AAG09054.1|AF294  82% gi|5031627|ref|NP_005499.1|     32%
gi|1354505|gb|AAB01981.1|       89% gi|6467141|dbj|BAA86968.1|      32%
gi|27924174|gb|AAH44963.1|      74% gi|20387076|emb|CAC85089.1|     32%
gi|4008586|emb|CAA76923.1|      74% gi|6467137|dbj|BAA86966.1|      32%
gi|6318165|emb|CAB60252.1|      66%
```

The aligned sequences for CXCR4 were then run through coarse grain TM predictions, and a summed graph of all even window sizes from 12 to 30 was produced. The summed graph was then averaged over the 5 nearest neighbors (5 residues lower and higher in number). This averaged, summed graph (see figure 1) was then used to predict the TM regions in fine detail. The fine predictions were then run though the capping program to determine the final TM predictions.

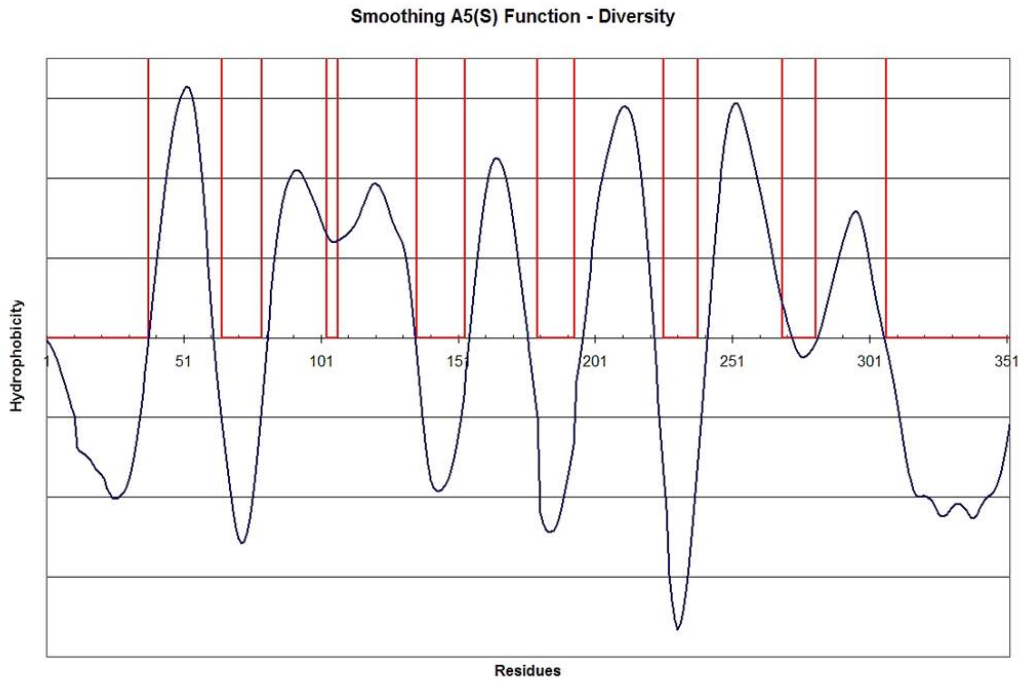

Figure 1 - Graph of the summation and averaged 5 nearest neighbors for the aligned profile of CXCR4. The red lines represent the residues not a part of the final TM predictions.

The final TM predictions for CXCR4 are shown below. The window sizes 20, 22, and 24 were indicated from this final TM prediction in the get_centers program for defining the hydrophobic moment (HPM) centers. These centers are: 51.667, 96, 120.333, 164.333, 212, 252.667, and 295.

```
NT       1 MEGISIYTSDNYTEEMGSGDYDSMKEPCFREENANFNK 38 (38)

TM 1    39 IFLPTIYSIIFLTGIVGNGLVILVMG 64 (26)
LP 1    65 YQKKLRSMTDKYRLH 79 (15)
```

```
TM 2    80 LSVADLLFVITLPFWAVDAVANW 102 (23)
LP 2   103 YFGNF 107 (5)

TM 3   108 LCKAVHVIYTVNLYSSVLILAFISLDRY 135 (28)
LP 3   136 LAIVHATNSQRPRKLLAE 153 (18)

TM 4   154 KVVYVGVWIPALLLTIPDFIFANVSE 179 (26)
LP 4   180 ADDRYICDRFYPND 193 (14)

TM 5   194 LWVVVFQFQHIMVGLILPGIVILSCYCIIISK 225 (32)
LP 5   226 LSHSKGHQKRKAL 238 (13)

TM 6   239 KTTVILILAFFACWLPYYIGISIDSFILLE 268 (30)
LP 6   269 IIKQGCEFENTVH 281 (13)

TM 7   282 KWISITEALAFFHCCLNPILYAFLG 306 (25)

CT     307 AKFKTSAQHALTSVSRGSSLKILSKGKRGGHSSVSTESESSSFHSS 352 (46)
```

3.2 Construction of the MembStruk Structures

CCR5: The MembStruk 4.1 defaults were used for the construction of the CCR5 receptor. The final structure after rigid body (RBMD) was run through the helical rotation scanning and found the following possible rotations as local minima: TM 1 (0,+80), TM 2 (0,-80,+90,180), TM 3 (0,+110), TM 4 (0), TM 5 (0), TM 6 (0), TM 7 (0,+110). The lowest energy structure was the original starting structure with no rotations.

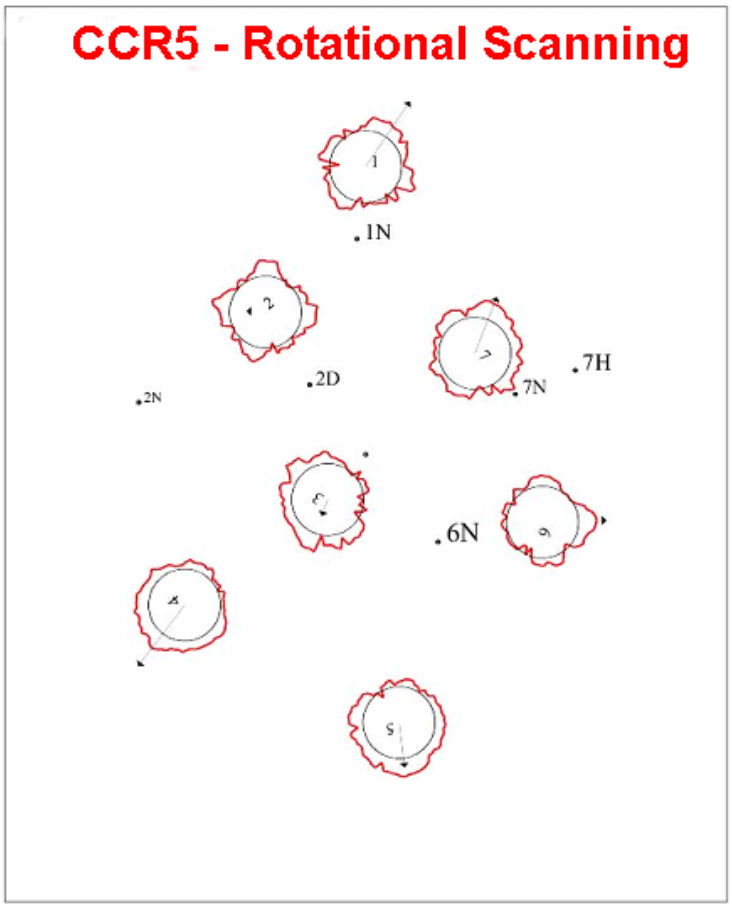

Figure 2 - Rotational scanning for the CCR5 Membstruk structure, residues shown are those hydrophilic residues in the middle 15 residues of the TM regions.

CXCR4: The structure was build using MembStruk 4.1 defaults. The coarse rotational optimization was done twice since the ASP on TM 2 was pointing towards the lipid after the first optimization. This ASP is also part of a conserved trio of residues that are often found in GPCRs. The trio is: ASN, ASN, ASP that form a bridge for TM7 – TM 1 – TM2. In this case, the ASP lies on TM 2 and was not located inside the protein barrel.

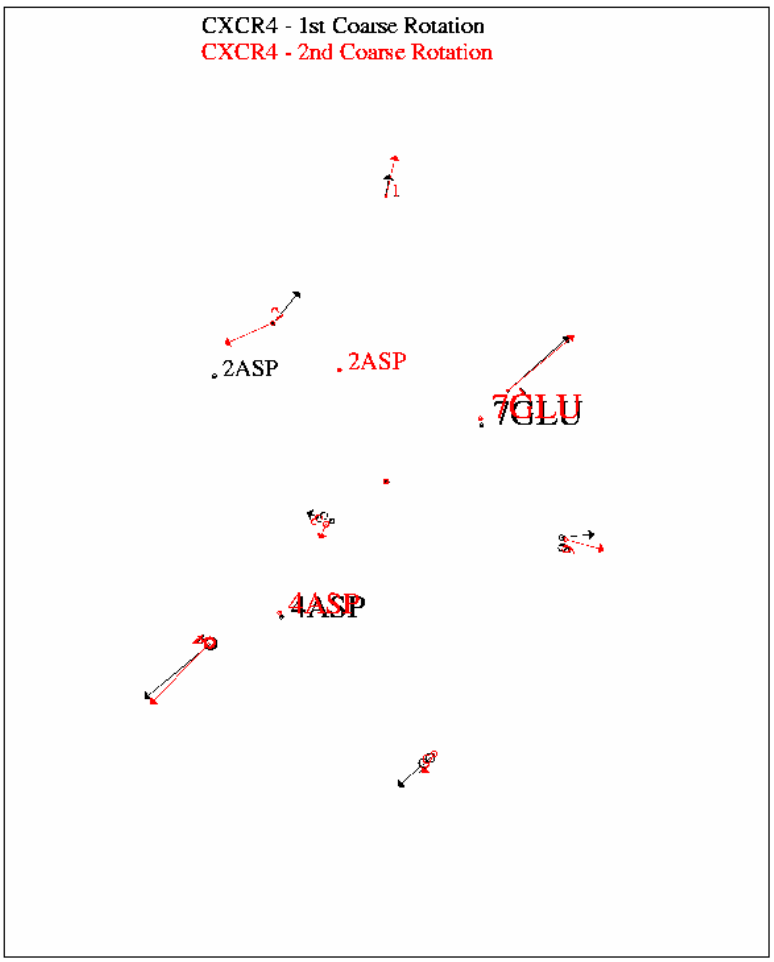

Figure 3 - MembComp graph of the final rotations after the 1st and 2nd run through the coarse rotational optimization step for CXCR4.

The structure was then run thought the fine grain rotational optimization program (rotmin). The final rotations were: TM 1 -5.0, TM 2 30.0, TM 3 -20.0, TM 4 15.0, TM 5 5.0, TM 6 -10.0, TM 7 20.0. Helix 2 was then rotated by -45 degrees using MembComp to bring the ASP back into the protein center. The analysis for TM 3 showed that no major rotation was needed to helix 3 (see Figure 4). The helical scanning (done during the beta phase of the helical scanning program) also showed that the original positions are part of a possible local minima.

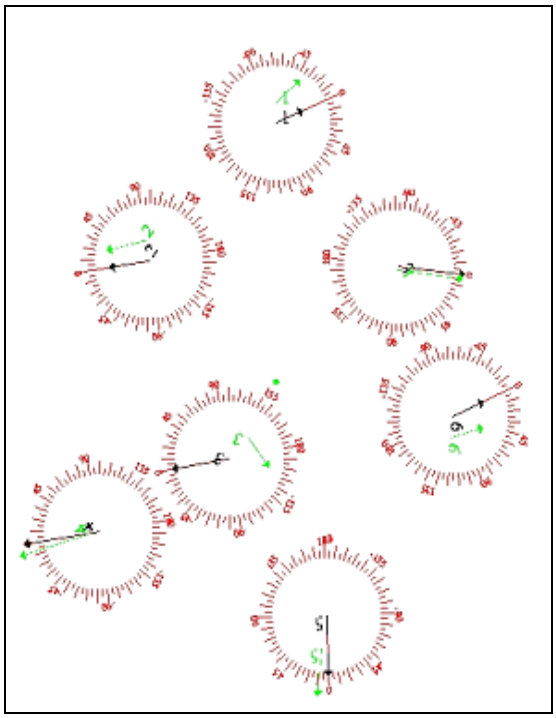

Figure 4 - Analysis of the hydrophobic extracellular (EC) half of TM 3 (black) and the intracellular (IC) half of TM 3 (green).

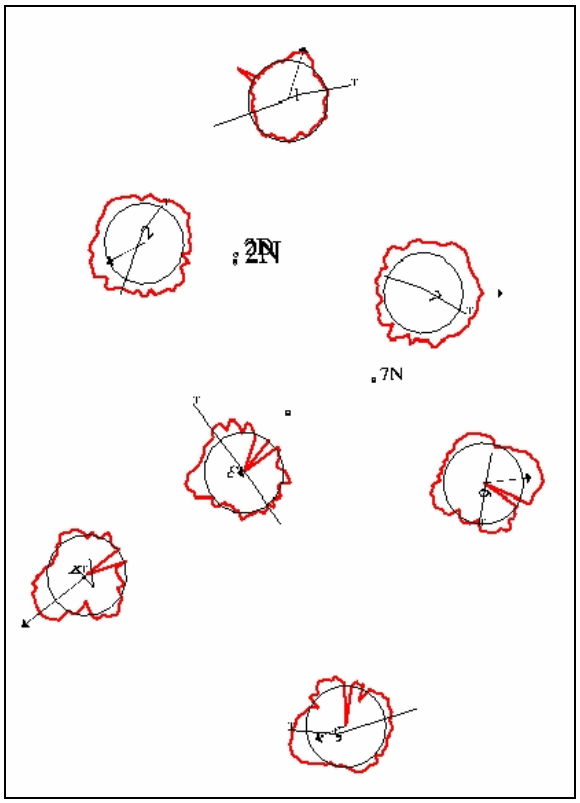

Figure 5 - Helical scanning of TM regions, with energies plotted. Shown are the Asp and Asn residues found on TM regions 2 and 7.

3.3 Proposed Improvements to MembStruk Structures

CCR5: The final structure for scanning looks good, but a few improvements that were unavailable when theses structures were first created might improve the structures. The hydrophobic analysis of the EC and IC halves of TM 3 would help to verify the correct placement of this helix's rotation. The helical scanning was done with a beta version of the program, and a better resolution of the scanning graph would help to verify if TM 7 is correctly rotated or if the Asn should be rotated closer into the TM 1, TM 2, TM 7 pocket that has been seen in the D2 structure (20). This verification of the rotation of TM 7 is even more important with the binding site scanning indicating that the TM 1, TM 2, TM 7 pocket is the binding site.

CXCR4: The main improvements here are in the possible rotations of TM 3 and TM 7. The analysis of TM 3's EC and IC halves indicate that helix 3 could be rotated up to -45 degrees and still maintain the hydrophobic moments in the right gaps. TM 7 might form an ASP-ASN-ASN bridge if rotated ~90 degrees, and the current helical scanning shows that helix 7 is very mobile. Re-running the helical scanning with finer detail might pin-point specific rotations that are local minima.

3.4 Scanning of the Receptors for Binding Sites

Each one of the ligands used in HierDock scanning was created with gasteiger charges. The ligands used for scanning of the CCR5 receptor are: Amd54, Tak-779, sch350581, sch351125, cis-pyrrolidine, and trans-pyrrolidine. The ligands used for scanning the CXCR4 receptor in HierDock are: Amd54, Anormed 1-3, Kureha 1, Takeda 1.

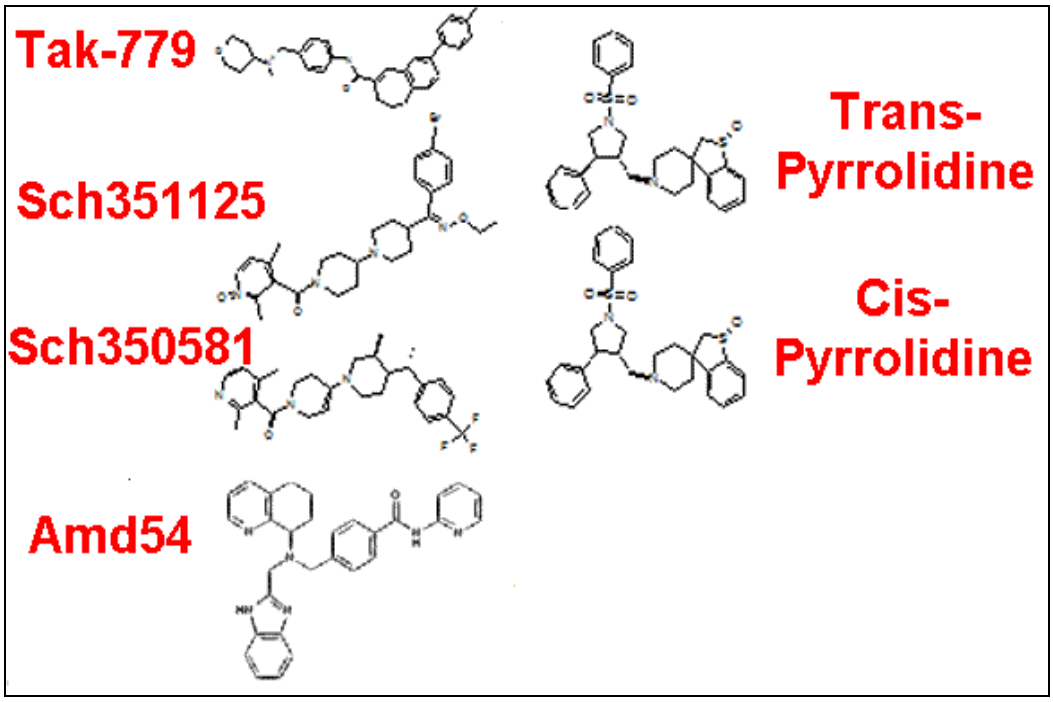

Figure 6 - Ligands used for binding site scanning on CCR5

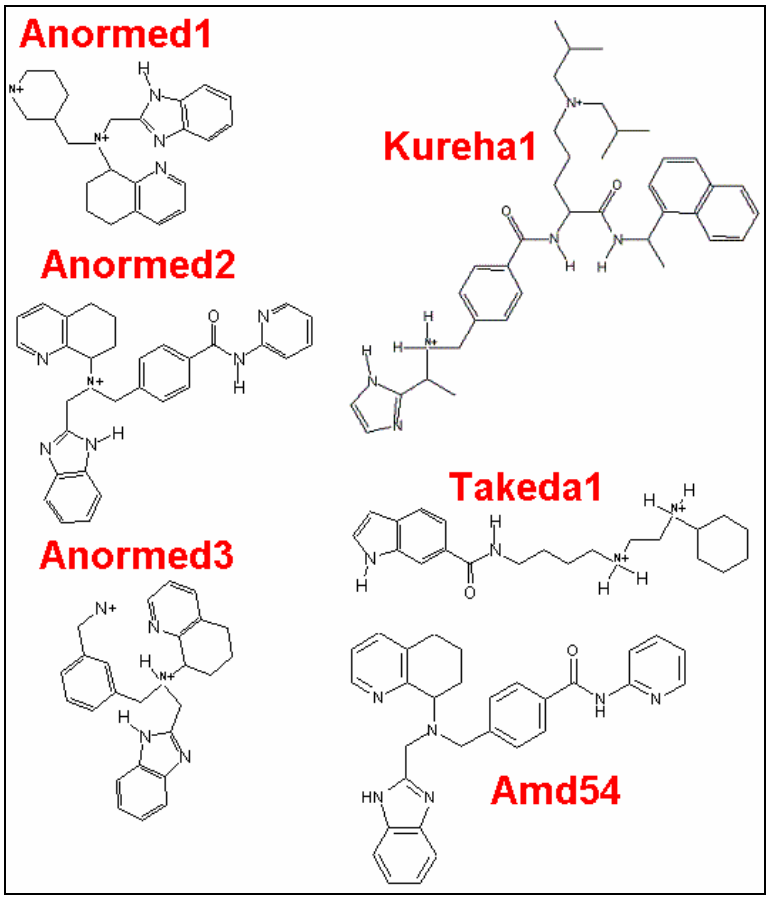

Figure 7 - Ligands used for HierDock scanning of CXCR4.

Each one of these ligands was run through the GrowBox scanning protocol of HierDock to discover the binding site. The results of the growbox scanning were then ranked according to buried surface and binding energy of conformations. These rankings are listed below:

```
CCR5
Amd54      90% - 49*(10) 45(6)   36(5)   1(3)  52(1)
Cis-pyr1   90% - 49*(10)  2(6)    1(6)  45(2)  53(1)
Sch350581 90% - 49(9)    1*(7)  45(5)   2(2)  19(1)
Sch351125 90% - 49(9)    1*(7)   2(3)  45(1)
Tak779     85% -  1*(7)  49(5)    2(3)  19(2)  37(2)
Trans-pyr190% - 49*(10)  1(3)   45(3)   2(2)

CXCR4
Amd54      85% - 26*(10)   4(4)   1(3)   5(1)
Anormed1   95% -  2(10)    1(6)  26*(5) 27(1)
Anormed2   85% - 26*(10)  27(3)   1(2)  22(1)
Anormed3   90% - 26*(10)   2(10)  1(10) 27(4)  22(1)
Kureha1    65% -  1*(2)   22(1)  21(1)  26(1)
Takeda1    80% -  1*(10)   2(10) 30(2)  21(1)  22(1)
```

Averaged Rankings

CCR5		CXCR4	
Box	Rank (Score)	Box	Rank (Score)
Box 49 –	1st (8)	Box 26 -	1st (6)
Box 1 -	2nd (8)	Box 1 -	2nd (7)
Box 2 -	3rd (4)	Box 2 -	3rd (7)
Box 45 -	3rd (4)	Box 27 -	4th (5)

The boxes chosen for CCR5 to determine the binding site were boxes 1 and 2 located in the TM 1-3, TM 7 pocket. Box 49 (located in the same region as boxes 1 and 2) was not used for the binding site determination since its center was located below the center of the protein causing all ligands to dock too close to the intracellular region of the protein which is not generally the correct spot (15). This binding site correlates well with mutational studies that describe several residues in this region as being important for binding. These residues are ‘antiviral activity – residues (TM region)’: Strong - L33, Y37 (TM 1), W86 (TM 2), Y108, T123 (TM3); Moderate - R31 (TM1), T82 (TM2), I198 (TM5), E282 (TM7); Borderline - F79 (TM2), L104 (TM3) (21-22).

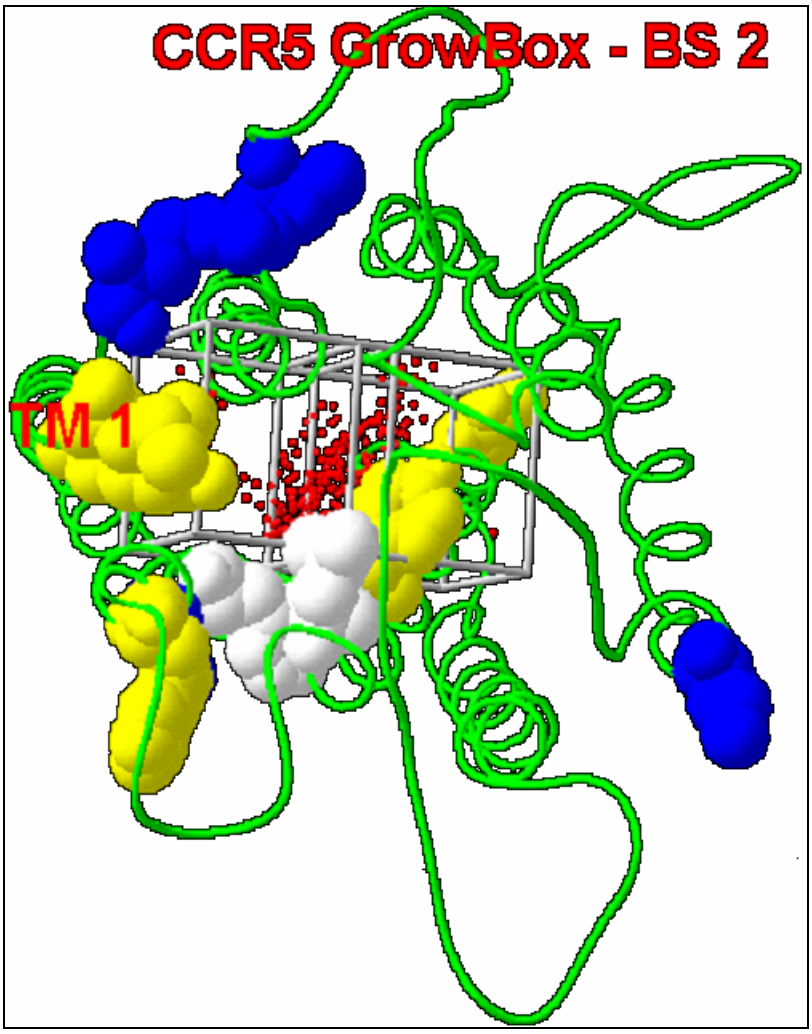

Figure 8 - Binding site obtained from GrowBox shown as grey boxes and red spheres. Mutations on residues that had strong antiviral activity are marked in yellow, the moderate residues marked in blue, and the borderline residues marked in white.

The binding site for CXCR4 was different, being located in the TM 3-6 pocket. The boxes used to form the binding site sphere set were boxes 26, 1, 27 that were all in this region. This binding site correlates well to evidence from mutational studies performed on this receptor. The mutational studies found that several mutations in the EC-II loop impair the co-receptor activity, and also that the residues D171 (TM 4), H203 (TM 5), D262 (TM 6), and E288 (TM 7) are important in antagonist binding (22-25).

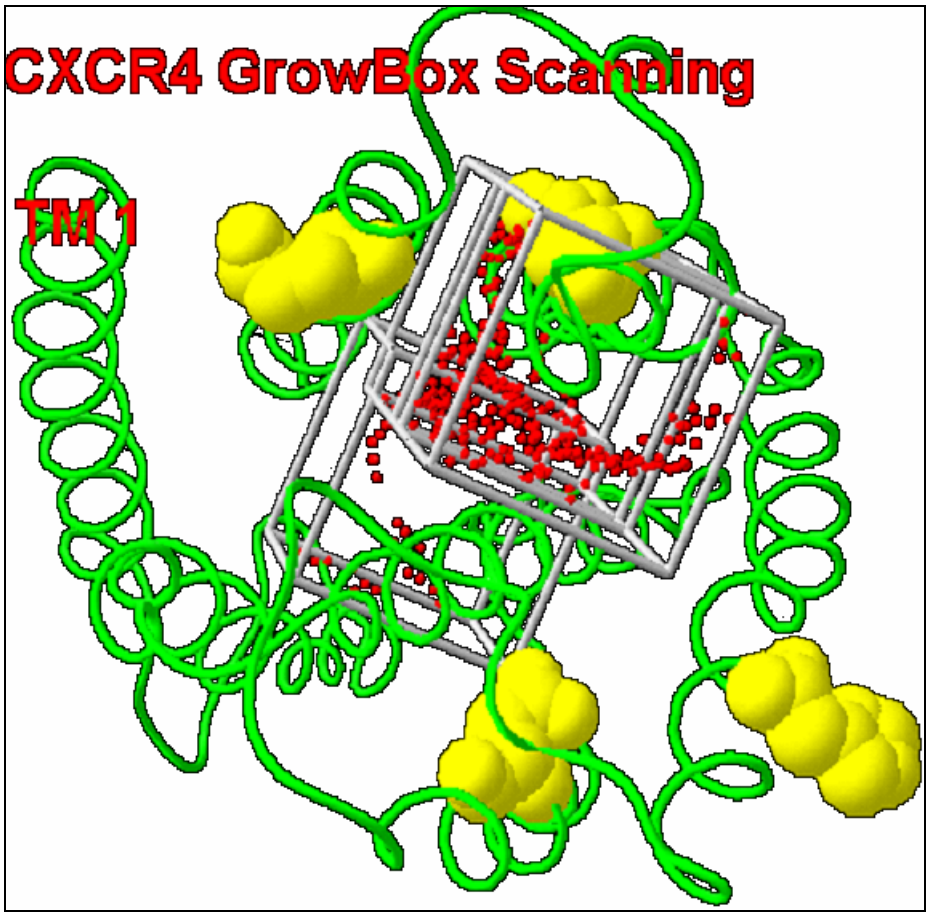

Figure 9 - Binding site found from GrowBox shown in the grey boxes and red spheres. Residues implicated in antagonist binding are colored yellow.

4.0 Discussion

These two receptors both developed during the transition between MembStruk 4.1 and 4.3 are great validation cases for the MembStruk and HierDock protocols in defining the correct binding location. The correlation between the mutational studies and the GrowBox scanning is very good, with the same areas being implicated in binding. This also provided a validation case for the new ranking method implemented in HierDock 2.5BS for GrowBox.

The ligand Amd54 that binds to both CCR5 and CXCR4 was found to bind the in the same locations as the other ligands. This suggests that either the two binding sites even though being on different sides of their receptors have similar characteristics, or that

Amd54 binds with one end in one receptors binding site and the opposite end in the other receptors pocket.

With the MembStruk structures becoming more and more accurate in detailing the 3D structural information of the selected GPCR, it is important to verify the accuracy of the function prediction capabilities of HierDock on these structures. In both cases, CCR5 and CXCR4 MembStruk structures are used to correctly identify the binding sites regions according to mutational studies. In particular, this paper shows the usefulness of the GrowBox procedure and the new ranking method for determining the best binding site.

5.0 References

1) Brechler, V., Handfield, D., Boissonneault, M., Tremblay, E., Houle, B., and Menard, L. (2004). Development of Radioligand Binding Assays for the Motilin Receptor Using ScreenReady Targets. *Application Note, Screenready Targets SRT 001. http://las.perkinelmer.com/content/RelatedMaterials/SRT-001%20(application%20note%20Motilin2).pdf*

2) Veazey1, R. S., Klasse, P. J., Ketas, T. J., Reeves, J. D., Piatak, Jr., M., Kunstman, K., Kuhmann, S. E., Marx1, P. A., Lifson, J. D., Dufour, J., Mefford, M., Pandrea, I., Wolinsky, S. M., Doms, R. W., DeMartino, J. A., Siciliano, S. J., Lyons, K., Springer, M. S., and Moore, J. P. (2003) Use of a Small Molecule CCR5 Inhibitor in Macaques to Treat Simian Immunodeficiency Virus Infection or Prevent Simian–Human Immunodeficiency Virus Infection. *Journal of Experimental Medicine.* Vol. 198(10), 1551-1562.

3) Anormed Inc. (2002) Novel heterocyclic compounds are modulators of chemokine receptors CXCR4 or CCR-5 useful for the treatment of HIV infection. Patent. WO-00222599 March 21.

4) Wallin, E. and von Heijne, G. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* 7, 1029-1038.

5) Mackay, C. R. (2001). Chemokines: immunology's high impact factors. *Nature Immunology* 2, 95-101

6) Gerard, C. R., B.J. (2001). Chemokines and disease. *Nature Immunology* 2, 108-115.

7) Horuk, R. (2003). Development and evaluation of pharmacological agents targeting chemokine receptors. *Methods* 29, 369-375.

8) Tagat, J. R., Steensma, R. W., McCombie, S. W., Nazareno, D. V., Lin, S. I., Neustadt, B. R., Cox, K., Xu, S., Wojcik, L., Murray, M. G., Vantuno, N., Baroudy, B. M., and Strizki, J. M. (2001). Piperazine-Based CCR5 Antagonists as HIV-1 Inhibitors. II. Discovery of 1-[(2,4-Dimethyl-3-pyridinyl)carbonyl]-4-methyl-4-[3(S)-methyl-4-[1(S)-[4-(trifluoro- methyl)phenyl]ethyl]-1-piperazinyl]-piperidine N1-Oxide (Sch-350634), an Orally Bioavailable, Potent CCR5 Antagonist. *J. Med. Chem.* 44(21), 3343-3346.

9) Palczewski, K., Kumasaka, T., Hori, T., Behnke, C., Motoshima, H., Fox, B., Trong, I., Teller, D., Okada, T., Stenkamp, R., Yamamoto, M. and Miyano, M. (2000). Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* 289, 739-745.

10) Teller, D., Okada, T., Behnke, C., Palczewski, K. and Stenkamp, R. (2001). Advances in Determination of a High-Resolution Three-Dimensional Structure of Rhodopsin, A Model of G-Protein-Coupled Receptors (GPCRs). *Biochemistry* 40, 7761-7772.

11) Archer, E., Maigret, B., Escrieut, C., Pradayrol, L., and Fourmy, D. (2003). Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends Pharmacol. Sci.* 24, 36-40.

12) Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and Goddard III, William A. (2002). Prediction of structure and function of G protein-coupled receptors. *Proc. Nat. Acad. Sci*. 99(20), 12622-12627.

13) Trabanino R.J., Hall, S.E., Vaidehi N., Floriano W.B., and Goddard III, W.A. (2004). First Principles Predictions of the Structure and Function of G-Protein Coupled Receptors: Validation for Bovine Rhodopsin. *BioPhys. J*. 86, 1904-1921.

14) Floriano W.B., Vaidehi, N., Singer M.S., Shepherd G.M., Goddard III, W.A. (2000). Molecular Mechanisms underlying differential odor responses of a mouse olfactory receptor. *P. Natl. Acad. Sci. USA* 97, 10712-10716.

15) Floriano, W. B., Vaidehi, N., and Goddard III, W. A. (2004). Making sense of olfaction through molecular structure and function prediction of olfactory receptors. *Chem. Senses* 29, 269-290.

16) Trabanino R. J. (2004). Caltech Chemistry PhD Thesis.

17) Altschul, S. F., Gish, W., Miller, W., Myers, W. E., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

18) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

19) Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). Clustal-W improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22, 4673–4680.

20) Kalani, M. Y. S., Vaidehi, N.;Hall, S. E., Trabanino, R. J., Freddolino, P. L., Kalani, M. A.;Floriano, W. B.;Kam, V. W. and Goddard III, W. A. (2004). "The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists." *Proc. Natl. Acad. Sci.* 101, 3815-3820.

21) Kazmierski, W.; Bifulco, N.; Yang, H.; Boone, L.; DeAnda, F.; Watson, C.; Kenakin, T. (2003). Recent Progress in Discovery of Small-Molecule CCR5 Chemokine Receptor Ligands as HIV-1 Inhibitors. *Bioorg. Med. Chem. 11,* 2663-2667.

22) Dragic, T. (2001). An overview of the determinants of CCR5 and CXCR4 co-receptor function. *Journal of General Virology* 82, 1807–1814.

23) Hatse, S., Princen, K., Gerlach, L., Bridger, G., Henson, G., De Clercq, E., Schwartz, T. W., and Schols, D., (2001). Mutation of Asp171 and Asp262 of the Chemokine Receptor CXCR4 Impairs Its Coreceptor Function for Human

Immunodeficiency Virus-1 Entry and Abrogates the Antagonistic Activity of AMD3100. *Molecular Pharmacology*. 60(1), 164-173.

24) Rosenkilde, M. M., Gerlach, L., Jakobsen, J. S., Skerlj, R. T., Bridger, G. J., and Schwartz, T. W. (2003). Molecular Mechanism of AMD3100 Antagonism in the CXCR4 Receptor. *Journal of Biological Chemistry*. 279(4), Iss. Jan. 23, 3033-3041.

25) Gerlach, L. O., Skerlj, R. T., Bridger, G. J., and Schwartz, T. W. (2001). Molecular Interactions of Cyclam and Bicyclam Non-peptide Antagonists with the CXCR4 Chemokine Receptor. *J. Bio. Chem.* 276(17), 14153-14160.