CReaTE

Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

http://create.canterbury.ac.uk

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g. Carey, D. (2018) Molecular dynamics study of melatonin binding to homology modelled Mel1a G-protein coupled receptor. M.Sc. thesis, Canterbury Christ Church University.

Contact: create.library@canterbury.ac.uk

Molecular Dynamics Study of Melatonin Binding to Homology Modelled Mel1a G-protein Coupled Receptor.

By

Daniel Carey.

Canterbury Christ Church University.

Thesis submitted for the Degree of MSc by Research.

2018.

Abstract:

The Mel1a G-coupled Protein receptor (GPCR) was modelled using the I-Tasser online web service. All-atom molecular dynamics was used to improve the structure. The primary ligand melatonin was docked to the structure post molecular dynamics and structurally aligned to the X-ray crystallographic structures of the $\beta2$ adrenergic and rhodopsin GPCR's, of the same family of proteins. A second set of all-atom molecular dynamics was undertaken with melatonin in the proposed active site which was parameterized ab initio in Gaussian16 to note any key conformational changes due to binding. The Mel1a GPCR becomes depolarized as a result of binding in the proposed active site by melatonin, based on Van der Waal interaction with amino acid residues on the extracellular side of the membrane (Ser176, Cys177, Tyr281 and Ser103).

<u>Contents</u>

Introduction.

Melatonin.

Melatonin also known as N-acetyl-5-methoxytryptamine is the primary hormone released from the pineal gland in the suprachiasmatic nucleus (SCN) (Carter and Juurlink, 2012). Melatonin is a derivative molecule of the amino acid tryptophan. This process follows a 4-step pathway, where tryptophan is converted into 5-hydroxytryptophan. This is performed by the enzyme tryptophan 5-monooxygenase. Following this, L-amino acid carboxylase catalyses the conversion of 5-hydroxytryptophan into the hormone serotonin (5-hydroxytryptamine). Another enzyme called Arylalkyl amine-*N*-acetyltransferase acetylates serotonin resulting in N-acetylserotonin, which is the immediate precursor molecule of melatonin. This process is performed in the pineal gland. The final enzyme responsible for the conversion to melatonin is hydroxy-indole-O-methyltransferase (also known as acetylserotonin methyltransferase) – which attaches a final methyl group to the OH group forming melatonin (Tan *et al.*, 2015) (Figure 1.1). The structure of melatonin was crystallised by (Wakahara, Fujiwara and Tomita, 1972).

The release of melatonin from the pineal gland is performed in a rhythmic fashion and peaks in the dark portion of the 24-hour day-night cycle performing a 'sleep cue'. The absence of light initiates the light-inhibited production of melatonin in the superchiasmatic nucleus (Dollins *et al.*, 1994). The role of melatonin in sleep regulation has led to research into the potential treatment of human disease, most prominently sleep disorders (Jan, Espezel and Appleion, 1994). Examples of this research show hippocampal restoration of neural precursor cell proliferation and has also shown to reduce cognitive defects caused by sleep wake cycle displacement. Which is more commonly known as jet-lag after long journeys in an eastward bound direction (Caspi, 2004). Melatonin has also been observed to have an influence on blood pressure regulation (hypertension), influencing endothelial dysfunction,

inflammation and reducing free radical burden (Dollins *et al.*, 1994). Abnormal melatonin synthesis has also been a research focus in behavioural disorders such as the autism spectrum where low levels of melatonin have been reported by independent groups described through review by Melke *et al.*, 2008. The sleep quality of people suffering from schizophrenia has also been shown to be enhanced by melatonin supplement (Shamir *et al.*, 2000). It has also been shown to improve sleep quality in several disease and physical states including but not limited to - insomnia, broken bones and the terminally ill (Bourne, Mills and Minelli, 2008). At current, melatonin is available as a health supplement which is available without a prescription but is cautioned as a short-term solution as less is known about long-term safety. The slow release melatonin drug Circadin is a prolonged release formulation of the hormone which is designed to mimic the melatonin release pattern during the night cycle and is taken orally. The target audience of Circadin is the elderly over the age of 55 for which it is currently licenced, as studies show that melatonin hormone release deteriorates with age (Lemoine and Zisapel, 2012). Melatonin is the primary ligand of the G-protein coupled receptors (GPCR) Mel1a (Slaugenhaupt *et al.*, 1995) and Mel1b (Reppert *et al.*, 1995) in humans, and many eukaryote species (also known as MT^1 and MT^2, or MTNR1A and MTNR1B).

G-Protein Coupled Receptors (GPCR).

Heterotrimic guanine nucleotide-binding protein (G-protein)-coupled receptors, are a family of cell surface proteins which respond to stimulus such as light, amines, hormones, peptides and even large proteins (Kobilka, 2007). They play a large variety of roles in the body. The protein family share a common architecture, consisting of a single polypeptide which is embedded in the membrane of the eukaryotic cell. Seven trans-membrane helix sections span the entire length of the membrane, which is why they are also commonly known as 7-trans-membrane receptors (7TM). They are the largest, most diverse group of membrane receptors in eukaryotes. Understanding of GPCR's has affected modern medicine and it is was estimated that 26.8% of approved drugs target the rhodopsin-like GPCR family of proteins in 2007 (Sriram and Insel, 2018). In 2017, 134 approved drugs are for the GPCR Family. The figure below shows an example of state change due to agonist binding in a GPCR and its following interactions with the heterotrimic g-proteins (Figure 1.2).
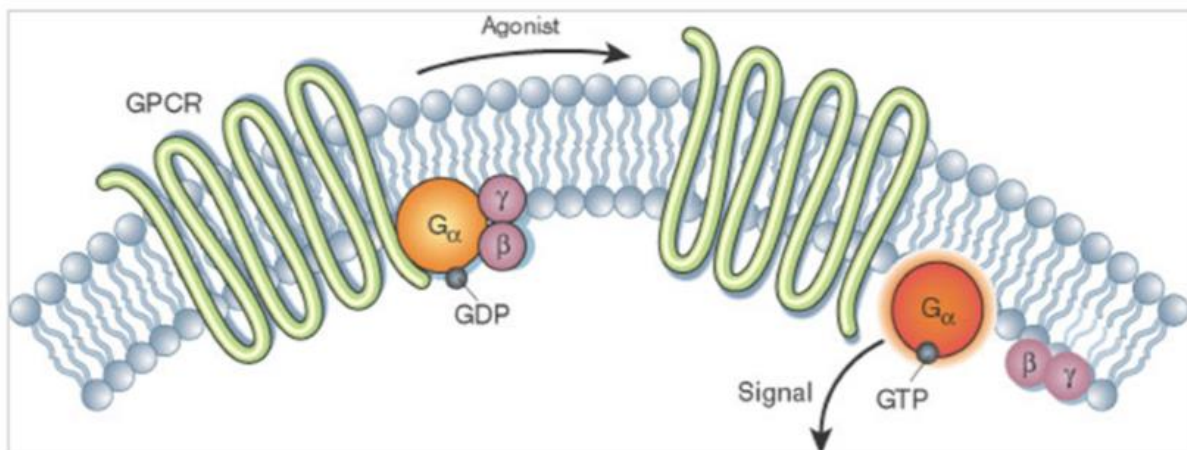
Figure 1.2, taken from, © 2002 Nature Publishing Group Li, J. et al. The Molecule Pages database. Nature **420**, 716-717 (2002). All rights reserved.

The first crystal structure of a GPCR was the bovine derived rhodopsin GPCR, which was diffracted with a resolution of 2.8Å and described as a highly organised structure with a conserved di-sulphide bridge (Palczewski *et al.*, 2000). Rhodopsin's are a large member of the subfamily and constitute ~90% of all GPCR's. Rhodopsin is a light sensitive receptor which is involved in phototransduction. It is found in the rods of the retina and when activated by light. The Active conformation of rhodopsin binds the rod cell G protein transducin, which catalyses the exchange of Guanine di-phosphate (GDP) for a guanine tri-phosphate (GTP), sending appropriate signals to the cell. The alpha subunit of the g-protein activates the effector enzyme cGMP phosphodiesterase, binding to its inhibitor subunit causing hyperpolarization of the ROS plasma membrane through hydrolysis of cyclic guanosine monophosphate (cGMP). Reduced flux of Na^+ ions through the cGMP-gate channels generates the neuronal signal in response to

light. Another crystal structure with a 2.4Å resolution was published in 2007 for the β2-adrenergic GPCR (Rasmussen *et al.*, 2011). The GPCR is responsible for cell signalling in the fight-or-flight response caused by its primary hormone epinephrine. The β2 adrenergic receptor signal is transmitted through adenylyl cyclase, which is an enzyme with key regulatory rolls in eukaryotic cells. This causes a signalling cascade which is initiated by the secondary messenger Cyclic adenosine monophosphate. This mediates physiological response resulting in smooth muscle relaxation and bronchodilation. By fusing the structure to a T4 lysosome sourced from *Escherichia coli,* which replaced the loop section of the third intracellular loop of the GPCR. It was then crystallized with the presence of the agonist carazolol. Static analysis of proteins however alone, only infer structure in an instant (Dror *et al.*, 2011). Attempts to see the dynamic change seen due to ligand binding have been attempted for protein complexes. When first characterized in the 1970's, GPCR's were often regarded as static signalling proteins active and in active states. Many studies since have shown GPCR's to be versatile in distinct conformation depending on bound ligand. An example of this has been performed with the β2 –adrenergic GPCR, Crystal structures provide snapshots in conformational character. From those, dynamic energy landscapes can be constructed (Nygaard *et al.*, 2014). Nuclear magnetic resonance (NMR) spectroscopy and Molecular dynamic (MD) simulations (Karplus and McCammon, 2002) are used to characterize conformational dynamics of the protein. The structures of resultant protein conformation can be cross referenced to provide details about function with the addition of certain stimulus.

Homology modelling.

Homology modelling of proteins based on the information available of similar structures presents a solution to X-ray crystallization of proteins (Dolan, Noah and Hurt, 2012). Since the GPCR is believed to share a similar architecture amongst its own family, it is reasonable to use previous data from same family proteins to predict a model of a protein. By using a known protein or DNA sequence which transcripts a protein, it is possible to build a representative model. The Swiss-model workspace is an example of a homology modelling web-based program which is used for 3-dimensional modelling. At the time of publishing in 2005, 33,000 experimentally determined protein structures were deposited in the protein databank. While the Uniprot protein knowledge database held more than 2.3 million protein sequences in the same

period (Arnold *et al.*, 2006). This leaves homology modelling a vast swathe of potential for structurally characteristic models. This method of building a 3-dimensional structure comprises steps which can lead to a representative model for detailed study. By building a model using the characteristic templates via sequence alignment, a model which is suitable can be chosen and modified manually post generation (Dolan, Noah and Hurt, 2012). This is performed by a selected modelling server like I-Tasser (Zhang, 2008) (iterative threading assembly refinement algorithm) uses the LOMETS (local meta-threading server) online web service (Wu and Zhang, 2007), which predicts a 3-dimensional structure based on target-to-template alignments from locally installed threading programs. A nucleotide sequence known to be responsible for an identified protein, can be taken from a website such as the protein data bank or the universal Protein Resource (UniProt) (Wu, 2006). Once obtained, I-Tasser searches for possible folds by four simple variants of PPA (profile to profile alignment threading algorithms) methods. Modelling is achieved using The Hidden Markov model (Karplus, Barrett and Hughey, 1998) PSI-BLAST (Altschul *et al.*, 1997), and the Needleman-wunch (Needleman and Wunsch, 1970) and smith-waterman (Smith and Waterman, 1981) alignment algorithms. Continuous fragments are then removed from the aligned regions and used to re-assemble full length models. Remaining structure such as loops are constructed by *ab initio* modelling. The SPICKER algorithm is used both as a simple strategy to identify protein folds by clustering protein structures, and for structure trajectories - which are generated during computer simulations (Smith and Waterman, 1981). This helps in predicting a native state for the protein model, using minimum energy confirmation rather than a model at a fixed temperature with a fixed energy state. I-Tassers scoring function *(C-score)* - based on the relative clustering structural density. The consensus significance score of multiple threading templates is presented to evaluate the accuracy of the I-TASSER predictions (Zhang, 2008).

Where M is the multiplicity of structures in the SPICKER cluster

$$\text{C-score} = \ln \left(\frac{M}{M_{\text{tot}}} \cdot \frac{1}{\langle\text{RMSD}\rangle} \cdot \frac{\prod_{i=1}^{4} Z(i)}{\prod_{i=1}^{4} Z_0(i)} \right)$$

Figure 1.3 – C–score equation for I–Tasser homology model quality.

- M_{tot} Is the number of I-Tasser structure decoys in the clustering.
- RMSD is the average Root Mean Square Deviation of the decoys to the cluster centroid.
- Z is the highest Z-score or the energy to mean in the unit of standard deviation of templates given by the PPA threading program.
- Z_0 Is the z-score cut-off specified for distinguishing good and bad templates.

The first two factors of the equation account for the structure convergence caused by SPICKER. This correlates with external restraints and inherent I-Tasser potentials. The third factor accounts for threading quality in alignments. The logarithm in the equation is to adjust C-score values to an even distribution. C-score accounts for confidence of alignment using multiple threading programs rather than one threading program. The equation has been shown to give a strong correlation when used to determine quality of predicted protein models. RMSD can sometimes be low for a structure - but it does not always correlate with quality of the model. The loops of a protein often being the highest degree of error in a digital model. Another issue with homology modelling is sequence identity. If the model is made using nucleotide sequence which has a low identity to known structures of the same type, the degree of error in the prediction process is larger.

Molecular Dynamics.

All-atom MD simulation of membrane proteins provides a computational tool to probe membrane proteins at scales ranging from femtoseconds to nanoseconds (Lindahl and Sansom, 2008). NAMD is a parallel MD code designed for high performance simulation of biological systems which uses a TCL file script (Phillips *et al.*, 2005). This code scales to hundreds of processors and works with the CHARMM forcefield (Brooks *et al.*, 2009), which uses empirical energy functions to model molecular systems and is focused on proteins, peptides, lipids nucleic acids, carbohydrates and small molecule ligands. It can be used with a variety of different canonical ensembles to represent the state of a biological system in thermal equilibrium. NPT is the temperature, pressure and absolute temperature ensemble (McDonald, 1972). It allows variation in volume, but keeps temperature in Kelvin, the number of atoms in the system and the pressure constant following the universal gas law. This ensemble plays an important role in chemistry as chemical reactions are usually carried out under constant pressure condition ((Yang *et al.*, 2014)). Molecules are simulated using physics and chemistry parameters to numerically solve newton's 2nd law of thermodynamics. The Langevin equation is an expansion of Brownian motion mechanics based on newton's second law (Olson, Chaudhury and Lee, 2011), with the addition of common phenomena including friction and random forces – which are not experienced in the vacuum environment of MD calculations (Figure 1.4).

$$M\dot{\upsilon} = F(r) - \gamma\upsilon + \sqrt{\frac{2\gamma k_B T}{M}} R(t),$$

Figure 1.4 – the Langevin equation use to solve newtons second law in CHARMM simulations. The variables are summarized below.

- M = Mass.
- F = Force.
- ɣ = Friction Co-efficient.
- υ = Random forces Co-efficient.
- K_b = The Boltzmann Constant.
- T = Temperature. *(In °Kelvin).*
- R(t) = Univariate Gaussian process.

The Boltzmann-constant is used to determine the thermal equilibrium of a system. Using the kinetic energy of every atom in the system (temperature in $°K$ is $\propto$ to the average kinetic energy in a molecular system).

$$f(\epsilon_k) = \frac{2}{\sqrt{\pi}} \frac{1}{(K_B T)^{\frac{3}{2}}} \sqrt{\epsilon_k} \exp(-\frac{\epsilon_k}{k_B t})$$

Following the equilibration analysis of the system molecular dynamics can be performed to analyse the complex nature of proteins.

Biological macromolecules interacting with various other complexes or small molecules, such as ligands with high specificity and affinity, are the basis of all processes in living organisms. Proteins are a class of molecule which relay messages and biological function by interfacing with other proteins or macromolecules to deliver a signal - providing function or response within the cell. Proteins play a vast majority of roles in the cell, ranging from structural (i.e. microtubules), mechanical (i.e. Muscular), biochemical and hormonal cell signalling. Protein signalling is realized through direct physical interaction via complex forces in its environment.

By quantifying said interactions and understanding molecular recognition between macromolecules - it is possible to make use of the detailed structural data obtained to aid in the process of designing new drugs to improve or alleviate symptoms of disorder or illness by being rationally designed for a dynamic model of a proteins active site. By modelling the Mel1a GPCR and its primary ligand melatonin, a representative descriptive dynamic model of the protein will be obtained with a proposed active site for its primary interacting ligand. No crystallographic structural data for the Mel1a GPCR exists. since no interactive data exists with either its primary ligand or in its biological environment at current, a dynamic model can be build using structural analogues of other GCPR's and dynamic data can be attained through this method.
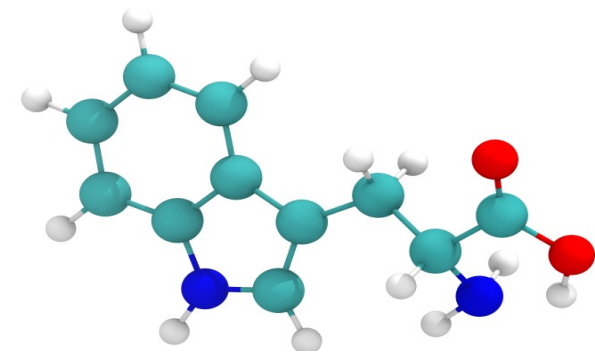
Tryptophan.

Tryptophan 5-monooxygenase, adds a oxygen-hydrogen hydroxyl group to the indole of tryptophan.
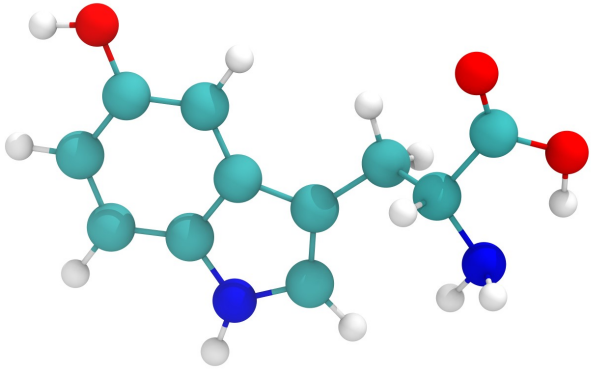
5-Hydroxytryptophan.

L-amino acid carboxylase, catalyses the conversion of 5-hydroxytryptophan into 5-hydroxytryptamine.
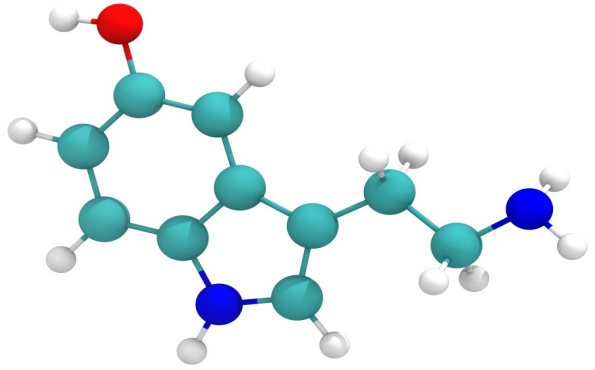
Serotonin.

Arylalkyl amine-N-acetyltransferase acetylates serotonin resulting in the addition of an acetyl group.
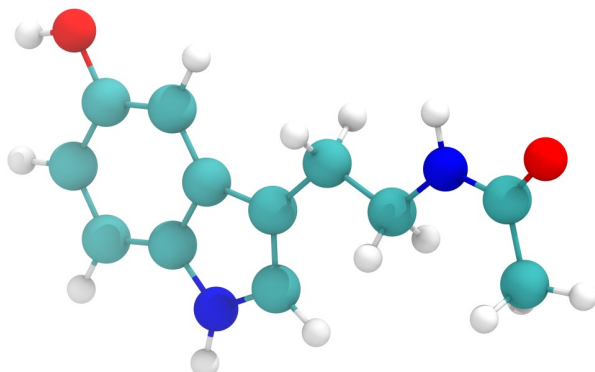
N-acetyl Serotonin.

hydroxy-indole-O-methyltransferase adds the final methyl group to the OH group attached the indole of the molecule.
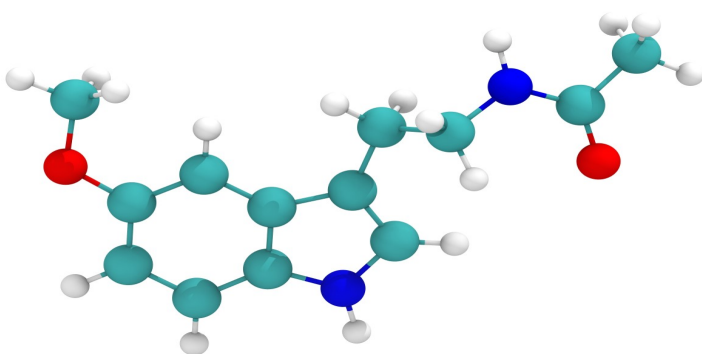
Melatonin.

The N-acetyl-5-methoxytryptamine (melatonin) molecule is synthesised.

Figure 1.1 - Melatonin Synthesis from tryptophan.

Materials and Methods

2.0 Generation of Initial Homology model.

No structural data exists for Mel1a therefore the protein amino acid sequence for the Mel1a G-coupled protein receptor (GCPR) was acquired from Uniprot accession number P48039 and saved as a fasta file (UniProt Consortium, 2018). The protein sequence used is shown below;

```
10          20          30          40          50

MQGNGSALPN  ASQPVLRGDG  ARPSWLASAL  ACVLIFTIVV  DILGNLLVIL

60          70          80          90          100

SVYRNKKLRN  AGNIFVVSLA  VADLVVAIYP  YPLVLMSIFN  NGWNLGYLHC

110         120         130         140         150

QVSGFLMGLS  VIGSIFNITG  IAINRYCYIC  HSLKYDKLYS  SKNSLCYVLL

160         170         180         190         200

IWLLTLAAVL  PNLRAGTLQY  DPRIYSCTFA  QSVSSAYTIA  VVVFHFLVPM

210         220         230         240         250

IIVIFCYLRI  WILVLQVRQR  VKPDRKPKLK  PQDFRNFVTM  FVVFVLFAIC

260         270         280         290         300

WAPLNFIGLA  VASDPASMVP  RIPEWLFVAS  YYMAYFNSCL  NAIIYGLLNQ

310         320         330         340         350

NFRKEYRRII  VSLCTARVFF  VDSSNDVADR  VKWKPSPLMT  NNNVVKVDSV
```

The sequence was then submitted into the online I-Tasser (iterative threading assembly refinement algorithm) server (available at -https://zhanglab.ccmb.med.umich.edu/I-TASSER/) developed by Zhang Lab, to attain an initial homology model (Zhang, 2008). I-Tasser uses the local meta-threading server (LOMETS) program which is locally installed to predict 3-dimensional structures from target-to-template alignments using locally installed threading programs including; FUGUE (Shi, Blundell and Mizuguchi, 2001) and HHSEARCH (Shi, Blundell and Mizuguchi, 2001). The second step excises the relevant fragments from PDB templates and reassembles them into full length models. Loop sections are built by ab-initio modelling. The SPICKER algorithm identifies near-native folds in protein structures by clustering protein structure decoy information on similar family proteins which are generated during computer simulation (Zhang and Skolnick, 2004). The structures with the lowest energy conformation are then selected. Four variants of profile to profile alignment threading

algorithms (PPA) the Hidden Markov (Karplus, Barrett and Hughey, 1998) PSI-BLAST (Altschul *et al.*, 1997), Needleman-Wunsch (Nordström *et al.*, 2011) and Smith-Waterman (Smith and Waterman, 1981) algorithms - are then used by the program to align protein sequences.

Five representative models were produced. The model with the lowest C-score, which is the estimated confidence score in the final model predicted by I-Tasser - is based on the quality of alignment with known family proteins, and convergence parameters of the structure produced. This returned a score of -0.01. A TM-Score (A proposed scale for measuring the similarity between two structures via root mean square deviation (RMSD) which is sensitive to local error) of 0.71 ± 0.11 and an estimated root mean square deviation (RMSD) of $6.5\text{Å} \pm 3.9\text{Å}$ (Figure 2.0.1) was chosen as an initial model for further study. This model was downloaded in PDB format from the online server. System preparation for molecular dynamics was carried out using Visual Molecular Dynamics (VMD) and the programs available in version 1.9.3 (Humphrey, Dalke and Schulten, 1996). The Mel1a homology model was then uploaded into Pymol and the molecular weight was calculated as 34.4 kDA including explicit hydrogen atoms.

A PSF file of the Mel1a GCPR was generated using VMD's Auto-PSF program for utilizing the CHARMM36 forcefield (Huang and MacKerell, 2013) which was sourced from the NamD additive forcefield downloads from the university of Maryland Mackerell lab homepage (available at- http://mackerell.umaryland.edu/charmm_ff.shtml). Using the forcefield the PSF file contains information regarding atom types, bond types, bond length and angle and dihedral and improper angles. The system was then solvated with the TIP3P model available in CHARMM36, with 8Å of H_2O (Hardy *et al.*, 2015). This was applied on all three-dimensional axis following a Boltzmann distribution Using VMD 1.9.3's autosolvate tool. (Dimensions - X-axis: 62Å, Y-axis: 62Å, Z-axis: 100Å). - using the Mel1a GCPR's dimensions as a box size indicator (Figure 2.0.2). The system was then ionized with a neutralizing concentration of 0.15mol/L of NaCl with a minimum distance from solute and between individual ions of 5Å using VMD 1.9.3's autoionize tool. The introduction of the ions was also added as a Boltzmann distribution (Figure 2.0.3). The system is neutralized to enable Particle Mesh Ewald function (Hardy *et al.*, 2015). The solvated and ionized protein was then placed into a 65Å (X + Y axis) 1-palmitoyl,2-oleoyl-sn-glycero-3-phosphocholine (POPC) bi-lipid membrane compatible with the CHARMM36 forcefield (Hardy *et al.*, 2015). The POPC membrane was aligned with the Mel1a Protein structure on the Z-axis. TIP3P water and POPC lipid molecules which were

closer to the protein or lipid-bilayer than 0.6Å were then removed using the VMD Tk Console. A final PDB and PSF set were then saved as the final system for molecular dynamics (Figure 2.0.4).

For minimization, heating and equilibration harmonic constraints to the carbon backbone of the protein (using the carbon-α atoms in each amino acid), were defined. A constraint was also applied to the phosphorus atoms of the bilipid membrane. In this fashion the side chains of the protein remain mobile allowing it to reach a converged minimum energy state, while keeping the overall orientation of the molecule aligned within the membrane. NAMD harmonic constraints limits the amount of kinetic energy that atoms experience thus preserving the overall conformation of the initial model until equilibration is complete.

2.1 Minimization of the Mel1a GCPR System.

The Mel1a system was subject to conjugate gradient minimization which is an iterative method that optimizes the geometry of all atoms within the system and finds the lowest energy conformation (Hardy *et al.*, 2015). For all NAMD calculations the CHARMM36 forcefield was utilised (Figure 2.1.1). The created system PDB and PSF file types are given to the INP file for the simulation to follow (Figure 2.1.2). The molecular system is run as an NPT ensemble (the number of particles (N), pressure (P) and temperature (T) are constant) at a pressure of 1.01325 bar and a temperature of 310 K.

Group pressure is responsible for allowing NAMD to calculate the pressure of the system using the kinetic energy term. *Langevinpiston* was set to on (Hardy *et al.*, 2015), and the following variables were used: *langevinPistonTarget* set to 200fs piston oscillation period, *LangevinPistonDecay* was set to 50fs (it is intentionally set smaller than the oscillation period to allow a larger degree random forces) and *LangvinPistonTemp* is set to 310 K (Figure.2.1.3). Timestep is set to 2 fs, *NonbondedFreq* is set to 2. (this command specifies how often a full electrostatics evaluation is performed), *fullElectFrequency* command dictates the amount of timesteps in fs between a full electrostatics evaluation and is set to 4 fs. *stepspercycle* commands the number of steps between each cycle between atom reassignments 20 steps are assigned (Figure 2.1.4). *rigidBonds* are set to water molecules which ignores the angles and bond lengths of the surrounding water molecules. In the CHARMM36 TIP3 model water is optimised to have fixed angle and bond lengths. *Cutoff* informs the system of the maximum distance between the furthest interacting surface of the protein and an exogenous molecule interaction should be recorded in terms of van der-Waal interaction and electrostatic potential (it is set to 12 Å), *switchDist* specifies the cutoff distance for calculations in Å, and *margin*

discerns the amount of space the system can expand or contract within. *Cell origin*, and *cellbasisvectors* are information given from a script file which generates the starting points for each axis of the system. *PMEGridSpacing* is set to 1 which activates Particle Mesh Ewald for electrostatics calculations in periodic systems (Figure 2.1.5). Langevin Variables are set to on. *Langevin Dampening* is set to 5 (the maximum friction and random forces co-efficient). Langevin for hydrogen molecules is off, preventing the system from using the Langevin algorithm for those atoms. *LangevinTemp* set to 310 K. Hydrogen molecules are set to rigid bonding in order to reduce computational load and assist in geometry optimization (Figure 2.1.6). Minimization steps are set to 2500, however the total steps are 2500 x 2 = 5000 due to the record step being every 2 fs. (Figure 2.1.7). All Calculations were performed on a 56 core Intel Xeon CPU E7-4850 v3 2.20GHz Linux server.

2.2 Heating of the Previously Minimized Mel1a GCPR system.

Once minimization was completed the Mel1a GCPR system was heated to bring the temperature of the system to 310 K. Initial co-ordinate and extended system files, created in the minimization calculation were used, as an input for geometry and call dimensions prior to system heating. Input file settings were similar to those used for minimization with a few exceptions specific for heating. *BinCoordinates* were specified as the optimized geometry co-ordinates derived from minimization. *Cellorigin* and *cellbasisvector* 1, 2 and 3 are hashed out and replaced with reference to the extended system file which provides this information from the minimization step. *Cutoff* for recorded interactions remain the same as do PME and PME gridspacing. Variables *WrapAll* and *Wrapwater* signify that water/ion molecules which fall outside the periodic boundary are not to be translated to the opposite side of the water box during heating. (Figure 2.2.1). *Langevin Dampening* coefficient was reduced from 5 to 1 1/ps (Lowering the friction and random forces co-efficient) the other Langevin parameters remained the same. Harmonic constraints were used as before in minimization. *Langevinpistonperiod*, *langevindecay and Langevinpistontarget* values were maintained identically to the minimization simulation. *Useflexiblecell* is set to "no" preventing the system from creating a 3-orthogonal dimensional system to fluctuate independently. *Useconstantarea* is set to no – allowing the system to fluctuate on all 3 axes. If it were enabled it would prevent x and y axis fluctuations, while allowing z-axis fluctuations. *Usegrouppressure* is enabled to allow the pressure to be calculated via the SHAKE method. (Figure 2.2.2). The initial temperature of the system was set to 0 degrees Kelvin and increased by an increment of 0.001 K per timestep

(Figure 2.2.3). The amount of timesteps was set to a total of 310,000.

2.3 Initial Equilibration of the Mel1a GCPR System.

An input file for equilibration of the Mel1a GCPR system was written with data gathered during the heating phase: coor (binary co-ordinates), xsc (extended system) and vel (velocities) files were referenced as the starting point for the equilibration step. The temperature of the system was set to 310 K (Figure 2.3.1). Harmonic constraints were also applied as used in the previous stages (Figure 2.2.1). Forcefield parameters files remain the same as the prior simulations using the CHARMM36 the amount of timesteps was increased to 5,000,000, limiting the equilibration time to 10ns (Figure 2.3.2).

2.4 Alternative Equilibration Simulation – ReassignFreq.

A second equilibration input file was created with some extra variables to try and refine the data to a narrower distribution of temperature. Using the algorithm *reassignFreq* – this is used to reassign the temperature of the entire system to a specified number. The algorithm randomly reassigns velocities within the system following a Maxwell distribution to the correct temperature. In this case 310 K for every timestep of the simulation. *reassigntemp* gives the temperature to reassign the velocities to which is set to 310K. *reassignhold* specifies the temperature to hold the simulation at (if the system deviates from that temperature in any step, It gets set back to 310 degrees Kelvin) *reassignincr* is set to its default value of 0, this disables the simulation from using simulated annealing or other slow cooling and heating methods specifying that *reassigntemp* and *reassignhold* are absolute values to be operated (Figure 2.4.1). The *vel* (velocity), *coor* (binary co-ordinates) and *xsc* (extended system) files from the heating run (Section 2.2.) performed before the first equilibration were sourced as initial data in the same fashion rather than continuing chronologically from the end of the initial equilibration. The same Mel1a GCPR system PDB and PSF files from system setup section 2.0 were sourced for this equilibration run. Harmonic constraints were utilized as previously described in figure 2.2.1. The calculation was run for 5,000,000 steps with a timestep value of 2 fs.

2.5 Alternative Equilibration Simulation – RescaleFreq.

A third equilibration calculation was undertaken using the variables for rescale frequency to relate the distribution of energy in the system in comparison to the previous two simulations. The equilibration was set to run for the same amount of time (10ns) as the previous 2 simulations following on from the same heating data obtained prior to the first equilibration run (shown in figure 2.2.1). The initial system PDB and PSF were used (section 2.0), and the initial harmonic constraints PDB with the Mel1a GCPR residues and bi-lipid phosphorus atoms limited in the same way (figure 2.2.1). *RescaleFreq* is similar to *reassignfreq* but rather than forcing the system to be at a constant temperature of 310 Kelvin, it allows deviation in temperature. All the velocities in the system are multiplied by the same value and chosen to shift the average kinetic energy so it corresponds to the correct temperature defined in the input file. Both algorithms for temperature control are mutually exclusive in their action. *resecaleFreq* determines how often the distribution of energies are rescaled to a specified value. *rescaletemp* defines the temperature to which the system is rescaled to.

2.6 10ns Molecular Dynamics Runs of the Rescale Equilibrated Mel1a GCPR system.

From the equilibration data, the rescale equilibrium data was used as it gave the best fit distribution of heat and velocity of the three types of simulation. The Molecular dynamics (MD) simulations were set up in 3 new input files. Labelled MD run1, run2, and run3 respectively (generating a triplicate dataset). The input for the three runs were identical in detail. The vel (velocity), coor (binary co-ordinates) and xsc (extended system) data was taken from the rescale equilibrium data and referenced as the starting point for the runs, harmonic constraints were hashed out to nullify them allowing the simulation freedom to explore conformational space. PDB and PSF inputs remained the same as section 2.0 and the variables for rescale were removed. The starting temperature was set to 310 degrees Kelvin (Figure 2.6.1). Timesteps remained totalling 10 nanoseconds duration for the simulation. (5,000,000 x 2 = 10,000,000) (Figure 2.6.2). langevinDampening was increased to 2 (Figure 2.6.2) following issues with simulation start. Some atoms within the simulation were "moving too fast" as stated by the CHARMM runtime environment. This is due to the FFT (Fast Fourier Transform) not being able to make a logical velocity path interpretation on the movement of atoms. Increasing the friction in the system slows these atoms to readable levels by the FFT algorithm as it is only recording once every 2 femtoseconds. Rigid bonds are turned off on the protein model allowing attached hydrogens to move free of constraint. Constraints persist on the water molecules

18

(TIP3P) in the environment.

2.7 CGENFF Melatonin Parameterization and Minimization.

The melatonin ligand was created in Avogadro (Hanwell *et al.*, 2012) using the crystal information from (Wakahara, Fujiwara and Tomita, 1972) and saved as a Mol2 file (Figure2.7.1). New folders were set up on the server for the ligand file to be placed with the appropriate files for parameterization using CHARMM General Forcefield (CGENFF) (Vanommeslaeghe *et al.*, 2010). MOL2CRD script (figure 2.7.2). was also placed in the file along with an input file written instructing the charm forcefield to use the generated melatonin mol2 file (2.7.3) to create parameters.

The output log file containing structural calculations (dihedrals, bond angles and lengths, atom-types parameterized by CGENFF) for the creation of melatonin is shown in figures 2.7.4 to 2.7.10.

2.8 Docking of CGenFF Parameterized Melatonin to MTNR1A G-coupled Protein Receptor.

The Mel1a GCPR co-ordinates from the final frame of each 30ns molecular dynamics run, were extracted from the system as a new PDB set of co-ordinates utilizing VMD 1.9.3. The initial homology PDB was also used for docking solutions using Chimera version 1.12 (Pettersen *et al.*, 2004) with the Autodock Vina (Trott and Olson, 2010) plugin for docking surface analysis .

A 3-dimensional box with dimensions X: 25Å Y: 25Å, by Z: 25Å which is shown as a green outline (figure 2.8.1) was placed over the extracellular portion of the protein which was placed in a vacuum. The parameters for docking were left at default. With the default options hydrogens are added to the structure and charges were merged. Non-polar charges were removed. No non-standard residues were present apart from the ligand – ignore non-standard residues was set to "false". The amount of binding modes was specified as 10 (ten docking positions) exhaustiveness of search and maximum energy difference were set to maximum values (exhaustiveness of search:8, Maximum energy difference 3). Docking was performed on the provided Opal Web service (Huang *et al.*, 2014) made available through the Chimera client autoDock Vina plugin to produce 10 solutions of the ligand in various combinations of co-ordinates for each PDB file (run 1,2,3 and homology). This data was saved and compiled into one PDBQT file for each of the different final frames of the previously mentioned protein simulations. From these solutions the best fitting solution was chosen for all protein frames (Figure 2.8.2). The final frame of the 30ns control structure run 2 was chosen for docking, as it had the lowest protein backbone RMSD. The co-ordinates of the ligand were then added into the final frame of 30ns molecular dynamics run 2. Necessary TIP3P water molecules within 2.5Å of the ligand were removed using the representation tab in VMD selection tools highlighting the entire MEL1a GCPR and bi-lipid membrane system, minus the index numbers of the water molecule atoms to be deleted. A new system PDB was saved to once all the molecules were removed.

The QM parameterized ligand geometry optimized PDB for the melatonin molecule were then edited to have the same co-ordinates as the docking solution - while keeping the correct topology attained through Gaussian16 (section 3.0). The optimized charges PSF and modified co-ordinate PDB generated from the quantum parameterization were merged with the system using VMD1.9.3's "Merge Structures". This compiles the co-ordinates of the Mel1a GCPR bi-

lipid membrane system PDB and PSF (topology information generated by the CHARMM forcefield though the FFtk plugin in VMD 1.9.3) with the ligands PDB and PSF structures which results in a complete molecular system ready for minimization.

2.9 30ns Molecular Dynamic Simulation of the Mel1a GCPR System.

Upon review of the data for the 10ns MD simulations, it was decided to extend the runs to 30 nanoseconds to give the system more time to search conformational space. An attempt to extend the simulations using the original 10 nanosecond runs was bid but failed due to atoms moving too fast similarly to the rescale and reassign equilibration runs this is due to the RATTLE algorithm (Andersen, 1983) which is used under the NPT ensemble to integrate the equations of motion in MD calculations. Like the SHAKE (Kräutler, Van Gunsteren and Hünenberger, 2001) algorithm it is used to verify the co-ordinates and the velocities of a molecule or system follow the constraints of the forcefield at each timestep. Instead, A new set of input files were created for fresh 30 nanosecond runs using the rescale equilibration data as a starting point (Figure 2.9.1) Allowing NAMD to record every 1 femtosecond of the simulation allowing for a larger degree of calculations at each timestep (Figure 2.9.2). *LangevinDampening* was set to 1 due to the change in the timestep variable as the random forces and friction co-efficient was increased previously to allow SHAKE and RATTLE to be able to calculate velocities and co-ordinates properly. The amount of timesteps was doubled to 30,000,000 (30ns) as data is now recorded every single femtosecond (Figure 2.9.3). Triplicate input files of the 30ns input files were created and placed into respective folders to generate triplicate data and was performed on the server identified previously.

2.10 Minimization of the Melatonin-Mel1a GCPR Docking System.

Due to the removal of water molecules which were closer than 1.5Å to melatonin it was necessary to repeat minimisation, heating and equilibration of the system prior to molecular dynamic calculations. Conjugant gradient minimization was performed using the NAMD input file variables as used for the initial molecular dynamics system without melatonin (Figure 2.10.1) however the custom parameter file for the melatonin molecule (made in section 3.0) with the associated PDB and PSF files defined for the system was used. The melatonin parameter file contains a novel set of parameters specifically calculated for melatonin (Section 3.0). A new set of harmonic constraints were also generated for the system. This was applied to the Carbon-α backbone of the Mel1a GPCR and bi-lipid membrane phosphorus atoms in a similar fashion to the control simulations (section 2.1 – Minimization of the Mel1a GPCR System.). This set of constraint PDB's is shown in figure 2.10.2, Minimisation was run for 2500 steps.

23

2.11 Heating of the Docked Mel1a-Melatonin System.

Similarly to the initial heating of the system minus melatonin, the same timesteps and parameters for the simulation were retained (section 2.2). The coor (binary co-ordinates) and xsc (extended system) from the conjugant gradient minimization from section 2.10 of the Mel1a-melatonin docking system were sourced in the heating input file for NAMD (Figure 2.11.1). Heating was carried out over 310,000 steps – in order to reach the specified temperature, of 310K, by increasing temperature at 0.001 kelvin per timestep (1 kelvin per 1000 steps).

2.12 Rescale Equilibration of the Mel1a-Melatonin Docking System.

Heating binary co-ordinates, velocity data and extended system data from the heating phase (section 3.2) was entered into a new input file for equilibration with the rescale algorithm which was used for the control calculations (Figure 2.12.1) where rescaling frequency was performed every timestep at a rescaleTemp of 310º K. This was performed using the same variables as the initial rescale equilibration (section 2.5). The docking system harmonic constraints mentioned in the previous sections 2.11 and 2.10, were also re-utilized. The only difference was the simulation time, which was run for 2,500,000 timesteps (5 ns at 2 fs per calculation step). The data was analysed using VMD1.9.3 to check for thermodynamic stability of the protein within system by Maxwell-Boltzmann distribution of velocity data and the NAMD plot analysis of the log file created in the calculation, before entering triplicate 30 ns molecular dynamic simulations.

2.13 30ns molecular dynamics simulation of Mel1a-Melatonin Docking System.

Three trajectories for the Mel1a/Melatonin docking system were carried out (adhering to the original parameters for the 30 ns control calculations). Timestep was set to 1 fs and harmonic constraints were disabled for the system. The langevin parameters were identical to the 30ns control runs were *langevinTemp* was set to 310K, *LangevinHydrogen* was set to off, and *langevinDampening* was set to 1. (Figure 2.13.1). The QM generated parameter file for melatonin was for all three MD simulations and rigid bonds were set to water only (Figure 2.13.2). Triplicate runs were performed with the *coor* (binary co-ordinates), *vel* (velocity) and *xsc* (extended system) files created in the rescale equilibration of the Mel1a/melatonin docking.
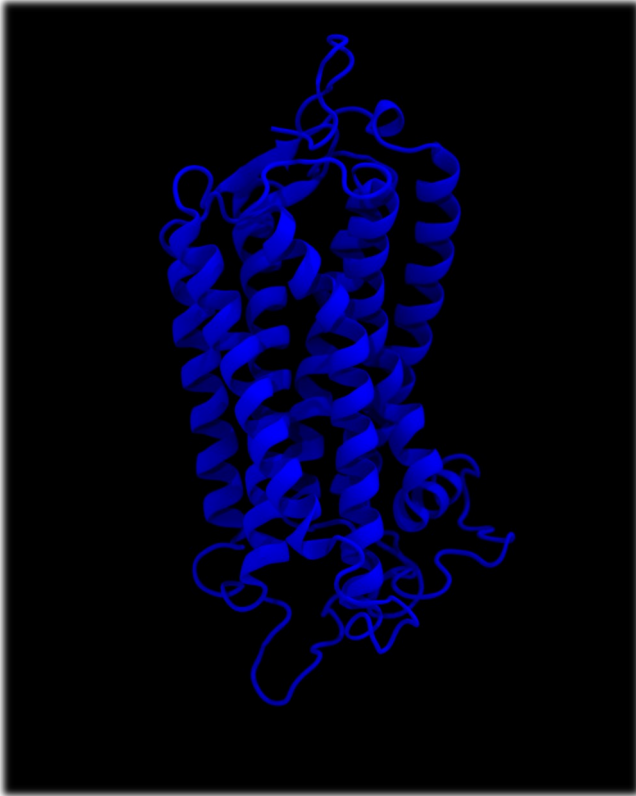
Figure 2.0.1. Mel1a Homology model generated in the I-Tasser online web server. Shown in Blue, new cartoon representation in VMD.
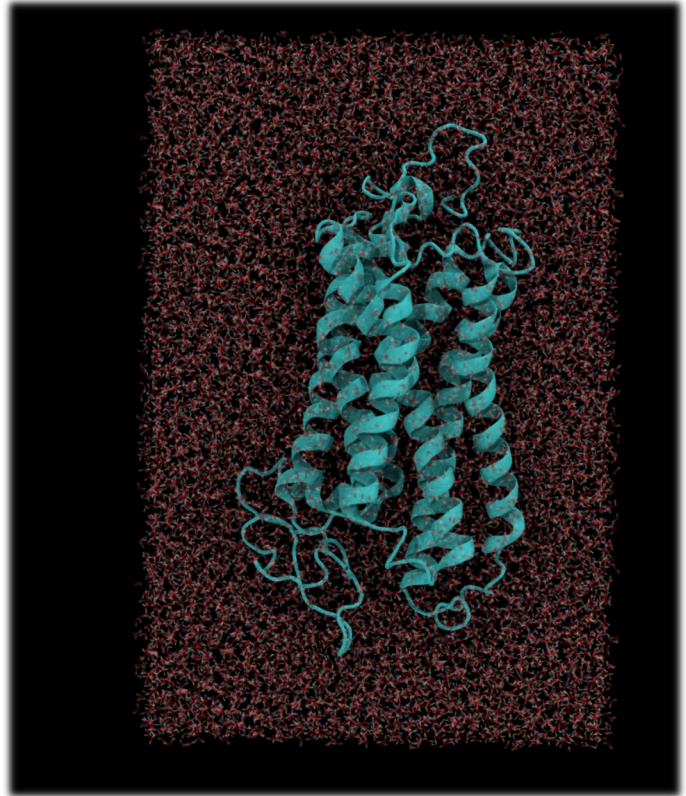

*Figure 2.0.2. Mel1a Homology model solvated in TIP3P water molecules. (*Dimensions - X-axis: 62Å, Y-axis: 62Å, Z-axis: 100Å). Protein shown in VMD - representation New Cartoon, Coloured in Cyan. Water is shown in representation liquorice.
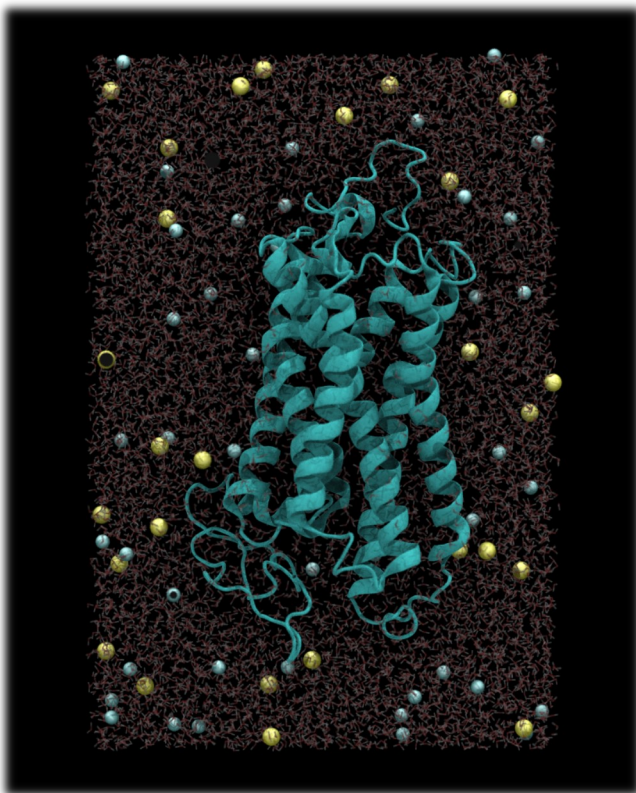

Figure 2.0.3. Mel1a Homology model solvated and ionized in TIP3P water molecules. Shown in VMD Protein representation New Cartoon in Cyan. TIP3P (water) Shown in liquorice, Ions shown in VDW Chlorine in blue, Sodium in yellow.
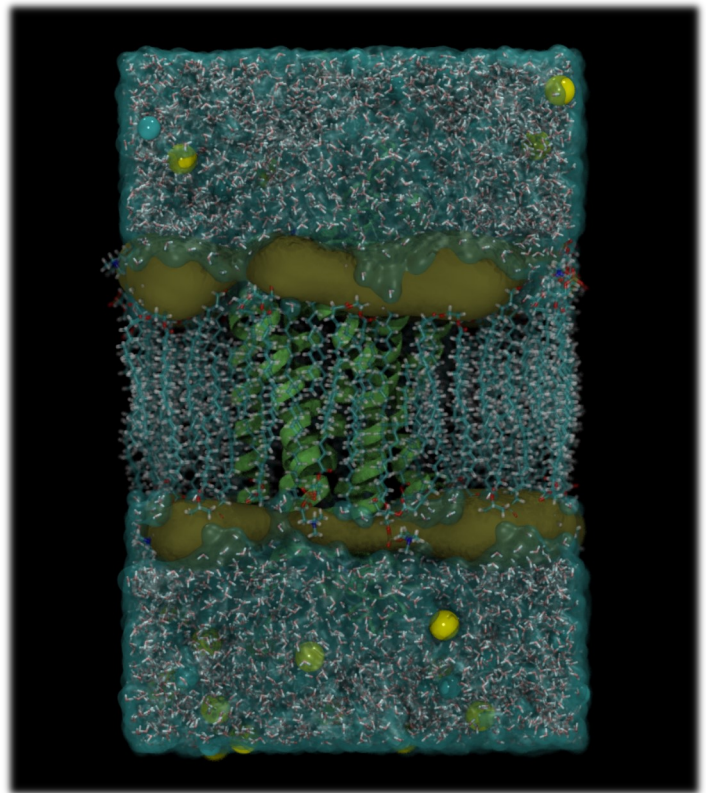

Figure 2.0.4, Homology model of protein in POPC bi-lipid membrane. Water represented in a Quicksurf transparent overlay coloured cyan and internal water represented in liquorice. Lipids represented in liquorice. Protein represented new cartoon, coloured in Green. Ions shown in VDW (Chlorine blue. Sodium, yellow). Phosphorus atoms shown in surf, coloured in Tan.

```
################################################################
## MINIMIZATION Mel1a GCPR, first run                        ##
################################################################
## FORCEFIELD PARAMATERS                                     ##
################################################################

paratypecharmm      on
parameters          /home/dc392/toppar/par_all36_prot.prm
parameters          /home/dc392/toppar/par_all36_na.prm
parameters          /home/dc392/toppar/par_all36_carb.prm
parameters          /home/dc392/toppar/par_all36_cgenff.prm
parameters          /home/dc392/toppar/par_all36_lipid.prm
parameters          /home/dc392/toppar/toppar_water_ions_namd.str
```

Figure 2.1.1. The Toppar directory contains molecule specific parameters for the system to follow during geometry optimization (Obtained with the NamD 2.12 additive force field files). NamD will not run a simulation unless it has all the required parameters for each molecule type.

```
################################################################
## STRUCTURE AND COORDINATES                                 ##
################################################################
structure           /home/dc392/structure/Mel1a_final.psf
coordinates         /home/dc392/coord/Mel1a_final.pdb


################################################################
## TEMPERATURE                                               ##
################################################################
temperature         310
```

Figure 2.1.2. PDB and PSF structures were referenced in the input file for minimization. The temperature for the minimization was set to 310 Kelvin.

```
################################################################
## CONSTANT PRESSURE CONTROL (VAR VOLUME)                    ##
################################################################
useGroupPressure        yes
useFlexibleCell         no
useConstantArea         no
langevinPiston          on
langevinPistonTarget    1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod    200.0
langevinPistonDecay     50.0
langevinPistonTemp      310
```

Figure.2.1.3 Group pressure is responsible for allowing NamD to calculate the pressure of the system using the kinetic energy taken from the system. useFlexibleCell and useConstantArea were not used for system minimization.
Langevinpiston is set to on, and the following variables were used, langevinPistonTarget set to 200fs (femtoseconds) as a piston oscillation period, LangevinPistonDecay was set to 50fs, it is intentionally set smaller than the oscillation period to allow a larger degree random forces. LangvinPistonTemp is set to 310 Kelvin, as a base figure to keep the pressure of the system adhered to.

```
#################################################################
## INTEGRATORS                                                 ##
#################################################################
timestep              2
nonbondedFreq         2
fullElectFrequency    4
stepspercycle         20
```

Figure 2.1.4 - timestep is set to 2, this instructs NamD to record data every 2fs of the simulation. NonbondedFreq is set to 2. – this command specifies how often a full electrostatics evaluation is performed. fullElectFrequency this command dictates the amount of timesteps in fs between a full electrostatics evaluation and is set to 4. stepspercycle commands the number of steps between each cycle between atom reassignments 20 steps are assigned.

```
#################################################################
## APPROXIMATIONS                                              ##
#################################################################
rigidBonds          water
rigidTolerance      0.00000001
cutoff              12
switching           on
switchdist          10
pairlistdist        14
margin              3
exclude             scaled1-4
1-4scaling          1.0    # 1.0 for Charmm, 0.833333 for Amber
PME                 on
cellOrigin          -0.049795910716056824 0.030537299811840057 0.23735018074512482
cellBasisVector1    73.94400024414063 0 0
cellBasisVector2    0 72.12999725341797 0
cellBasisVector3    0 0 103.47800064086914
PMEGridSpacing      1
```

Figure 2.1.4. - rigidBonds are set to water molecules which tells the program to ignore the angles and bond lengths of the surrounding water molecules (TIP3P). Cutoff informs the system of the maximum distance between the furthest interacting surface of the protein and an exogenous molecule interaction should be recorded in terms of Van der-Waal interaction and electrostatic potential. – it is set to 12 Angstroms. switchDist specifies the cutoff distance for calculations. Margin discerns the amount of space the system can expand or contract within. Cell origin, and cellbasisvectors are information given from a script file which generates the starting points for each axis of the system. PMEGridSpacing is set to 1 which activates Particle Mesh Ewald for electrostatics calculations in periodic systems.

```
#################################################################
## TEMP AMD PRESSURE COUPLING                                  ##
#################################################################
langevin              on
langevinTemp          310
langevinDamping       5
langevinHydrogen      off
```

Figure 2.1.5 —Langevin Variables are set to on. Langevin Dampening is set to 5—the maximum friction and random forces coefficient. Langevin for hydrogen molecules is off, preventing the system from using the Langevin algorithm for those atoms.Langevin variables were switched to on. - langevinTemp set to 310 Kelvin. Hydrogen molecules are set to rigid bonding. To reduce computational load and assist in geometry optimization.

```
#################################################################
## EXECUTION SCRIPT                                            ##
#################################################################
minimize 2500
```

Figure 2.1.6 – Minimization steps are set to 2500, however the total steps are 2500 x 2 = 5000 due to the record step being every 2 femtoseconds.

```
################################################################
## HEATING OF mel1a constraints 1                            ##
################################################################
## CONSTANTS                                                  ##
################################################################


cutoff 12.0
pairlistdist 14.0
switching on
switchdist 10.0
PME on
PMEGridspacing 1
wrapAll on
wrapWater on

##System definition##

coordinates /home/dc392/coord/Mel1a_final.pdb
structure   /home/dc392/structure/Mel1a_final.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT.

binCoordinates /home/dc392/min/minimise1/mel1a_eqi_min1.dcd.restart.coor
#binVelocities  /home/dc392/min/minimise1/mel1a_eqi_min1.dcd.restart.vel

extendedSystem /home/dc392/min/minimise1/mel1a_eqi_min1.dcd.restart.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED

#cellOrigin         -0.049795910716056824 0.030537299811840057 0.23735018074512482
#cellBasisVector1   73.94400024414063 0 0
#cellBasisVector2   0 72.12999725341797 0
#cellBasisVector3   0 0 103.47800064086914

##Conditions##


##harmonic constraints##
constraints on
consref    /home/dc392/par/Mel1a_firstrun_constraints.pdb
conskfile /home/dc392/par/Mel1a_firstrun_constraints.pdb
constraintScaling 2
consexp   2
conskcol B
```

Figure 2.2.1. - .PDB and .PSF files remain the same for the heating phase. binCoordinates are added to the system to give the system a new set of optimized geometry co-ordinates. Velocities from the minimization simulation are not carried across and are hashed out in the input as they offer no beneficial figures in minimization. Cellorigin and cellbasisvector 1 2 and 3 are also hashed out as they are no longer used due to the change in positioning of atoms during the system during minimization. Cutoff for recorded interactions remain the same as do PME and PME gridspacing. Variables WrapAll and Wrapwater signify that water/ion molecules which fall outside the periodic boundary are not to be translated to the opposite side of the water box during heating.

```
##Conditions##


##harmonic constraints##
constraints on
consref    /home/dc392/par/Mel1a_firstrun_constraints.pdb
conskfile /home/dc392/par/Mel1a_firstrun_constraints.pdb
constraintScaling 2
consexp   2
conskcol B

##Output Parameters##

binaryoutput yes
outputname /home/dc392/inp/heat/mel1a_eqi_heat1
outputenergies 500
outputtiming 500
outputpressure 500
binaryrestart yes
dcdfile        /home/dc392/inp/heat/mel1a_eqi_heat1.dcd
dcdfreq 1000
XSTFreq 1000
restartfreq 1000
restartname /home/dc392/inp/heat/mel1a_eqi_heat1.restart

##Thermostat Parameters##
langevin on
langevintemp 60
langevinHydrogen     off
langevindamping 1

##Barostat Parameters##

usegrouppressure yes
useflexiblecell no
useConstantArea no
langevinpiston on
langevinpistontarget 1.01325
langevinpistonperiod 200
langevinpistondecay 50
langevinpistontemp 60
```

Figure 2.2.2.— constraints, langevin parameters and output parameters for heating

```
##################################################################
# MD Protocol                                                    #
##################################################################

seed                4307
temperature         0
reassignFreq        1
reassignIncr        0.001
reassignHold        310
```

Figure 2.2.3 - Temperature parameters . Using a random seed, temperature is started at 0 degrees kelvin and increased by an increment of 0.001 Kelvin per timestep. 310,000 steps are required to reach 310 Degrees Kelvin.

```
##########################################################
## EQUILIBRATION OF Mel1a in membrane                   ##
##########################################################
## CONSTANTS                                             ##
##########################################################

cutoff 12.0
pairlistdist 14.0
switching on
switchdist 10.0
PME on
PMEGridspacing 1
wrapAll on
wrapWater on

##System definition##

coordinates          /home/dc392/coord/mel1a_final.pdb
structure            /home/dc392/structure/mel1a_final.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT.

binCoordinates /home/dc392/inp/heat/mel1a_eqi_heat1.coor
binVelocities  /home/dc392/inp/heat/mel1a_eqi_heat1.vel


extendedSystem /home/dc392/inp/eqi/mel1a_eqi_heat1.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED


##Conditions##
#temperature 310

##harmonic constraints##
constraints on
consref    /home/dc392/par/Mel1a_firstrun_constraints.pdb
conskfile /home/dc392/par/Mel1a_firstrun_constraints.pdb
constraintScaling 2
consexp   2
conskcol B
```

Figure 2.3.1. – Structure and co-ordinate files remain the same using the PDB and PSF files created in VMD. binCoordinates, BinVelocities and extendedSystem draw the data created in the heating simulation for use in the equilibration simulation. The temperature of the system is set to 310 Kelvin, harmonic constraints are upheld for the equilibration simulation.

```
##Force Field Parameters##

paratypecharmm on
parameters           /home/dc392/toppar/par_all36_prot.prm
parameters           /home/dc392/toppar/par_all36_na.prm
parameters           /home/dc392/toppar/par_all36_carb.prm
parameters           /home/dc392/toppar/par_all36_cgenff.prm
parameters           /home/dc392/toppar/par_all36_lipid.prm
parameters           /home/dc392/toppar/toppar_water_ions_namd.str
exclude scaled1-4
1-4scaling 1.0|
rigidbonds all

#Implicit Solvent Parameters

#gbis off
#alphaCutoff           14.0
#ionConcentration      0.15

##Script##

run 5000000
```

Figure 2.3.2 - Forcefield parameters used for the system in minimization, heating and equilibration rigid bonds for the system are set to on, restricting movement of the hydrogen atoms in the simulation. The amount of timesteps in increased to 5,000,000 to reach 10 nanoseconds of time (2 x 5,000,000 = 10,000,000).

```
##System definition##

coordinates          /home/dc392/coord/mel1a_final.pdb
structure            /home/dc392/structure/mel1a_final.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT.

binCoordinates /home/dc392/inp/heat/mel1a_eqi_heat1.coor
binVelocities  /home/dc392/inp/heat/mel1a_eqi_heat1.vel


extendedSystem /home/dc392/inp/heat/mel1a_eqi_heat1.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED


##Conditions##
#temperature 310

##reassignment parameters##
reassignfreq 1
reassigntemp 310
reassignhold 310
reassignincr 0
```

Figure 2.4.1. - Reassign equilibration input file details.

```
##System definition##

coordinates         /home/dc392/coord/mel1a_final.pdb
structure           /home/dc392/structure/mel1a_final.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT.

binCoordinates /home/dc392/inp/heat/mel1a_eqi_heat1.coor
binVelocities  /home/dc392/inp/heat/mel1a_eqi_heat1.vel


extendedSystem /home/dc392/inp/heat/mel1a_eqi_heat1.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED


##Conditions##
#temperature 310|

##rescale parameters##
rescaleFreq 1
rescaleTemp 310
margin 1.5

##harmonic constraints##
constraints on
consref    /home/dc392/par/Mel1a_firstrun_constraints.pdb
conskfile /home/dc392/par/Mel1a_firstrun_constraints.pdb
constraintScaling 2
consexp   2
conskcol B
```

Figure 2.5.1. Input file for the rescale equilibration run, variables remain similar to the first two equilibration runs but reassign variables are removed and replaced with rescale parameters.

Figure 2.6.1- Parameters for Molecular dynamics simulations at 10 nanoseconds.

```
##Output Parameters##

binaryoutput yes
outputname mel1a_MD_10ns_refreq_control
outputenergies 500
outputtiming 500
outputpressure 500
binaryrestart yes
dcdfile mel1a_MD_10ns_refreq_control.dcd
dcdfreq 2000
XSTFreq 2000
restartfreq 2000
restartname mel1a_MD_10ns_refreq_control.restart

##Thermostat Parameters##
langevin on
langevintemp 310
langevinHydrogen     off
langevindamping 2

##Barostat Parameters##

usegrouppressure yes
useflexiblecell no
useConstantArea no
langevinpiston on
langevinpistontarget 1.01325|
langevinpistonperiod 200
langevinpistondecay 100
langevinpistontemp 310

##Integrator Parameters##

timestep 2
firstTimestep 0
fullElectFrequency 2
nonbondedfreq 1

##Force Field Parameters##

paratypecharmm on
parameters          /home/dc392/toppar/par_all36_prot.prm
parameters          /home/dc392/toppar/par_all36_na.prm
parameters          /home/dc392/toppar/par_all36_carb.prm
parameters          /home/dc392/toppar/par_all36_cgenff.prm
parameters          /home/dc392/toppar/par_all36_lipid.prm
parameters          /home/dc392/toppar/toppar_water_ions_namd.str
exclude scaled1-4
1-4scaling 1.0
rigidbonds water

#Implicit Solvent Parameters

#gbis off
#alphaCutoff         14.0
#ionConcentration    0.15

##Script##

run 5000000
```

Figure 2.6.2. - Input variables for 10ns calculation of Mel1a GCPR bi-lipid system.
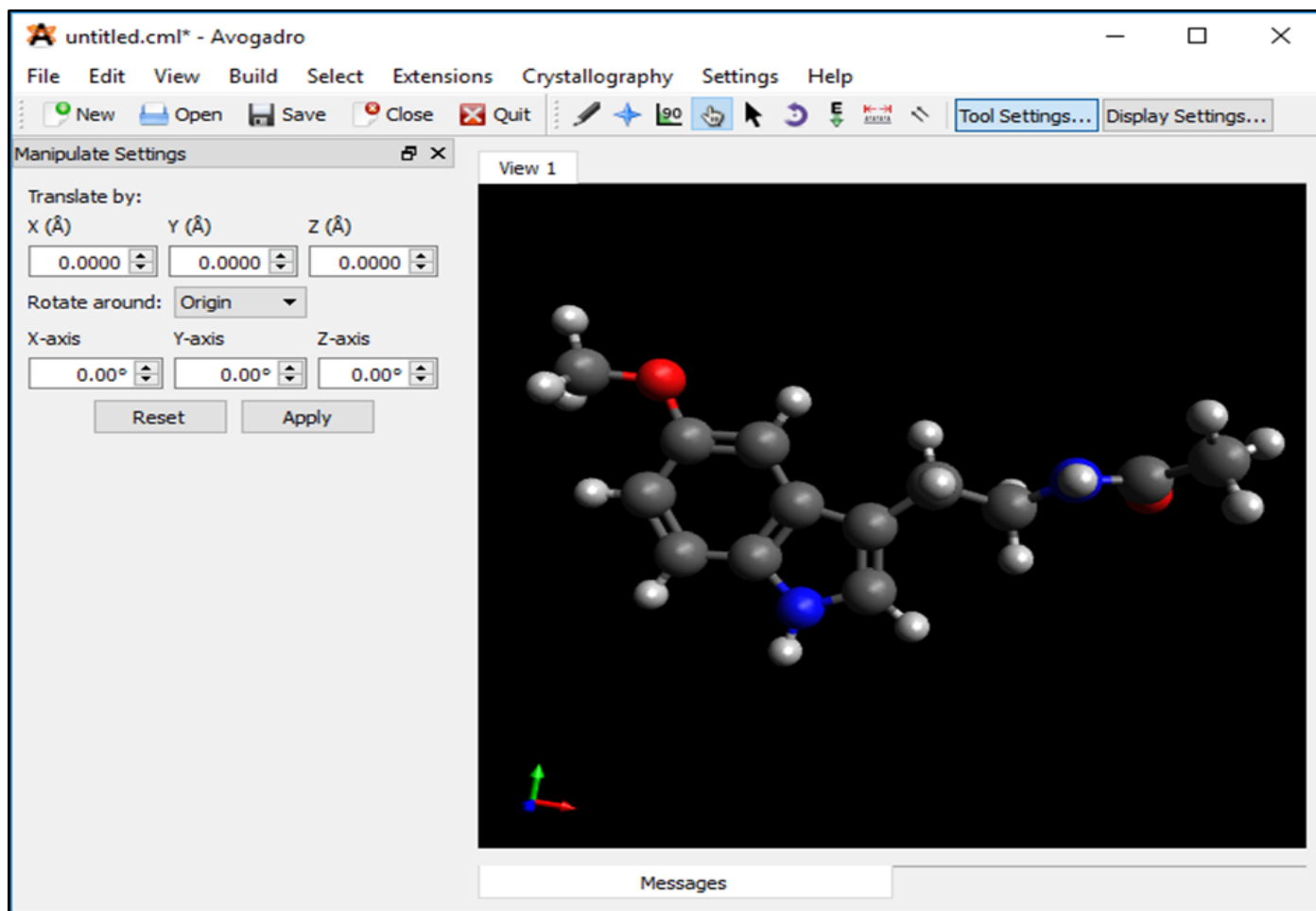
Figure 2.7.1 – melatonin created in Avogadro following the crystal structure attained from (reference) saved in .mol2 format for use with the CGenFF forcefield for minimization and parameterization.

```sh
#!/bin/sh
if [ $# -ne 3 ] ; then
  echo "Usage: `basename $0` <template crd> <mol2 file> <output crd>"
  exit 1
fi

awk -v crdin=$1 '
    BEGIN{while (getline < crdin && substr($0,1,1)=="*");
      print}
    /^@<TRIPOS>ATOM/{cpy=1;getline}
    (cpy && NF>5){x=$3;y=$4;z=$5;getline < crdin;
      printf("%10i%10i  %-10s%-8s%20.10f%20.10f%20.10f  %-5s%6i        %20.10f\n",
        $1,$2,$3,$4,x,y,z,$8,$9,$10)}
    /^@<TRIPOS>BOND/{exit}' $2 > $3
```

Figure 2.7.2. MOL2CRD file contents used to create the parameters through GCenFF. Gives the usage command for CGenFF this contains the Melatonin.inp, Melatonin.mol2 file and an output name which is specified in the command.

```
* Structure Optimization
* Needs external variable "resi"
*

ioformat extended

read rtf card name top_all36_cgenff.rtf
read para card flex name par_all36_cgenff.prm
stream @resi.str

read sequence @resi 1
bomlev -1 ! for 3-membered rings
generate @resi first none last none setup warn
bomlev 0

write coor card name @resi_temp.crd
system "`echo ./mol2crd @RESI_temp.crd @RESI.mol2 @RESI_init.crd | awk '{print tolower($0)}'`"
read coor card name @resi_init.crd

write coor pdb name @resi_init.pdb

!minimize structure
MINI CONJ nstep 200 nprint 20 inbfrq 1000 cutnb 999.
MINI NRAP nstep 50 tolgrd 0.00001

!save minimized geometry
write coor card name @resi_min.crd
write coor pdb name @resi_min.pdb

open write unit 10 card name @resi_min.xplor_ext.psf
write psf xplo unit 10 card
```

Figure 2.7.3. – structure input file used to instruct CGenFF to use the Mel.mol2 file and MOL2CRD file to create parameters. It also uses the CHARMM forcefield files top_all36 cgenff.rtf and par_all36_cgenff.prm to create a PDB file for the ligand. It also instructs under the command @resi (in the case of the melatonin file, this is known as MEL) it writes a minimized .crd and .pdb file before finishing with the .psf. which is written in XPLOR format.

```
                Chemistry at HARvard Macromolecular Mechanics
                (CHARMM) - Free Version 42b1        August 15, 2017
                                  654
        Copyright(c) 1984-2014  President and Fellows of Harvard College
                            All Rights Reserved
        Current operating system: Linux-4.4.0-116-generic(x86_64)@neuromancer
                    Created on  4/5/18 at 19:44:52 by user: dc392

            Maximum number of ATOMS:    360720, and RESidues:        120240
 RDTITL> * STRUCTURE OPTIMIZATION
 RDTITL> * NEEDS EXTERNAL VARIABLE "resi"
 RDTITL> *

 CHARMM>

 CHARMM>     ioformat extended
 MISCOM> Expanded I/O format is used.

 CHARMM>

 CHARMM>     read rtf card name top_all36_cgenff.rtf
 VOPEN> Attempting to open::top_all36_cgenff.rtf::
 MAINIO> Residue topology file being read from unit  90.
 TITLE> * ------------------------------------------------------------------- *
 TITLE> *          CGENFF: TOPOLOGY FOR THE CHARMM GENERAL FORCE FIELD V. 3.0.1   *
 TITLE> *                   FOR SMALL MOLECULE DRUG DESIGN                      *
 TITLE> * ------------------------------------------------------------------- *
 TITLE> *
 VCLOSE: Closing unit   90 with status "KEEP"

 CHARMM>     read para card flex name par_all36_cgenff.prm
 VOPEN> Attempting to open::par_all36_cgenff.prm::

         PARAMETER FILE BEING READ FROM UNIT 90
 TITLE> * ------------------------------------------------------------------ *
 TITLE> *          CGENFF: PARAMETERS FOR THE CHARMM GENERAL FORCE FIELD V. 3.0.1  *
 TITLE> *                   FOR SMALL MOLECULE DRUG DESIGN                     *
 TITLE> * ------------------------------------------------------------------ *
 TITLE> *
 PARMIO> NONBOND, HBOND lists and IMAGE atoms cleared.
 VCLOSE: Closing unit   90 with status "KEEP"
```

Figure 2.7.4 - CGenFF's output during the minimization and parameterization process. This shows reading of the CHARMM files top_all36_cgenff.rtf and par_all36_cgenff.prm.

```
CHARMM>      stream @resi.str
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel.str::
OPNLGU> Unit 99 opened for READONLY access to mel.str

                      INPUT STREAM SWITCHING TO UNIT    99
RDTITL> * TOPPAR STREAM FILE GENERATED BY
RDTITL> * CHARMM GENERAL FORCE FIELD (CGENFF) PROGRAM VERSION 1.0.0
RDTITL> * FOR USE WITH CGENFF VERSION 3.0.1
RDTITL> *
Parameter: IN1 <- "" <empty>

CHARMM>

CHARMM>      read rtf card append
MAINIO> Residue topology file being read from unit  99.
RDTITL> * TOPOLOGIES GENERATED BY
RDTITL> * CHARMM GENERAL FORCE FIELD (CGENFF) PROGRAM VERSION 1.0.0
RDTITL> *

CHARMM>

CHARMM>      read param card flex append

            PARAMETER FILE BEING READ FROM UNIT 99
RDTITL> * PARAMETERS GENERATED BY ANALOGY BY
RDTITL> * CHARMM GENERAL FORCE FIELD (CGENFF) PROGRAM VERSION 1.0.0
RDTITL> *
PARMIO> NONBOND, HBOND lists and IMAGE atoms cleared.

CHARMM>      RETURN
VCLOSE: Closing unit    99 with status "KEEP"

                      RETURNING TO INPUT STREAM     5
```

Figure 2.7.5 – shows the confirmation of the resname MEL by the script. Topologies being generated by the general forcefield is started.

```
CHARMM>

CHARMM>      read sequence @resi 1
Parameter: RESI -> "MEL"

CHARMM>      bomlev -1 ! for 3-membered rings

CHARMM>      generate @resi first none last none setup warn
Parameter: RESI -> "MEL"
NO PATCHING WILL BE DONE ON THE FIRST RESIDUE
NO PATCHING WILL BE DONE ON THE LAST  RESIDUE
GENPSF> Segment   1 has been generated. Its identifier is MEL.
PSFSUM> PSF modified: NONBOND lists and IMAGE atoms cleared.
PSFSUM> Summary of the structure file counters :
        Number of segments    =        1   Number of residues   =          1
        Number of atoms       =       33   Number of groups     =          1
        Number of bonds       =       34   Number of angles     =         58
        Number of dihedrals   =       76   Number of impropers  =          1
        Number of cross-terms =        0
        Number of HB acceptors =       0   Number of HB donors  =          0
        Number of NB exclusions =       0   Total charge =   -0.00000

CHARMM>      bomlev 0

CHARMM>

CHARMM>      write coor card name @resi_temp.crd
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel_temp.crd::
RDTITL> SYSTEM "`echo ./mol2crd MEL_temp.crd MEL.mol2 MEL_init.crd | awk '{print tolower($0)}'`"
RDTITL> No title read.
VCLOSE: Closing unit    90 with status "KEEP"
VCLOSE: Closing unit    90 with status "KEEP"

CHARMM>      system "`echo ./mol2crd @RESI_temp.crd @RESI.mol2 @RESI_init.crd | awk '{print tolower($0)}'`"
Parameter: RESI -> "MEL"
Parameter: RESI -> "MEL"
Parameter: RESI -> "MEL"
Invoking: `echo ./mol2crd MEL_temp.crd MEL.mol2 MEL_init.crd | awk '{print tolower($0)}'`

CHARMM>      read coor card name @resi_init.crd
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel_init.crd::
        SPATIAL COORDINATES BEING READ FROM UNIT 90
TITLE> 33  EXT
RDTITL> No title read.
VCLOSE: Closing unit    90 with status "KEEP"
```

Figure 2.7.6. – the CGenFF output file reading the atoms, bonds, angles summarising the residue as 1 and giving a measure of dihedral and improper angles found by the forcefield before writing a temporary init.CRD, prior to minimization.

```
CHARMM>

CHARMM>    write coor pdb name @resi_init.pdb
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel_init.pdb::
RDTITL>
RDTITL> No title read.
 Write CHARMM-pdb format
VCLOSE: Closing unit   90 with status "KEEP"

CHARMM>

CHARMM>    !minimize structure
CHARMM>    MINI CONJ nstep 200 nprint 20 inbfrq 1000 cutnb 999.

NONBOND OPTION FLAGS:
    ELEC      VDW      ATOMs    CDIElec  FSHIft   VATOm    VFSWIt
    BYGRoup  NOEXtnd  NOEWald
CUTNB  =999.000 CTEXNB =999.000 CTONNB =995.000 CTOFNB =997.000
CGONNB =  0.000 CGOFNB = 10.000
WMIN   =  1.500 WRNMXD =  0.500 E14FAC =  1.000 EPS    =  1.000
NBXMOD =      5
There are        0 atom  pairs and        0 atom  exclusions.
There are        0 group pairs and        0 group exclusions.
<MAKINB> with mode   5 found      92 exclusions and    68 interactions(1-4)
<MAKGRP> found       0 group exclusions.
Generating nonbond list with Exclusion mode = 5
== PRIMARY == SPACE FOR      529 ATOM PAIRS AND        0 GROUP PAIRS
NBONDA>>  Maximum group spatial extent (12A) exceeded.
  Size is       13.54 Angstroms and starts with atom:        1
  Please check group boundary definitions.


General atom nonbond list generation found:
    436 ATOM PAIRS WERE FOUND FOR ATOM LIST
      0 GROUP PAIRS REQUIRED ATOM SEARCHES




CONJUG> An energy minimization has been requested.

NCGCYC =         100   NSTEP  =          200
PCUT   =   0.9999000   PRTMIN =            1
STEP   =   0.0200000   TOLFUN =    0.0000000
TOLGRD =   0.0000000   TOLITR =          100
TOLSTP =   0.0000000

MINI MIN: Cycle      ENERgy      Delta-E        GRMS    Step-size
MINI INTERN:        BONDs      ANGLes      UREY-b   DIHEdrals   IMPRopers
```

Figure 2.7.7 – writing of an initial PDB before minimization is performed using the resname specified (MEL) it measures the spatial dimensions of the ligand and concludes it is larger than 12 angstroms at 13.54 angstroms. CGenFF states how many bond combinations have been found for the ligands structure. Minimization is then initiated by the forcefield.

```
MINI MIN: Cycle       ENERgy      Delta-E         GRMS   Step-size
MINI INTERN:           BONDs       ANGLes       UREY-b   DIHEdrals    IMPRopers
MINI EXTERN:          VDWaals        ELEC       HBONds         ASP         USER
           ---------    ---------    ---------    ---------    ---------
MINI>         0     -2.68627      0.00000     10.30825      0.00000
MINI INTERN>         7.91387     10.76762      0.27492      2.21775      0.01350
MINI EXTERN>         8.66354    -32.53747      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>        20    -15.92331     13.23704      1.69578      0.05845
MINI INTERN>         2.02383      4.46163      0.35023      1.20049      0.01179
MINI EXTERN>        10.58311    -34.55438      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>        40    -18.18476      2.26146      0.88898      0.03041
MINI INTERN>         1.93577      3.65116      0.22931      0.99464      0.00106
MINI EXTERN>         9.24055    -34.23725      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>        60    -18.76544      0.58067      0.58007      0.02631
MINI INTERN>         1.76201      3.46333      0.21157      0.98041      0.00076
MINI EXTERN>         9.10122    -34.28474      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>        80    -18.93169      0.16626      0.14233      0.00459
MINI INTERN>         1.77527      3.66701      0.22721      0.96010      0.00001
MINI EXTERN>         8.73190    -34.29320      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>       100    -18.94889      0.01719      0.03272      0.00084
MINI INTERN>         1.78262      3.64520      0.22937      0.95127      0.00000
MINI EXTERN>         8.74762    -34.30497      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>       100    -18.94889      0.01719      0.03272      0.00000
MINI INTERN>         1.78262      3.64520      0.22937      0.95127      0.00000
MINI EXTERN>         8.74762    -34.30497      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>       120    -18.95096      0.00208      0.03701      0.00199
MINI INTERN>         1.78822      3.66610      0.22950      0.95037      0.00000
MINI EXTERN>         8.72079    -34.30594      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>       140    -18.95255      0.00158      0.02967      0.00186
MINI INTERN>         1.78626      3.67434      0.23089      0.94938      0.00001
MINI EXTERN>         8.71836    -34.31178      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>       160    -18.95338      0.00083      0.01753      0.00089
MINI INTERN>         1.78161      3.66841      0.23117      0.94915      0.00000
MINI EXTERN>         8.72813    -34.31185      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>       180    -18.95368      0.00030      0.01040      0.00112
MINI INTERN>         1.78332      3.67158      0.23098      0.94887      0.00000
MINI EXTERN>         8.72306    -34.31149      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
MINI>       200    -18.95381      0.00012      0.00434      0.00036
MINI INTERN>         1.78601      3.66752      0.23131      0.94870      0.00000
MINI EXTERN>         8.72480    -34.31213      0.00000      0.00000      0.00000
           ---------    ---------    ---------    ---------    ---------
```

Figure 2.7.8 – summary table of all Van der-waal, bond angles, dihedrals and impropers during minimization in CGenFF on melatonin.

```
CHARMM>    MINI NRAP nstep 50 tolgrd 0.00001

NONBOND OPTION FLAGS:
    ELEC      VDW       ATOMs    CDIElec  FSHIft   VATOm    VFSWIt
    BYGRoup  NOEXtnd  NOEWald
CUTNB  =999.000 CTEXNB =999.000 CTONNB =995.000 CTOFNB =997.000
CGONNB =  0.000 CGOFNB = 10.000
WMIN   =  1.500 WRNMXD =  0.500 E14FAC =  1.000 EPS    =  1.000
NBXMOD =      5
There are      436 atom  pairs and      160 atom  exclusions.
There are        0 group pairs and        0 group exclusions.
Generating nonbond list with Exclusion mode = 5
== PRIMARY == SPACE FOR     529 ATOM PAIRS AND        0 GROUP PAIRS
NBONDA>> Maximum group spatial extent (12A) exceeded.
  Size is       12.66 Angstroms and starts with atom:       1
  Please check group boundary definitions.


General atom nonbond list generation found:
     436 ATOM PAIRS WERE FOUND FOR ATOM LIST
       0 GROUP PAIRS REQUIRED ATOM SEARCHES




 NRAPH> An energy minimization has been requested.

 NSTEP   =          50   NPRINT =          20
 STEP    =   0.0200000   TFREQ  =  50.0000000
 TOLGRD  =   0.0000100   TOLSTP =   0.0000000

MINI MIN: Cycle      ENERgy      Delta-E      GRMS     Step-size
MINI INTERN:         BONDs       ANGLes      UREY-b   DIHEdrals    IMPRopers
MINI EXTERN:         VDWaals       ELEC      HBONds         ASP         USER
---------      ---------   ---------   ---------   ---------   ---------
MINI>       0  -18.95381    -0.00000    0.00434     0.00000
MINI INTERN>     1.78601     3.66752    0.23131     0.94870     0.00000
MINI EXTERN>     8.72480   -34.31213    0.00000     0.00000     0.00000
---------      ---------   ---------   ---------   ---------   ---------

 NRAPH> Minimization exiting with gradient tolerance ( 0.0000100) satisfied.

NRAP MIN: Cycle      ENERgy      Delta-E      GRMS     Step-size
NRAP INTERN:         BONDs       ANGLes      UREY-b   DIHEdrals    IMPRopers
NRAP EXTERN:         VDWaals       ELEC      HBONds         ASP         USER
---------      ---------   ---------   ---------   ---------   ---------
NRAP>       3  -18.95384     0.00004    0.00000     0.02000
NRAP INTERN>     1.78579     3.66545    0.23116     0.94863     0.00000
NRAP EXTERN>     8.72658   -34.31146    0.00000     0.00000     0.00000
---------      ---------   ---------   ---------   ---------   ---------

CHARMM>

CHARMM>    !save minimized geometry
CHARMM>    write coor card name @resi_min.crd
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel_min.crd::
RDTITL> WRITE COOR PDB NAME MEL_MIN.PDB
RDTITL> No title read.
VCLOSE: Closing unit   90 with status "KEEP"
VCLOSE: Closing unit   90 with status "KEEP"
```

Figure 2.7.9. – After minimization is finished in CGenFF, the new geometry is saved in a file called @resi_min.pdb (new co-ordinates after minimization) this creates the file MEL_MIN.PDB.

```
CHARMM>

CHARMM>      !save minimized geometry
CHARMM>      write coor card name @resi_min.crd
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel_min.crd::
RDTITL> WRITE COOR PDB NAME MEL_MIN.PDB
RDTITL> No title read.
VCLOSE: Closing unit   90 with status "KEEP"
VCLOSE: Closing unit   90 with status "KEEP"

CHARMM>      write coor pdb name @resi_min.pdb
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel_min.pdb::
RDTITL>
RDTITL> No title read.
 Write CHARMM-pdb format
VCLOSE: Closing unit   90 with status "KEEP"

CHARMM>

CHARMM>      open write unit 10 card name @resi_min.xplor_ext.psf
Parameter: RESI -> "MEL"
VOPEN> Attempting to open::mel_min.xplor_ext.psf::
OPNLGU> Unit 10 opened for WRITE access to mel_min.xplor_ext.psf

CHARMM>      write psf xplo unit 10 card
RDTITL> No title read.
```

Figure 2.7.10 – CgenFF saves a new CRD and PSF file with the pretext @resi_min.CRD, and @resi_min.xplor_ext.psf. @resi is MEL, therefore the new filenames are; MEL_MIN.CRD, Mel_MIN.xplor_ext.psf.

These files are now ready with topology and co-ordinate information for a docking simulation.
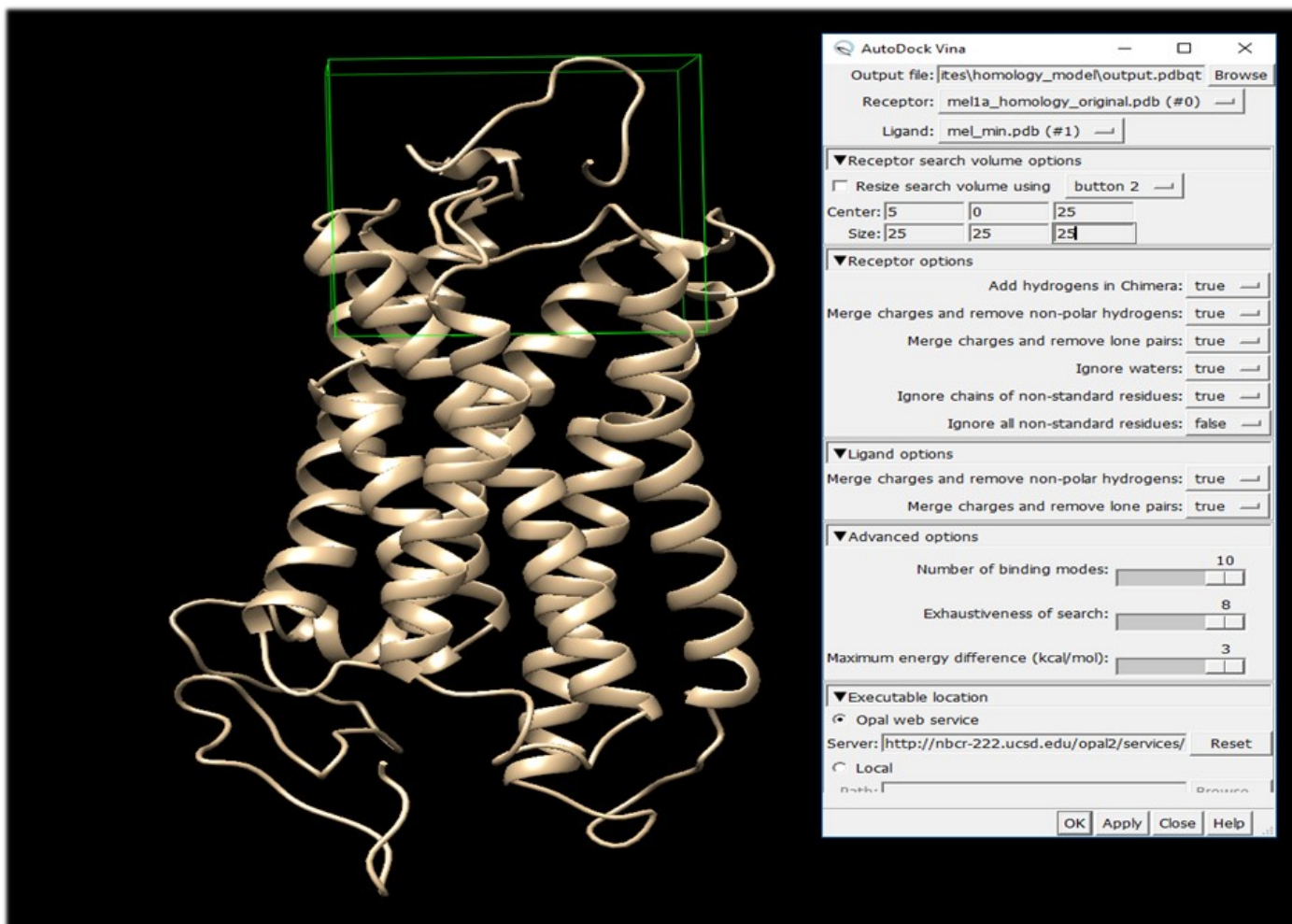
Figure 2.8.1 – Docking was performed on the provided Opal Web service made available through the Chimera client.
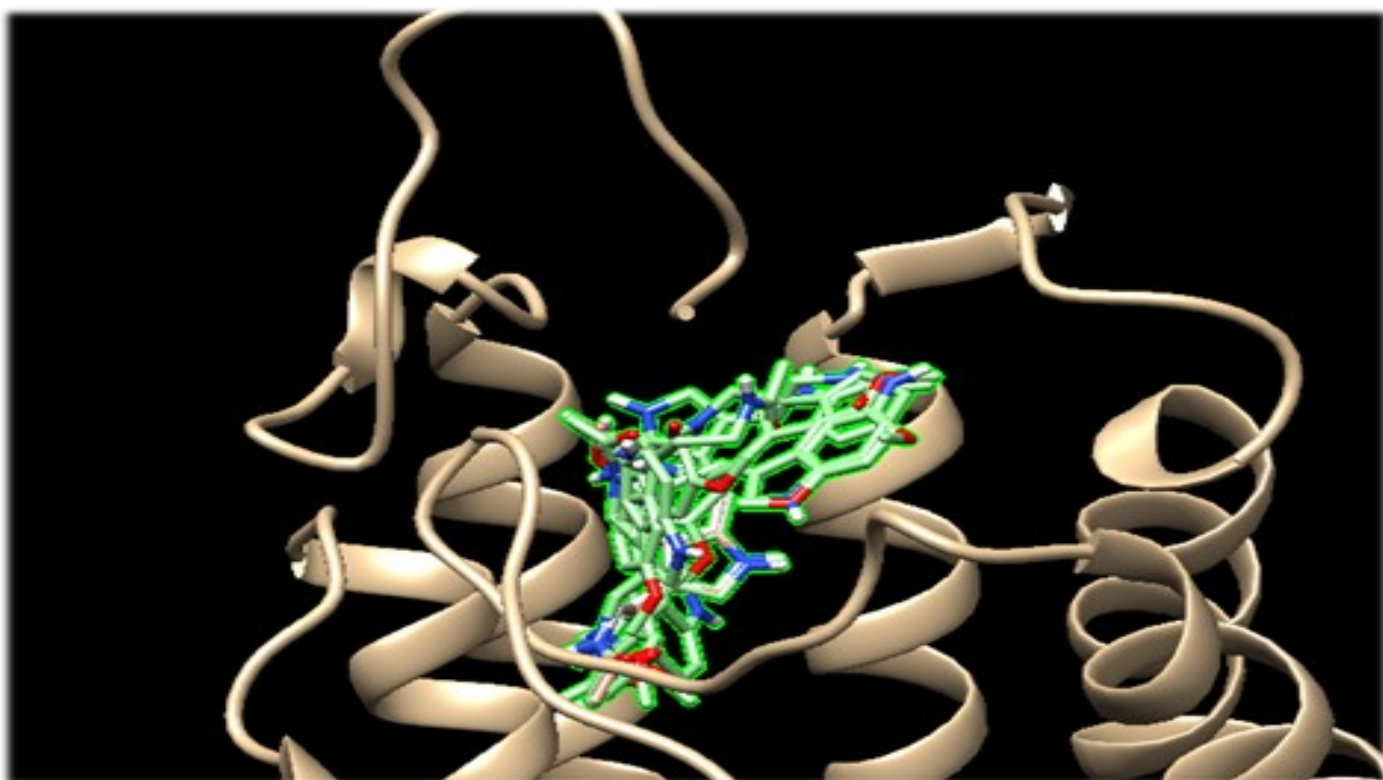
Figure 2.8.2 – Docking solutions shown in Chimera 1.12 using sticks for the 10 Melatonin results and cartoon for the protein.

```
########################################################################
## MD of Mel1a GCPR - Control NO ligands. 30Nanoseconds - refreq eqi.   ##
########################################################################
## CONSTANTS                                                            ##
########################################################################

cutoff 12.0
pairlistdist 14.0
switching on
switchdist 10.0
PME on
PMEGridspacing 1
wrapAll on
wrapWater on

##System definition##

coordinates /home/dc392/coord/mel1a_final.pdb
structure   /home/dc392/structure/mel1a_final.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT. (eqi files are required to perform MD)

binCoordinates /home/dc392/inp/eqi/eqi_rescale/mel1a_firstrun_eqi_310k_refreq.coor
binVelocities  /home/dc392/inp/eqi/eqi_rescale/mel1a_firstrun_eqi_310k_refreq.vel

extendedSystem /home/dc392/inp/eqi/eqi_rescale/mel1a_firstrun_eqi_310k_refreq.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED


##Conditions##
#temperature 310

##harmonic constraints## OFF
#constraints on
#consref   /home/dc392/par/
#conskfile /home/dc392/par/
#constraintScaling 2
#consexp  2
#conskcol B
```

Figure 2.9.1 - input file for the 30 nanosecond (ns) input files for pdb and psf remain the same, rescale equilibration data is used for initial starting point for the simulation identical to the 10ns simulations. Harmonic constraints are disabled.

```
##Output Parameters##

binaryoutput yes
outputname mel1a_MD_refreq_control_30ns_run1
outputenergies 500
outputtiming 500
outputpressure 500
binaryrestart yes
dcdfile mel1a_MD_refreq_control_30ns_run1.dcd
dcdfreq 2000
XSTFreq 2000
restartfreq 2000
restartname mel1a_MD_refreq_control_30ns_run1.restart

##Thermostat Parameters##
langevin on
langevintemp 310
langevinHydrogen      off
langevindamping 1

##Barostat Parameters##

usegrouppressure yes
useflexiblecell no
useConstantArea no
langevinpiston on
langevinpistontarget 1.01325
langevinpistonperiod 200
langevinpistondecay 100
langevinpistontemp 310

##Integrator Parameters##

timestep 1
firstTimestep 0
fullElectFrequency 2
nonbondedfreq 1
```

Figure 2.9.2 - Timestep is changed to 1, so the FFT algorithm records data for every step instead of every 2 steps as to compensate for the atoms which are moving too fast for FFT to make a logical velocity and c-ordinate trajectory. Langevin Dampening was lowered back to 1 decreasing the amount of friction experienced by the system. As this is the case timesteps will also have to be doubled to 30,000,000 as there is no longer the need to divide the timesteps by 2.

```
##Force Field Parameters##

paratypecharmm on
parameters          /home/dc392/toppar/par_all36_prot.prm
parameters          /home/dc392/toppar/par_all36_na.prm
parameters          /home/dc392/toppar/par_all36_carb.prm
parameters          /home/dc392/toppar/par_all36_cgenff.prm
parameters          /home/dc392/toppar/par_all36_lipid.prm
parameters          /home/dc392/toppar/toppar_water_ions_namd.str
exclude scaled1-4
1-4scaling 1.0
rigidbonds water


#Implicit Solvent Parameters

#gbis off
#alphaCutoff          14.0
#ionConcentration     0.15


##Script##

run 30000000
```

Figure 2.8.3 - in identical fashion to the 10 nanosecond molecular dynamics runs, the force field parameters remain the same since the FFT algorithm is now recording every 1 femtosecond, the number of run steps is doubled to compensate for the recording of data being doubled.

```
########################################################################
## MINIMIZATION Mel1a GCPR + melatonin, first run                    ##
########################################################################
## FORCEFIELD PARAMATERS                                             ##
########################################################################

paratypecharmm      on
parameters          /home/dc392/toppar/par_all36_prot.prm
parameters          /home/dc392/toppar/par_all36_na.prm
parameters          /home/dc392/toppar/par_all36_carb.prm
parameters          /home/dc392/toppar/par_all36_cgenff.prm
parameters          /home/dc392/toppar/par_all36_lipid.prm
parameters          /home/dc392/toppar/toppar_water_ions_namd.str
parameters          /home/dc392/mel/par/mel1_final.par

########################################################################
## STRUCTURE AND COORDINATES                                         ##
########################################################################
structure           /home/dc392/structure/final_docking_system.psf
coordinates         /home/dc392/coord/final_docking_system.pdb

########################################################################
## TEMPERATURE                                                       ##
########################################################################
temperature         310

########################################################################
## TEMP AMD PRESSURE COUPLING                                        ##
########################################################################
langevin            on
langevinTemp        310
langevinDamping     5
langevinHydrogen    off

########################################################################
## CONSTANT PRESSURE CONTROL (VAR VOLUME)                            ##
########################################################################
useGroupPressure        yes
useFlexibleCell         no
useConstantArea         no
langevinPiston          on
langevinPistonTarget    1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod    200.0
langevinPistonDecay     50.0
langevinPistonTemp      310

########################################################################
## OUTPUTS                                                           ##
########################################################################
outputname          mel1a_dock_min
outputEnergies      40
restartfreq         500
DCDfreq             20
outputTiming        50
wrapAll             on
wrapNearest         on
binaryoutput        yes
binaryrestart       yes

########################################################################
## INTEGRATORS                                                       ##
########################################################################
timestep            2
nonbondedFreq       2
fullElectFrequency  4
stepspercycle       20
```

Figure 2.10.1 -The gradient descent model was used for 2500 timesteps with a LangevinDampening co-efficient of 5. Due to the removal of water molecules and the addition of the Melatonin ligand in the new PDB and PSF files, it is necessary to re-minimize and heat the system before equilibration, as distribution of potential energy, heat and velocity has changed as a result.

```
##harmonic constraints##
constraints on
consref   /home/dc392/par/docking_constraints.pdb
conskfile /home/dc392/par/docking_constraints.pdb
constraintScaling 2
consexp  2
conskcol B
```

Figure 2.10.2 - A new constraints file which limits the energy of carbon alphas in the protein and the phosphorus atoms of the membrane were created due to the altered/removed atoms within the system mentioned previously to uphold the same environment in minimization.

```
################################################################
## HEATING OF mel1a constraints docking run 1               ##
################################################################
## CONSTANTS                                                ##
################################################################


cutoff 12.0
pairlistdist 14.0
switching on
switchdist 10.0
PME on
PMEGridspacing 1
wrapAll on
wrapWater on

##System definition##

coordinates /home/dc392/coord/final_docking_system.pdb
structure   /home/dc392/structure/final_docking_system.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT.

binCoordinates /home/dc392/inp/min/min_dock/mel1a_dock_min.coor
#binVelocities  /home/dc392/inp/min/min_dock/mel1a_dock_min.vel

extendedSystem /home/dc392/inp/min/min_dock/mel1a_dock_min.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED

#cellOrigin          NA for heating.
#cellBasisVector1    na
#cellBasisVector2    na
#cellBasisVector3    na

##Conditions##


##harmonic constraints##
constraints on
consref    /home/dc392/par/docking_constraints.pdb
conskfile /home/dc392/par/docking_constraints.pdb
constraintScaling 2
consexp  2
conskcol B

##Output Parameters##

binaryoutput yes
outputname mel1a_eqi_heat_dock
outputenergies 500
outputtiming 500
outputpressure 500
binaryrestart yes
dcdfile      mel1a_eqi_heat_dock.dcd
dcdfreq 1000
XSTFreq 1000
restartfreq 1000
restartname mel1a_eqi_heat_dock

##Thermostat Parameters##
langevin on
langevintemp 60
langevinHydrogen     off
langevindamping 1
```

Figure 2.11.1 –Heating input file phase. Langevin Dampening is set to 1 down from 5 in minimization. New docking .pdb and .psf files are also sourced.

```
#################################################################
## EQUILIBRATION OF docking Mella in membrane - reassignFreq run   ##
#################################################################
## CONSTANTS                                                      ##
#################################################################

cutoff 12.0
pairlistdist 14.0
switching on
switchdist 10.0
PME on
PMEGridspacing 1
wrapAll on
wrapWater on

##System definition##

coordinates          /home/dc392/coord/final_docking_system.pdb
structure            /home/dc392/structure/final_docking_system.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT.

binCoordinates /home/dc392/inp/heat/heat_dock/mella_eqi_heat_dock.coor
binVelocities  /home/dc392/inp/heat/heat_dock/mella_eqi_heat_dock.vel


extendedSystem /home/dc392/inp/heat/heat_dock/mella_eqi_heat_dock.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED


##Conditions##
#temperature 310

##rescale parameters##
rescaleFreq 1
rescaleTemp 310
margin 5

##harmonic constraints##
constraints on
consref    /home/dc392/par/docking_constraints.pdb
conskfile /home/dc392/par/docking_constraints.pdb
constraintScaling 2
consexp  2
conskcol B

##Output Parameters##

binaryoutput yes
outputname mella_docking_310k_rescale
outputenergies 500
outputtiming 500
outputpressure 500
binaryrestart yes
dcdfile mella_docking_310k_rescale.dcd
dcdfreq 2000
XSTFreq 2000
restartfreq 2000
restartname mella_docking_310k_rescale.restart

##Thermostat Parameters##
langevin on
langevintemp 310
langevinHydrogen     off
langevindamping 2
```

Figure 2.12.1 – Equilibration input file. the system crashed reporting periodic cell boundary limits being exceeded. As a result. The margin of the simulation was increased from 1.5 to 5 to counteract the issues seen. The Langevin Dampening coefficient was also increased to 2, to give the system a larger friction and random forces variable.

```
###############################################################################
## MD of Mel1a and melatonin GCPR - Control NO ligands. 30Nanoseconds - refreq eqi.    ##
###############################################################################
## CONSTANTS                                                                   ##
###############################################################################

cutoff 12.0
pairlistdist 14.0
switching on
switchdist 10.0
PME on
PMEGridspacing 1
wrapAll on
wrapWater on

##System definition##

coordinates /home/dc392/coord/final_docking_system.pdb
structure   /home/dc392/structure/final_docking_system.psf

#FOR MINIMISATION BIN VEL AND BINCOORD ARE HASHED OUT. (eqi files are required to perform MD)

binCoordinates /home/dc392/inp/eqi/dock_rescale/mel1a_docking_310k_rescale.coor
binVelocities  /home/dc392/inp/eqi/dock_rescale/mel1a_docking_310k_rescale.vel

extendedSystem /home/dc392/inp/eqi/dock_rescale/mel1a_docking_310k_rescale.xsc

#EITHER AN xsc FILE OR CELL BASIS VECS&ORIGIN NEED TO BE UNHASHED


##Conditions##
#temperature 310

##harmonic constraints## OFF
#constraints on
#consref   /home/dc392/par/
#conskfile /home/dc392/par/
#constraintScaling 2
#consexp  2
#conskcol B

##Output Parameters##

binaryoutput yes
outputname mel1a_MD_dock_melatonin_30ns_run1
outputenergies 500
outputtiming 500
outputpressure 500
binaryrestart yes
dcdfile mel1a_MD_dock_melatonin_30ns_run1.dcd
dcdfreq 2000
XSTFreq 2000
restartfreq 2000
restartname mel1a_MD_dock_melatonin_30ns_run1.restart

##Thermostat Parameters##
langevin on
langevintemp 310
langevinHydrogen    off
langevindamping 1
```

Figure 2.13.1 – Input file for molecular dynamics on the docking system.

```
##Barostat Parameters##

usegrouppressure yes
useflexiblecell no
useConstantArea no
langevinpiston on
langevinpistontarget 1.01325
langevinpistonperiod 200
langevinpistondecay 100
langevinpistontemp 310

##Integrator Parameters##

timestep 1
firstTimestep 0
fullElectFrequency 2
nonbondedfreq 1

##Force Field Parameters##

paratypecharmm on
parameters          /home/dc392/toppar/par_all36_prot.prm
parameters          /home/dc392/toppar/par_all36_na.prm
parameters          /home/dc392/toppar/par_all36_carb.prm
parameters          /home/dc392/toppar/par_all36_cgenff.prm
parameters          /home/dc392/toppar/par_all36_lipid.prm
parameters          /home/dc392/toppar/toppar_water_ions_namd.str
parameters          /home/dc392/mel/par/mel1_final.par
exclude scaled1-4
1-4scaling 1.0
rigidbonds water

#Implicit Solvent Parameters

#gbis off
#alphaCutoff         14.0
#ionConcentration    0.15

##Script##

run 30000000
```

Figure 2.13.2 – Parameters for docking MD simulation, mel1_final.par was added to the parameters section.

3.0. Quantum Mechanical (QM) Parameterization of Melatonin.

The PDB file previously created in Avogadro for melatonin was used as an initial co-ordinate file for the forcefield toolkit (FFTK) in VMD. Using VMD's built in Force Field Tool Kit (FFtk) an initial PSF was generated which was entered into the *buildpar* section of FFTK for Gaussian16. Atom types were defined using the CGenFF forcefield and the protein datafile for CHARMM36. Properties for the ligand were viewed/edited in Molfacture, also available through VMD1.9.3. Once this step is completed it is turned into new PDB and PSF files. The basis set used for parameterization was MP2/6-31G* (Jurečka *et al.*, 2006).

3.0.1 Geometry Optimization.

An initial parameter file was generated in FFTK (Jurečka *et al.*, 2006). and an input Gaussian 16 geometry file was generated under the *opt.geometry* tab in FFtk. This generates a new set of optimized co-ordinates and low energy geometry conformation of the molecule.

3.0.2 Water Interaction Properties of Melatonin.

Water interaction sites were then calculated for the molecule in FFtk. This builds parameters for the donor (interaction of oxygen in water) and acceptor (interactors with hydrogen in water) sites of the ligand for hydrogen bonding atoms/interactions. The base name is taken from the original topology data (.psf) and geometry optimized .pdb. The QM water interaction data was then utilised in charge optimisation of the molecule. The water interaction section utilized basis sets HF/6-31G* for water interaction, HF single point used the HF/6-31G* basis set, and MP2 single point utilized MP2/6-31G*.

3.0.3 Charge optimization of Melatonin.

The water interaction data was then used in the charge optimisation of melatonin and an output log was specified to write new charge parameters to. Initial charges under the constraints tab were set to -0.1 as they cannot be 0. Aliphatic hydrogens were set to +0.09 as an initial charge in accordance with the CHARMM rules. A charge sum for the molecule was calculated based on the initial charge input (low bounds are set to -1.0, while high bounds are set to 1.0 to give a range of charge distribution for each atom type as a scaffold for optimisation). The Gaussian log files containing the water interaction results were then loaded as QM target data. In addition to each water interaction HF single point energy (at the theory level of all water interactions) and MP2 single point energy (a higher level of theory) was supplied. As such, the MP2 single

point data also provided dipole information for the molecule. Optimisations were performed using the simulated annealing. After the optimization of the charges the data was saved as an updated .psf which includes the optimized charge data along with all previously attained water interaction data prepared using the HF and MP2 basis sets.

3.0.4 Hessian Bond Calculations for the Melatonin molecule and bond optimization using basis set MP2/6-31G*.

Data calculated during geometry optimization, the in-progress parameter file and Gaussian16 checkpoint files were utilised in order to calculate the Hessian for bonds and angles. Bond lengths, angles and force constants were then optimised via molecular mechanics algorithms against the Hessian QM target data. Hessian data is generated using the excited and ground state of bonds through convergence of the results attained.

3.0.5 Scanning of torsion bonds and atoms to generate QM target data.

Data from the previous steps (charge optimized .psf, geometry optimized .pdb) is loaded into the scan torsions tab of FFtk. Along with the hessian data generated in the previous step which was saved into the parameter (.par) file. This data is utilized to identify the dihedral angles present in the molecule. No equivalent dihedrals (based upon molecular symmetry) are present in the molecular structure of melatonin, so none were removed. Scan radius is assigned in increments of +/-90 ^{o} to provide 180 degrees of data per dihedral calculation with a 10° window size. Two exceptions to this were applied for the CT2 molecules found in the structure which were additionally scanned across +/-180-^{o}. This was done to account for the greater molecular flexibility in this region which produced a greater range of the dihedral angles.

3.0.6 Optimization of torsional angles and phase shift.

Optimization of dihedral angles was then performed using the data generated and saved in the optimized charges PSF, optimized bonds PSF, and files optimized previously. Scan torsions data was also sourced - to the QM target data section to identify dihedrals for torsional refinement. Parameter files were sourced from the CHARMM forcefield for proteins and general forcefield (CGenFF) also included is the in-progress PAR file for the ligand. Once the dihedrals had been optimized. The file is saved and merged into the PAR file ready to be used. Dihedral angles were optimised initially without an energy cut off, force constants were reset (kmax equal to 3) and then a cut off of 20 kcal/mol was applied to all further calculations. All calculations were performed as simulated annealing and melatonin conformations greater than 20Kcal/mol received a weighting of zero. The MM calculations were then compared to the QM

Hamiltonian in order to assess the progress of the optimisation where the lowest possible RMSE was obtained. In addition, reference was made to analogous force constants and phase shifts from CHARMM while periodicities were set in accordance with the known molecular nature (i.e. 2 for planar rings).

Results.

4.1 Characterization of the I-Tasser Homology model of the Mel1a G-Protein Coupled Receptor.

The I-tasser Generated Mel1a GCPR homology model was characterised and coloured based on secondary structure identification for ease of understanding the overall structure in VMD1.9.3. (Figures 4.1.1, 4.1.2, 4.1.3). Characterization of the GCPR will allow comparison with known structures of the same family Rhodopsin and β2 adrenergic GCPR's which have been solved via X-ray crystallography. Initial characterization is also used as a backwards reference to any changes which may happen under molecular dynamic simulation.

The seven alpha helix sections of the I-Tasser homology model for Mel1a are shown in Figure 4.1.1.

Alpha Helix A, Asp19 – Asn55.

Comprised of 37 residues, a singular 180-degree helical structure is formed between Asp19 to Ala21 leading to a short loop section which contains pro23. this continues as an alpha-helix until residue Asn55. Cys32 a residue responsible for disulphide bonds has no partner cysteine in close proximity- the nearest being Cys289 in alpha helix-turn-helix G.

Alpha Helix B, Asn63 - Asn90.

An alpha helix shape is maintained in the first 12 residues until reaching Leu74, where the helix takes a larger radius of Å for the next 10 residues and has a pi-helix appearance. Included in the residues are a double-barrel Val75 and Val76. This gives the helix A slight angular turn between approximately Ile78 and Tyr79. The new angle is followed from this point where a more typical alpha helical conformation returns continuing until residue Asn90.

Alpha Helix C, Tyr97 - Ser132.

Cys100 is pointed inward towards the core of the protein and forms a disulphide bond with Cys177, this would make the section very stable and infer that it is kept in a tight conformation for a biological purpose. Cys127 faces outward into what would be the membrane. The closest cysteine residues sit on alpha helix E - at Cys146 which is more than 15Å away and alpha helix D at Cys206 which is separated by A minimum of 7Å from the cysteine residues on alpha helix B. The cysteine residues suggest that alpha helix C is less conformationally flexible than that of the other helixes.

Alpha Helix D, Lys142 - Arg164.

Leading from Lys142 the alpha helix follows a uniform 3 10 helix shape. Having a cysteine molecule (Cys146) with no disulphide bond, although it has the potential to create a bond with Cys127 following molecular dynamics of the protein. At Leu156, the helix exhibits a slight turn towards alpha helix C. This is further perpetuated by Pro161 a residue known for inducting turns in secondary structure, towards the end of the helix prior to the structure changing into a loop section (Beta Loop E).

Alpha Helix E, Ser185 - Leu215.

After the first 4 residues Ser185 to Ile189 the helical shape widens and there is an angular turn which leads the remaining helix away from the core of the protein. After Pro199 the helix returns to the classic alpha helix conformation. Which extends all the way to Leu215. Met200 sits outside the core of the protein in the bi-lipid cell environment. Which may show incorrect orientation of the helix. Furthermore, Cys206 does not have a nearby partner cystine in any other section of the secondary structure. Its closest cysteine neighbour is over 7Å away on alpha helix C.

Alpha Helix F, Leu229 - Val261.

An alpha helical structure is evident until residue Phe247 where a 90-degree bend is observed. This kink is 6 residues long ending in Pro253 before returning to an alpha helical conformation between residues Leu254 to Val261. No interactions are present with surrounding parts of the protein, so the kink is a property of the primary sequence.

Alpha Helix-Turn-Helix G, Asn274 - Ile309.

A typical helix structure is seen for the first 13 residues (Asn274 to Asn287). A slight bend is promoted towards helix A at which Cys289 appears to be at the closest point. This leads to the impression that it is closer to Cys32 in orientation. However, this may be due to the current orientation of the helix which will change under simulation conditions. The helix-turn-helix portion is seen between the two isoleucine residues (Ile 293 and 294) which are followed by Tyr295 and Gly296. Residues Leu297 and 298 are seated before the final turn begins. Asn299 sits at the centre turning between the helix sections. After the turn the remaining amino acids follow a standard alpha helical conformation from asn299 to Ile309.

Loop sections A to E and beta Turn A in the Mel1a GPCR are seen in Figure 4.1.2, these sections are highly mobile at the end of either side of the protein (extracellular and intracellular). Mapping of the loop regions is important as it is going to change significantly in molecular dynamic (MD) simulation. Hydrophobic parts of the loop sections are likely to pack into conformations after calculations have been performed.

Beta Loop A. Met1-Gly18.

The extracellular Beta loop A spans 18 residues and at Leu8 and Pro9 there is a turn that corresponds to the uppermost point of the entire protein on the extracellular side of the membrane. The loop then traverses towards the core of the protein until residue Pro14. where a beta-turn-like conformation follows on to Val15.

Beta Loop B, Lys56-Gly62.

The loop contains 6 residues. It sits on the intracellular side of the Mel1a GCPR. It provides transition between alpha helix A, and alpha helix B.

Beta Loop C – Asn91-Gly96.

The extracellular loop is a transitional structure between alpha helix B and alpha helix C. Containing 5 residues. Asn91 transitions from alpha helix to loop section, Gly96 sits before the start of alpha helix D.

Beta Loop D Leu133 – Ser141.

Initiated at Leu 133, the loop transitions between alpha Helix D and alpha Helix E completing at residue Ser141 on the intracellular side of the protein. The centre section of the loop takes on a circular conformation. Starting at Lys134 and progressing in a circular motion until Tyr139. Before straightening out prior to alpha helix E with double Ser140 and 141 residues.

Beta Loop E Ala165 - Thrrr167.

This extracellular loop transitions between alpha helix E and beta turn A, consisting of 3 residues Ala165, Gly166 and Thr167.

Beta Turn A 168 – 179.

Beta turn A is a beta sheet on the extracellular side, between residues Leu168 and Ala171 after which Pro172 induces a turn. This is followed by a short loop section which includes 2 residues (Arg173, Ile174) before returning to a beta sheet conformation at Tyr175. This sheet contains a cysteine (Cys177) residue which has formed a disulphide bond with Cys100 on alpha helix C, inferring a restaint in positioning for the beta turn in this area. The second sheet ends at Phe179. Hydrogen bond interactions are present between the two sheet sections between Gln169 oxygen (O) and Thr178's HG1, Asp171's oxygen atom(O) and Ile174's (HN) interact. Pro179 has a hydrogen bond interaction with the HN atom on the Ile174 residue - likely inducing the turn between the beta sheets.

Figure 4.1.3 shows the remaining secondary structure loop sections from the I-Tasser Mel1a GCPR structure.

Beta Loop F Ala180-Ser184.

Beta loop F contains 4 residues(Ala180 -Ser184). It transitions between Beta turn A and alpha helix F on the extracellular side of the Mel1a GCPR.

Beta Loop G – 216 – 228.

Formed of 12 residues, loop G resides on the intracellular portion of the protein. The structure contains 2 Proline residues, Pro223 and Pro227 which seem to cause a bend between this length of loop. Lys226 is held in place by a hydrogen bond between Pro223's O atom, and Lys226's HD1 atom.

Beta Loop H – Ala262 – Pro273.

Containing 11 residues, beta loop H starts at Ala 262 on the end of alpha helix F. Pro265 induces a turn. Molecular dynamics simulation would likely cause a more packed conformation. Leading to a less exposed methionine residue. Furthering on from Met268, Pro270 this causes another fold in the loop section. The remaining 3 residues follow the modified direction by Proline towards helix-turn-helix G, the final alpha helix in the proteins structure.

Beta Loop I – Ile310 toVal350.

Loop I resides on the intracellular side of the Mel1a GCPR and is comprised of 40 residues. It begins at the end of helix turn helix G starting at Ile310. Four residues after the initial residue, Cys314 is exposed and has no obvious di-sulphide bonding partner. it is possibly related to protein-protein interaction in which a di-sulphide bond is formed between another molecule and MEL1A. Due to Cysteine molecules being on the intracellular side of the protein, it is more likely it is involved in protein-protein interaction with another sulphur containing Cysteine residue on an exogenous molecule. Further along the loop Pro335 and Pro337 inferring the latter half of the loop may be more tightly packed. Two residues further along the loop Met339 is present. Hydrophobic in nature is also presents the assumption that this section of loop is packed towards the centre of the intracellular side of the protein.

The I-Tasser model appears to have a general rhodopsin like GPCR family shape, there are clear intracellular and extracellular loop sections, and the central core is comprised of Alpha helix sections. A structural alignment with known X-ray diffracted structures of GPCR's will give a general comparison between a confirmed crystal structure and the homology model.

4.2 Structural Alignment with X-ray Structures of Human Rhodopsin and β2-adregernic receptor.

The homology model for Mel1a was aligned with the X-ray crystal structures for the Rhodopsin and β2-adregernic GCPR's for comparative analysis after removing all other molecules and residues not related to the structures. This was performed in Pymol 2.1.1 (Delano, 2002)(Figure 4.2.1). RMSD of the alignment is given in angstroms and provides a measure of structural variation between the crystal structures and the prediction of Mel1a made by the i-Tasser server, the GCPR's are shown in carton. RMSD between Mel1a and β2 Adrenergic receptors is 1.006Å. RMSD between Mel1a and Rhodopsin receptors is 2.170Å. The I-tasser structure resembles both the B2 adrenergic and rhodopsin families, the loop sections on the extracellular side more closely follow the profile of the rhodopsin crystal structure, but share similar architecture in the core helix sections of the protein, the lower RMSD figure for the adrenergic receptor is likely due to the missing sections of loop secondary structure, which should me located on the intracellular side, but are not present - due to the T4 lysosome fusion which was required for crystallization of the structure. This fusion replaced the loop sections between TM4, TM5, and TM6 on the receptor. The final loop section on the intracellular side of rhodopsin is also missing due to the crystallization process, which would be comparative to loop I. Alpha Helixes E and F sit differently to the rhodopsin model. Which explains the higher RMSD between the structures.

4.3 Electrostatic Analysis of the Initial I-Tasser Homology Model.

The homology model for Mel1a was studied using Pymol for electrostatic surface map analysis by converting the provided .PDB file downloaded from the I-Tasser server. The .PDB was converted to a PQR file using the Pymol plugin for APBS electrostatics (Dolinsky *et al.*, 2004). Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM. The charge map which was set to -/+ 10 (default is 5) the charge information is created using the AMBER forcefield (Figure 4.3.1) (Rodríguez-Acosta *et al.*, 2016).

The extracellular side of the protein shows an electronegative pocket which may be a potential active site for melatonin. The extracellular side is mostly positive/neutral in charge on the surface with a small pocket. Which appears to be mostly negatively charged for the conformation predicted by I-Tasser. The side cut view of the protein displays the depth of both

pockets in a cut away section. Showing neither pockets are connected from the intracellular to extracellular side. The intracellular side of Mel1a is mostly electropositive. This could be a potential site for interaction with intracellular proteins. This is supported by electrostatic analysis of the β2 adrenergic and rhodopsin GPCR's (data not shown). which have a similar electrostatic distribution of charge which supports the electrostatics of the I-Tasser model of the Mel1a GPCR, where it is seen in both models that the extracellular side has an electronegative pocket, and the intracellular side has a large electropositive surface. The electrostatic surface of GNAi3 – a G-protein believed to interact with all three of the proteins, Mel1a, Rhopdopsin and the β2 adrenergic, is electronegative on a large proportion the surface. (data not shown).

Summary.
The I-Tasser model for the Mel1a GCPR shows high similarity to both the rhodopsin and β2 adrenergic GPCR's, with close architectural similarity in the trans-membrane core. It also has a similar electrostatic charge distribution. The characterization also confirms that the model has 7-Transmembrane sections which traverse the bi-lipid membrane as a GPCR is described.

Figure 4.1.1.

Characterization of the I-Tasser Homology Model Colour Map and Alpha Helix sections of the MEL1A GCPR .

MEL1A G-CPR Homology Model.

Displayed in VMD 1.9.3 - Drawing Method "New Cartoon", Material "AOChalky" and under the Colouring Method "ColorID" in various colours to segregate each section of the structure.
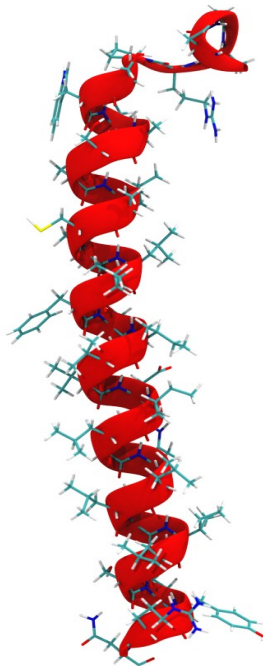
Alpha Helix A , Asp19 - Asn55.
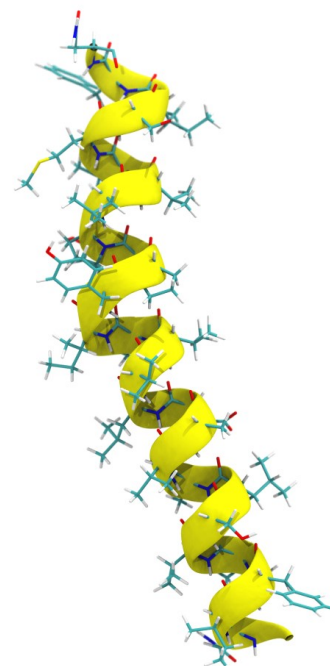
Alpha Helix B, Asn63 - Asn90.
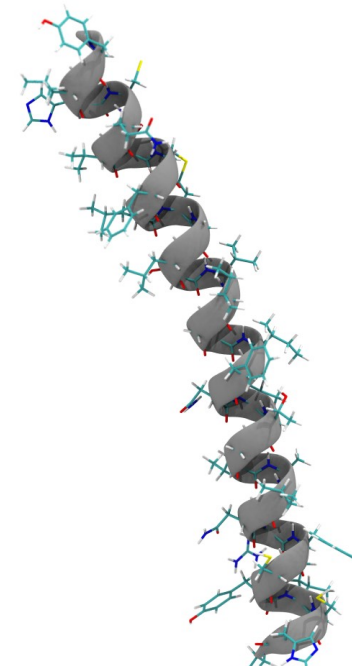
Alpha Helix C Tyr97 - Ser132
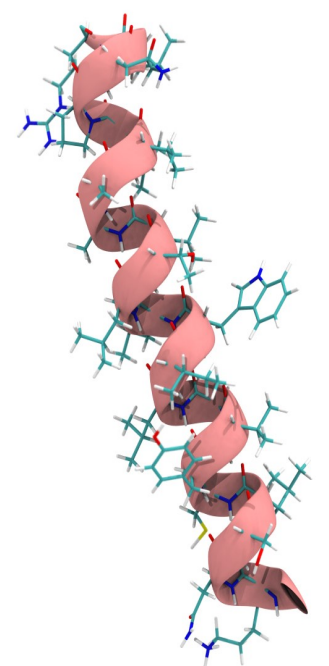
Alpha Helix D, Lys142 - Arg164
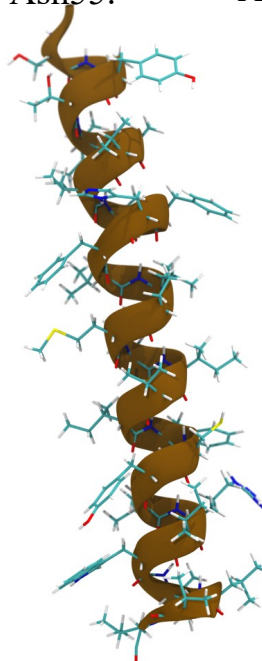
Alpha Helix E, Ser185 - Leu215.
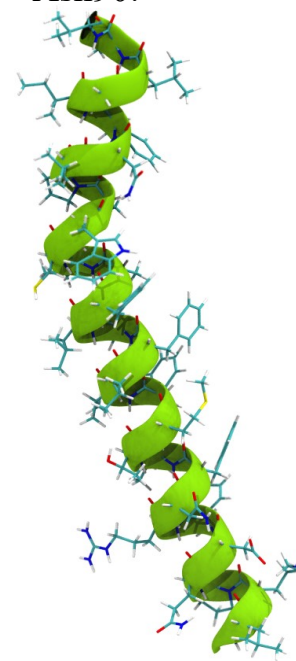
Alpha Helix F, Leu229 - Val261.
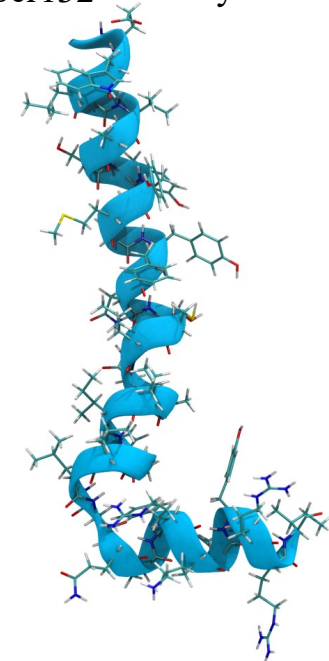
Alpha Helix-Turn-Helix G. Asn274 - Ile309.
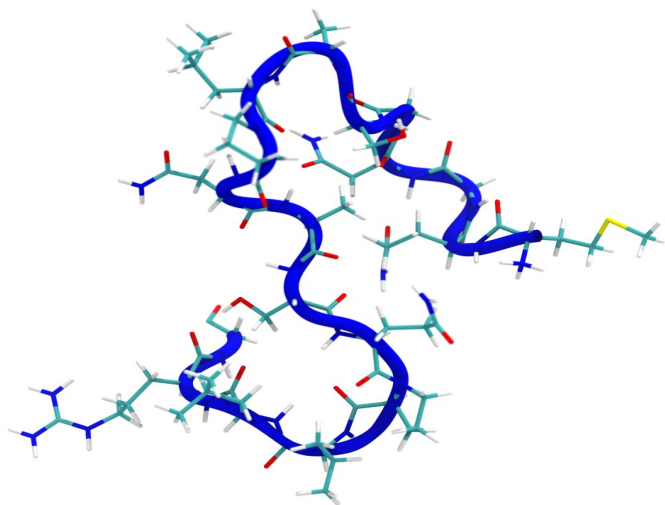
Beta Loop A,
Met1 - Gly18.
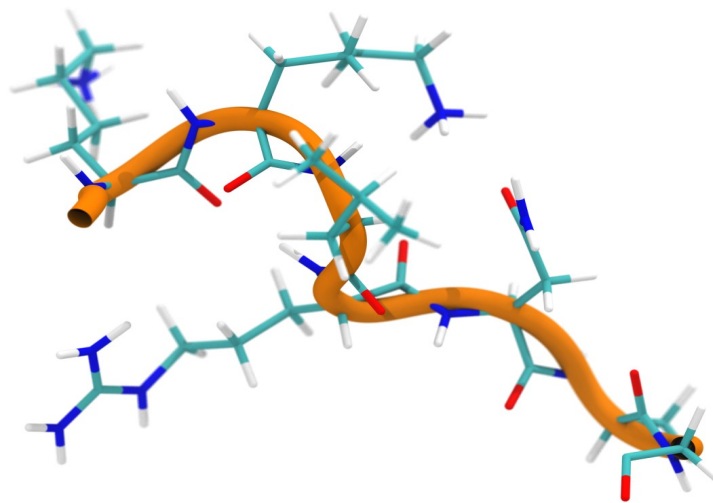
Beta Loop B,
Lys56 - Gly62.
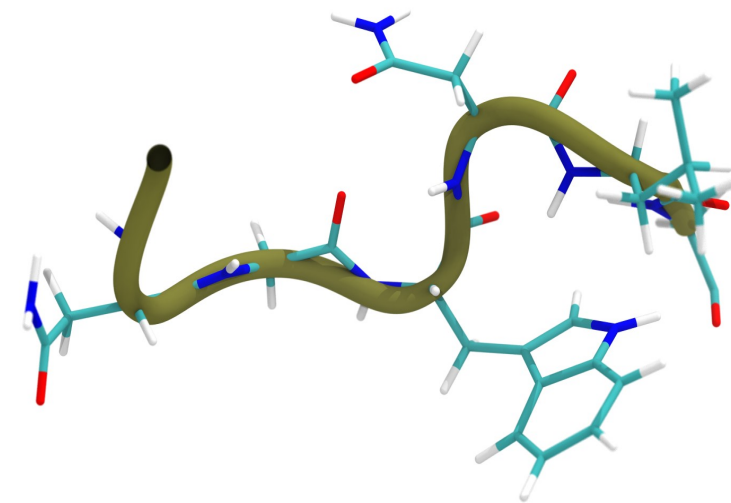
Beta Loop C,
Asn91 - Gly96.
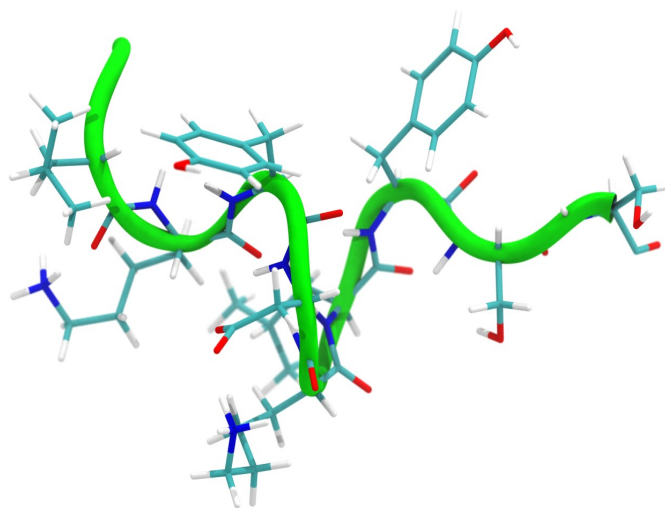
Beta Loop D,
Leu133 - Ser141.
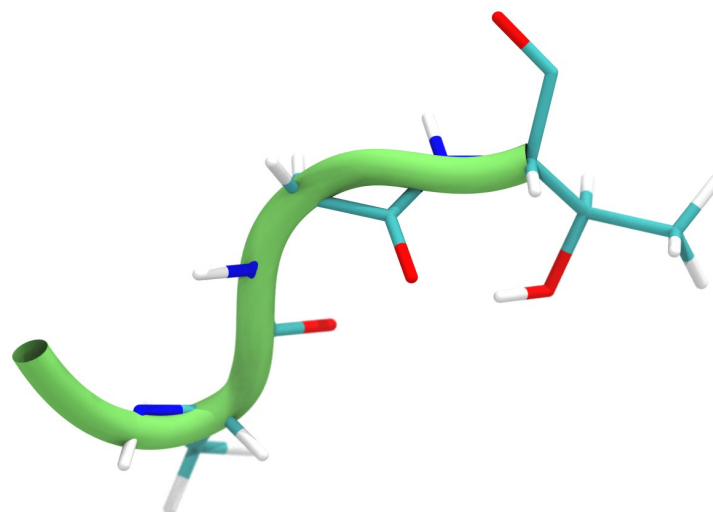
Beta Loop E,
Ala165 - Thr167.
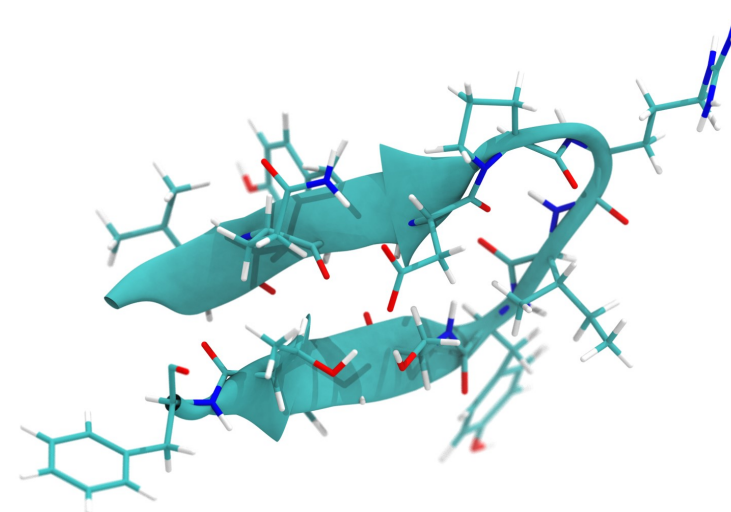
Beta Turn A,
Leu168 - Phe179.

Figure 4.1.2   *MEL1A Characterization, Beta Loop and Turns Part 1.*
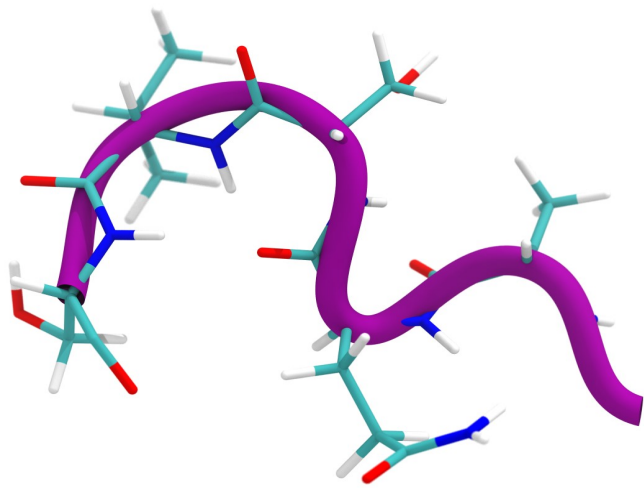
Loop F, Ala180 - Ser184
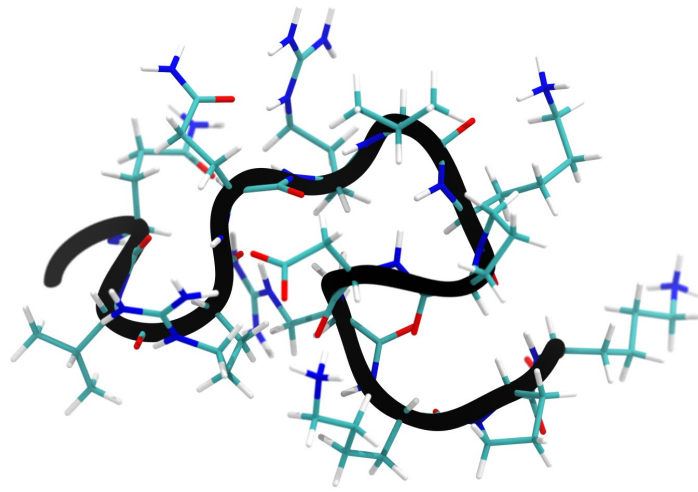
Loop G, Glu216 - Lys228
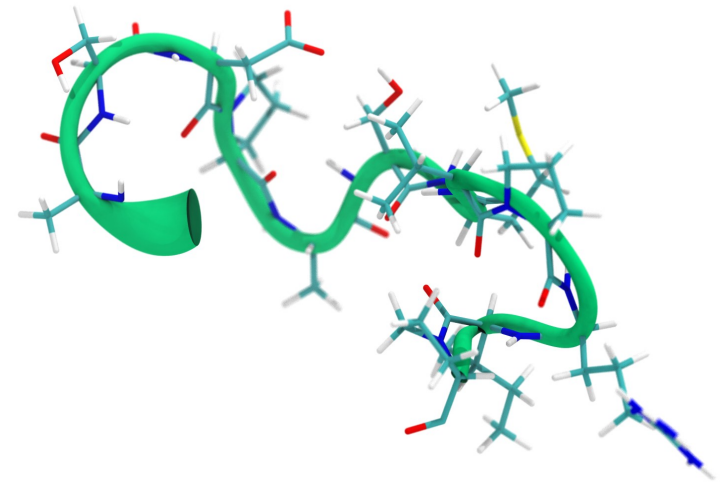
Loop H, Ala262 - Pro273
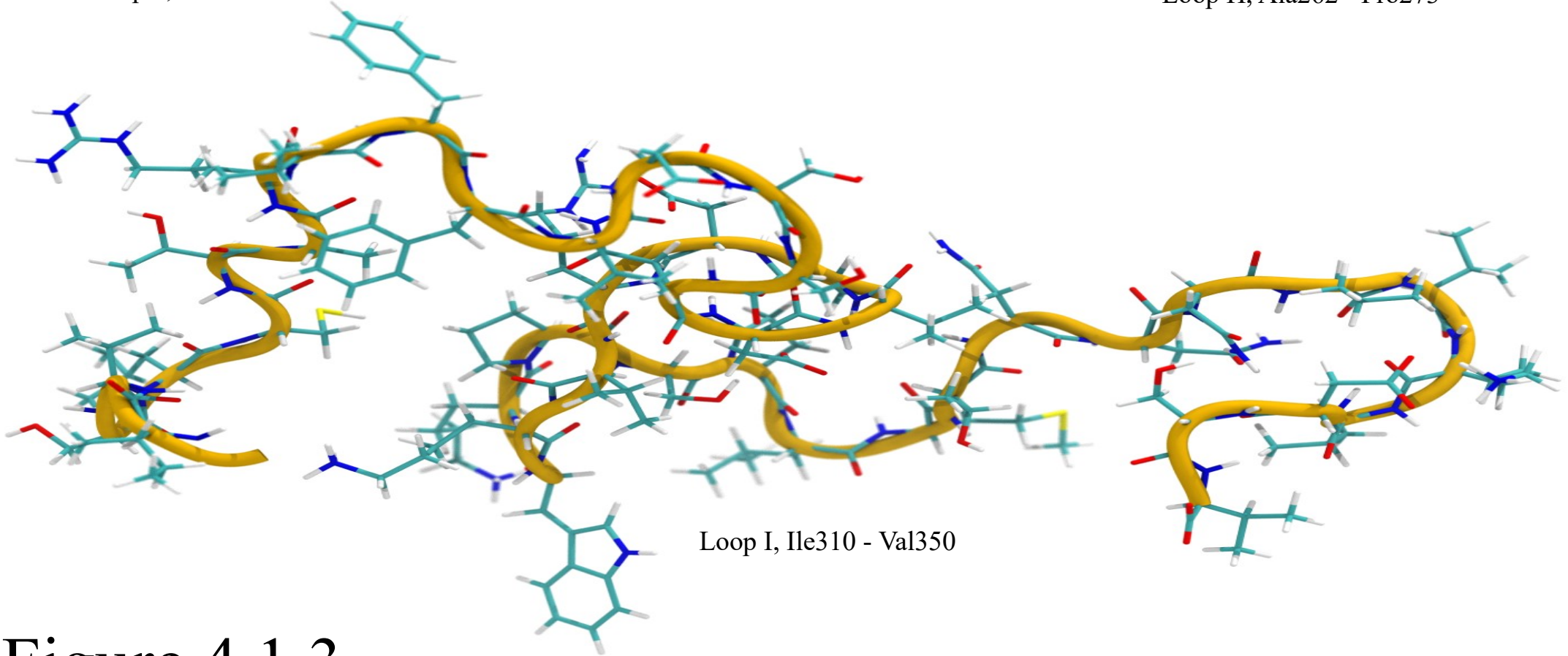
Loop I, Ile310 - Val350

Figure 4.1.3 *MEL1A Loop Sections Part 2*

Initial I-Tasser Homology model of the Mel1a and β2 Adrenergic receptors.

Initial I-Tasser Homology model of the Mel1a and Rhodopsin Receptors.
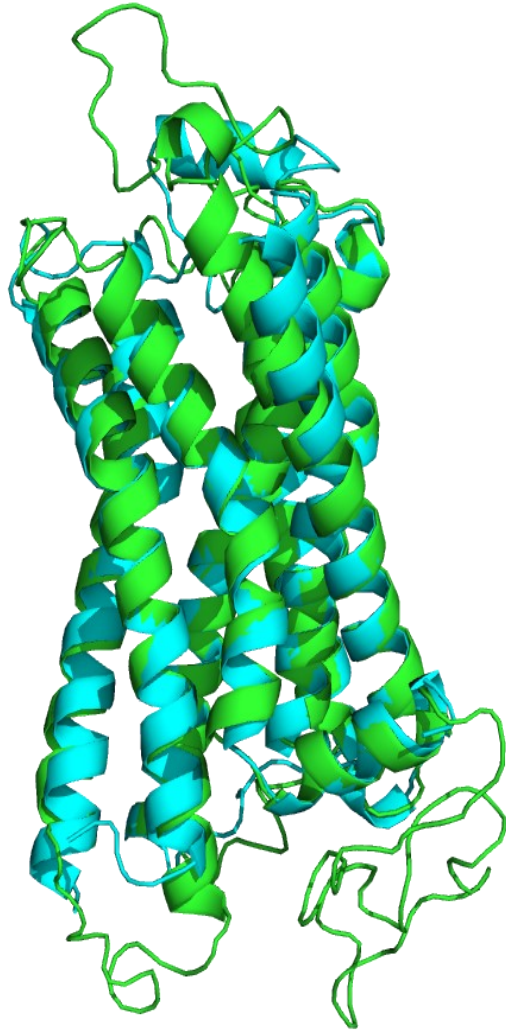
1. Mel1a GCPR shown in Green, X-ray crystal structure of β2-adregernic GCPR shown in Cyan. RMSD between the structures is 1.006Å. Models are shown in representation : cartoon
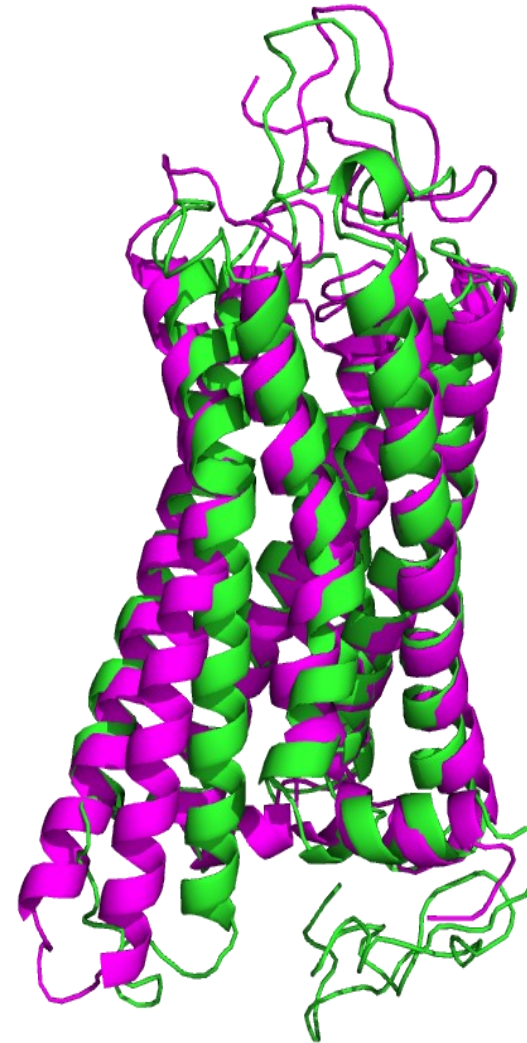
2. Mel1a GCPR shown in Green, Rhodopsin GCPR shown in Magenta. RMSD between the structures is 2.170Å Models are shown in representation : cartoon

Figure 4.2.1 Structural alignment to X-ray Crystal Structures of β2-Adregernic Receptor and Rhodopsin G-Coupled Protein Receptors.

Extracellular View and Electronegative Pocket.
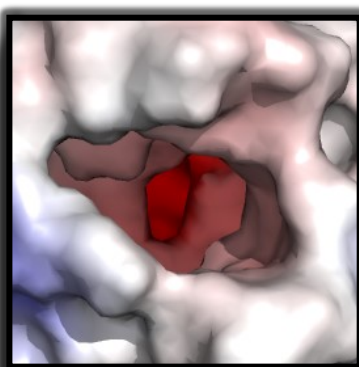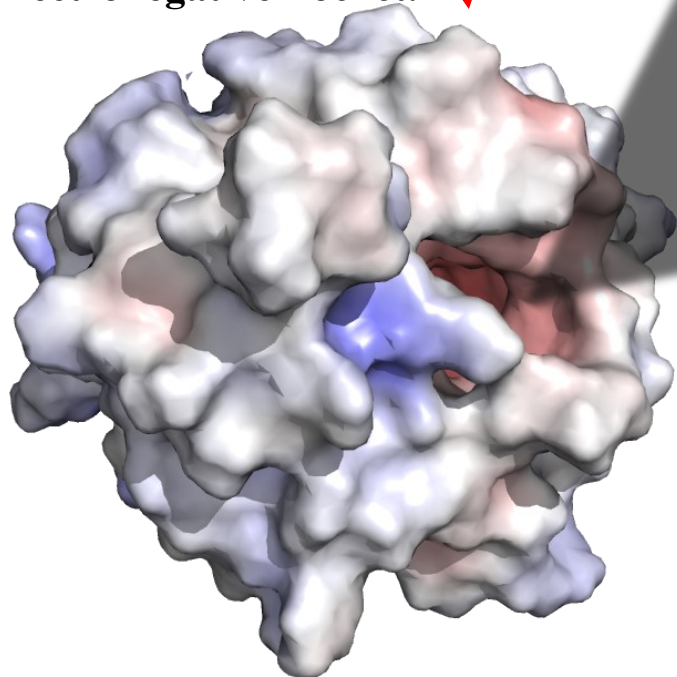
(Above) Electronegative pocket which could potentially be a binding site for small molecules.

Pocket image taken at 45 degree angle from original image (left).

Intracellular view.
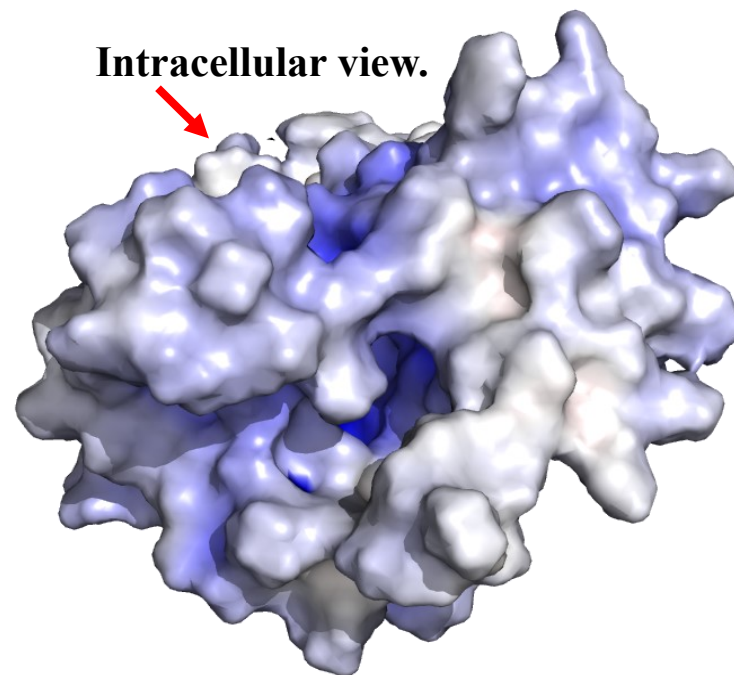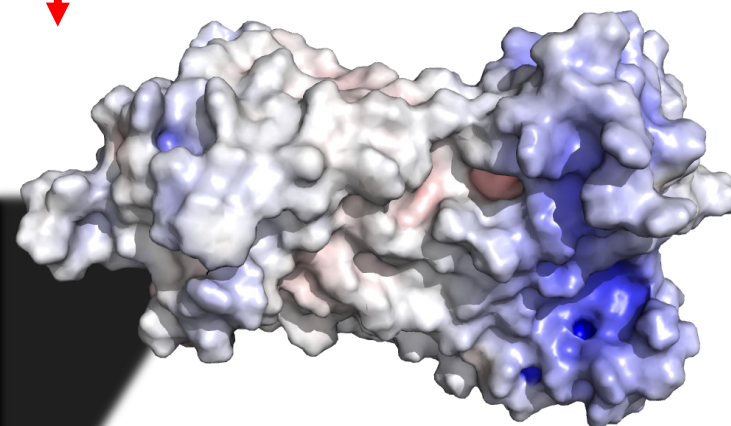
−10.000   10.000

−10.000   10.000

Side View (membrane portion).

Electrostatics images rendered in Pymol using the APBS electrostatics plugin utilizing the AMBER forcefield—to assess the surface interaction properties of Mel1a. Areas displayed in red are electronegative, Areas shown in blue are positive. Neutral areas have a white appearance.
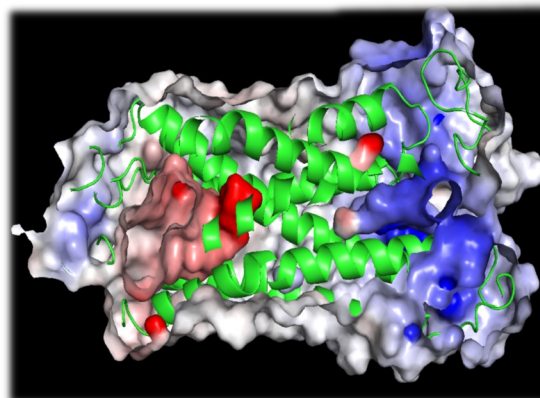
Side Slice of Mel1A to show intracellular and extracellular pockets. (blue intracellular, red extracellular)

Figure 4.3.1.   - Electrostatic analysis of the Initial I-Tasser Homology Model of the Mel1A GPCR.

5.0 Minimization of Mel1a GPCR System by Conjugant Gradient.

Data extracted from the NAMD log file and was analysed using VMD 1.9.3 NAMD plot module with a stride of 1, incorporating the entire dataset for the minimization calculation (frames 0 to 2,500).

Bond energy for the entire system had an average over the 2500 timesteps of 680.21 Kcal/mol with an initial starting energy of 4457.61 Kcal/mol followed by an initial spike in energy to 14,046 Kcal/mol between timeframes 15 and 20 within the initial 50 timesteps), and a final bond energy of 658.76 Kcal/mol in frame 126 (or timestep 2500). The calculated RMSD for the protein backbone was as follows: average of 0.134Å, a minimum of 0.046Å, a maximum of 0.154Å and a standard deviation of 0.013Å. The RMSD visualiser tool was also used to plot a chart over time showing convergence at the 50th frame for the selection 'protein' (Figure 5.0.1). Bond angles have an average of 3,126.17 Kcal/mol with an initial value of 8,739.38 Kcal/mol and a final value of 3,822.01 Kcal/mol. Dihedral angles show an average of 7,510.41 Kcal/mol, an initial of 8,523.76 Kcal/mol and a final frame figure of 8,115.71 Kcal/mol. Improper data shows an average value of 77.35 Kcal/mol, an initial value of 121.62 Kcal/mol and a final frame value of 90.89167 Kcal/mol Figure. Potential energy of the entire system has an average value of 25,856,661.43 Kcal/mol which reduces significantly from its initial value of 99,999,999,999.99 Kcal/mol and results in a final value of -82,206.3318 Kcal/mol (Figure 5.0.2).

<u>Summary</u>.

The data suggests that the addition of the POPC membrane, ligands and solvent, the system needed minimization to conform, as the potential energy was high. It has minimized all energy variables in the system and given the POPC membrane a degree of randomization. As Van Der Waal (VDW) interaction forces from the GPCR structure and the POPC membrane rebalance, the potential energy decreases and becomes negative. This is due to the way the system was built in VMD, the POPC membrane is placed over the protein and lipids were removed that were closer than 0.6Å.

5.1. Heating of the Mel1a GCPR Control System to 310K.

Data was extracted from the NAMD log file and was analysed using VMD 1.9.3's NAMD plot module with a stride of 1, incorporating the entire dataset for heating (steps 0 to 310,000). RMSD for the protein backbone was calculated using the RMSD trajectory tool module in VMD 1.9.3 and has the following values: an average of 0.128Å, minimum 0.015Å, maximum 0.189Å and a standard deviation of 0.041Å. Bond energy data over the simulation of the Mel1a GCPR system gives an average figure of 1,463.64 Kcal/mol, with an initial of 472.1 Kcal/mol and a final figure of 2,529.75 Kcal/mol (Figure 5.1.1). Potential energy shows an average figure of -99,921.81 Kcal/mol with an initial reading of -113,045.54 Kcal/mol and a final figure of -82,061.99 Kcal/mol. Van Der-Waal shows an average interaction energy of 8,065.20 Kcal/mol, an initial of 9,321.47 Kcal/mol and a final of 4,656.60 Kcal/mol (Figure 5.1.2) for the entire system.

Summary.

Temperature data shows the increase in degrees Kelvin over the timesteps, starting with the initial timestep at 0 K, and finishing at 307.74 K as an average figure for the entire system (Figure 5.1.3). This shows that the protein system is ready for equilibration at 310 K

5.2 Rescale Equilibration of the Mel1a Bi-Lipid Membrane System Analysis.

After equilibration was completed an RMSD trajectory analysis was performed on the data accrued. The average RMSD of the protein backbone was 0.255Å with a standard deviation of 0.004Å over 2500 timesteps. (minimum 0.243Å maximum 0.267Å). Using the timeline plugin in the VMD 1.9.3 the protein was coloured by RMSD (Figure 5.2.1) visually combining the data of the atom trajectories of residues over the course of the calculation. Maximum movement of the protein is evenly distributed since the use of harmonic constraints preserves the structural integrity of the model during equilibration. This is evident in that fact that the greatest RMSD is 0.91Å. Regions in the protein of high RMSD areas are shown in red, while regions of low RMSD in drawn in blue. A Maxwell-Boltzmann temperature distribution of the Mel1a System was also performed in order to assess the velocity distribution of the system (Figure 5.2.2). From this the temperature can then be derived using the sum of all kinetic energy within the data. The velocity data was compiled into a graph and then normalized. The temperature of the system following the equation showed a system temperature of 310.023º K with a narrow distribution of temperature.

NAMD log file data was also analysed to check system stability using the NAMD plot module available in VMD 1.9.3. Average Kinetic energy in the system was 28,376.06 Kcal/mol with an initial value of 28,171.63 Kcal/mol and a final value of 28,490.36 Kcal/mol. The temperature in K of the system shows an average value of 309.99 K, with an initial temperature of 307.73 degrees Kelvin, and a final temperature of 311.21 K (Figure 5.2.3). Potential energy in the system is on average -86,395.81 Kcal/mol. Average Van der Wall interaction energy is 3,304.05 Kcal/mol with an initial value of 4,656.6 Kcal/mol and a final value of 3,197.06 Kcal/mol (Figure 5.2.4). Bond energy of the entire system through equilibration shows an average figure of 2,457.94 Kcal/mol with an initial figure of 2,529.75 Kcal/mol and a final figure of 2,461.34 Kcal/mol (Figure 5.2.5).

Summary.

Limited deviation in the above variables suggest that the system has reached relative equilibration at the desired temperature of 310 Kelvin within the confines of the simulations timesteps. The RMSD overlay suggests there are no visible heat spots in the system, while the data involving kinetic and potential energy, Van der Waal interactions and bond energy support the previously shown Maxwell-Boltzmann data and that the temperature stays stable for the duration of the calculation.

5.3 Analysis of the Mel1a 10ns Trajectories.

The final frame from each set of trajectory data was analysed using Pymol 2.1.1 for electrostatic analysis and structural alignment. RMSD trajectory data was obtained using VMD version 1.9.3 and visual inspection of the trajectories was also undertaken using VMD to note any zones in the structure where consistent conformational change was observed. The beta2 adrenergic and rhodopsin X-ray models were also used as reference for structural comparison.

Mel1a 10 ns trajectory data was uploaded into VMD 1.9.3 for interpretation. Using the trajectory analysis tool, the trajectory data was aligned (As the subject to be analysed is the protein structure alone, which in MD conditions moves in the respective X, Y and Z co-ordinates independently) to the protein reference frame (frame 0). Run 1 showed an average RMSD of 4.613Å with a minimum of 0.942Å a maximum of 5.663Å, and a standard deviation of 0.799Å. Run 2 gave an average RMSD of 4.417Å, a minimum RMSD of 0.925Å a maximum at 5.326Å, with a standard deviation of 0.766Å. Run 3 exhibited an average RMSD of 4.674Å a minimum of 0.948Å and maximum of 6.361Å with a standard deviation of 1.08Å. The VMD timeline plugin was used to analyse RMSD of the protein within the system and the output data was used to colour the protein in each of the three MD runs by RMSD (Figure 5.3.1). Areas in red represent high deviation in the proteins residue positioning. Blue areas show residues which do not deviate far from their initial positioning and are mostly seen in the of the core of the protein in all three calculation datasets. The data suggests the core of the protein is stable (alpha helix sections). Most movement is seen in the loop sections which are exposed to solute. Although the maximum RMSD is similar to the previous two runs, it appears the core of the protein is searching conformational space to a higher degree than the previous runs on the end portions of a few alpha helix sections. Although most of the high RMSD remains in the loops the RMSD in the helix sections suggesting a higher degree of deviation from the initial conformation. This suggests it needs more time to search conformational space.

Run 3 shows conformational change in the helix-turn-helix G (Figure 4.1.1) in comparison to the previous two 10ns runs (runs 1 and 2). Although the RMSD of the entire protein is not significantly different to the other sets of trajectory data (Run 1: 4.613Å, run 2: 4.417Å, run 3: 4.674Å) most of the conformational motion is observed in the loop sections before, in between and after the 7 helix trans-membrane portions of the protein on the intracellular and extracellular sides. These sections are exposed to solvent and therefore it is expected that movement would be greater in these regions as loop sections are intrinsically flexible.

RMSD data against simulation frame (Figure 5.3.2) is also presented for each trajectory and the data shows three similar profiles. A deviation in RMSD is observed in run 3 where a RMSD of 2Å is evident between frames 1500 to 2500, compared to runs 1 and 2.

The log files from the three 10 ns Mel1a trajectory datasets were analysed using the NamdPlot module available in Vmd 1.9.3 and representative charts of the data were generated. The resultant figures from all three trajectories are shown in the table below (Table 5.3.1) data can also be viewed in graph form in figures 5.5.1 and 5.5.2 for 10ns for run 1, 5.10.1, 5.10.2 for 10 ns run 2 and 5.15.1 and 5.15.2 for 10 ns run 3;

Table 5.3.1	Initial.	Average.	Final.
Kinetic energy (10ns run 1).	28,433.70 Kcal/mol	31,180.79 Kcal/mol	31,102.7 Kcal/mol
Kinetic energy (10ns run 2).	28,433.59 Kcal/mol	31,361.43 Kcal/mol	31,419.24 Kcal/mol
Kinetic energy (10ns run 3).	28,433.7 Kcal/mol	31,183.43 Kcal/mol	31,343.69 Kcal/mol
Potential energy (10ns Run 1).	-87,607.42 Kcal/mol	-84,993.63 Kcal/mol	-84,909.59 Kcal/mol
Potential energy (10ns Run 2).	-87,607.42 Kcal/mol	-86,178.61 Kcal/mol	-87,607.42 Kcal/mol
Potential energy (10ns Run 3).	-87,607.42 Kcal/mol	-84,990.66 Kcal/mol	-85,119.47 Kcal/mol
Bond energy (10ns Run 1).	2,553.98 Kcal/mol	6,335.65 Kcal/mol	6,481.68 Kcal/mol
Bond energy (10ns Run 2).	2,553.98 Kcal/mol	5,516.88 Kcal/mol	5,569.43 Kcal/mol
Bond energy (10ns Run 3).	2,553.98 Kcal/mol	6,333.26 Kcal/mol	6,353.42 Kcal/mol
Temperature in Kelvin (10ns Run 1)	280.16 K	307.2 K	306.46 K
Temperature in Kelvin (10ns Run 2)	280.16 K	309.01 K	309.58 K
Temperature in Kelvin (10ns Run 3)	280.16 K	307.23 K	308.83 K

Table 5.3.1 – A summary of the data collected from all 10 ns molecular dynamics runs.

5.4 Mel1a 10ns Run 1.
5.6 Visual analysis of the Mel1a protein structure in 10ns Run 1.

Upon visual analysis of the protein in MD run 1 (Figure 5.6.1), the following characteristics were observed. Mel1A loop A (on the extracellular side of the membrane) moves closer to the core of the protein as the simulation progresses, resulting in a packed conformation at the end of the trajectory. Cys100 and Cys177 form a disulphide bridge between beta loop turn A and alpha helix C limiting the movement of this loop section throughout the calculation. Helix-turn-helix G bends on the upper residues between Tyr285 and Leu290 in the final frame of the simulation. On the intracellular side of the membrane loop I (the longest loop section on the intracellular side) residue Cys314 seems to be traversing the membrane layer in an outwards direction from the protein core. Met339 is in an unfavourable position (due to being hydrophobic) in solute for the entire duration of the run and extends further from the core as the simulation continues initially 16.70Å distance from Asn55 on alpha helix A, and a final frame distance of 24.21Å from the same residue. In conjunction with this conduct, loop G moves away from the core of the protein on the intracellular side. Lys222 which moves away from the core, measured from Val214 on alpha helix E: start distance of 13.25Å and a final frame distance of 18.01Å.

5.7 Structural Alignment Analysis of Trajectory 1.

Structural alignments with the original homology model for the first trajectory can be seen in Figure 5.7.1. The *align* command was used in Pymol 2.1.1 to structurally align the initial Mel1a homology model (from I-Tassier) shown in green, with the final frame of the 10ns Mel1a trajectory 1data set shown in red (Part A). The RMSD between the structures is 3.03Å with most of the change being observed in the loop sections of the Mel1a protein. This suggests that the proteins loop sections (post homology modelling) were not in a favourable conformation and that the 10ns data set has allowed the structure to dynamically search for more appropriate loop conformers. Part B of figure 4.7.5 shows the structural alignment between the β2-adrenergic GCPR (cyan) and the first 10ns run of Mel1a (red). the RMSD between the structures is 2.626Å which is comparatively smaller between part A and part B. However, the loop sections of the protein are still largely very divergent. In the crystal structure of the β2 adrenergic GCPR many of the loop domains were not resolved most likely due to their conformational flexibility. Therefore, the lack of loop data in this X-ray structure may account

for the lower RMSD between it and the Mel1a since the calculation is now focused on the transmembrane helices. Part 3 shows the structural alignment between the rhodopsin receptor (shown in blue) and the final frame of Mel1a run 1 (red). With an alignment RMSD of 3.517Å similarly most of the difference seen in the structures is within the loop domains of these structures and like the crystal structure of β2 – sections of loop structures are missing on the intracellular domain of the model.

Since crystal data for the GCPR superfamily is limited it is possible that the homology model was created using data of both the rhodopsin and β2 adrenergic GCPR's. Both receptors also sit within the Rhodopsin-like receptor family (Mel1a within subfamily A9, β2 in subfamily A17). And may account for the low RMSD figures, and the high similarity in the core sections of the GCPR's

5.8 Electrostatic Analysis of Mel1a 10ns Run 1.

Electrostatic surface analysis was performed using Pymol. The .PDB was converted to a PQR file using the Pymol plugin for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM. Molecular surface visualisation was altered to show electrostatic charge potential of -/+ 10 (from the default value of 5). Areas shown in red are electronegative. Areas shown in white are neutral while blue areas of the surface are positive. Dimensions of the electrostatic surface are shown in figure 5.8.1. The extracellular portion of the protein shows a hue of small electro negative and positive charges but is mostly neutral except for an electronegative pocket which traverses through the core of the protein to the intracellular side where it finishes. The side profile of the protein which is situated within the POPC membrane of the simulation is predominantly neutral. Both the extracellular and intracellular portions of the protein can be seen in the side view and positive charge is more visible on the intracellular side.

5.9 Mel1a 10ns Run 2.

Visual inspection of the Mel1a GCPR was undertaken using VMD to note any key characteristic conformations similar to the first 10ns run (Figure 5.11.1). Both the intracellular and extracellular loops seemed better packed overall than the initial calculation (run 1). The di-sulphide bond seen in the initial calculation is present in the second simulation between loop A and helix C (Cys100 and Cys177). Met339 on the intracellular loop I, is still exposed to solute in the same fashion as the previous run. Cys314 also located on loop I, remains unpaired and searching the surface of the membrane showing similar behaviour to the first run moving away from the core of the protein. There is a large pocket on the extracellular side of the protein which is very clearly visible. Loop D has conformed into an alpha helical structure and no longer resembles a loop. On the extracellular side Met1 on loop A and Met268 on loop H are exposed to solvent.

5.12 Structural Alignment Analysis Run 2.

Following the same analysis pattern in Pymol performed for Run 1 (section 5.7), Mel1a 10ns run2 was structurally aligned with the homology model, β2 adrenergic and rhodopsin GCPR's (figure 5.12.1). Part A shows the 10ns Run2 in red, and the homology model of Mel1a in green. The RMSD between the structures is 3.203Å. in comparison to 10ns run 1, a lot of the deviation is seen in the loop portions of the protein which is expected for the same previously stated reasons in section 5.7. while the core sees little conformational movement there are some visible areas in the core alpha helix sections conformational change has occurred. Part B shows the structural alignment between Mel1a run2's final frame which is shown in red and the β2 adrenergic GCPR shown in cyan. The deviation between the structures is 2.626Å. Although the RMSD is lower than the structural alignment with the homology model of the same protein (part A), the crystallised 2 adrenergic GCPR is missing loops as previously mentioned and likely accounts for the lower figure (section 5.7). Part 3 shows the structural alignment of Mel1a Run2's final frame and the x-ray crystal structure of Rhodopsin. The RMSD between the structures is 3.517Å which is slightly higher than parts A and B. Loop D of Mel1a shows a similar conformation to the rhodopsin loop of the same orientation. Pymol chooses to define the Mel1a loop d structure as a loop rather than helical but has a very similar conformation to the rhodopsin helix which it is aligned to. Loop variation is again similarly to the other structural alignments and higher in comparison to the central core of the protein except for loop

D. This implies that the Mel1a GCPR is in the same family as the crystallised structures as it has little deviation in the central core.

From the final frame of the 2nd 10ns calculation, the Mel1a GCPR was removed from the system and saved as a new PDB for electrostatic surface analysis in Pymol using the APBS electrostatics plugin. The .PDB was converted to a PQR file for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM. The charge for the overlay was set to +/- 10 as done previously in all electrostatic surface maps. The resultant model is shown in figure 5.13.1. a large electronegative cavity is visible on the external side of the cell wall. Much larger than seen in the homology model electrostatics model (Figure 4.3.1) and the first 10ns run (Figure 5.8.1). the central channel which traverses through the core of the protein in the final frame of 10ns run 1 is not correlative with the final frame of run 2, suggesting a second conformation. As seen in the first run and the homology electrostatics analysis, the membrane portions of the core are persistently neutral in charge.

5.14 Mel1a 10ns Run 3.

5.16 Visual Analysis of the Mel1a Protein Structure in 10ns Run 3.

Visual inspection of the Mel1a GCPR was undertaken using VMD to note any key characteristic conformations similar to the first and second 10ns runs (Figure 5.16.1). Cys314 on the intracellular side of the GCPR consistently moves away from the core of the protein traversing the POPC membrane in the third dataset in the same fashion as the first two 10ns runs. Similarly, Met339 is exposed to solute on Loop I but the rest of the loop seems to be closer packed to the core of the protein. Loop G exhibits similar behaviour to 10ns Run 1 extending away from the core of the protein but to a lesser extent (initial distance between VAL214 on loop G and Leu222 on helix E of 11.82Å) at a final distance of 15.43Å. on the extracellular side, Met1 and Met268 sit exposed to solvent. The initial loop (loop A) Met1 (N-terminus of the protein) packs into the central core of the GCPR by the end of the simulation. Helix-Turn -Helix G condenses on the Z-axis at Asn299 to Ile309 and it moves closer to the previous part of the helix (Cys289 to Leu298) and folds in on itself. The benzene portions of Tyr295, Phe302, and Tyr306 stack by the end of the simulation in a diagonal conformation.

5.17 Structural Alignment Analysis Run 3.

The final frame of the 3rd Mel1a GCPR calculation was saved as a new PDB file with the solvent and bi-lipid membrane POPC layer removed in VMD. Pymol was then used to structurally align the initial homology model, the β2 adrenergic and rhodopsin GCPR's. the structural alignments are shown in figure 5.17.1. Part A shows the structural alignment between the initial homology model of Mel1a shown in green with the 10ns MD model being shown in red. The RMSD between the structures is 3.612Å. the overlay suggests that for the most part. Dynamic change is again similarly to the data from runs 1 and 2, most of the variation resides in the loop sections on both sides of the 7TM trans-membrane portion of the protein showing little deviation in the core of the protein. The alignment Part B shows the β2 adrenergic GCPR in cyan. RMSD between the aligned protein structures is 3.340Å. The core of the aligned structures shows a similar conformation. The loop sections are varied and Loop D on the Mel1a structure is represented as a Helix structure which is absent from the β2 structure and still appears as a loop. On the extracellular side (top half of the structure) the β2 adrenergic receptor has a horizontal Helix section within the loops which does not exist within the Mel1a structure. Part C shows the alignment of the Rhodopsin crystal structure in Blue with the 3rd 10ns MD run. The RMSD between the aligned structures is 4.449Å. Loop sections with the exception of Loop D show large variation in the structural alignment. Rhodopsin and the Mel1a structure share the helix-like loop section. the helix turn helix G in Mel1a is spun on its axis with the end of the final turn in the helix facing away from the proteins core, unlike rhodopsin which sits closer to the core.

5.18 Electrostatic Analysis of Mel1a 10ns Run 3.

The final frame of the 3rd 10ns MD run was saved as a new PDB file with the membrane, ions and solvent removed. Using Pymol, the APBS electrostatics plugin was used to make a surface map of electrostatic potential. The .PDB was converted to a PQR file for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM (Figure 5.18.1). The extracellular side of Mel1a has no obvious electronegative pockets and is predominantly neutral in charge unlike 10ns runs 1 and 2 which had electronegative pockets of different sizes. This may be due to the N-terminus of the protein conforming inside the protein's central core as mentioned in the visual analysis. There is however an electronegative channel present in the

core of the protein similarly to 10ns run 1 (Figure 5.8.1). The intracellular portion of the Mel1a GCPR shows the cavity leading to the central electronegative channel and is mostly positive in charge on the side view harbours a slight negative charge compared to runs 1 and 2 which had neutral charge suggesting a different type of conformational change which exhibits negative potential in the core of the protein, possibly describing a function. It is likely a second conformation is being shown in 10ns run 3.

Summary.

After analysis of the structural characteristics of Mel1a post 10ns MD calculation. The largest change in conformation was seen in the intracellular and extracellular loop sections of the protein structure. It was decided to extend in triplicate the Molecular Dynamics simulations to 30 nanoseconds to allow the system to explore conformational space for longer (10ns to 30ns). Increasing the length of the simulation gives the system longer to find new conformations. The second 10 ns MD run displayed a second conformation. The extracellular side of Mel1a conformed presenting a large electronegative cavity, along with a distinct difference in bond energy of approximately 1000 Kcal/mol. However, all other variables remained within relative range of each other, supporting the need for a longer runtime of the system.
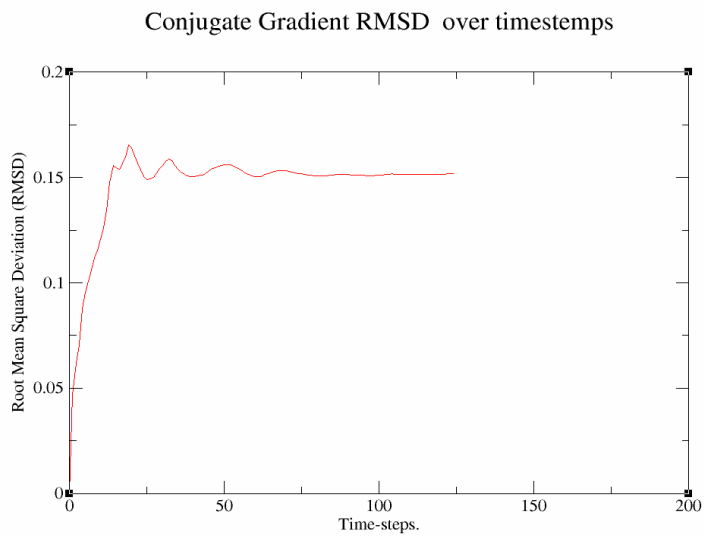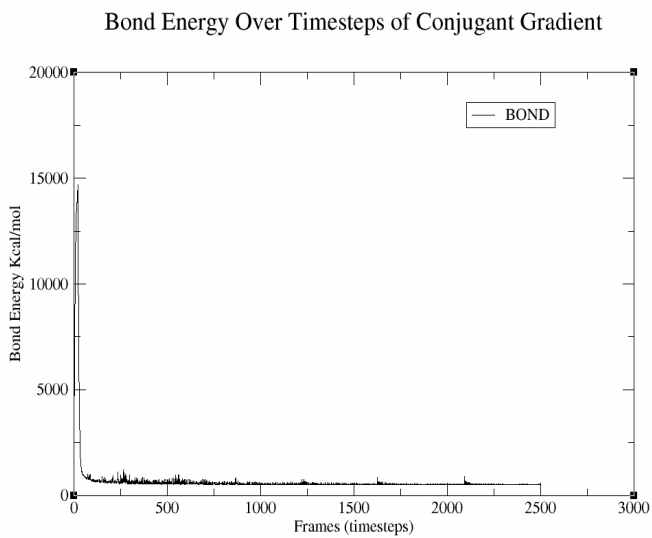
Figure 5.0.1– The above graphs show bond energy of the entire system (left) and root mean square deviation (RMSD) during the minimization simulation fopr the protein residues selected in the VMD RMSD visualization tool.
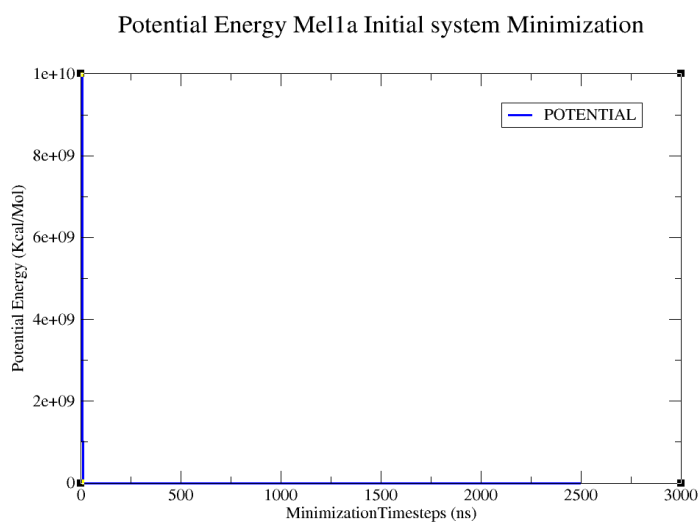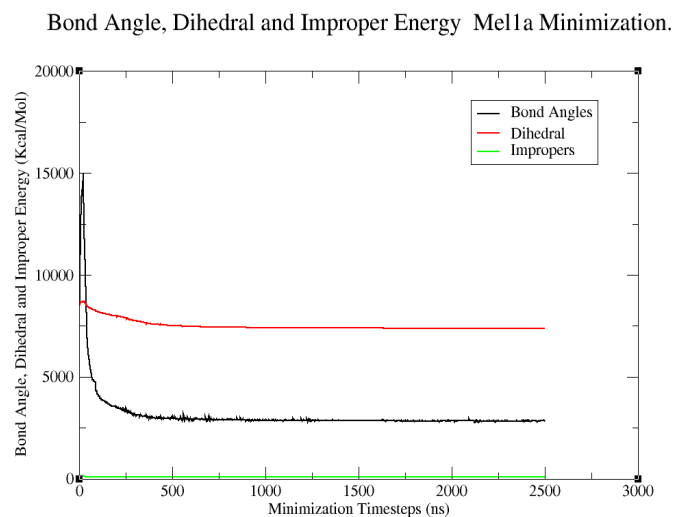
Figure 5.0.2 - Bond angle, Dihedral and Improper data from the minimization (left graph) and potential energy of the entire system (right graph).
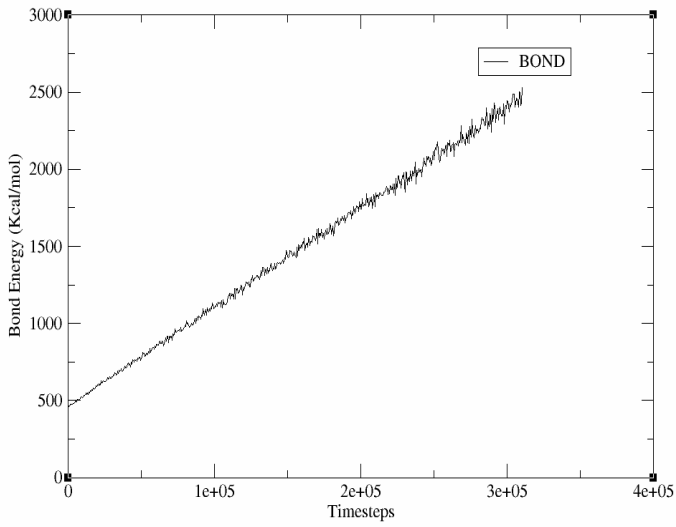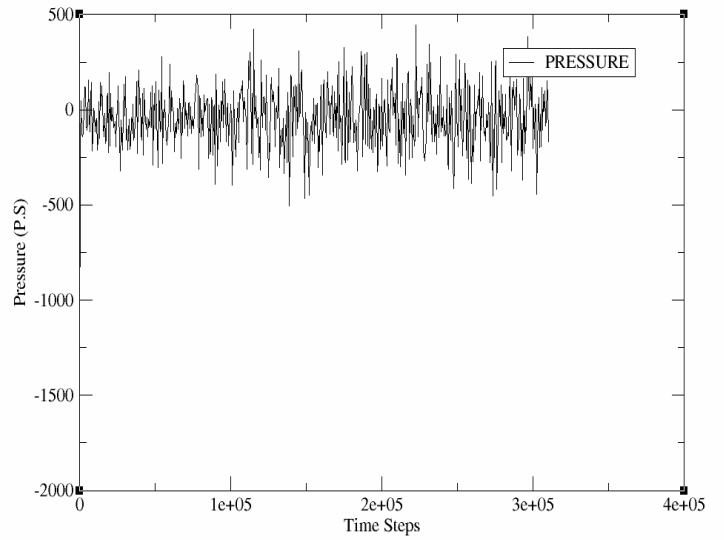
Figure 5.1.1- Bond energy over the timesteps of the heating phase for the entire Mel1a system (left) . The right graph displays to constant pressure variable attained through the NPT ensemble for which pressure should be constant the graph shows an average constant pressure as the ensemble describes.
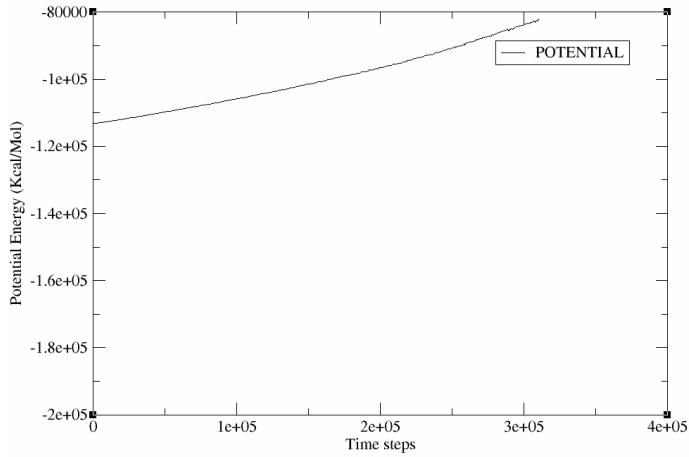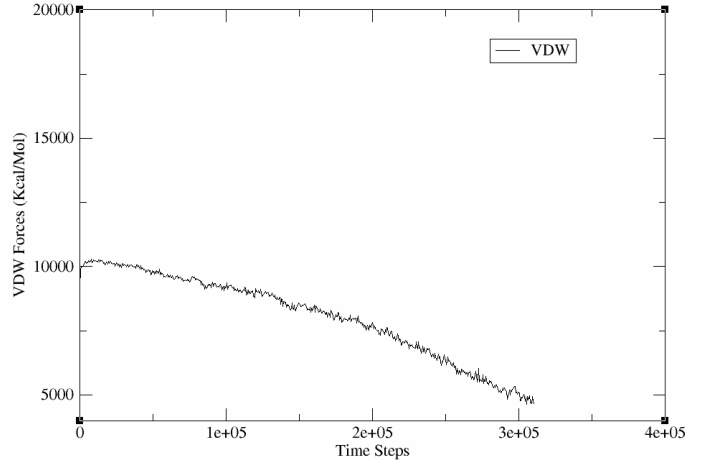
Figure 5.1.2 - Potential energy (left) throughout the heating phase. (Right) Van der Waal interaction data.
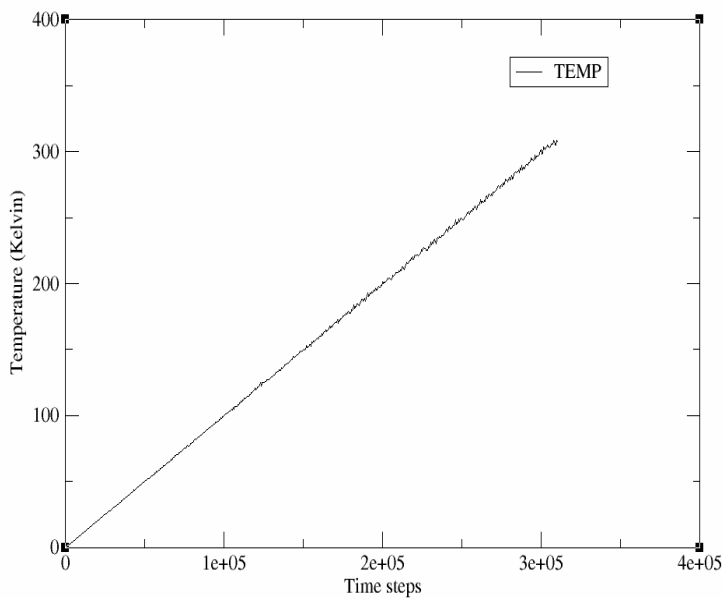
Figure 5.1.3 - Temperature in kelvin over 310,000 timesteps.
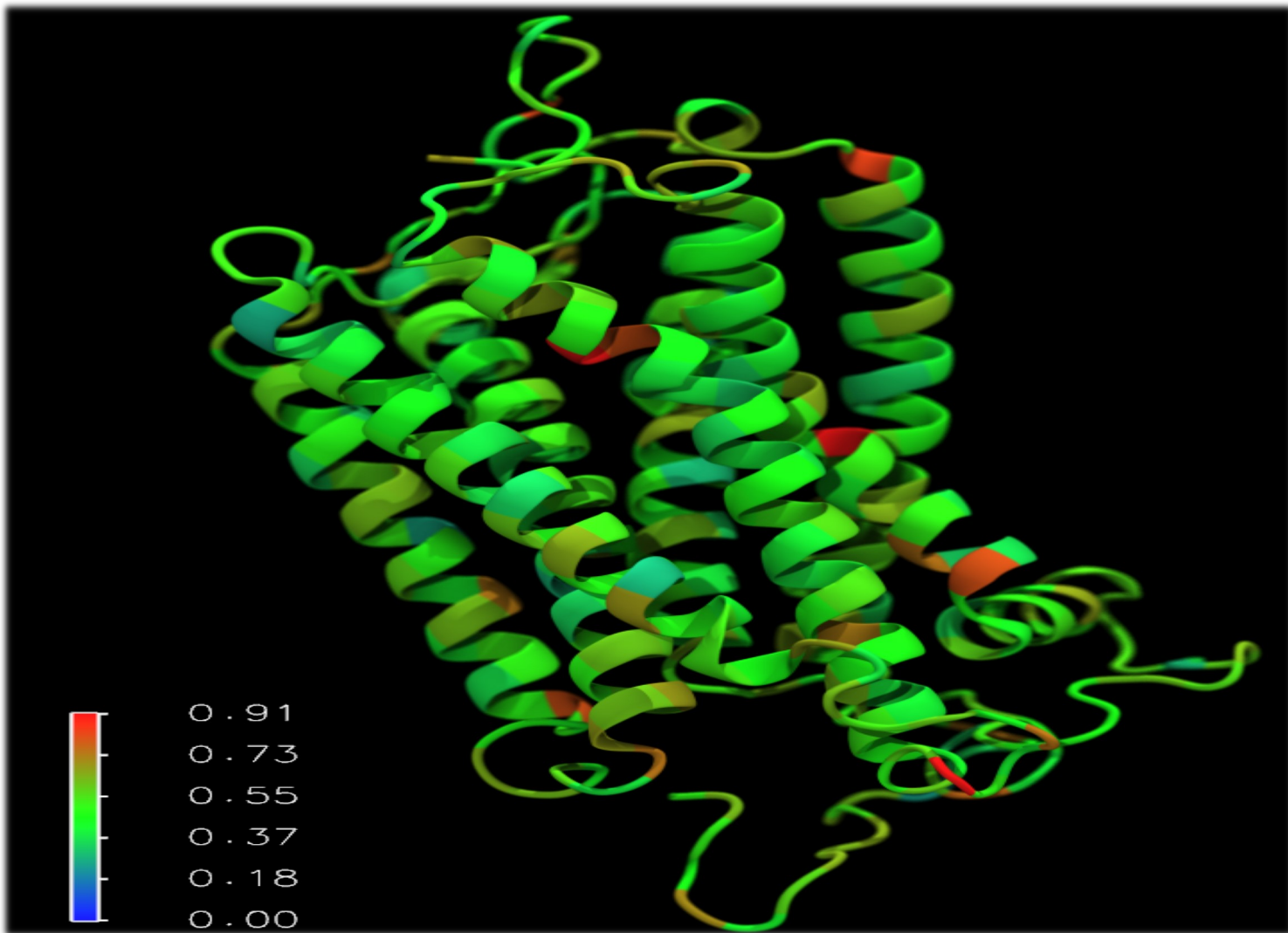
Figure 5.2.1 - Mel1a Coloured by RMSD achieved with the created DCD file of the simulation, drawing method New Cartoon. The scale bar represents the minimum and maximum RMSD in the structure.
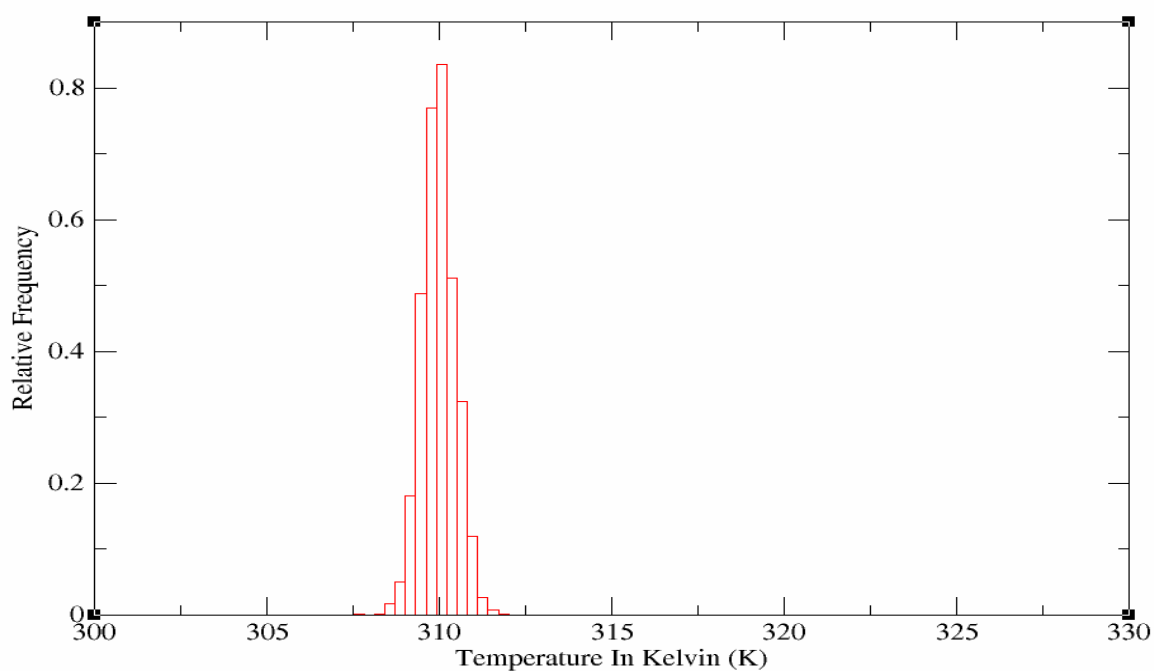
Figure 5.2.2—Temperature distribution in the Mel1a GCPR bi-lipid membrane system. The y-axis describes the frequency of temperature distribution. The X-axis shows the temperature in Kelvin (K).
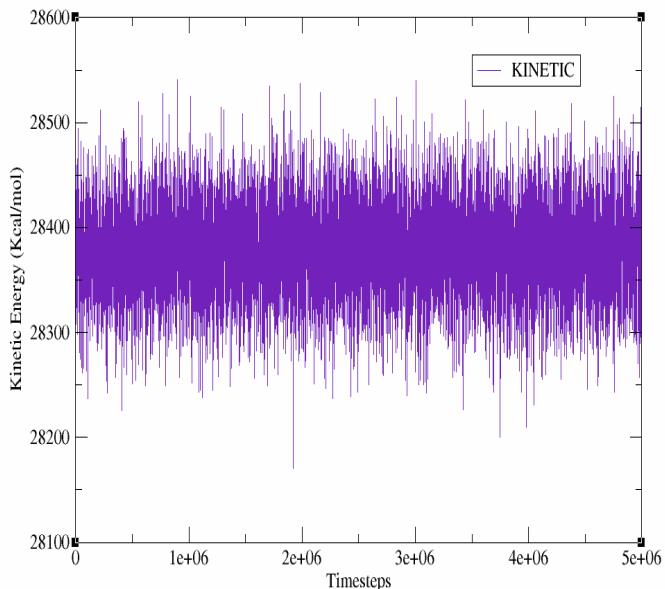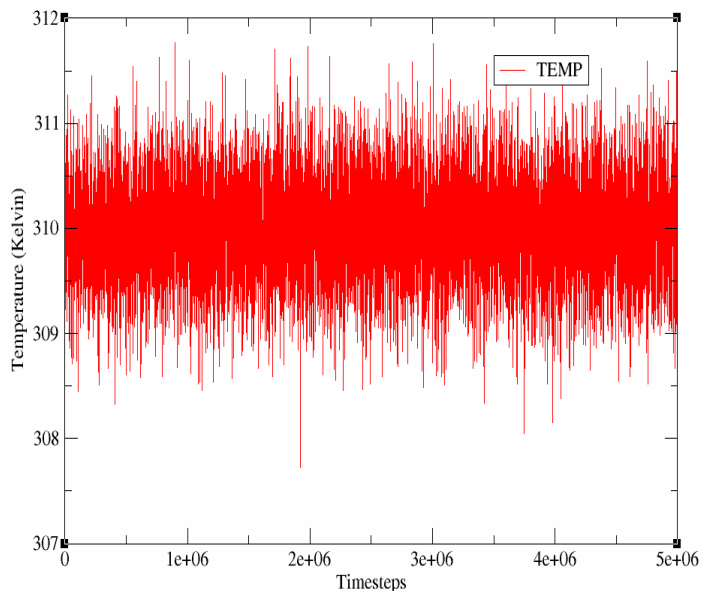
Figure 5.2.3 - Kinetic energy data from the entire Mel1a GCPR Bi-Lipid Membrane System (left) Average temperature of the system through out the calculation of the entire system (right).
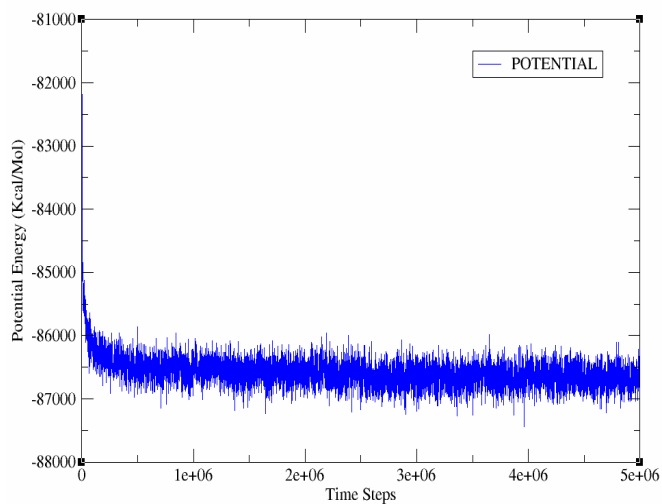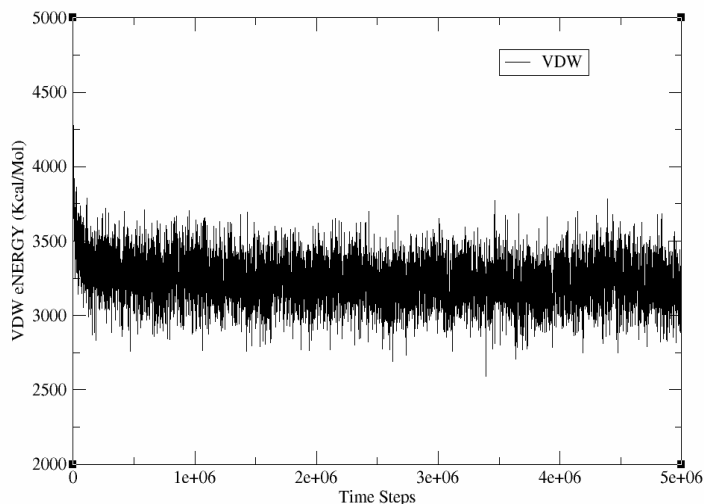
Figure 5.2.4 - Potential energy chart for the equilibration of the Mel1a GCPR bi-lipid membrane systes (left graph). Van Der Waal interaction energy plot for the previously mentioned system(right graph)
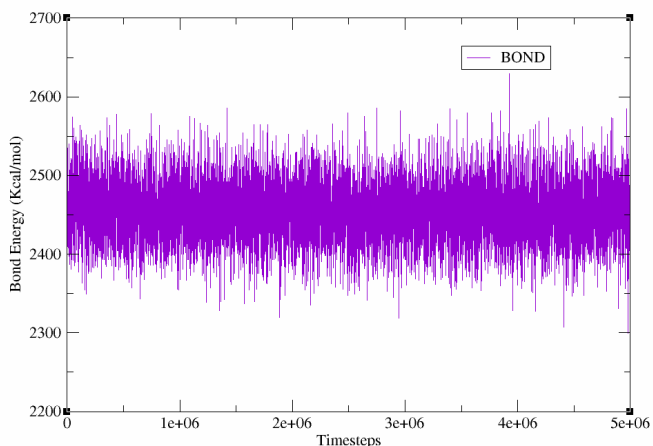
Figure 5.2.5 - Bond energy over the duration of the simulation for the previously mentioned system (left).
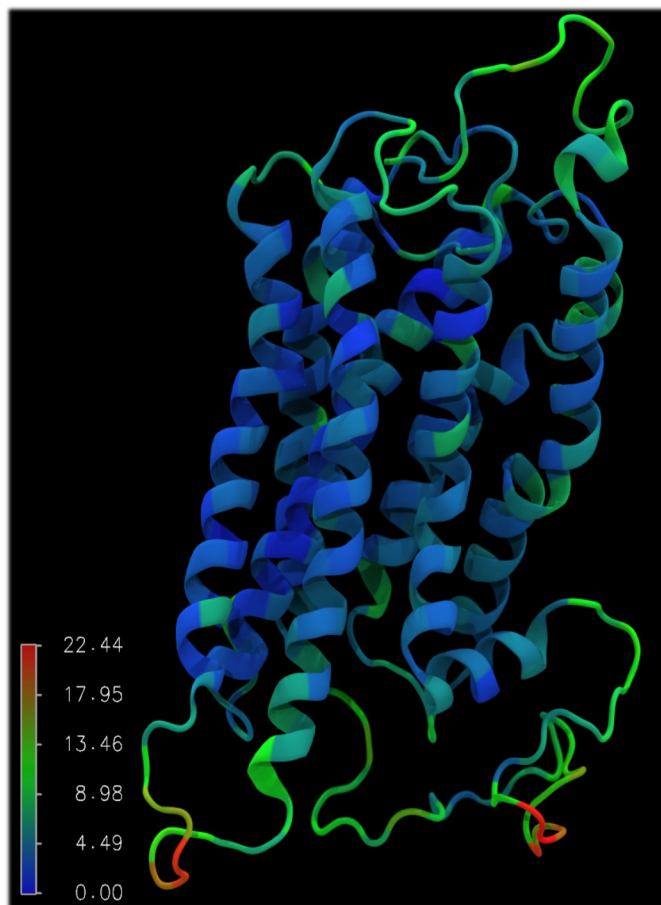
Run 1 10ns RMSD (root mean square deviation) trajectory by colour of Molecular Dynamics.
Red resembles high areas of movement Max 21.26 Å, blue resembles low areas of movement min of 0.00Å.
Protein is shown in Colouring Method: Trajectory > User, Representation: New Cartoon.
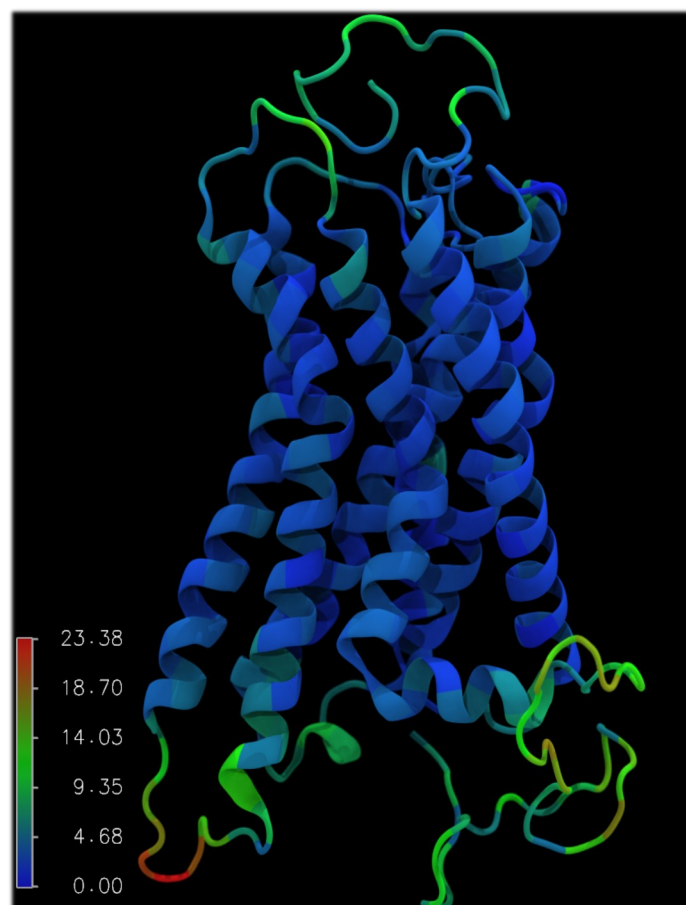
Run2 10ns RMSD (root mean square deviation) trajectory by colour.
High RMSD is seen in loop sections similarly to run 1, stability is seen in the core of the protein within the alpha helix sections. Most deviation in conformation is seen in the loop sections of the protein (maximum of 23.38Å).
Protein is shown in Colouring Method: Trajectory > User, Representation: New Cartoon.
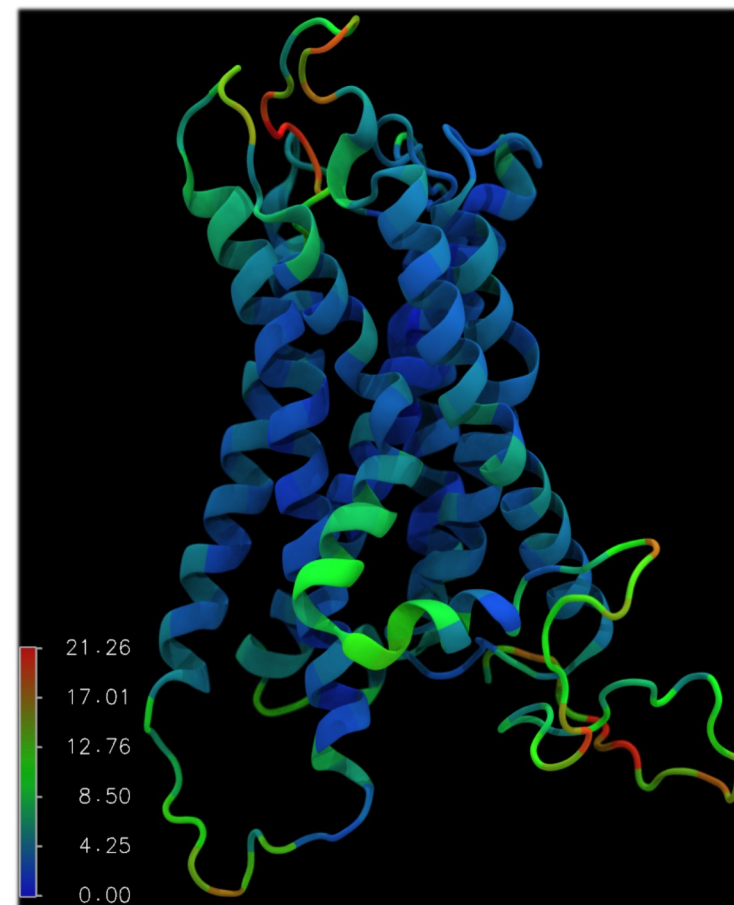
Run 3 10ns RMSD (Root Mean Square Deviation) trajectory by colour overlay.
Red resembles high RMSD, blue represents low RMSD. Maximum RMSD seen is 21.26Å.
Protein is shown in Colouring Method: Trajectory > User, Representation: New Cartoon.

Figure 5.3.1 – 10ns MD of MEL1A - Coloured by Trajectory RMSD.
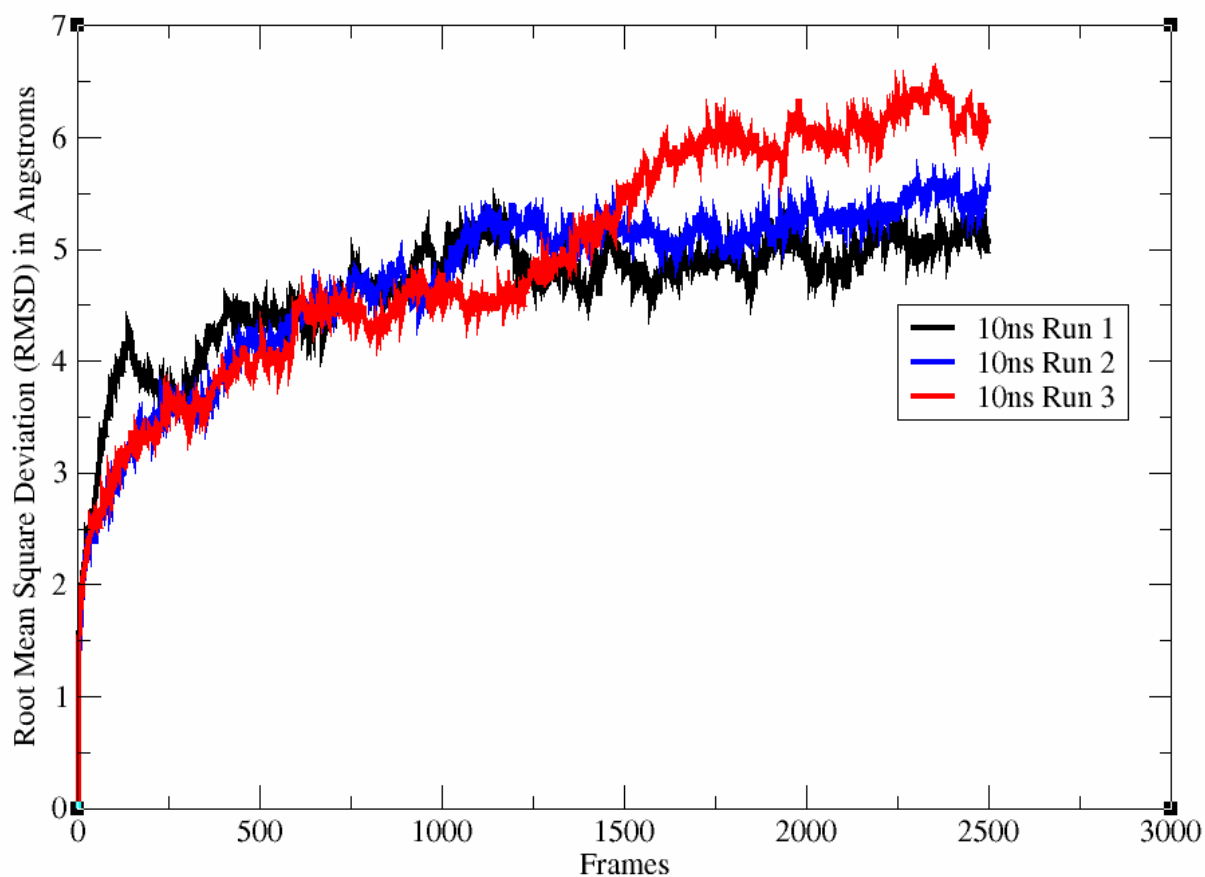
Figure 5.3.2 – RMSD plot for the 10 nanosecond runs or Mel1a. Each colour represents an different RMSD plot for the three simulations. The data impresses a need for longer simulations as the system appears to still be searching conformational space.

Bond Energy 10ns MD Run 1
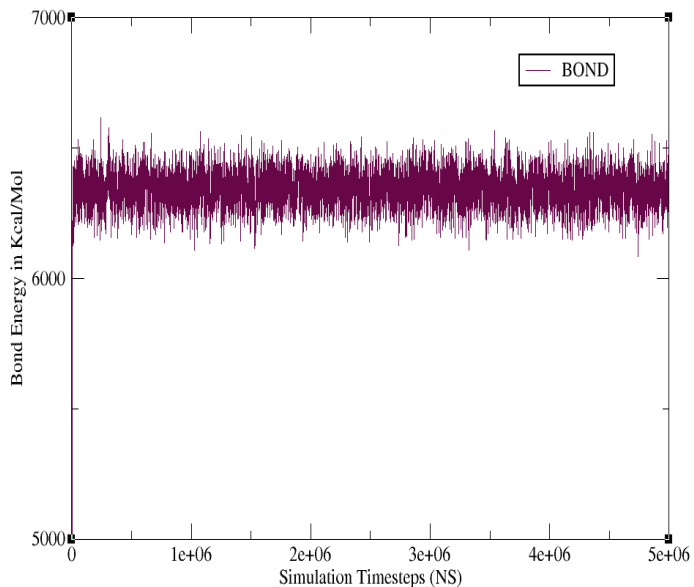
Temperature 10ns MD Run 1
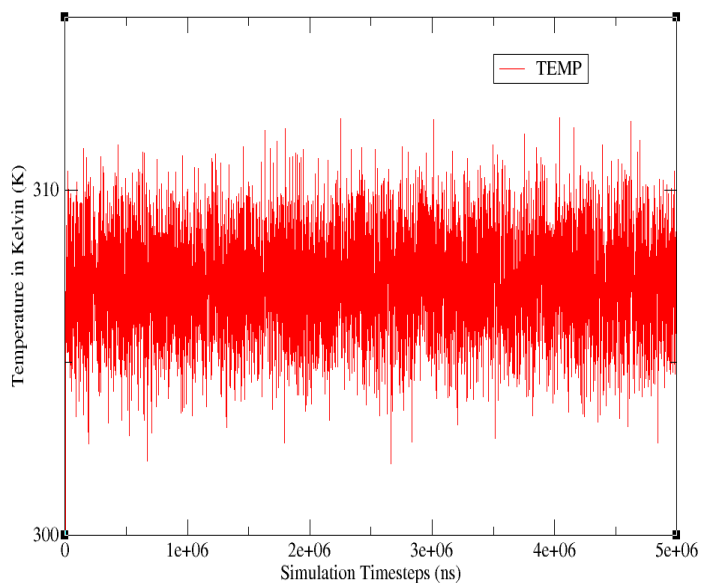
5.5.1 Bond energy (left) and temperature data for the Mel1a system (Run1).

Kinetic Energy 10ns MD Run 1
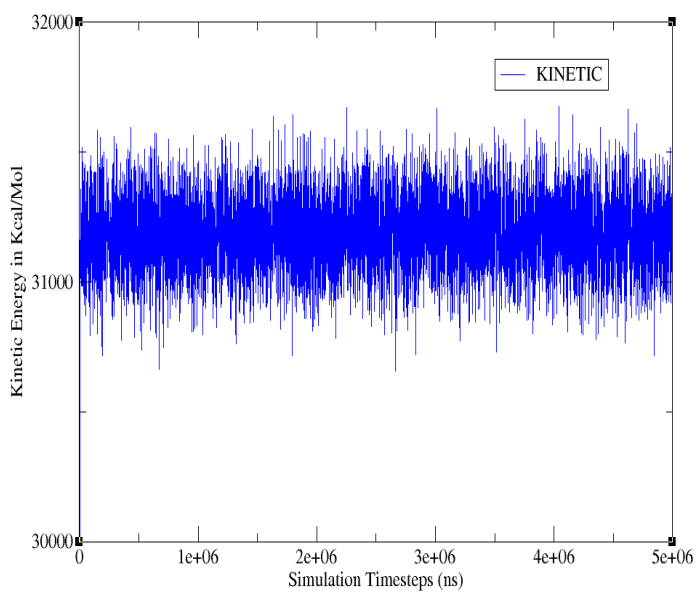
Potential Energy 10ns MD run 1
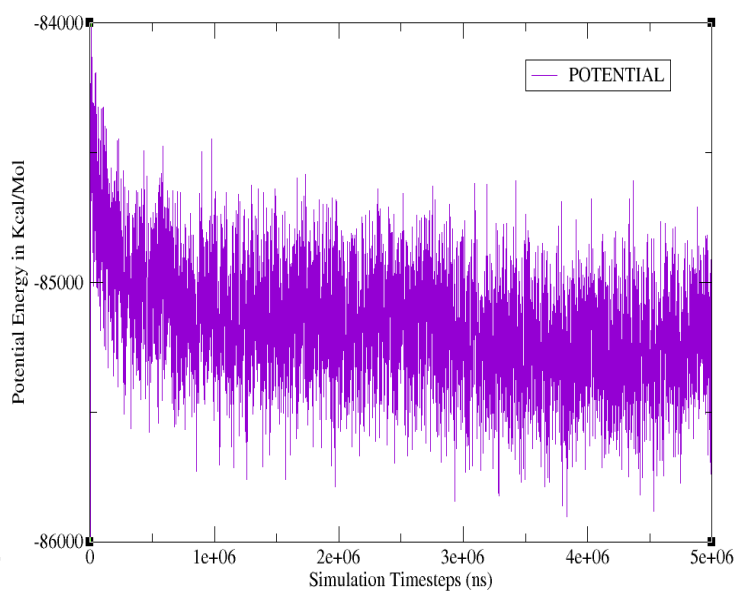
5.5.2 —Kinetic energy data (left) and Potential energy data (right) from the Mel1a system (run1).
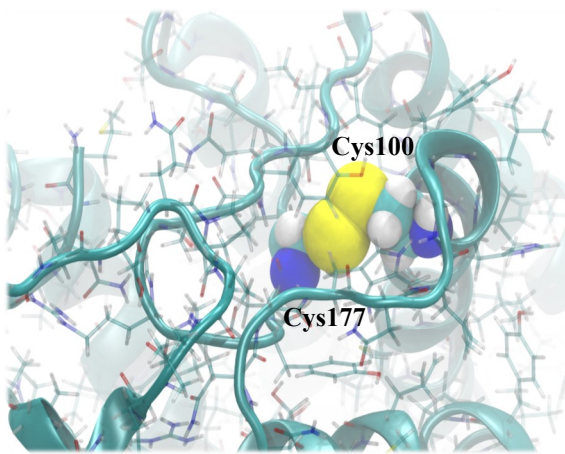
Disulphide bond between Cys100 and Cys177 on the extracellular side of Mel1a GCPR, in 10ns Trajectory 1.

Met339 and Cys314 exposed to solvent on the intracellular side of the membrane (above).

Met339 moves away from Helix A from an initial figure of 16.70Å to 24.21Å by calculation end (above).

The bend in Helix-turn-helix G in 10ns trajectory 1 in the final frame.
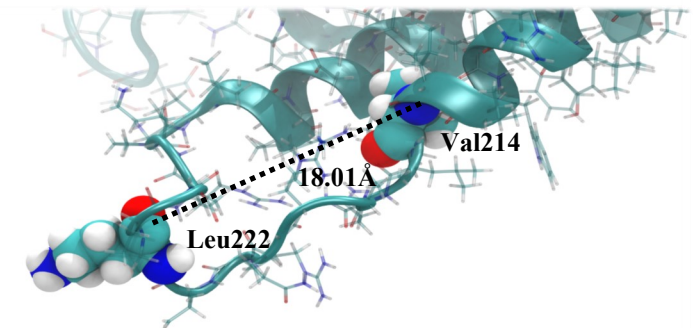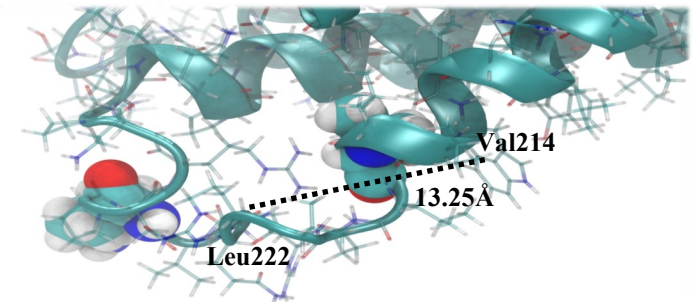
Leu222 moves from an initial distance of 13.25 to 18.01 by calculation end.

Figure 5.6.1—Visual analysis of Mel1a 10ns Trajectory 1.

MD 10ns Run 1 Mel1a and initial I-tasser Homology model of Mel1a GCPR.
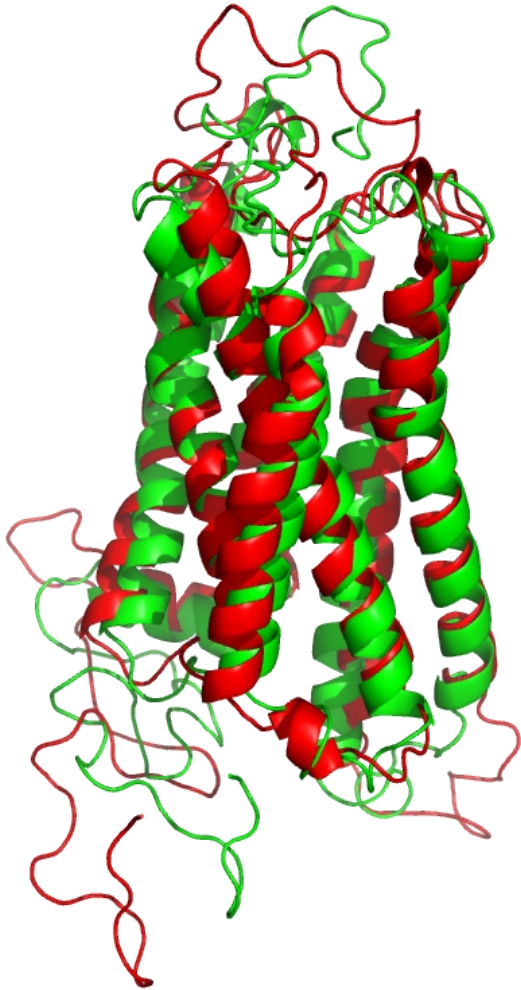
MD 10ns Run 1 Mel1a and X-ray β2 Adrenergic GCPR'.

MD 10ns Run 1 Mel1a and Rhodopsin GCPR.

Part A. The initial I-Tasser Mel1a GPCR Homology shown in green overlaid with the final frame of the 10ns Mel1a GPCR, run 1 shown in red. The deviation (RMSD) between the structures is 3.203Å

Part B. The β2 adrenergic GPCR shown in cyan structurally aligned the with 10ns Mel1a MD run 1, shown in red. deviation between the structures is 2.626Å
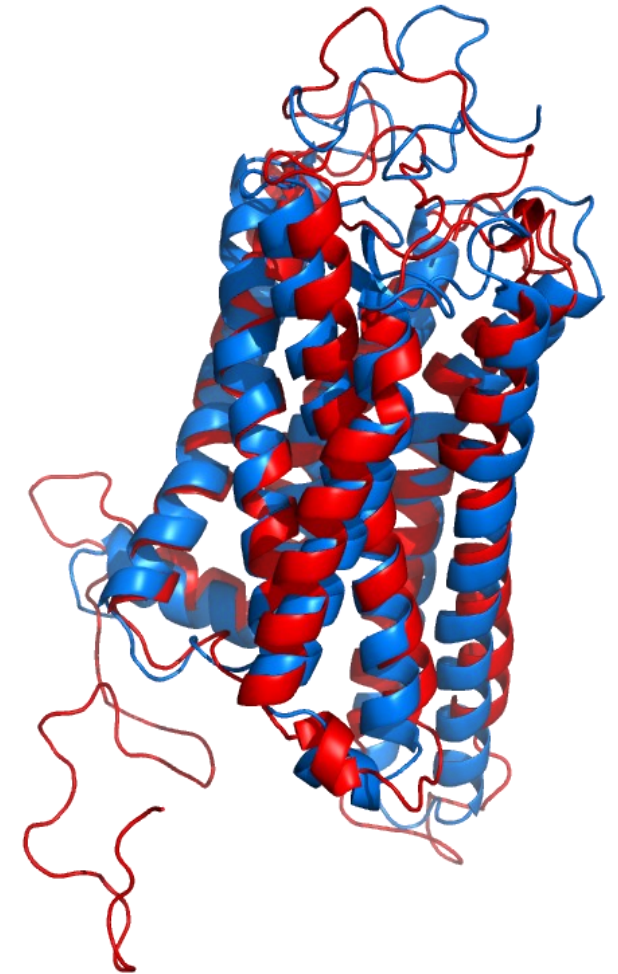
Part C. Structural alignment between the Rhodopsin GPCR and the final frame of the 10ns MD Mel1a GPCR, run 1. the structures have an RMSD of 3.517Å

Figure 5.7.1 — Mel1a 10ns MD Run 1 Structural Alignment with the Initial Homology model of Mel1a, 2-adrenergic and rhodopsin GPCR's.

Extracellular Side (and pocket).

Intracellular Side.
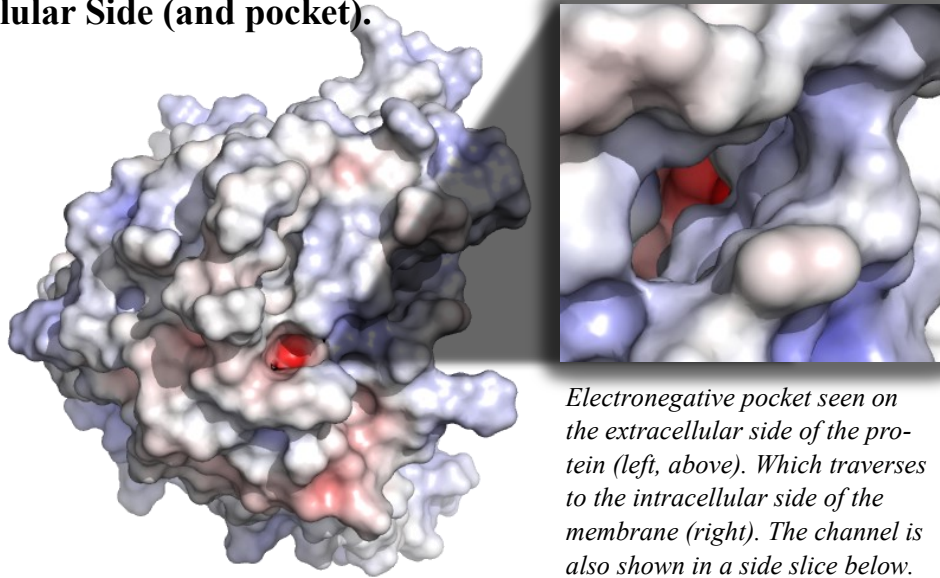
Electronegative pocket seen on the extracellular side of the protein (left, above). Which traverses to the intracellular side of the membrane (right). The channel is also shown in a side slice below.
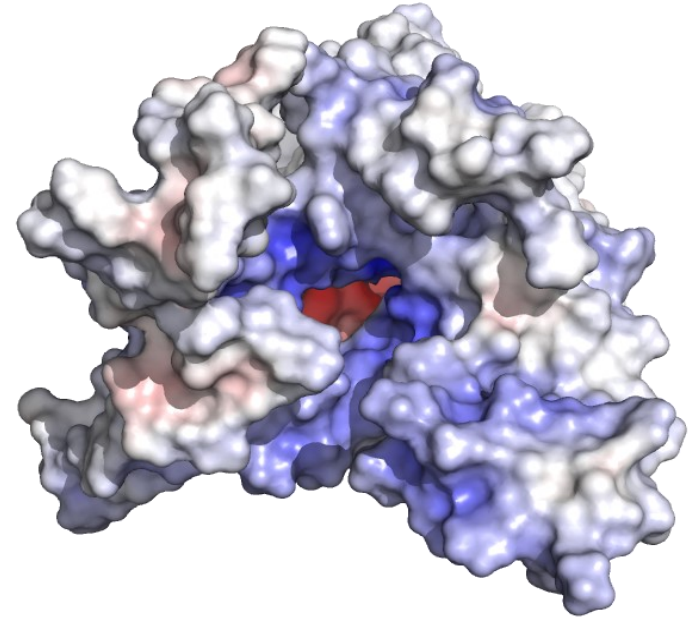
−10.000   10.000

−10.000   10.000

Electrostatic surface analysis was generated using Pymol and the APBS electrostatics plugin. Red areas are electronegative, blue are positive, white are zones with neutral charge. The final frame of the Mel1a GCPR 10ns Molecular dynamics simulation was utilized for analysis.
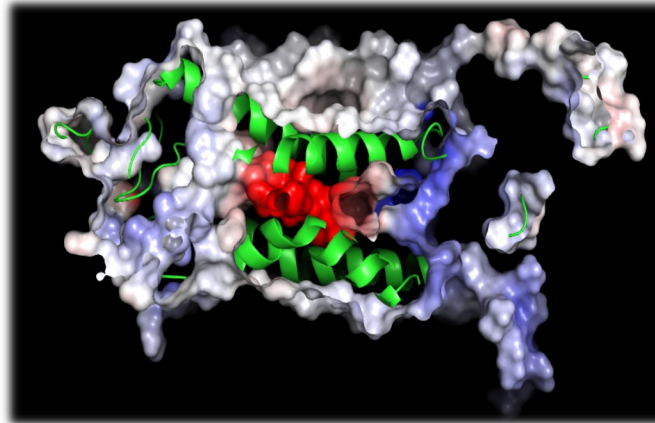
Side slice of the protein showing interior electronegative channel which runs from the external to internal sides of the membrane.

Side View (membrane portion).
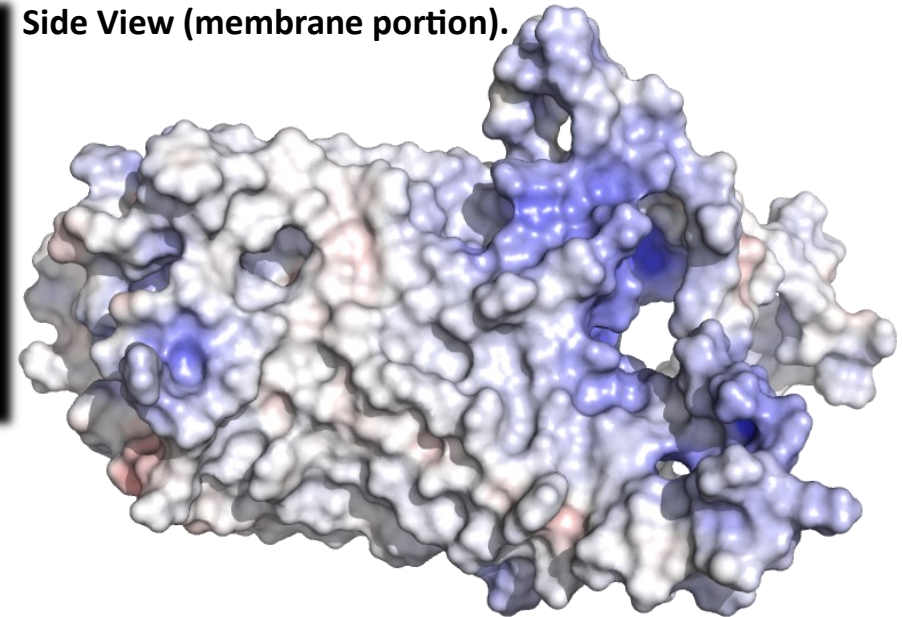
Figure 5.8.1 - Electrostatics of Mel1a GCPR final frame 10ns MD Run 1.
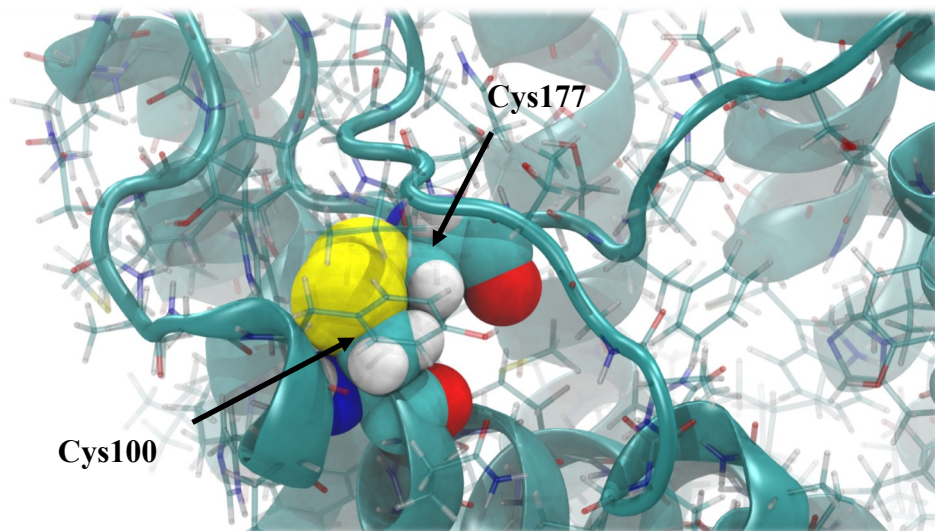
The Disulphide Bond between Cys100 on helix C and Cys177 on beta turn A on the extracellular side of the membrane.
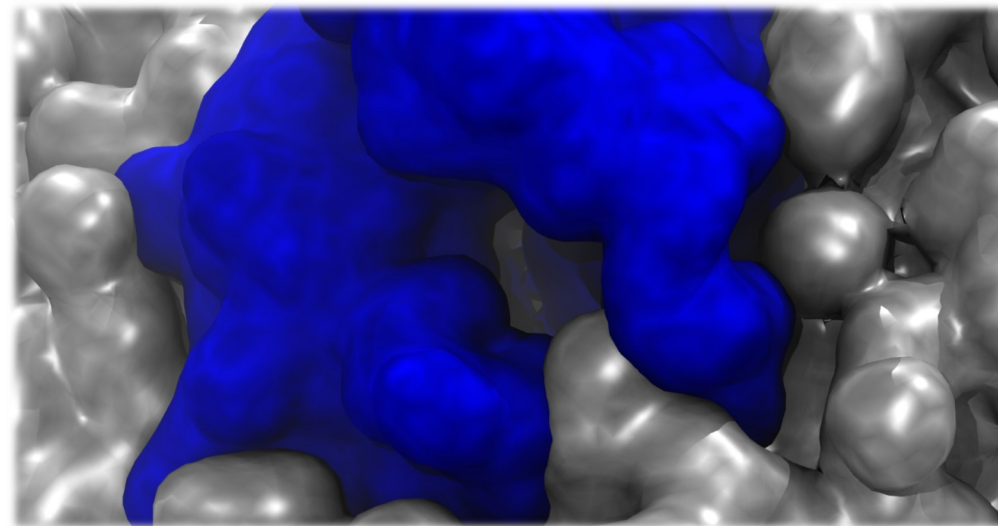
Extracellular Cavity of the Mel1a GPCR, on the external side of the membrane

Loop D takes a Helix-like Form (left).

Met1 (the C-terminus) on loop A and Met268 on Loop H are exposed to solvent. (right)
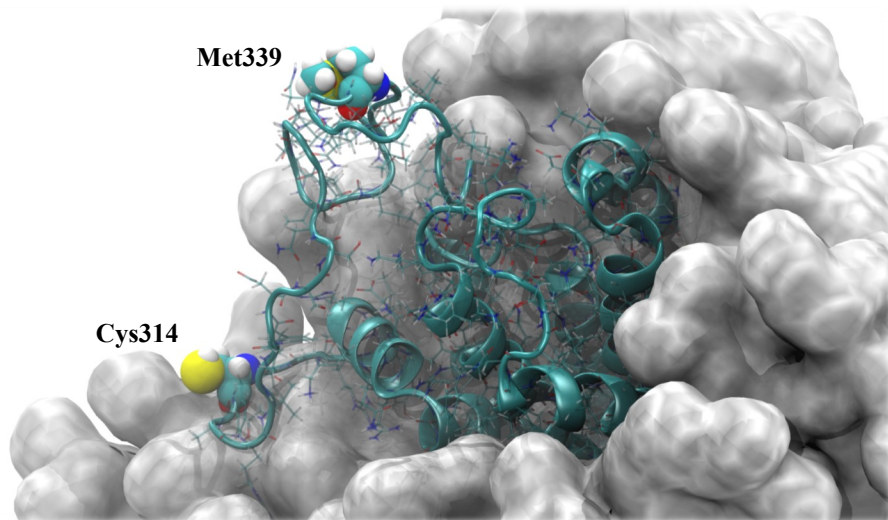
Met339 exposed to solute and Cys314 on the intracellular side moving away from the core of the protein along the membrane surface.
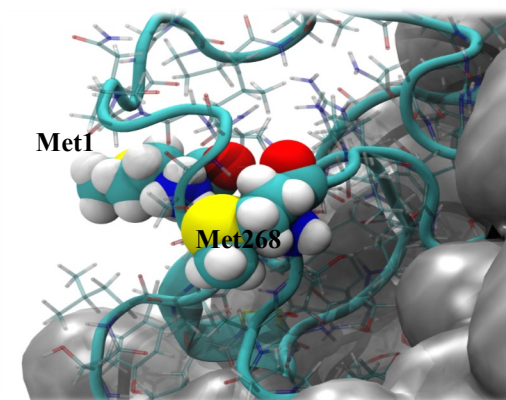
Figure 5.11.1—Visual analysis of 10ns Trajectory 2 of Mel1a

MD 10ns Run 2 Mel1a and initial I-tasser Homology model of Mel1a GCPR.
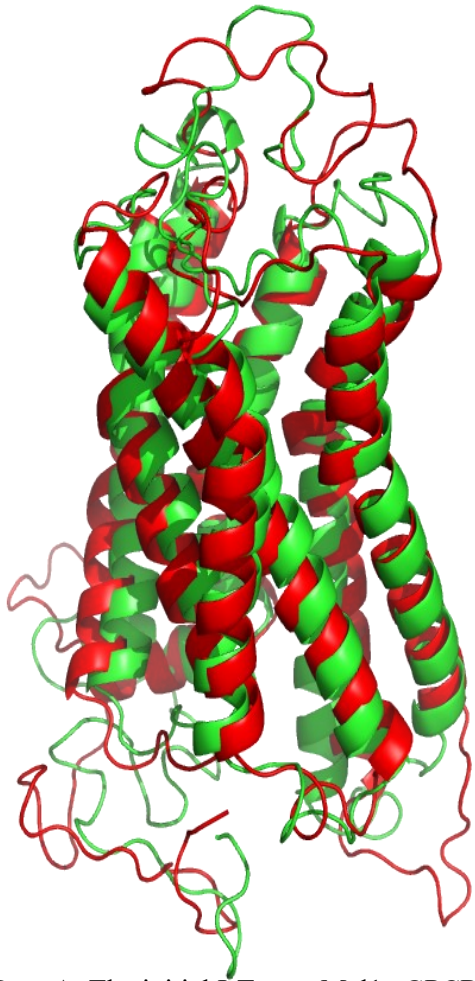
MD 10ns Run 2 Mel1a and X-ray β2 Adrenergic GCPR'.

MD 10ns Run 2 Mel1a and Rhodopsin GCPR.

Part A. The initial I-Tasser Mel1a GPCR Homology shown in green overlaid with the final frame of the 10ns Mel1a GPCR run 2, shown in red. The deviation (RMSD) between the structures is 3.197Å
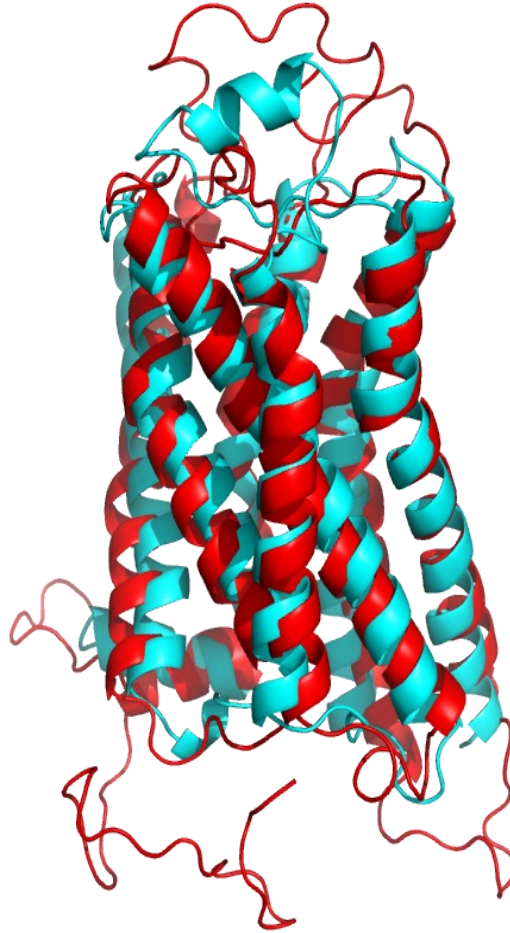
Part B. The β2 adrenergic GPCR shown in cyan structurally aligned the with 10ns Mel1a MD run 2, shown in red. deviation between the structures is 2.678Å
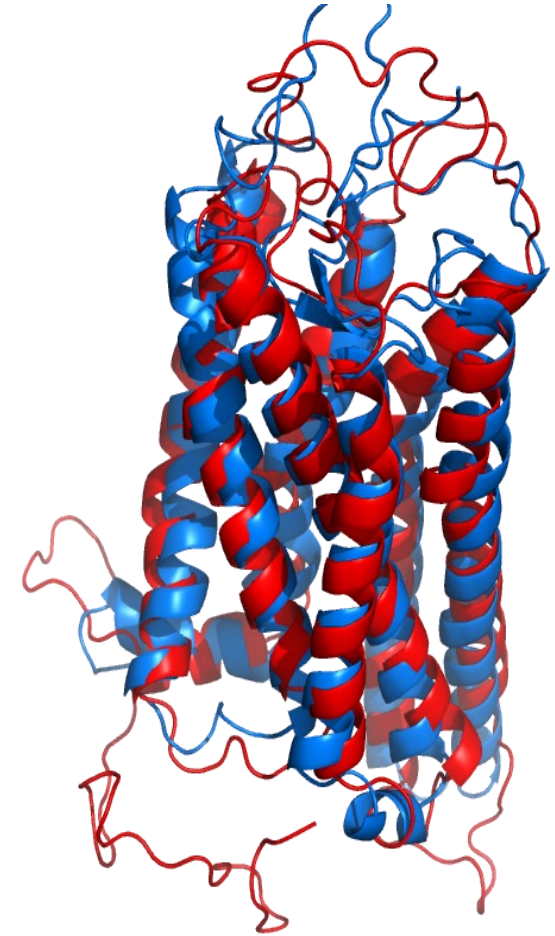
Part C. Structural alignment between the Rhodopsin GPCR and the final frame of the 10ns MD Mel1a GPCR, run 2. the structures have an RMSD of 3.375Å

Figure 5.12.1 - Mel1a 10ns MD Run 2 Structural Alignment with the Initial Homology model of Mel1a, β2-adrenergic and rhodopsin GPCR's.

Extracellular Side.
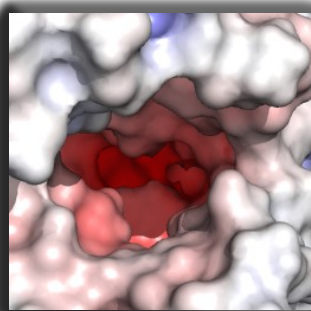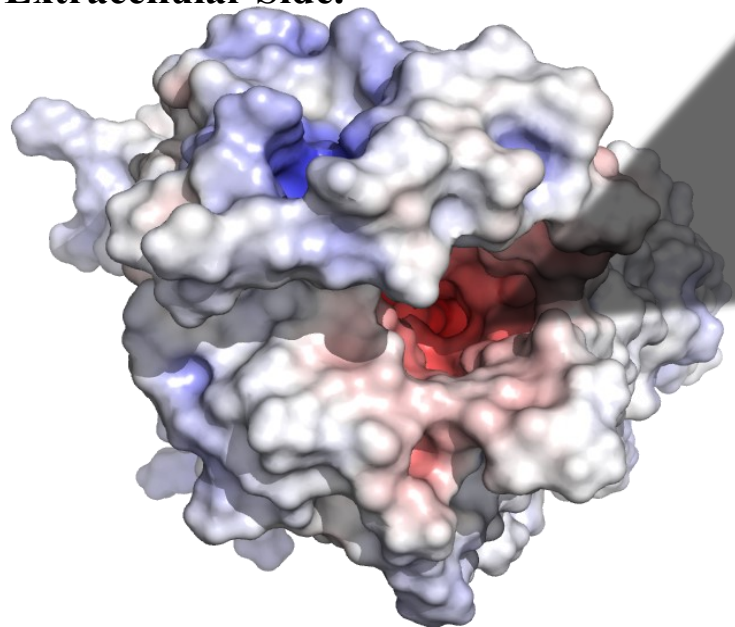
Cropped picture of extracellular cavity (left).

Intracellular Side.
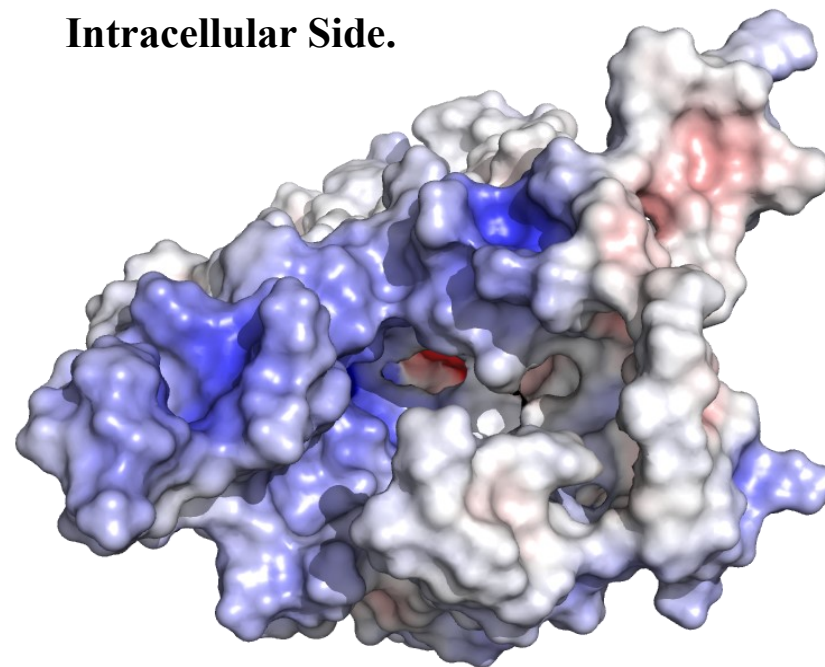
Electrostatic surface analysis was generated using Pymol and the APBS electrostatics plugin. Red areas are electronegative, blue are positive. White are zones with neutral charge. The final frame of the second Mel1a GCPR 10ns Molecular dynamics simulation was utilized for analysis after being separated from the bi-lipid membrane and solvent in the system.
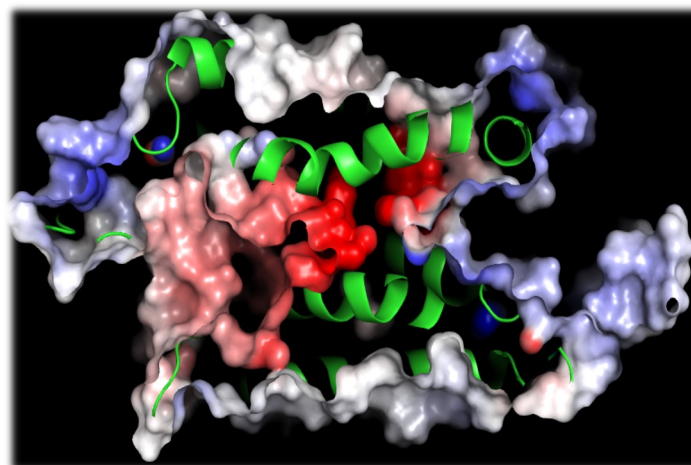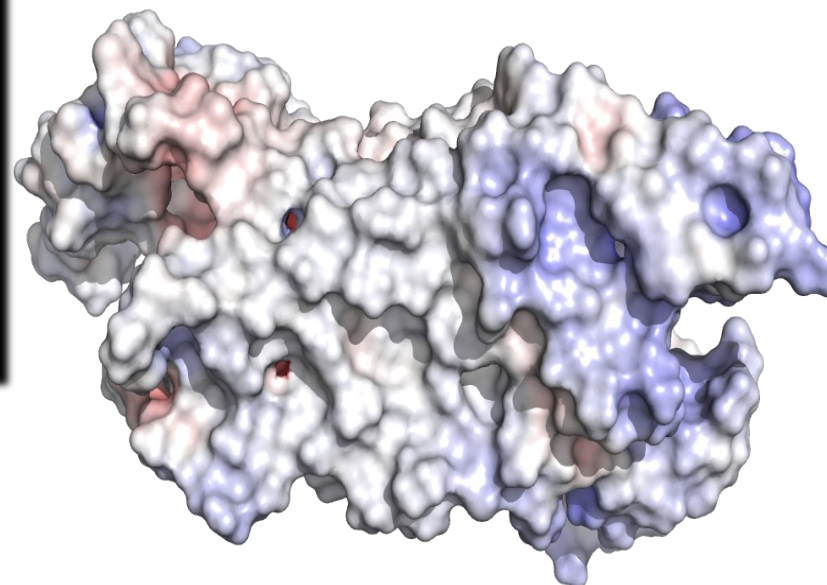
Unlike 10ns Run1, 10ns Run 2 has no channel which traverses from the intracellular to the extracellular, although electronegativity is still seen.

Side View (membrane portion)

Figure 5.13.1 – Electrostatics of Mel1a GCPR final frame 10ns MD Run 2.
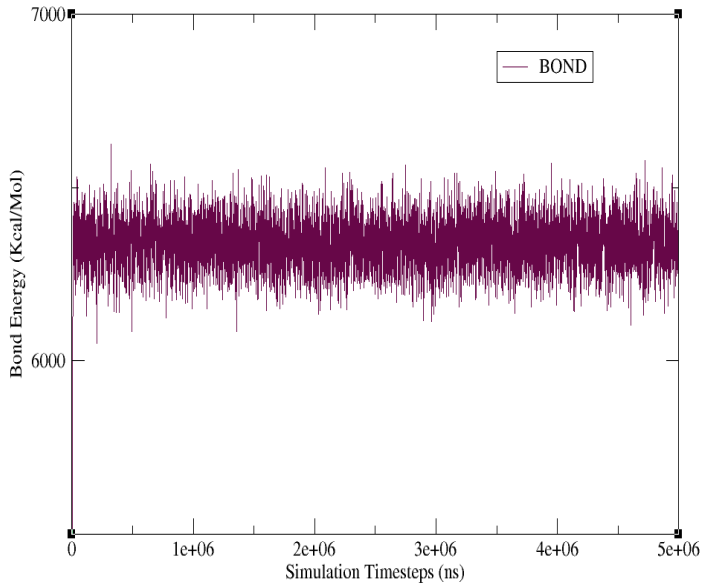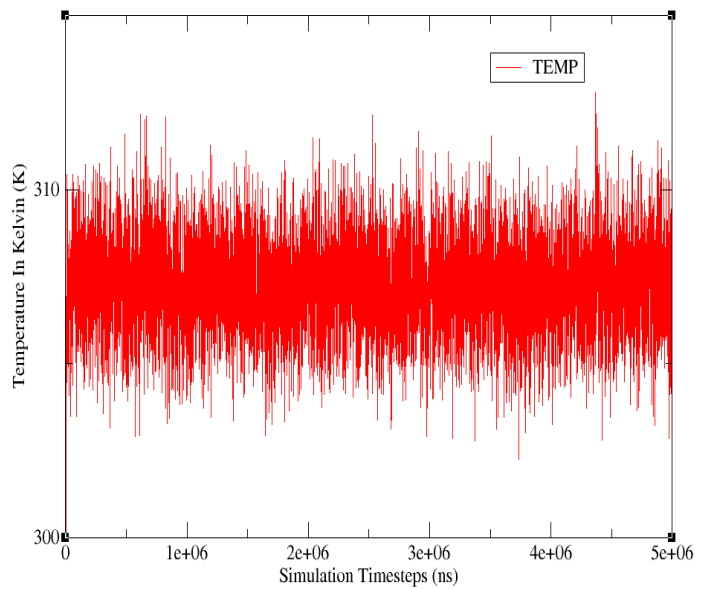
Figure 5.15.1—bond energy data in the Mel1a GCPR membrane system (left), temperature of the previously mentioned system in Kelvin (right).
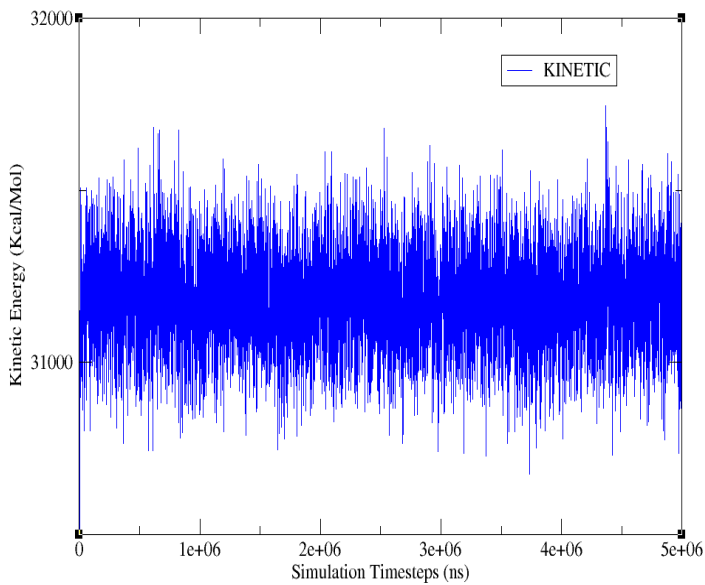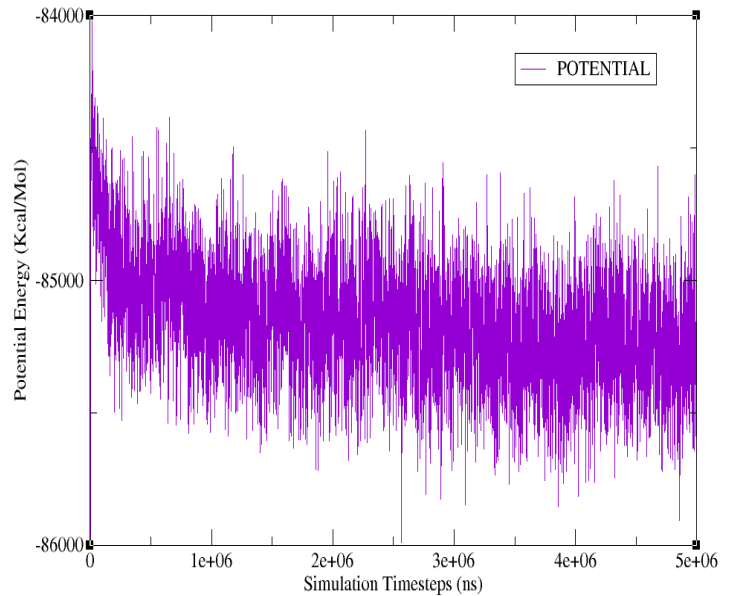
5.15.2 —Kinetic energy in the Mel1a GCPR system (left) negative potential energy in the MEl1a GCPR system (right).
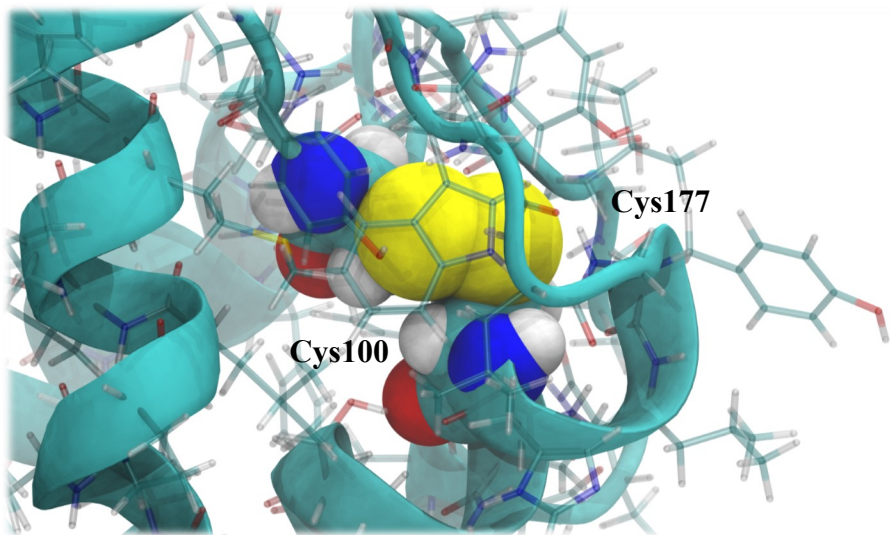
The Di-sulphide bond between Cys100 on Helix C and Cys177 on beta loop A.
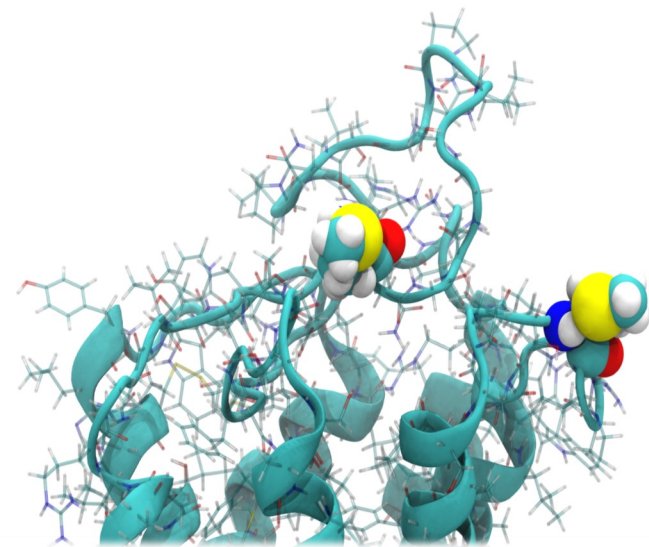
Met1 (C-terminus) on loop A and Met268 on the extracellular side of Mel1a, on loop H.— exposed to solvent.
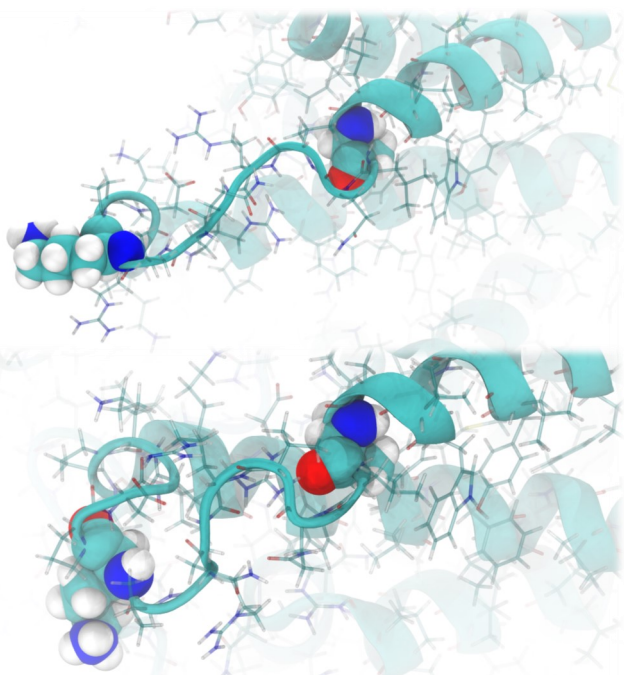
Leu222 on loop G extends towards solute from the core of the protein. (measured from Val214)
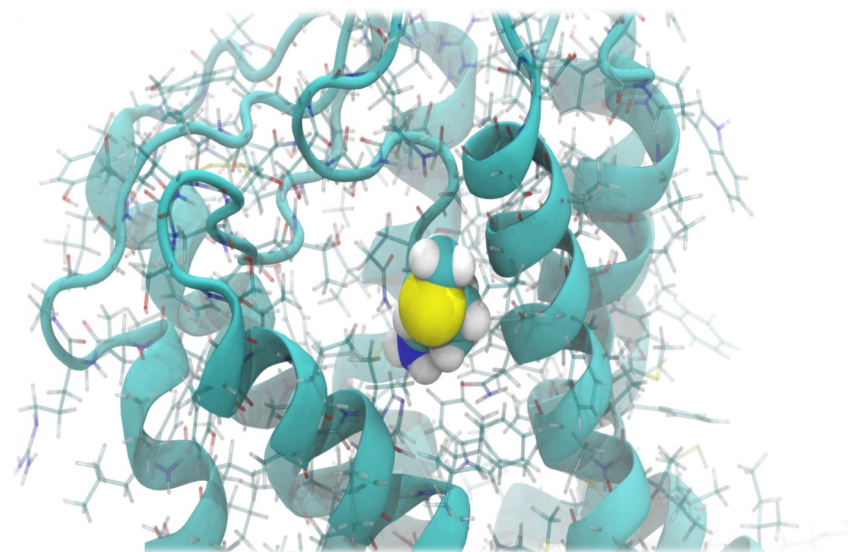
Met1 (C-terminus) on loop A compacted into the core of the Mel1a GCPR by the final frame.

Figure 5.16.1—Visual Analysis of 10ns trajectory 3 of the Mel1a GPCR.

MD 10ns Run 3 Mel1a and initial I-tasser Homology model of Mel1a GCPR.
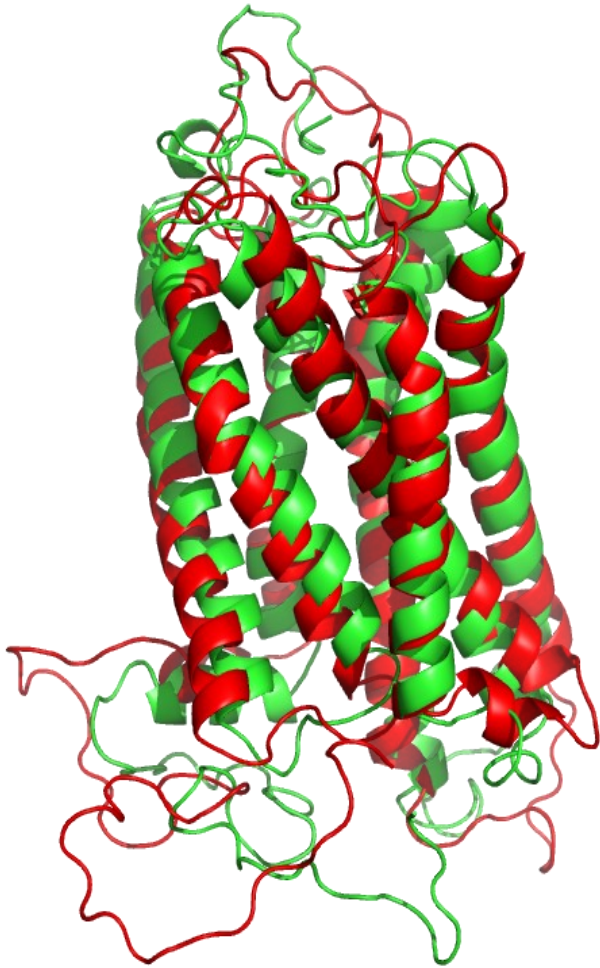
MD 10ns Run 3 Mel1a and X-ray β2 Adrenergic GCPR'.

MD 10ns Run 3 Mel1a and Rhodopsin GCPR.

Part A. The initial I-Tasser Mel1a GPCR Homology shown in green overlaid with the final frame of the 10ns Mel1a GPCR, run 3 shown in red. The deviation (RMSD) between the structures is 3.612Å
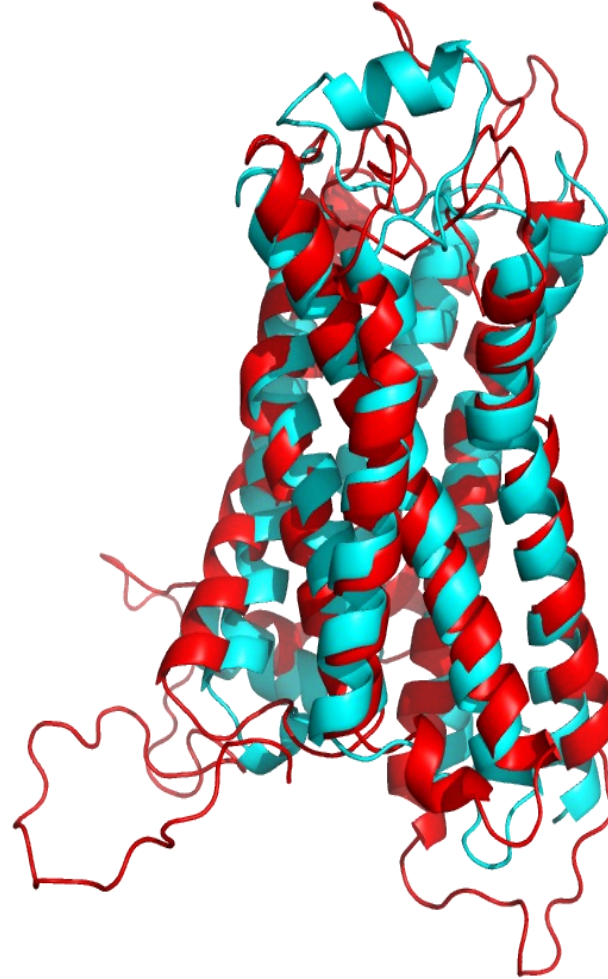
Part B. The β2 adrenergic GPCR shown in cyan structurally aligned the with 10ns Mel1a MD run 3, shown in red. deviation between the structures is 3.348Å

Part C. Structural alignment between the Rhodopsin GPCR and the final frame of the 10ns MD Mel1a GPCR, run 3. the structures have an RMSD of 4.449Å

Figure 5.17.1 - Mel1a 10ns MD Run 3 Structural Alignment with the Initial Homology model of Mel1a, 2-adrenergic and rhodopsin GPCR's.

Extracellular side
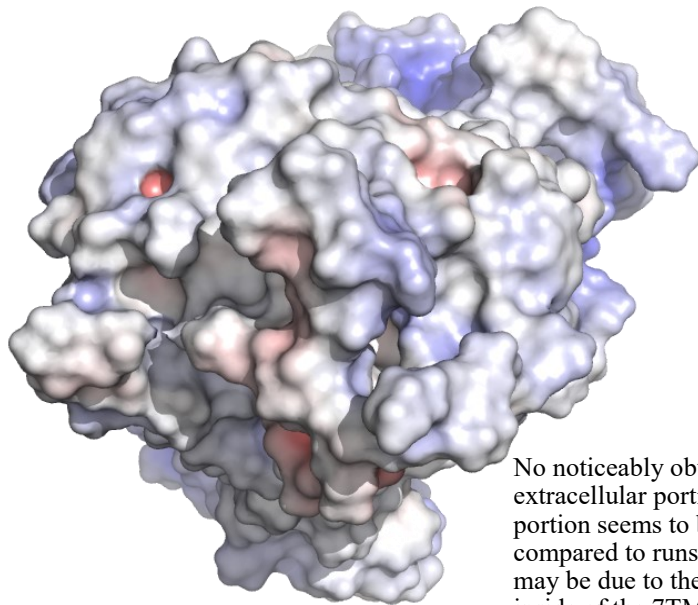
No noticeably obvious electronegative pocket is seen in the extracellular portion of the protein and most of the extracellular portion seems to be neutral due to conformational differences compared to runs 1 and 2 of the Mel1a system at 10ns. This may be due to the MET1 residue of loop A conforming to the inside of the 7TM core of the GCPR causing a different charge distribution on the extracellular side.

Electrostatic surface analysis was generated using Pymol and the APBS electrostatics plugin. Red areas are electronegative, blue are positive, white are zones with neutral charge. The final frame of the Mel1a GCPR 10ns Molecular dynamics simulation run 3 was utilized for analysis after removing it from the membrane system in VMD 1.9.3.

Intracellular Side
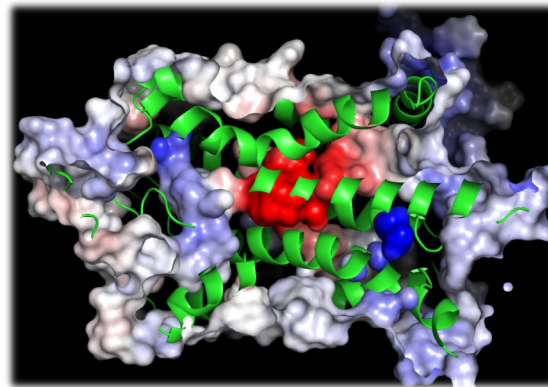
Side slice view of Mel1a, shows the electronegative channel which traverses through the central core of the protein. Also seen in 10ns Run 1.
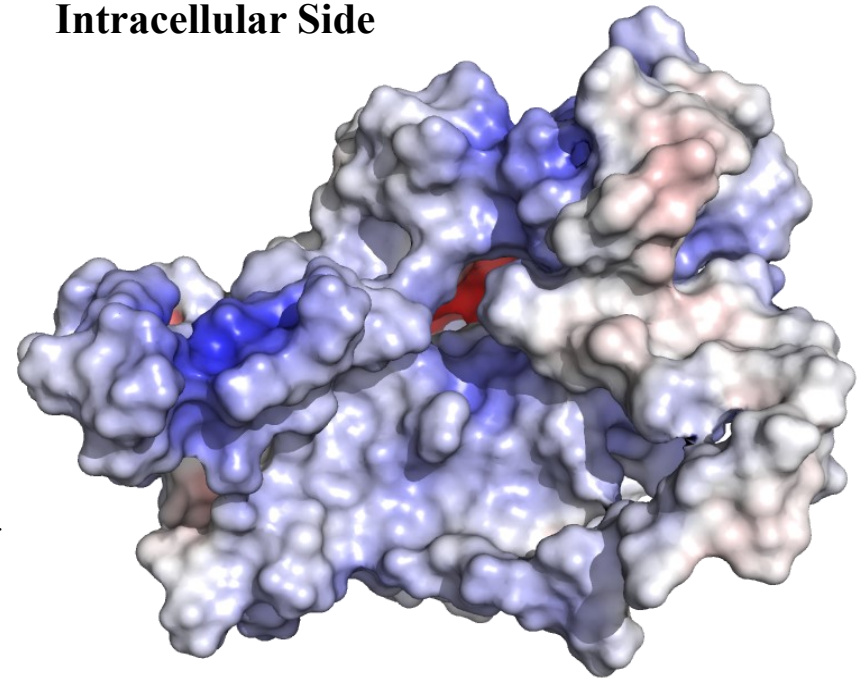
Side View (membrane portion)
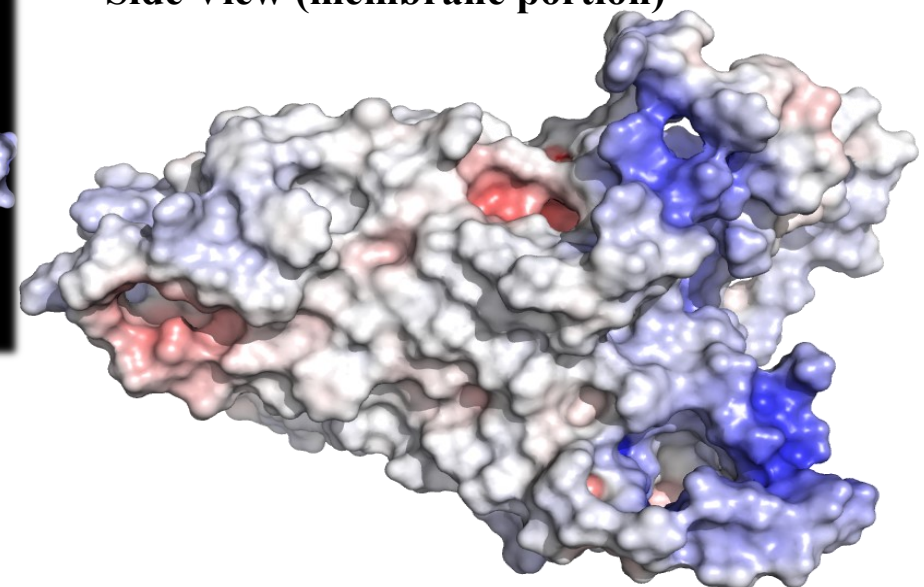
Figure 5.18.1 - Electrostatics of Mel1a GCPR final frame 10ns MD Run 3.

6.0 - 30 Nanosecond Molecular Dynamic Analysis of the Mel1a GPCR Bi-lipid Cell Membrane System.

Data from the triplicate 30 nanosecond trajectories was individually uploaded into VMD 1.9.3 for analysis. The trajectory data was coloured by RMSD for all three set of trajectory data. In addition, the log file was also analysed using the NamD plot plugin available in VMD to check the potential energy components of the systems throughout the simulations. Ramachandran plots were also generated for the final frame of each triplicate run to assess the secondary structure of the protein, and whether the residues were in typical conformations for the type of secondary structure. Structural alignment figures were made using Pymol comparing the final frame of the simulations of all 3 runs with the I-Tasser homology model of Mel1a, X-ray crystal structures of Rhodopsin and β2-adregenic GPCR's. Electrostatic surface analysis of each dataset was also performed using Pymol and the APBS plugin. These steps were followed in order to establish a detailed analysis of the protein as a representative model system.

Using the VMD RMSD trajectory tool the frames were aligned to the initial frame (frame 0) of the calculation, and the protein backbone was used to analyse the RMSD of the structure. 30ns trajectory 1 had an average RMSD of 5.609Å, a standard deviation of 0.892Å, a minimum RMSD of 0.832Å and a maximum RMSD of 7.012Å. 30ns trajectory 2 had an average RMSD of 5.412Å , a standard deviation of 0.833Å, a minimum RMSD of 0.820Å and a maximum RMSD of 6.913Å. 30ns trajectory 3 had an average RMSD of 6.137Å, a standard deviation of 1.408Å, a minimum RMSD of 0.839Å and a maximum RMSD of 8.393Å.

The timeline plugin available in VMD 1.9.3 was used to analyse the RMSD data over the duration of the 30ns calculation. The trajectory was loaded into VMD with a stride of 15 allowing for 1000 frames of the simulation to be loaded into the program. This process was performed for all three 30ns trajectories and made into a figure showing trajectory data coloured by RMSD (Figure 6.0.1). The high RMSD values shown in red are seen in the loop sections of the Mel1a GPCR (seen previously in the 10ns simulations figure 5.3.1). The overall RMSD of the protein is higher with certain areas reaching a value of 33.14Å in 30ns trajectory 1 which is primarily observed at the Met1 N-terminus residue which is loose in solute on the extracellular side of the protein. A maximum RMSD of 36.50Å in 30ns trajectory 2 is observed primarily on the intracellular side of the GPCR however this region of the protein appears to be the most conformationally stable in terms of RMSD according to the colour of the region in the figure. A maximum RMSD of 43.46Å is observed in the 30ns trajectory 3 which is again predominately on the intracellular side of the GPCR and represents the highest maximum

RMSD of the triplicate 30ns trajectories. Trajectory 3's Notable areas of RMSD are in the Helix F towards the intracellular side and towards the C-terminus of Loop I which has the highest RMSD of the entire structure according to the trajectory colourization. This data is expected since the time of the calculation has been increased to 30 ns (from the initial 10ns) which should allow the system to more fully explore conformational space. The helical transmembrane core of Mel1a, in all three trajectories, remains extremely stable with only small relative movement (RMSD) located at the ends of some of the 7TM regions.

Ramachandran plots were generated using the final frame of each MD trajectory (Figure 6.0.2) and show that the majority of the protein residues are found within expected secondary structure areas of the chart. A small portion of residues were observed outside the loop section of the Ramachandran's in all trajectories. This is likely due to the highly mobile character of the loop domains. Transition residues are also likely in the sections of the plot which do not conform to secondary structure types as they transition between one type of secondary structure to the next (Ie. alpha helix to loop section).

Analysis of the Mel1a 30 ns Trajectories.

The log files from the three 30 ns Mel1a trajectory datasets were analysed using the NamdPlot module available in Vmd 1.9.3 and representative charts of the data were generated. The resultant figures from all three trajectories are shown in the table below (Table 6.0.1). Data can also be viewed in graph form in figures 6.1.1, 6.1.2 and 6.1.3 for 30ns for run 1, 6.5.1, 6.5.2 and 6.5.3 for 30 ns run 2 and 6.10.1, 6.10.2 and 6.10.3 for 30 ns run 3;

Table 6.1.1	Initial.	Average.	Final.
Kinetic energy (30ns run 1).	28,433.59 Kcal/mol	31,331.96 Kcal/mol	31,355.04 Kcal/mol
Kinetic energy (30ns run 2).	28,433.59 Kcal/mol	31,347.03 Kcal/mol	31,515.34 Kcal/mol
Kinetic energy (30ns run 3).	28,433.59 Kcal/mol	31,349.74 Kcal/mol	31,482.67 Kcal/mol
Potential energy (30ns Run 1).	-87,607.43 Kcal/mol	-86,286.507 Kcal/mol	-86,232.59 Kcal/mol
Potential energy (30ns Run 2).	-87,607.42 Kcal/mol	-86,197.92 Kcal/mol	-86,208.45 Kcal/mol
Potential energy (30ns Run 3).	-87,607.42 Kcal/mol	-86,213.9 Kcal/mol	-86,349.28 Kcal/mol

Bond energy (30ns Run 1).	2,553.98 Kcal/mol	5,497.81 Kcal/mol	5,484.58 Kcal/mol
Bond energy (30ns Run 2).	2,553.98 Kcal/mol	5,496.59 Kcal/mol	5,530.10 Kcal/mol
Bond energy (30ns Run 3).	2,553.98 Kcal/mol	5,496.75 Kcal/mol	5,519.9 Kcal/mol
Temperature in Kelvin (30ns Run 1)	280.16 K	308.71 K	308.94 K
Temperature in Kelvin (30ns Run 2)	280.16 K	308.87 K	310.53 K
Temperature in Kelvin (30ns Run 3)	280.16 K	308.89 K	310.2 K
Van der Waal interaction energy (30ns run 1)	3,309.59 Kcal/mol	2,726.01 Kcal/mol	2,783 Kcal/mol
Van der Waal interaction energy (30ns run 2)	3,309.59 Kcal/mol	2,710.42 Kcal/mol	2,631.01 Kcal/mol
Van der Waal interaction energy (30ns run 3)	3,309.59 Kcal/mol	2,714.37 Kcal/mol	2,591.42 Kcal/mol

Table 6.0.1 – A summary of the data collected from all 30 ns molecular dynamics runs.

6.2 Visual Analysis of the Mel1a Protein Structure in 30ns Trajectory 1.

The DCD file for the MD calculation was loaded into VMD 1.9.3 with a stride of 1 and therefore incorporated all 15,000 frames of the simulation (Figure 6.2.1). In a similar fashion to the 10ns simulations of Mel1a, the disulphide bond between CYS100 on Helix C and CYS177 on beta turn A is present. Loop D behaves as a loop throughout the simulation, unlike 10ns run 2 (section 5.11). CYS314 shows similar behaviour to all the previous 10ns MD runs traversing away from the core of the protein in Loop I whilst staying close to the POPC membrane. The sulphur atom is oriented towards the lipid bi-layer to avoid contact with solvent (water and ions). MET339 remains exposed to solute in the same fashion as the previous runs. Met 1 initially sits close to the centre of the core, but as the simulation continues it breaks away from the core and reconfirms above Helix A by the end of the calculation.

6.3 Structural Alignment Analysis of Mel1a from 30ns trajectory 1.

The final frame of trajectory 1 was saved in VMD (as a pdb file) with the solvent, ions and lipid bilayer removed. It was then loaded into Pymol where structural alignment with the

original I-Tasser model, Beta 2 Adrenergic and Rhodopsin was carried out (Figure 6.3.1) . For the I-Tasser model, part A, The RMSD between the structures is 3.813Å. Most of the deviation seen between the aligned structures is observed in the loop sections of the GPCR's. There are differences in the orientation of some of the helix sections of the protein. Most notably in the Helix-turn-helix G where both the final helix section before loop I is facing away from the protein at a different angle and further into the core at the upper half (before loop H) of the helix in comparison to the homology model in part A. All the other alpha helix sections have slightly different angular orientations but follow a similar path. Alpha helix F sits further from the core of the Mel1a GPCR, Helix A is more tilted towards Helix-Turn-Helix G. which via molecular dynamics simulation has conformationally changed. In Figure 6.3.1, Part B the structural alignment with the β2 adrenergic GPCR X-ray crystal structure is shown. The RMSD between the structures is 3.119Å. This is slightly less than the difference between the alignment with the homology model (Part A) and may be because the β2 adrenergic receptor is missing intracellular loop sections. The helix-turn-helix G domain tilts inwards on the upper extracellular side, and as a result the intracellular side tilts away from the core of Mel1a. As seen in all structural alignments part B has high RMSD in the loop domains of both GPCR's. There is also no structural similarity between the upper short extracellular helix observed on the β2 adrenergic receptor and Mel1a's extracellular loops where no helix exists at this location. Part C in figure 6.3.1 shows the structural alignment of trajectory 1 final frame with the Rhodopsin GPCR. The RMSD between the protein structures is 3.234Å. The loop domains of the aligned GPCR's are again different in conformation but are more similar than that of the structurally alignment observed with the β2 adrenergic receptor (Part B, figure 6.3.1). Helix-Turn-Helix G is still pointed away from the protein core in comparison to Rhodopsin on the 'Loop I' end, and more buried in the core at the 'Loop H' end. This shows that the Mel1 GPCR is conformationally different from the β2 adrenergic and rhodopsin GPCRs' but shares the family resemblance as a GPCR. The stability in the core also shows the Mel1a GPCR under simulation conditions does not deviate significantly inside the bi-lipid membrane.

6.4 Electrostatic Surface analysis of Mel1a - 30ns Run 1.

The final frame of 30ns MD on run 1 was saved as a new PDB file through VMD 1.9.3 and input into Pymol. The .PDB was converted to a PQR file using the Pymol plugin for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface

using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM. Using (Figure 6.4.1), an electrostatic surface map was created with a charge distribution map +/- 10 . Red areas of the surface map show electronegative regions of the protein while blue indicates electropositive regions. White colouring represents neutral areas in the protein. The extracellular side of the protein has no obvious electronegative cavity in comparison to the Mel1a GPCR 10 ns trajectory 1 (Figure 5.8.1), 2 (Figure 5.13.1), and the I-tasser homology model (Figure 4.3.1). However, there are electronegative patches observed on the extracellular surface. The intracellular side of Mel1a has a large central electronegative cavity which is not directly connected to the extracellular side as was observed in the earlier 10ns trajectories (Figures 5.8.1 and 5.13.1). Of note, this electronegative cavity is surrounded by an electropositive "shroud" provided by the loop sections of the protein. A membrane positioned side view of Mel1a shows that it is electronically neutral (white).

6.6 Visual Analysis of the Mel1a Protein Structure in 30ns Run 2.

The DCD file was loaded into VMD 1.93 with a stride of 1 incorporating all 15,000 frames for visual analysis (Figure 6.6.1 and 6.6.2). The CYS100-CYS177 di-sulphide bond persists in 30ns trajectory 2 as with all previous runs. MET 1 is initially buried towards the core of the GPCR however, as the simulation progresses it moves out into solute before returning to a similar conformation. MET268 moves from the outside of the protein to the inside core of the protein as MET1 and returns to its initial conformation. As the simulation progresses and between THR37 and VAL40 on alpha helix A the structure changes conformation, producing a short loop section between the alpha helix A domain. Similar phenomena are seen in the VAL84 and MET86 portion of alpha helix B before it transitions into Loop C where the residues are almost uncoiled but hold a pi helix-like conformation from Pro80 and after the turn in Helix B ending at Phe89. Between LEU133 and SER140 on loop D the beta loop has changed form to a helix-like conformation before transitioning into alpha Helix D (Unlike the previous trajectory in Section 6.2). Conformational change is seen on loop E between GLN219 and LYS226 before transitioning into alpha helix F where it has an almost helical form. Helix-Turn-Helix collapses on the Z-axis - This is observed at Asn299 where the helix bends in another direction and continues to residue ILE309. The conformation of Helix-turn-helix G collapses on the Z-axis of the display from 42.93Å to 38.48Å (distance measured from residue GLU274:HG2 at the beginning of the helix-turn-helix measured in relative distance from

ILE309:O) causing significant conformational deviation as the trajectory progresses to the final frame. Similarly, as observed in all trajectories, the intracellular residue on loop I - CYS314 is moving away from the core of the protein and the sulphur atom is oriented towards the POPC membrane away from solvent. Met 339 is similarly still exposed to solvent on the intracellular side. This seems to 'spread' the entirely to loop I across the POPC membrane away from the core of the protein.

6.7 Structural Alignment Analysis of Mel1a from 30ns Run 2.

The final frame of the Mel1a bilipid cell membrane system (30ns trajectory 2) was saved in VMD as a pdb file with the solvent, ions and lipid bilayer removed (Figure 6.7.1). It was then loaded into Pymol and structurally aligned with the initial I-Tasser homology model of Mel1a (Orange) shown in Part A. The RMSD between the aligned structures is 4.194Å. In addition to the previously seen variation in the loop sections, large parts of the alpha helix core deviate in conformation. Most notably Helix A which has a central loop section and Helix-turn-helix G which has a sharper 'Z' shape conformation reducing its size on the Z-axis. Most of the helix does not structurally align as it has done in the previous 10ns calculations nor the homology structural alignments. (Section 4.1 and section 4.7). Part B in figure 6.7.1 shows the structural alignment of the β2 adrenergic GPCR (shown in Cyan) with the final frame of 30ns run 2 (shown in green). The RMSD between the structures is 3.563Å, which is comparatively lower than the RMSD between the initial homology model and 30ns Run 2's final frame. But as noted previously loop sections are missing from the crystal structure for the β2 adrenergic GPCR. Most of the helix sections follow a similar pattern except for alpha helix A which bends significantly at the centre loop between residues THR37 and VAL40 towards helix-turn-helix G. The helix-turn-helix G sections of the core which vary considerably compared to previous structural alignments the Asn299 to Ile309 section sits above the similar helix shape seen in the β2 adrenergic GPCR. Loop D has a helical representation in Pymol which is not shared by the β2-adrenergic GPCR. Part C in figure 6.7.1 shows the structural alignment with the rhodopsin GPCR X-ray crystal structure. RMSD between the aligned structures is 3.885Å. The structure of rhodopsin seems to be more like the final product of the 30ns simulation. Alpha helixes E and F bend inwards towards the core of Mel1a on the extracellular side more so than rhodopsin's structure. Helix A is very different as in part A and part B with a loop section splitting it into two separate helixes. The extracellular side of Helix B also shows a large degree of conformational difference as the upper half bends in towards the core.

6.8 Electrostatic Surface analysis of Mel1a - 30ns Run 2.

The final frame of 30ns MD on run 2 was saved as a new PDB file through VMD 1.9.3 and input into Pymol. The .PDB was converted to a PQR file using the Pymol plugin for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM (Figure 6.8.1) an electrostatic surface map was created with a charge distribution map +/- 10. The extracellular side of the Mel1a GPCR has a shallow electronegative pocket visible. The intracellular side shows a large electropositive surface and an electronegative cavity which is covered by loop sections this cavity traverses to the central core of the protein structure but does not connect with the intracellular side similarly to both figures 4.7.11 and the initial homology model in figure 4.1.5, which have a central channel which is not connected from the intracellular side to the extracellular side. The side view shows a mostly neutral (white) extracellular profile. The intracellular is more electropositive. This is also seen in the side slice of Mel1a which shows the disconnected channel.

6.11 Visual Analysis of the Mel1a Protein Structure in 30ns Trajectory 3.

The DCD file for the third 30ns run of Mel1a was loaded into VMD 1.9.3 with a stride of one incorporating all 15,000 frames for visual analysis (Figures 6.11.1 and 6.11.2). On the extracellular portion of the Mel1a Met1 and MET268 are closely positioned to the core of the protein, ASN10 on loop A is interacting with the solvent and extending loop A from MET1's position. The disulphide bond between CYS100 on alpha helix C and CYS177 is present as in all the other molecular dynamics simulations performed. Helix B appears to maintain the pi-helix conformation, with the Pro80 to Asn90 section on the extracellular side moving out away from the core of the Mel1a GPCR. Loop G on the intracellular side moves outward from the core of the protein (VAL221 on loop G in distance from LEU215 on Helix F which is initially 13.78Å extending to 17.75Å by the final frame of the simulation. CYS314 clings to the POPC membrane moving away from the core of the protein, MET339 also is as seen similarly in all MD runs (10ns to 30ns) exposed to solute but appears to relocate closer to the POPC membrane as the simulation progresses. This leaves loop I spread across a large area of the intracellular side of the membrane.

The final frame of the Mel1a bi-lipid cell membrane system (30ns Run 31) was saved in VMD 1.9.3 with the solvent and lipid bi-layer removed (Figure 6.12.1). It was then loaded into Pymol and structurally aligned with the initial I-Tasser homology model of Mel1a (Orange) shown in Part A. The RMSD between the structures is 4.031Å. Most of the alpha helix 7tm regions follow the same profile as the homology model. The exceptions being helix-turn-helix G which the turn helix (ASN299 to ILE309) is pointed more towards the core of the structure. The extracellular loop sections follow a similar shape to the homology model in comparison to other structural alignments. The initial start of alpha helix F sits further away from the core of the protein. compared to the homology model the intracellular side is pushed further out from the core. Loop G which transitions between sits further out into the intracellular area as a result. The intracellular loops deviate largely from the homology model loop which is expected due to being exposed to solvent. Part B shows the structural alignment with the β2 adrenergic GPCR (shown in cyan) and the final frame of the 30ns Mel1a model (shown in green), the deviation (RMSD) between the structures is 3.671Å. the loop sections are greatly varied with the extracellular portion showing a completely different conformation, also missing the helix-like coil in present the b2 adrenergic GPCR. Helix F sits lower in the Mel1a Structure than the latter, resulting in the helix being closer to the intracellular side of the protein. Loop G as a result is sitting much further out into the solute region. Helix-turn-helix G sits further in towards the core of the protein, while the end of helix A is pushed further away from the core of the protein out into what would be the POPC membrane. The lower RMSD figure is likely representative of the missing loop sections from the β2 adrenergic receptor due to being fused to a T4 lysosome in order to crystallise. Part C shows the structural alignment of the Rhodopsin GPCR X-ray crystal structure (shown in magenta), aligned with the final frame of 30ns run 3 of Mel1a GPCR (shown in green). The deviation between the structures (RMSD) is 3.706Å. High deviation in the loop areas is seen on both the intra and extracellular sides of the protein. Helix E and F sit higher than the comparative helixes in the rhodopsin structure. The lower half of helix F near the extracellular side bends towards Helix E. The extracellular upper half of Helix-turn-helix G sits further away from the core of the protein and the intracellular side on the final turn of the helix (ASN299 to ILE309) orients towards Helix A.

The final frame of 30ns MD on run 2 was saved as a new PDB file in with the solvent and POPC membrane removed through VMD 1.9.3 and uploaded into Pymol. The .PDB was

converted to a PQR file using the Pymol plugin for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM (Figure 6.12.1). A distribution map of +/- 10 charge was created. The extracellular side of Mel1a is a mix of neutral positive and negative zones with a small electronegative channel that follows through to the other side of the membrane. The core is similarly seen in figures 4.7.4 and 4.7.15 for the 10ns calculations. The intracellular portion is electropositive in the loop sections with a partially neutral set of zones. The side profile of Mel1a is predominantly neutral. The intracellular side visible in the side profile is electropositive. The extracellular side visible in the side profile has varied areas of positive negative and neutral charge.

Summary.

The data acquired from the 30 ns trajectory runs shows increased RMSD when backwards compared to the 10 ns trajectory data (section 5.0). Each of the trajectory datasets retains the GPCR architecture but conformationally changes from the initial I-Tasser homology model. The core of the Mel1a GPCR remains extremely stable (an RMSD below 3Å measured from the residues of the alpha helix sections) in all 3 trajectories with most of the activity appearing in the intracellular and extracellular loop sections. The mel1a GPCR has distinct differences from the rhodopsin and β2 adrenergic receptor. The electrostatic data received from each 30ns trajectory dataset also suggests that the extracellular side of the protein is negative, while the intracellular side is predominantly electropositive. The disulphide bond seen on the extracellular side of mel1a (Cys100-Cys177) appears in all 10ns and 30ns trajectories to this point suggesting it is a region with a definable biological purpose (Ferguson *et al.*, 1996).
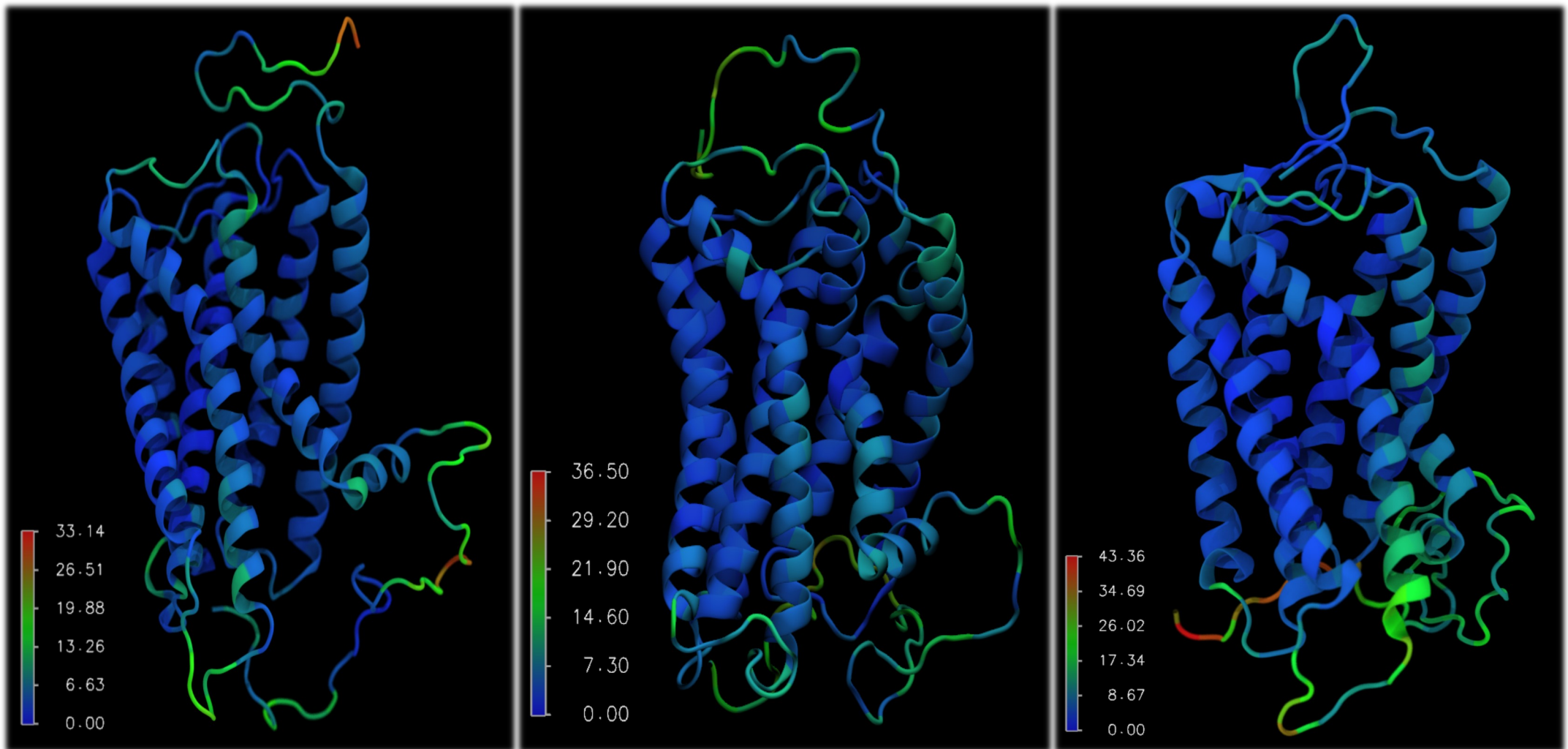
30ns Run 1 Mel1a GCPR RMSD Trajectory Overlay.
Red resembles high areas of movement Max 33.14Å, blue resembles low areas of movement (min of 0.00Å). The Mel1a GCPR is shown in Using the trajectory data from the DCD file and displayed in New Cartoon.

30ns Run2 Mel1a GCPR RMSD Trajectory Overlay.
High RMSD is seen in loop sections similarly to 30ns Run 1, The Mel1a GCPR is shown using the trajectory data from the DCD file and shown in New Cartoon in VMD 1.9.3.

30ns Run 3 Mel1a GCPR RMSD Trajectory Overlay.
The maximum RMSD is 5Å (maximum of 43.36Å) higher in relation to the previous 30ns runs. The Mel1a GCPR is shown using the trajectory data from the DCD file and represented in New Cartoon.

Figure 6.0.1– 30ns Molecular Dynamics of MEL1A - RMSD coloured by trajectory.

30 Nanosecond Molecular Dynamics Run 1 Ramachandran Plot.

30 Nanosecond Molecular Dynamics Run 3 Ramachandran Plot.

30 Nanosecond Molecular Dynamics Run 3 Ramachandran Plot.
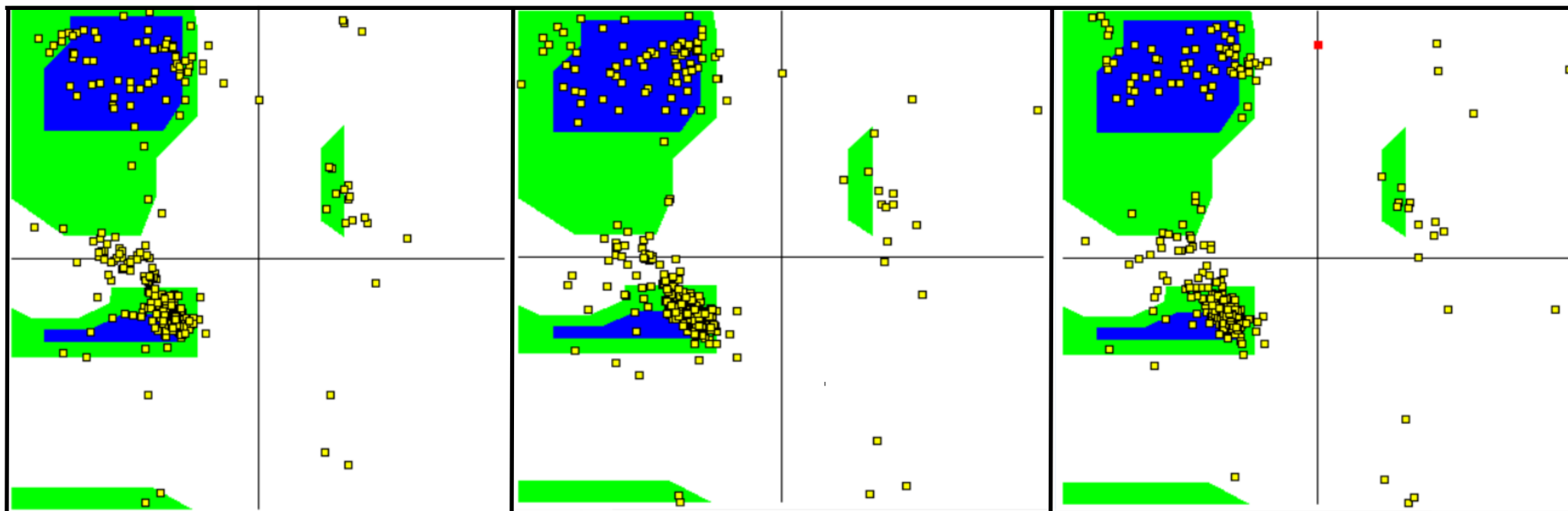
30 Nanosecond Control RMSD Plot Over Time
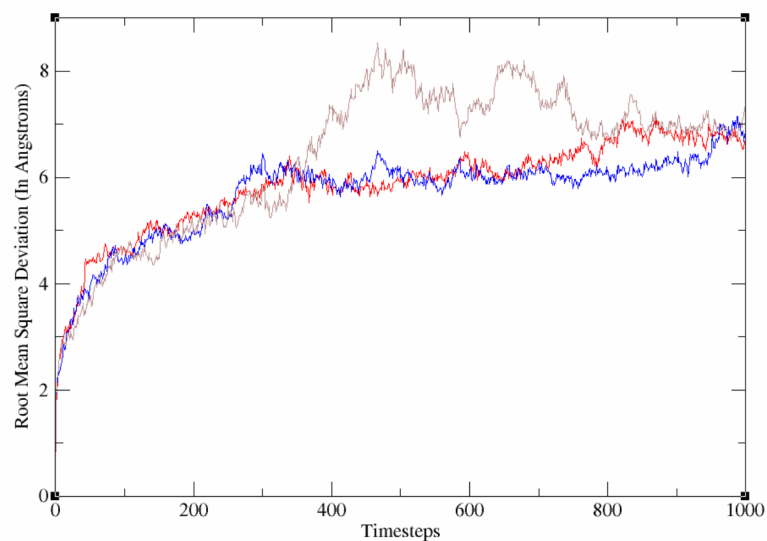
Figure 6.0.2 - Ramachandran plot analysis and RMSD over time in all three Mel1a GCPR 30ns MD.

30 Nanosecond Molecular Dyanmics Run 1 Temperature Over Timesteps
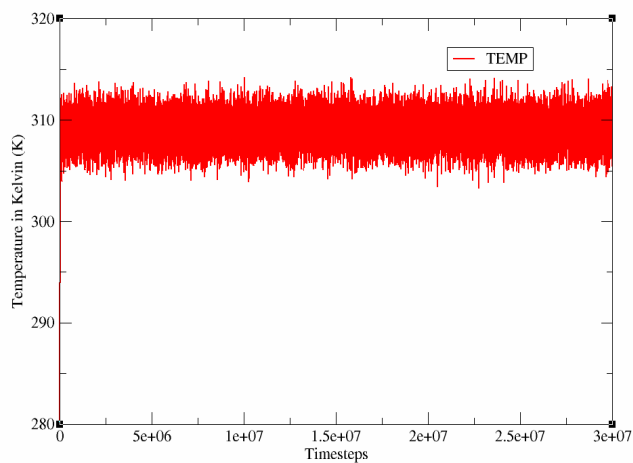
Kinetic Energy Over Timesteps Molecular Dynamics 30ns Run 1
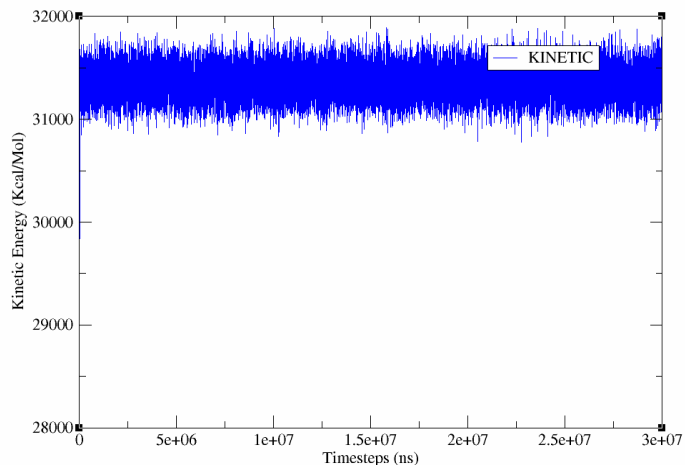
Figure 6.1.1 - Temperature of the Mel1a bi-lipid molecular dynamics system (left). Kinetic energy of the previously mentioned system (right).

Potential Energy Over Timesteps Molecular Dynamics 30ns Run 1
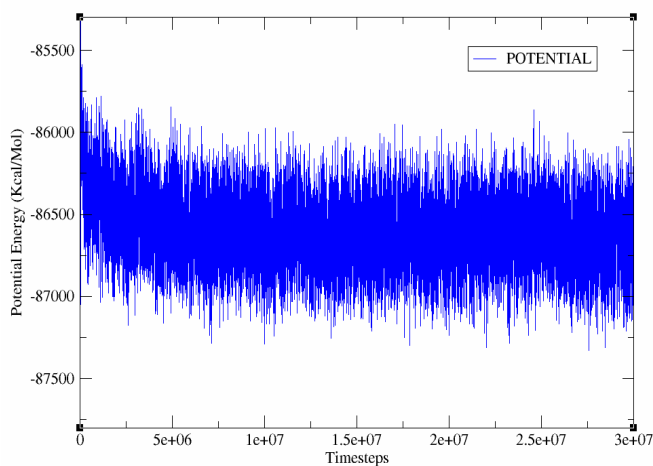
Van der Waals interaction over timesteps. Molecular Dynamics 30ns Run1
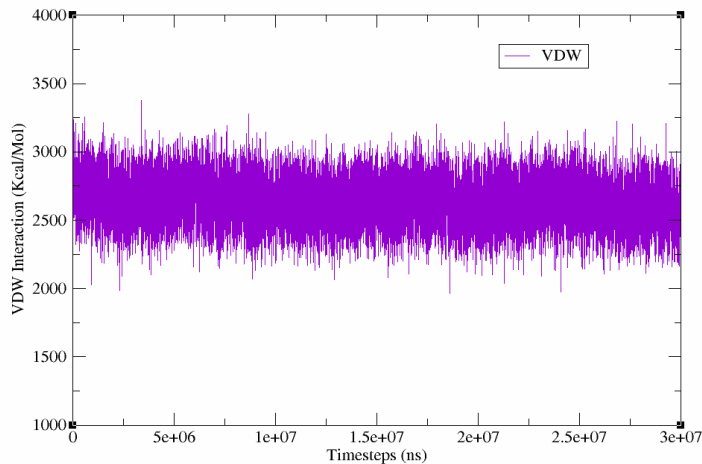
Figure 6.1.2 - Potential energy in the previously mentioned system (left graph) Van Der Waal interaction energy in the previously mentioned system.

Figure 6.1.3 - Bond Energy of the Mel1a bi lipid cell membrane environment (left).

Cys314 on the intracellular section of the Mel1a GCPR moves away from the core of the protein traversing the POPC membrane.

Cys314.

Met339.

Disulphide bond between Cys100 on Helix C, and Cys177 on beta turn A.
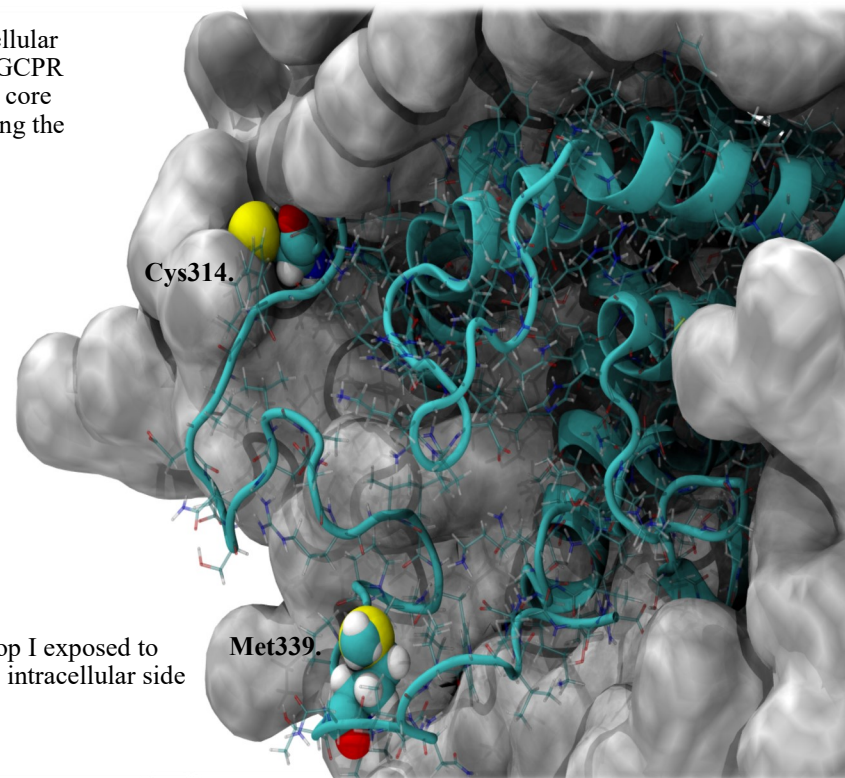
Met339 on loop I exposed to solvent on the intracellular side of Mel1a.
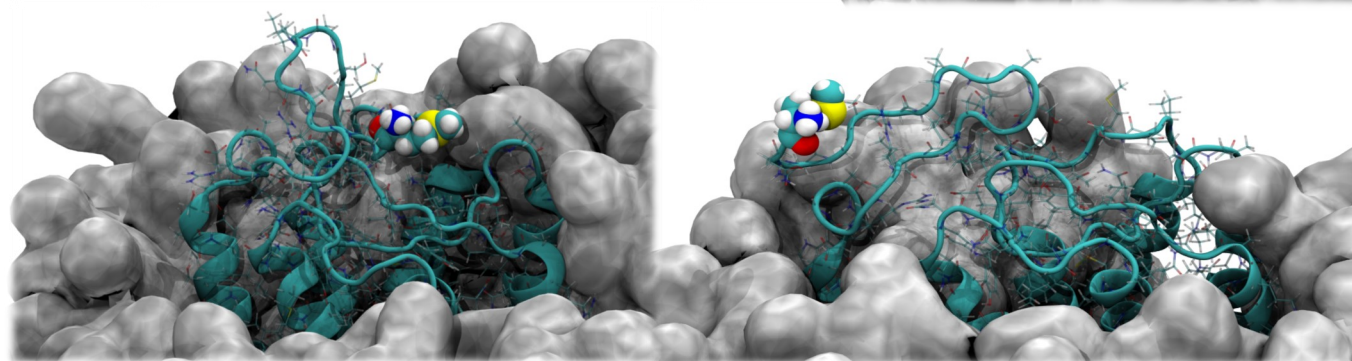
Met1 in the initial frame of the calculation sits close to the core, before moving out into solute and reconforming along with half of Loop A to the top of Helix A by the end of the calculation

Pi-Helix B (Shown in orange) retains its shape in trajectory two, starting at Pro80 and continuing to Asn90.

Figure 6.2.1—Visual analysis of Mel1a GCPR 30ns Trajectory 1.

MD 30ns Run 1 Mel1a and initial I-tasser Homology model of Mel1a GCPR.

MD 30ns Run 1 Mel1a and X-ray β2 Adrenergic GCPR'.

MD 30ns Run 1 Mel1a and Rhodopsin GCPR.
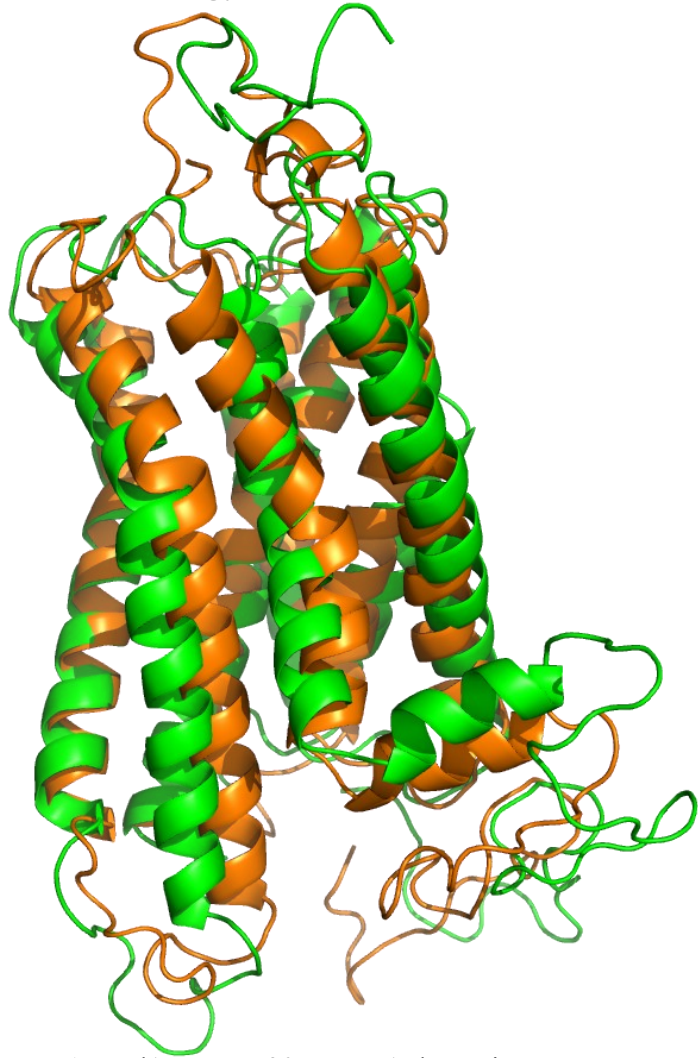
Part A. Mel1a GCPR 30ns Run 1 shown in Green, I-Tasser's Homology model of Mel1a GCPR shown in Orange. RMSD between the structures is 3.813Å.
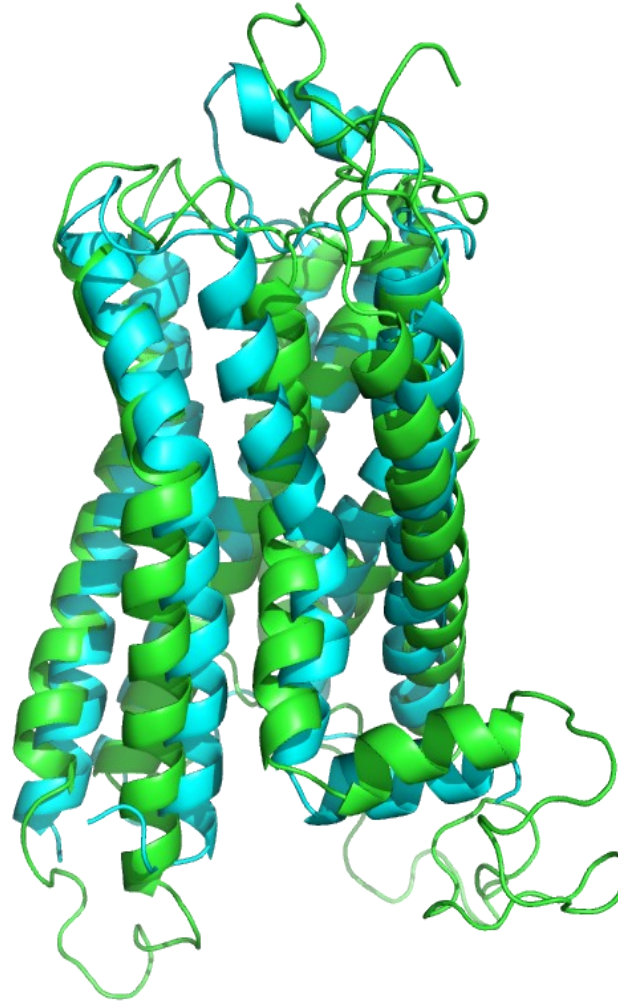
Part B. Mel1a GCPR 30ns Run 1 shown in Green, X-ray crystal structure of β2-adregernic GCPR shown in Cyan. RMSD between the structures is 3.119Å
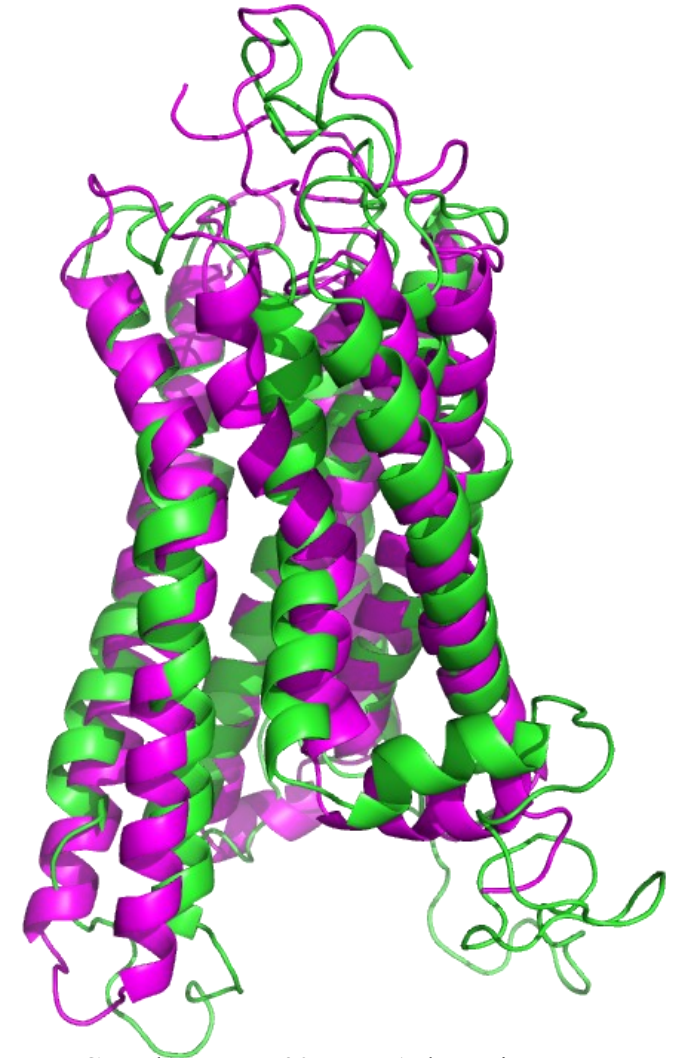
Part C. Mel1a GCPR 30ns Run1 shown in Green, X-ray Crystal Rhodopsin GCPR shown in Magenta. RMSD between the structures is 3.234Å

Figure 6.3.1 - Structural alignment of 30 ns run 1 to initial I-Tasser Homology Model of Mel1a and X-ray Crystal Structures of β2-Adregernic and Rhodopsin G-Coupled Protein Receptors (GCPR's).

Extracellular View. ↓

Intracellular View and Electronegative Cavity. ↓

(Electro-negative cavity on the internal side of the membrane. (Above).
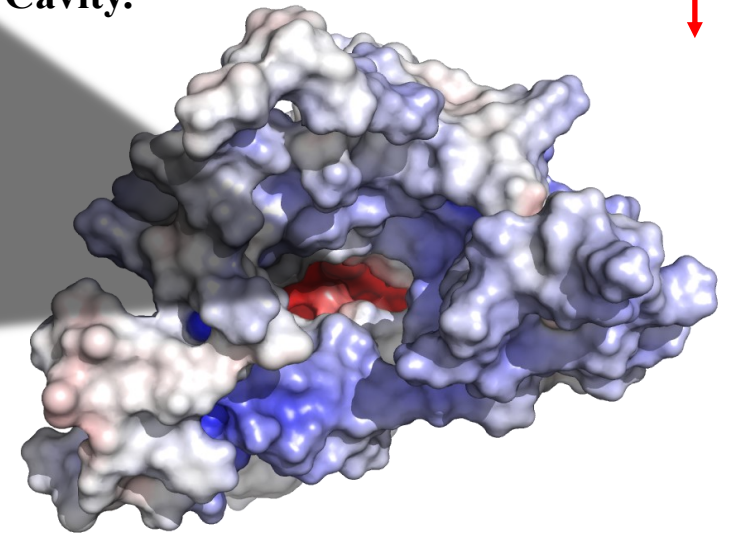
Side View (membrane portion). ↓
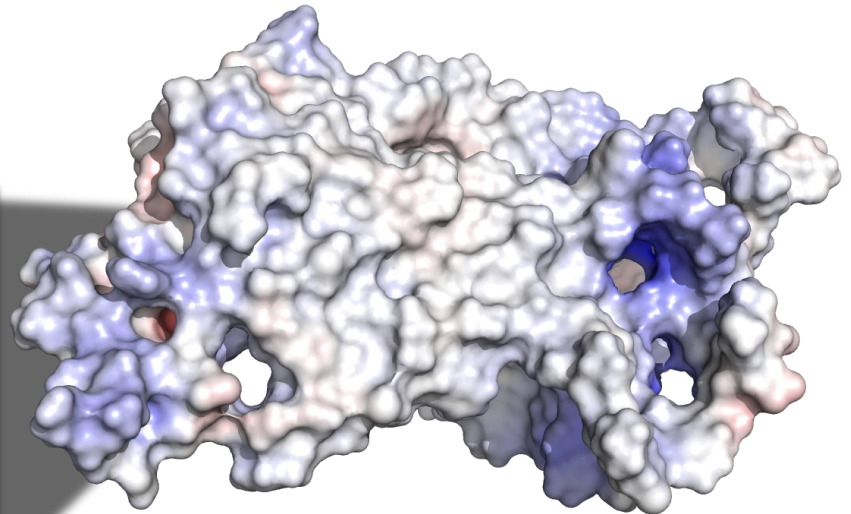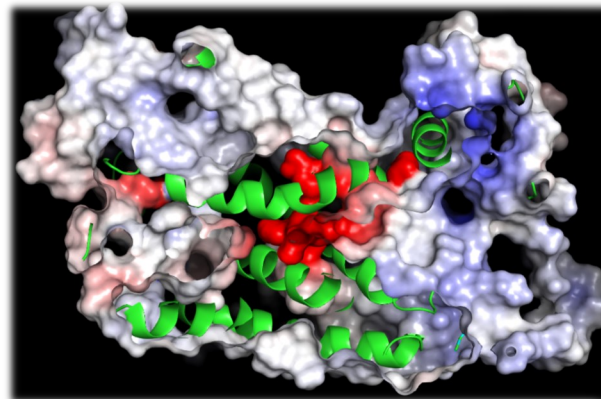
Electrostatics images rendered in Pymol using the APBS electrostatics plugin, utilizing the AMBER forcefield to assess the surface interaction properties of the first 30 nanosecond run. The final frame of the simulation is saved through VMD prior to being loaded into Pymol. Areas displayed in red are electro-negative, Areas shown in blue are positive. Neutral areas have a white appearance.

Side Slice of Mel1a to show electronegative pocket in the core of the GCPR.

Figure 6.4.1 - Electrostatic analysis of Mel1a MD run 1, 30ns.

Kinetic Energy Over Timesteps 30ns Molecular Dynamics Run 2
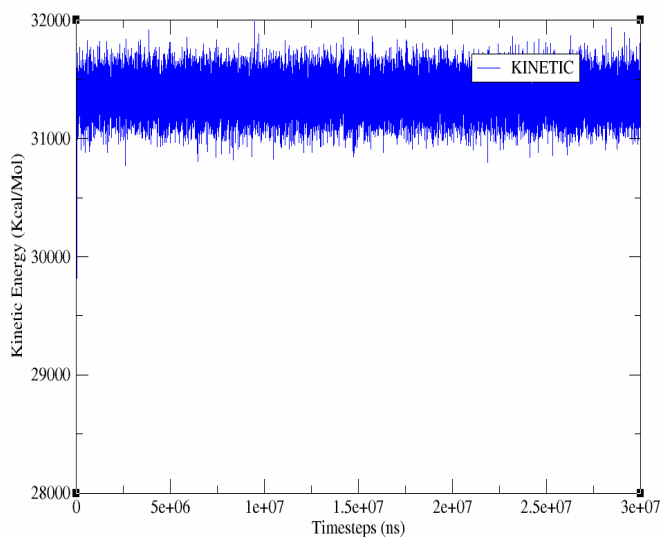
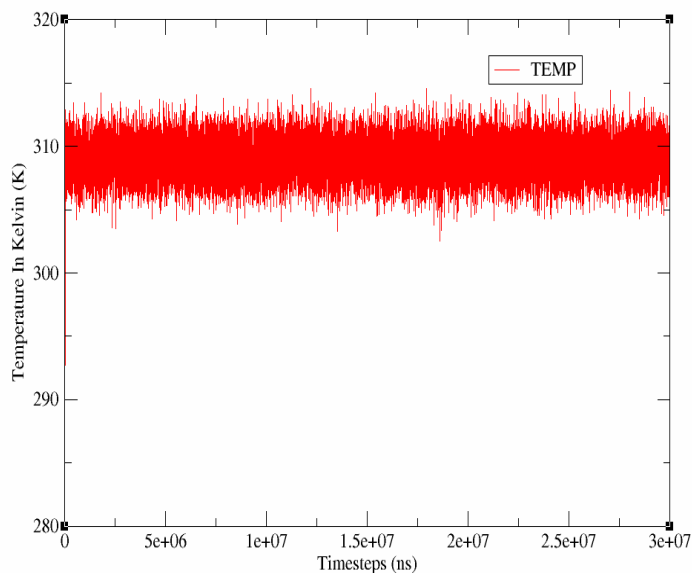Temperature Over Timesteps 30ns Molecular Dynamics Run 2
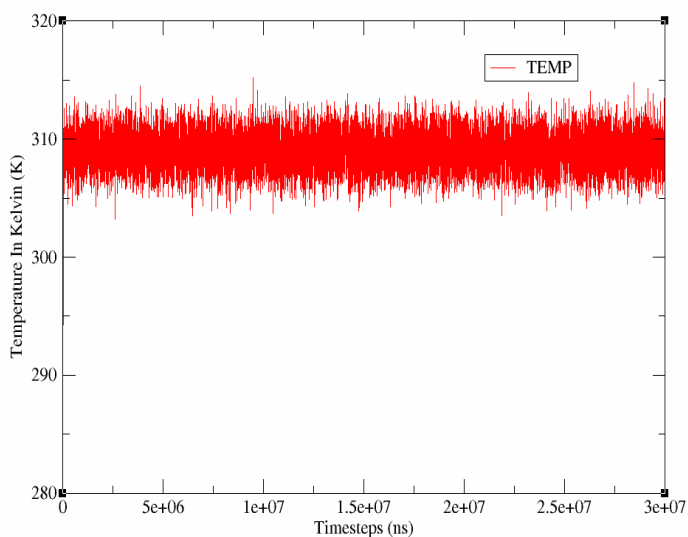
Figure 6.5.1—kinetic energy in the Mel1a bi-lipid system 30ns Run 2 (left) , and temperature in Kelvin of the previously mentioned system (right).

Van der Waal Interaction, Over Timesteps 30ns Molecular Dynamics run2
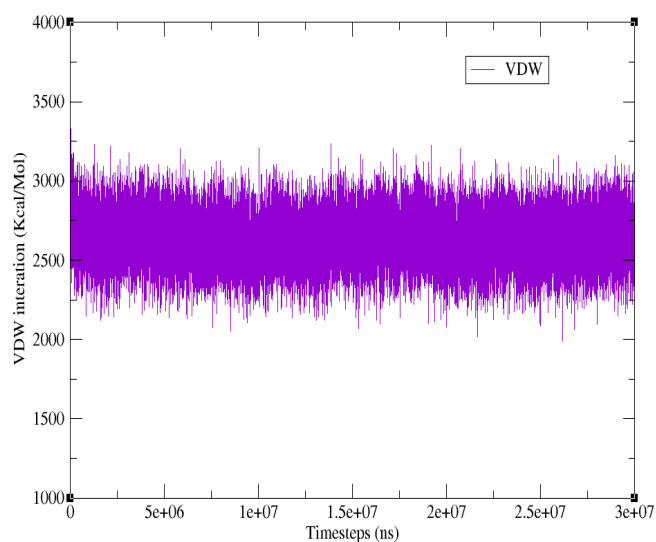
Potential Energy Over Timesteps, 30ns Molecular Dynamics Run 2
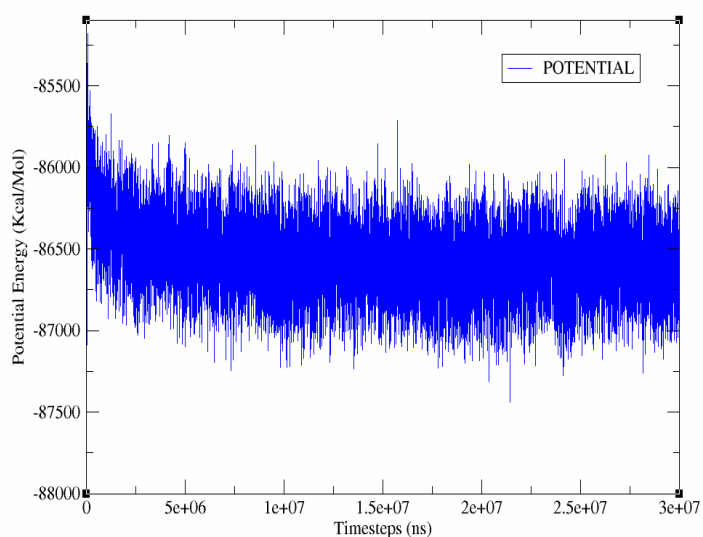
Figure 6.5.2 - Van Der Waal Interaction energy in the entire Mel1a GCPR membrane system (left). Potential energy in the system (right).

Figure 6.5.3 - Bond Energy in the system (left).
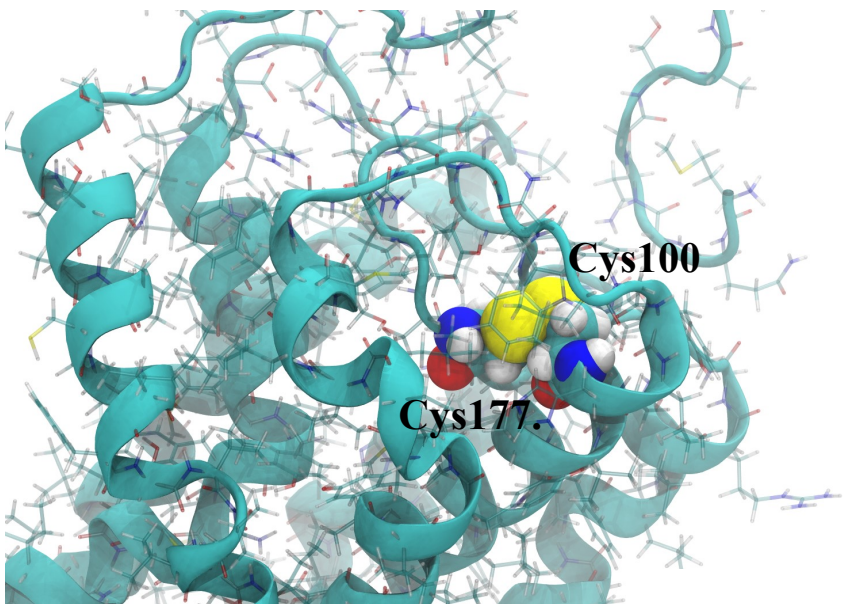
Disulphide bond present in 30ns Mel1a GCPR, Trajectory 2 between Cys100 on Helix C and Cys177 on Beta Turn A.

Met1 which breaks away from the core of the protein and then reconfirms on the surface of the extracellular side as the calculation ends.

Alpha Helix A is split into two helix sections between Thr37 and Val40. (above)

Figure 6.6.1—Visual analysis of the 30ns Mel1a GCPR trajectory 2.

Phe89.

Pro80.

Cys314

Met339

Loop I spreads out over the POPC membrane on the intracellular side. (highlighted in orange).

Lys226

Gln219
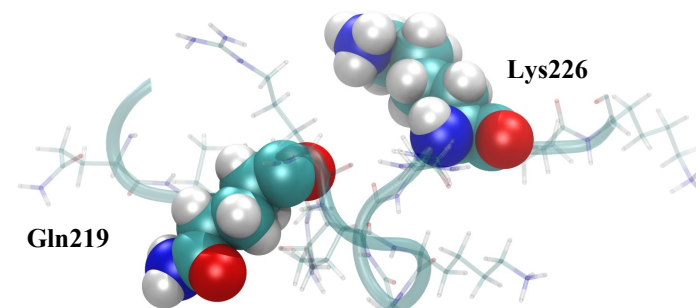
Loop E helix-like conformation before Helix F.

Proline (Pro80) induced bend in Helix B in the final frame of the calculation. Leading to Phe89

Gln274.

42.93Å

Asn299

Ile309.

Gln274.

38.48Å

Asn299.
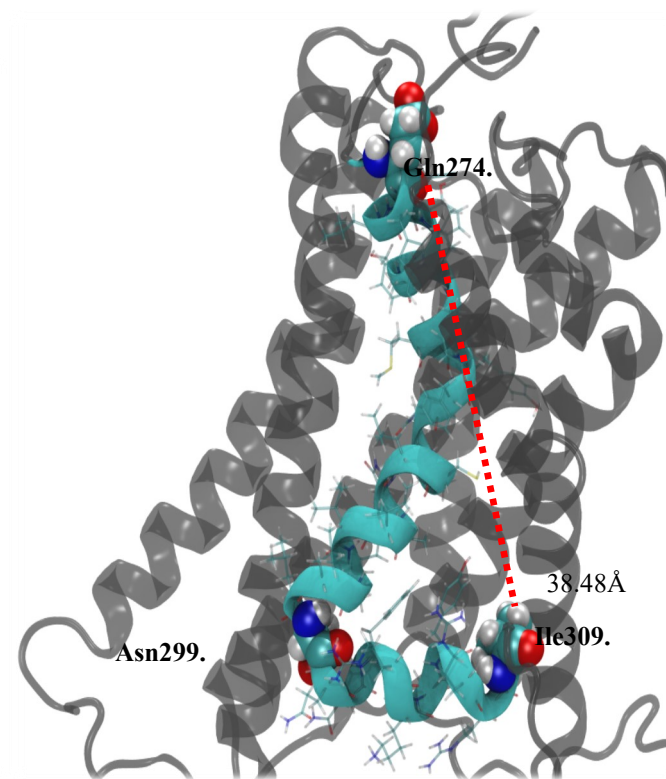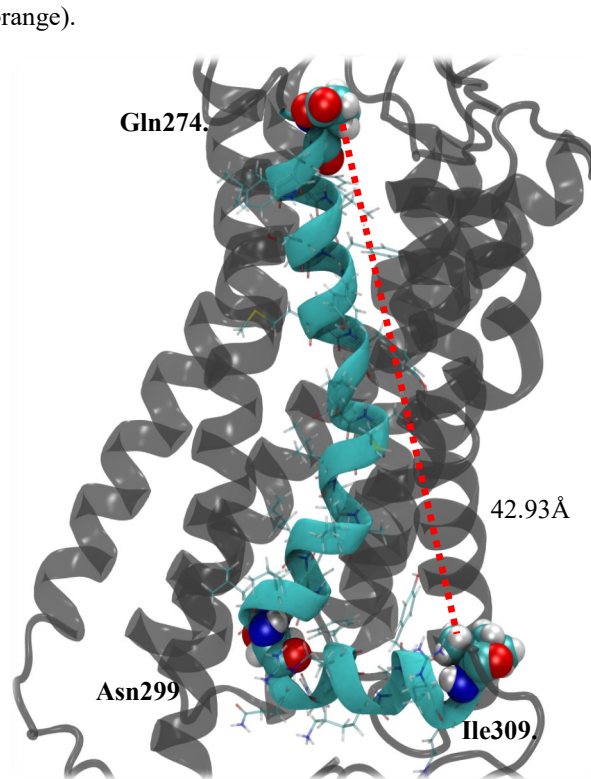
Ile309.

Measured from Gln274 to Ile309, Helix-Turn Helix G collapses on its axis from 42.94 (left) to 38.48 (right).

Figure 6.6.2—Visual analysis of the 30ns Mel1a GCPR trajectory 2.

MD 30ns Runs Mel1a and X-ray Initial I-TASSER Homology model of Mel1a GCPR.

MD 30ns Run 2 Mel1a and X-ray β2 Adrenergic GCPR.

MD 30ns Run 2 Mel1a and Rhodopsin GCPR.
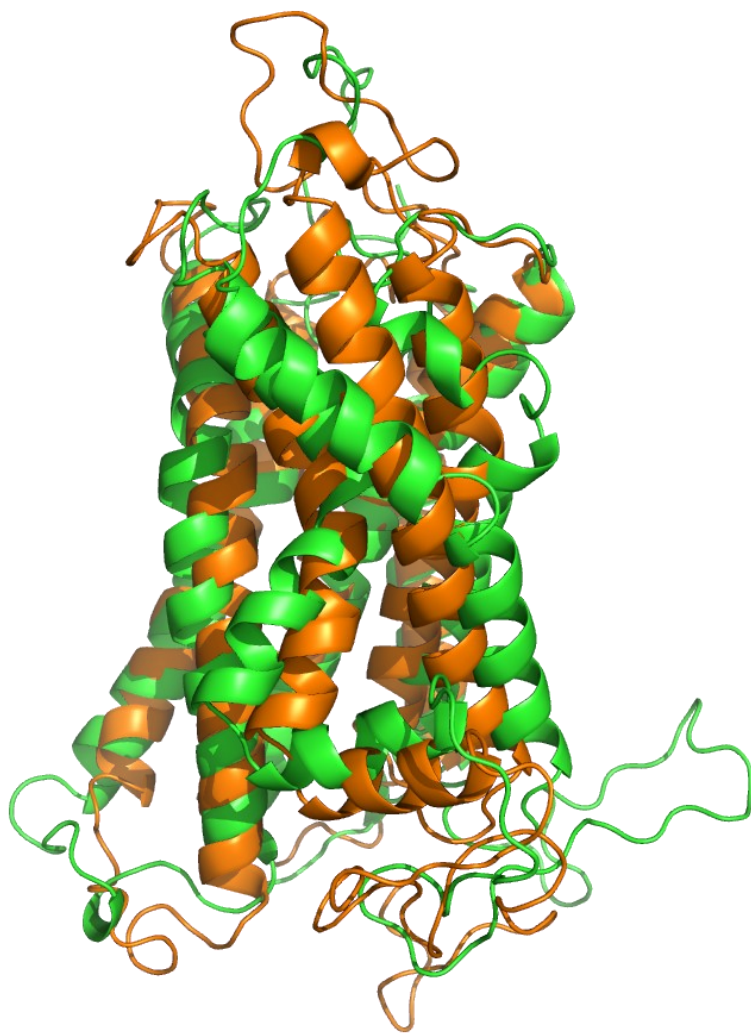
Part A -Mel1a GCPR run 2 Shown in Green, initial I-TASSER Homology model GCPR shown in Orange. RMSD between the Structures is 4.194Å.
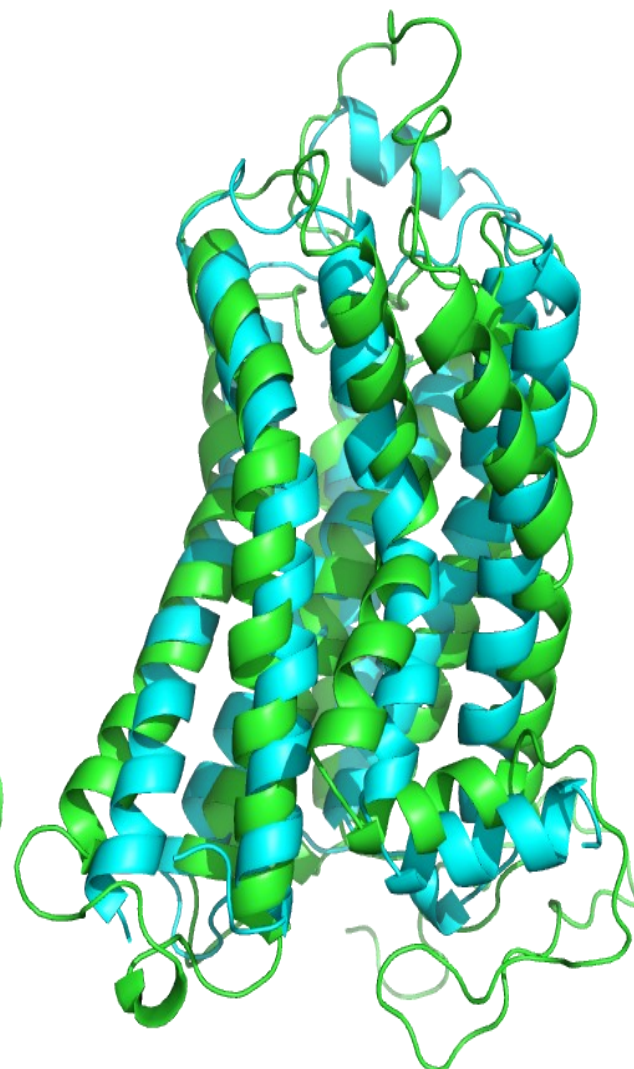
Part B -Mel1a GCPR run 2 Shown in Green, X-ray crystal Structure of β2-adregenic GCPR shown in Cyan. RMSD between the Structures is 3.563Å.
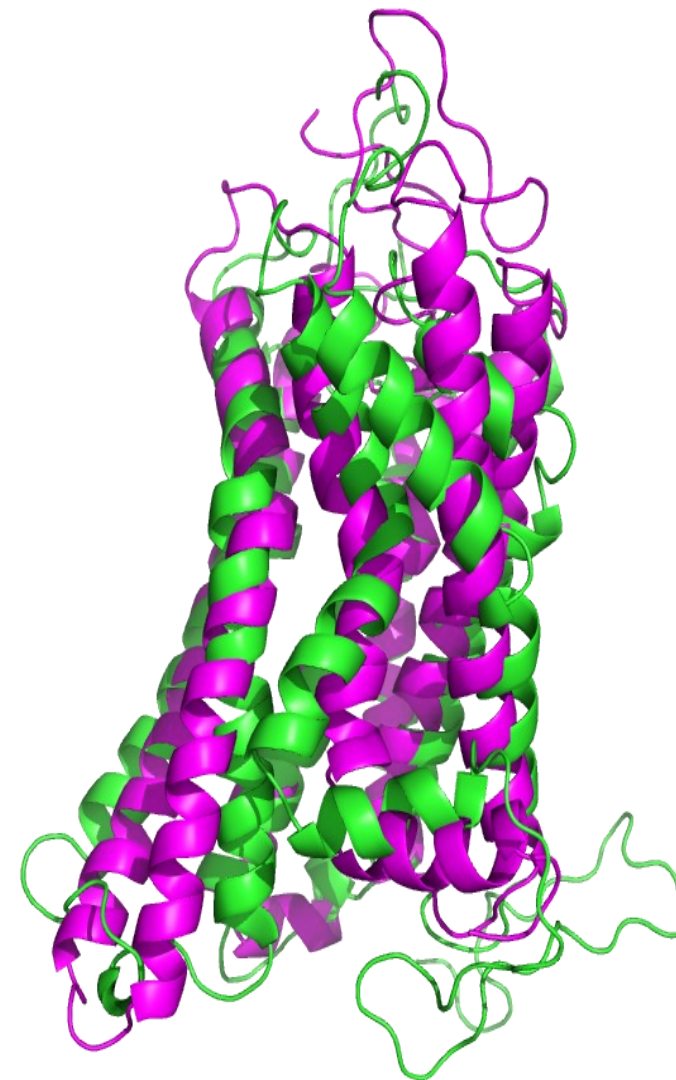
5. Mel1a GCPR 30ns Run 2 shown in Green, X-ray Crystal Rhodopsin GCPR shown in Magenta. RMSD between the structures is 3.885Å.

Figure 6.7.1 Structural alignment of Mel1a 30 ns Run 2 to X-ray Crystal Structures of β2-Adregernic and Rhodopsin G-Coupled Protein Receptors (GCPR's).

Extracellular View. ↓

Intracellular View and Internal Cavity. ↓
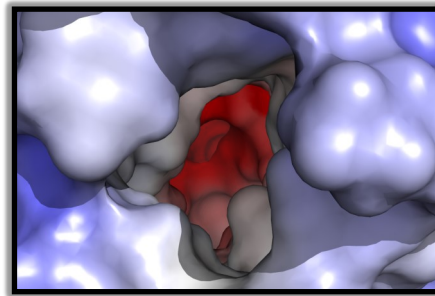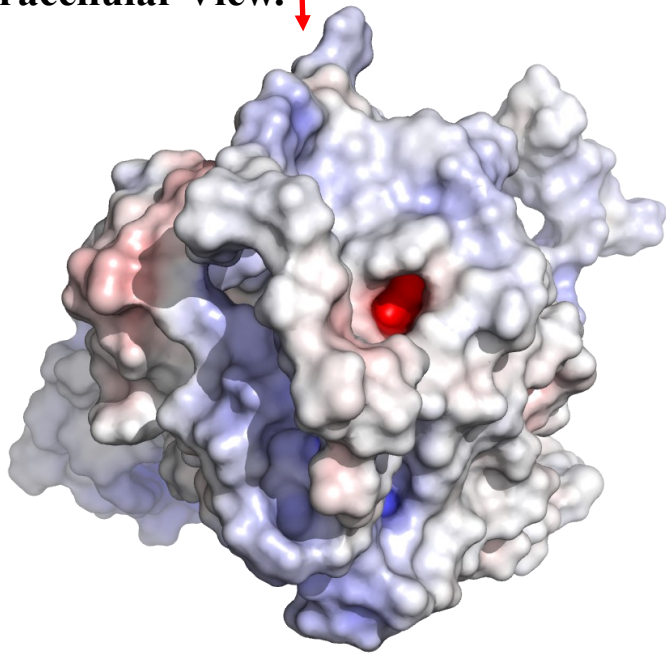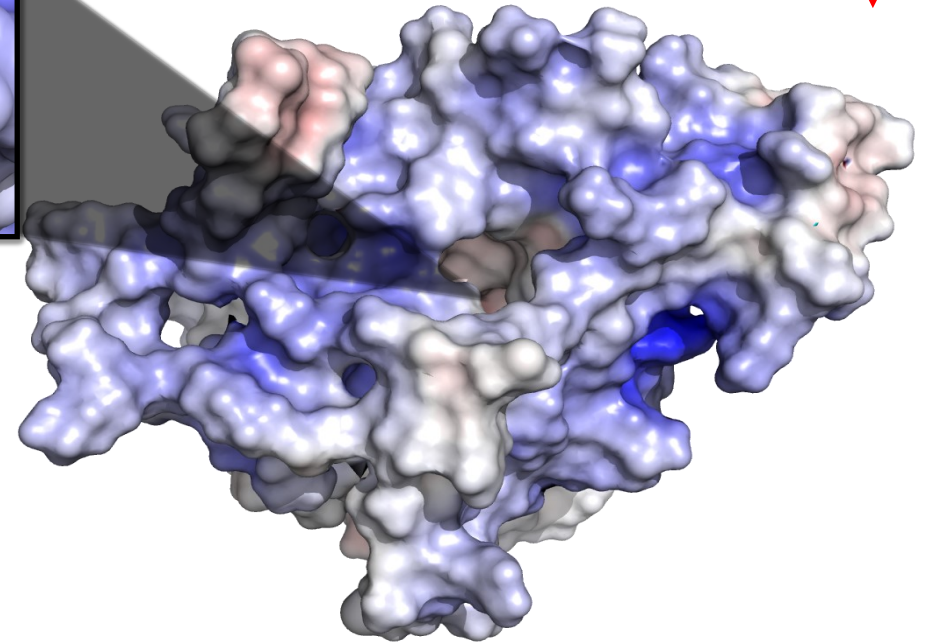
Intracellular Cavity from Run 2 takes a different conformational state in comparison to run 1 - but still persists on the internal side of the membrane. (above)

Electrostatics surface rendered in Pymol using the APBS electrostatics plugin, utilizing the AMBER forcefield to assess the surface interaction properties of the second 30 nanosecond run. The final frame of the 30ns simulation is saved through VMD prior to being loaded into Pymol. Areas displayed in red are electro-negative, areas shown in blue are positive. Neutral areas have a white appearance.

Side View (Membrane Portion). ↓
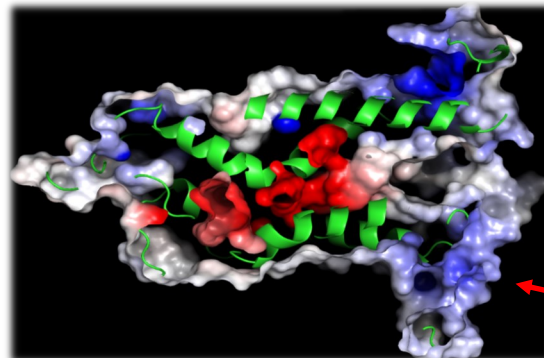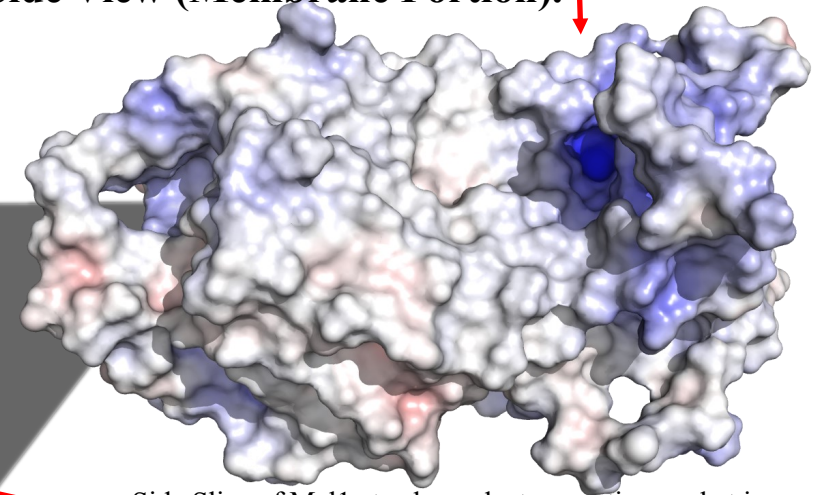
Side Slice of Mel1a to show electronegative pocket in the core of the GCPR.

Figure 6.8.1 Electrostatic Analysis of Mel1a MD run 2, 30ns.

Kinetic Energy Over Timesteps 30ns Molecular Dynamics Run 3
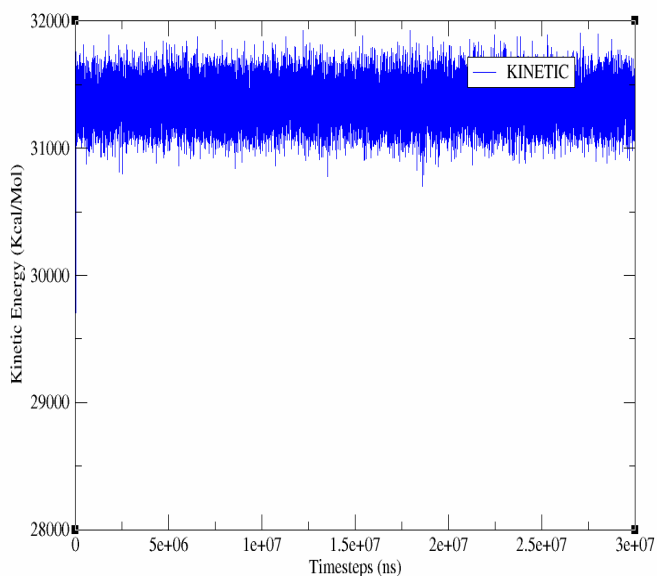
Temperature Over Timesteps 30ns Molecular Dynamics Run 3

Figure 6.10.1 - kinetic energy in they Mel1a 30ns System (left) and temperature in Kelvin of the Mel1a 30ns system, trajectory 3 (right)

Van der Waal Interactions Over Timesteps 30ns Molecular Dynamics run3
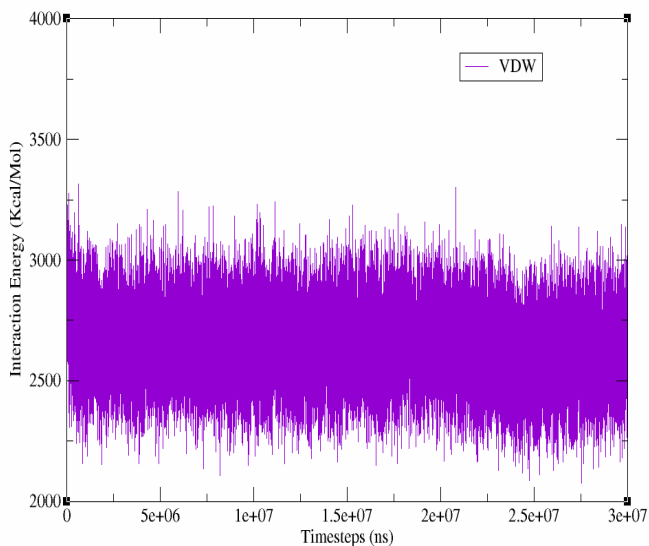
Figure 6.10.2 - Van Der Waal interaction energy in the Mel1a GCPR system (left) and negative potential energy (right).

Figure 6.10.3 - Bond energy in the Mel1a GCPR system.
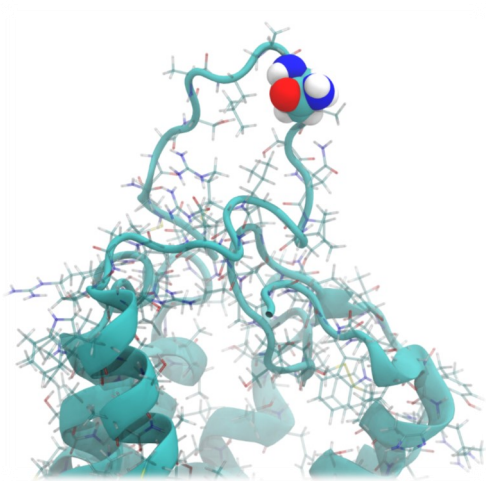
Loop A, focus on Asn10 is conformational searching in the solvent-ion environment throughout the calculation.
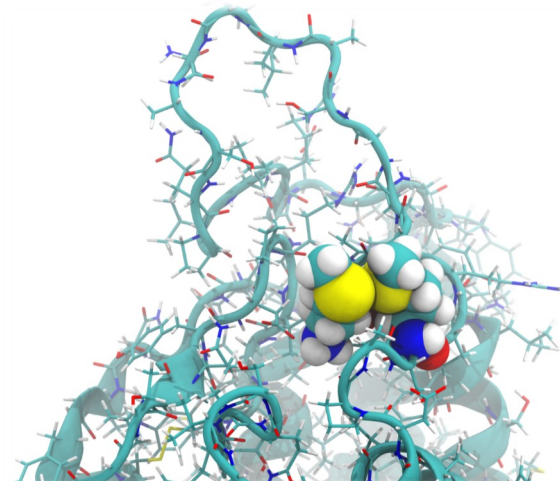
Met1 on loop A and met268 are seated closely on the extracellular helix sections throughout the calculation.
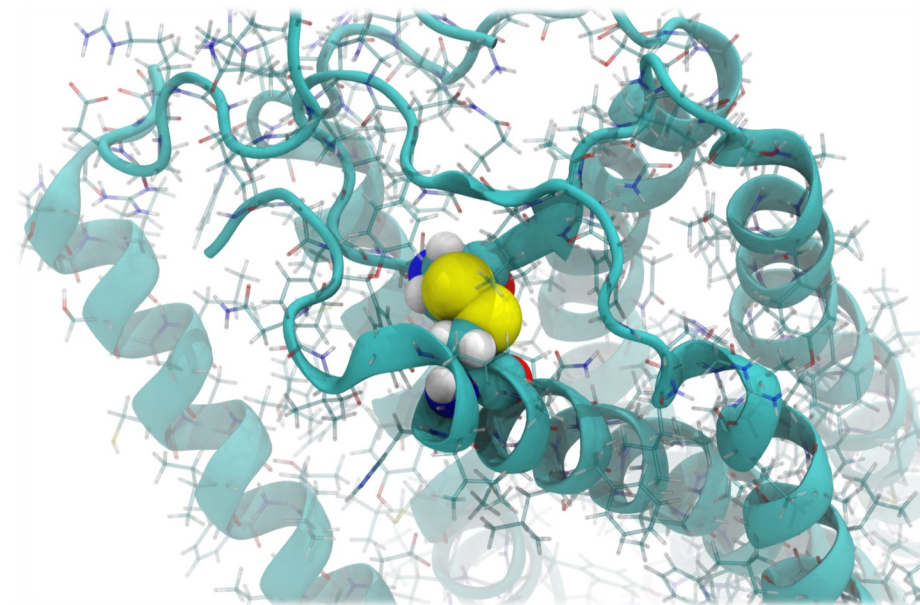
The disulphide bond between Cys100 on helix C and cys177 on beta turn A persists in the 3rd 30ns trajectory for the Mel1a GPCR.
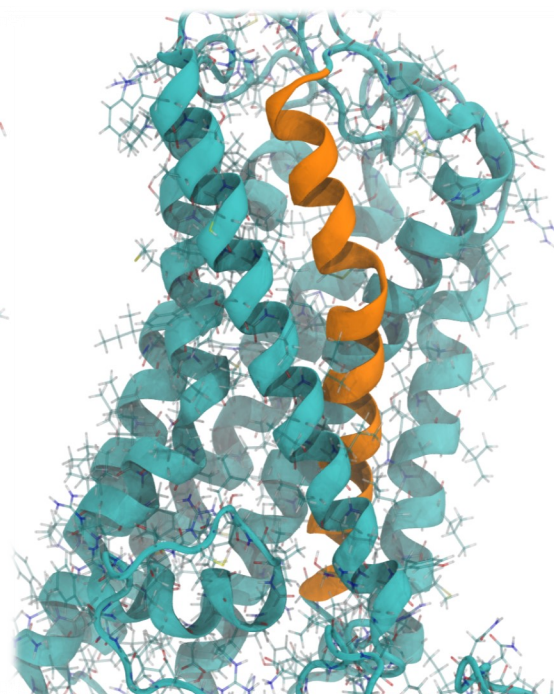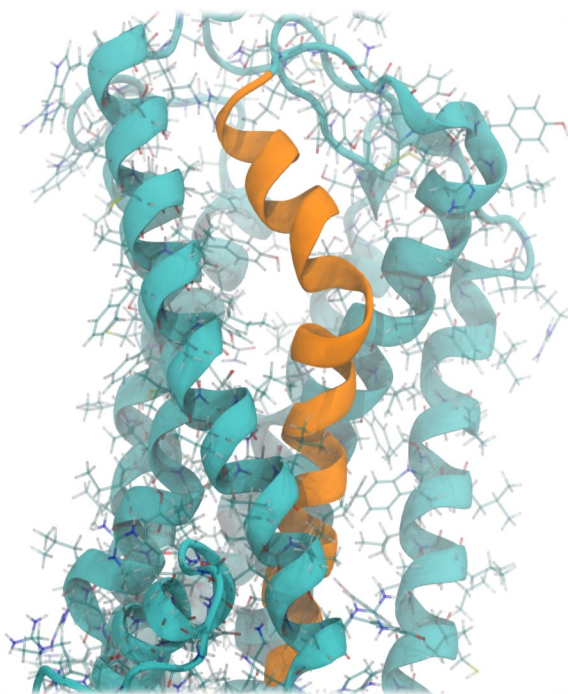
The upper section of Helix B (pro80 to Asn90) remains in pi-helix conformation throughout the calculation, but moves away from the core as the calculation continues.

Figure 6.11.1—Visual analysis of the 30ns Mel1a GCPR trajectory 3.
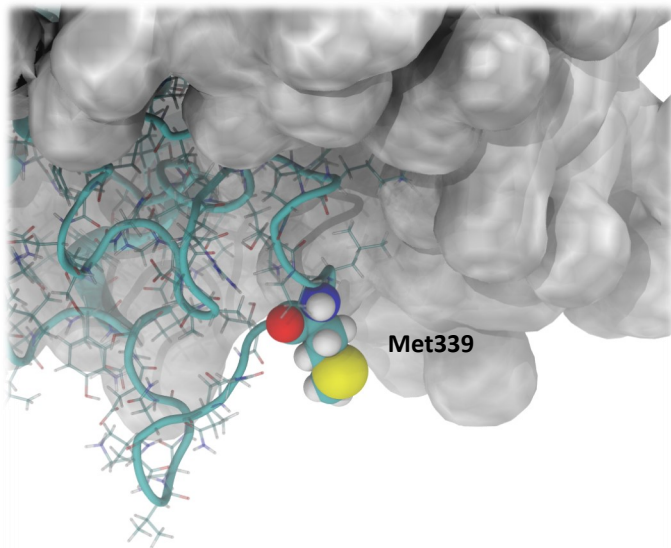
Met339 exposed to solute on the intracellular side of the bili-pid membrane.

Cys314 and the early residues in Loop I retracted into the membrane with the central core via the conformational change of Helix-turn-helix G
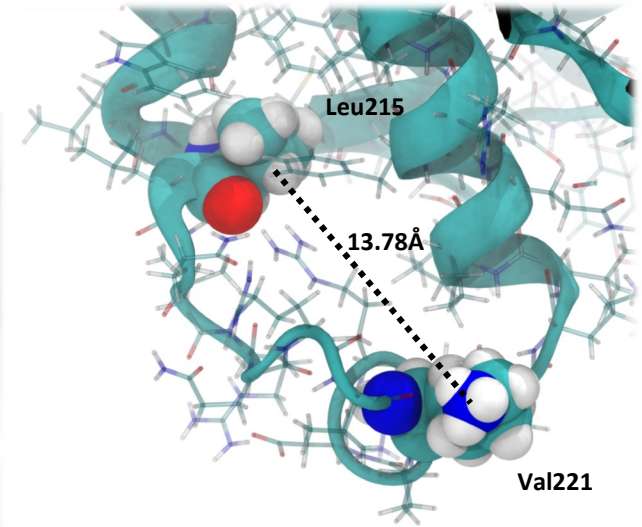
Initially loop G on the intracellular side sits 13.78Å apart (measured from Leu215 on helix F to Val221 on loop G).

Met339 later in the simulation sitting closer to the bilipid membrane on the intracellular side.
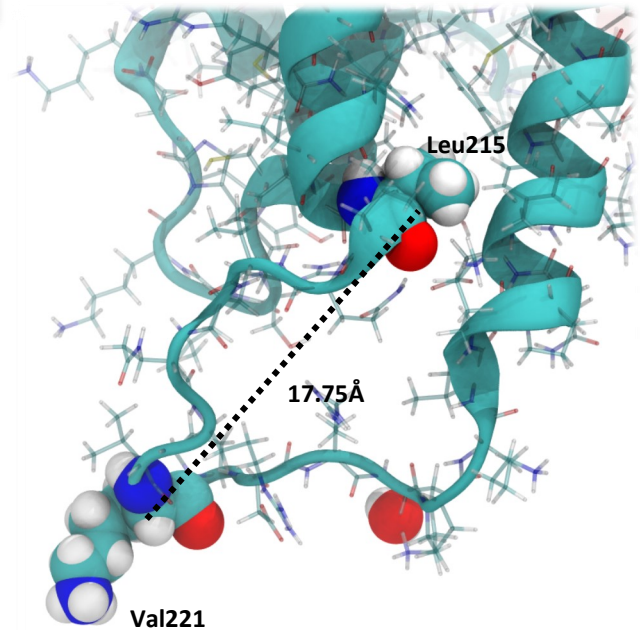
Loop G on the intracellular side sits 17.75Å apart (measured from Leu215 on helix F to Val221 on loop G). By the end of the trajectory dataset.

Figure 6.11.2—Visual analysis of the 30ns Mel1a GCPR trajectory 3.

MD 30ns Runs Mel1a and initial I-TASSER homology model of Mel1a GCPR.

MD 30ns Runs Mel1a and X-ray β2 Adrenergic GCPR.

MD 30ns Run 1 Mel1a and Rhodopsin GCPR.
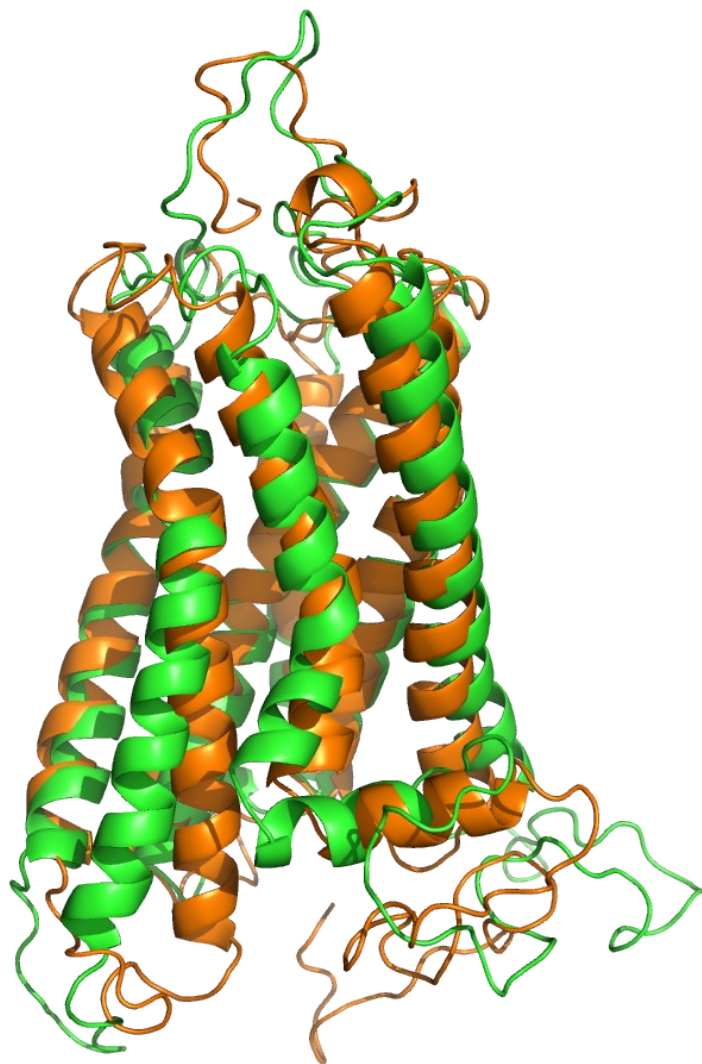
Part A—Mel1a GCPR 30ns run 3 show in in Green, Initial I-Tasser Homology Model of Mel1a GCPR shown in Orange. RMSD between the structures is 4.031Å
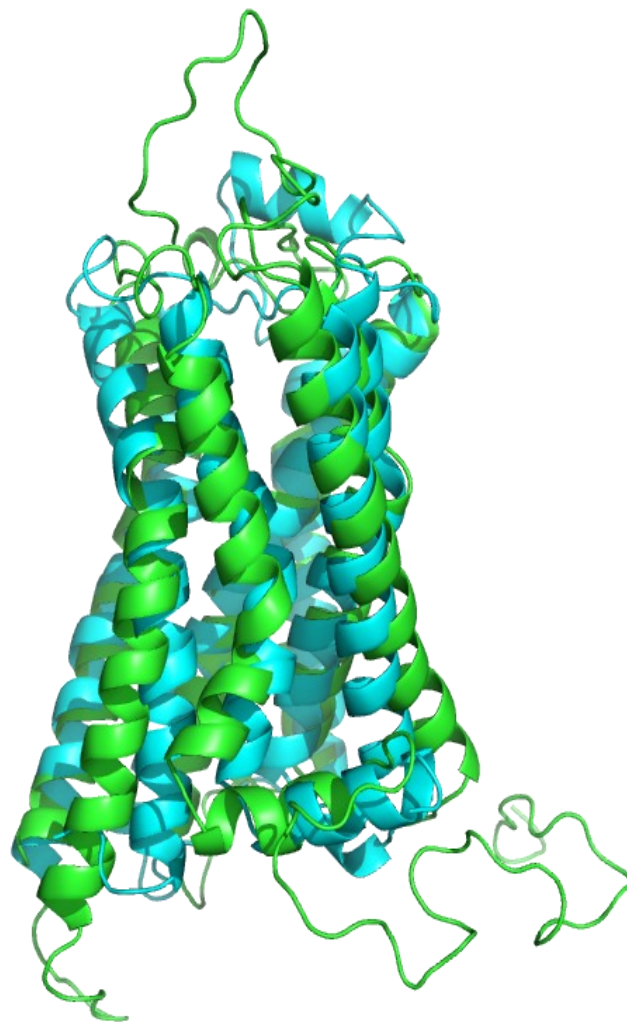
Part B—Mel1a GCPR 30ns run 3 show in in Green, X-ray crystal structure of β2-adregenic GCPR shown in Cyan. RMSD between the structures is 3.671Å
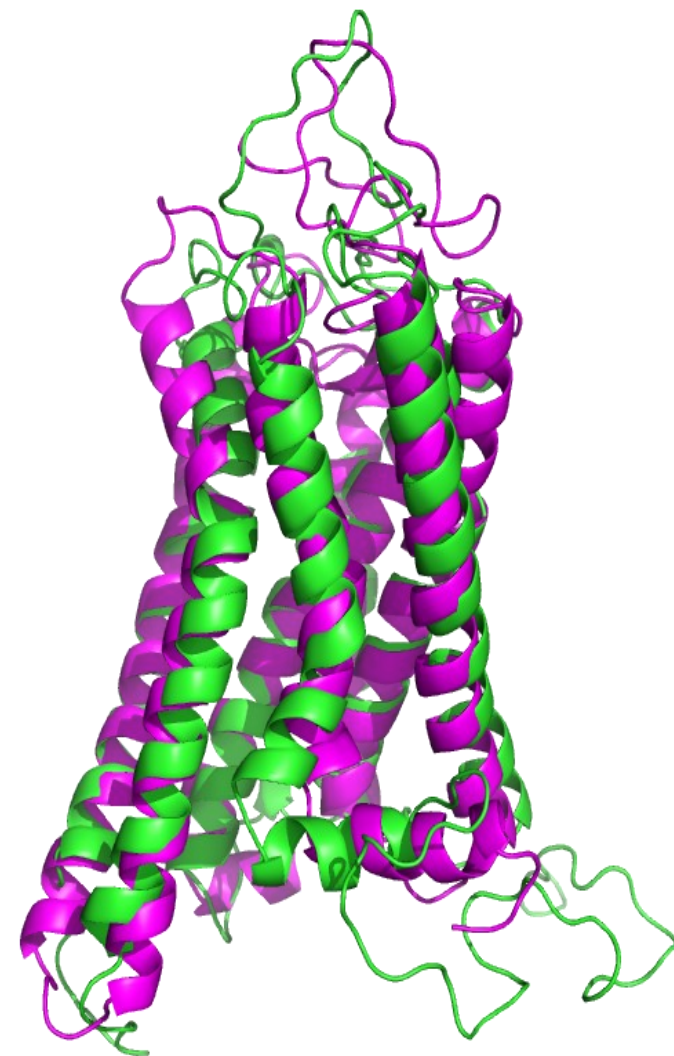
6. Mel1a GCPR 30ns Run 3 shown in Green, X-ray Crystal Rhodopsin GCPR shown in Magenta. RMSD between the structures is 3.706Å

Figure 6.12.1 - Structural alignment of the Mel1a GCPR, 30ns Run 3 to X-ray Crystal Structures of β2-Adrenegic and Rhodopsin GCPR's.

Extracellular View and Cavity.
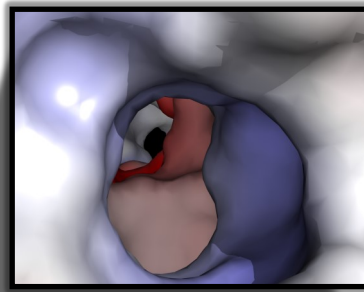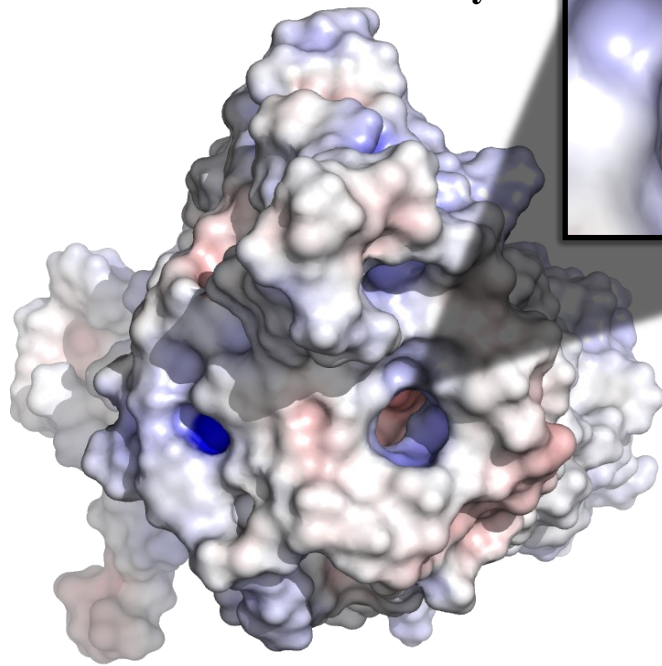
Extracellular cavity (left) which follows through the core of the protein to the Intracellular side of the protein (displayed below)

Intracellular View and Cavity.
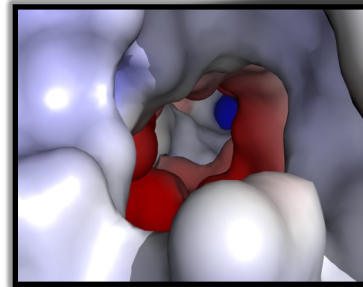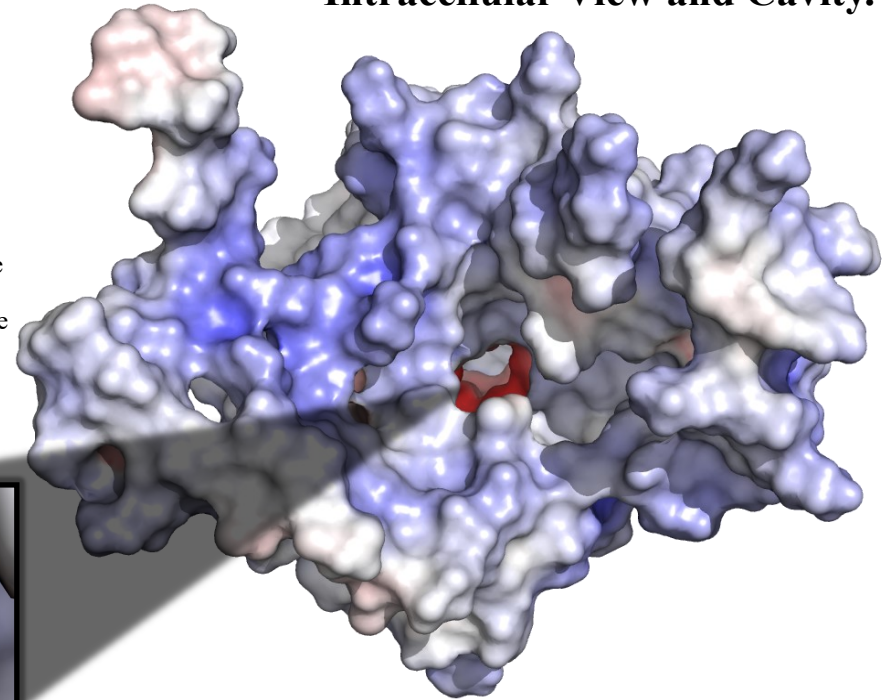
Electrostatics surface rendered in Pymol using the APBS electrostatics plugin, utilizing the AMBER forcefield to assess the surface interaction properties of the third 30 nanosecond run. The final frame of the 30ns simulation is saved through VMD prior to being loaded into Pymol. Areas displayed in red are electro-negative, areas shown in blue are positive. Neutral areas have a white appearance. Unlike simulations 1 and 2 - the protein has a central electronegative channel which travels from one side of the membrane to the other.
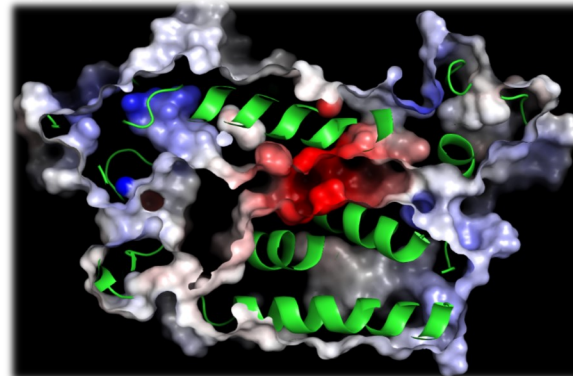
Cutaway side view of the protein core to show the electronegative central channel within.

Side View (membrane Portion).
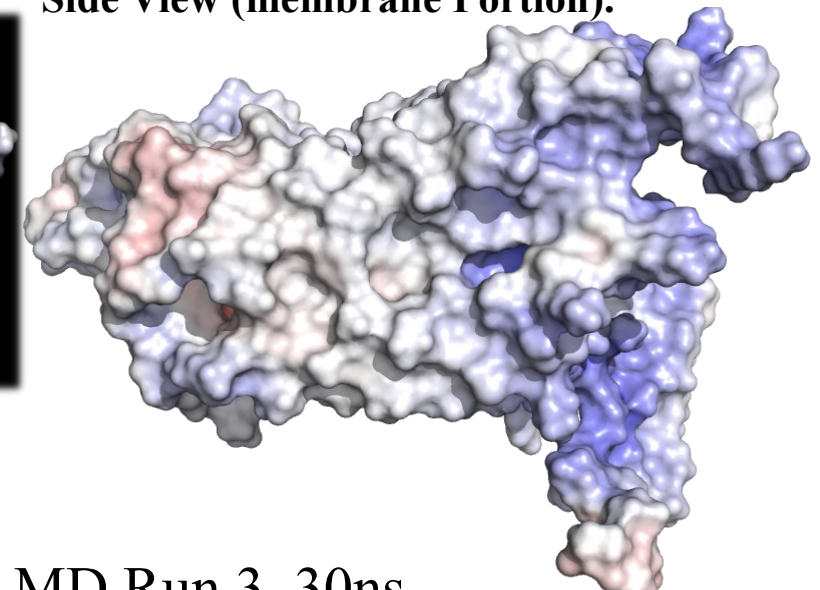
Figure 6.13.1 Electrostatic Analysis of the Mel1a GCPR MD Run 3, 30ns.

7.0 - Gaussian16 Ligand Generation Analysis and Comparison to 1972 X-Ray Crystal Structure of Melatonin.

The Melatonin ligand, parametrised via the FFtk (Force Field Tool Kit) plugin in VMD 1.9.3 was compared to the crystal structure of melatonin published in 1972 by (Wakahara, Fujiwara and Tomita, 1972). Bond lengths, angles and dihedrals were cross referenced with the crystal structure in order to validate the generated forcefield. All the CHARMM36 atom types mentioned in the tables below are referenced in Figure 7.0.1, with a key for Melatonin explaining the atom types of the structure. This chapter contains data for the Gaussian16 structure generated using quantum mechanical methods (Frisch *et al.*, 2016), the crystal structure data PDB, and the post minimization data of the parameterised structure using NAMD. This is shown in tables 1 bond angles, 2 bond lengths and table 3 which shows bond dihedrals. A reference figure for the QM derived and X-ray crystal versions of the ligand is shown in Figure 7.0.2. The X-ray structure is planar, when compared to the QM structure which varies in the N-acetyl section of the molecule, the C10-C10 atoms are angular sitting in an upwards direction from the planar indole. The Oxy-methyl section is also angled and not planar like the diffracted version of melatonin.

7.1 Minimization of the Gaussian16 Parameterized Melatonin Ligand in CHARMM36.

A conjugant gradient minimization was performed using NAMD and sourcing the FFTK generated forcefield parameter file from the G16 geometry optimization (PDB) which was run on 10 threads for a total of 5000 timesteps. The melatonin ligand was placed into a TIP3P water box with 10Å padding around the physical dimensions of the molecule using the auto-solvate tool available in VMD1.9.3. The bond angles, dihedrals, and length data were charted against simulation time. The graphs were split up into sub sections of the ligand for ease of interpretation.

Oxy-Methyl portion of the ligand.

Melatonin's bond angles in the oxy-methyl section of the ligand are shown in figure 7.1.3. initial fluctuation is seen at the start of the simulation. However, after the first 500 timesteps the data levels out showing a stable angular conformation for all atom-types. C7-O1-C9 displays stability at 123.55°, C6-C7-O1 shows angular stability at 104.14° and C8-C7-O1 shows angular stability at 129.52°. The final values of the simulation can be seen in table 1.

The oxy-methyl bond lengths (seen in figure 7.1.4) in the graph show an average stability after a settling period spanning the first 500 timesteps of the minimization. C9-01 reduces from 1.35Å- 1.45Å initially and by the end of the minimization to a bond length of 1.41Å. The C7-O1 bond similarly settles in the first 500 timesteps balancing in a similar length rage of 1.35Å to 1.45Å. The final value for this bond length is 1.40Å. (see table 2)

in figure 7.1.5 dihedral properties of the oxy-methyl section of melatonin throughout the minimization begin settle after the first 250 timesteps and. by 1000 timesteps they have stabilised. C8-C7-O1-C9 shows a 47.92° angle after the minimization. Dihedral angle C6-C7-O1-C9 similarly stabilised after. Final dihedral angle values are shown in table 3.

Tryptophan-Derived Indole Section of the ligand.

During minimization of the ligand, the bond angle data displays a settling period of 500 timesteps. (figure 7.1.6) After which the angles settle as melatonin water box reaches a minimized state. Data for the final bond angles of the indole can be seen in table 1

After a settling period of 500 timesteps, the data settles showing narrower information about

bond lengths. Since the indole has double bonds in the ring structures which change, the lengths periodically change causing a spike in the data. This is due to the CHARMM General Forcefield RTF file which alters the bond lengths of the indole to simulate resonance. The final bond lengths can be seen in table 2.

5 N-acetyl Section of the ligand.

Most of the bond angle data settles after 500 timesteps (Figure 7.1.7)- except for the N2-C10-C11 angle, which fluctuates throughout the minimization. Presenting the impression that it is an area which is highly adaptive to its solvent environment in addition to the rest of the melatonin molecule and has many angular conformations.

During the first 500 steps bond length fluctuation (Figure 7.1.8) reduces significantly from bond lengths between 1.8Å and 0.2Å, after which bond levels seem to stabilize between 1.6Å and 0.6Å implying minimization of the ligand. Final figures for individual bond lengths can be seen in table 2. Since many different types of bond exist in the chart not all lengths are in the same range.

Dihedral Data for the N-acetyl section of Melatonin is shown in figure 7.1.9. The dihedral angles settle after the first 500 timesteps of minimization except for those which contain C10 and C11 atoms. They exhibit a degree of conformational flexibility and seem to take longer to converge than the remainder of the structure at around 1000 timesteps. The dihedral bonds seen in C10-C11-N2-C12 which also contain C11 and C12 are steadier after 500 timesteps possibly being more stable on that particular plane. This flexibility may be due to interactions with water or other parts of the ligand such as the O2 molecule which is double bonded to C12 or the hydrogen atoms closest to the n-acetyl section of the molecule on C8. It may also be a product of resonance in the indole section of melatonin. This same flexible data is not seen in dihedrals which contain only C10 or C11 suggesting this observation is linked to these paired atoms. Final values for all the dihedral angles can be seen in table 3.

Table 1:- A summary of all the Bond Angles Found in the Structure of Melatonin.

Bond Angles	G16 geometry optimized structure of melatonin. (basis set: MP2/6-31G*)	Crystal Melatonin	NAMD Minimized.
C9 – O1 - C7	116.36°	116.95°	123.55°
O1 – C7 - C6	124.88°	124.40°	104.14°
O1 – C7 - C8	113.68°	114.63°	129.52°
C7 – C6 – C5	121.31°	122.13°	127.35°
C6 – C5 – C1	117.51°	117.05°	112.82°
C5 – C1 - N1	130.81°	131.11°	129.27°
C5 – C1 – C2	122.01°	121.99°	122.82°
C2 – C1 – N1	107.18°	106.90°	107.89°
C1 – N1 – C4	109.35°	108.94°	105.88°
N1 – C4 – C3	109.63°	110.24°	113.84°
C1 – C2 – C3	107.43°	107.85°	108.61°
C7 – C8 – C2	118.09°	118.41°	114.55°
C8 – C2 – C3	132.89°	132.71°	129.76°
C8 – C2 – C1	119.64°	119.44°	121.58°
C2 – C3 – C10	126.73°	125.43°	121.80°
C4 – C3 – C10	126.69°	128.53°	133.87°
C4 – C3 – C2	106.40°	106.04°	103.72°
C3 – C10 – C11	111.28°	113.89°	120.02°
C10 – C11 – N2	112.99°	110.90°	124.22°
C11 – N2 - C12	120.71°	122.55°	111.33°
N2 – C12 - C13	115.52°	117.10°	115.23°
N2 – C12 - O2	122.14°	121.05°	121.40°
O2 – C12 – C13	122.26°	121.85°	119.97°

See Figure 7.0.1 - Melatonin Key for Bond Identification.

Table 2:- The Bond Lengths Between the Atoms of the Backbone Structure in Angstroms (Å) of Melatonin.

Bond Lengths.	G16 geometry optimized structure of melatonin. (basis set: MP2/6-31G*)	Crystal Melatonin (1972).	NAMD Minimized.
C9 – O1	1.42Å	1.41Å	1.41Å
O1 – C7	1.38Å	1.38Å	1.40Å
C7 – C6	1.42Å	1.40Å	1.38Å
C7 – C8	1.39Å	1.37Å	1.38Å
C6 – C5	1.38Å	1.37Å	1.39Å
C5 – C1	1.40Å	1.39Å	1.32Å
C1 – C2	1.42Å	1.40Å	1.45Å
C8 – C2	1.41Å	1.41Å	1.41Å
C1 – N1	1.38Å	1.37Å	1.36Å
N1 – C4	1.38Å	1.38Å	1.39Å
C4 – C3	1.38Å	1.35Å	1.37Å
C2 – C3	1.43Å	1.42Å	1.40Å
C3 – C10	1.49Å	1.49Å	1.49Å
C10 – C11	1.53Å	1.50Å	1.55Å
C11 – N2	1.45Å	1.45Å	1.41Å
N2 – C12	1.37Å	1.33Å	1.34Å
C12 – O2	1.23Å	1.24Å	1.23Å
C12 – C13	1.52Å	1.49Å	1.50Å

Hydrogen bond lengths are not mentioned due to the not being present in the PDB co-ordinate structure file for the 1972 crystal structure for Melatonin. See Figure 7.0.1 - Melatonin Key for atom identification.

Table 3:- Dihedral Angles present in the Ligand Melatonin.

Dihedrals.	G16 geometry optimized structure of melatonin. (basis set: MP2/6-31G*)	Crystal Melatonin.	NAMD Minimized.
C9 – O1 – C7 – C8	-0.16°	0.07°	47.92°
C9 – O1 – C7 – C6	179.71°	-179.45°	-100.07°
C2 – C3 – C10 – C11	74.83°	174.67°	75.72°
C4 – C3 – C10 – C11	-99.59°	6.33°	-114.86°
C3 – C10 – C11 – N2	-179.26°	171.76°	-124.73°
C10 – C11 – N2 – C12	76.72°	-170.64°	-90.75°
C11 – N2 – C12 – C13	178.92°	176.37°	-165.37°

C11 – N2 – C12 – O2	-4.25°	-4.09°	35.46°
O2 – C12 – C13 – HM6	-81.37°	-73.85°	138.79°

See Figure 7.0.1 - Melatonin Key for Dihedral Location Information.

Summary.

Differences are seen in the NAMD minimized version of the data generated with Gaussian16 most significantly in dihedral angles. These differences are due to the interaction with solvent since the ligand has been parameterized in a vacuum devoid of external influence, dihedrals angles are expected to change. Gaussian 16 optimizes small molecules to be used under a specific forcefield by matching the data to the forcefield. The differences between the minimized melatonin and X-ray crystal version of the same molecule will inherently have different bond lengths angles and dihedrals due to the crystallization conditions used to create the lattice. The parts of the ligand which are not as flexible fit into the same range in terms of dihedral angle, bond angle and length given the atom types. The N-acetyl section of the ligand appears to be the most flexible section of the molecule, which is why the data is varied in these atoms (table 3). Bond lengths between the crystal structure, the minimized structure and the Gaussian16 optimized versions are highly similar. The lengths vary in very minor lengths between 0.01Å and 0.03Å (table 2). The melatonin bond angles vary slightly in the Oxy-methyl section and the N-acetyl section but do not deviate significantly in the indole section (table 1). The optimized ligand is representative of Melatonin. After analysis of the NAMD minimized ligand it was docked into the previously performed Mel1a GCPR 30 nanosecond - Run 2 and saved as a new PDB and PSF file ready for minimization.
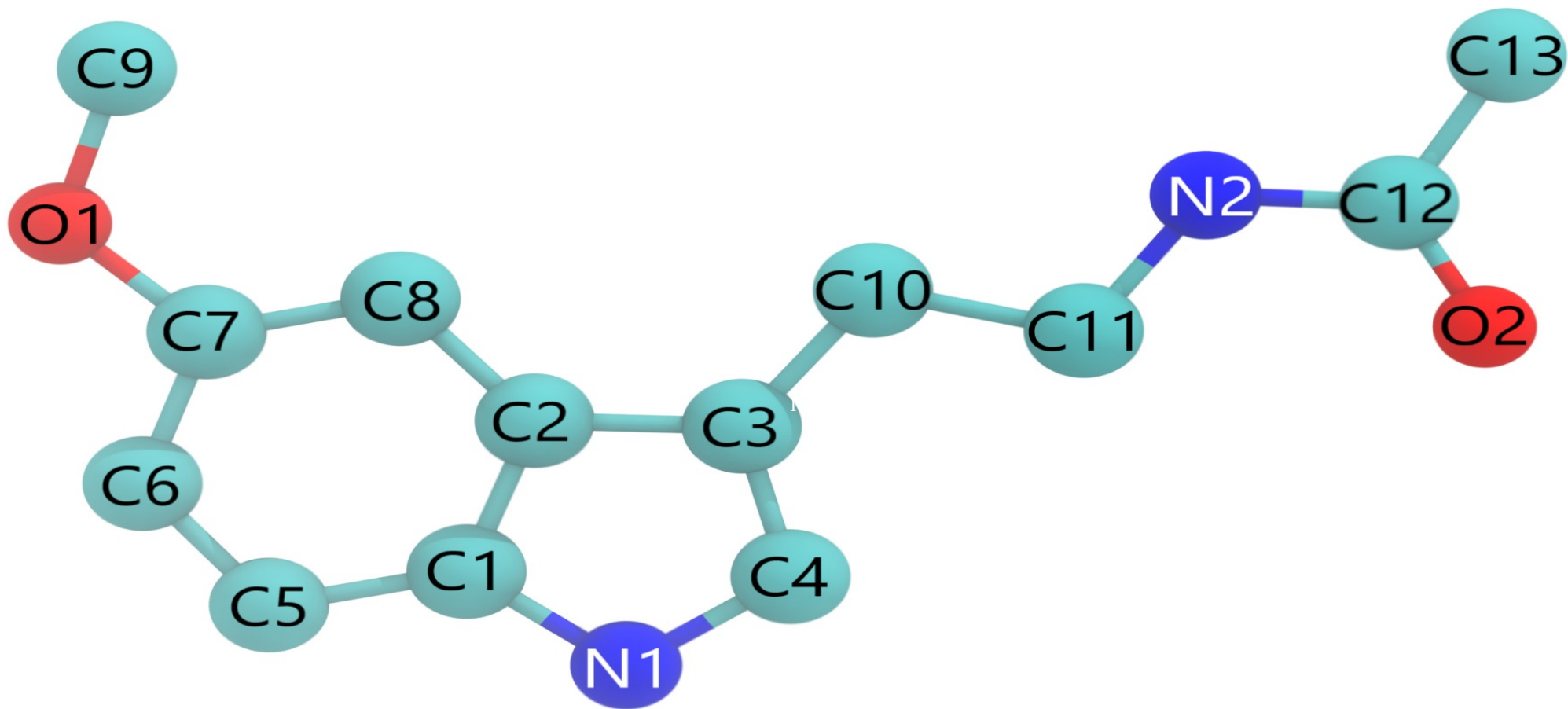
Figure 7.0.1 Reference Melatonin Atom Diagram for Bond length, Bond Angle and Dihedral analysis .
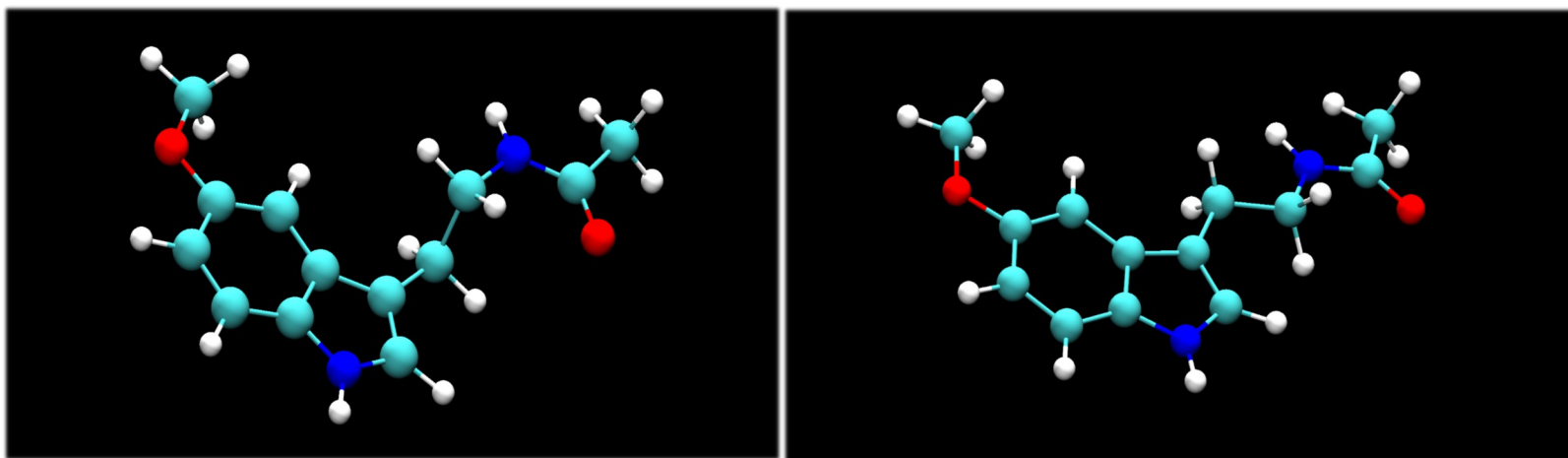
7.0.2- Gaussian16 Generated QM melatonin (Left) and 1972 X-Ray Crystal Structure of Melatonin (Right).

Bond Angles Guassian16 Melatonin Minimization (Oxymethyl Group)
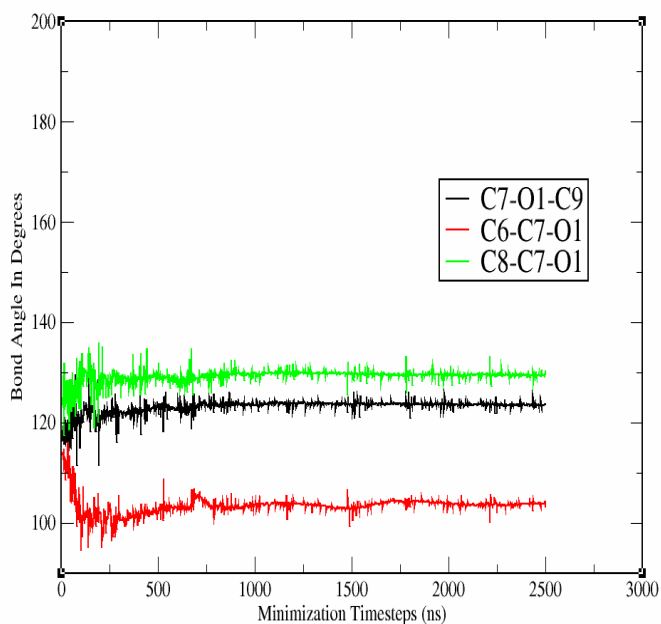
Figure 7.1.3 - Bond Angles in the Oxy-methyl section of the ligand

The final values of the simulation can be seen in *Table 1:- A summary of all the bond angles found in the structure of melatonin.*

Gaussian16 Melatonin Minimization Oxymethyl Group Bond Lengths
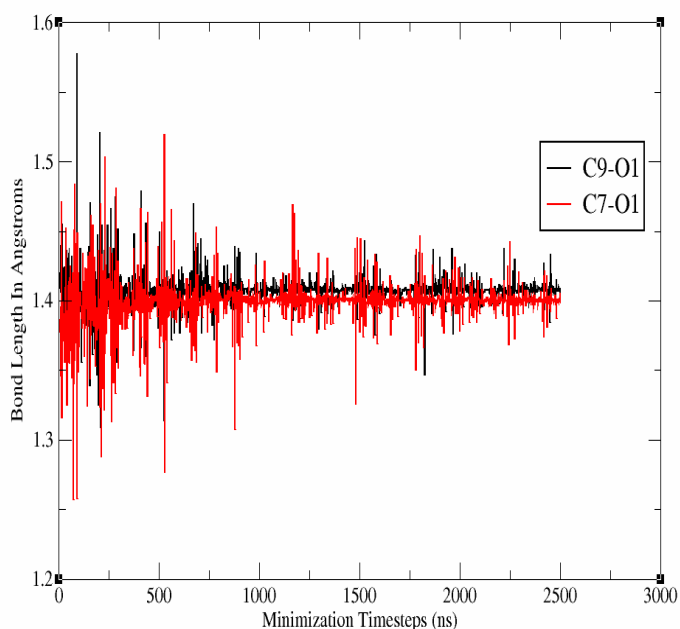
Figure 7.1.4 - The Oxy-Methyl bond lengths

The final values for the bond lengths are shown in *Table 2:- A Table listing All the Bond Lengths Between the Atoms of the Backbone Structure in Angstroms (Å).*

Oxymethyl Section Dihedrals Guassian16 Minimization Over Timesteps
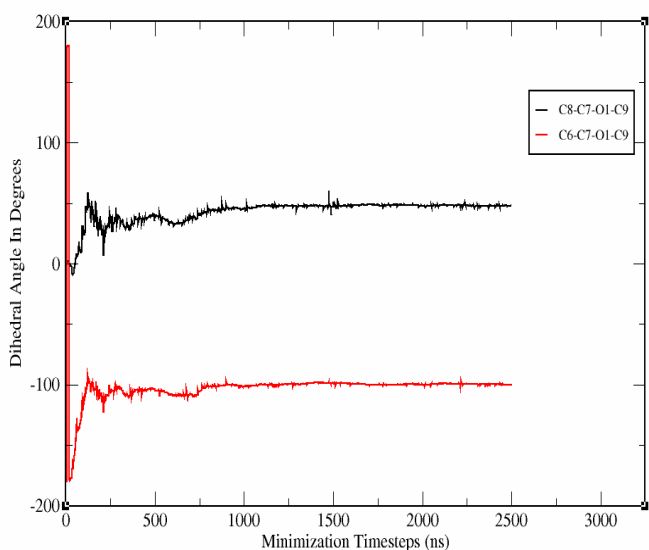
Figure 7.1.5 - Dihedral Properties of the oxymethyl section of Melatonin throughout the minimization of the ligand in a water box

The final figures are shown in *Table 3: Dihedral Analysis present in the ligand melatonin.*
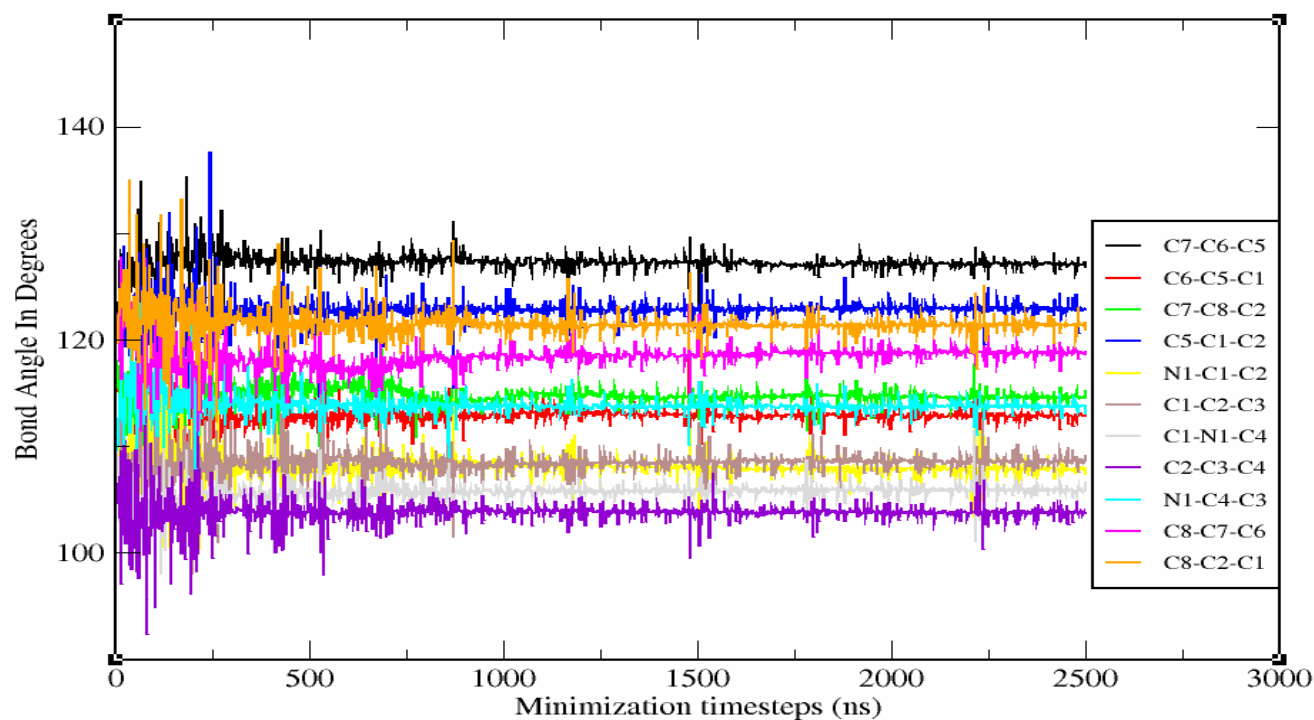
Figure 7.1.6 - Bond angles in the indole ring section of Melatonin during minimization.

Figure 7.1.7 - Bond lengths within the indole ring section of Melatonin over the timesteps of minimization.

Gaussian16 Melatonin Minimization Bond Angles Carbon Chain.
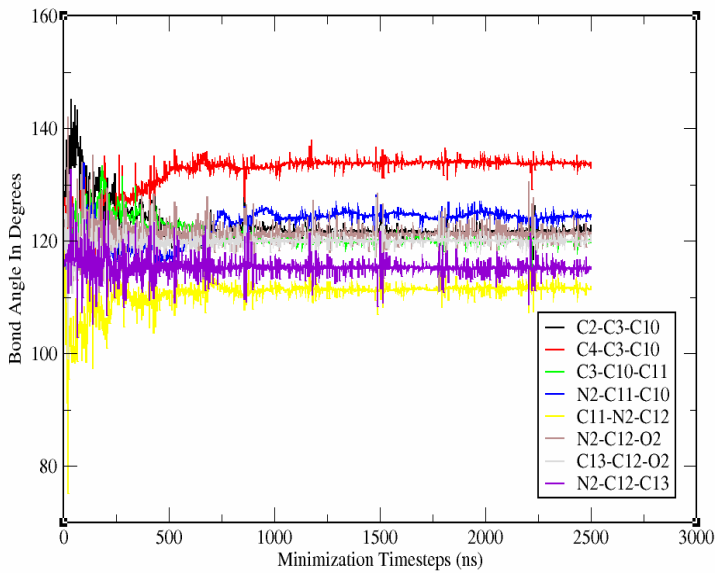
Figure 7.1.8 - Bond angle data for the N-acetyl section of Melatonin.

Gaussian16 Melatonin Minimization Carbon Chain Bond Lengths
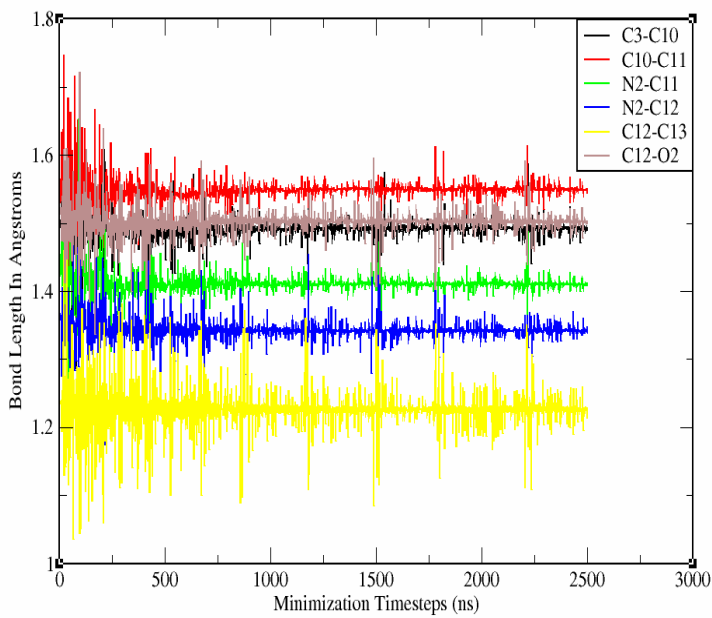
Figure 7.1.9 - Bond length data for the N-acetyl section of Melatonin.

Carbon Chain Section Dihedrals Guassian16 Minimization Over Timesteps
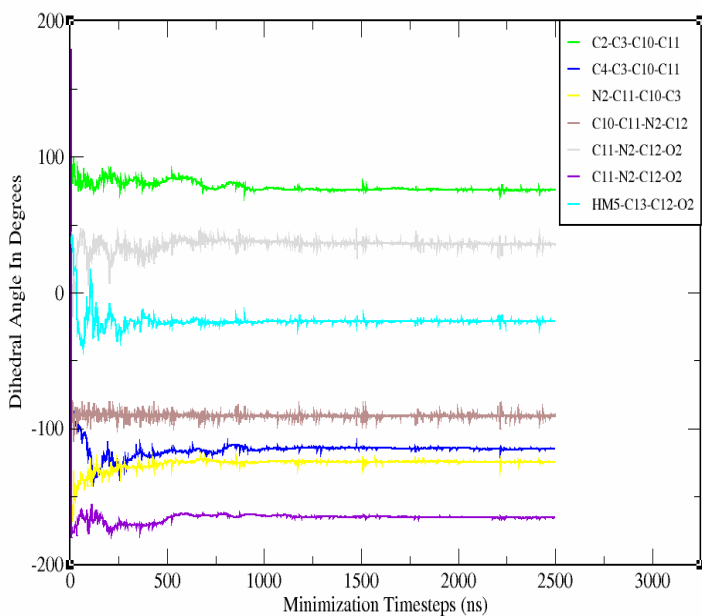
Figure 7.1.10 - Dihedral Data for the N acetyl section of Melatonin.

8.0 – Minimization and Heating of the Mel1a GPCR/Melatonin Docking System.

The melatonin molecule was docked to the final frame of the second 30ns trajectory using Chimera 1.1.2 and Autodock Vina. Since the system has changed, the Mel1a-melatonin system needs re-minimization and heating, followed by equilibration to prepare it for MD analysis. The addition of melatonin to the system allows observation of the Mel1a GPCR's conformational state, and to determine if the chosen docking co-ordinates cause a binding effect on the extracellular side of the Mel1a GPCR.

Data was loaded from the NAMD log file for the melatonin-Mel1a bound complex and analysed in VMD 1.9.3 using the NamD plot module with a stride of 1 (therefore loading the entire trajectory).

Bond angle data gives an average figure of 2,680.76 Kcal/mol over the calculation, initial bond angle energy is 10,477.06 Kcal/mol the final timestep displays a value of 2,605.28 Kcal/mol. This is recorded at such high levels due to the melatonin molecule being introduced to the CHARMM Forcefield. Dihedral energy shows an average of 6,802.65 Kcal/mol with an initial value of 8,196.43 Kcal/mol, and a final value at the end of the calculation of 6,597.27Kcal/mol. improper energies in the system give an average figure of 42.05 Kcal/mol, with an initial value of 200.6 Kcal/mol and a final value of 40.59 Kcal/mol (Figure 8.0.1) . Average bond energy in the docking system was 275.67 Kcal/mol with an initial value of 5,572.7 Kcal/mol this is also likely to be caused by the addition of melatonin, the final value was 521.94 Kcal/mol. The potential energy provided an average figure of -76,823 Kcal/mol with an initial value of -116,062.86 Kcal/mol and a final value of -119,026.58 Kcal/mol (Figure 8.0.2).

RMSD data was generated using VMD 1.9.3's built in RMSD visualizer tool. The Mel1a/Melatonin docking system (selection protein, backbone) stabilised at approximately frame 90 of the minimisation (Figure 8.0.1). An RMSD trajectory analysis was also carried out on the protein backbone using VMD's trajectory tool with the selection "protein", the data had an average RMSD of 0.239Å, a minimum of 0.057Å, maximum of 0.285Å and a standard deviation of 0.053Å. The mel1a GPCR back was again harmonically constrained, but the melatonin molecule was not. The bond angle data clearly shows an initially large figure. Which is due to the addition of melatonin and appears to be provoking a reaction from the Mel1a GPCR.

Summary.

Potential energy in the system is significantly more negative than the initial minimization of Mel1a (section 5.0.) This may be due to the addition of the melatonin molecule, present in the system showing an interaction which affects the Mel1a GPCR. Bond angle, dihedral and improper energy values are also lower in reference to the previous minimization at the end of the calculation. Protein backbone trajectory RMSD is also higher in the Mel1a-melatonin minimization than the non-melatonin minimization calculation. Local RMSD within 10Å of the Melatonin residue is an average of 0.125Å, with a minimum RMSD of 0.05Å and a maximum RMSD of 0.137Å the standard deviation is 0.015Å. this may account for the higher RMSD seen in the Mel1a GPCR. It also suggests that the melatonin molecule is in a desirable position on the extracellular side.

8.1 Heating of the MEL1a -Melatonin system.

Data was extracted from the NAMD log file and was analysed using VMD 1.9.3's NAMD plot module with a stride of 1, incorporating the entire data set for heating (steps 0 to 310,000). Bond energy has an average figure of 1,476.71 Kcal/mol with an initial value of 473.59 Kcal/mol and a final value of 2,536.52 Kcal/mol. Average potential energy of the Mel1a-melatonin docking system is -104,640.53 Kcal/mol with an initial potential energy of -118,972 Kcal/mol and a final potential energy of -86,449.55 Kcal/mol (Figure 8.1.1). The temperature data shows an initial temperature of 0 K and a final figure of 309.75 K. The average Van Der Waal interaction energy is 6,796.82 Kcal/mol, with an initial of 8,150.05 Kcal/mol, and a final calculation value of 3,859.32 Kcal/mol (Figure 8.1.2) Van Der Waal energy decreases over the heating phase as the mel1a-melatonin system is given temperature (therefore kinetic energy). This happens as the Mel1a GPCR begins to conform to the binding effect of melatonin seen in the minimization stage and VDW interaction energy. The RMSD trajectory tool (via selection of the protein backbone) showed the average RMSD of the simulation to be 0.226Å with a minimum of 0.015Å, a max of 0.370Å and a standard deviation of 0.091Å. RMSD remains low after minimization and heating due to the harmonic constraints applied to the Mel1a GPCR, and the limited time it can conformationally adjust as the kinetic energy in the system increases along with temperature.

8.2 – Equilibration of the Mel1a-Melatonin Docking System.

After heating trajectory data was analysed for equilibration of the Mel1a-melatonin system using VMD's built in RMSD trajectory tool. The average RMSD (protein backbone) was 0.852Å (min of 0.496Å, max of 0.982Å) with a standard deviation of 0.084Å. This RMSD data was colour mapped onto the model in order to visualise it directly in the context of the protein (Figure 8.2.1) The core of the protein is stable in the central transmembrane core (an average RMSD of below 1.5Å) with the majority of higher RMSD being in the loop sections. A Maxwell Boltzmann distribution of temperature was also generated using the data from the log file in order to assess the if the system had approached thermal equilibrium (Figure 8.2.2). In addition, the log file from the trajectory was analysed for kinetic energy, bond energy, Van Der Waal energy and total potential energy.

The average kinetic energy in the Mel1a GPCR-melatonin docking system was 28,380.32 Kcal/mol with an initial figure of 28,220.39 Kcal/mol and a final value of 28,347.58 Kcal/mol, which is 142.78 Kcal/mol less than the final kinetic energy reading in the rescale equilibration run in the non-melatonin Mel1a GPCR system shown in Figure 8.2.3, which does not represent a significant change with the presence of Melatonin. Potential energy in the data shows an average of -88.913 Kcal/mol, with an initial value of -86.386 Kcal/mol and a final value of -89,136.66 Kcal/mol. In comparison to the initial Mel1a equilibration figures (-86,395.28 Kcal/mol at the final frame of equilibration) potential energy became more negative in the melatonin bound state by -2,741.38 Kcal/mol (Figure 8.2.3). This is good evidence suggesting that melatonin in the proposed site causes a conformational change in the Mel1a GPCR. Average temperature is 309.97 K, with an initial value of 308.23 K and a final calculation of 309.62 K. Bond energy in the docking system shows an average of 2491.69 Kcal/mol, with an initial figure of 2476.57 Kcal/mol and a final value of 2,520.29 Kcal/mol (Figure 8.2.4). Van der Waal interaction energy shows an average over the calculation of 2,955.09 Kcal/mol, an initial figure of 3,997.90 Kcal/mol and a final calculated figure of 2,734.49 Kcal/mol (Figure 8.2.5)

Van Der Waal Interaction energy fluctuates initially followed by a downward trend towards stability at the 1.5-millionth timestep. Fluctuation in the system is likely due to the presence of the melatonin molecule causing the Mel1a GPCR to undergo conformational change due to ligand binding. This would explain the VDW energy variation followed by a region of stability. Since harmonic constraints are applied during equilibration then it is assumed that the Mel1a may not be in its ideal conformation for melatonin binding although note that the constraint is

not applied to the ligand which is free to move within the binding site during equilibration.

After equilibration analysis, the system was moved forward into molecular dynamic simulation for 30 nanoseconds.

8.3. 30 ns Molecular Dynamics of the Mel1a GPCR - Melatonin System.

Data from triplicate 30 ns trajectories were individually uploaded into VMD 1.9.3 for analysis and RMSD analysis and colour mapping of this data was generated (Figure 8.3.1). In addition, the log file was also analysed using the NAMD plot plugin to check components of the potential energy function. Ramachandran plots were also generated for the final frame of each triplicate run (Figure 8.3.2). Electrostatic surface analysis of each dataset was performed using Pymol and the APBS plugin. Additionally, structural alignments of the i-tasser homology model of mel1a, the final frame of the second 30ns trajectory, the β2 adrenergic and rhodopsin GPCR's were also performed in Pymol and These steps were performed to note any key conformational changes between the unbound Mel1a GPCR and the Mel1a bilipid system in the presence of melatonin at its proposed binding site. Finally, a cluster analysis was performed in VMD on each trajectory to visualise protein conformational clusters and close to the melatonin molecule in the binding site, ideally looking for conformational change due to melatonin.

Plots of RMSD against time (Figure 8.3.3) for all three trajectories show consistent conformational exploration with one exception in trajectory 3, which samples a comparatively different conformational space between frames 200 500. After frame 500, trajectory 3 finds a stable conformation like that observed in the other simulations. The Ramachandran plots are taken from the final frame of all three docking runs and show most of the residues in the Mel1a GPCR structure adhere to expected types of secondary structure with few outliers, which suggest some residues in the structure are in undesirable conformations or are transition residues which are intermediaries between different types of secondary structure form.

The log files from the three separate calculations of the Mel1a-melatonin system were analysed using VMD 1.9.3's NAMD plot module and data was turned into representative graphs (Figures 8.4.1, 8.4.2, 8.4.3 for run 1, 8.10.1, 8.10.2 and 8.10.3 for run 2 and 8.16.1, 8.16.2 and 8.16.3) the summarized values from the data are shown in table 8.3.1 below;

Table 8.3.1	Initial.	Average.	Final.
Kinetic energy (30ns docking run 1).	28,404.82 Kcal/mol	31,350.5 Kcal/mol	31,321.62 Kcal/mol
Kinetic energy (30ns docking run 2).	28,404.82 Kcal/mol	31,352.15 Kcal/mol	31,434.78 Kcal/mol
Kinetic energy (30ns docking run 3).	28,404.82 Kcal/mol	31,346.73 Kcal/mol	31,359.2 Kcal/mol
Potential energy (30ns docking Run 1).	-89,292.58 Kcal/mol	-86,878 Kcal/mol	-87,051.5 Kcal/mol
Potential energy (30ns docking Run 2).	-89,292.58 Kcal/mol	-86,858 Kcal/mol	-86,953.14 Kcal/mol
Potential energy (30ns docking Run 3).	-89,292.58 Kcal/mol	-86,853.86 Kcal/mol	-86,528 Kcal/mol
Bond energy (30ns docking Run 1).	2,505.7 Kcal/mol	5,509.25 Kcal/mol	5,576.95 Kcal/mol
Bond energy (30ns docking Run 2).	2,505.7 Kcal/mol	5,507.44 Kcal/mol	5,521.15 Kcal/mol
Bond energy (30ns docking Run 3).	2,505.7 Kcal/mol	5,507.95	5,548.26 Kcal/mol
Temperature in Kelvin (30ns docking Run 1)	279.8 K	308.82 K	308.55 K
Temperature in Kelvin (30ns docking Run 2)	279.8 K	308.83 K	309.65 K
Temperature in Kelvin (30ns docking Run 3)	279.8 K	308.78 K	308.9 K
Van der Waal interaction energy (30ns docking run 1)	2,494.6 Kcal/mol	2,678.68 Kcal/mol	2,597 Kcal/mol
Van der Waal interaction energy (30ns docking run 2)	2,494.6 Kcal/mol	2,671.68 Kcal/mol	2,458.05 Kcal/mol
Van der Waal interaction energy (30ns docking run 3)	2,494.6 Kcal/mol	2,671.68 Kcal/mol	2,458.05 Kcal/mol

***Table 8.3.1 – a summary of the data collected from all 30 ns docking MD runs.**

8.5 Visual Analysis of the System Run 1.

The DCD file was loaded into VMD 1.9.3, with a stride of 1 incorporating all 15,000 frames of the simulation (Figure 8.5.1). The extracellular portion of Mel1a has the same di-sulphide bond as seen in all MD runs between CYS100 on alpha helix C with CYS177 of beta turn A. Residue MET1 (the N-terminus) is in contact with the top of the extracellular helix sections until frame 6,950 out of 15,000 where it relocates into solvent for the duration of the calculation. A portion (Pro80 to Asn90) of alpha helix B near the extracellular side moves away from the ligand, altering the upper half of the helix resulting in the structure straightening, helix B was the extracellular region which resembled a pi-helix conformation in the 30ns Mel1a GPCR simulations. Alpha helix A moves towards melatonin moving closer to the core of the protein on the extracellular side. The central core of the protein remains exceptionally stable throughout the simulation (core RMSD of below 6Å). In frame 1418 (out of 15,000) SER103HG:1 sits 1.99Å from the MEL1:O2 atom, SER176 is placed 5.35Å from MEL1's O2 atom and 2.92Å away from MEL1's HN1 atom but the residue shifts between the mentioned O2 atom and the HN1 atom throughout the calculation and appears to be interacting through Van Der Waal interaction. Melatonin's N acetyl section (Figure 7.0.1, section 7.0) is seated between the oxygen atoms of CYS100 and CYS 177. Melatonin HM5 atom (bonded to atom C13) is making a Van Der Waal interaction with the CYS177 at distance of 2.47Å (final frame, distance measured from the centre of the HM5 melatonin atom to the centre of the O atom in CYS177). Melatonin atom HM6 (bonded to atom C13) is in range of the CYS100 oxygen atom with an interaction distance 2.92Å. The loop sections on the intracellular and extracellular side of the protein show less overall movement during the trajectory (compared to the Mel1a simulations at 30ns). One exception is the C-terminus end (VAL350) in loop I which is exposed to solvent. CYS314 and MET 339 are embedded in the POPC membrane with the sulphur atom directed towards the bi-lipid membrane due to being hydrophobic on both residues.

8.6 Structural Alignment Analysis Run 1.

The final frame of the trajectory for melatonin-Mel1a GPCR was saved as a PDB file and uploaded into Pymol for structural alignment (figure 8.6.1). This was to describe any conformational changes caused by the addition of melatonin to the previously calculated structures of Mel1a in the absence of melatonin (30ns trajectory 2) and any key conformational differences between the β2 adrenergic and rhodopsin crystal structures. Figure 8.6.1 Part A

shows the structural alignment of melatonin-Mel1a (shown in green) with the initial homology

model of Mel1a produced using I-Tasser (shown in orange), with an RMSD of 4.295Å. The

structures are similar in overall GPCR architecture but there is considerable variation across the

loop domains of each model. For example, some of the helix domains remain in approximately

the same position as the initial homology model with some exceptions. Alpha helix A bends in

towards helix-turn-helix G. The alpha helix B has a completely different conformation to the

homology model and is now facing away from the core of the protein after transitioning from

loop B. Alpha helix-turn-helix G is compressed into a smaller volume than in the I-tasser model

while alpha helix F in melatonin-Mel1a, has a straighter conformation. Most of the dynamic

change in the core of the melatonin-Mel1a is seen at the apex of each transmembrane helix

closest to the melatonin molecule, while the centre stays relatively stable. The loop sections, as

in all structural alignments, do not align in a recognisable pattern. Figure 8.6.1 Part B shows the

structural alignment of the 30ns Mel1a GPCR trajectory 2final frame (shown in red), with the

final frame of the melatonin-Mel1a (shown in green). The RMSD between the structures is

2.982Å which is comparatively lower than the i-Tasser homology model structural alignment in

part A. Some obvious conformational changes are visible. Alpha helix A in mel1a-melatonin

leading on from loop A is bent in an inwards direction towards the core of the protein

compared to the final frame of the 30ns Mel1a (Trajectory 2) model due to the presence of

melatonin. The loop domain in the centre of alpha helix A persists in nearly the same

conformation in both structures. Alpha helix B in the Mel1a-melatonin model also bends

outwards from the core, retaining its helical conformation, unlike helix B in the control run

(trajectory 2) which loses its 3-10 helix conformation and becomes more pi-helix in appearance

as it transitions into loop C. Loop I has a different conformation compared to the 30ns control

(trajectory 2) and loop sections sit closer to the intracellular core of the protein. Figure 8.6.1

Part C shows the structural alignment with the β2 adrenergic GPCR (shown in cyan) to Mel1a-

melatonin trajectory 1. The RMSD between the structures is 3.970Å. Unlike the previous

alignments there are many conformational differences that are observed primarily in the helical

domains. Alpha helix A has a central loop and has a completely different form compared to the

β2 adrenergic receptors equivalent helix. Helix B sits further out from the core compared to the

β2 adrenergic counterpart this is caused by the presence of melatonin in the mel1a-melatonin

docking structure. The lower half of helix-turn-helix G is much more compressed in the Z-axis.

The helix section in the extracellular loop area of the β2 adrenergic structure is not present in

Mel1a-melatonin structure. The loop domains on the intracellular side are not resolved in the

β2 adrenergic receptor crystal structure. All the alpha helices hold a different conformational

position and the RMSD is higher than when compared to the I-tasser homology model or the Mel1a, but the GPCR conformation is overall still conserved. In comparison to the 30ns control and initial I-Tasser homology structural alignments, it appears the binding of melatonin alters the conformation of the Mel1a GPCR. Figure 8.6.1 Part D shows the structural alignment of Mel1a/-melatonin trajectory 1 (shown in green) with the rhodopsin GPCR (shown in magenta). The RMSD between the two structures is 4.664Å. Nearly all the helical domains of the rhodopsin GPCR share a different conformational positioning to the melatonin-Mel1a model. Most notable are alpha helix-turn-helix G which follows a different conformational positioning to the point where they no longer are in an overlapping alignment. Alpha helix B also sits further away from the core of the protein. Helix F sits closer to the core on the intracellular side. Helix A on Mel1a has a central loop section leading to a conformational deviation in comparison to the helix on the rhodopsin structure. Intracellular loop sections conform very differently, but due to most of the highly mobile loop domains being missing from the crystal structure, it is hard to compare loop I in mel1a to the intracellular side of Rhodopsin, the extracellular side has higher similarity to rhodopsin, but is still notably different in helix A and B. the structural alignments show that the addition of melatonin elicits a conformational change in the Mel1a GPCR. while there is conformational change, the overall architecture of the receptor is still recognizable.

8.7 Electrostatic Surface Analysis Run 1.

The Mel1a GPCR was saved as PDB and uploaded into Pymol for electrostatic surface analysis. The .PDB was converted to a PQR file using the Pymol plugin for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM, with a charge pf +/- 10 (Figure 8.7.1). There is an electronegative central channel which traverses from the extracellular side to the intracellular side of the protein which is filled with water. The side profile of the protein shows that the membrane portion of the GPCR is neutral in charge while the intracellular portion is positive with small areas of electronegativity. The extracellular domain has no polarization in charge. This could mean that the binding of melatonin to the causes the extracellular side to depolarize, delivering a signal to the intracellular side, via conformational change. It is documented that alpha subunits interact with the intracellular

portion of GPCR's in order to send a signal to the cell upon conformational change of the receptor.

8.8 Cluster Analysis and Unique Hydrogen Bond Analysis of Melatonin and Extracellular Loop Sections Run 1.

The DCD file for the trajectory 1 of melatonin bound to the MEL1a GPCR was loaded into VMD with a stride of 10. Using the WMC PhysBio Clustering plugin, similar positioning of residues in the dataset are clustered visually (Figure 8.8.1). For cluster analysis the region defined was all protein residues within 10Å of melatonin (resname MEL1) as its focal point. The cutoff distance was set to 3Å and set to RMSD with a weight of 1 for the cluster analysis programme. Figure 8.8.1 Part A shows the residues that are clustered at less than a 10Å distance from the melatonin residue throughout the 1,500 frames which are loaded into VMD and the appearance shows little mobility (local average RMSD within 10Å of melatonin of 2.9Å with a stride of 10) in the GPCR due to melatonin over the 30ns period of the calculation and all 1500 frames fitted within the 3Å cutoff. In addition to this, a hydrogen bond donor and acceptor analysis was performed on all residues within 5Å of melatonin (resname Mel1). A detailed log was created showing all unique hydrogen bonds within this region. Part B of figure 8.8.1 shows the results. Unique hydrogen bond analysis for all 1,500 frames in VMD returned a 32.13% occupancy hydrogen bond between Ser103-side-OG donor interaction with acceptor MEL1-side-O2. In addition, 21.47% Occupancy was seen between donor Ser176-side-OG and acceptor MEL1-side-C12. An occupancy of 21% was also seen in hydrogen bonding between donor Ser103-side-OG and acceptor MEL1-side-C12, and a final hydrogen bond occupancy of 13.20% between Ser176-side-OG donor and MEL1-side-C5. Van Der Waal interaction analysis was also performed visually in VMD to note any orbital interaction between residues and melatonin during the docking calculation. Initial frame Interaction between Ser87's OG atom at a distance of 2.28Å from the MEL1:HN1 atom.

The data provides the impression that melatonin's binding is via VDW interaction, and that the main residues in the Mel1a GPCR responsible are Cys100 and Cys177 in the disulphide bond, Ser83 on helix B. Lastly, residues Ser103 on helix C and Ser176 on beta turn A.

8.11 Visual Analysis of the System Run 2.

The DCD file for the first 30ns calculation was loaded into VMD 1.9.3 with a stride of 1 incorporating all 15,000 frames of the simulation (figure 8.11.1). The MET1 N-terminus on the extracellular side of the protein (loop A) remains consistently localized. The disulphide bond between Cys100 (on alpha helix C) and Cys177 (on beta turn A) is present as in all the MD calculations. Little conformational movement is seen in the extracellular loop sections of the protein. The intracellular loop sections of Mel1a also remain stable and packed except for Loop I. CYS314 recedes into the membrane as helix-turn-helix G reconfirms shortening the position of conformation between Tyr285 and Cys289. The Tyr282 residue (HH atom) moves closer to the melatonin residue and seems to interact through water interaction (TIP31435) to HA1 (Hydrogen bonded to the C5 of the benzene structure of the indole in melatonin). The Asn299 to ILE309 section of helix-turn-helix G compressing towards the core of the protein seems to cause the initial loop I residues to relocate into the membrane and appears to restrict solvent interaction. Met339 however still sits within a pocket of the POPC membrane with the hydrophobic portion of the residue avoiding solvent interaction. The remaining loop sections of the extracellular side show little conformational exploration.

In comparison to the first Mel1a-melatonin trajectory dataset, the Mel1a GPCR seems more stable and less conformationally changed by the binding effect of melatonin in similar areas previously mentioned in section 8.5 (The extracellular helix portions). The largest changes are seen in Helix-turn Helix G. and the most RMSD is seen in the intracellular Loop sections of the Mel1a GPCR. The core of Mel1a is stable with an average RMSD of 3.5Å.

8.12 Structural Alignment Analysis Run 2.

The final frame of the second 30ns run of the docked melatonin and Mel1a GPCR system was removed from the bi-lipid system and saved as a fresh PDB file in VMD and uploaded into Pymol (Figure 8.12.1). Structural alignment was performed on the final frame of the protein to compare the conformation of MEL1a-melatonin with its initial homology model, the final frame of the 30ns control runs without melatonin, the β2 adrenergic and rhodopsin GPCR's. (Figure 8.12.1) Part A shows the alignment between the final frame of the 30ns Mel1a-melatonin docking run 2 (shown in green) and the initial I-Tasser homology model of Mel1a (shown in orange). the RMSD between the structures is 4.709Å. Neither the intracellular or extracellular loop sections share similarity in conformation, however the final frame of run 2 seems overall more closely conformed to the core on the extracellular side. Alpha helix A has a loop split in the middle of the helix and sits further away from the core of the protein in

comparison to the initial I-Tasser homology model. Alpha helix B also shares a divided loop section on the extracellular side prior to transitioning into loop C, which then folds into a helix section for the remaining residues after Pro80. Alpha helix F follows a similar form but moves away from the core of the protein towards the intracellular side. Alpha helix-turn-helix G is more compressed on the Z-axis and has a sharper turn which pushes the intracellular portion of the helix out from the core, as a result the end section ASN299 to ILE309 sits further from the core than in the I-Tasser homology model. Loop I progresses from the end of Helix-turn-helix G in an upwards direction towards the GPCR's helix A, before navigating towards the intracellular solvent environment. Part B of figure 8.12.1 shows the structural alignment between the final frame of the Mel1a 30ns MD run 2, and the final frame of the 30ns Mel1a-melatonin docking trajectory 2. The RMSD between the structures is 3.427Å, which is comparably smaller than the deviation between the i-Tasser homology model of Mel1a shown in part A. The extracellular portion of alpha helix A sits further up and closer to the melatonin residues location than the comparable helix of the control run. Helix A also sits further out that the control runs comparative helix from the central core of the protein. Helix B also sits further away from the core than the non-melatonin control. The extracellular loop sections resemble each other in similar conformations with minor spatial deviations. Part C of figure 8.12.1 shows the structural alignment between the β2 adrenergic GPCR (shown in cyan) and the Mel1a-melatonin trajectory data set 2 (shown in green) the RMSD between the structures is 4.241Å. Most of the helix sections do not share similar positioning. Alpha helix A sits further away from the core of the protein than the equivalent helix on the β2 adrenergic receptor. While the upper portion of helix A sits closer to Helix-Turn-Helix G. the extracellular short helix section seen in the β2 adrenergic GPCR is not seen in the Mel1a-melatonin structure. Helix B still has the loop like section on the extracellular side pointing the top of the helix towards the core, where melatonin is situated. Part D of figure 8.12.1 shows the structural alignment between Rhodopsin (shown in magenta) and the mel1a-melatonin trajectory dataset 2. The RMSD between the structures is 4.107Å. most of the helix sections between the Mel1a-melatonin structure and the rhodopsin structure follow a similar architecture. Helix A sits further from the core in the upper extracellular region after the loop which splits it on the Mel1a-melatonin structure. Helixes E and F sit closer to the extracellular side of the protein than the equivalent helixes of rhodopsin. Helix-Turn-Helix G has a bending conformation between Asn299 and Ile309 which traverses towards the core of the protein in the Mel1a-melatonin structure, whereas the equivalent structure of rhodopsin is straight. The extracellular loop sections follow very different conformation.

8.13 Electrostatic Surface Analysis Run 2.

The Mel1a GPCR was removed from the Mel1a/melatonin Docking system using VMD 1.9.3 and saved as a separate PDB in isolation before being uploaded into Pymol for electrostatic surface analysis. The .PDB was converted to a PQR file using the Pymol plugin for APBS electrostatics. Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM, with a charge of +/- 10, using the AMBER forcefield (Figure 8.13.1). The extracellular side of the Mel1a GPCR has a central crater-like pocket which is mostly neutral in charge with 3 small electronegative pockets contained within. Two of the three are only negative on the surface, while another continues into the central core of the protein. This traverses to the extracellular cavity, which expands into a very wide channel that finishes on the exterior side. The intracellular side is a mix of electropositive and neutral in charge a loop section stands out from the core of the protein and the electronegative central channel forming an almost "horseshoe" shaped section of the protein which is on the lower half of the intracellular side imagine in figure 8.13.1. The side profile is mostly neutral with weak electronegative patches. The intracellular side is mostly electropositive, and the extracellular side is mostly neutral with a mix of positive and negative zones. It appears the addition of Melatonin depolarises the extracellular side of Mel1a, while the intracellular side remains electropositive.

8.14 Cluster Analysis and Unique Hydrogen Bond Analysis of Melatonin and Extracellular Loop Sections Run 2.

The DCD file for the first Docking run of melatonin to the MEL1a GPCR was loaded into VMD with a stride of 10. Using the WMC PhysBio Clustering plugin, similar positioning of residues in the dataset are clustered visually. For cluster analysis the region defined was all protein residues within 10Å of melatonin (resname MEL1) as its focal point. The cutoff distance was set to 3Å and set to RMSD with a weight of 1 for the cluster analysis programme. Part A of Figure 8.14.1 shows the resultant cluster analysis of the 10Å area around the melatonin molecule. High stability of the surrounding structure is seen with an average RMSD of 2.346Å. Unique hydrogen bond analysis was performed on the 1,500 frames taken from the DCD file with a cutoff of 5Å from the melatonin molecule (resname Mel1) to determine occupancy of any unique hydrogen bonds between the ligand and the protein (Part B, Figure 8.14.1). A hydrogen bond occupancy of 31.13% between donor Ser176-side-OG and acceptor

MEL1-side-C12 was the highest occupancy recorded by the analysis. Residue Ser103-side-OG also had a donor interaction occupancy of 12.93% with acceptor MEL1-side-O2. Cys177-main-N had an occupancy of 11.80% with acceptor MEL1-side-O2. Ser176-side-OG has a donor interaction with MEL1-side-C5 acceptor with an occupancy of 10.33%. Van Der Waal interaction was also assessed visually using VMD. The final frame of the calculation saw a separation distance between TYR281:HB1 and MEL1:O1 of 2.42Å. Another interaction was seen in between SER87:OG and MEL1:HN1 at a distance of 2.84Å. At frame 350/1,500 a 2.73Å separation VDW interaction was seen between TYR281 and MEL1:O2. SER87:OG also showed interaction at a distance of 2.96Å from MEL1:HN1. In frame 1350, MEL1:O2 sits at 1.86Å from SER176:HG1. The initial frame of the calculation shows CYS177:O 2.68Å from MEL1:HG1, and TYR281:HB1 2.5Å from MEL1:O1.

Analysis of the proposed site around Melatonin indicates that most of the interaction happens via unique Hydrogen bond interaction, with Ser176, Ser 103, and the disulphide partner Cys177 (on beta turn A) showing occupancy between 32% and 10% in hydrogen bond analysis. Other interactions are seen via VDW analysis with the residues Tyr281 and Ser87.

8.17 Visual Analysis of the System Run 3.

The DCD file for the first docking 30ns of the Mel1a and melatonin calculation was loaded into VMD 1.9.3, with a stride of 1 incorporating all 15,000 frames of the simulation (Figure 8.17.1). Extracellular loop A is closely conformed to the core of Mel1a for the duration of the calculation but changes positioning in reference to ASN91 on loop C (first residue of loop C). VAL15 on loop A moves relative to ASN91 from 4.53Å at the beginning of the simulation to 23.51Å in the final frame. The rest of the extracellular loop sections show little deviation. A small amount of conformational change is seen in the extracellular sections of alpha helix A which move towards the melatonin ligand (residues Pro23 to Ile38). Helix-turn-helix G moves away from the melatonin molecule. Alpha helix B slightly away from melatonin. The central core of the GPCR is highly stable (under 4Å in the helix core section of Mel1a) except for the mentioned interaction with melatonin seen in the helixes. The intracellular loop G is initially close to helix E and F and progressively searches conformational space resulting in a position further away from the core of the protein. From reference residue PHE234 on helix F, ARG220 initially sits 5.84Å away, by the end of the calculation it is separated from PHE234 by 17.39Å. Loop I is not very mobile and seems to pack to the central core of the protein (packed to helix

turn helix G and alpha helixes A,B and D. MET339 is still packed into the POPC membrane with the sulphur atom avoiding solvent.

The third trajectory analysis reveals the protein is more stable due to the addition of Melatonin in the proposed active site for Mel1a. the core remains very stable (4Å in the helix core) and little RMSD in the loop sections, especially Loop I, which in comparison to the previous runs is much less mobile, and seems to stay close to the intracellular portion of Mel1a. Loop G shows the largest Activity on the intracellular side. This may be due to conformation adjustment that is initiating a signalling pose to the interior of the cell membrane.

8.18 Structural Alignment Analysis Run 3.

The final frame of the third 30ns run of the docked melatonin and Mel1a GPCR system was removed from the bi-lipid system and saved as a fresh PDB file in VMD and uploaded into Pymol (Figure 8.18.1). Structural alignment was performed on the final frame of the protein to compare the conformation of MEL1a-melatonin with its initial homology model, the final frame of the 30ns control runs without melatonin, the β2 adrenergic and rhodopsin GPCR's. Part A in figure 8.18.1 shows the structural alignment of the Mel1a/melatonin docking system (shown in green), with the initial I-Tasser generated homology model (shown in orange). The RMSD between the aligned structures is 4.198Å. Helix A of the docking system only slightly overlays with the counterpart helix on the homology model and maintains the central loop not seen in the i-tasser generated structure. The first half of the helix after loop A bends due to the loop section further towards helix-turn-helix G. Helix-turn-helix G also has a different orientation and the ASN299 to ILE309 portion of the helix after the final turn is pointed inwards towards the core. Helix F sits higher towards the extracellular side of the GPCR when compared to the homology model. The entire helix is located further outwards away from the core but maintains its form. Helix E sits more towards the intracellular side of the membrane maintaining its 3-10 helix form. Helix B's top half before transitioning into Loop C loses helix form and separates into a loop shape before regaining shape for one whole turn prior to helix C. the loop I domain on the intracellular side (loops A, C, E, F and H) are more packed towards the central core in comparison to the homology model. Loop I's residues also appear closer to the membrane in positioning and 'higher' than some of the core helix sections. Loop G is further away from the core of the GPCR than seen in the homology model. Figure 8.18.1 Part B shows the structural alignment of the final frames of 30ns Control Mel1a Run 2 (red) with the final frame of the 30ns docking Mel1a/melatonin run 2 (green). The RMSD between the two structures of the Mel1a GPCR is 2.674Å. The similarity between the aligned structures is more

apparent in part B than aligned with the initial I-tasser homology model of Mel1a in part A. the extracellular loop sections share a similar conformation and are close to the core except for Loop A which sits further away from the core in both the 30ns Mel1a control and 30ns Mel1a/melatonin docking GPCR structures. Helix A and helix B in the control sit further away from the extracellular side than the comparatively seen helixes in the docking GPCR. the central loop seen in the helix is seen in both the non-melatonin and Mel1a-melatonin models. Helix B overlays closely with the early half of the helix after transitioning from loop B but deviates in the upper half of the helix as it turns into a loop (residues Met86 to Phe89). The upper half of the docking GPCR's helix B sits further away from the core than the comparative helix B on the control model. This is likely due to influence from the melatonin molecule which has been placed near helix B/C in the docking simulation. Helix-turn-helix G sits lower overall than the comparative control model, and the final helix after the turn (ASN299 – ILE309) sits at a slightly lower angle than the controls equivalent residues. Helixes E and F share the same form as the equivalent ones on the control with slight deviation in the extracellular portions of both helixes. The extracellular side of Helix C also sits closer to the core than the control run. Figure 8.18.1 Part C shows the structural alignment between the Mel1a melatonin docked GPCR and the β2 adrenergic GPCR, the RMSD between the structures is 4.191Å. Loops on the Extracellular side of the protein do not show similarity and all the extracellular alpha helix sections follow a different conformation. Alpha helix F folds in towards the core of the extracellular side while helixes G and A are further out from the core. Part D shows the structural alignment between the Mel1a-melatonin trajectory 3 structure and the Rhodopsin GPCR. Helix-turn-helix G has a different orientation to the equivalent helix on rhodopsin's structure. The ASN299 to ILE309 section faces in towards the core of the protein whereas the equivalent section of rhodopsin does not. Alpha helix E on the intracellular side sits further away from the core of the protein. alpha helix F sits closer to the core of Mel1a, and further upwards towards the extracellular side of the structure. Loop sections are missing from the intracellular side of Rhodopsin due to the loops being highly mobile and difficult to crystallize, little comparison is possible due to this. But loop G is longer on the Mel1a docking model than its equivalent loop on the Rhodopsin GPCR and reaches further to the intracellular side of the protein.

The structural alignments show that the Mel1a-melatonin GPCR has deviated significantly from the 2 adrenergic and rhodopsin GPCR's due to the melatonin which is docked to Mel1a.

which has gone through conformational change while maintaining the general architecture of a GPCR.

8.19 Electrostatic Surface Analysis Run 3.

The Mel1a GPCR was removed from the third iteration of the Mel1a/melatonin Docking system using VMD 1.9.3 and saved as a separate PDB in isolation before being uploaded into Pymol for electrostatic analysis. The .PDB was converted to a PQR file using the Pymol plugin for APBS electrostatics (Figure 8.19.1) . Grid spacing was set to 0.5, electrostatics were rendered as a Connolly surface using the Non-Linear Poisson-Boltzmann equation, Utilising a protein dielectric of 2.0 and a solvent dielectric of 78.0. At a Temperature 310K using an ion concentration of 150mM, with a charge of +/- 10, using the AMBER forcefield. Red areas of the surface map indicate electronegative portions of the protein, while blue areas represent positive areas of the protein. White zones represent areas of the GPCR which are neutral in charge. The extracellular side of the protein is a mixture of electro positive, negative and neutral areas with no obvious cavities and has no polar appearance. The intracellular side of the protein is mostly electropositive and has a small central cavity which is deep within the centre of the extracellular side. In the side profile of the protein an electronegative cavity is seen which runs towards the intracellular portion of the protein. This leads to the intracellular cavity seen in the centre of the intracellular side. The side slice also shows the central core does not traverse from the extracellular to intracellular side of the GPCR in this final frame of docking run 3.

The electrostatic surface map of the Mel1a-melatonin trajectory dataset 3 shows that the extracellular side has become de-polarized due melatonin binding to Mel1a, the intracellular side remains positive which implying that the binding action of melatonin sends a signal to the interior of the cell wall for appropriation by other molecules to continue the signal.

8.20 Cluster Analysis and Unique Hydrogen Bond Analysis of Melatonin and Extracellular Loop Sections Run 3.

The DCD file for the third trajectory of Mel1a-melatonin was loaded into VMD with a stride of 10. Using the WMC PhysBio Clustering plugin, similar positioning of residues in the dataset

are clustered visually. For cluster analysis the region defined was all protein residues within 10Å of melatonin (resname MEL1) as its focal point. The cutoff distance was set to 3Å and set to RMSD with a weight of 1 for the cluster analysis programme. Part A of Figure 8.20.1 shows the resultant cluster analysis of the 10Å area around the melatonin molecule (Local average RMSD of 2.193Å). Similarly, to 30ns docking runs 1 and 2, Little deviation is seen in the nearby sections of Mel1a (Part A). In addition to the cluster analysis unique hydrogen bond analysis was undertaken on the 1,500 frames loaded into VMD (Figure 8.20.1, Part B). The highest hydrogen bond occupancy of 30.87% was seen between donor SER103-side-OG and acceptor MEL1-side-O2. An occupancy of 21.67% was seen between donor CYS177-main-N and acceptor MEL1-side-O2. SER176-side-OG had A donor occupancy of 18.73% with MEL1-side-C12. SER176-side-OG also had a hydrogen bond with MEL1-side-C5 with an occupancy of 11.07%. VDW interaction was also visually analysed for the same stride of 10 on the DCD file for the docking calculation. Contact is seen between TYR281:HB1 at a distance of 2.51Å from MEL1:O1 in the initial frame, other VDW interactions are between SER103 and MEL1:O2 at 1.95Å and SER87:OG at a distance of 2.41Å from MEL1:HN1 in the initial frame, by frame 430/1,500, TYR281:HB1 has a VDW interaction distance of 1.65Å from MEL1:O1, SER103:HG3 interacts with MEL1:O2 at a distance of 1.90Å. In Frame 1049/1,500, SER103:HG3 is seen 1.65Å from MEL1:O2 and TYR281:HB1 is placed 2.49Å from MEL1:O1. Frame 1105 sees SER176 2.42Å from MEL1:O2. The final frame of the calculation shows SER176 2.42Å from MEL1:O2.

Summary.

The proposed docking site for melatonin seems to be mostly achieved by VDW interaction, with small amounts of hydrogen bond occupancy. The local RMSD around melatonin is an average 2.193Å showing the site to be very stable throughout the 1500 frames. The residues involved are similar in all the trajectory datasets for the Mel1a-melatonin system promoting that Ser176, Cys177, Tyr281, and Ser103 are most prominently involved in the conformational uptake of melatonin.
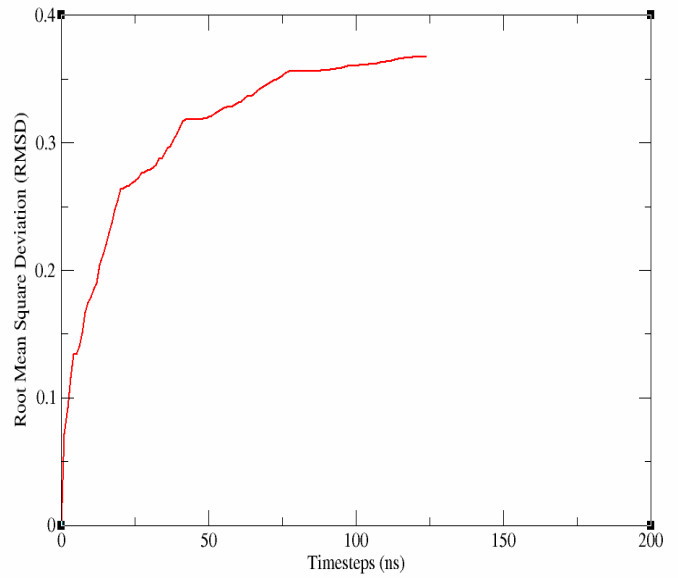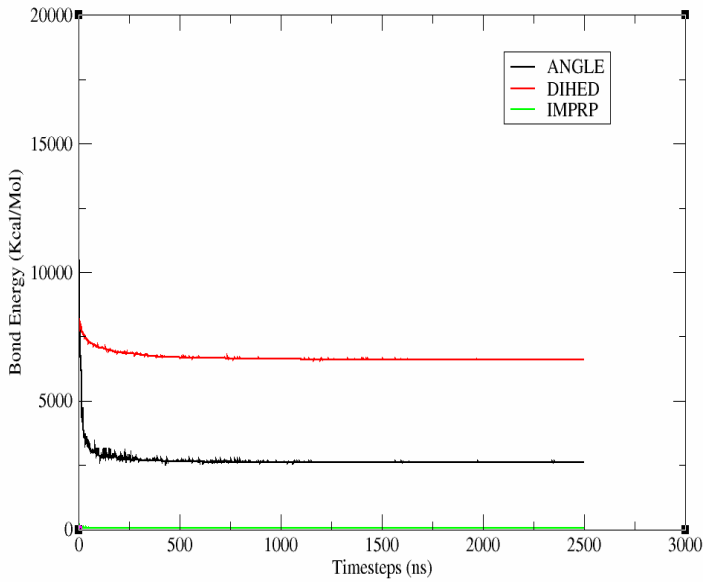
Figure 8.0.1 - Bond angle, dihedral and improper data for the Mel1a/Melatonin Docking system (Left) RMSD plot of selection "protein" in the RMSD visualiser tool in VMD 1.9.3 (Right)

Bond Energy over timesteps Docking Minimization by Conjugant Gradient   Docking System Minimization by Conjugant Gradient Potential Energy
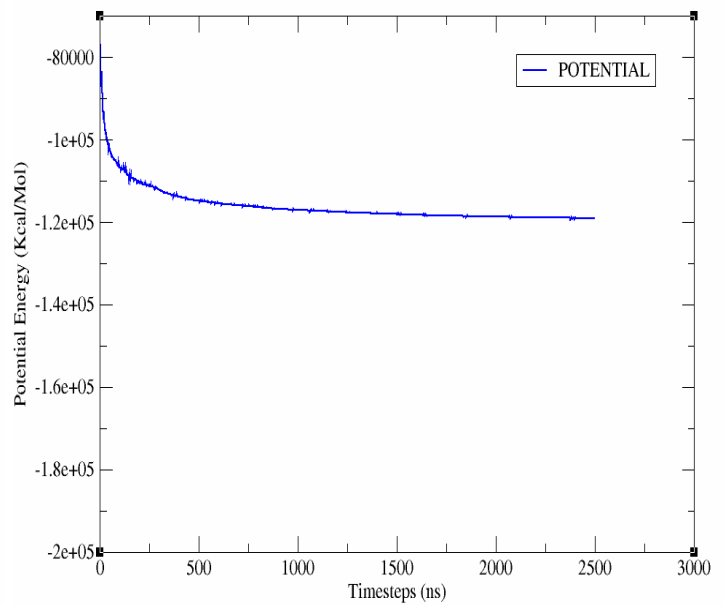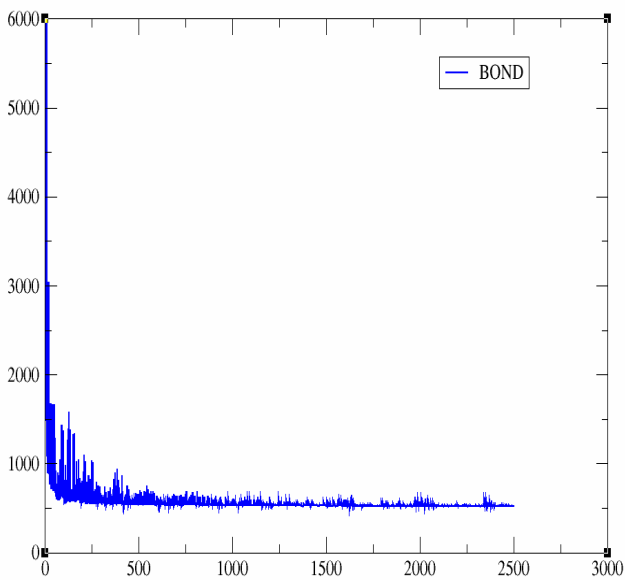
Figure 8.0.2 - Bond energy of the Mel1a/Melatonin docking system (Left). Potential energy of the previously mentioned system (Right)

Bond Energy During Docking Heating Steps
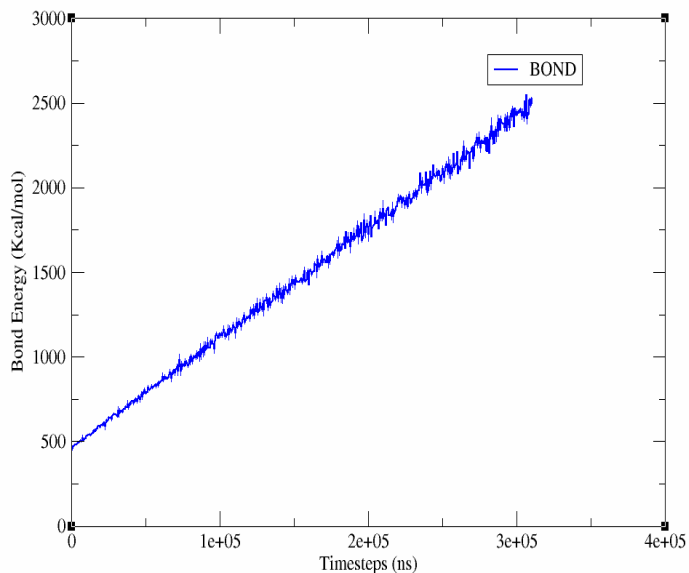
Potential Energy Over Timesteps Heating of Docking System
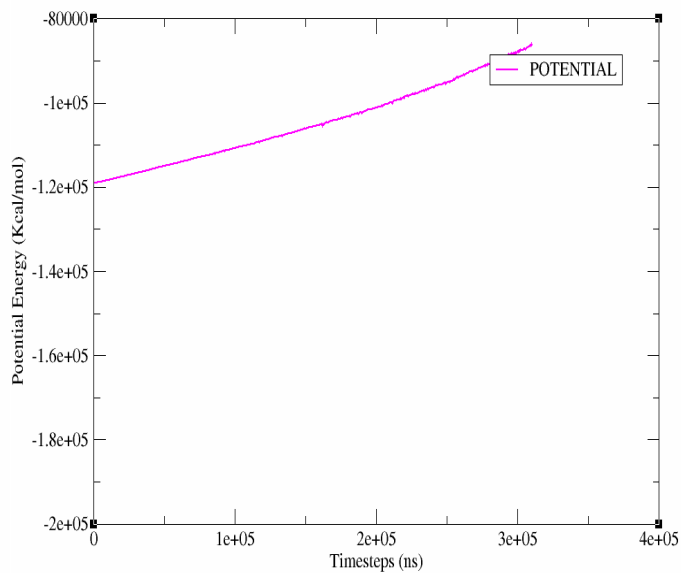
Figure 8.1.1– Bond energy of the Mel1a/Melatonin Docking system during calculation (Left). Potential energy plot for the previously mentioned system throughout the heating phase (Right).

Heating Of Docking System Over Timesteps
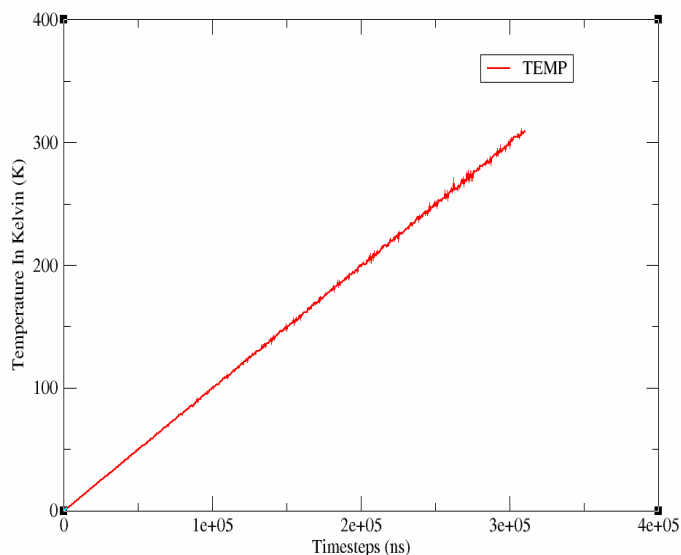
Van der Waal Interaction Energy During Heating Of Docking System
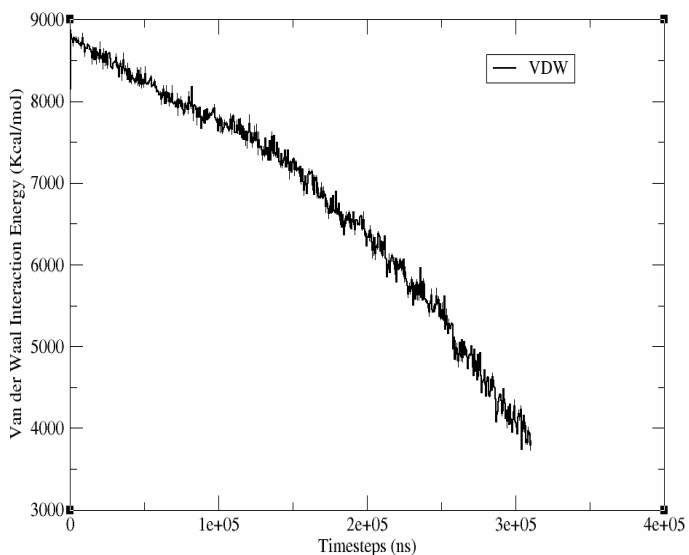
Figure 8.1.2 - The temperature in Kelvin of the Mel1a/Melatonin Docking System (left) Van Der Waal interaction Data for the previously mentioned docking system (Right)
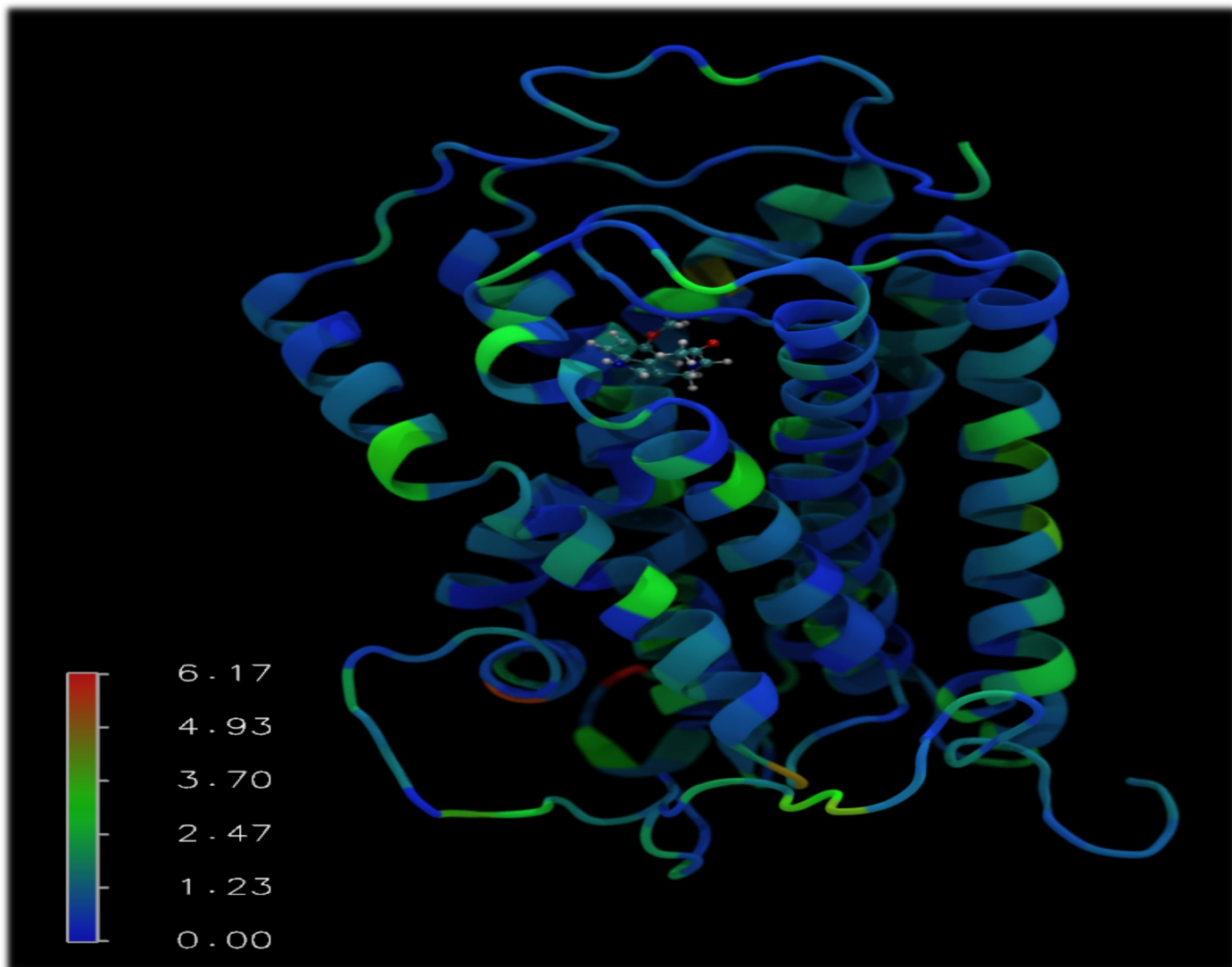
Figure 8.2.1- RMSD overlay of Mel1a with docked QM Melatonin. Coloured in RBG with red being high RMSD, blue being low and medium being intermediary the system exhibits a maximum RMSD of 6.17Å. Protein is shown VMD using the trajectory data from the calculation located in the dcd file Representation style : New Cartoon.

Docking Equilibration Temperature Distribution
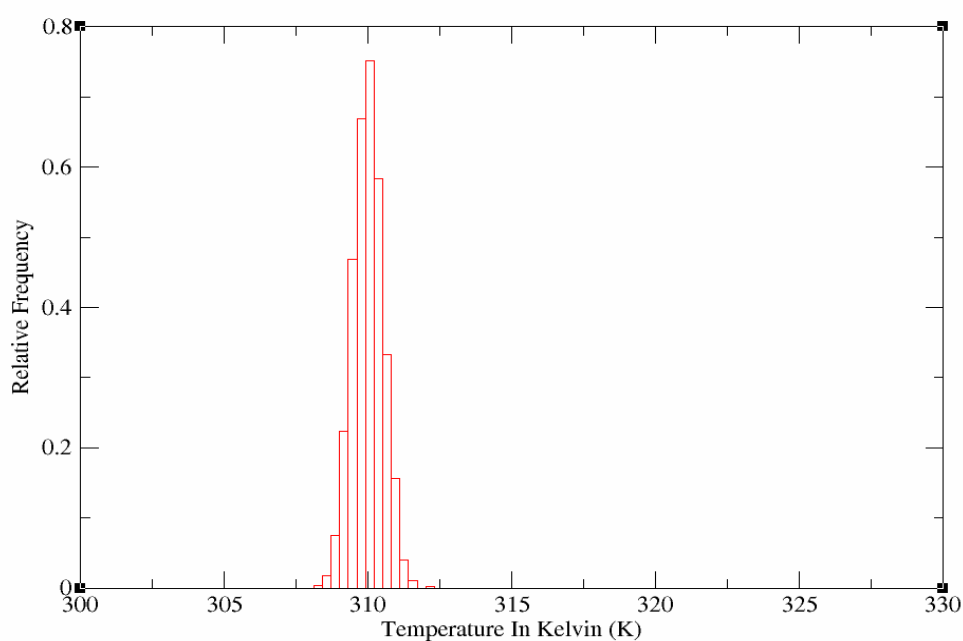
Figure 8.2.2 - Heat distribution in the docking system. Y-axis: Relative frequency describes the portion of system at a specific temperature, X-Axis shows the temperature in Kelvin (K).

Figure 8.2.3 - Kinetic energy in the entire Mel1a GCPR/melatonin docking system. (left) Potential energy in the previously mentioned system (right).

Figure 8.2.4 - The temperature of the Mel1a/melatonin GCPR simulation (left) Bond energy of the previously mentioned system. (right)

Figure 8.2.5 - Van Der Waal interaction energy (left)

Mel1a With Docked Melatonin 30ns Run 1 Mel1a GCPR RMSD Trajectory Overlay.
Red resembles high areas of movement Max 33.14Å, blue resembles low areas of movement (min of 0.00Å).
The Mel1a GCPR is shown in Using the trajectory data from the DCD file and displayed in New Cartoon.
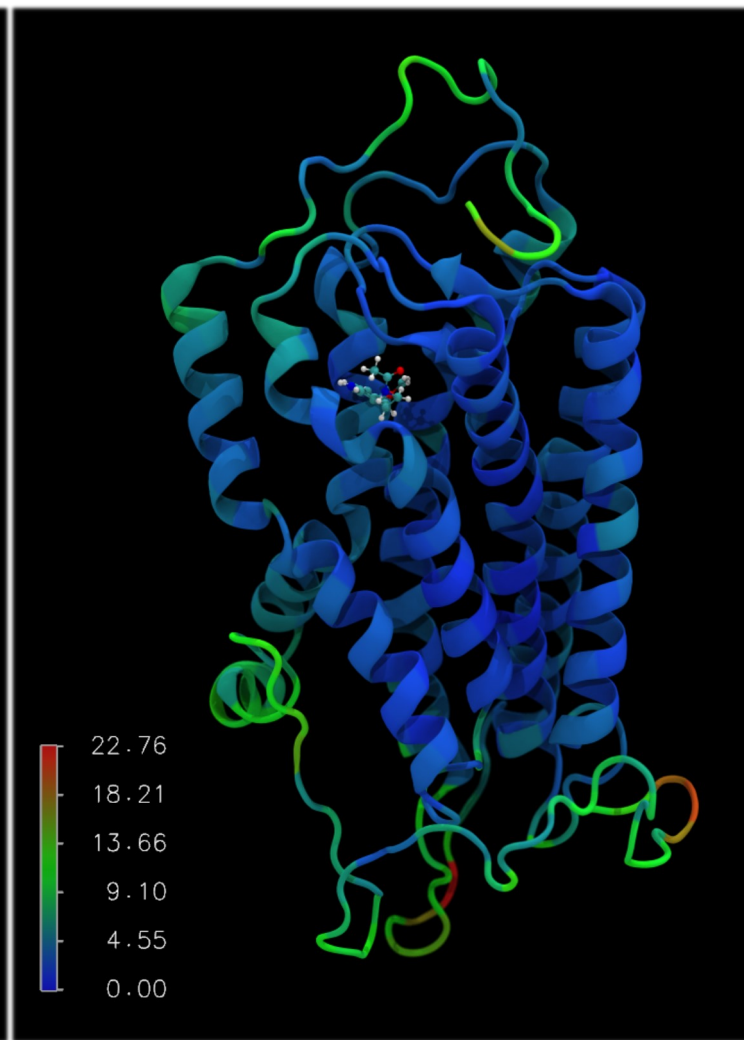
Mel1a With Docked Melatonin 30ns Run 2 Mel1a GCPR RMSD Trajectory Overlay.
Red resembles high areas of movement Max 33.14Å, blue resembles low areas of movement (min of 0.00Å).
The Mel1a GCPR is shown in Using the trajectory data from the DCD file and displayed in New Cartoon.
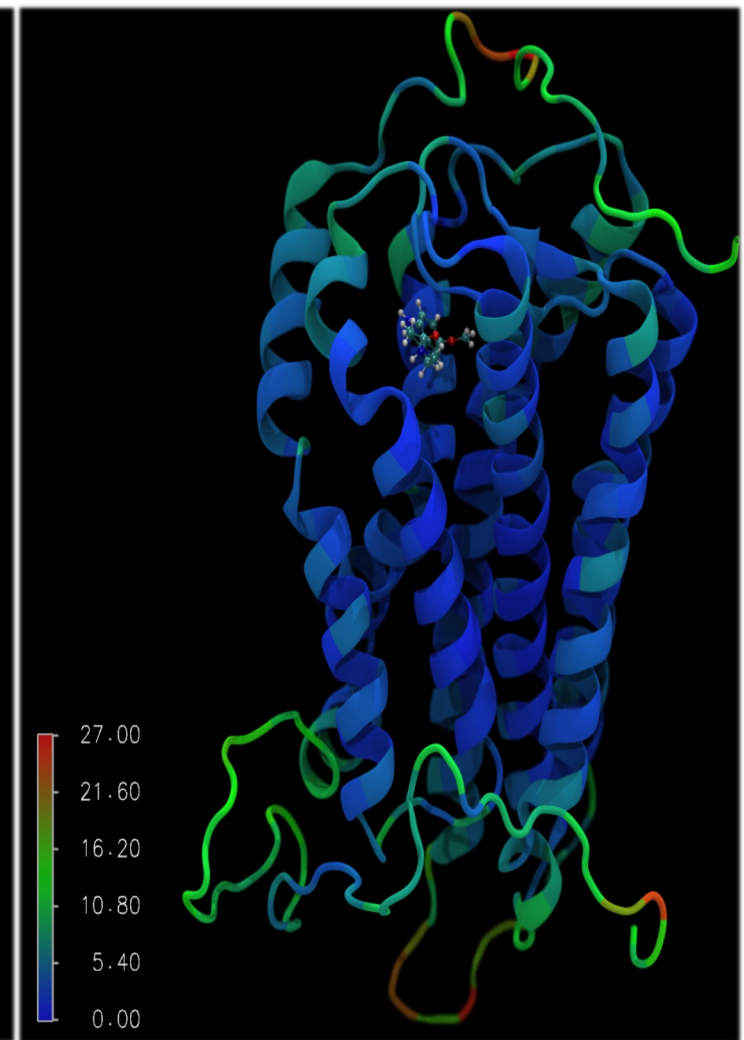
Mel1a With Docked Melatonin 30ns Run 3 Mel1a GCPR RMSD Trajectory Overlay.
Red resembles high areas of movement Max 33.14Å, blue resembles low areas of movement (min of 0.00Å).
The Mel1a GCPR is shown in Using the trajectory data from the DCD file and displayed in New Cartoon.

8.3.1– trajectory by colour (RMSD) for triplicate 30ns Mel1a/Melatonin Docking MD runs 1, 2 and 3.

30 Mel1a Docking Run 1 Ramachandran (Final Frame).
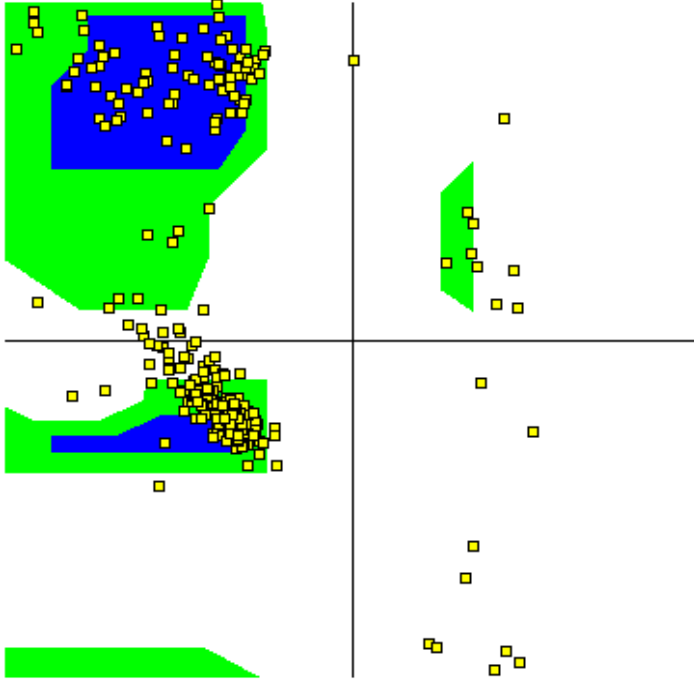
30 Mel1a Docking Run 2 Ramachandran (Final Frame).
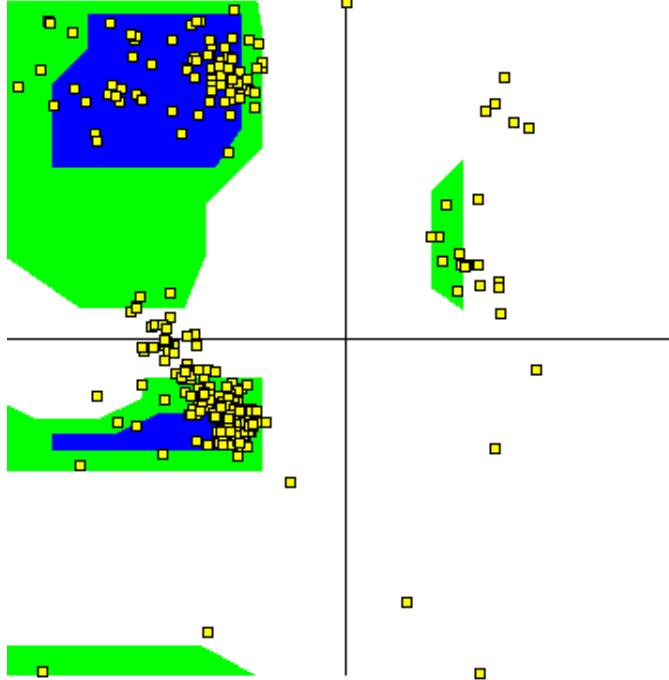
30 Mel1a Docking Run 2 Ramachandran (Final Frame).
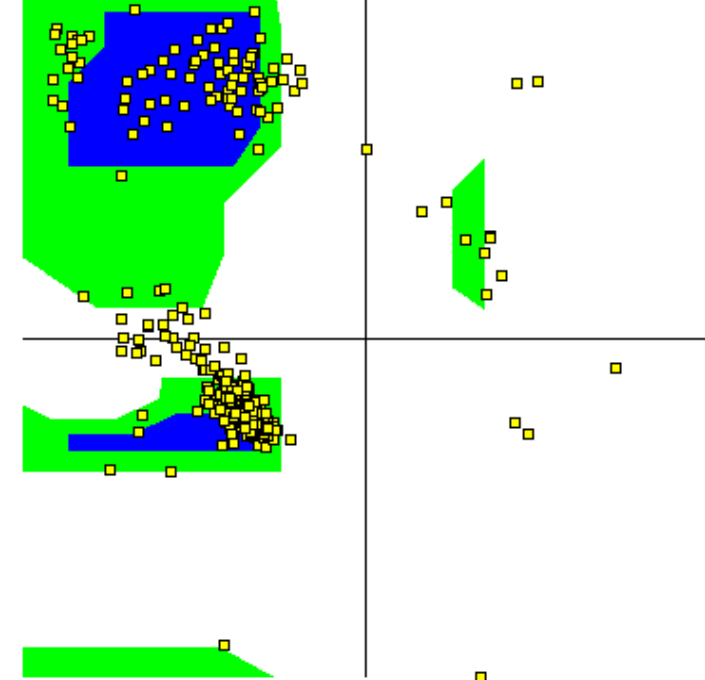
Docking MD All Runs RMSD Over 30 Nanoseconds.
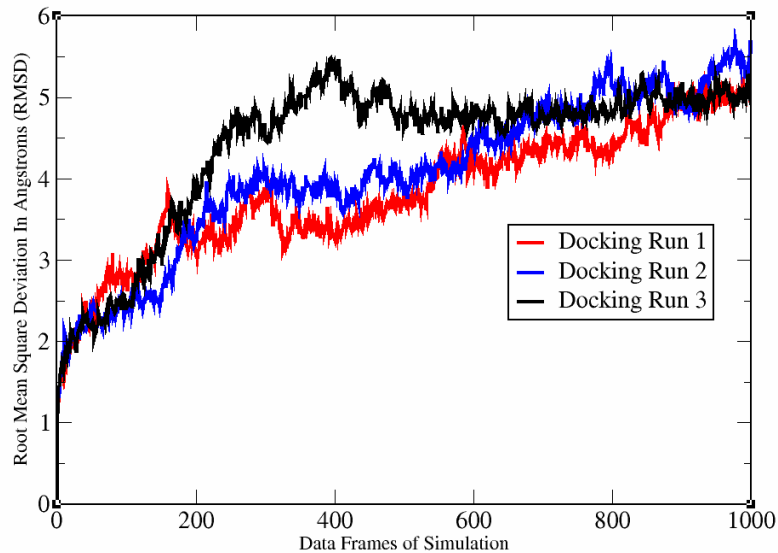
RMSD trajectory plot for all three docking runs (left) show consistent conformational exploration with one exception in run3, which samples a comparatively different conformational trajectory.

Figure 8.3.2 - RMSD trajectory and Ramachandran plots for the final frames of MEL1a, 30ns Docking runs 1,2 and 3.
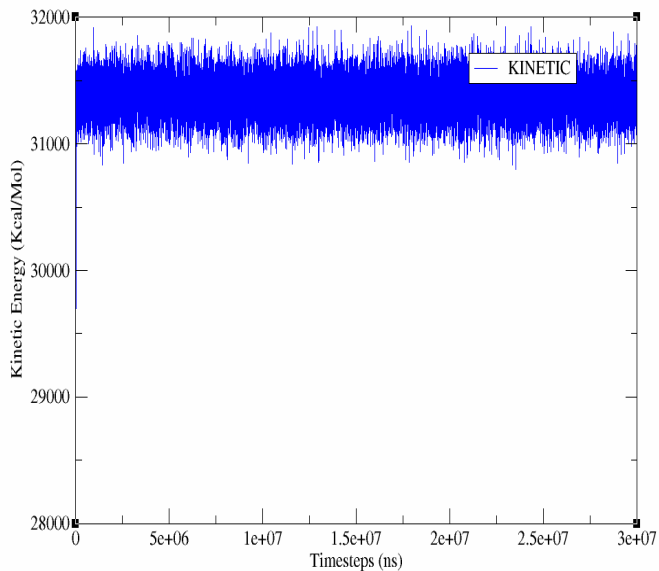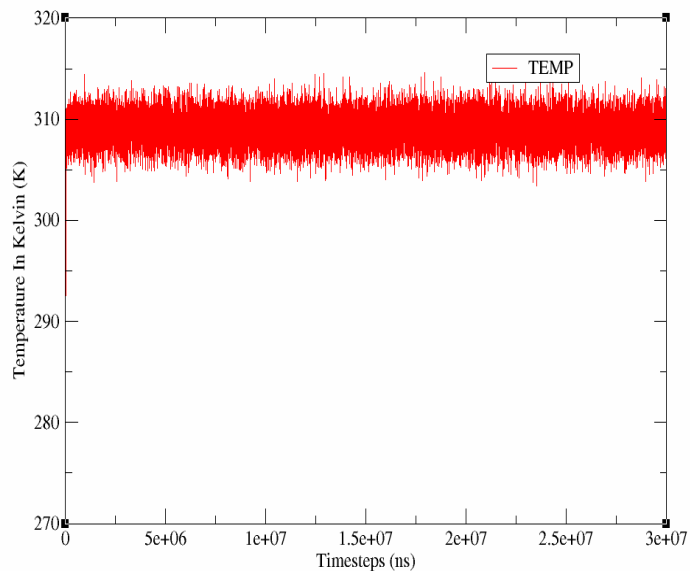
Figure 8.4.1— Kinetic energy in the Mel1a and melatonin docking system 30ns Run 1 (left) Temperature of the system in kelvin (right).
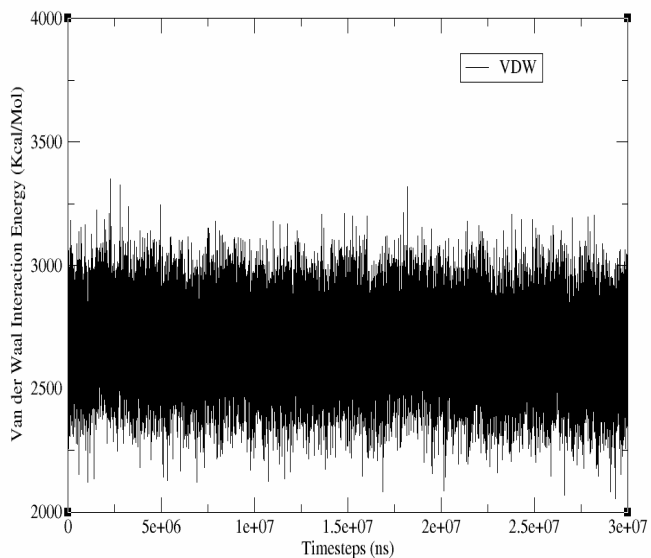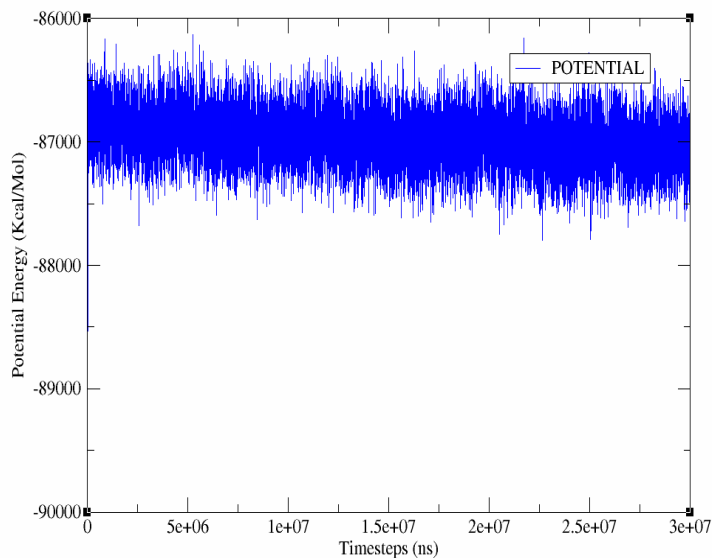
Figure 8.4.2 - Van Der Waal interaction energy plot of the previously mentioned system (left) Potential Energy Plot of the same system (Right).

Figure 8.4.3 - Bond energy in the previously mentioned system..
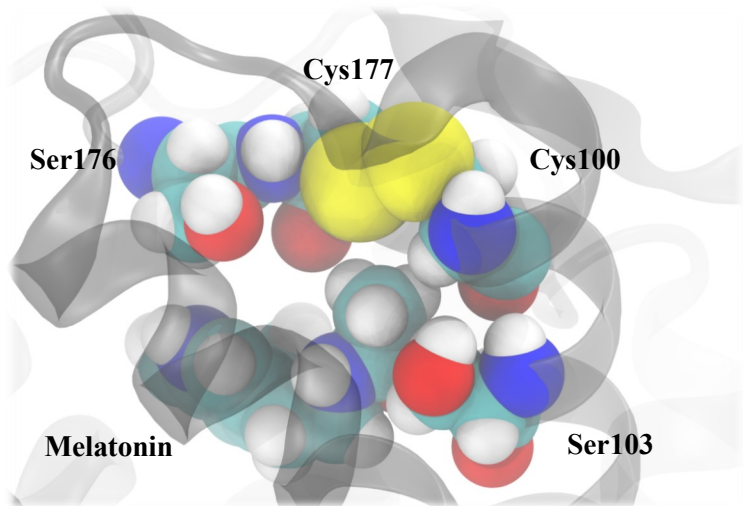
Ser103, Cys100 on helix C, Cys177 and Ser176 in VDW range of Melatonin in the final frame of the calculation.
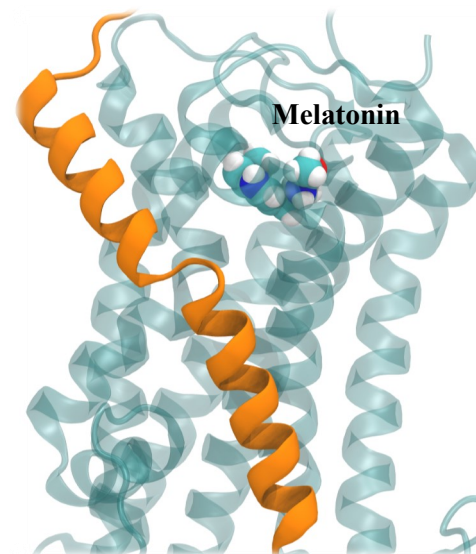
Helix A at the beginning of the calculation (left) and helix A at the end of the calculation (right). The extracellular region has conformed towards the Melatonin residue.
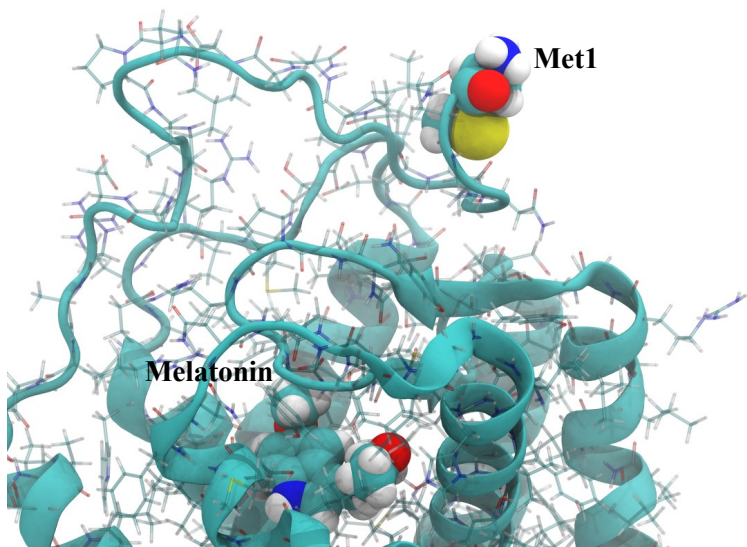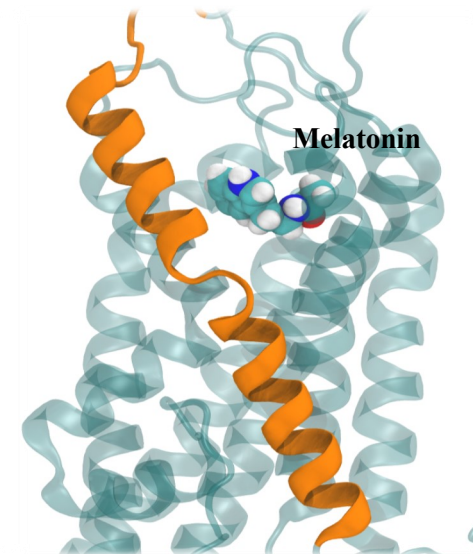
Met1 at frame 6,950 where it moves away from the protein core into the solvent on the extracellular side.
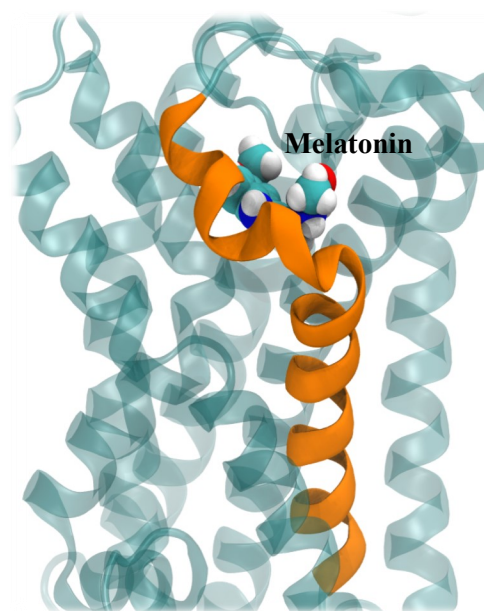
Pi-helix B at the beginning of the calculation conforms close to the melatonin ligand (left), as the simulation reaches the final frame, pi-helix B conforms away from the ligand but retains its pi-helix shape (right).
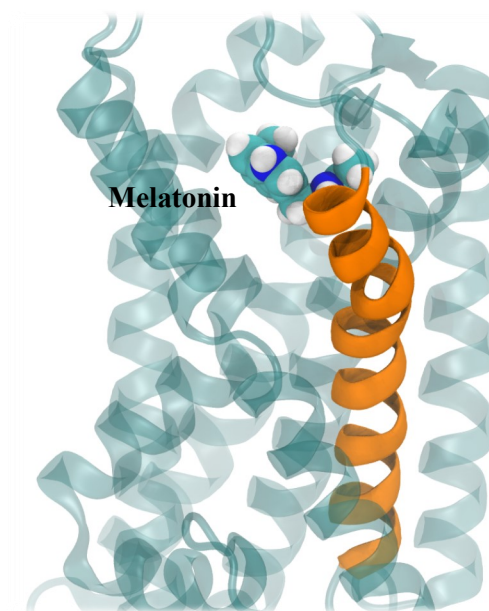
Figure 8.5.1-Visual Analysis of the 30ns Mel1a GCPR-melatonin trajectory 1.

Structural Alignment Between the I-Tasser Homology Model of Mel1a and the Final Frame of Mel1a/Melatonin Docking 30ns Run 1.

Structural Alignment Between the Mel1a 30ns Control run 2 and the Final Frame of Mel1a/Melatonin Docking 30ns Run 1.

Structural Alignment Between X-ray crystal structure of the β2 adrenergic GCPR and the Final Frame of Mel1a/Melatonin Docking 30ns Run 1.

Structural Alignment Between the x-ray crystal structure of Rhodopsin and the Final Frame of Mel1a/Melatonin Docking 30ns Run 1.

Part A. - the 30ns Mel1a/melatonin docking structure (green) aligned with the initial homology model of Mel1a (orange). The RMSD between the structures is 4.295Å.

Part B. - The 30ns Mel1a/Melatonin docking structure (green) aligned with the final frame of the control 30ns final frame of Run 2 (red). the RMSD between the structures is 2.982Å

Part C. - The 30ns Mel1a/Melatonin docking structure (green) aligned with the X-ray crystal structure of the β2 adrenergic GCPR (Cyan). The RMSD between the structures is 3.970Å

Part D. - The 30ns Mel1a/Melatonin docking structure (green) aligned with the X-ray crystal structure of Rhodopsin (magenta). The RMSD between the structures is 4.664Å.

Figure 8.6.1
- Structural alignment of the 30ns Docking run 1 with Mel1a and melatonin against the initial homology and 30ns control (Run 2) structure of Mel1a, and the X-ray crystal structures of the β2 adrenergic and Rhodopsin GCPR's.

Extracellular View and Cavity.

Electronegative extracellular pocket on the outermost part of the protein.

Intracellular View and Cavity.

Intracellular pocket on the internal part of the protein. On the inside of the cell wall.
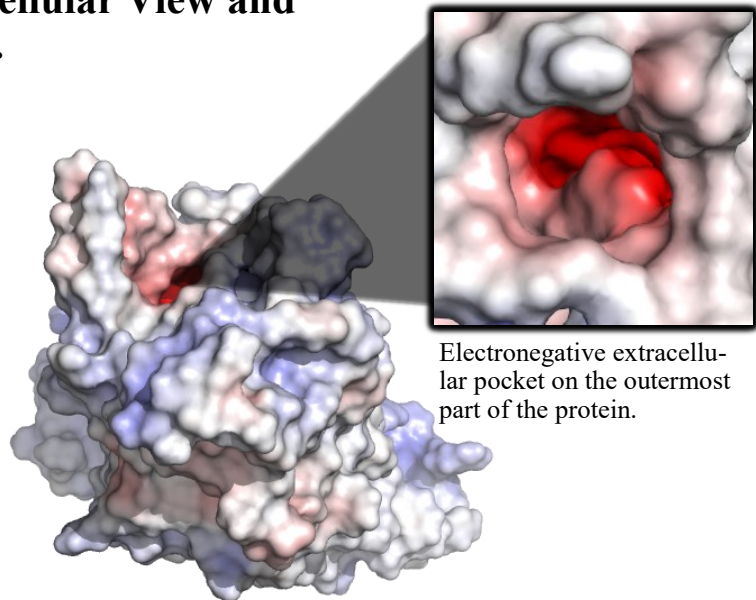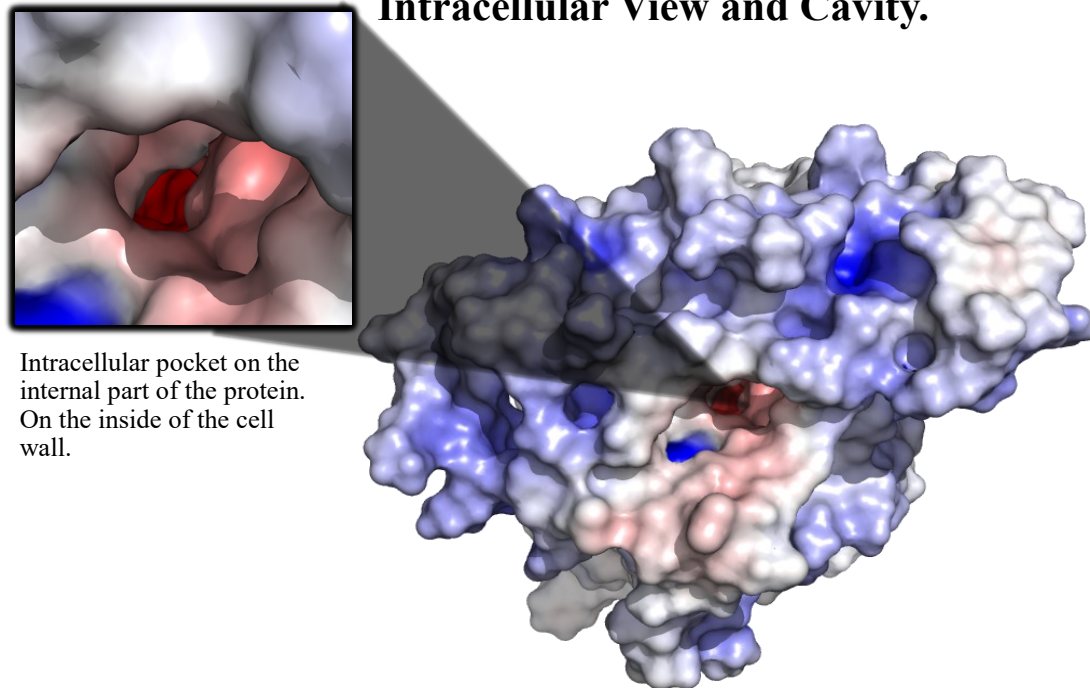
Electrostatic surface analysis rendered in Pymol using the final frame of the simulation for mel1a/melatonin docking run 1 which is saved via VMD1.9.3, utilizing the AMBER forcefield. Areas in red display electronegative surfaces, areas in blue represent positive regions. White areas show areas of the protein which are neutral in charge.
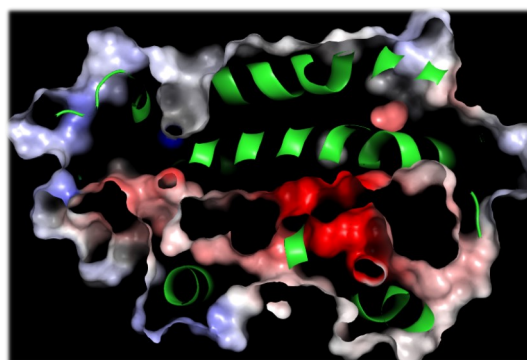
Cutaway side view of the electronegative channel which runs from the external side to the internal side of the membrane.

Side View (membrane Portion).
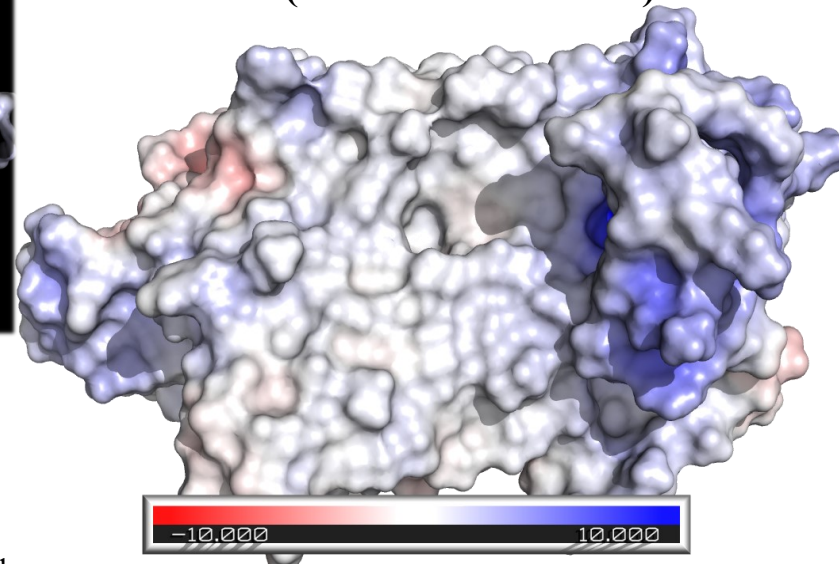
Figure 8.7.1 – Electrostatic Surface Analysis of 30ns Docking Mel1a Run 1.
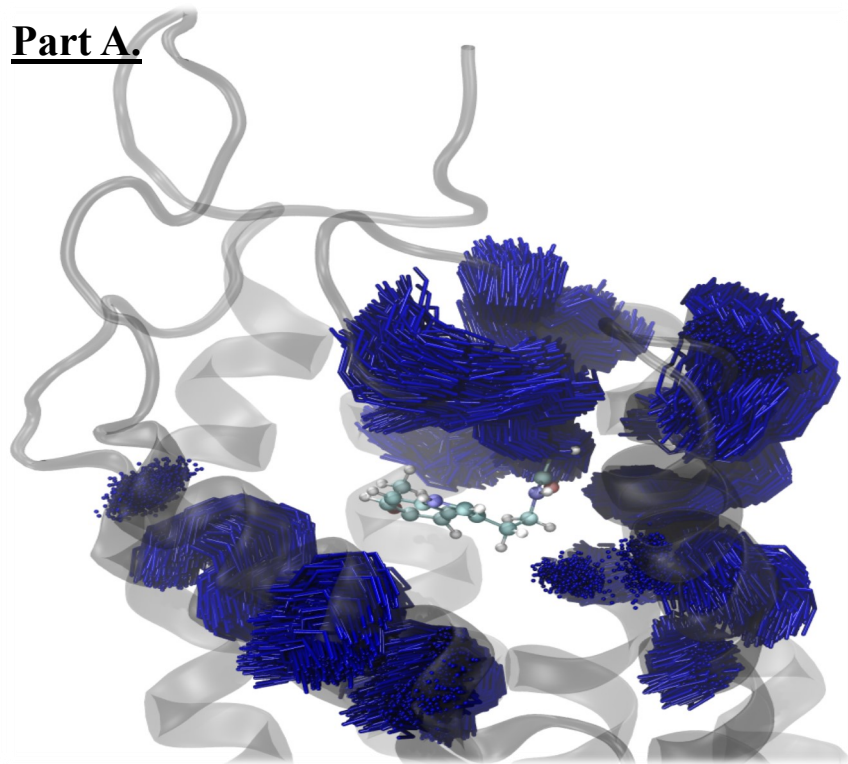
Part A.

Cluster analysis using the WMC PhysBio plugin in V.M.D 193, all 1,500 frames fit within the 3Å Range. Local average RMSD from the melatonin residue is 2.9Å

Part B. Unique Hydrogen Bond Analysis.

```
Found 10 hbonds.
donor                 acceptor     occupancy
SER176-Side-OG        MEL11-Side-C5    13.20%
CYS177-Main-N         MEL11-Side-O2     4.40%
SER103-Side-OG        MEL11-Side-O2    32.33%
SER103-Side-OG        MEL11-Side-C12   21.00%
MEL11-Side-N1         VAL84-Main-O      0.07%
SER176-Side-OG        MEL11-Side-C1     0.27%
MEL11-Side-N1         SER87-Side-OG     9.53%
SER176-Side-OG        MEL11-Side-C12   21.47%
SER176-Side-OG        MEL11-Side-O2     8.07%
MEL11-Side-N1         SER176-Side-OG    0.07%
```

Part C. - VDW interaction Figures between melatonin and surrounding residues.
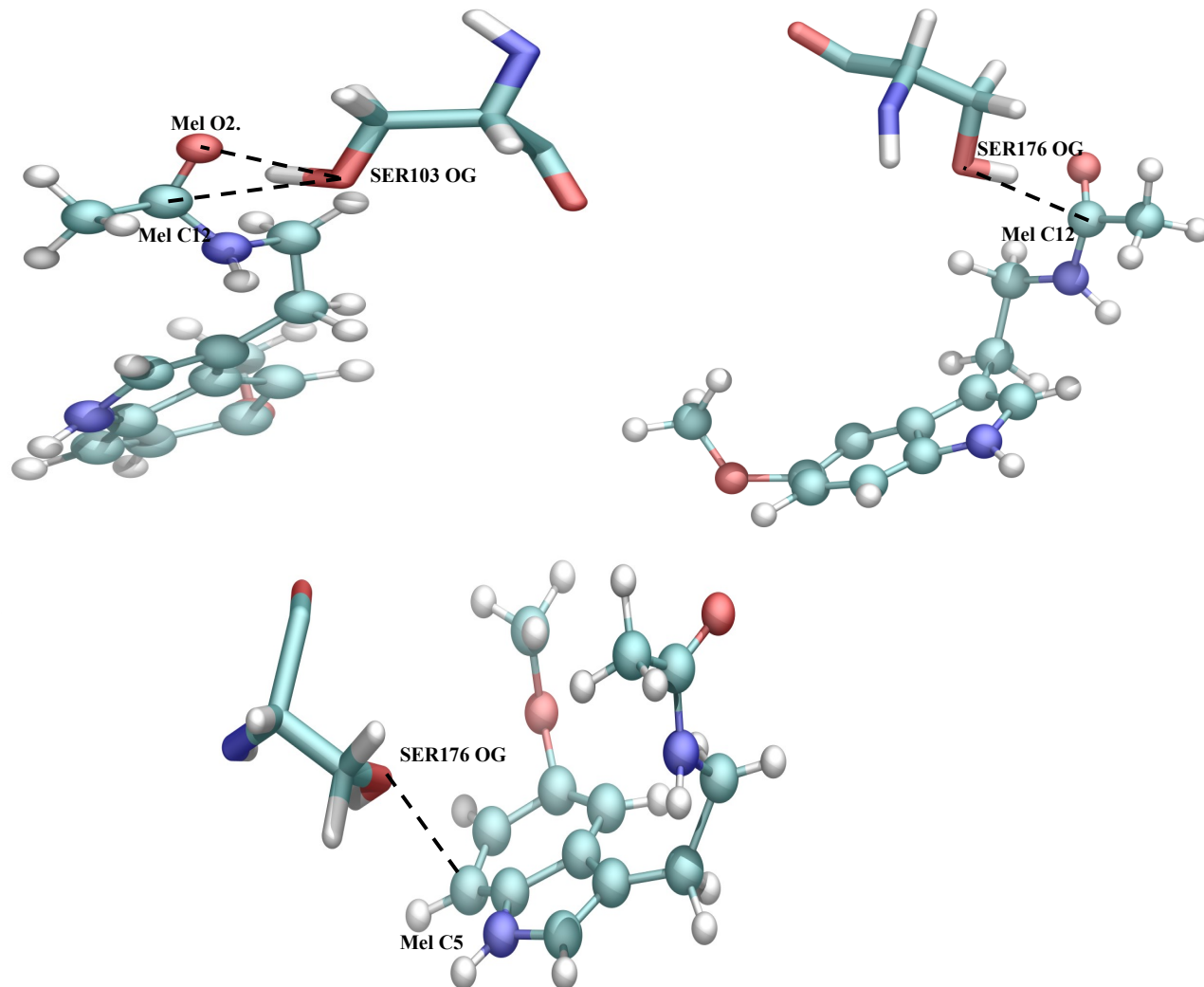
Figure 8.8.1
- Cluster analysis of MEL1a and Melatonin, Unique Hydrogen Bond Analysis. 30ns Docking Run 1.
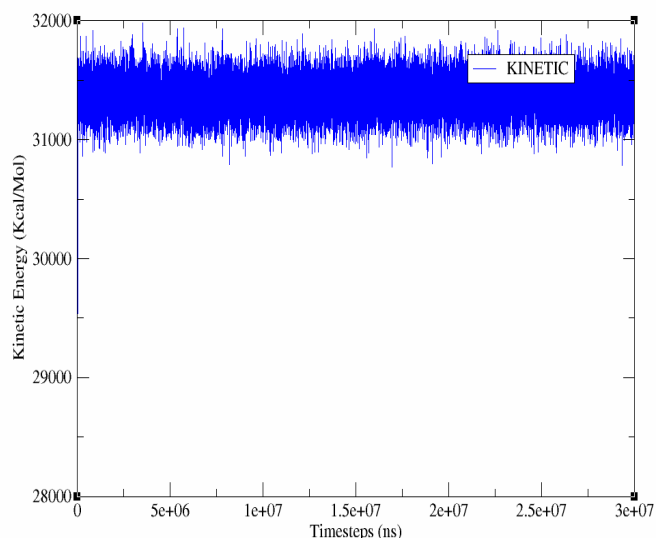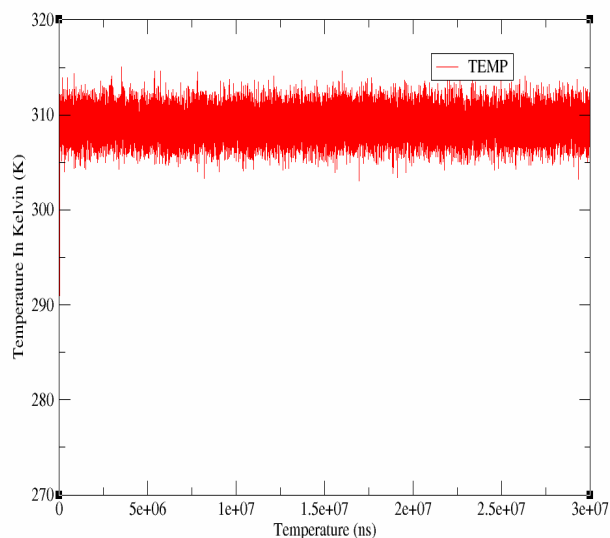
Figure 8.10.1—kinetic energy in the Mel1a/Melatonin 30ns Docking system run 2 (left) and temperature in Kelvin in the previously mentioned system (right).
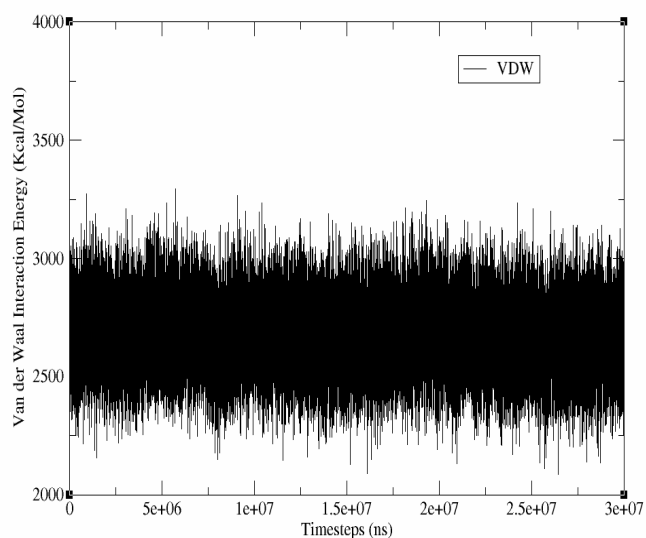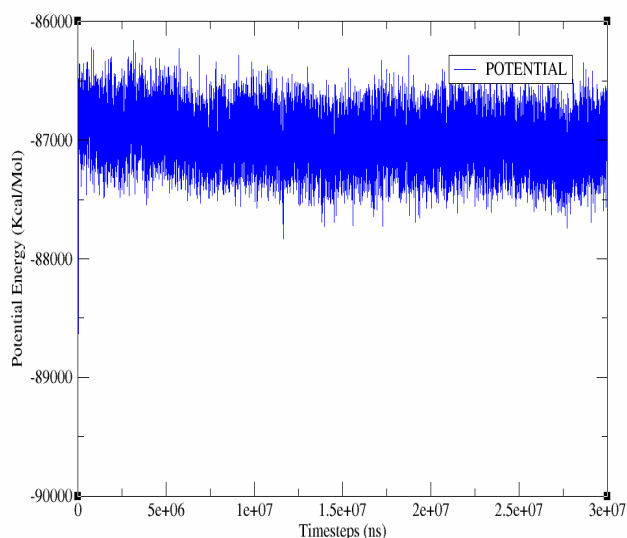
Figure 8.10.2—Van Der Waal Interaction energy (left) in the Mel1a/Melatonin system (run2) and potential energy in the previously mentioned system (right).

8.10.3 - Bond energy of the Mel1a/melatonin docking system (left)
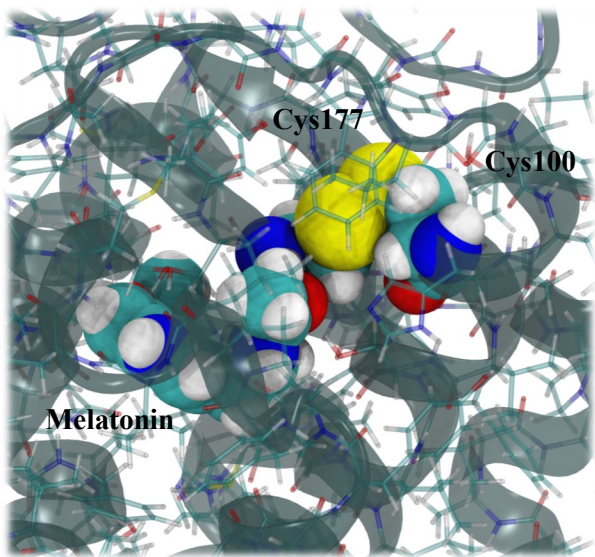
Disulphide bond between Cys100 (Helix C) and Cys177 (Beta Turn A) interacting with the docked Melatonin Molecule in the final frame.
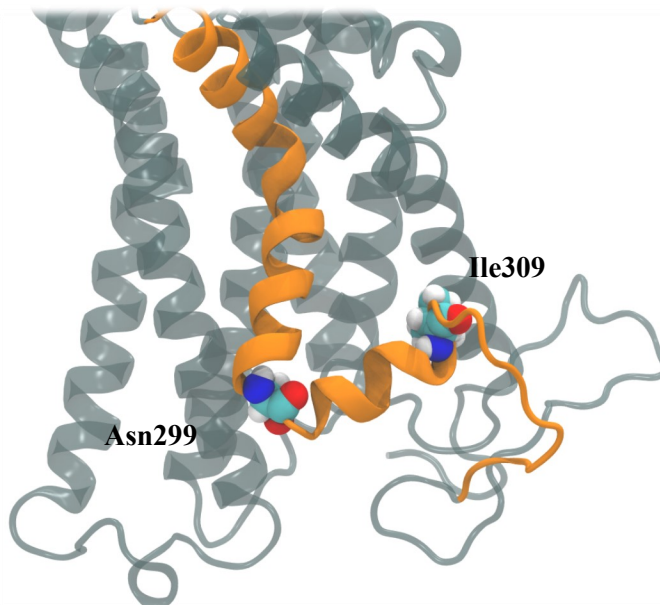
Helix Turn Helix G retracts into the membrane from the intracellular side, restricting the initial residues of loop I from the solvent environment.
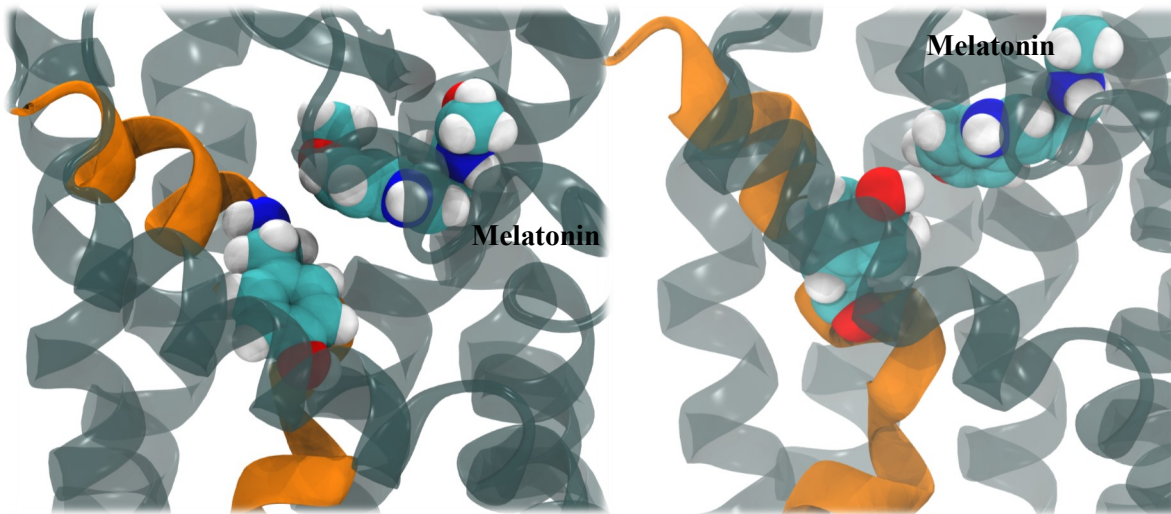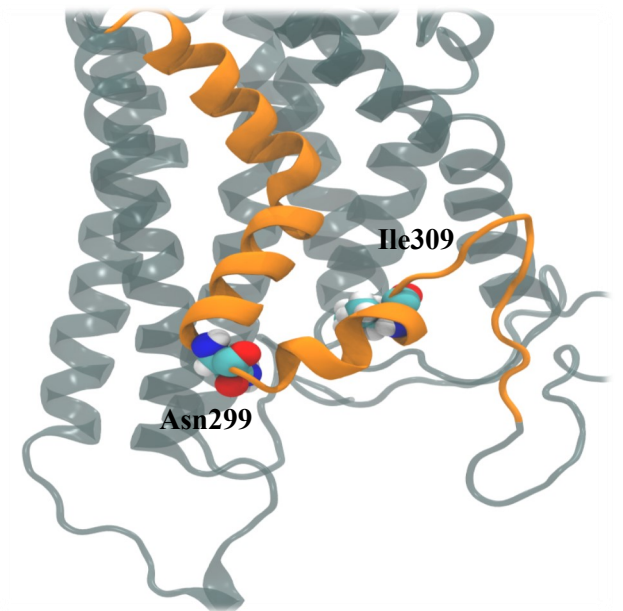
Tyr282 interacting with Melatonin, in the initial frames flips from one polar side of the Tyr282 residue to the other.
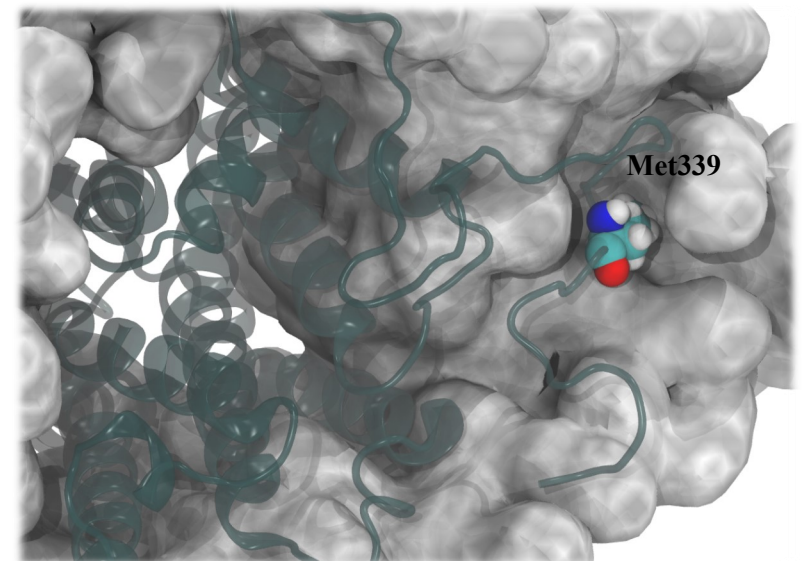
Met339 located in a pocket of the POPC membrane in the final frame, spreading Loop I away from the GPCR.

Figure 8.11.1-Visual Analysis of the 30ns Mel1a GCPR-melatonin trajectory 2.

Structural Alignment Between the I-Tasser Homology Model of Mel1a and the Final Frame of Mel1a/Melatonin Docking 30ns Run 2.

Structural Alignment Between the Mel1a 30ns Control run 2 and the Final Frame of Mel1a/Melatonin Docking 30ns Run 2.

Structural Alignment Between X-ray crystal structure of the β2 adrenergic GCPR and the Final Frame of Mel1a/ Melatonin Docking 30ns Run 2.

Structural Alignment Between the x-ray crystal structure of Rhodopsin and the Final Frame of Mel1a/Melatonin Docking 30ns Run 2.

Part A. - the 30ns Mel1a/melatonin docking structure (run 2, green) aligned with the initial homology model of Mel1a (orange). The RMSD between the structures is 4.607Å.
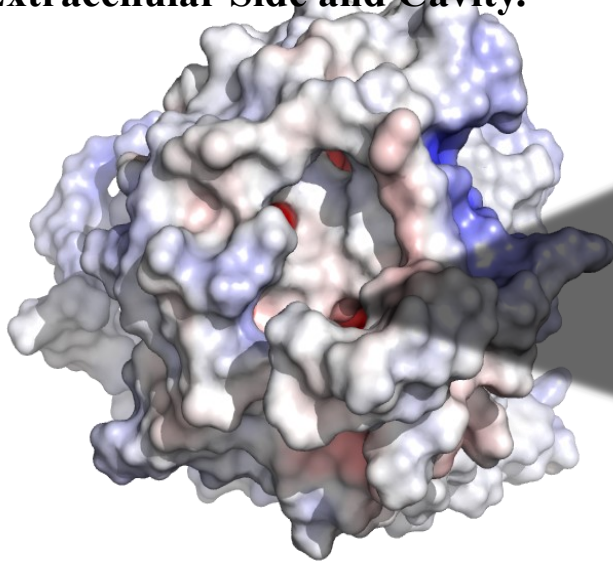
Part B. - The 30ns Mel1a/Melatonin docking structure (green) aligned with the final frame of the control 30ns final frame of Run 2 (red). the RMSD between the structures is 3.427Å

Part C. - The 30ns Mel1a/Melatonin docking structure (run 2, green) aligned with the X-ray crystal structure of the β2 adrenergic GCPR (Cyan). The RMSD between the structures is 4.241Å

Part D. - The 30ns Mel1a/Melatonin docking structure (run 2, green) aligned with the X-ray crystal structure of Rhodopsin (magenta). The RMSD between the structures is 4.107Å.

Figure 8.12.1 - Structural alignment of the 30ns Docking run 2 with Mel1a and melatonin against the initial homology and 30ns control (Run 2) structure of Mel1a, and the X-ray crystal structures of the β2 adrenergic and Rhodopsin GCPR's.

Extracellular Side and Cavity.

Intracellular side and Cavity.

Intracellular cavity has opened up into a wide electronegative channel (right) which continues through to the extracellular side of the protein.
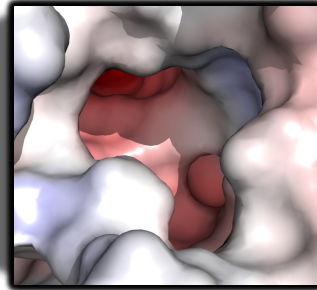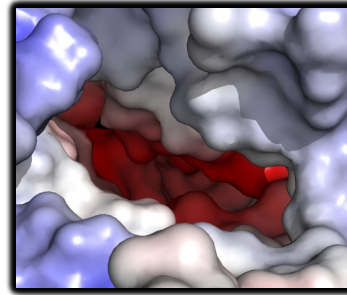
Extracellular cavity on the external side or the Mel1a (above, left), surrounded by mostly neutral areas in run 2. electronegative surface is greatly reduced. Binding of melatonin may cause re-striction of the active site to other biomolecules.
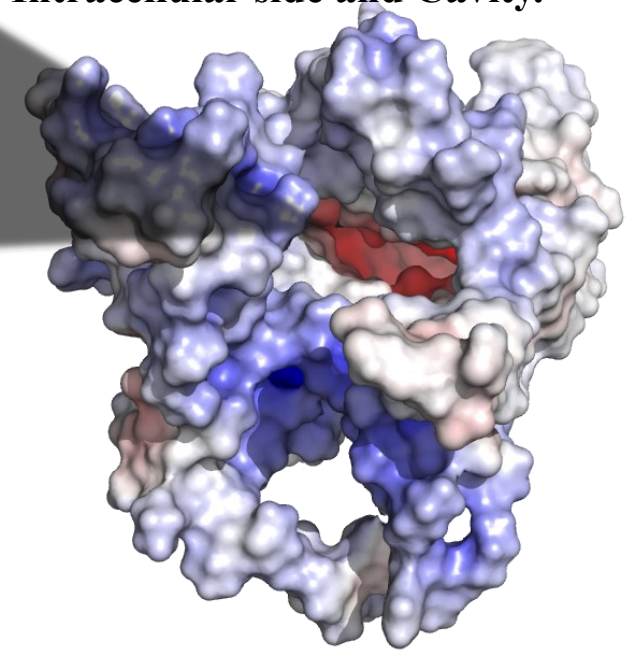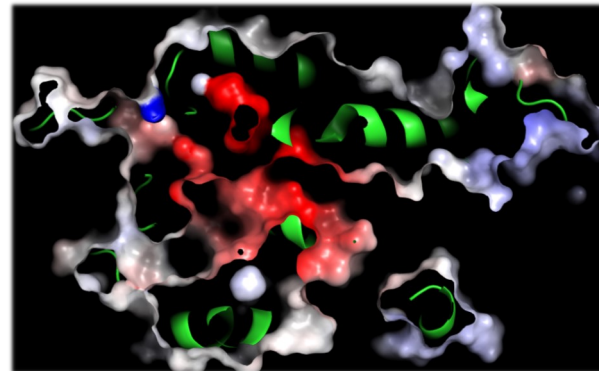
Electrostatics images rendered in Pymol using the APBS electrostatics plugin, utilizing the AMBER forcefield to assess the surface inter-action properties of the second 30 nanosecond run. The final frame of the simulation is saved through VMD prior to being loaded into Pymol. Areas displayed in red are electro-negative, Areas shown in blue are positive. Neutral areas have a white appearance. 30ns Docking Run 2 has a larger channel than the first run. But the external side has less exposed electronegative surface that the first (30ns Run 1).

Side view (membrane portion).

Side slice of Mel1a Shows the wide electronegative channel passing through the centre of the protein from the extracellular to intracellular sides of the membrane.
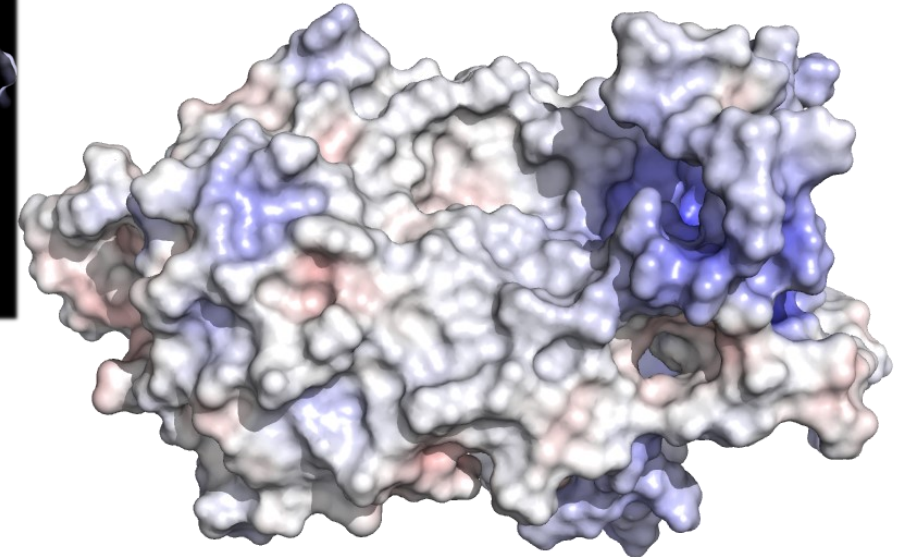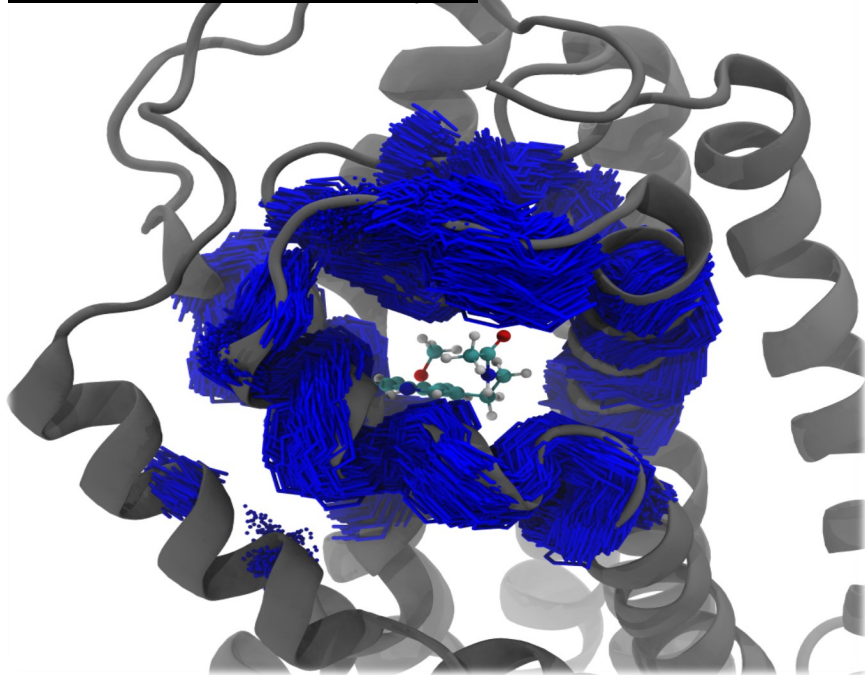
Figure 8.13.1 - Electrostatic Surface Analysis of 30ns Docking Mel1a Run 2.

Part A.—Cluster Analysis

Cluster analysis using the WMC PhysBio plugin in V.M.D 1.9.3, all 1,500 frames fit within the 3Å Range. Local average RMSD 10Å from the melatonin residue is 2.346Å

Part B Unique Hydrogen Bond Analysis.

```
Found 12 hbonds.
donor              acceptor       occupancy
SER176-Side-OG     MEL11-Side-C5    10.33%
CYS177-Main-N      MEL11-Side-O2    11.80%
SER103-Side-OG     MEL11-Side-O2    12.93%
SER103-Side-OG     MEL11-Side-C12    4.07%
MEL11-Side-N1      SER87-Side-OG     0.27%
SER87-Side-OG      MEL11-Side-C5     4.07%
SER176-Side-OG     MEL11-Side-C1     0.07%
MEL11-Side-N1      LEU83-Main-O      0.33%
SER176-Side-OG     MEL11-Side-C12   32.13%
SER87-Side-OG      MEL11-Side-N1     0.33%
SER87-Side-OG      MEL11-Side-C1     0.60%
SER176-Side-OG     MEL11-Side-O2     2.80%
```

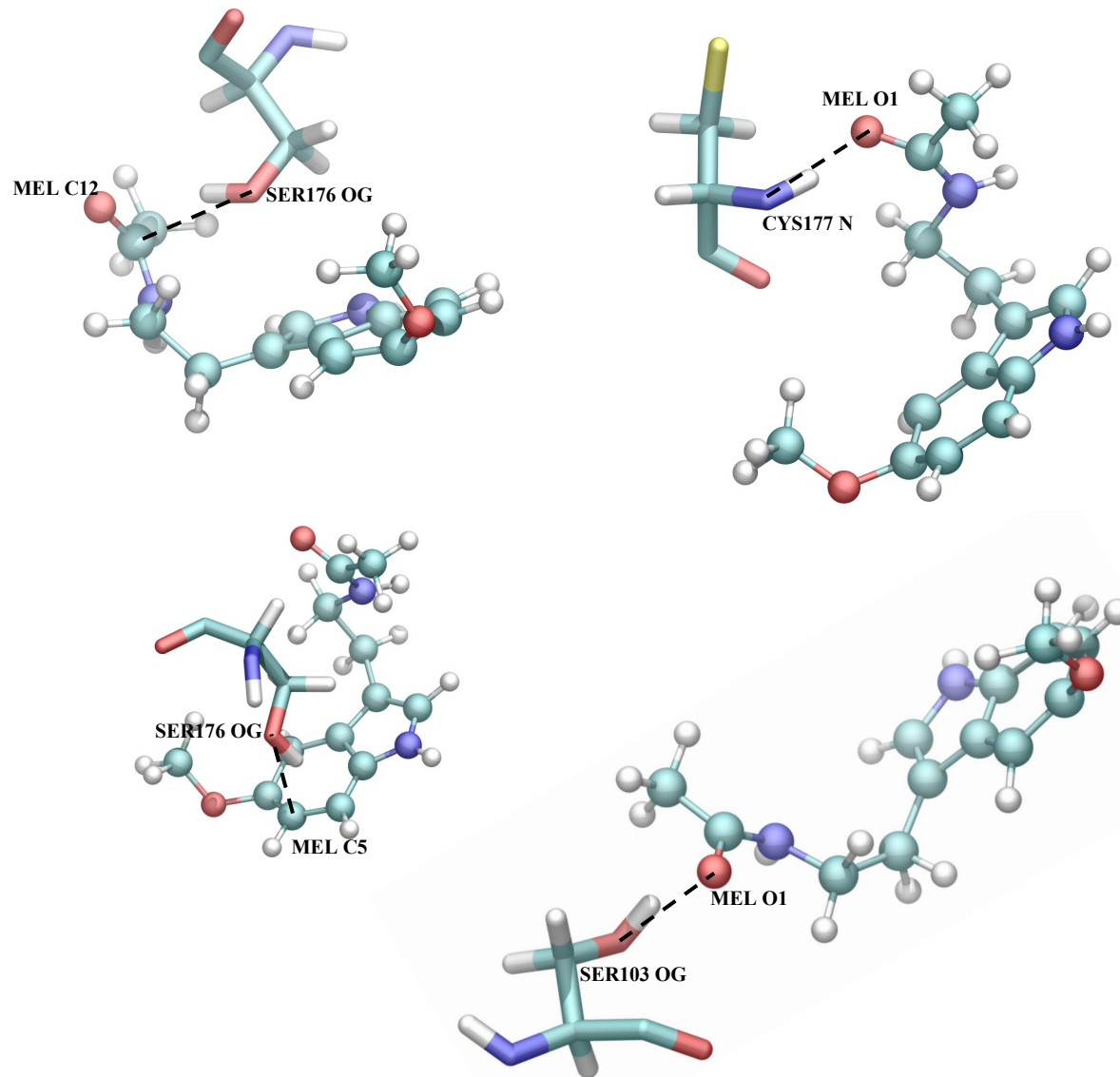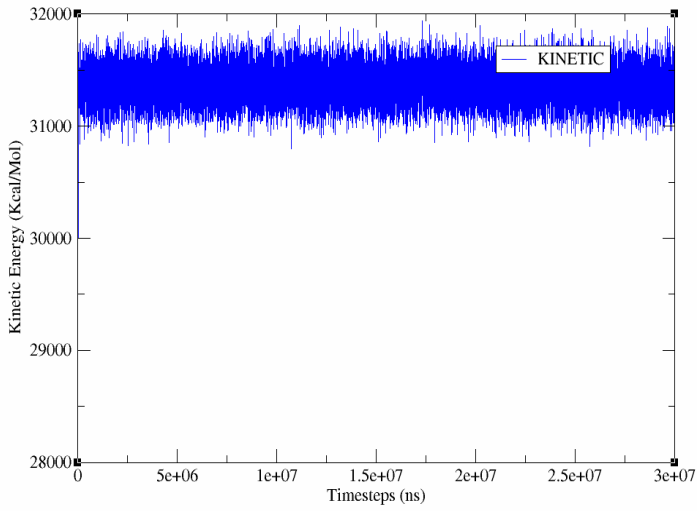# Part C. - VDW interaction Figures between melatonin and surrounding residues.

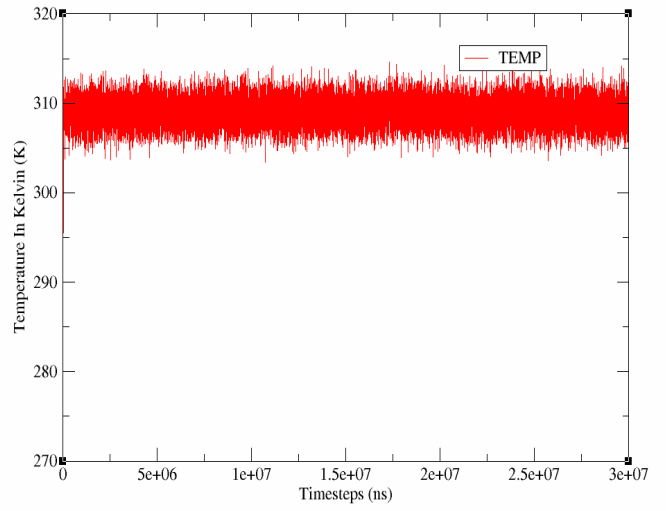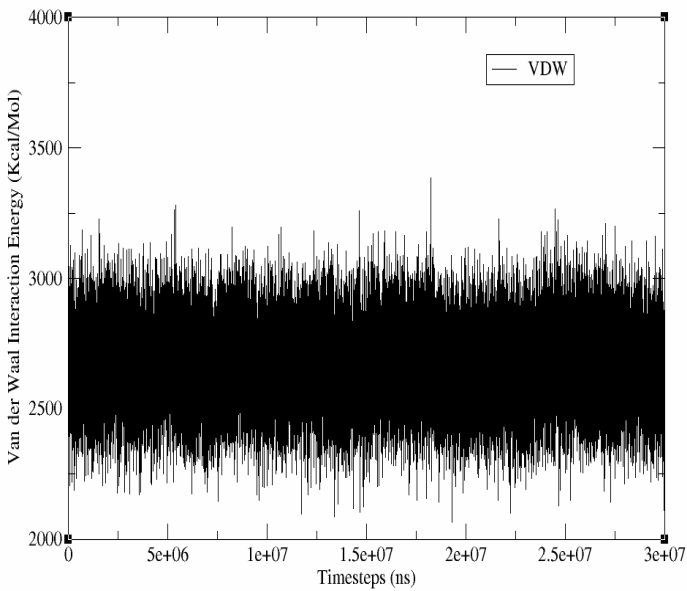# Figure 8.14.1 - Cluster analysis and unique H-Bonds of Mel1a with Melatonin. Docking run 2.

Figreu 8.16.1—kinetic energy over the entire Mel1a/melatonin docking system simulation (left) and temperature of the previously mentioned system..
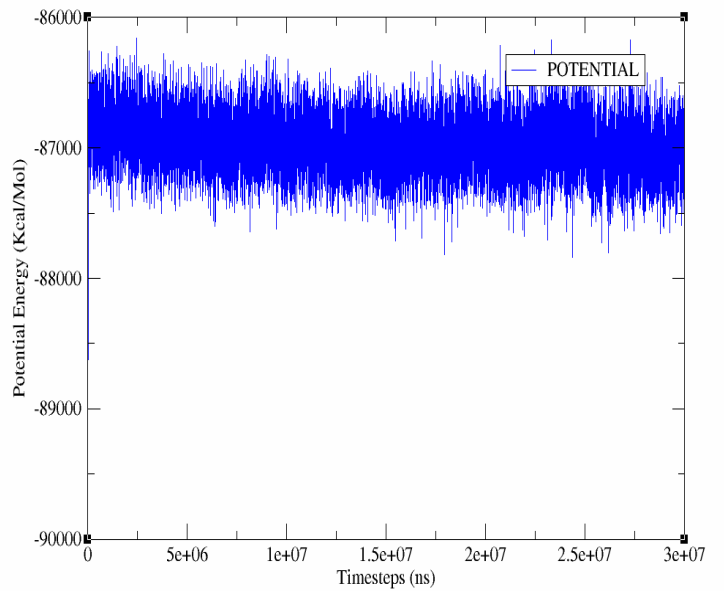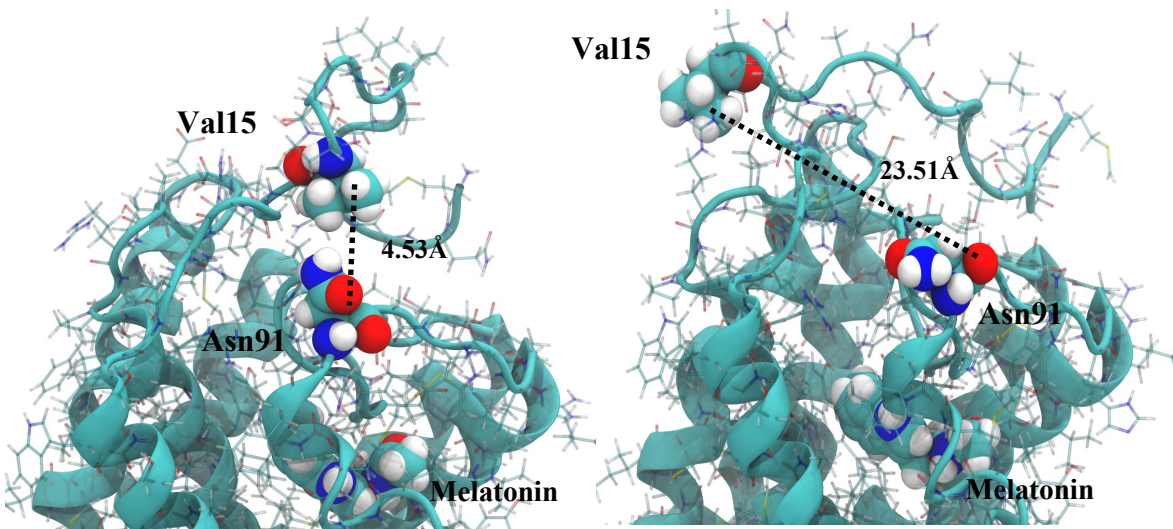

Figure 8.16.2—Van Der Waal interaction energy of the whole system in the previously mentioned docking system (left), and potential energy in the same docking system.

8.16.3 - Bond energy in the Mel1a/melatonin docking system.

Val15 moves from 4.53Å in the initial frame (left) to 23.51Å by the final frame reconforming along with Loop A on the extracellular side (right).
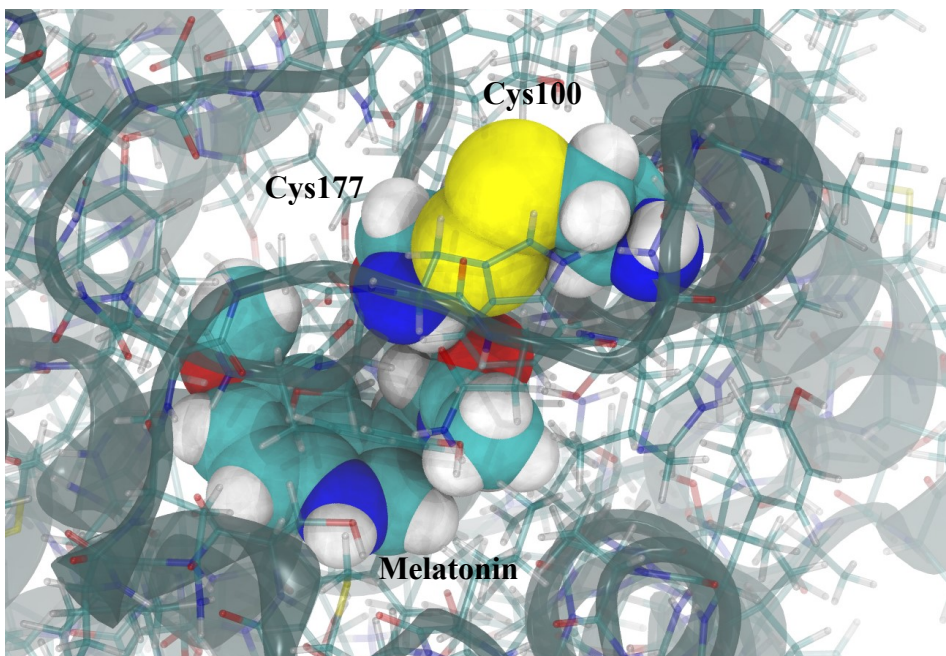
Disulphide bond between Cys100 (Helix C) and Cys177 (Beta Turn A) interacting with the docked Melatonin Molecule in the final frame.
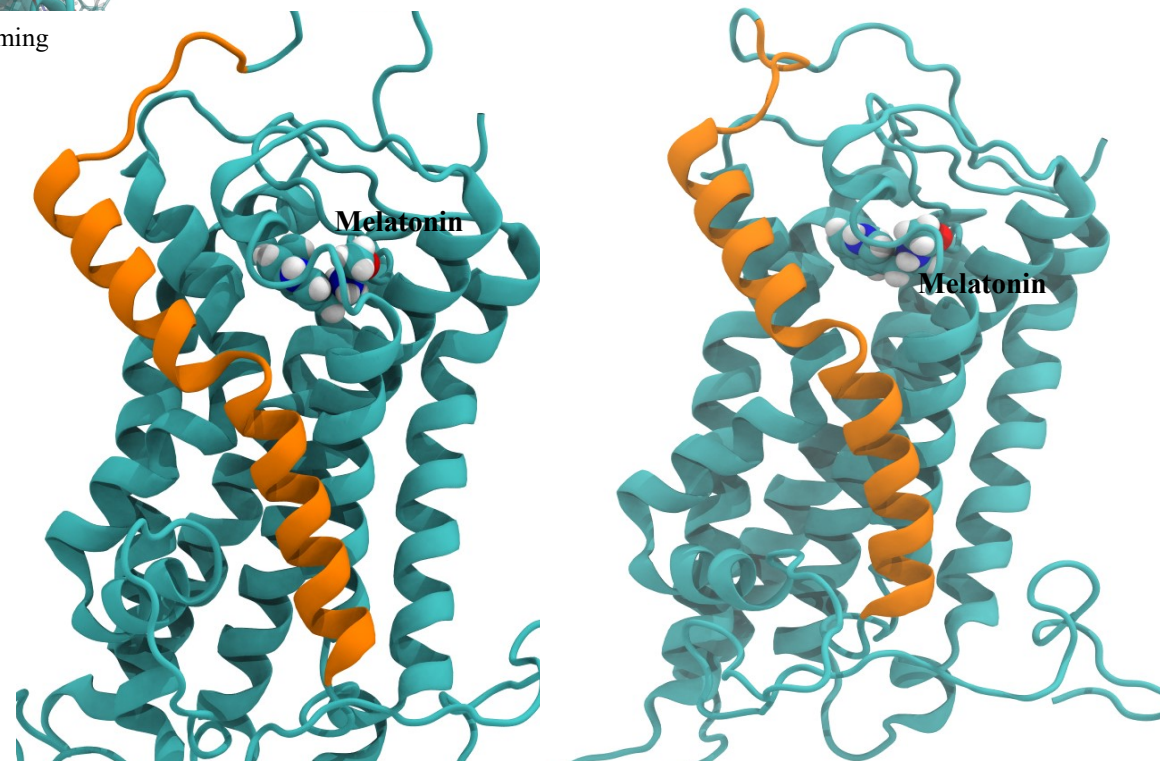
Alpha Helix A (highlighted in orange) in the initial frame of the simulation (left) helix A conforming closer to the Melatonin Residue in the final frame of the calculation (right).

Figure 8.17.1-Visual Analysis of the 30ns Mel1a GCPR-melatonin trajectory 3.

Structural Alignment Between the I-Tasser Homology Model of Mel1a and the Final Frame of Mel1a/Melatonin Docking 30ns Run 3.

Structural Alignment Between the Mel1a 30ns Control Run 2 and the Final Frame of Mel1a/Melatonin Docking 30ns Run 3.

Structural Alignment Between X-ray crystal structure of the β2 adrenergic GCPR and the Final Frame of Mel1a/ Melatonin Docking 30ns Run 3.

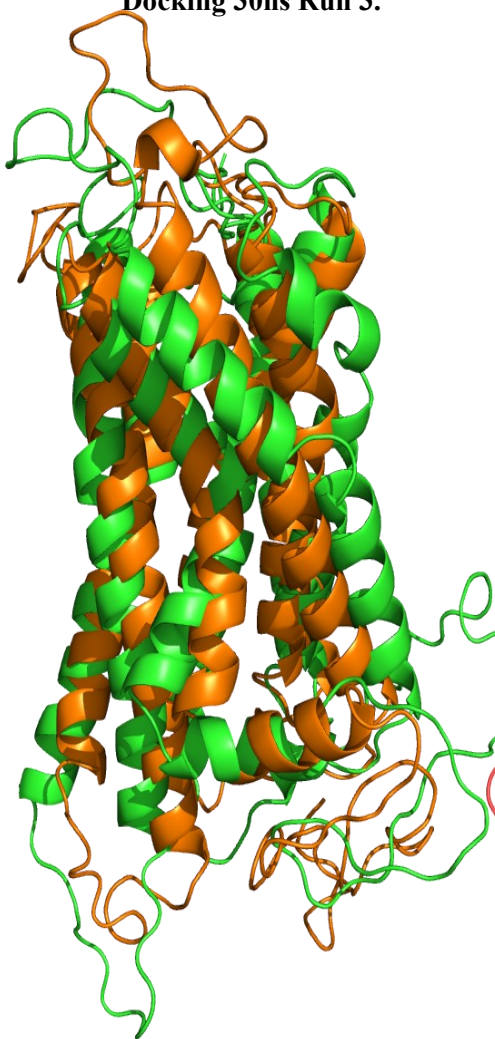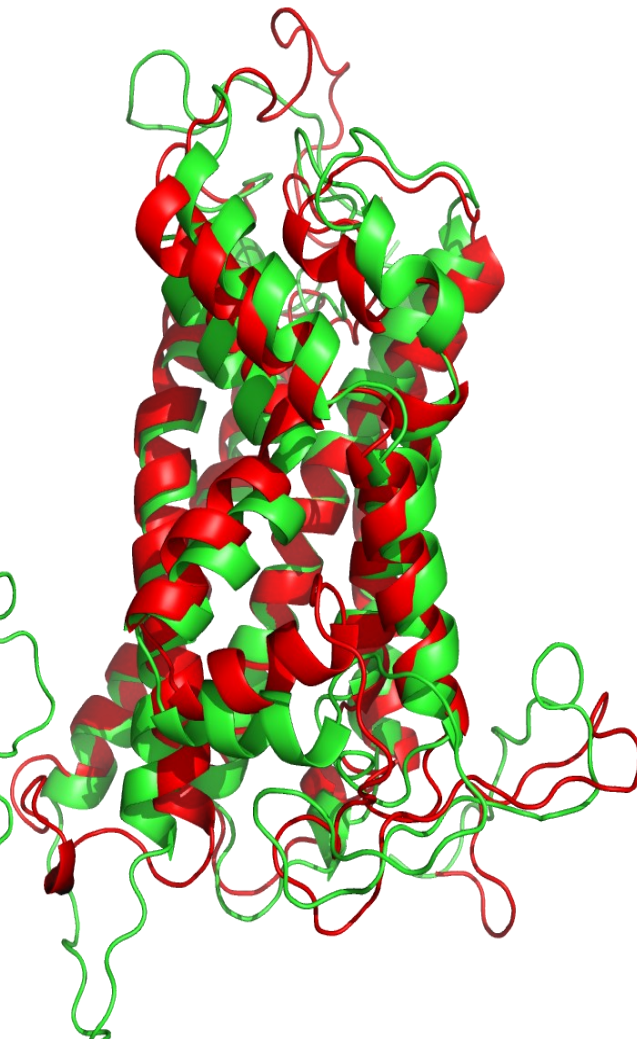**Structural Alignment Between the x-ray crystal structure of Rhodopsin and the Final Frame of Mel1a/Melatonin Docking 30ns Run 3.**
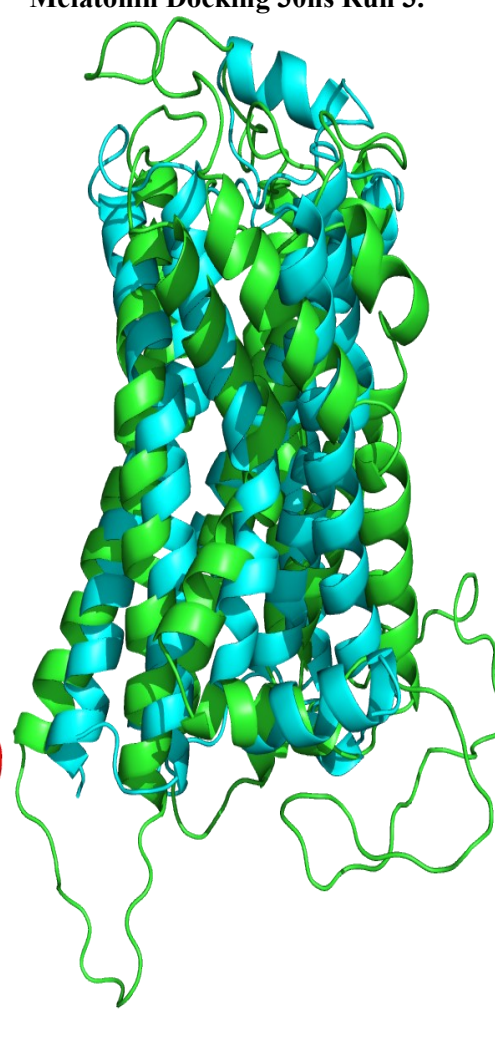
Part A. - the 30ns Mel1a/melatonin docking structure (run 3, green) aligned with the initial homology model of Mel1a (orange). The RMSD between the structures is 4.198Å.

Part B. - The 30ns Mel1a/Melatonin docking structure (run 3, green) aligned with the final frame of the control 30ns final frame of Run 2 (red). the RMSD between the structures is 2.674Å

Part C. - The 30ns Mel1a/Melatonin docking structure (run 3, green) aligned with the X-ray crystal structure of the β2 adrenergic GCPR (Cyan). The RMSD between the structures is 3.944Å

Part D. - The 30ns Mel1a/Melatonin docking structure (run 3, green) aligned with the X-ray crystal structure of Rhodopsin (magenta). The RMSD between the structures is 4.191Å.

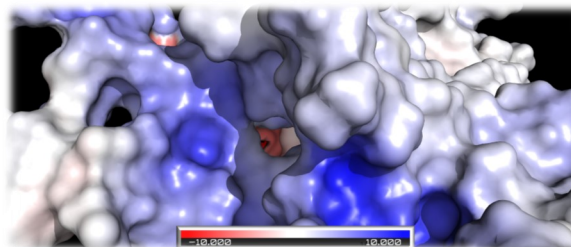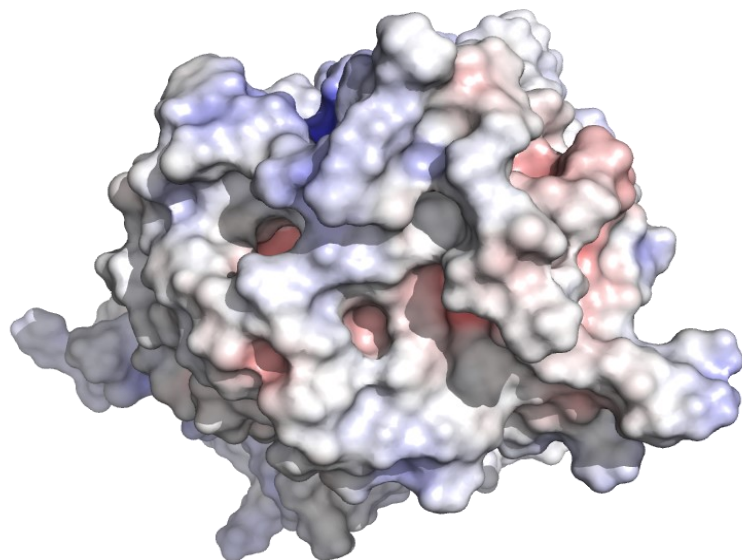Figure 8.18.1 - Structural alignment of the 30ns Docking run 3 with Mel1a and melatonin against the initial homology and 30ns control (Run 2) structure of Mel1a, and the X-ray crystal structures of the β2 adrenergic and Rhodopsin GCPR's.

Extracellular View Of Protein.

Intracellular View of protein.

The external and internal sides have no obvious cavities unlike the other 2 docking runs through the core of the protein. But there is a cavity on the side of the protein (Below) which does not enter the central core section of the molecule. This is seen in the intracellular view and side view images on the right upper and lower images.

Electrostatics images rendered in Pymol using the APBS electrostatics plugin, utilizing the AMBER forcefield to assess the surface interaction properties of the third 30 nanosecond docking run. The final frame of the simulation is saved through VMD prior to being loaded into Pymol. Areas displayed in red are electronegative, Areas shown in blue are positive. Neutral areas have a white appearance.
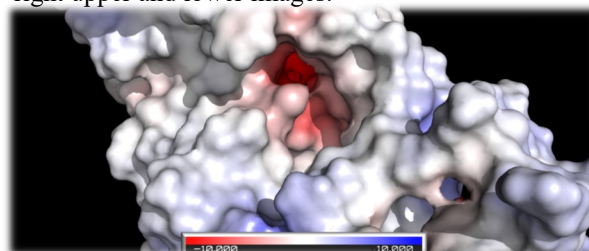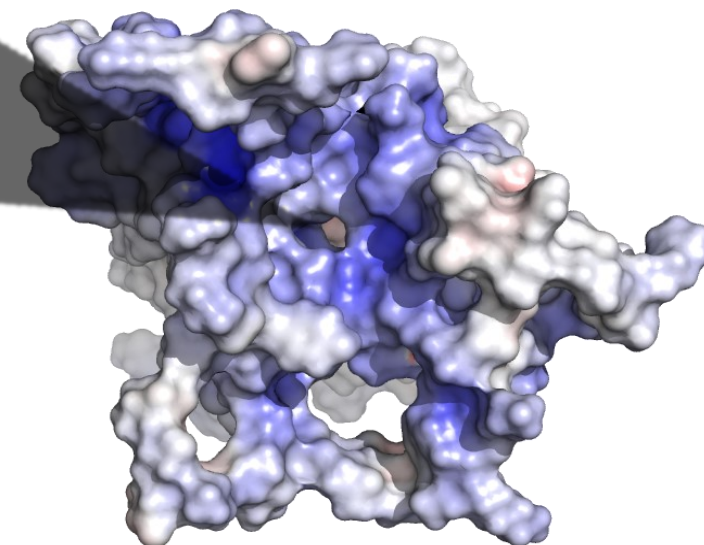
Side View (Membrane Portion).
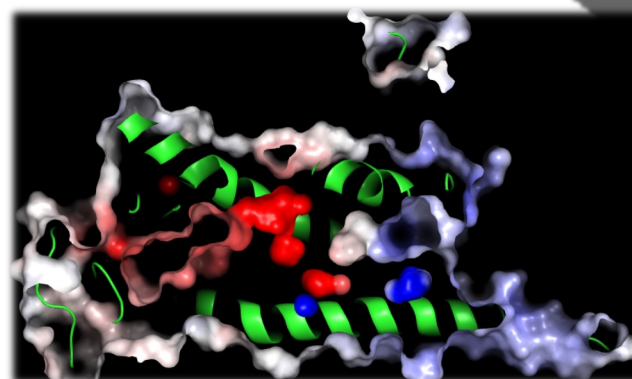
There is no traversing internal channel (above) in the core of this electrostatic analysis, but cavities are still present on the extracellular and intracellular sides - although there is no obvious route into the cavities.. There is an electronegative portion on the side of the protein in the bi-lipid membrane of the system. (right).

Figure 8.19.1 — Electrostatic surface analysis of 30ns Docked Mel1a Run 3.

Part A—Cluster Analysis.
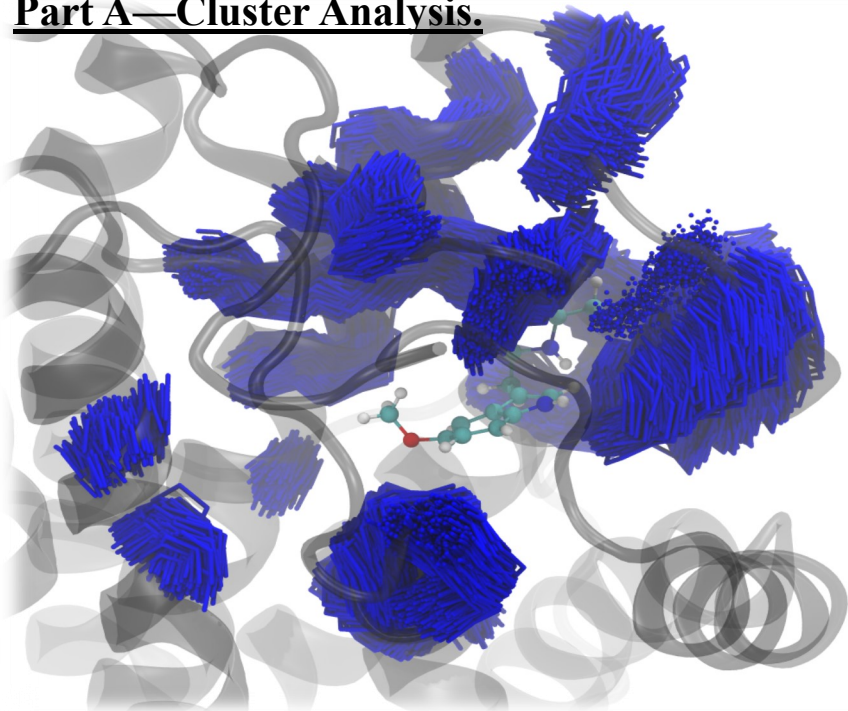
Cluster analysis using the WMC PhysBio plugin in V.M.D 1.9.3, all 1,500 frames fit within the 3Å Range. Local average RMSD 10Å from the melatonin residue is 2.193Å

Part C. - VDW interaction Figures between melatonin and surrounding residues.

Part B Unique Hydrogen Bond Analysis.

```
Found 9 hbonds.
donor              acceptor      occupancy
SER103-Side-OG     MEL11-Side-C12    8.53%
SER176-Side-OG     MEL11-Side-C5     11.07%
SER103-Side-OG     MEL11-Side-O2     30.87%
CYS177-Main-N      MEL11-Side-O2     21.67%
SER176-Side-OG     MEL11-Side-C1     0.47%
MEL11-Side-N1      VAL84-Main-O      0.20%
MEL11-Side-N1      SER87-Side-OG     1.73%
SER176-Side-OG     MEL11-Side-C12    18.73%
SER176-Side-OG     MEL11-Side-O2     2.00%
```
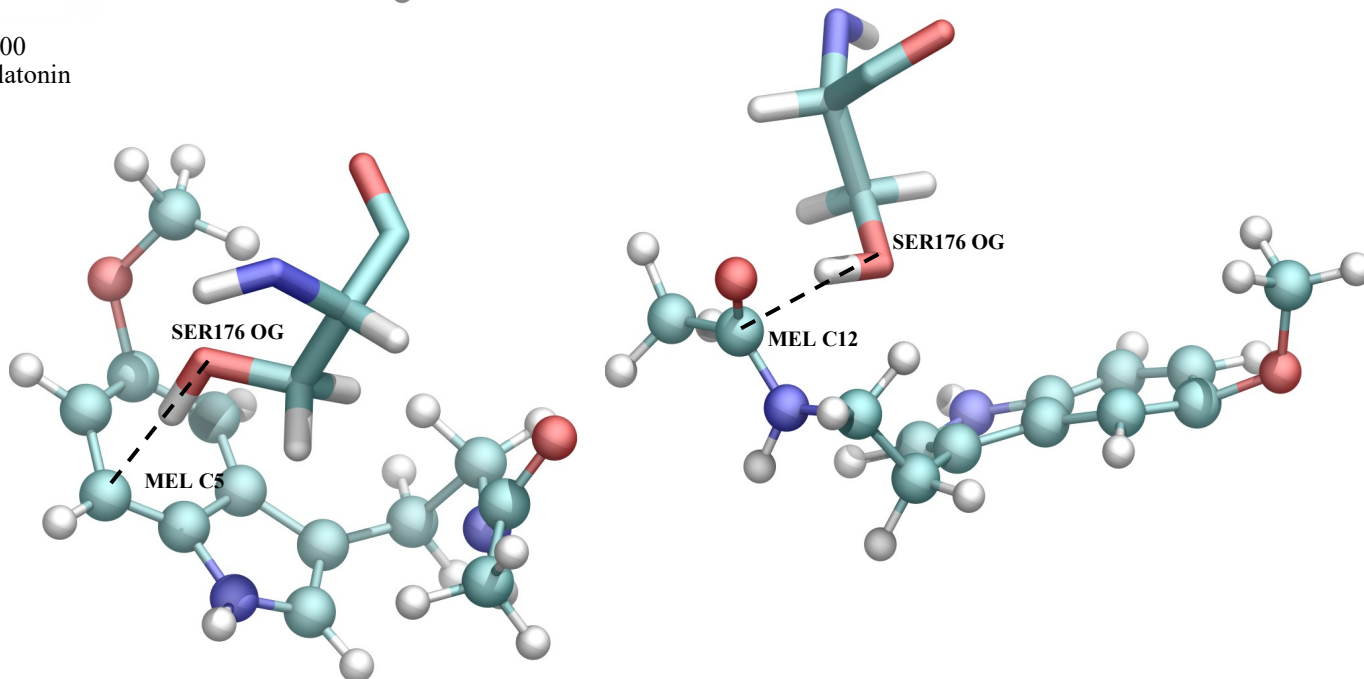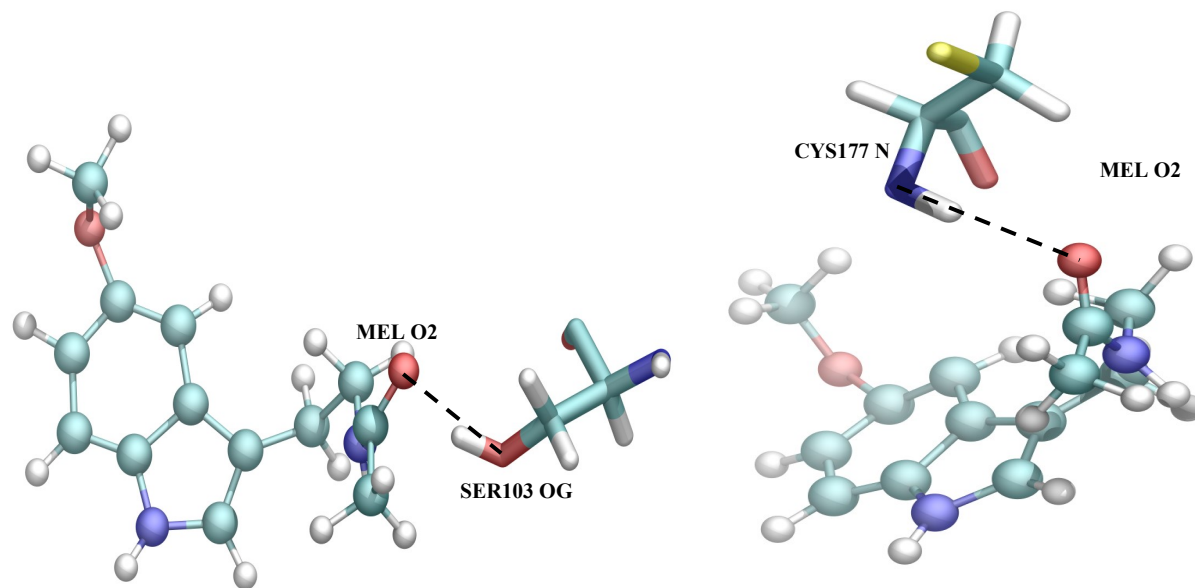
Figure 8.20.1 - Cluster analysis and H-bond analysis of 30ns MEL1a and Melatonin, 30ns Docking run 3.

Discussion.

A highly stable representative model of the GPCR Mel1a was achieved through all-atom molecular dynamics after initial homology modelling on the I-tasser web server, while appearing closely related in architecture to the Rhodopsin and B2 adrenergic GCPR's. Triplicate datasets show a myriad of different conformations in both the 10 ns and 30ns trajectory datasets. A high degree of stability is seen in the membrane core of the Mel1a GCPR while the intracellular and extracellular loops show a larger degree of free movement (section 5.0) with the simulation times being increased from 10 ns to 30 ns, there was an overall increase in RMSD seen in the model, repeating the stability in the core and high deviation in the intracellular and extracellular portions while retaining the GPCR common characteristics when compared to Rhodopsin and the B2 adrenergic receptors (section 6.0).

The Melatonin ligand which was generated through Gaussian 16 represents a close match for the X-ray crystal version retrieved by Wakahara et al in 1972 and presents a more accurate version of the ligand than CGenFF could predict. Melatonin, with the addition of the hessian basis set data (MP2-36G*). It is more likely to ellicit an accurate response from the homology model of Mel1a than its CGenFF counterpart with notable differences in its Dihedral data being noted in a vacuum, and post monimization in solvent may lead to a more accurate final result than achieveable through CHARMM.

After the addition of melatonin to the system it is clear that the active site on the extracellular side of the protein and is based around the characteristic disulphide bond common to the GPCR family of proteins (Cys100 and Cys177). This appears to be mostly achieved by Van der Waal interaction after docking site analysis (section 8.0). RMSD within the Mel1a GPCR remained stable when backwards compared to the 10 and 30 ns trajectory data (sections 6.0 and 8.0).

With further generated trajectory data for the bound and unbound states described in the results section, commonality may be found between the datasets which better describe an average conformation. It appears that the binding of melatonin on the extracellular side of the membrane leaves the Mel1a GPCR de-polarized, while the intracellular side remains in a positive state. In the 30 ns Mel1a-melatonin trajectory a transmembrane negative portion is visible, implying it may traverse the surface of the POPC membrane possibly looking for an interaction with a partner membrane bound protein as a result (Ferguson *et al.*, 1996). This depolarization is likely the first step in a signal which it relayed to the interior of the cell membrane, which is then acted upon by further proteins within the circadian rhythm pathway.

The molecular dynamics study of Mel1a and its primary agonist melatonin describes the difficult nature of molecular description in dynamic models. Every simulation has a set of trajectories which is unique to the previous, which means every simulation is different to the last. The action of melatonin which is seen on the Mel1a GCPR in this study is a controlled environment - in isolation from the rest of the pathway from which the receptor is derived. Consequently, only the binding interaction between the ligand and its receptor can be described. This inherently asks more questions than it answers. As the ligand shows interaction with the protein, what does the binding action of melatonin cause in the internal side of the membrane?

It is known through experiment that GCPRs have heterotrimeric guanine nucleotide-binding proteins (G-proteins) anchored to the internal side of the lipid membrane. Awaiting stimulus on the internal side of the bi-lipid membrane consisting of an alpha, beta and gamma subunits (Rosenbaum, Rasmussen and Kobilka, 2009). This may explain the non-conformity seen in the results of the initial MD simulations where Mel1a is in a state of 'conformational flux' in the control trajectories. It may also be resolved by simulation time being extended further for more conformational exploration. There is no signal to be delivered further from the receptor, nor a

clear open pathway to induct a signal in this system after the addition of melatonin. Additional study involving g-proteins will yield results which further the understanding of conformational change in the system which has been built.

The absence of any clear active site for melatonin in the control section of the results also generates a new question. Since the action of rhodopsin-like G-protein coupled receptors is to transmit a signal like a switch, it is conceivable the binding of proteins anchored on the internal side of the membrane cause conformational change for external ligands to be appropriated by the Mel1a GPCR. Arrestins are small family of proteins important for signal transduction and are known to act on GCPR's like Mel1a post activation, to return them to an inactive state (Ferguson *et al.*, 1996). A trajectory analysis of an arrestin family protein bound to the glycosylated intracellular side of a GCPR could potentially also reveal an active site on the proteins extracellular region following protein-protein interaction (Hirsch *et al.*, 1999). Proteins are a network system which are comprised of many different types of biomolecules. Further study of the signal transduction pathway inclusive of Mel1a will provoke answers into the full understanding of the GCPR and its partner interactions.

The data accrued through this study can be taken further by comparatively using simple compounds which have been tailored through the drug discovery process, such as tasimelteon, agomelatine and ramelteon which were designed to target the Mel1a GCPR - to study the dynamic effect each drug causes in backwards comparison to Melatonin – further extending knowledge into the possible efficacious and side effects (attrition) on the all atom scale.

Bibliography.

Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic acids research*. Oxford University Press, 25(17), pp. 3389–3402.

Andersen, H. C. (1983) 'Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations', *Journal of Computational Physics*. Elsevier, 52(1), pp. 24–34.

Arnold, K. *et al.* (2006) 'The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling', *Bioinformatics*, 22(2), pp. 195–201. doi: 10.1093/bioinformatics/bti770.

Bourne, R. S., Mills, G. H. and Minelli, C. (2008) 'Melatonin therapy to improve nocturnal sleep in critically ill patients: Encouraging results from a small randomised controlled trial', *Critical Care*. BioMed Central, 12(2), p. R52. doi: 10.1186/cc6871.

Brooks, B. R. *et al.* (2009) 'CHARMM: The biomolecular simulation program', *Journal of Computational Chemistry*. Wiley-Blackwell, 30(10), pp. 1545–1614. doi: 10.1002/jcc.21287.

Carter, M. D. and Juurlink, D. N. (2012) 'Melatonin', *Canadian Medical Association Journal*. Canadian Medical Association, 184(17), pp. 1923–1923. doi: 10.1503/cmaj.111765.

Caspi, O. (2004) 'Melatonin for the prevention and treatment of jet lag', *Alternative Therapies in Health and Medicine*. John Wiley & Sons, Ltd, 10(2), pp. 74–78. doi: 10.1002/14651858.CD001520.

Delano, W. L. (2002) 'PyMOL: An Open-Source Molecular Graphics Tool', *CCP4 Newsletter On Protein Crystallography*, 700, pp. 82–92. Available at: https://www.ccp4.ac.uk/newsletters/newsletter40/11_pymol.pdf.

Dolan, M. A., Noah, J. W. and Hurt, D. (2012) 'Comparison of common homology modeling algorithms: Application of user-defined alignments', in Orry, A. J. W. and Abagyan, R. (eds) *Methods in Molecular Biology*. Totowa, NJ: Humana Press, pp. 399–414. doi: 10.1007/978-1-61779-588-6_18.

Dolinsky, T. J. *et al.* (2004) 'PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations', *Nucleic acids research*. Oxford University Press, 32(suppl_2), pp. W665–W667.

Dollins, A. B. *et al.* (1994) 'Effect of inducing nocturnal serum melatonin concentrations in daytime on sleep, mood, body temperature, and performance', *Proceedings of the National*

Academy of Sciences of the United States of America, 91(5), pp. 1824–1828. doi: 10.1073/pnas.91.5.1824.

Dror, R. O. *et al.* (2011) 'Activation mechanism of the β2-adrenergic receptor', *Proceedings of the National academy of sciences of the USA*, 108(46), pp. 18684–18689. doi: 10.1073/pnas.1110499108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1110499108.

Ferguson, S. S. G. *et al.* (1996) 'Role of P-Arrestin in Mediating Agonist-Promoted G Protein-Coupled Receptor Internalization', *Science*. American Association for the Advancement of Science, 271(January), pp. 3–6.

Frisch, M. J. *et al.* (2016) 'Gaussian˜16 {R}evision {B}.01'.

Hanwell, M. D. *et al.* (2012) 'Avogadro: An advanced semantic chemical editor, visualization, and analysis platform', *Journal of Cheminformatics*, 4(8), p. 17. doi: 10.1186/1758-2946-4-17.

Hardy, D. J. *et al.* (2015) 'Multilevel summation method for electrostatic force evaluation', *Journal of Chemical Theory and Computation*. American Chemical Society, 11(2), pp. 766–779. doi: 10.1021/ct5009075.

Hirsch, J. A. *et al.* (1999) '˚ Crystal Structure of Visual Arrestin : The 2 . 8 A A Model for Arrestin ' s Regulation', *Cell*. Cambridge, Mass.: MIT Press, 1974-, 97(2), pp. 257–269.

Huang, C. C. *et al.* (2014) 'Enhancing UCSF Chimera through web services', *Nucleic acids research*. Oxford University Press, 42(W1), pp. W478–W484.

Huang, J. and MacKerell, A. (2013) 'CHARMM36 all atom additive protein force field: Validation based on comparison to NMR data', *J Comput Chem*. Wiley Online Library, 34(25), pp. 2135–2145. doi: 10.1002/jcc.23354.

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'VMD: Visual molecular dynamics', *Journal of Molecular Graphics*, 14(1), pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.

Jan, J. E., Espezel, H. and Appleion, R. E. (1994) 'the Treatment of Sleep Disorders With Melatonin', *Developmental Medicine & Child Neurology*. Blackwell Publishing Ltd, 36(2), pp. 97–107. doi: 10.1111/j.1469-8749.1994.tb11818.x.

Jurečka, P. *et al.* (2006) 'Benchmark database of accurate (MP2 and CCSD (T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs', *Physical Chemistry Chemical Physics*. Royal Society of Chemistry, 8(17), pp. 1985–

1993.

Karplus, K., Barrett, C. and Hughey, R. (1998) 'Hidden Markov models for detecting protein homologies', *Bioinformatics*, 14(10), pp. 846–856. Available at: file:///d:/journals/4783.pdf.

Karplus, M. and McCammon, J. A. (2002) 'Molecular dynamics simulations of biomolecules', *Nature Structural Biology*. Nature Publishing Group, 9(9), pp. 646–652. doi: 10.1038/nsb0902-646.

Kobilka (2007) 'G protein coupled receptor structure and activation', *Biochimica et biophysica acta*, 1768(4), pp. 794–807. doi: 10.1016/j.bbamem.2006.10.021.

Kräutler, V., Van Gunsteren, W. F. and Hünenberger, P. H. (2001) 'A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations', *Journal of computational chemistry*. Wiley Online Library, 22(5), pp. 501–508.

Lemoine, P. and Zisapel, N. (2012) 'Prolonged-release formulation of melatonin (Circadin) for the treatment of insomnia', *Expert Opinion on Pharmacotherapy*. Taylor & Francis, 13(6), pp. 895–905. doi: 10.1517/14656566.2012.667076.

Lindahl, E. and Sansom, M. S. (2008) 'Membrane proteins: molecular dynamics simulations', *Current Opinion in Structural Biology*, 18(4), pp. 425–431. doi: 10.1016/j.sbi.2008.02.003.

McDonald, I. R. (1972) 'NpT-ensemble monte carlo calculations for binary liquid mixtures', *Molecular Physics*. Taylor & Francis, 23(1), pp. 41–58. doi: 10.1080/00268977200100031.

Melke, J. *et al.* (2008) 'Abnormal melatonin synthesis in autism spectrum disorders', *Molecular Psychiatry*. Nature Publishing Group, 13(1), pp. 90–98. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17505466%5Cnpapers2://publication/uuid/CCF0C39F-0F6B-4B4E-9602-E24B1F0E1105.

Needleman, S. B. and Wunsch, C. D. (1970) 'A general method applicable to the search for similarities in the amino acid sequence of two proteins', *Journal of Molecular Biology*, 48(3), pp. 443–453. doi: 10.1016/0022-2836(70)90057-4.

Nordström, K. J. V *et al.* (2011) 'Independent HHsearch, Needleman–Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families', *Molecular biology and evolution*. Oxford University Press, 28(9), pp. 2471–2480.

Nygaard, R. *et al.* (2014) 'The dynamic process of β2 -adrenergic receptor activation', *Cell*,

152(3), pp. 532–542. doi: 10.1016/j.cell.2013.01.008.The.

Olson, M. A., Chaudhury, S. and Lee, M. S. (2011) 'Comparison between self-guided langevin dynamics and molecular dynamics simulations for structure refinement of protein loop conformations', *Journal of Computational Chemistry*. Wiley Subscription Services, Inc., A Wiley Company, 32(14), pp. 3014–3022. doi: 10.1002/jcc.21883.

Palczewski, K. *et al.* (2000) 'Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor', *Science*, 289(5480), pp. 739–745. doi: 10.1126/science.289.5480.739.

Pettersen, E. F. *et al.* (2004) 'UCSF Chimera—a visualization system for exploratory research and analysis', *Journal of computational chemistry*. Wiley Online Library, 25(13), pp. 1605–1612.

Phillips, J. C. *et al.* (2005) 'Scalable molecular dynamics with \uppercase{NAMD}', *J. Comput. Chem.* Wiley Online Library, 26(16), pp. 1781–1802. doi: Doi 10.1002/Jcc.20289.

Rasmussen, S. G. F. *et al.* (2011) 'Crystal structure of the β2adrenergic receptor-Gs protein complex', *Nature*, 477(7366), pp. 549–557. doi: 10.1038/nature10361.

Reppert, S. M. *et al.* (1995) 'Molecular characterization of a second melatonin receptor expressed in human retina and brain: the Mel1b melatonin receptor', *Proc Natl Acad Sci U S A*, 92(19), pp. 8734–8738. doi: 10.1073/pnas.92.19.8734.

Rodríguez-Acosta, A. *et al.* (2016) 'Biological and biochemical characterization of venom from the broad-banded copperhead (Agkistrodon contortrix laticinctus): Isolation of two new dimeric disintegrins', *Animal Biology*, 66(2), pp. 173–187. doi: 10.1163/15707563-00002495.

Rosenbaum, D. M., Rasmussen, S. G. F. and Kobilka, B. K. (2009) 'The structure and function of G-protein-coupled receptors', *Nature*. Nature Publishing Group, 459(7245), pp. 356–363. doi: 10.1038/nature08144.

Shamir, E. *et al.* (2000) 'Melatonin improves sleep quality of patients with chronic schizophrenia.', *The Journal of clinical psychiatry*, 61(5), pp. 373–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10847313.

Shi, J., Blundell, T. L. and Mizuguchi, K. (2001) 'FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties1', *Journal of molecular biology*. Elsevier, 310(1), pp. 243–257.

Slaugenhaupt, S. A. *et al.* (1995) 'Mapping of the gene for the mel1a-melatonin receptor to

human chromosome 4 (MTNR1A) and mouse chromosome 8 (MTNR1A)', *Genomics*, 27(2), pp. 355–357. doi: 10.1006/geno.1995.1056.

Smith, T. F. and Waterman, M. S. (1981) 'Identification of common molecular subsequences', *Molecular Biology*, 147(1), pp. 195–197. doi: 10.1016/0022-2836(81)90087-5.

Sriram, K. and Insel, P. A. (2018) 'GPCRs as targets for approved drugs: How many targets and how many drugs?', *Molecular Pharmacology*, p. mol.117.111062. doi: 10.1124/mol.117.111062.

Tan, D. X. *et al.* (2015) 'Melatonin as a potent and inducible endogenous antioxidant: Synthesis and metabolism', *Molecules*. Multidisciplinary Digital Publishing Institute, 20(10), pp. 18886–18906. doi: 10.3390/molecules201018886.

Trott, O. and Olson, A. J. (2010) 'AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading', *Journal of computational chemistry*. Wiley Online Library, 31(2), pp. 455–461.

UniProt Consortium, T. (2018) 'UniProt: the universal protein knowledgebase', *Nucleic Acids Research*. Oxford University Press, 46(5), pp. 2699–2699. doi: 10.1093/nar/gky092.

Vanommeslaeghe, K. *et al.* (2010) 'CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields', *Journal of Computational Chemistry*. Wiley Subscription Services, Inc., A Wiley Company, 31(4), pp. 671–690. doi: 10.1002/jcc.21367.

Wakahara, A., Fujiwara, T. and Tomita, K. (1972) 'The crystal and molecular structure of melatonin, n-acetyl-5-methoxytryptamine', *Chemistry Letters*. The Chemical Society of Japan, 1(11), pp. 1139–1142.

Wu, C. H. (2006) 'The Universal Protein Resource (UniProt): an expanding universe of protein information', *Nucleic Acids Research*. Oxford University Press, 34(90001), pp. D187–D191. doi: 10.1093/nar/gkj161.

Wu, S. and Zhang, Y. (2007) 'LOMETS: A local meta-threading-server for protein structure prediction', *Nucleic Acids Research*, 35(10), pp. 3375–3382. doi: 10.1093/nar/gkm251.

Yang, L. Q. *et al.* (2014) 'Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms', *Journal of Biomolecular Structure and Dynamics*. Taylor & Francis, 32(3), pp. 372–393. doi: 10.1080/07391102.2013.770372.

Zhang, Y. (2008) 'I-TASSER server for protein 3D structure prediction', *BMC Bioinformatics*, 9(1), p. 40. doi: 10.1186/1471-2105-9-40.

Zhang, Y. and Skolnick, J. (2004) 'SPICKER: A clustering approach to identify near-native protein folds', *Journal of Computational Chemistry*. Wiley Subscription Services, Inc., A Wiley Company, 25(6), pp. 865–871. doi: 10.1002/jcc.20011.