

HOW INTRINSIC DYNAMICS AND ALLOSTERY IN G-PROTEIN COUPLED  
RECEPTORS (GPCRS) DEFINE FUNCTIONAL SELECTIVITY?

by

Özge Duman

B.S., Chemical Engineering, Boğaziçi University, 2018

Submitted to the Institute for Graduate Studies in  
Science and Engineering in partial fulfillment of  
the requirements for the degree of  
Master of Science

Graduate Program in Chemical Engineering  
Boğaziçi University

2020

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my thesis supervisor Prof. Türkan Halilođlu for her guidance, support, patience, and friendship. It was a great experience to work with her and I have learned a lot from her wisdom. I would never achieve such success in my studies without her faith in me, continuous support, and understanding.

I wish to express my sincere appreciation for the members of the thesis committee, Asst. Prof. Betül Uralcan and Prof. Özlem Keskin, for accepting to be a member of the thesis committee, devoting their valuable time to read and comment on my thesis.

I would like to give my deepest thanks to Yiđit Kutlu, Büşra Özgüney, Burçin Acar, Zeynep Erge Akbaş Buz and for their continuous support, patience and friendship, especially to Yiđit Kutlu for being there for me in difficult times and keeping me motivated. Also, I would like to thank other members of the PRC family; Gökçehan Kara and Fatih Özhelvacı for their help.

I also thank my dear friends Selma Başıbüyük for helping me edit my thesis and emotional support especially during the pandemic, Fatmanur Tiryaki for her friendships, and emotional support during my thesis period.

Finally, I wish to thank my grandmother, my mother, and my father for their endless love and always supporting me to make it happen.

It was a great opportunity for me that my research has been supported by Tubitak (118Z915).

## ABSTRACT

# HOW INTRINSIC DYNAMICS AND ALLOSTERY IN G-PROTEIN COUPLED RECEPTORS (GPCRS) DEFINE FUNCTIONAL SELECTIVITY?

G-protein Coupled Receptors (GPCRs) are the largest family of the of sensory proteins which play an important role in many diseases related to signaling pathways. Thus, understanding GPCR activation mechanism is a corner stone in drug development. However, the transition pathways between inactive/active states remains not fully explored. In this thesis, a new approach, ANM-LD methodology, that combines the Anisotropic Network Model (ANM) and Langevin Dynamics (LD) is performed to decipher the transition pathways in the activation process and determine the key dynamic factors underlying the conformational changes of GPCR family. Towards this goal, three different receptor families (M2R,  $\beta$ 2AR, and NTSR1) are studied. The predicted transition pathways reveal that the intrinsic dynamic modes of the inactive state enable the conformational changes essential for the activation. As a result, the distances between TM3-TM6 and TM3-TM7 helices and the path of some functionally important motifs describe the major pathway that show commonality and differences in all three GPCRs. The agonist binding pocket interactions differ markedly for different subspecies, which are reflected in the selected dynamic modes of motion. When M2R is compared with NTSR1 and  $\beta$ 2AR, the weaker packing interactions result in more flexible segments in the connector of NTSR1 and  $\beta$ 2AR resulting in the TM3-TM6 distance to decay slower. Although a common agonist binding site does not exist, the conserved DRY, and NPxxY motifs are the common drivers of the GP binding site. The hinges of the most dominant dynamic modes coincide functionally important residues, constitute the dynamic infrastructure of allosteric interactions and enables the communication between the extracellular and intracellular sides of the structure.

## ÖZET

# İÇSEL DİNAMİKLER VE ALLOSTERİ G-PROTEİN KENETLİ RESEPTÖRLERDE (GPCR) FONKSİYONEL SEÇİCİLİĞİ NASIL TANIMLAR?

G-protein Kenetli Reseptörler (GPCR'ler), sinyal yollarıyla ilgili birçok hastalıkta önemli bir rol oynayan en büyük sinyal proteini ailesidir. Bu nedenle, GPCR aktivasyon mekanizmasını anlamak, ilaç geliştirmede bir kilit taşıdır. Ancak, inaktif/aktif durumlar arasındaki geçiş yolları tam olarak keşfedilmemiştir. Bu tezde, aktivasyon sürecindeki geçiş yollarını deşifre etmek ve konformasyonel değişikliklerin altında yatan anahtar dinamik faktörleri belirlemek için Anisotropik Ağ Modeli (ANM) ile Langevin Dinamiklerini (LD) birleştiren yeni bir yaklaşım olan ANM-LD metodolojisi uygulanmaktadır. GPCR ailesi. Bu amaç doğrultusunda, üç farklı reseptör ailesi (M2R,  $\beta$ 2AR, NTSR1) incelenmiştir. Öngörülen geçiş yolları, inaktif durumun içsel dinamik modlarının aktivasyon için gerekli olan konformasyonel değişiklikleri mümkün kıldığını ortaya koymaktadır. Sonuç olarak, TM3-TM6 ve TM3-TM7 sarmalları arasındaki mesafeler ve işlevsel olarak önemli bazı motiflerin patikaları, üç GPCR'nin ortaklığı ve farklılıkları gösteren ana patikayı tanımlar. Agonist bağlama cebi etkileşimleri, farklı alt türler için önemli ölçüde farklılık gösterir ve bu, seçilen dinamik hareket modlarının farklılıklarında yansıtılır. M2R, NTSR1 ve  $\beta$ 2AR ile kıyaslandığında, daha zayıf paketleme etkileşimleri NTSR1 ve  $\beta$ 2AR konektöründe daha esnek segmentlere sebep olur ve bu da TM3-TM6 mesafesinin daha yavaş azalmasına neden olur. Ortak bir agonist bağlanma sahası mevcut olmamasına rağmen, korunan DRY ve NPxxY motifleri, GP bağlanma sahasının ortak yürütücüleridir. En baskın dinamik modların menteşeleri, işlevsel olarak önemli rezidüer ile çakışır, allosterik etkileşimlerin dinamik altyapısını oluşturur ve yapının hücre dışı ve hücre içi tarafları arasındaki iletişimi sağlar.

## LIST OF FIGURES

Figure 1.1.	Schematic representation of muscarinic acetylcholine receptor (PDB ID: 3UON) found in <i>homo sapiens</i> . . . . .	4
Figure 2.1.	The procedure applied in ANM-LD simulations. . . . .	12
Figure 3.1.	The structure and corresponding regions of muscarinic acetylcholine receptors. . . . .	15
Figure 3.2.	The wire diagram of M2 receptor. . . . .	15
Figure 3.3.	Functionally important sites of M2 receptor. . . . .	17
Figure 3.4.	Investigated cases in ANM-LD simulations of M2 receptor. a) Alignment of 4MQS to 3UON(RMSD: 2.66 Å). b) Alignment of 4MQT to 3UON(RMSD: 2.44 Å). . . . .	19
Figure 3.5.	RMSD difference through the simulations of M2 receptor from 3UON to 4MQS . . . . .	21
Figure 3.6.	RMSD difference of selected runs through the simulations of M2 receptor from 3UON to 4MQS . . . . .	22
Figure 3.7.	Functionally important distances through the transition pathway by ANM-LD simulations of 3UON to 4MQS (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	24

Figure 3.8.	Functionally important distances through the transition pathway by ANM-LD simulations of 3UON to 4MQS (final/target structures are shown as diamond, and the distances of snapshots are shown as spheres). . . . .	25
Figure 3.9.	The RMSD of important residues through the activation of M2 receptor from 3UON to 4MQS (final/target structures are shown as diamond, and the RMSD of snapshots are shown as circles). . .	26
Figure 3.10.	Selected Mode Mapping a) DF=0.10-0.80 Å All Modes $R_{cut}$ =13 Å. b) DF=0.10-1.00 Å All Modes $R_{cut}$ =13 Å. c)DF=0.35 Å Modemax=100 $R_{cut}$ =13 Å. d)DF=0.40 Å Modemax=100 $R_{cut}$ =13 Å. .	28
Figure 3.11.	ANM mode frequency of parallel simulations from 3UON to 4MQS	30
Figure 3.12.	ANM mode 1 shape of the inactive state of M2 receptor . . . . .	31
Figure 3.13.	ANM mode 5 shape of the inactive state of M2 receptor . . . . .	31
Figure 3.14.	Demonstration of the minimum fluctuating residues, the hinge residues of ANM mode 1 and 5 (dark blue and sphere) on the structure. a) Hinges of ANM mode 1 and 5. b) Hinges of ANM mode 1. c) Hinges of ANM mode 5. . . . .	32
Figure 3.15.	Demonstration of ANM mode 1 and 5 hinges (dark blue), functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a)ANM mode 1 and 5 b)ANM mode 1 c)ANM mode 5. . . . .	33
Figure 3.16.	Cross correlation of residues in the activation pathway of M2 receptor from 3UON to 4MQS. . . . .	34

Figure 3.17. RMSD difference through the simulations of M2 receptor from 3UON to 4MQT. . . . .	36
Figure 3.18. RMSD difference of selected runs through the simulations of M2 receptor from 3UON to 4MQT. . . . .	37
Figure 3.19. Functionally important distances through the conformation transition/activation pathway obtained from ANM-LD results of 3UON to 4MQT (initial and final/target structures shown as diamond, snapshots shown as circles). . . . .	38
Figure 3.20. RMSD of important motifs through the activation of M2 receptor from 3UON to 4MQT (The RMSD of final/target structures are shown as diamond, and the RMSD of snapshots obtained from ANM-LD simulations are shown as circles). . . . .	39
Figure 3.21. RMSD of important residues through the activation of M2 receptor from 3UON to 4MQT (The RMSD of final/target structures are shown as diamond, and the RMSD of snapshots obtained from ANM-LD simulations are shown as circles). . . . .	41
Figure 3.22. Selected Mode Mapping a)DF=0.10-0.80 Å All Modes $R_{cut}$ =13 Å. b) DF=0.60 Å Modemax=100 $R_{cut}$ =13 Å. c) DF=0.40 Å Modemax=100 $R_{cut}$ =13 Å. d) DF=0.35 Å Modemax=100 $R_{cut}$ =13 Å. .	42
Figure 3.23. ANM mode frequency of parallel simulations from 3UON to 4MQT.	44
Figure 3.24. ANM mode 2 shape of the inactive state of M2 receptor. . . . .	45
Figure 3.25. ANM mode 3 shape of the initial state of M2 receptor. . . . .	46

Figure 3.26.	Demonstration of hinge residues of ANM mode 2 and 3 (dark blue and sphere) on the structure. a) Hinges of ANM mode 2 and 3. b) Hinges of ANM mode 2. C) Hinges of ANM mode 3. . . . .	47
Figure 3.27.	Demonstration of ANM mode 2 and 3 hinges (dark blue), functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a)ANM mode 2 and 3 b)ANM mode 2 c)ANM mode 3. . . . .	48
Figure 3.28.	Difference vector between initial (3UON) and final (4MQS – 4MQT) structure. . . . .	49
Figure 3.29.	Comparison of selected ANM modes in two cases. a) ANM mode 1 and ANM mode 2 b) ANM mode 1 and ANM mode 3 c) ANM mode 5 and ANM mode 2 b) ANM mode 5 and ANM mode 3. . .	51
Figure 3.30.	Cross correlation map of the activation pathway of M2 receptor from 3UON to 4MQT. . . . .	52
Figure 3.31.	Snapshots generated by ANM-LD simulations from inactive (3UON) to active (4MQS) structure. . . . .	53
Figure 3.32.	Snapshots generated by ANM-LD simulations from inactive (3UON) to active (4MQT) structure. . . . .	53
Figure 3.33.	Important residues (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, T386, and E382) (dark blue and sphere) demonstrated on the snapshots of the simulations from 3UON to 4MQS. . . . .	55

Figure 3.34.	The TM3-TM6 distance between R121 (dark blue and sphere) and T386 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQS. . . . .	56
Figure 3.35.	The TM3-TM7 distance between R121 (dark blue and sphere) and Y440 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQS. . . . .	56
Figure 3.36.	Important residues (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, T386, and E382) (dark blue and sphere) demonstrated on the snapshots of simulations from 3UON to 4MQT. . . . .	57
Figure 3.37.	TM3-TM6 distance between R121 (dark blue and sphere) and T386 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQT. . . . .	58
Figure 3.38.	TM3-TM7 distance between R121 (dark blue and sphere) and Y440 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQS. . . . .	58
Figure 3.39.	Alignment of intermediate structures obtained by ANM-LD simulations of both cases. . . . .	59
Figure 3.40.	Alignment of final structures obtained by ANM-LD simulations of both cases. . . . .	60
Figure 3.41.	Conformational changes in ECL2 allosteric site through the transition pathways generated by ANM-LD simulations. . . . .	63
Figure 3.42.	The structure and corresponding regions of $\beta$ 2-Adrenergic receptor.	65

Figure 3.43.	The wire diagram of corresponding regions and residue numbers of $\beta$ 2-Adrenergic receptor. . . . .	65
Figure 3.44.	Functionally important sites of $\beta$ 2-Adrenergic receptor. . . . .	67
Figure 3.45.	Investigated cases in ANM-LD simulations of $\beta$ 2-Adrenergic receptor. a)3SN6 to 2RH1 (RMSD: 2.93 Å).b)3P0G to 2RH1 (RMSD: 2.61 Å).c)3SN6 to 3D4S (RMSD: 3.22 Å).d)3P0G to 3D4S (RMSD: 2.71 Å). . . . .	69
Figure 3.46.	The RMSD of $\beta$ 2 adrenergic receptor from 2RH1 to 3SN6 in the ANM-LD simulations. . . . .	71
Figure 3.47.	The RMSD of selected simulations of $\beta$ 2 adrenergic receptor from 2RH1 to 3SN6 in the ANM-LD simulations. . . . .	72
Figure 3.48.	Functionally important distances through the transition pathway obtained from ANM-LD results of 2RH1 to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	73
Figure 3.49.	The RMSD of important motifs through the activation of M2 receptor from 3UON to 4MQS (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	75
Figure 3.50.	The RMSD of important residues through the activation of $\beta$ 2 adrenergic receptor from 2RH1 to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	76

Figure 3.51. Selected Mode Mapping of $\beta$ 2AR simulations from 2RH1 to 3SN6 a) DF=0.60 Å Modemax=30 $R_{cut}$ =8. b)DF=0.40 Å Modemax=30 $R_{cut}$ =13. c) DF=0.60 Å Modemax=30 $R_{cut}$ =13. d) DF=0.10-0.40 Å All Modes $R_{cut}$ =13 . . . . .	78
Figure 3.52. ANM mode frequency of parallel simulations from 2RH1 to 3SN6.	80
Figure 3.53. Difference vector between initial (2RH1) and final (3SN6) structure.	81
Figure 3.54. Mean square fluctuations of selected ANM modes of $\beta$ 2AR simulations from 2RH1 to 3SN6. . . . .	81
Figure 3.55. ANM mode 11 shape of the initial state of $\beta$ 2 adrenergic receptor.	82
Figure 3.56. ANM mode 7 shape of the initial state of $\beta$ 2 adrenergic receptor. .	83
Figure 3.57. ANM mode 24 shape of the initial state of $\beta$ 2 adrenergic receptor.	84
Figure 3.58. Demonstration of hinge residues of ANM mode 7, 11, and 24 (dark blue and sphere) on the structure a) Hinges of ANM mode 7, 11, and 24. b) Hinges of ANM mode 7 c) Hinges of ANM mode 11. d) Hinges of ANM mode 24. . . . .	85
Figure 3.59. Demonstration of ANM mode 7, 11, and 24 hinges (dark blue) and important regions (red and sphere), and the intersection (dark blue and sphere). a)ANM modes 7, 11, and 24 b)ANM mode 7 c)ANM mode 11 d)ANM mode 24. . . . .	86
Figure 3.60. Cross correlation map of the activation pathway of $\beta$ 2 adrenergic receptor from 2RH1 to 3SN6. . . . .	87

Figure 3.61.	Alignment of active structure (3P0G) used in this case with the active structure (3SN6) used in the first case (RMSD: 1.55 Å). . . . .	88
Figure 3.62.	RMSD difference through the simulations of $\beta 2$ adrenergic receptor from 2RH1 to 3P0G. . . . .	90
Figure 3.63.	RMSD difference of selected runs through the simulations of $\beta 2$ adrenergic receptor from 2RH1 to 3P0G. . . . .	91
Figure 3.64.	Functionally important distances through the transition pathway obtained from ANM-LD results of 2RH1 to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	92
Figure 3.65.	RMSD of important motifs through the activation of $\beta 2$ adrenergic receptor from 2RH1 to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	94
Figure 3.66.	RMSD of important residues through the activation of $\beta 2$ adrenergic receptor from 2RH1 to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	95
Figure 3.67.	Selected Mode Mapping of $\beta 2$ AR simulations a)DF=0.60 Å Modemax=30 $R_{cut}$ =8 Å. b)DF=0.60 Å Modemax=100 $R_{cut}$ =8 Å. c)DF=0.60 Å Modemax=30 $R_{cut}$ =13 Å. d)DF=0.10-0.40 Å All Modes $R_{cut}$ =13 Å. . . . .	96
Figure 3.68.	ANM mode frequency of parallel simulations from 2RH1 to 3P0G. . . . .	98
Figure 3.69.	The difference vector between initial (2RH1) and final (3SN6 – 3P0G) structures. . . . .	99

Figure 3.70.	ANM mode 9 shape of the initial state of $\beta 2$ adrenergic receptor. .	100
Figure 3.71.	Demonstration of hinge residues of ANM mode 9 (dark blue and sphere) on the structure. . . . .	101
Figure 3.72.	Demonstration of ANM mode 9 (dark blue) and functionally important regions (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, L272, and E268) (red and sphere) on the structure. . . . .	102
Figure 3.73.	Cross correlation map of the activation pathway of $\beta 2$ adrenergic receptor from 2RH1 to 3P0G. . . . .	103
Figure 3.74.	Alignment of inactive structure (3D4S) used in this case with the inactive structure (2RH1) used in the first case (RMSD: 0.53 Å). .	104
Figure 3.75.	The RMSD difference through the simulations of $\beta 2$ adrenergic receptor from 3D4S to 3SN6. . . . .	107
Figure 3.76.	The RMSD difference of selected runs through the simulations of $\beta 2$ adrenergic receptor from 3D4S to 3SN6. . . . .	108
Figure 3.77.	Functionally important distances through the transition pathway obtained from ANM-LD results of 3D4S to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres). .	110
Figure 3.78.	RMSD of important motifs through the activation of $\beta 2$ adrenergic receptor from 3D4S to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	111

Figure 3.79. RMSD of important residues through the activation of $\beta 2$ adrenergic receptor from 3D4S to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	112
Figure 3.80. Selected Mode Mapping a)(NMD) DF=0.40 Å Modemax=100 $R_{cut}$ =13 b)(NMD) DF=0.60 Å Modemax=30 $R_{cut}$ =8 c)DF=0.35 Å Modemax=30 $R_{cut}$ =8 d)DF=0.40 Å Modemax=100(R12) $R_{cut}$ =13. . . .	114
Figure 3.81. ANM mode frequency of parallel simulations from 3D4S to 3SN6.	116
Figure 3.82. Difference vector between initial (2RH1 - 3D4S) and final (3SN6) structure. . . . .	117
Figure 3.83. Mean square fluctuations of selected ANM modes of $\beta 2$ AR simulations from 3D4S to 3SN6. . . . .	118
Figure 3.84. ANM mode 1 shape of the initial state of $\beta 2$ adrenergic receptor. .	119
Figure 3.85. ANM mode 4 shape of the initial state of $\beta 2$ adrenergic receptor. .	120
Figure 3.86. ANM mode 9 shape of the initial state of $\beta 2$ adrenergic receptor. .	121
Figure 3.87. Demonstration of hinge residues of ANM mode 1, 4, and 9 (dark blue and sphere) on the structure a) Hinges of ANM mode 1, 4, and 9. b) Hinges of ANM mode 1. c) Hinges of ANM mode 4. d) Hinges of ANM mode 9. . . . .	122

Figure 3.88. Demonstration of selected ANM modes 1, 4, and 9 hinges (dark blue) functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a) ANM mode 1, 4, and 9 with b) ANM mode 1 with c) ANM mode 4 with d) ANM mode 9 with. . . . .	123
Figure 3.89. Cross correlation map of the activation pathway of $\beta 2$ adrenergic receptor from 3D4S to 3SN6. . . . .	124
Figure 3.90. RMSD difference through the simulations of $\beta 2$ adrenergic receptor from 3D4S to 3P0G. . . . .	127
Figure 3.91. RMSD difference of selected runs through the simulations of $\beta 2$ adrenergic receptor from 3D4S to 3P0G. . . . .	128
Figure 3.92. Functionally important distances through the transition pathway obtained from ANM-LD results of 3D4S to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	129
Figure 3.93. RMSD of important motifs through the activation of $\beta 2$ adrenergic receptor from 3D4S to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	131
Figure 3.94. RMSD of important residues through the activation of $\beta 2$ adrenergic receptor from 3D4S to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	132
Figure 3.95. Selected Mode Mapping of $\beta 2$ AR simulations a) DF=0.10-0.60 Å All Modes $R_{cut}$ =10 Å. b) DF=0.60 Å Modemax=30 $R_{cut}$ =10. c) DF=0.35 Å Modemax=100 $R_{cut}$ =10 Å. d) DF=0.35 Å Modemax=100 $R_{cut}$ =13 Å. . . . .	134

Figure 3.96. ANM mode frequency of parallel simulations from 3D4S to 3P0G.	136
Figure 3.97. Difference in square vector distance between initial (2RH1-3D4S) and final (3P0G) structure. . . . .	137
Figure 3.98. Difference vector between initial (3D4S) and final (3SN6 - 3P0G) structure. . . . .	138
Figure 3.99. Mean square fluctuations of selected ANM modes of $\beta$ 2AR simulations from 3D4S to 3P0G. . . . .	139
Figure 3.100. ANM mode 1 shape of the initial state of $\beta$ 2 adrenergic receptor. .	140
Figure 3.101. ANM mode 2 shape of the initial state of $\beta$ 2 adrenergic receptor. .	140
Figure 3.102. Demonstration of hinge residues of ANM mode 1 and 2 (dark blue and sphere) on the structure. a) Hinges of ANM mode 1 and 2. b) Hinges of ANM mode 1. c) Hinges of ANM mode 2. . . . .	141
Figure 3.103. Demonstration of ANM mode 1 and 2 hinges (dark blue), functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a) ANM mode 1 and 2 b) ANM mode 1 c) ANM mode 2. . . . .	142
Figure 3.104. Cross correlation map of the activation pathway of $\beta$ 2 adrenergic receptor from 3D4S to 3SN6. . . . .	143
Figure 3.105. Snapshots generated by ANM-LD simulations from inactive (2RH1) to active (3SN6) structure. . . . .	144

Figure 3.106. Snapshots generated by ANM-LD simulations from inactive (2RH1) to active (3P0G) structure. . . . .	145
Figure 3.107. Snapshots generated by ANM-LD simulations from inactive (3D4S) to active (3SN6) structure. . . . .	145
Figure 3.108. Snapshots generated by ANM-LD simulations from inactive (3D4S) to active (3P0G) structure. . . . .	146
Figure 3.109. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3SN6. . . . .	147
Figure 3.110. The TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3SN6 . . . . .	148
Figure 3.111. The TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3SN6 . . . . .	149
Figure 3.112. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3P0G. . . . .	150
Figure 3.113. TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3P0G. . . . .	151

Figure 3.114. TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3P0G. . . . .	151
Figure 3.115. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6. . . . .	152
Figure 3.116. TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6. . . . .	153
Figure 3.117. TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6. . . . .	153
Figure 3.118. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 3D4S to 3P0G. . . . .	154
Figure 3.119. TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6. . . . .	155
Figure 3.120. TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3P0G. . . . .	155

Figure 3.121. Alignment of intermediate structures obtained by ANM-LD simulations of the simulations in which 2RH1 is used as inactive structure.	156
Figure 3.122. Alignment of final structures obtained by ANM-LD simulations of the simulations in which 2RH1 is used as inactive structure. . . . .	157
Figure 3.123. The structure and corresponding regions of NTSR1. . . . .	160
Figure 3.124. The wire diagram of NTS1 receptor. . . . .	161
Figure 3.125. Functionally important sites of NTSR1. . . . .	163
Figure 3.126. Alignment of active structure (5T04) to inactive structure (3ZEV) (RMSD: 2.28 Å). . . . .	164
Figure 3.127. The RMSD difference through the simulations of NTS1R from 3ZEV to 5T04. . . . .	166
Figure 3.128. The RMSD difference of selected runs through the simulations of NTS1R from 3ZEV to 5T04. . . . .	167
Figure 3.129. Functionally important distances through the transition pathway by ANM-LD simulations of 3ZEV to 5T04 (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	168
Figure 3.130. The RMSD of important motifs through the activation of NTSR1 from 3ZEV to 5T04 (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	169

Figure 3.131. RMSD of important residues through the activation of NTSR1 receptor from 3ZEV to 5T04 (initial and final/target structures shown as diamond, snapshots shown as spheres). . . . .	170
Figure 3.132. Selected Mode Mapping of NTS1R simulations from 3ZEV to 5T04 a) DF=0.10-1.00 Å All Modes $R_{cut}=10$ . b) DF=0.60 Å Mode-max=30 $R_{cut}=13$ . . . . .	171
Figure 3.133. ANM mode frequency of parallel simulations from 3ZEV to 5T04.	172
Figure 3.134. Difference vector between initial (3ZEV) and final (5T04) structure.	173
Figure 3.135. Mean square fluctuations of selected ANM modes of NTS1R simulations from 3ZEV to 5T04. . . . .	174
Figure 3.136. ANM mode 1 shape of the inactive state of NTS1R. . . . .	175
Figure 3.137. ANM mode 2 shape of the inactive state of NTS1R. . . . .	175
Figure 3.138. ANM mode 5 shape of the inactive state of NTS1R. . . . .	176
Figure 3.139. ANM mode 8 shape of the inactive state of NTS1R. . . . .	177
Figure 3.140. Demonstration of hinge residues of ANM mode 1, 2, 5, and 8 (dark blue and sphere) on the structure a)ANM modes 1, 2, 5, and 8 b)ANM mode 1. c)ANM mode 2 d)ANM mode 5 e)ANM mode 8.	178
Figure 3.141. Demonstration of ANM modes 1, 2, 5, and 8 hinges (dark blue),important regions (red and sphere),the intersection (dark blue and sphere) a)Modes 1, 2, 5, and 8 b)Mode 1 c) Mode 2 d)Mode 5 e)Mode 8. .	179

Figure 3.142. Cross correlation map of the activation pathway of NTS1R from 3ZEV to 5T04. . . . .	180
Figure 3.143. Snapshots generated by ANM-LD simulations from inactive (3ZEV) to active (5T04) structure. . . . .	182
Figure 3.144. Important residues (DRY Motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L303) (dark blue and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3ZEV to 5T04. . . . .	182
Figure 3.145. TM3-TM6 distance between R167 (dark blue and sphere) and L303 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3ZEV to 5T04. . . . .	183
Figure 3.146. TM3-TM7 distance between R167 (dark blue and sphere) and Y369 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3ZEV to 5T04. . . . .	183
Figure 3.147. Alignment of inactive structures of M2R, $\beta$ 2AR, and NTSR1. . . . .	184
Figure 3.148. Multiple Sequence Alignment performed on inactive structures of M2R, $\beta$ 2AR, and NTSR1. . . . .	186
Figure 3.149. Comparison of collective variable analysis through the transition pathway of M2R, $\beta$ 2AR, and NTS1R. . . . .	187

## LIST OF TABLES

Table 3.1.	Residues and corresponding regions of functionally important sites of M2 receptor. . . . .	16
Table 3.2.	Summary of used structure in ANM-LD simulations of M2 receptor.	18
Table 3.3.	ANM-LD Result Summaries . . . . .	20
Table 3.4.	ANM-LD Result Summaries. . . . .	35
Table 3.5.	Key collaborating residues in allosteric sites based on ANM mode and cross correlation analysis. . . . .	62
Table 3.6.	Residues and corresponding regions of functionally important sites of $\beta$ 2-Adrenergic receptor. . . . .	66
Table 3.7.	Summary of used structure in ANM-LD simulations of $\beta$ 2-Adrenergic receptor. . . . .	68
Table 3.8.	ANM-LD Result Summaries. . . . .	70
Table 3.9.	ANM-LD Result Summaries. . . . .	89
Table 3.10.	ANM-LD Result Summaries. . . . .	106
Table 3.11.	ANM-LD Result Summaries. . . . .	126
Table 3.12.	Residues and corresponding regions of functionally important sites of NTS1R. . . . .	162

Table 3.13.	Summary of used structure in ANM-LD simulations of NTS1 receptor.	164
Table 3.14.	ANM-LD Result Summaries. . . . .	165
Table 3.15.	Similarity Matrix (in percent) of inactive structures of M2R, $\beta$ 2AR, and NTSR1. . . . .	185
Table 3.16.	Identity Matrix (in percent) of inactive structures of M2R, $\beta$ 2AR, and NTSR1. . . . .	185

## LIST OF SYMBOLS

$a_i$	Acceleration of particle i
$D_F$	Deformation factor
$d_i$	Distance between the coordinates of atom i in two structures
$D_i$	Distance difference vector
$F_i$	Force exerted on particle i
H	Hessian matrix
$k_B$	Boltzmann constant
$m_i$	Mass of particle i
$N$	Number of atoms
$O_{ij}$	Overlap value of $j^{th}$ eigenvector with the difference vector for cycle i
$R$	Random Forces
$R_{ij}^0$	Equilibrium distance between residues i and j
$R_{cut}$	Cut-off radius
$R_i$	Position vector of residue i
$R_{ij}$	Instantaneous distance between residues i and j
$step_{min}$	Number of minimization steps
$step_{sim}$	Number of simulation steps
t	Time
T	Absolute temperature
$u_k$	$k^{th}$ eigenvector
V	Potential function
$\Delta R_i$	Fluctuation of residue i
$\gamma_f$	Force constant for harmonic spring
$\lambda_k$	$k^{th}$ eigenvalue
$v_i$	Velocity of atom i

## LIST OF ACRONYMS/ABBREVIATIONS

ANM	Anisotropic Network Model
ANM-LD	Anisotropic Network Model Guided Langevin Dynamics
ENM	Elastic Network Model
GNM	Gaussian Network Model
GPCRs	Guanine Binding Protein Coupled Receptors
ICL	Intracellular Loop
LD	Langevin Dynamics
MD	Molecular Dynamics
MSF	Mean Square Fluctuations
PDB	Protein Data Bank
RMSD	Root-Mean-Square Deviation
Sander	Simulated Annealing with NMR-Derived Energy Restraints
SF	Square Fluctuations
TM	Transmembrane

## 1. INTRODUCTION

Guanin-binding protein coupled receptors (GPCRs) are the largest family of sensory proteins of eukaryotes, including families of glutamate, rhodopsin, Frizzled/Taste2, Adhesion and Secretin according to the GRAFS classification system [1]. These proteins are membrane proteins responsible for triggering intracellular signaling cascades.

There are more than 800 membrane proteins in humans, and they play a role in the formation of many diseases related to signaling pathways (cancer, obesity, sensory disorders, immunological, viral infections, cardiovascular and inflammatory diseases, etc.). GPCR targeted drugs constitute one-third of the drugs in the pharmaceutical market, so they have become the most popular target class for drug discovery [2]. Drug binding to GPCRs may alter the intracellular signaling profile, induce, block, or modulate signaling effects.

The function of GPCRs depends on how they move, in other words, their ability to act in creating different conformational and dynamic states. In this way, they transmit signals by conformational changes in the structure [1]. Extracellular ligands bounded to the extracellular side of GPCR allows G-protein to bind to the intracellular side of GPCR. This allosteric mechanism is quite complex, where the ligands can stimulate alternative intracellular signaling pathways in different ways. Many GPCRs have multiple ligand binding sites. Different ligands can generate different dynamic responses on the same GPCR system. These behaviors can change the target system to which the receptor binding to G-protein and transmits the signal.

Molecular simulations complement X-ray crystallography and NMR-spectroscopy studies, provide detailed information on the dynamic behavior of the protein [1,3]. The crystalline structures generally represent only one stable conformational state of the protein, and the dynamic information obtained from the NMR spectrum often includes the most frequently observed dynamic conditions [3,4]. However, biomolecules visit a

number of conformational and dynamic states in a functional process, and a significant proportion of these cases are not stable to determine experimentally. Therefore, in order to solve the molecular mechanism of signal transduction, and the ligand-partner protein selectivity observed in this process by GPCRs, it is necessary to study the dynamic behavior from small movements (femtosecond-nanosecond) to large conformational (millisecond and beyond) changes [1]. There are online platforms such as GPCRmd [5] which allows the researchers to visualize, investigate and share the MD simulation results of GPCRs. Highly conserved GPCR regions, ligand binding site, toggle switch, ionic lock, PIF motif, NPxxY motif are investigated through the MD trajectory. Additionally, hydrogen bonding network, sodium-ion interaction to orthosteric ligand binding site and allosteric ligand binding site are demonstrated for 26 receptor subtypes of class A [6]. The water bridge formation considered as an indicator of GPCR activation is demonstrated via water mediated interactions. Chen et al. demonstrated the allosteric signal transduction mechanism with MD simulations in their study. They have brought to the literature that microswitches, which show significant conformational changes during activation, form an allosteric network between the intracellular domain and the extracellular domain [7].

Molecular dynamics (MD) simulations, especially at atomic level, have become more powerful in recent years with the increase of computer power and simulation algorithms and the subtle potential energy functions that better represent the basic physics. Starting from the static experimental structure, these simulations predict the movement of each atom in a receptor and the movement of each atom in the molecules in which the receptor interacts. However, MD simulations are known to be inadequate to exemplify movements involving a high number of atoms and corresponding to long-term structural changes where classical MD simulations have limitations [1].

GPCRs are in many different conformational and dynamic states, including active and inactive states to perform their functions. Conformations close to the minimum energy in equilibrium can be obtained experimentally because they are stable and long lasting. Nevertheless, the conformation between these stable states is unstable and

short-lived. The conformational transition pathways with unstable conformations are the difficult to discover with experimental techniques. However, transition pathways are important to quest activation mechanism and dynamics at molecular level. In this thesis, it is aimed to study the dynamic behavior of GPCR systems using Anisotropic Network Model Guided Langevin Dynamics (ANM-LD) methodology [8]. ANM-LD methodology is an effective hybrid methodology which combines intrinsic dynamics of ANM and external forces via Langevin Dynamics. In ANM-LD methodology, two states of the structure are required, and, in each step, generated structure is perturbed with the selected ANM mode until the RMSD converges between the iteratively created structure and the target structure.

As mentioned above, GPCRs are divided into six classes (Class A-F) according to the classical A-F classification system. Class A, B and C constitute the main classes, and class A is the largest of the classes of Rhodopsin-like receptors [9]. There are more than 800 G-protein coupled receptors in homo sapiens alone, and according to GPCR database information, there are more than 15000 receptors known in living organisms [10]. Although there are so many receptors, all GPCRs have similar three-dimensional structures. All receptors have a transmembrane domain consisting of seven  $\alpha$ -helices connected by three extracellular and three intracellular loops, which are necessary for the transduction of a signal across the cell membrane as seen in Figure 1.1 [11,12]. These seven helices are named TM1-TM7 (transmembrane helices), extracellular loops ECL1-ECL3 (extracellular loops), and intracellular loops ICL1-ICL3 (Intracellular loops).

The regions in which GPCRs exhibit large conformational changes when activated are similar, but the binding of a ligand to GPCR may affect the dynamics of GPCR in various ways; It can change the time spending on each of the conformational states of the GPCR and increase or decrease the rate of transitions between different conformational states. On the other hand, binding of an intracellular partner (eg, a G protein or arrestin), dimerization with another GPCR, post-translational modification, or a change in pH / lipid composition may also have similar effects [13,14].

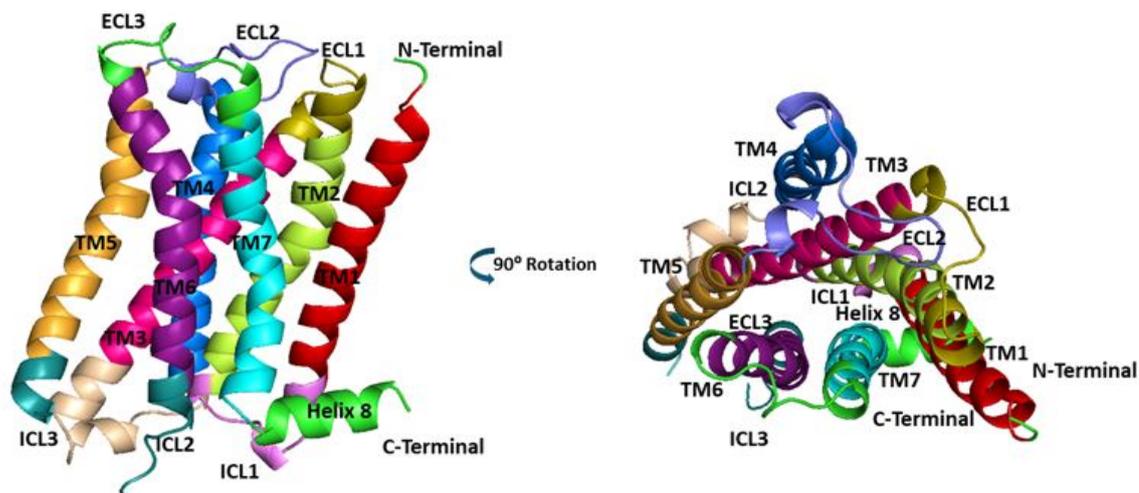


Figure 1.1. Schematic representation of muscarinic acetylcholine receptor (PDB ID: 3UON) found in *homo sapiens*.

The dynamics of the GPCR throughout its life will reflect a combination of structural changes due to the internal dynamics of the receptor and its external perturbations. Global conformational changes that affect the entire structure seem to be particularly important. It is taught that the transmembrane helices, in particular the helices TM 5-7, play a critical role in transmitting a signal across the membrane [1]. The transmembrane helices of the GPCR can adopt not only a single active state and a single inactive state, but also a large number of different conformational states having significant effects for receptor signaling. Conformational changes tend to be the most at the intracellular side of the GPCR, while the extracellular region of the receptor also shows conformational changes. The ligand binding region is associated with conformational changes of the intracellular side, but the structure has undergone significant changes in the extracellular side.

Conformational change in GPCR allows an extracellular signal to be transmitted across the cell membrane to form an intracellular response. Such a conformational change is necessary for GPCR activation in which the GPCR assumes a G protein-binding conformational state. For example, the  $\beta$ 2-adrenergic receptor ( $\beta$ 2AR) binding

to the G protein shows the greatest conformational changes during activation at the intracellular side (TM5-TM7) by rearrangements in the ligand binding pocket (TM5-TM6). An allosteric residue network between the TM3, TM5, TM6 and TM7 regions serves as the center for detecting and responding to conformational change in the  $\beta$ 2AR [15, 16]. Binding of an agonist to the ligand binding pocket induces binding of the intracellular G protein. Conversely, G protein binding will also enhance agonist binding; that is, an agonist binds with a higher interest to a GPCR in the active state [16]. On the other hand, MD simulations show that the intracellular mapping interface is in equilibrium between active and inactive conformation even in the absence of agonist [15]. The presence of a bound agonist helps to stabilize the active conformation of the binding pocket.

GPCRs in both inactive and active states appear to show similar conformational changes between different subspecies [1]. However, the activation of different GPCRs varies in various ways [15]. The agonist binding pocket interactions differ markedly between GPCRs, indicating that different GPCRs have evolved to recognize very different ligands. The ligand-binding sites of peptide receptors such as opioid receptors are shallower than aminergic receptors, since peptide particles are much larger molecules than endogenous ligands [17]. Zhau et al used 45 different receptors in their study on the activation of class A receptors [18]. As a result, rearrangement of 34 conserved residue pairs during activation is observed via residue-residue contact score. These determined residues overlap with the CWxP, PIF, sodium-ion pocket, NPxxY, and DRY motif. Using six different receptors, including muscarinic acetylcholine receptor and  $\beta$ 2-Adrenergic receptor, a common activation pathway has been brought to the literature. According to this proposed activation pathway, a common ligand-residue contact is not found. This is the biggest indicator of the selective binding of subtypes of GPCRs [19–21]. However, as a result of binding the agonist, the receptor undergoes a similar conformational change and reaches the conformational state of active structure. This allosteric interaction starts with the CWxP motif, continues with the sodium-ion pocket, PIF motif and NPxxY motif, and finally, rearrangement is seen in the DRY motif and the activation of the G-protein is accomplished.

The aim of this thesis is to study the structural dynamics of homologous proteins in sub-types of the GPCR family, the relationship between the dynamic behavior in a given time scale and the activation mechanism. The structural and dynamic foundations of the allosteric mechanism underlying this dynamic behavior can be studied in the context of the differences and the common characteristics between the GPCR family members by means of functional selectivity. For this purpose, it is aimed to determine the effect of intrinsic dynamics of GPCR systems, and how the dynamic potential of the structure generalizes and differentiates these structural transitions between different ligands and homologs. In this direction, it is aimed to explore the conformational transition paths of three different receptor families (muscarinic acetylcholine receptor,  $\beta$ 2-adrenergic receptor, neurotensin 1 receptor) with known inactive and active structures; the activation pathways, the most dominant dynamic modes that are essential, allosteric interactions embedded in the pathways, and the selective signal transduction mechanism and dynamic factors that determine selectivity in GPCRs between homologous proteins of the GPCRs.

## 2. MATERIALS AND METHODS

### 2.1. Elastic Network Models and Normal Mode Analysis

Always place some text after headings before putting a graphics into a section as seen in Proteins are not stable in their steady state and can undergo random changes in dynamic equilibrium states (global minimums), bringing together different conformational assemblages around these equilibrium states. These changes are associated with the collective, cooperative, and global movements that bring out the functions of proteins. Functionally important movements can be predicted by solving the topologies of motion modes with Normal Mode Analysis (NMA). These collective, cooperative, and global movements are often associated with slow modes. Elastic Network Models (ENMs) are simplified normal mode analyzes in which amino acids are defined as nodes connected to their neighbors by harmonic springs. In ENMs, detailed amino acid potentials are not important, rather one-dimensional, general harmonic movements between interacting atoms are considered. All of these; It provides computational efficiency to ENMs and helps to discover the intrinsic dynamic behaviors of proteins that are necessary for them to perform their functions by proposing a specific and realistic solution for each protein to identify conformational changes [22–25]. Gaussian Network Model (GNM) and Anisotropic Network Model (ANM) [26, 27] are the two most important ENMs Models [22]. These two different network approaches are used in this thesis.

#### 2.1.1. Anisotropic Network Model

Anisotropic Network Model (ANM) is one of the three-dimensional ENM in which directional preferences are taken into consideration by considering real-life anisotropic changes of proteins instead of isotropic changes [26–28]. Anisotropic variations may contain the components of the position vector independently. Therefore, GNM has  $N-1$  mode, while ANM has  $3N-6$  mode. The main gain of ANM is the ability to create alternative conformations in the immediate neighbors of the given structure

by deformation of the initial structures along the collective modes by using difference vectors [22, 29]. The entire potential of the system is calculated by Equation 2.1 given below. In this equation,  $R_{ij}^0$  is the instantaneous distance between residues  $i$  and  $j$ , and  $\gamma_f$  is the force constant.

$$V_{ANM} = \frac{1}{2} \sum_{ij} \gamma_{f,ij}(R_{ij} - R_{ij}^0) \quad (2.1)$$

Changes of equilibrium positions of residue pairs  $i$  and  $j$   $\langle R_{ij} - R_{ij}^0 \rangle$  and the correlation between them is calculated by decomposition of 3N-6 mode of Hessian Matrix (H) and then reconstruction of this matrix. The Hessian Matrix is distorted by the displacement of the mode  $k$ , and the correlation is calculated with Equation 2.2.

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{3k_B T}{\gamma} \text{tr}[H^{-1}] = \frac{3k_B T}{\gamma_f} \sum_{k=1}^N \text{tr}[\lambda_k^{-1} u_k u_k^T]_{ij} \quad (2.2)$$

## 2.2. Molecular Dynamic Simulations and Langevin Dynamics

Molecular dynamic (MD) simulation is a computer aided simulation method to study the dynamics of atoms and molecules. In this method, Newton's equation of motion is solved iteratively. To be able to perform MD simulations, the position vectors and energy function are required. The position vector of protein's initial structure can be obtained by NMR spectroscopy or X-ray crystallography. The force field is used to determine the applied force on each atom. Since time intervals are generated, new positions of atoms and molecules can be determined by using Newton's 2<sup>nd</sup> law for each time interval and the procedure is continues iteratively, until the target structure is obtained [30, 31].

Newton's 2<sup>nd</sup> law is the basis for the analysis of molecular dynamic simulations and given in Eq.2.3. In this equation, m is mass and a is the acceleration of particle i.

$$F_i = m_i a_i \quad (2.3)$$

The acceleration can be calculated with Equation 2.4. by taking the second derivative of the position vector. The force acting on the molecule can be written as the gradient of the potential energy, which is given in Equation 2.5.

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \quad (2.4)$$

$$F_i = -gradV_i = -\Delta V_i \quad (2.5)$$

Newton's equation of motion is rearranged as a function of time. In this rearrangement, the gradient of potential energy is written instead of force acting on the particle and Equation 2.6 is obtained.

$$\frac{d^2 r_i}{dt^2} = -\frac{1}{r_i} \frac{\partial V(r_1, r_2, \dots, r_N)}{\partial r_i} = a_i \quad (2.6)$$

Maxwell-Boltzmann distribution is used to determine the acceleration profile of the atoms in the initial position. For this purpose, Equation 2.7 and Equation 2.8 is used. In this equation, T is absolute temperature in Kelvin; v is the velocity of the i atom in the x direction, m is the mass of the particle i;  $k_b$  is Boltzmann constant.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T}\right] \quad (2.7)$$

$$p = \sum_{i=1}^N m_i v_i = 0 \quad (2.8)$$

The kinetic energy of the system and the absolute temperature are related. Hence, the absolute temperature  $T$  can be determined by Equation 2.9. In this equation,  $N$  is the total number of atoms,  $\langle v^2 \rangle$  is the square of the average velocity and  $m$  is mass of particle  $i$ .

$$T = \frac{2}{3Nk_B} \sum_{i=1}^N \frac{1}{2} m_i \langle v_i^2 \rangle \quad (2.9)$$

The Langevin equation is a stochastic differential equation that describes the dynamics of a subset of the degrees of freedom of a physical system.

The friction coefficient and the force due to the zero-mean collision between solvent particles and molecule particles are considered to be independent of time. When the friction coefficient is taken as zero, the Langevin equation becomes equal to the equation underlying the molecular dynamics. For this reason, adding the Langevin equation to the Molecular Dynamics equation improves the method by taking external effects into account [32]. Molecular Dynamics equation with external effects is given in Equation 2.10. In this equation  $R$  is random forces and  $m$  is mass of particle  $i$ ,  $\gamma$  is the force constant,  $r$  is the position vector.

$$m_i \frac{d^2 r_i}{dt^2} = -\frac{\partial v}{\partial r_i} - m_i \gamma \frac{dr_i}{dt} + R_i(t) \quad (2.10)$$

### 2.3. ANM Driven Langevin Dynamics

ANM driven Langevin Dynamics (ANM-LD) simulation is one of the effective hybrid methodologies applying LD simulations with ANM predictive intrinsic functional

dynamics to produce conformations between two states [8]. Cooperative and global movements that are the cause of functional transitions can be observed by restricting certain modes and changing simulation parameters.

In order to apply the ANM-LD methodology, the initial and target structure must be known. Since the ANM-LD procedure is iterative, in each simulation step the generated structure is perturbed with the selected ANM mode. The convergence to the target structure is determined by calculating the RMSD between the structure created in each iteration and the target structure. The RMSD equation is calculated by Equation 2.11. In this equation,  $N$  is the total number of atoms and  $d$  is the distance between the formed structure and the target structure.

$$RMSD = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}} \quad (2.11)$$

When selecting the ANM mode, the overlap value is calculated by using the difference vector between the target structure and the generated structure and the global mode. The overlap value ( $O$ ) is calculated by Equation 2.12 by using global mode( $u$ ) and difference vector ( $D$ ) between the generated and target structure.

$$O_{i,j} = \frac{|u_j \cdot D_i|}{|u_j| \cdot |D_i|} \quad (2.12)$$

ANM mode is selected as the mode with the highest overlap value. This structure is perturbed by the selected ANM mode and the deformation factor which is one of the simulation parameters. The structure formed as a result of this procedure is used as the initial structure of the next step. This methodology continues until the RMSD converges between the iteratively created structure and the target structure. The procedure applied in the ANM-LD simulations is demonstrated in Figure 2.1.

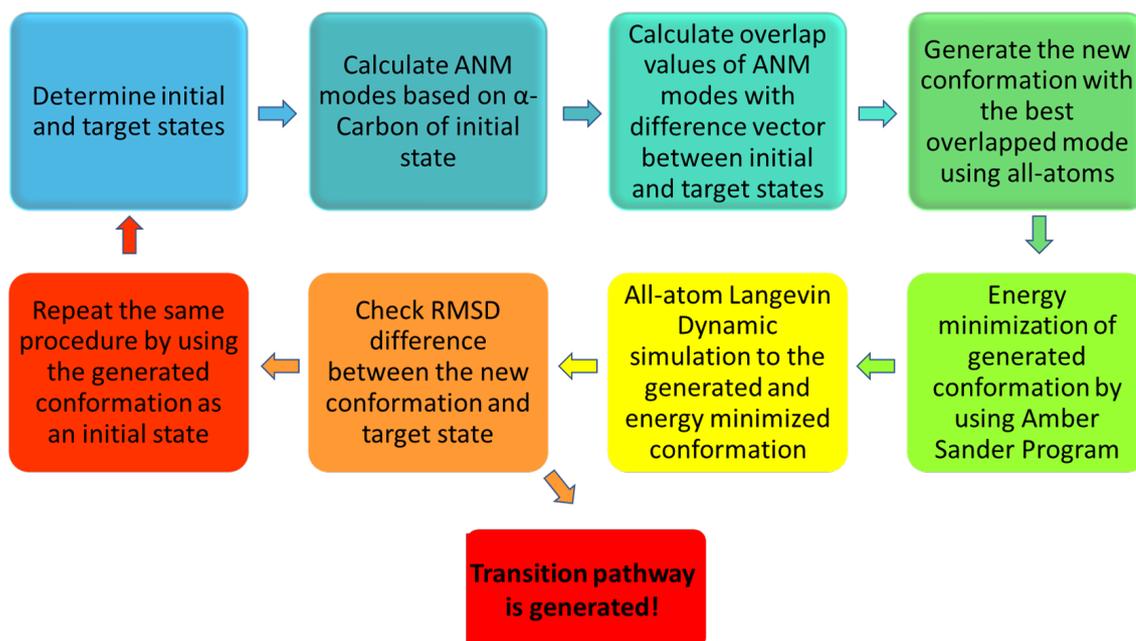


Figure 2.1. The procedure applied in ANM-LD simulations.

Before ANM-LD simulation is started, the protonation states of the charged residues are calculated for initial and final structure by using PDB2PQR server [33]. In ANM-LD simulations, some parameters are determined before the simulation is started. These parameters are deformation factor (DF), cut-off distance ( $R_{cut}$ ), maximum number of modes (modemax) which belong to the ANM part of the ANM-LD methodology; absolute temperature (T), step of energy minimization (stepmin), and step of LD simulation (stepsim). LD part of ANM-LD simulation methodology is performed with Amber ff03-ff10 [34, 35].

In the unconstrained ANM-LD method, while calculating the activation path between the initial and target structure, the difference vector between the structure created in each cycle and the target structure is calculated. This difference vector is compared with the slow ANM modes of the generated structure and the structure is deformed by selecting the ANM modes with maximum overlap and a new structure is created. After examining the unconstrained ANM-LD simulations, it will be possible to

reveal the hidden modes during the structural transition by blocking the most selected ANM modes in restricted ANM-LD simulations in order to examine whether there are different transition paths.

### 3. RESULTS and DISCUSSION

#### 3.1. Muscarinic Acetylcholine Receptor

Muscarinic acetylcholine receptors belong to Class A of G protein-coupled receptors and play an important role in nervous systems. For this reason, muscarinic acetylcholine receptors are targeted for the treatment of many central and parasympathetic nervous system related diseases, such as Alzheimer's disease, schizophrenia, Parkinson disease [36]. Here, M2R, which one of the five subtypes (M1R-M2R-M3R-M4R-M5R) of muscarinic receptor, is investigated [37]. Although sequence similarity of these subtypes is very high, they connect different G-proteins to transfer signal.

Acetylcholine enables the activation of M2 receptor by  $G_{i/o}$  (inhibitory G-protein) to regulate acetylcholine-activated  $K^+$  channels and inhibit adenylyl cyclase activity to decrease adenosine 3',5'-cyclic monophosphate (cAMP) production [37, 38].

Muscarinic acetylcholine receptors have seven transmembrane helices (TM1-TM2-TM3-TM4-TM5-TM6-TM7), three extracellular loops (ECL1-ECL2-ECL3), and three intercellular loops (ICL1-ICL2-ICL3) like other receptors in GPCR family. Also, M2R ends with Helix 8 which is prevalent in Rhodopsin-like receptors [36]. The structure and corresponding regions of M2R is given in Figure 3.1.

The residue number of M2 receptor starts with 20 and ends with 456. Since there are 161 missing residues between 216 and 377, residue number of M2 receptor is 276. The corresponding regions of residues is given in Figure 3.2 [39].

Inactive and active states of GPCRs seem to show correlative conformational changes between various subspecies. Nevertheless, the activation of various GPCRs differs in a few different ways. In this section, M2 receptor is used to investigate the activation mechanism of GPCRs. The most important characteristic of GPCR

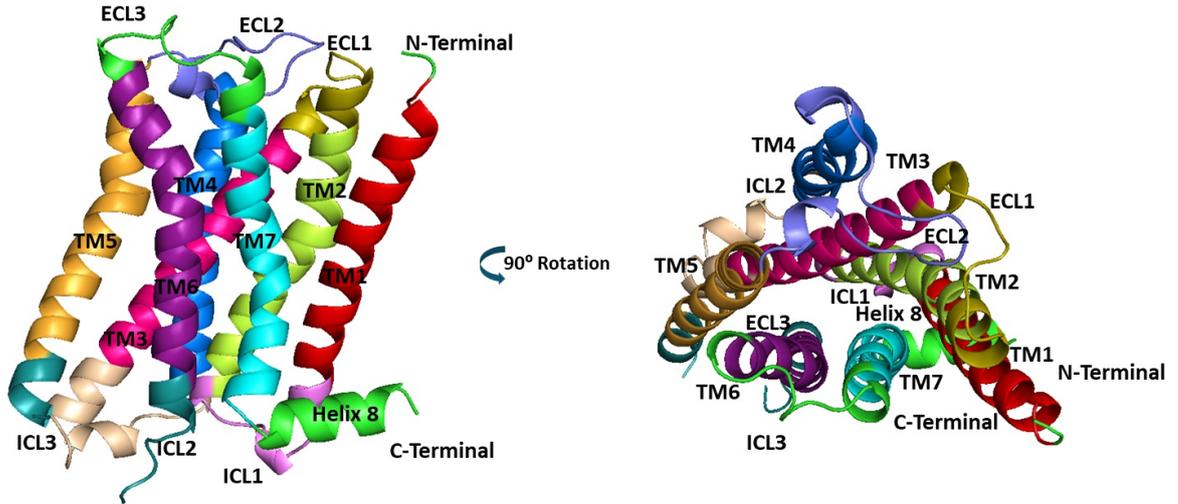


Figure 3.1. The structure and corresponding regions of muscarinic acetylcholine receptors.

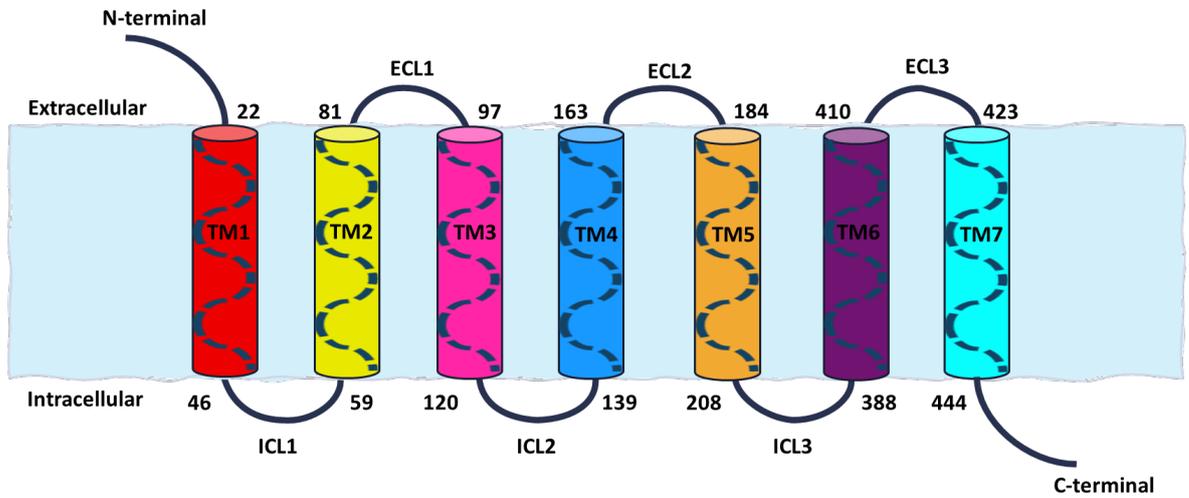


Figure 3.2. The wire diagram of M2 receptor.

activation is the formation of a cage for the accommodation of G-protein  $\alpha$  subunit. During activation, conformational changes happen particularly in three unique locales that are called as ligand-binding site, connector region, and G-protein binding site. Through the cage formation, intracellular side of TM6 moves outward, TM5 moves slightly outward and NPxxY motif of TM7 rearranges [40]. The distance between  $R121^{3.50}$  (DRY Motif) in TM3 and  $T386^{6.34}$  in TM6 is increases while the distance between  $R121^{3.50}$  (DRY Motif) in TM3 and  $Y440^{7.53}$  (NPxxY Motif) in TM7 decreases (Ballesteros–Weinstein residue numbering [41]) [36, 42]. The conformational change in ligand binding site is originated from  $W400^{6.48}$  shifting towards to  $F195^{5.47}$  and  $V199^{5.51}$ . The connector region is placed between the ligand binding site and G-protein binding site where  $V111^{3.40}$  interacts with  $P198^{5.50}$ . In the G-protein binding site the interactions between  $S118^{3.47} - Y206^{5.58}$  and  $I117^{3.46} - Y440^{7.53}$  enable the inward movement of  $Y206^{5.58}$  and  $Y440^{7.53}$ . Additionally, the salt bridge between  $R121^{3.50}$  and  $E382^{6.30}$  is broken through the activation process [36, 42, 43].

Functionally important residues and corresponding regions of M2R are given in Table 3.1 and Figure 3.3.

Table 3.1. Residues and corresponding regions of functionally important sites of M2 receptor.

Important Sites	Residue Names	Region
TM3 – TM6 Distance	R121 – T386	ICL2 – ICL3
TM3 – TM7 Distance	R121 – Y440	ICL2 – TM7
Salt Bridge	R121 – E382	ICL2 – ICL3
DRY Motif	D120 – R121 – Y122	ICL2
NPxxY Motif	N436 – P437 – Y440	TM7
Connector	V111 – P198	TM3
Ligand Binding Site	W400	TM6
GP Binding Site	Y206 – Y440	TM5 – TM7

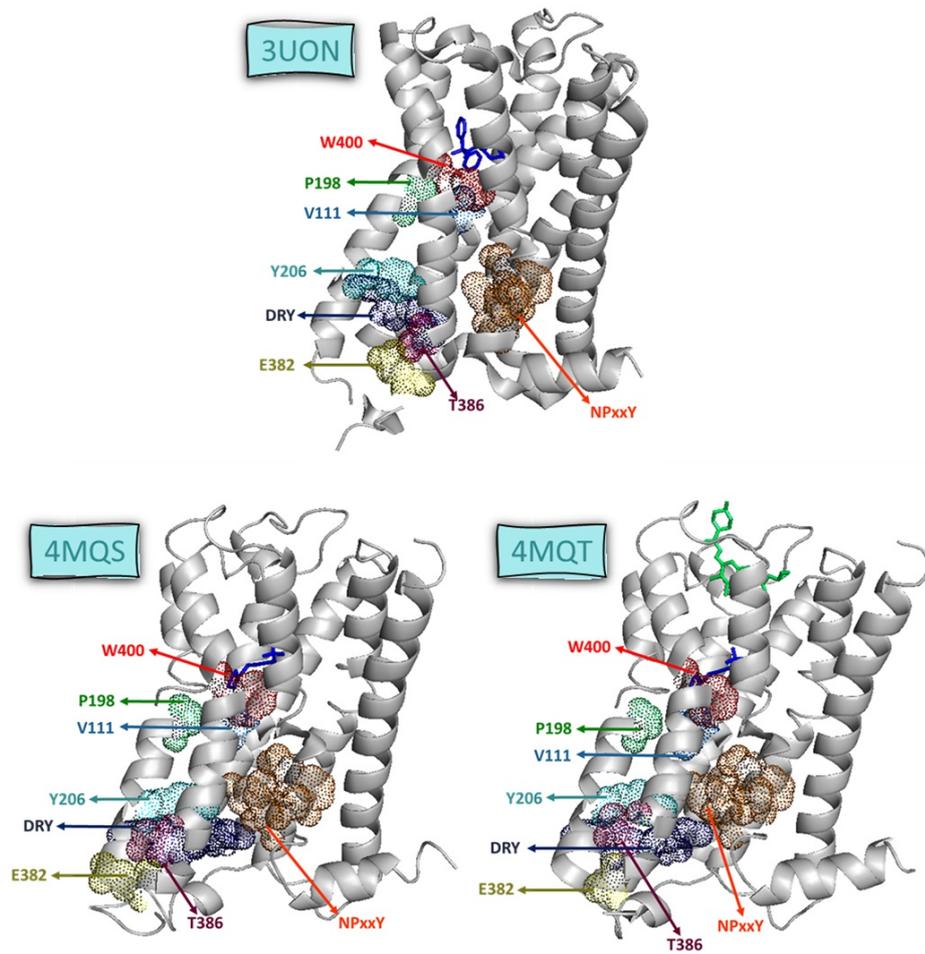


Figure 3.3. Functionally important sites of M2 receptor.

### 3.1.1. ANM-LD Simulations of M2 Receptor

In ANM-LD simulations of M2 receptor, one inactive and two active structures are used, which are obtained from Protein Data Bank with PDB IDs 3UON; 4MQS and 4MQT, respectively. 3UON is the inactive state of M2 muscarinic acetylcholine receptor bound to an antagonist [43]. Two different cases are examined to see the activation process of M2 receptor. The structure of active human M2 muscarinic acetylcholine receptor bound to the agonist iperoxo (PDB ID: 4MQS, [43]) and the structure of active human M2 muscarinic acetylcholine receptor bound to the both agonist iperoxo and allosteric modulator LY2119620 (PDB ID: 4MQT, [40]) are used as active structures in these cases. Properties of M2 receptor structures are given in Table 3.2 [40, 43].

Table 3.2. Summary of used structure in ANM-LD simulations of M2 receptor.

PDB ID	Ligand & Complex	Organism	Mutation	Structure
3UON	QNB	<i>Homo Sapiens</i>	-	
4MQS	IXO	<i>Homo Sapiens</i>	-	Differences in allosteric sites (horizontal conformation of W422)
4MQT	IXO & Allosteric Modulator	<i>Homo Sapiens</i>	-	Differences in allosteric sites (vertical conformation of W422)

To explore the activation process of the M2 receptor, two different cases (Case 1 & Case 2) are studied. These cases are from inactive structure (3UON) to active structure (4MQS), and from inactive structure (3UON) to active structure (4MQT), which are given in Figure 3.4.

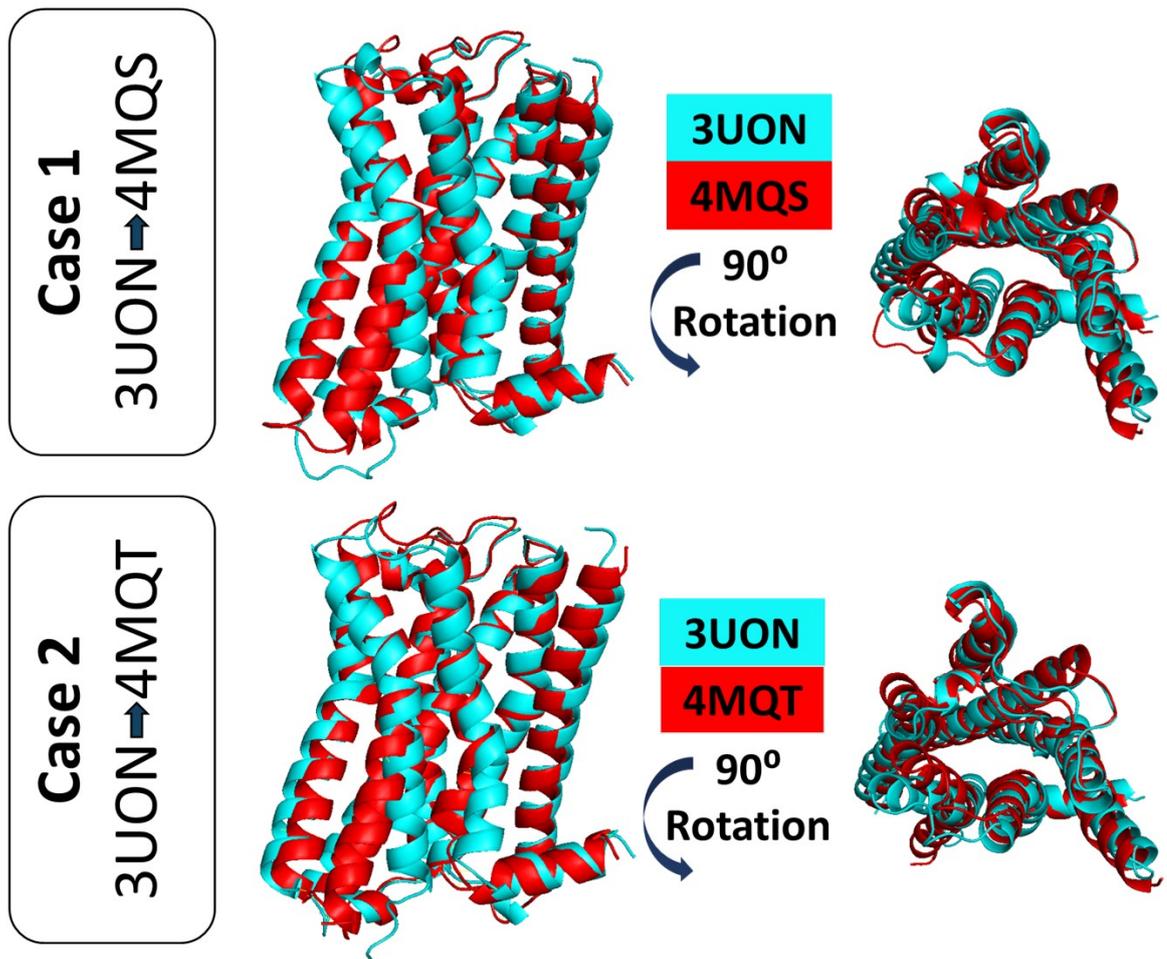


Figure 3.4. Investigated cases in ANM-LD simulations of M2 receptor. a) Alignment of 4MQS to 3UON (RMSD: 2.66 Å). b) Alignment of 4MQT to 3UON (RMSD: 2.44 Å).

3.1.1.1. ANM-LD Simulations from 3UON to 4MQS. ANM-LD simulations of M2 receptor from inactive state (3UON) to active state (4MQS) are performed to examine the transition pathway of activation mechanism. Different simulations are prepared with different  $R_{cut}$  values of 10 Å, 13 Å, and 18 Å; maximum mod number of 30, 100, and all modes; DF values of 0.35 Å, 0.40 Å, 0.60 Å, adoptive between 0.1 Å and 0.4 Å - 0.8 Å - 1.0 Å. In the preparation step of initial and final structures for the simulation, R216-I217 and N1002-Y1161 are deleted from inactive structure; K19 and chain B are deleted from active structure. ANM-LD result summary is given in Table 3.3.

Table 3.3. ANM-LD Result Summaries

Simulation Name	Cycle Number	RMSD (Å)	The Most Selected ANM Modes
$D_F = 0.10 - 0.80\text{Å}$ All Modes $R_{cut} = 13\text{Å}$	43	1.32	1, 5, 9
$D_F = 0.10 - 1.00\text{Å}$ All Modes $R_{cut} = 13\text{Å}$	40	1.53	1, 2, 5, 9
$D_F = 0.35\text{Å}$ Modemax = 100 $R_{cut} = 13\text{Å}$	70	1.06	1, 11
$D_F = 0.40\text{Å}$ Modemax = 100 $R_{cut} = 13\text{Å}$	79	1.14	1, 11, 19

The RMSD along the transition pathway generated by ANM-LD simulation from inactive to active structure is investigated and shown in Figure 3.5. RMSD of initial structure is calculated as 2.66 Å, and RMSD of final generated structure scatter between and 1.06 Å and 2.10 Å. The RMSD profile through the transition pathway with best convergence is shown in Figure 3.6.

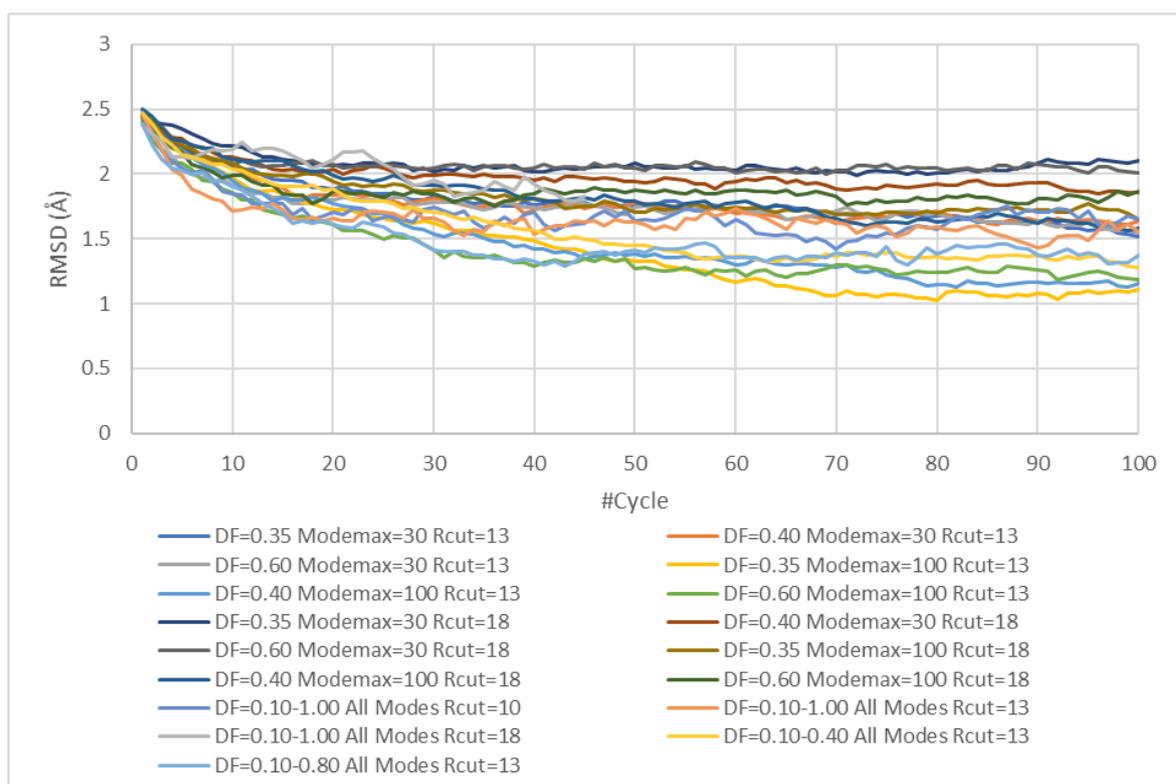


Figure 3.5. RMSD difference through the simulations of M2 receptor from 3UON to 4MQS

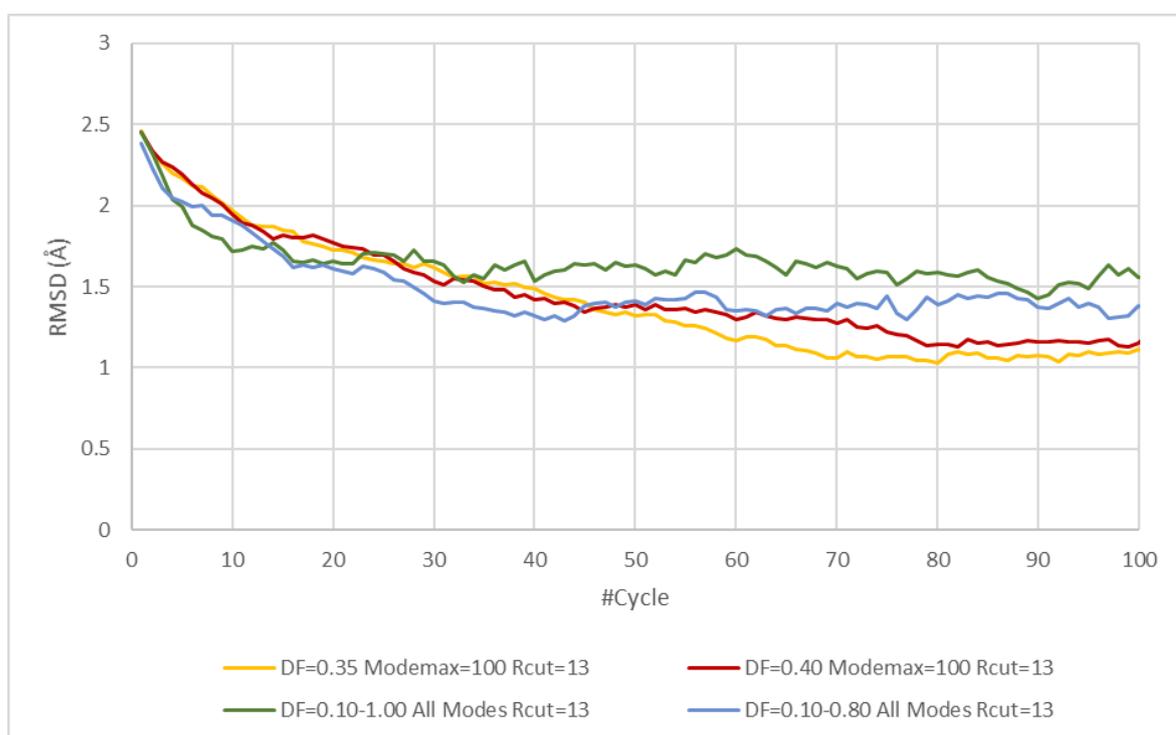


Figure 3.6. RMSD difference of selected runs through the simulations of M2 receptor from 3UON to 4MQS

Although the RMSD difference is not significant, some key changes occur in the structure during the activation pathway of M2R. These key changes are examined in order to determine how the activation takes place; in which order the structural changes occur.

- The TM3-TM6 and TM3-TM7 Distances and Salt Bridge Breakage.

First, the TM3-TM6 and TM3-TM7 distances, and salt bridge breakage of R121 and Y440 are investigated through the transition pathway in Figure 3.7, which play an important role in activation process [36]. In the calculations of the TM3-TM6 and TM3-TM7 distances,  $C\alpha$  atoms of residues R121-T386 and R121-Y440 are used, respectively. To observe salt bridge breakage, the distance between the side chains of R121 and E382 is considered. The TM3-TM6 distance is 8.4 Å and 13.8 Å respectively, for the initial and final/target structures [36]; and the TM3-TM7 distance is 17.6 Å and 11.5 Å, respectively, for the initial and final/target structures. When the last five snapshots of two best converged ANM-LD simulations from seventeen parallel simulations are considered, the TM3-TM6 distance is 14.3 Å, and 11.2 Å; the TM3-TM7 distance is 12.9 Å and 12.2 Å. Additionally, the distance between atoms forming the salt bridge is determined as 10.8 Å and 16.3 Å for the initial and final structures, respectively. In the best two simulations, the distance reaches 19.2 Å and 17.3 Å. As shown in Figure 3.7, salt bridge breakage occurs in first twenty cycles within the time window of the activation pathway of M2 receptor.

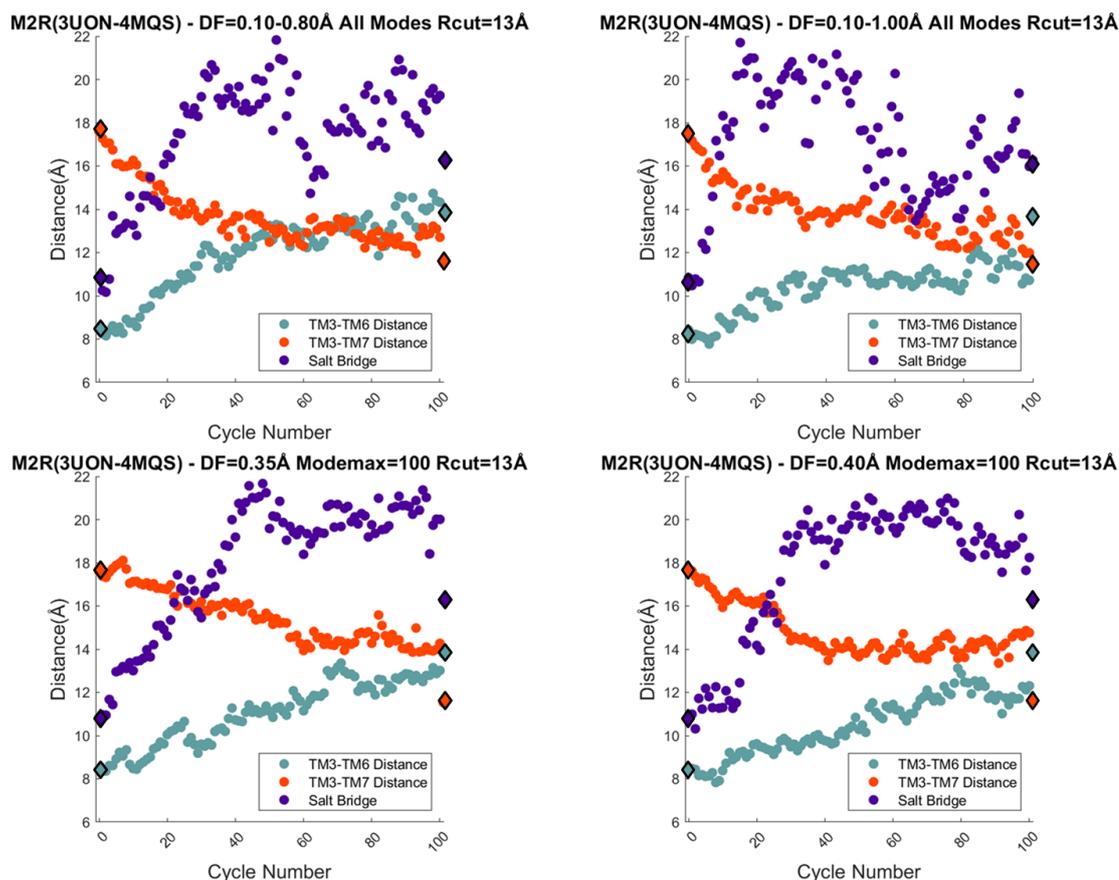


Figure 3.7. Functionally important distances through the transition pathway by ANM-LD simulations of 3UON to 4MQS (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the distance profile obtained by ANM-LD simulations is examined, it is seen that the first event that occur during the activation is the salt bridge breakage. Although the TM3-TM6 and TM3-TM7 distances are approaching the final/target structure simultaneously, the TM3-TM7 distance decreases faster.

- DRY and NPxxY Motifs.

The RMSD of the important motifs DRY (D120 – R121 – Y122) and NPxxY (N436 – P437 – Y440) with respect to the initial structure is calculated by using  $C\alpha$

atoms through the activation process and given in Figure 3.8. The RMSD of DRY and NPxxY motifs of the active structure are used as reference value when analyzing the RMSD profile of important motifs and are determined as 1.6 Å and 3.2 Å, respectively. When the last five snapshots of the ANM-LD simulation with parameter  $DF=0.10-0.80$  Å All Modes  $R_{cut}=13$  Å are considered, the RMSD of DRY motif is calculated as 2.3 Å, and 2.2 Å; the RMSD of NPxxY motif is calculated as 2.1 Å, and 2.4 Å.

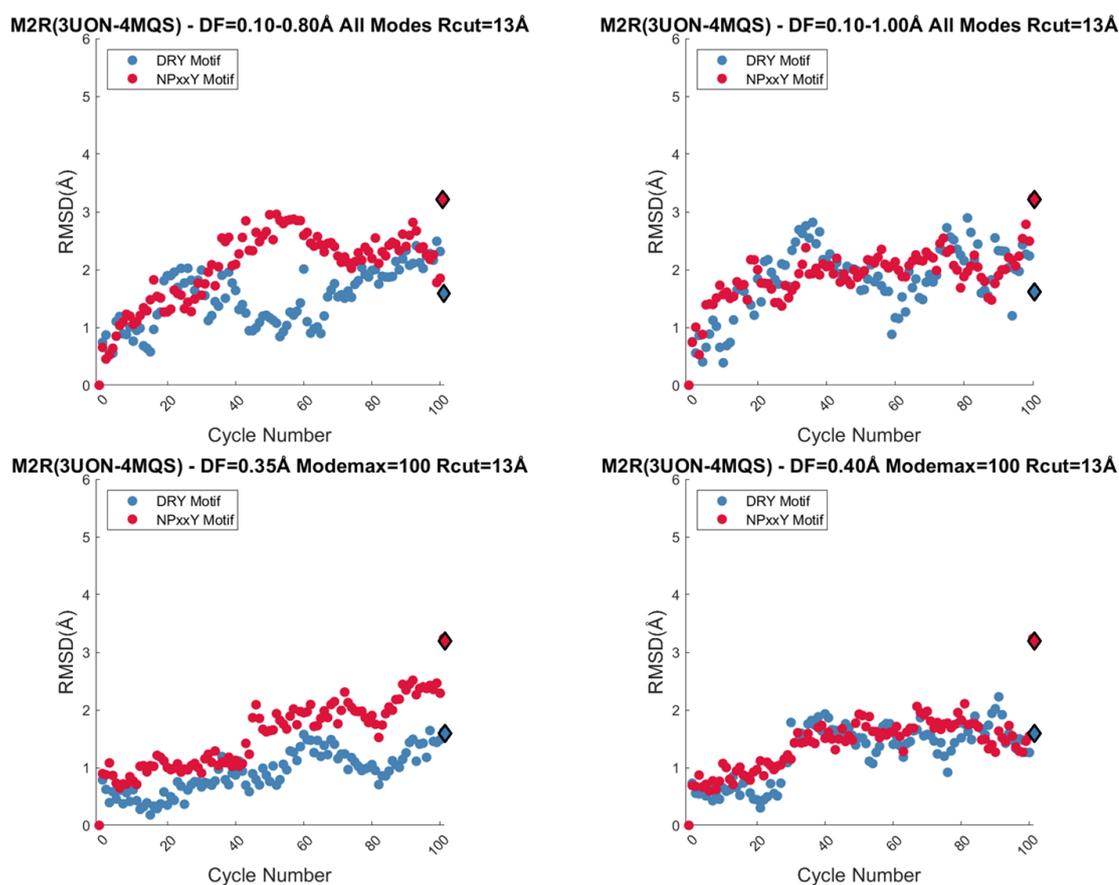


Figure 3.8. Functionally important distances through the transition pathway by ANM-LD simulations of 3UON to 4MQS (final/target structures are shown as diamond, and the distances of snapshots are shown as spheres).

- Ligand Binding Site, Connector Region, and G-protein Binding Site.

The RMSD profiles of the connector site (V111 and P198), the ligand (QNB) binding site (W400) and the G-protein binding site (Y206) from their initial positions are investigated by using  $C\alpha$  atoms through the activation process and given in Figure 3.9.

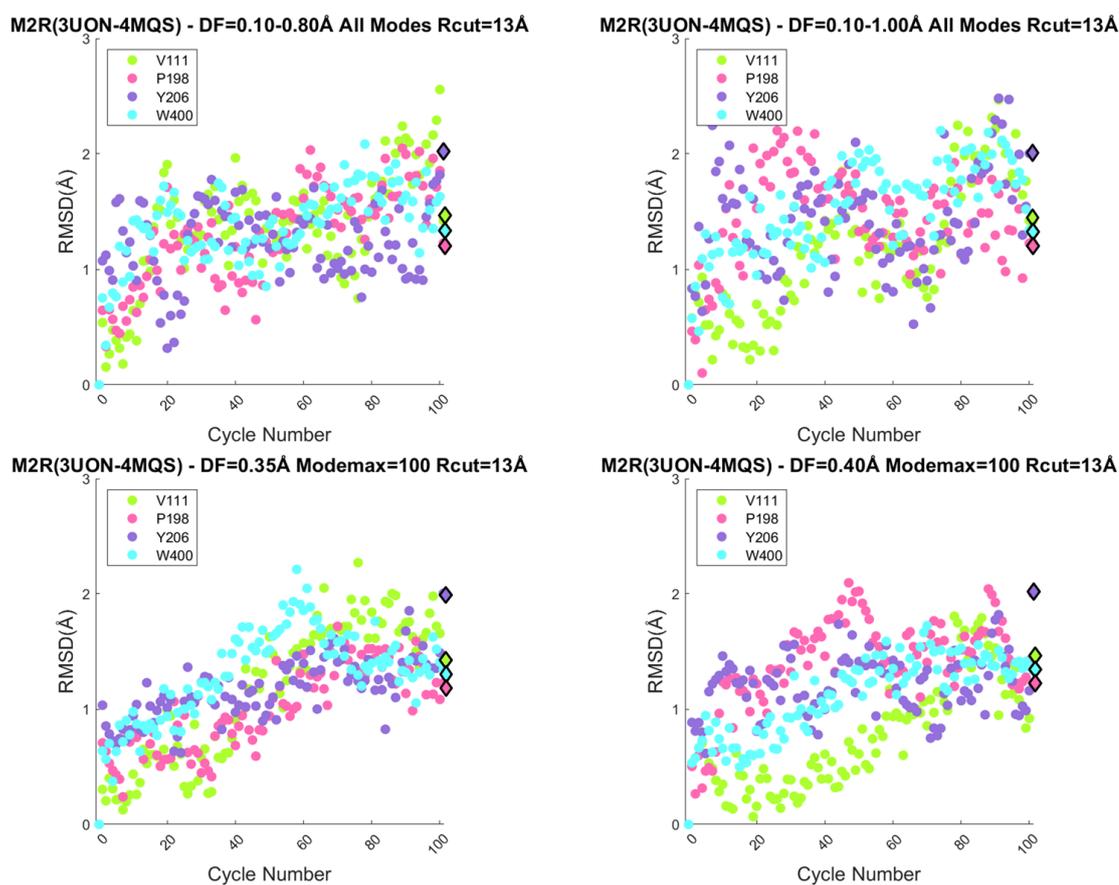


Figure 3.9. The RMSD of important residues through the activation of M2 receptor from 3UON to 4MQS (final/target structures are shown as diamond, and the RMSD of snapshots are shown as circles).

The RMSD between the initial and active/target structures is 1.4 Å, 1.2 Å, 2.0 Å, and 1.3 Å for V111, P198, Y206, and W400, respectively. When the last five snapshots of two best converged ANM-LD simulations are considered, the RMSD of V111 is

calculated as 2. Å, and 1.8 Å; the RMSD of P198 is calculated as 1.7 Å, and 1.4 Å; the RMSD of Y206 is calculated as 1.7 Å, and 1.9 Å; the RMSD of W400 is calculated as 1.6 Å, and 1.8 Å.

- Key Modes of Motion.

The selected ANM modes that enable the transition pathway from the inactive to active structures provides clues about the intrinsic dynamics and mechanism of activation. For this purpose, the most selected ANM modes are determined to disclose the dynamic determinants during the activation process. ANM mode shapes are displayed as mean square fluctuations. Although ANM modes are calculated for every generated snapshot through the transition pathway, the selected modes are mapped to the dynamic modes of the initial structure which allows us to find out how often the modes – the same type of motion- are selected and what the overlap values of these modes are based on the initial structure. The overlap value is calculated between the modes selected in each step and the slowest thirty modes of the first structure. The mode with the highest overlap value is determined as the corresponding mode in the initial structure. However, the modes with an overlap value of less than 0.5 are not used in determining the corresponding modes, since mean square fluctuations of these modes does not overlap with the corresponding modes of the initial structure. Selected mode mapping and overlap values are given in Figure 3.10.

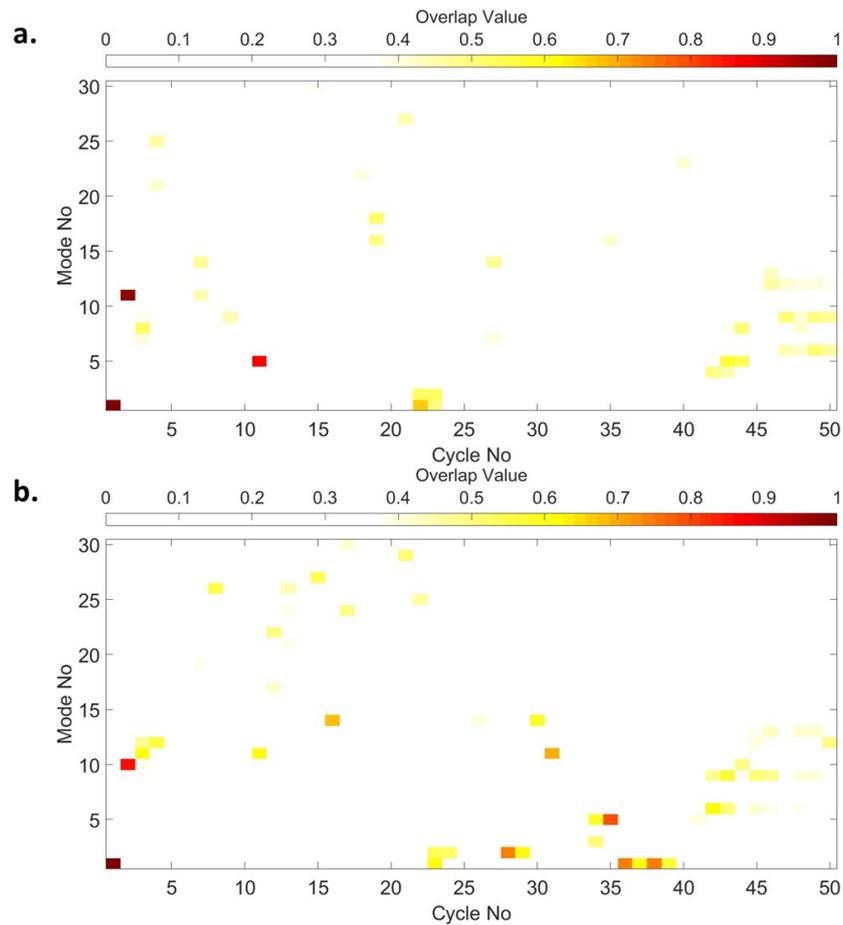


Figure 3.10. Selected Mode Mapping a) DF=0.10-0.80 Å All Modes  $R_{cut}=13$  Å. b) DF=0.10-1.00 Å All Modes  $R_{cut}=13$  Å. c) DF=0.35 Å Modemax=100  $R_{cut}=13$  Å. d) DF=0.40 Å Modemax=100  $R_{cut}=13$  Å.

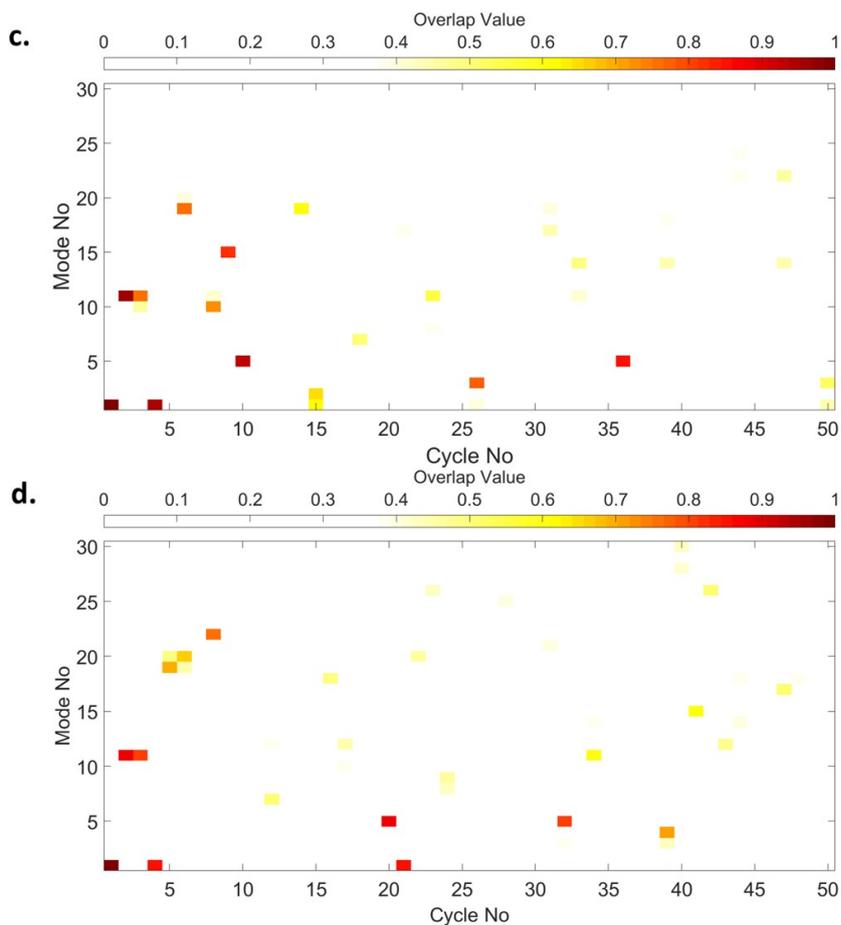


Figure 3.10. Selected Mode Mapping a) DF=0.10-0.80 Å All Modes  $R_{cut}=13$  Å. b) DF=0.10-1.00 Å All Modes  $R_{cut}=13$  Å. c) DF=0.35 Å Modemax=100  $R_{cut}=13$  Å. d) DF=0.40 Å Modemax=100  $R_{cut}=13$  Å. (cont.)

The selected ANM modes are investigated according to frequency and overlap value which are shown in Figure 3.11. According to these criteria ANM modes 1 and 5 are selected as the key modes of motion activation.

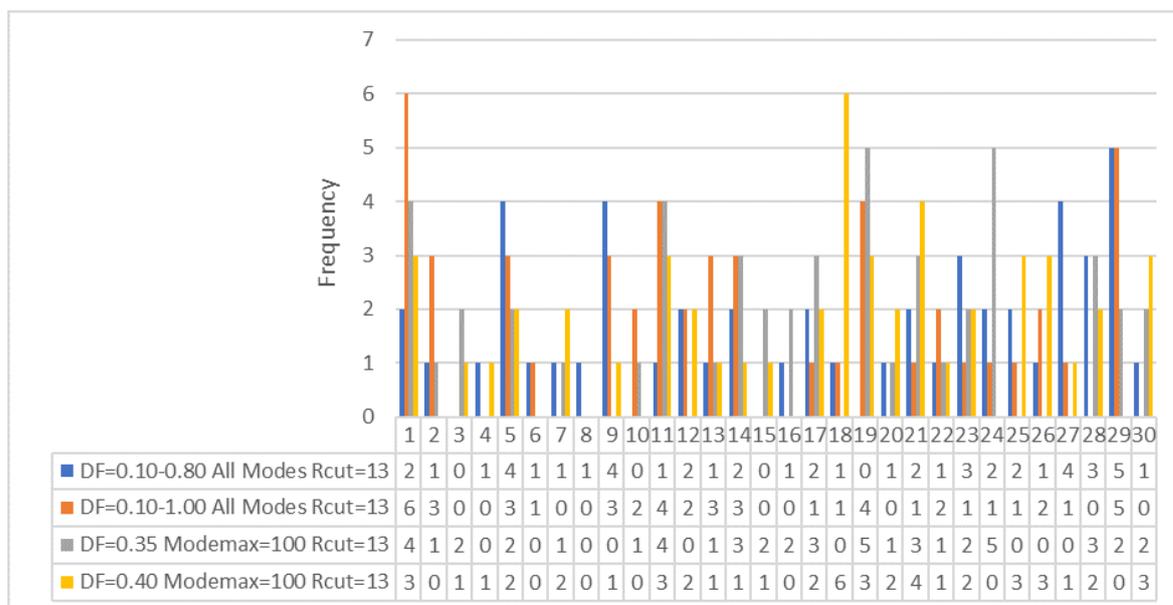


Figure 3.11. ANM mode frequency of parallel simulations from 3UON to 4MQS

The shape of ANM modes are represented as square fluctuations. Minimum points in the mean square fluctuation graph are taken into consideration to determine the hinge residues that play an important role in regulating the motion. Hinge residues of ANM mode 1 are determined as S32-L35, S76, L100-Y104, D120-Y122, K134-R135, I159, F181-A185, I209, T386, V407, Y426, and Y430 as shown in Figure 3.12.

Another important selected mode is ANM mode 5 in the activation of M2 receptor. The hinge residues of ANM mode 5 are V24, I48, Q55, L100, D120-Y122, A140, L154, D173-C176, A191, Y196, V385-T386, N404, T423, Y426, F447, and T450 as shown in Figure 3.13.

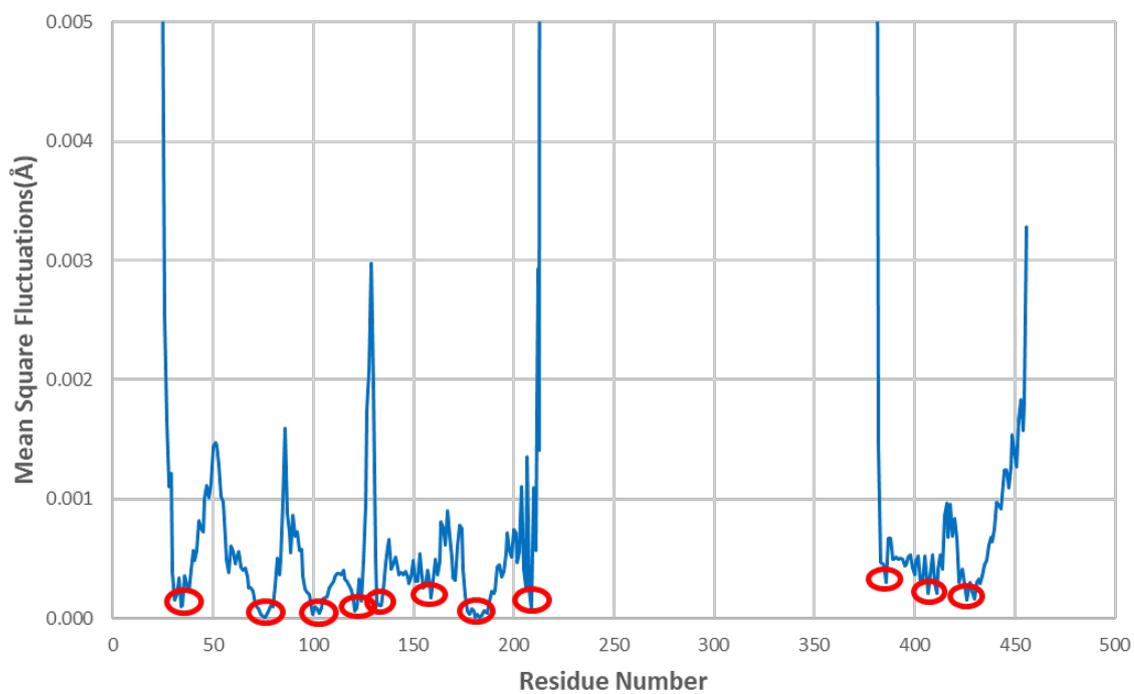


Figure 3.12. ANM mode 1 shape of the inactive state of M2 receptor

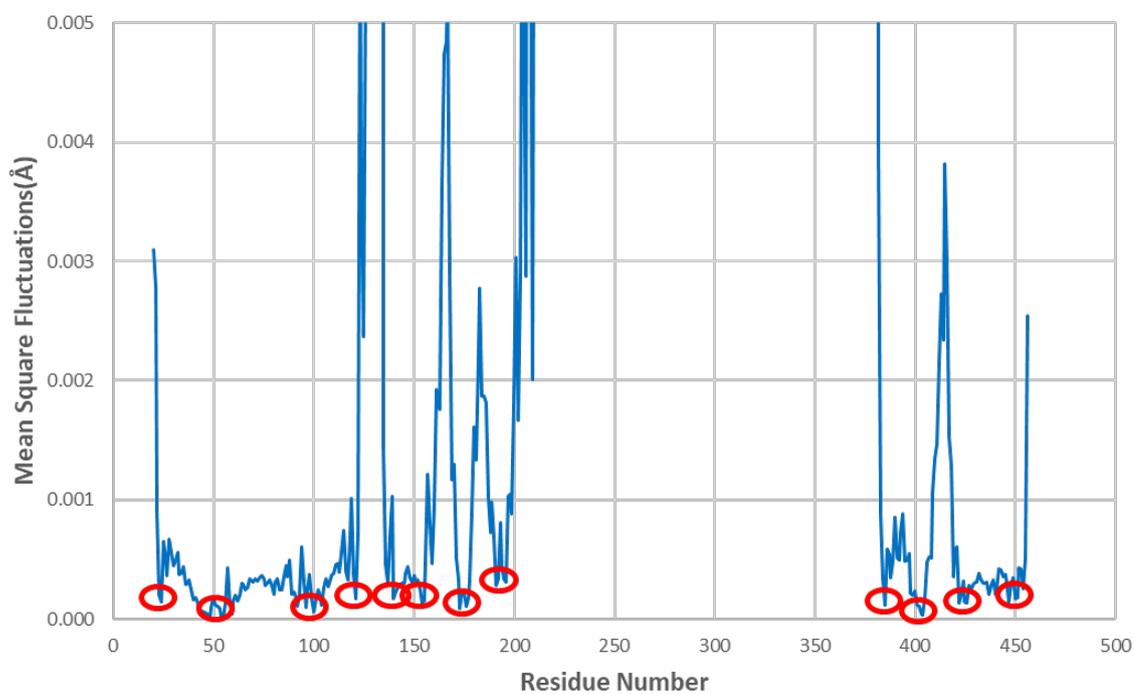


Figure 3.13. ANM mode 5 shape of the inactive state of M2 receptor

The hinge residues of ANM mode 1 and 5 are demonstrated on the inactive structure of M2 receptor in Figure 3.14.

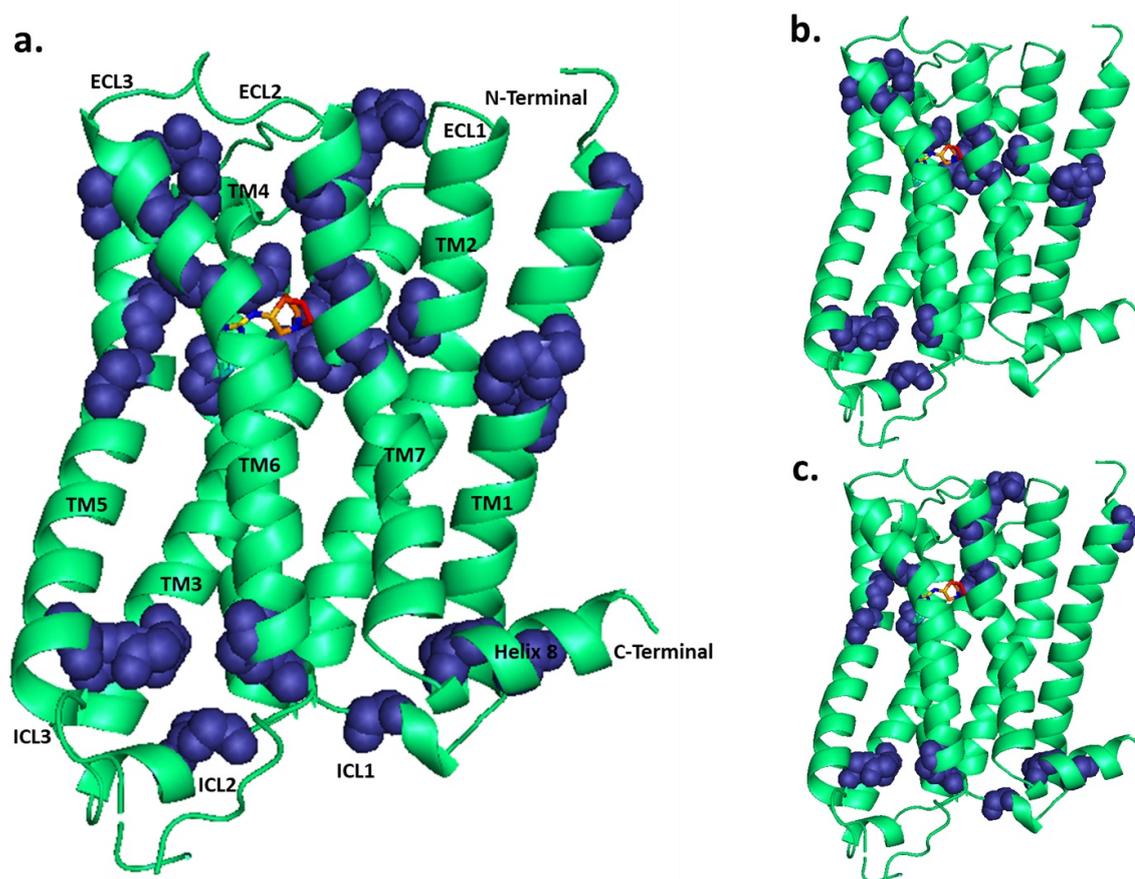


Figure 3.14. Demonstration of the minimum fluctuating residues, the hinge residues of ANM mode 1 and 5 (dark blue and sphere) on the structure. a) Hinges of ANM mode 1 and 5. b) Hinges of ANM mode 1. c) Hinges of ANM mode 5.

When the hinge residues are examined on the structure, it is observed that the most effective regions during the activation process are intracellular loops (ICL1-ICL2), extracellular loop (ECL2), and residues located around the ligand (QNB) binding site (W400). During the activation process of M2 receptor, W400 shifts toward to F195 and V199. This change appears to be affected by the hinge residues of ANM mode 5 N404, Y196, Y426, and T423 in the extracellular side and the hinge residues of ANM

mode 1 R121, Y122, and T386 in the intracellular side as stated in literature [36]. This is an indication that these two ANM modes reflect the movement in the ligand (QNB) binding site and G-protein binding site. Additionally, Y196 which is a hinge residue of ANM mode 5, enables a strong communication between TM5 and TM6 by interacting with F195 and N404 [36].

Functionally important regions (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, T386, and E382) (red and sphere) and the hinge residues of ANM mode 1 and 5 (deep blue) are demonstrated on the inactive structure in Figure 3.15.

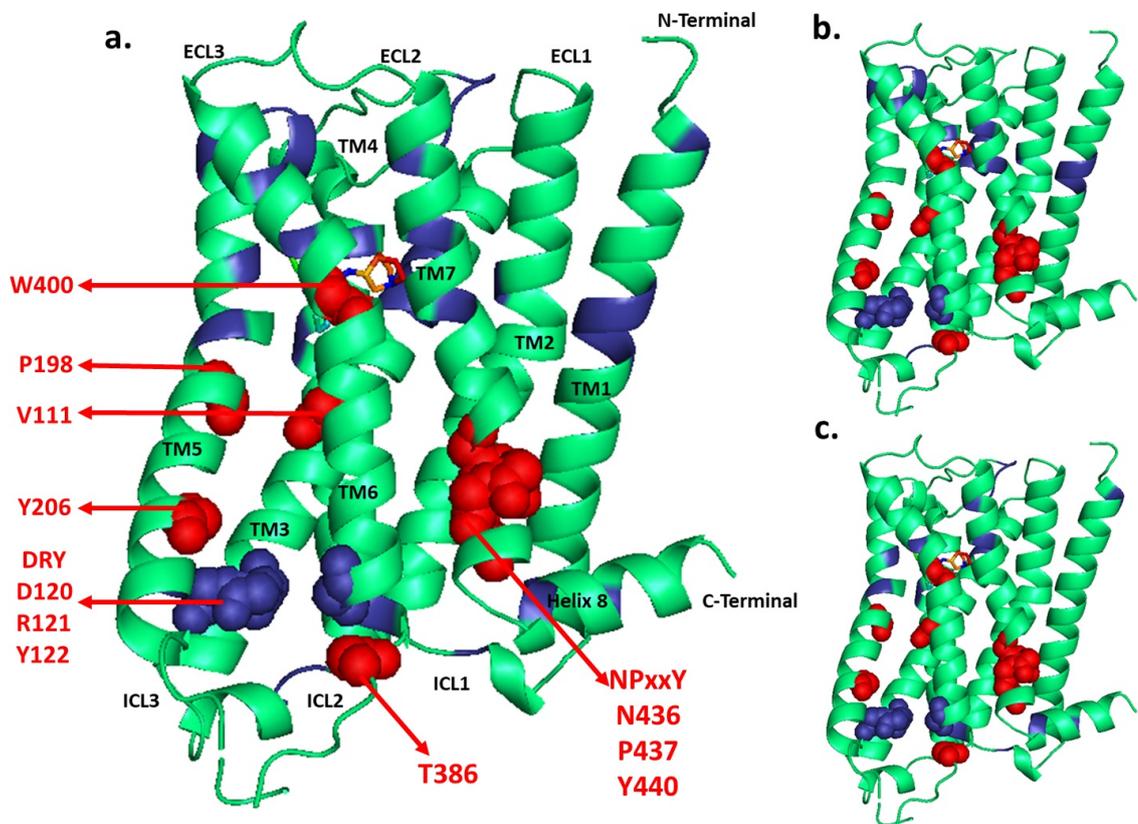


Figure 3.15. Demonstration of ANM mode 1 and 5 hinges (dark blue), functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a)ANM mode 1 and 5 b)ANM mode 1 c)ANM mode 5.

- Cross-correlation Analysis.

Normalized cross-correlation analysis is used to identify the correlated residues. The normalized correlations are in the range  $[-1, 1]$ . The residues having the cross-correlation values below zero are negatively correlated residues, and cross-correlation values above zero are positively correlated residues. Positively correlated residues which corresponds to the range  $[0, 1]$  in normalized cross correlation analysis have a greater significance in driving the motion. Normalized cross-correlation analysis is performed using the intermediate conformation in the activation pathway of M2 receptor and demonstrated in Figure 3.16.

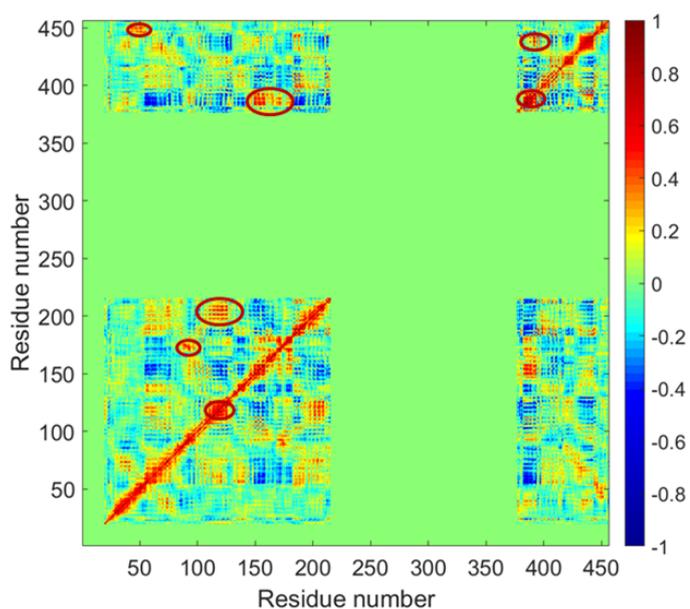


Figure 3.16. Cross correlation of residues in the activation pathway of M2 receptor from 3UON to 4MQS.

The residue pairs with the correlation value below 0.4 are identified as poorly correlated residues and with the correlation value higher than 0.6 are identified as highly correlated residues [36]. Highly correlated residues are determined as S118 & V125, S118 & S215, L393 & A441-L442, and I392 & A438. It is observed that, S118

in TM3 helix is highly correlated with V125 in ICL2 and S215 in ICL3. Additionally, I392 and L393 located in TM6 helix are correlated with the residues nearby NPxxY motif (A441-L442) in TM7.

**3.1.1.2. ANM-LD Simulations from 3UON to 4MQT.** ANM-LD simulations are executed to investigate the activation mechanism of M2 receptor from the inactive structure (3UON) to the active structure (4MQT). Parallel ANM-LD simulations are prepared with different maximum mod number of 30, 100, and all modes;  $R_{cut}$  values of 10 Å, 13 Å, and 18 Å; DF values of 0.35 Å, 0.40 Å, 0.60 Å, adaptive between 0.1 Å and 0.8 Å - 1.0 Å. In the preparation phase of initial and final structures, S215-I217, P377-P378, and N1002-Y1161 are deleted from initial structure; K19, C457, and chain B are deleted. The result summary of ANM-LD simulations is given in Table 3.4.

Table 3.4. ANM-LD Result Summaries.

Simulation Name	Cycle Number	RMSD (Å)	The Most Selected ANM Modes
$D_F = 0.10 - 0.80\text{Å}$ All Modes $R_{cut} = 13\text{Å}$	47	1.32	2, 5, 6, 8
$D_F = 0.60\text{Å}$ Modemax = 100 $R_{cut} = 13\text{Å}$	50	1.20	1, 2, 3, 8, 10
$D_F = 0.40\text{Å}$ Modemax = 100 $R_{cut} = 13\text{Å}$	55	1.15	2, 8, 10
$D_F = 0.35\text{Å}$ Modemax = 100 $R_{cut} = 13\text{Å}$	56	1.16	2, 9

Root mean square deviations (RMSD) through the activation of M2 receptor is calculated and given in Figure 3.17. RMSD of initial structure is calculated as 2.4 Å. RMSD of final generated structure scatter between 2.1 Å and 1.2 Å. The RMSD graph of the simulations with best convergence is shown in Figure 3.18.

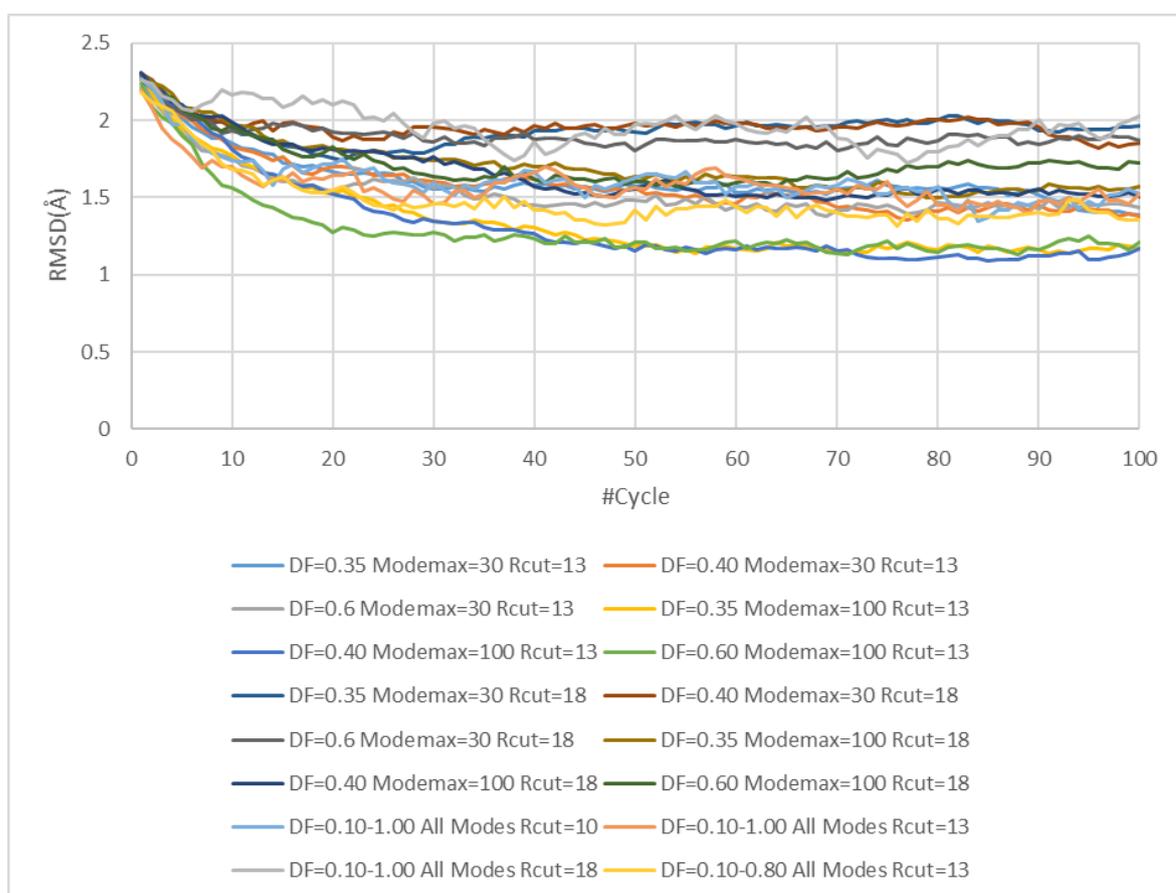


Figure 3.17. RMSD difference through the simulations of M2 receptor from 3UON to 4MQT.

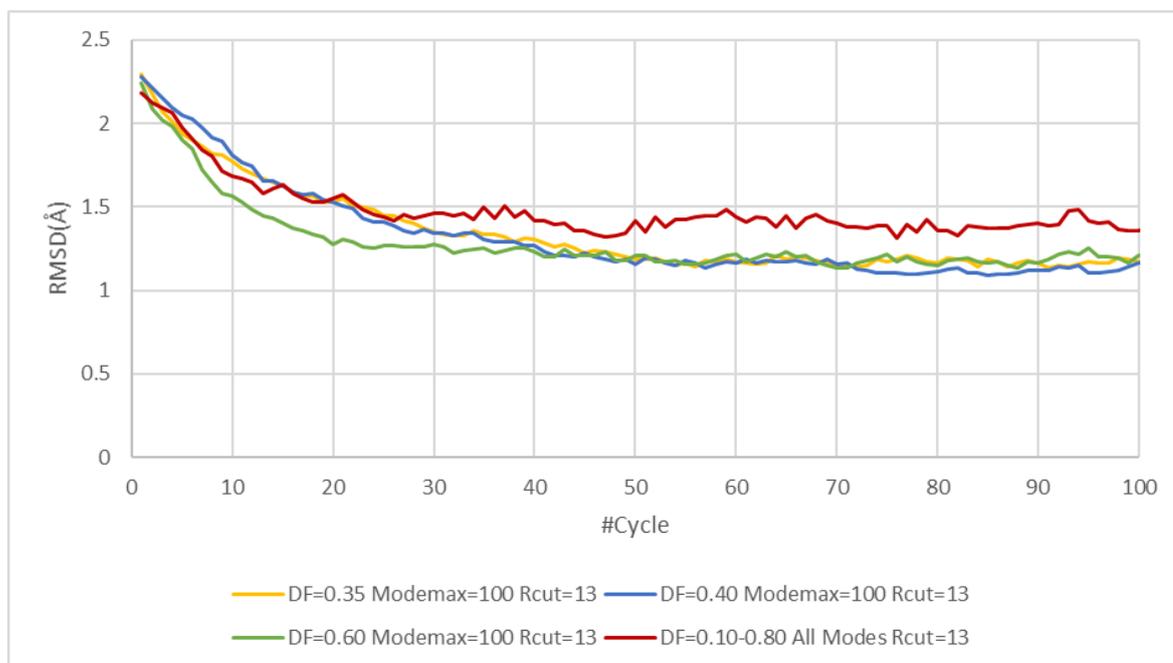


Figure 3.18. RMSD difference of selected runs through the simulations of M2 receptor from 3UON to 4MQT.

It was mentioned above that some important structural changes occurred during the activation of the M2 receptor. In this case, the TM3-TM6 distance, the TM3-TM7 distance, salt bridge breakage, RMSD of DRY and NPxxY motif and RMSD of functionally important residues are calculated.

- TM3-TM6 and TM3-TM7 Distances and Salt Bridge Breakage.

The TM3-TM6 and TM3-TM7 distances and salt bridge breakage of R121 and Y440 are investigated through the transition pathway in Figure 3.19. In the investigation of the TM3-TM6 and TM3-TM7 distances  $C\alpha$  atoms of residues R121-T386 and R121-Y440 are used, respectively. The TM3-TM6 distance is 8.4 Å and 13.8 Å, and the distance between TM3-TM7 residues is 17.6 Å and 11.6 Å for the inactive and active structures, respectively. Additionally, the distance between atoms forming the salt bridge is determined as 10.8 Å and 16.3 Å for initial and final/target structures,

respectively. When the last five snapshots of the two best converged ANM-LD simulations from sixteen parallel simulations are considered, the TM3-TM6 distance is 12.3 Å, and 13.2 Å; the TM3-TM7 distance is 13.6 Å, and 14.2 Å; the distance between the residues forming salt bridge is determined as 16.1, Å and 18.2 Å.

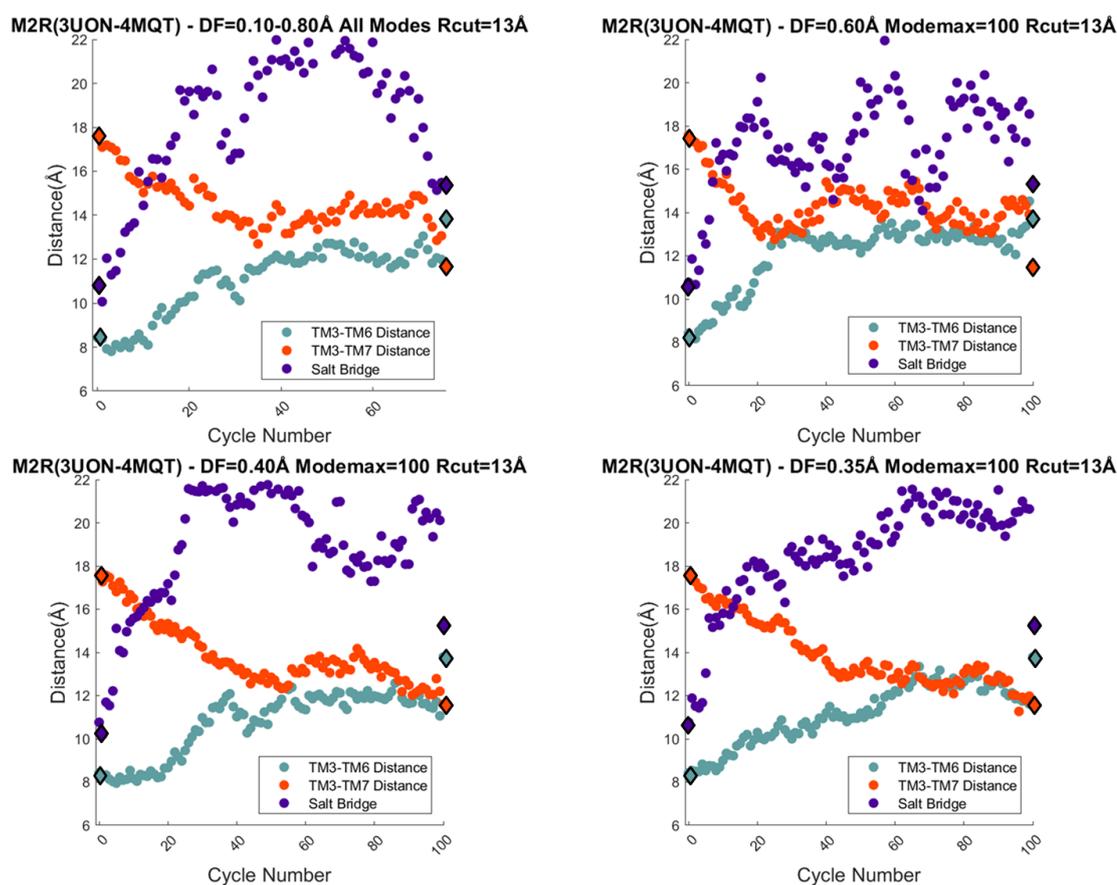


Figure 3.19. Functionally important distances through the conformation transition/activation pathway obtained from ANM-LD results of 3UON to 4MQT (initial and final/target structures shown as diamond, snapshots shown as circles).

The simulation with parameters of  $DF=0.10-0.80$  Å All Modes  $R_{cut} = 13$  Å, shows the transition during the salt bridge breakage very well. In the first twenty-five cycle, the salt bridge is broken, and the distance is reached to 20.6 Å. Then, the distance slowly decreases to 15.5 Å, which is calculated as 15.3 Å for active structure. TM3-TM6 and TM3-TM7 distance profile though the activation show similarity with the

simulation from 3UON to 4MQS. Around 40<sup>th</sup> cycle, distances of TM3-TM6 and TM3-TM7 approach 11.0 Å and 12.8 Å and remain stable for few cycles. Then, they slowly reach to the distances of the active structure.

- DRY and NPxxY Motifs.

RMSD of DRY (D120 – R121 – Y122) motif and NPxxY (N436 – P437 – Y440) motif based on the initial structure is calculated by using C $\alpha$  atoms and given in Figure 3.20.

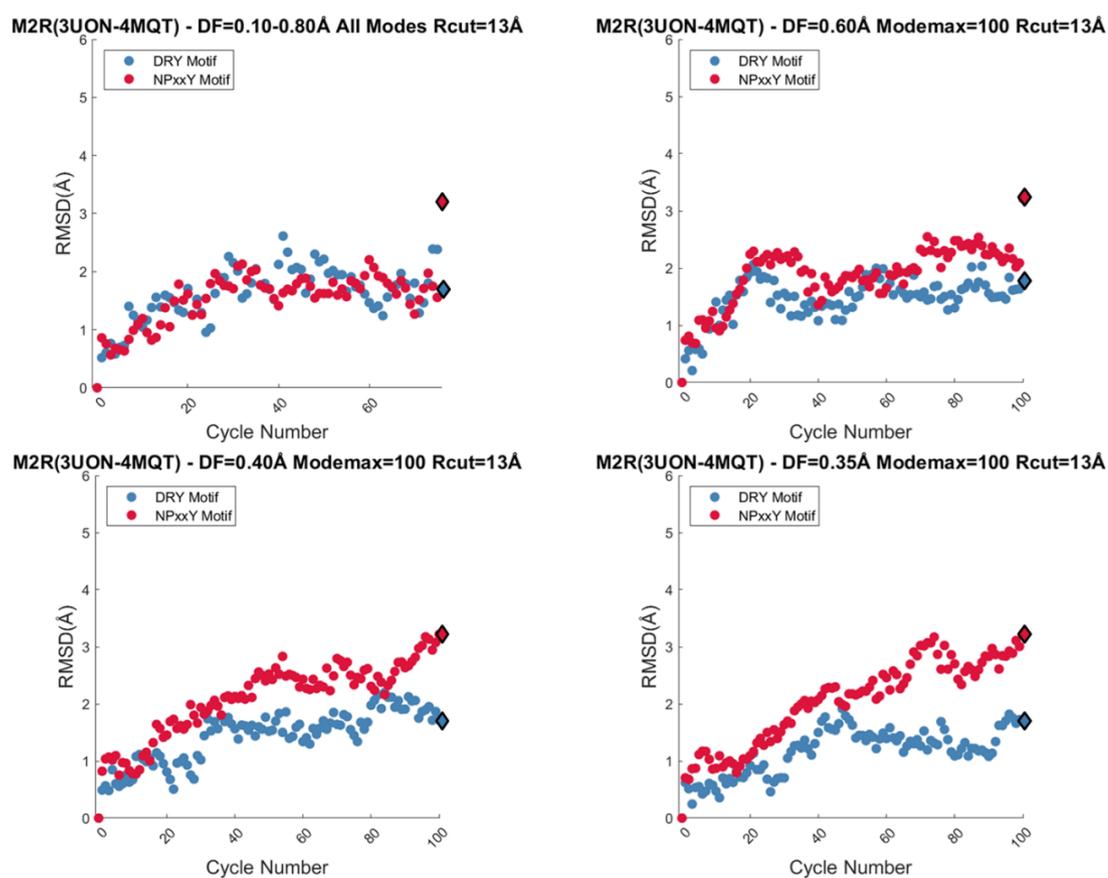


Figure 3.20. RMSD of important motifs through the activation of M2 receptor from 3UON to 4MQT (The RMSD of final/target structures are shown as diamond, and the RMSD of snapshots obtained from ANM-LD simulations are shown as circles).

The RMSD of DRY and NPxxY motifs are 1.7 Å and 3.2 Å for the active structure (4MQT), while they are calculated as 1.6 Å and 3.2 Å for the active structure (4MQS). The RMSD of NPxxY motif is almost the same in the two cases, however there is a difference of 0.1 Å change in the RMSD value of DRY motif. When the last five snapshots of the ANM-LD simulation with parameters  $DF=0.10-0.80$  Å All Modes  $R_{cut}=13$  Å are considered, the RMSD of DRY motif is calculated as 1.8 Å, and 1.9 Å; the RMSD of NPxxY motif is calculated as 3.1 Å.

- Ligand Binding Site, Connector Region, and G-protein Binding Site.

The RMSD profile according to initial structure of connector site (V111 and P198), ligand (QNB) binding site (W400) and G-protein binding site (Y206) is examined by using  $C\alpha$  atoms along the activation pathway and given in Figure 3.21. The RMSD between the initial and active/target structures is 1.6 Å, 1.2 Å, 2.0 Å, and 1.3 Å for V111, P198, Y206, and W400, respectively. Although there is 0.1 Å difference in the RMSD of V111 between the active cases (4MQS and 4MQT), other RMSD values have approximately same value, which is caused from the distinctness in the ligand binding site. When the last five snapshots of two best converged ANM-LD simulations are considered, the RMSD of V111 is calculated as 2.0 Å, and 1.8 Å; the RMSD of P198 is calculated as 1.7 Å, and 1.4 Å; the RMSD of Y206 is calculated as 1.7 Å, and 1.9 Å; the RMSD of W400 is calculated as 1.6 Å, and 1.8 Å.

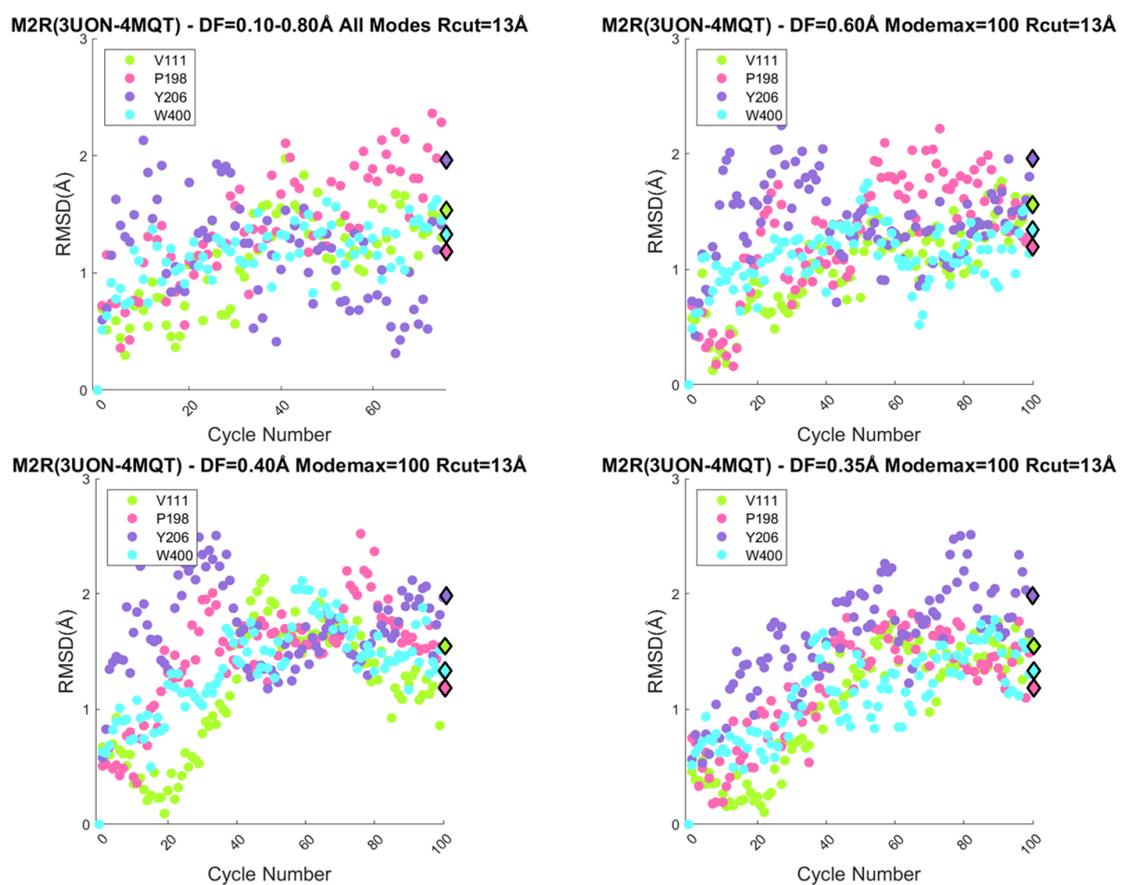


Figure 3.21. RMSD of important residues through the activation of M2 receptor from 3UON to 4MQT (The RMSD of final/target structures are shown as diamond, and the RMSD of snapshots obtained from ANM-LD simulations are shown as circles).

- Key Modes of Motion.

The selected ANM modes through the transition pathway are investigated according to frequency and overlap value, which are shown in Figure 3.22-3.23. ANM modes 2 and 3 are selected as the key modes of motion underlying activation.

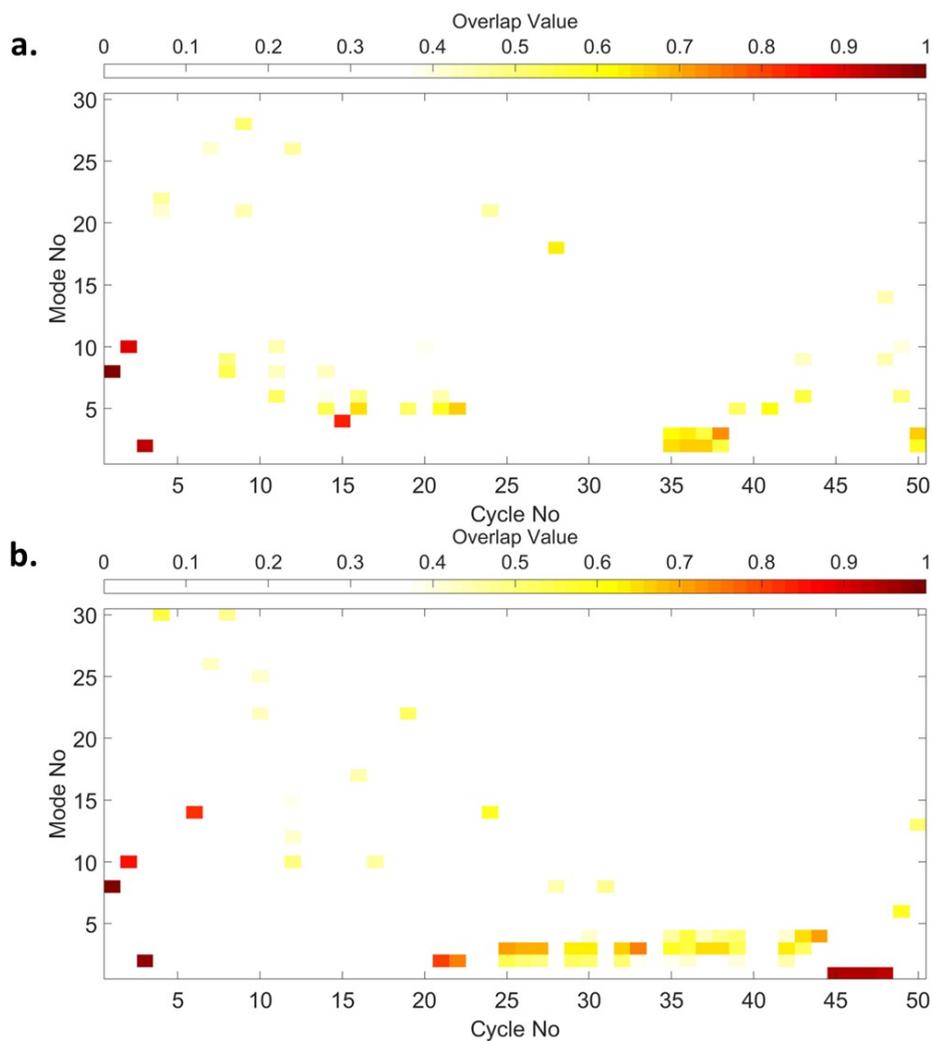


Figure 3.22. Selected Mode Mapping a) DF=0.10-0.80 Å All Modes  $R_{cut}=13$  Å. b) DF=0.60 Å Modemax=100  $R_{cut}=13$  Å. c) DF=0.40 Å Modemax=100  $R_{cut}=13$  Å. d) DF=0.35 Å Modemax=100  $R_{cut}=13$  Å.

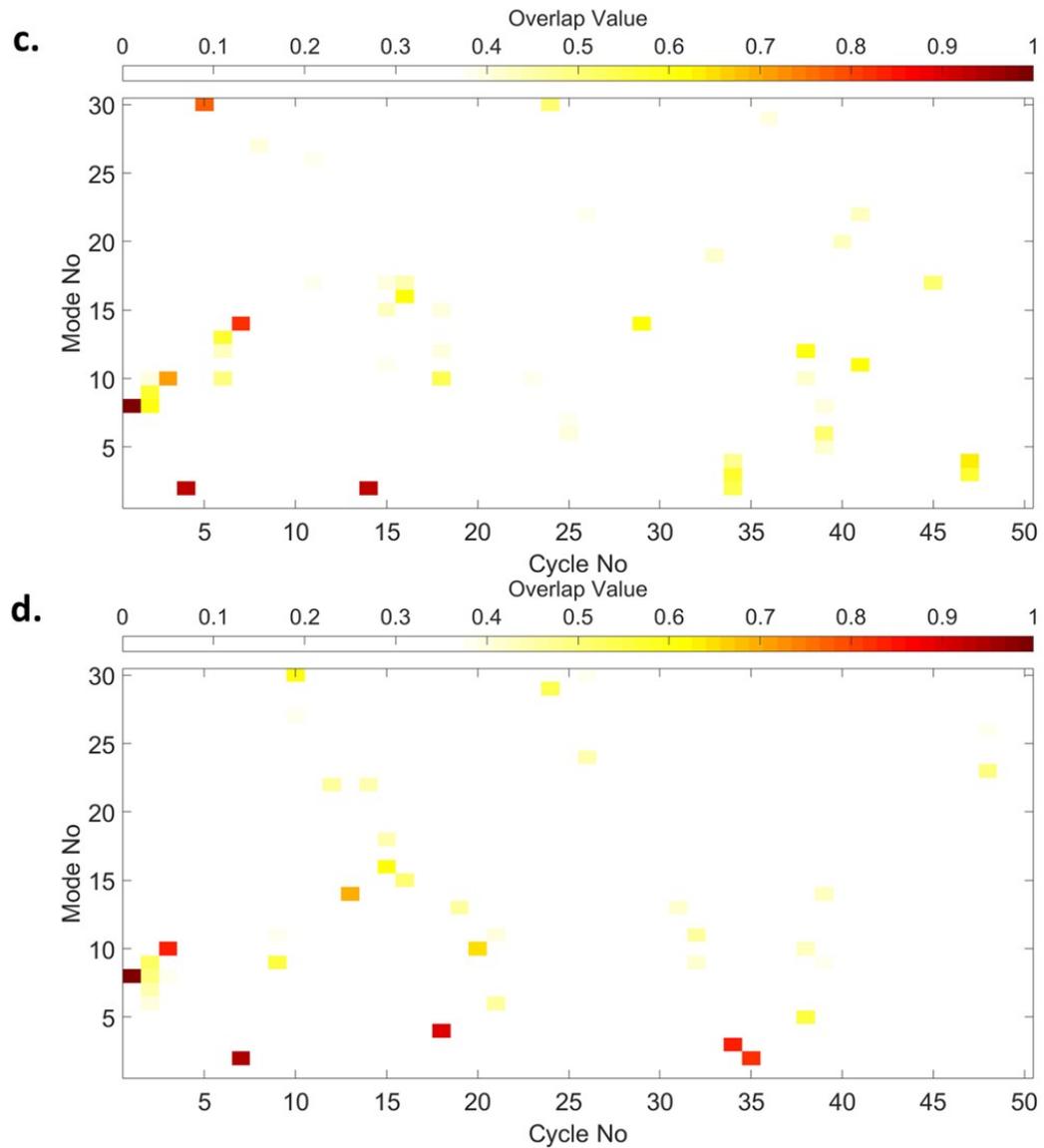


Figure 3.22. Selected Mode Mapping a) DF=0.10-0.80 Å All Modes  $R_{cut}=13$  Å. b) DF=0.60 Å Modemax=100  $R_{cut}=13$  Å. c) DF=0.40 Å Modemax=100  $R_{cut}=13$  Å. d) DF=0.35 Å Modemax=100  $R_{cut}=13$  Å. (cont.)

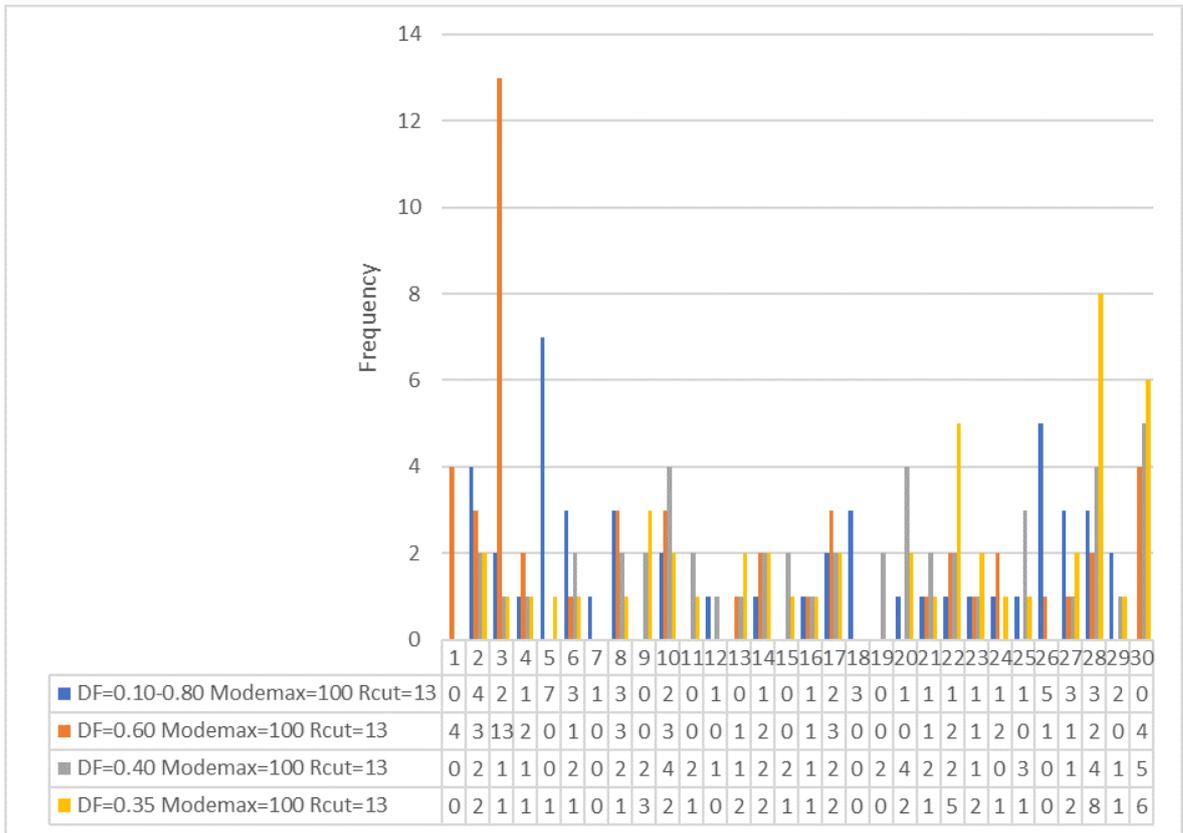


Figure 3.23. ANM mode frequency of parallel simulations from 3UON to 4MQT.

The most selected modes are ANM modes 2 and 3. The hinges residues are the minimum fluctuating residues in the selected ANM modes. Hinge residues of ANM mode 2 are determined as I38, N58, I72-G73, D103, D120-R121, R135, W148, H208, T386, and Y430 as given in Figure 3.24.

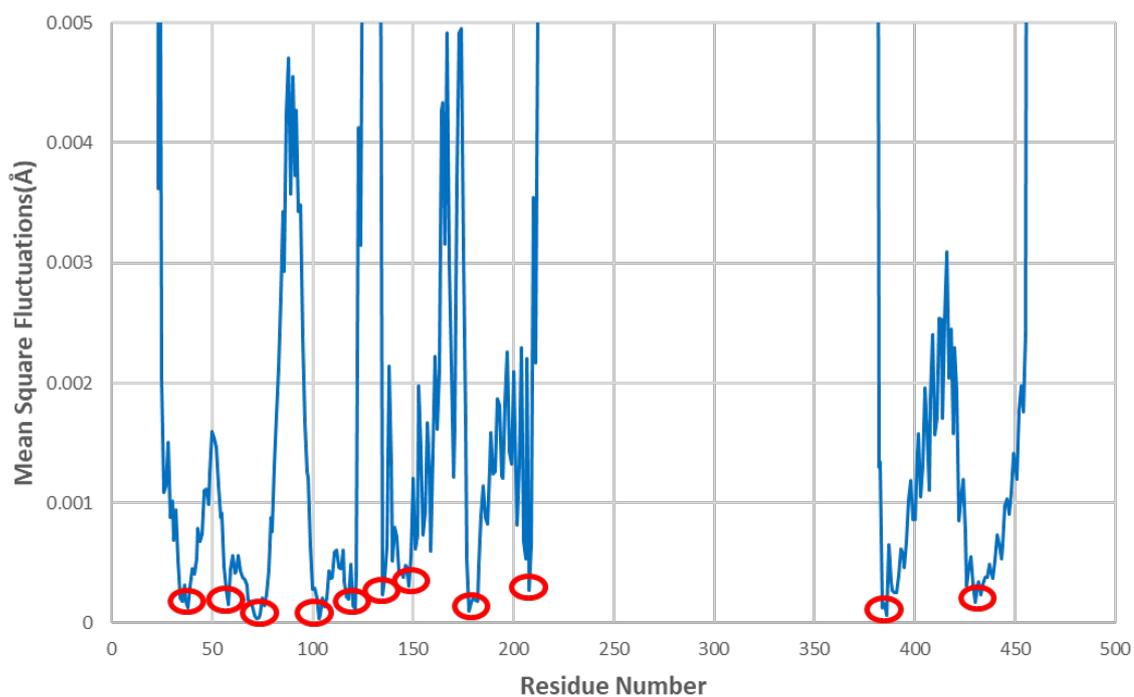


Figure 3.24. ANM mode 2 shape of the inactive state of M2 receptor.

ANM mode 3 is also one of the most preferred modes through the activation of M2 receptor. Hinge residues of ANM mode 3 are determined as F25, G40, N58, L62, Y104, V149-S151, V171, Y177, I200, V385, T388, I392-I397, and P437-Y440 as shown in Figure 3.25.

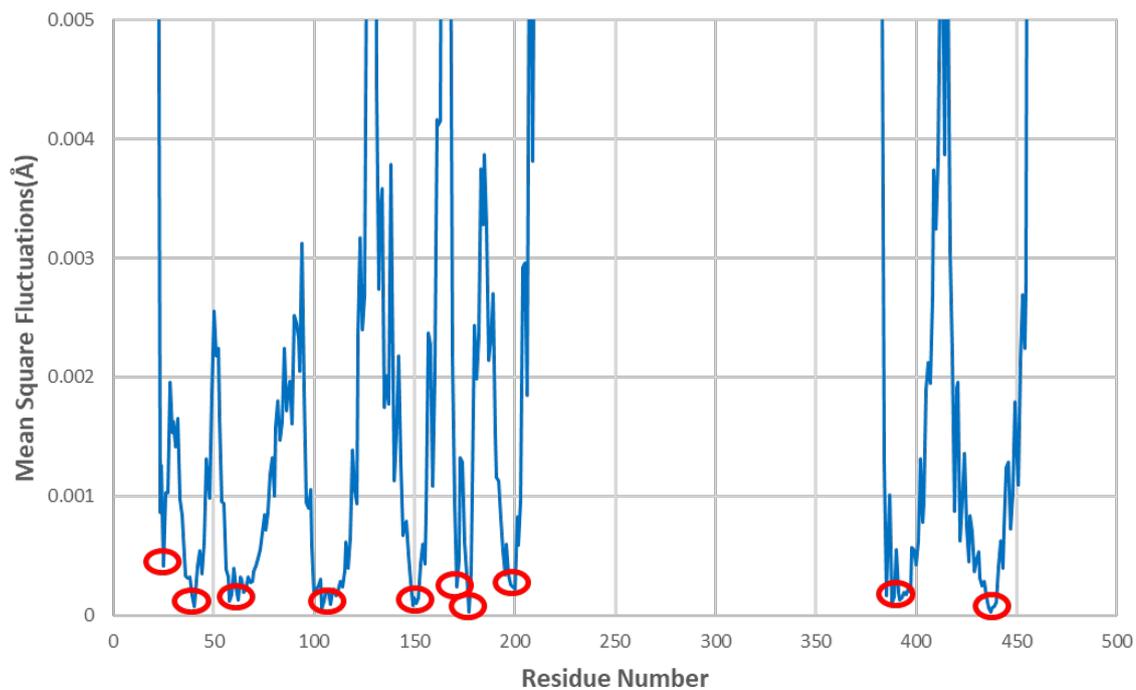


Figure 3.25. ANM mode 3 shape of the initial state of M2 receptor.

The hinge residues of ANM mode 2 and 3 are demonstrated on the inactive structure of M2 receptor in Figure 3.26.

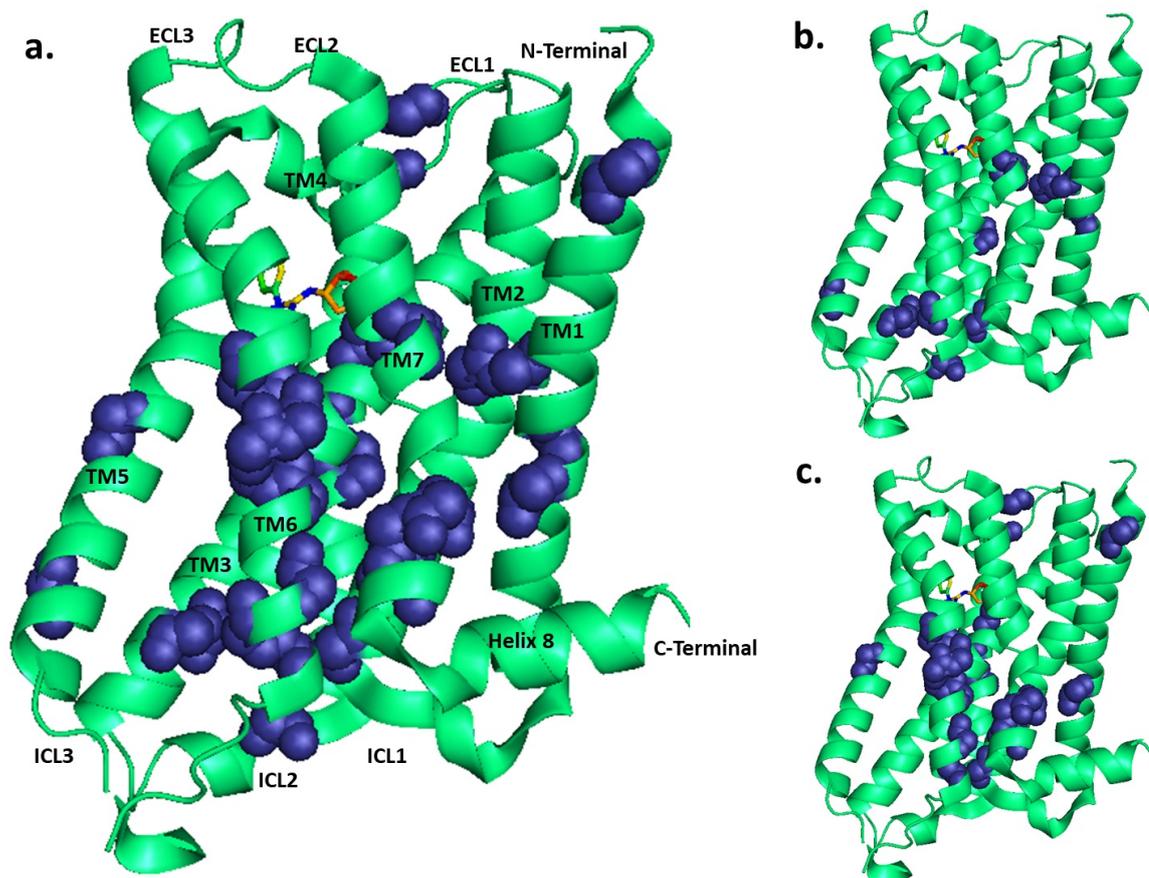


Figure 3.26. Demonstration of hinge residues of ANM mode 2 and 3 (dark blue and sphere) on the structure. a) Hinges of ANM mode 2 and 3. b) Hinges of ANM mode 2. C) Hinges of ANM mode 3.

When hinge residues are examined on the structure, it can be deduced that hinge residues are located around the ligand (QNB) binding site, intracellular loop (ICL2), extracellular loop (ECL2), and at the end of TM3, TM6, and TM7 helices. Unlike the simulations using 4MQS as the active structure, in this case T430 and P437-Y440 in TM7 region are determined as hinge residues, which can be the reason for closer results in the RMSD of NPxxY motif in TM7 and TM3-TM7 distance calculations. Additionally, hinge residues overlap with the residues in orthosteric ligand-binding site. The aromatic modulator located between Y177 in ECL2, hinge residue of ANM mode

3, and W422 in TM7 [40]. Other residues identified as hinge residues that undergo conformational changes during the activation of M2 receptor are Y104 and D103 [40].

Functionally important regions (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, T386, and E382) (red and sphere) and the hinge residues of ANM mode 2 and 3 (deep blue) are demonstrated on the initial structure in Figure 3.27.

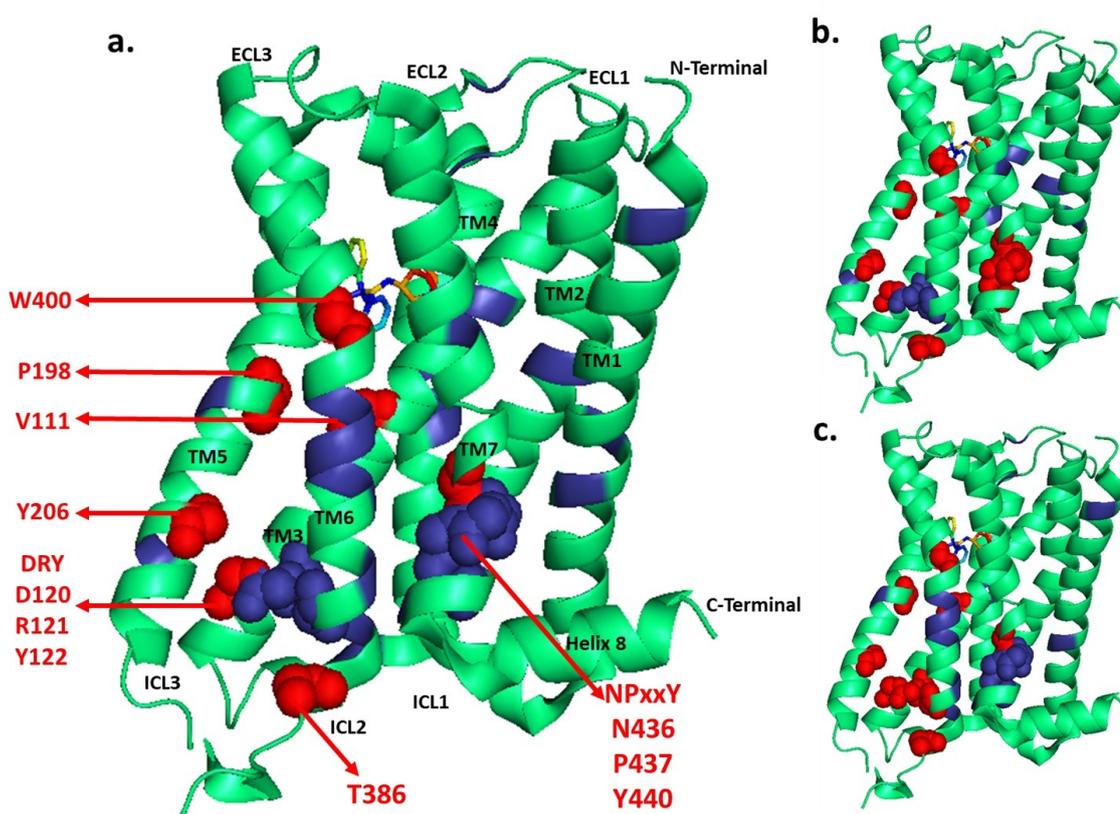


Figure 3.27. Demonstration of ANM mode 2 and 3 hinges (dark blue), functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a)ANM mode 2 and 3 b)ANM mode 2 c)ANM mode 3.

The difference vectors of the two different cases used to examine the activation of the M2 receptor are given in Figure 3.28. When the mean square fluctuations

profiles are examined, it is observed that the difference vectors mostly overlap with each other, while there are slightly differences between two profiles. W422 shows differences between the cases, since W422 has different side chain orientation. Additionally, Y426 in TM7 is another distinctness in mean square fluctuation profile because it interacts with allosteric modulator in the active structure (4MQT). Moreover, T16, V23-V29, L129, S182, S213, K383, G425, N432, A445, and H453 differ between two cases.

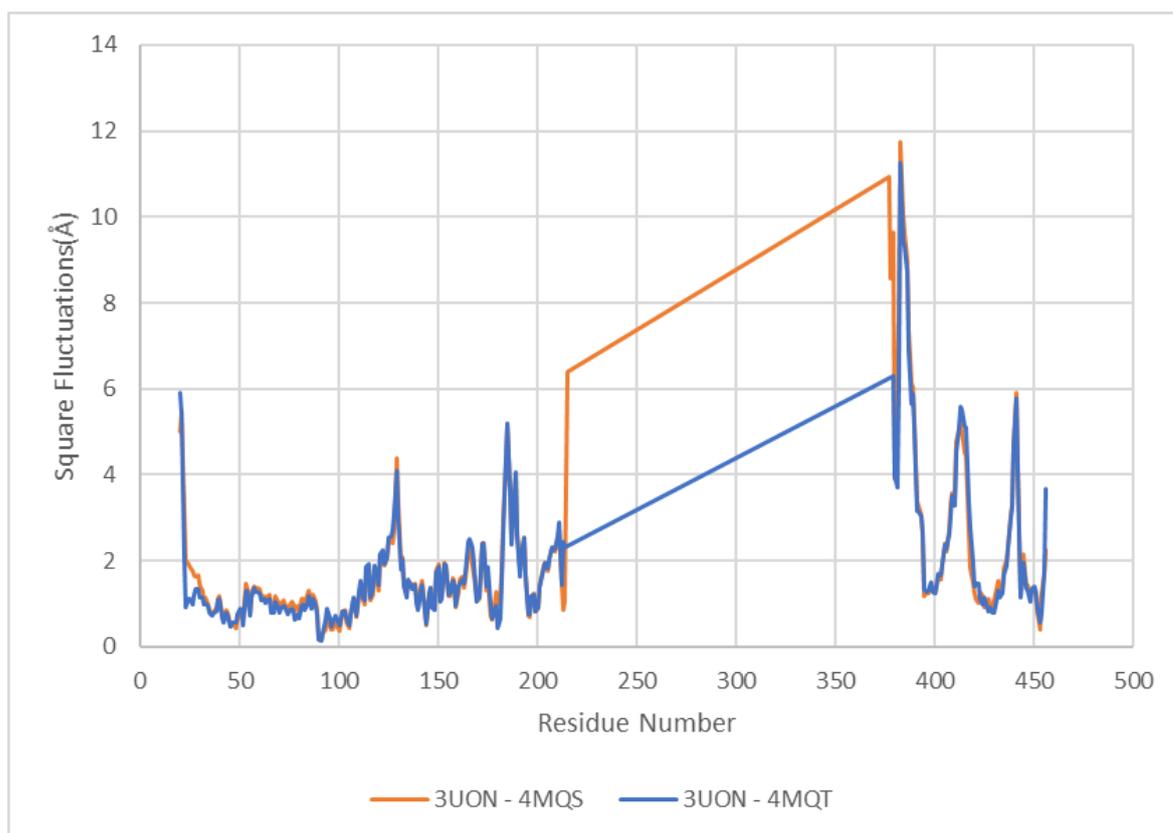


Figure 3.28. Difference vector between initial (3UON) and final (4MQS – 4MQT) structure.

When the most selected modes with highest overlap value are examined, it is observed that different ANM modes are selected in two cases, although ANM modes are determined based on the initial structure. However, MSF profile of ANM modes selected in these two cases show similarities. It can be noted that ANM mode 1 selected in the simulation from 3UON to 4MQS shows similarity with the ANM mode 2 selected

in the simulation from 3UON to 4MQT as given in Figure 3.29. There is a difference in residues at the end of TM6 and in ECL3, which correlates with the differences in the difference vector profile. ANM mode 5 of the simulation from 3UON to 4MQS is quietly different from ANM mode 3 of the simulation from 3UON to 4MQT. It can be deduced from the mean square fluctuations profiles of two cases, that ANM mode 1 of the simulation from 3UON to 4MQS shows similar pattern with ANM mode 2 of the simulation from 3UON to 4MQT, and ANM mode 5 of the simulation from 3UON to 4MQS shows similar pattern with ANM mode 3 of the simulation from 3UON to 4MQT from the stand point of minimum and maximum points. The most obvious difference in ANM mode analysis between Case 1 and Case 2 of M2 receptor in the orthosteric ligand-binding site, which is affected by the presence of activator in 4MQT. In simulations where the 4MQT structure to which the activator is attached is used as the target structure, the hinges of the selected ANM modes are observed more intense in the core region of TM5, TM6 and TM7, which plays an important role in transmitting signal across the membrane. It has also been observed that the evolutionarily conserved NPxxY motif is designated as the hinge point of selected ANM modes in the simulation from 3UON to 4MQT. This is an indication that the activator enables the changes more efficiently in the transition pathway.

In the collective variable analysis, the distance between the residues forming the salt bridge in the inactive structure in the simulation from 3UON to 4MQS, reaches to 22 Å and then drops to 14 Å, after then it fluctuates until approaching the distance value of the target structure. However, the distance between the residues forming the salt bridge in the inactive structure in the simulations from 3UON to 4MQT reaches to 22 Å approximately is the same cycle number as the first case, after that it remains constant for a while and then drops to the distance value in the target structure. During the stabilization of the distance value between the residues forming the salt bridge in the inactive structure, the TM3-TM6 and the TM3-TM7 distances stay also steady. This situation may be the reason of these snapshots represented the intermediate structure.

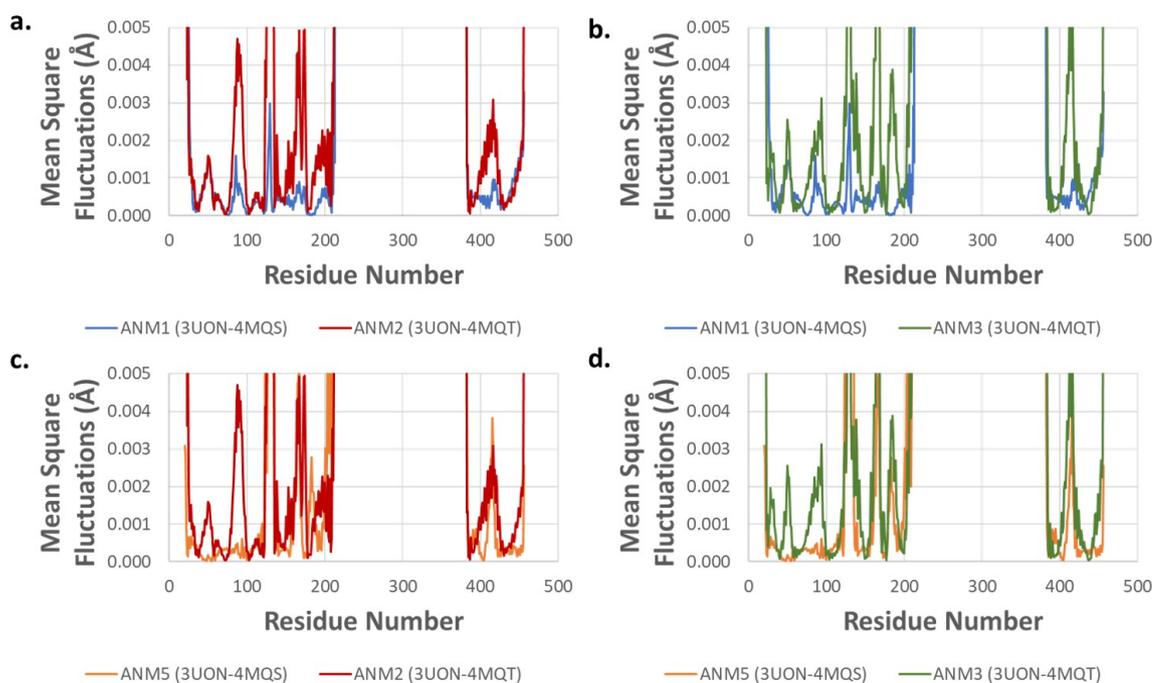


Figure 3.29. Comparison of selected ANM modes in two cases. a) ANM mode 1 and ANM mode 2 b) ANM mode 1 and ANM mode 3 c) ANM mode 5 and ANM mode 2 b) ANM mode 5 and ANM mode 3.

- Cross-correlation Analysis.

Normalized cross-correlation analysis is done through the activation pathway of M2 receptor and demonstrated in Figure 3.30.

According to cross correlation analysis, highly correlated residues are identified as D39&N432 and P157&T386. Since T386 is located at the end of TM6 and opening the gap between the end of TM3 and TM6 is crucial for the activation of the M2 receptor, it can be concluded that P157 plays an important role in activation. Additionally, the strong correlation between in D39 and N432 is unique in this case. In the simulations from 3UON to 4MQS, there is no correlation between N432 located nearby NPxxY motif and any other residues. Moreover, in the difference vector analysis, it is seen that there is a distinctness in N432 between two cases. This means that in the active

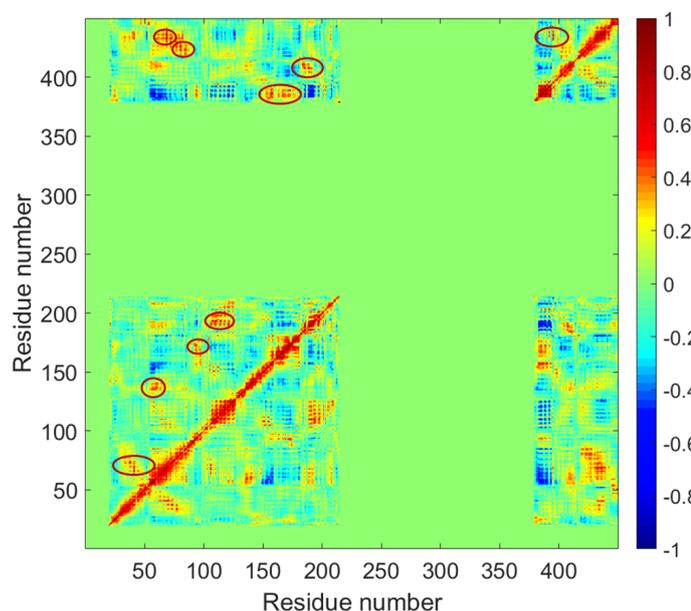


Figure 3.30. Cross correlation map of the activation pathway of M2 receptor from 3UON to 4MQT.

structure (4MQT) to which the activator is attached, the G-protein binding site is derived by TM1 helix according to cross correlation analysis.

### 3.1.2. Activation Pathway of M2 Receptor

The conformational change in GPCR allows transmitting an extracellular signal across the cell membrane to generate an intracellular response. Such a conformational change is necessary for GPCR activation where the GPCR assumes a G protein-binding conformational state. M2 receptor shows the greatest conformational changes in this part on the intracellular side of TM5, TM6 and TM7 during activation through rearrangements in the ligand binding site in TM5, TM6 and TM7. In the activation process of M2 receptor intracellular side of TM6 moves outward, TM5 moves slightly outward and NPxxY motif of TM7 rearranges [40]. Additionally, ionic lock between  $R131^{3.50}$  and  $E382^{6.30}$  is broken through the activation process [36, 42, 43]. In order to understand the dynamic behavior underlying the activation and to determine the

intermediate structures in the activation path from 3UON to 4MQS – 4MQT, some snapshots obtained during the simulation are examined and given in Figure 3.31-3.32.

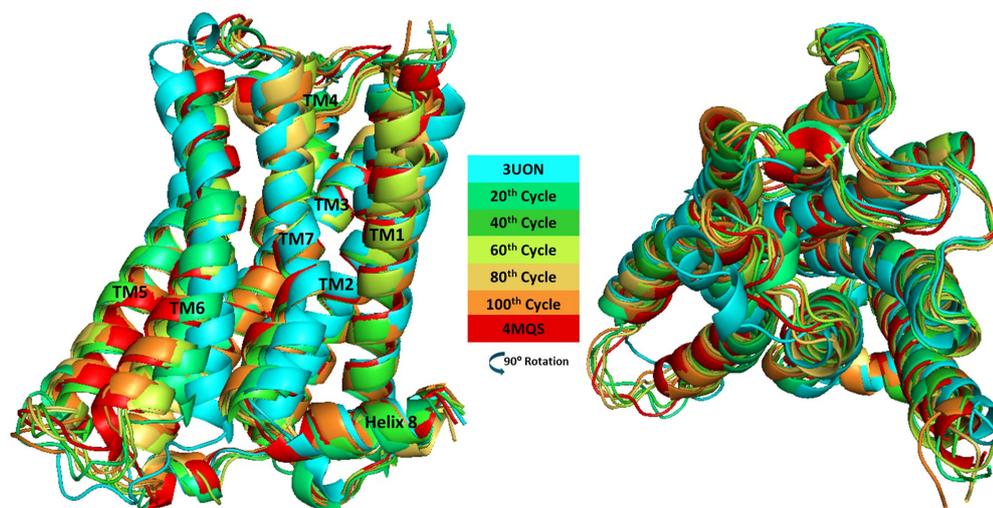


Figure 3.31. Snapshots generated by ANM-LD simulations from inactive (3UON) to active (4MQS) structure.

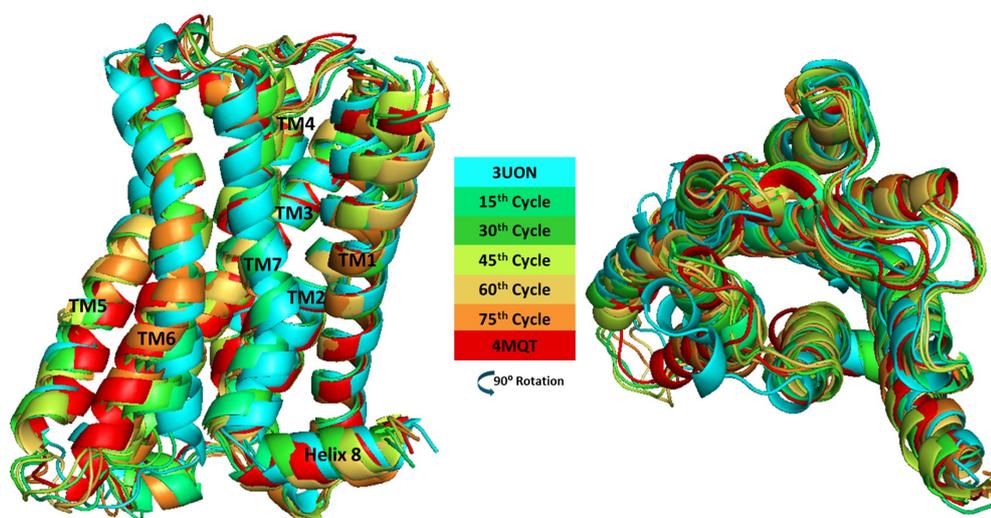


Figure 3.32. Snapshots generated by ANM-LD simulations from inactive (3UON) to active (4MQT) structure.

The intermediate conformations in between inactive and active states generated by ANM-LD simulations are examined one by one through the activation pathway. The important residues on the snapshots obtained from the simulations from inactive structure (3UON) to active structure (4MQS), TM3-TM6 and TM3-TM7 distances are demonstrated in Figure 3.33, 3.34 and 3.35. On account of the breakage of ionic lock in the first twenty cycles, TM6 starts to move towards to the TM5 and the distance between the residues at the end of TM3 and TM6 is increased by 2.1 Å and TM3-TM7 distance is decreased by 3.2. When the intermediate structures estimated via conventional molecular dynamics (cMD) with microsecond-timescale in the literature are examined [36], it is found that TM3-TM6 distance of intermediate structure is 10.0 Å, TM5 is unchanged and NPxxY motif in TM7 moves slightly towards to the TM3 [36]. Considering these conditions, the structure obtained in the 20<sup>th</sup> cycle can be defined as one of the intermediate structures. When the snapshot in the 40th cycle is observed, it is seen that Y206 (G-protein binding site) and T386 at the end of TM6 are located very close with each other. In the 60<sup>th</sup> and 80<sup>th</sup> cycles of the simulation, The TM3-TM6 and TM3-TM7 distances remain approximately constant, while the TM5 continues to move outward. In addition, an outward opening is observed in the extracellular loop (ECL1) of TM2. The distinctness between the active structure and generated final structure from ANM-LD simulations occurs in the NPxxY motif in TM7. Therefore, TM3-TM7 distance does not fully converge to the active structure.

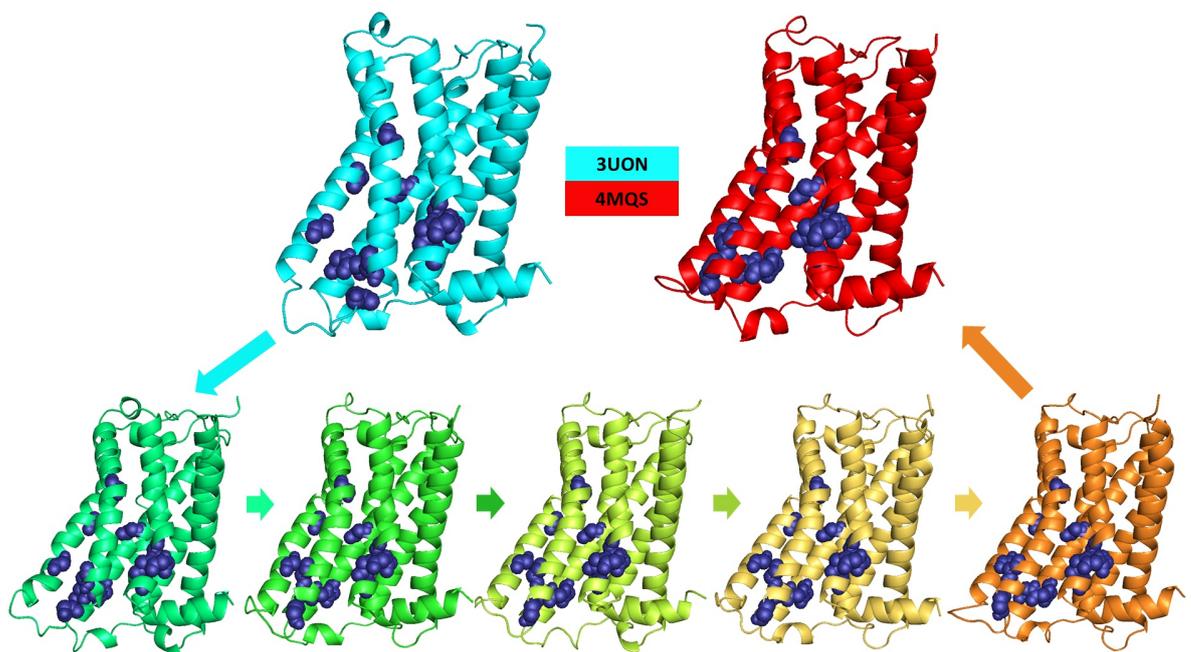


Figure 3.33. Important residues (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, T386, and E382) (dark blue and sphere) demonstrated on the snapshots of the simulations from 3UON to 4MQS.

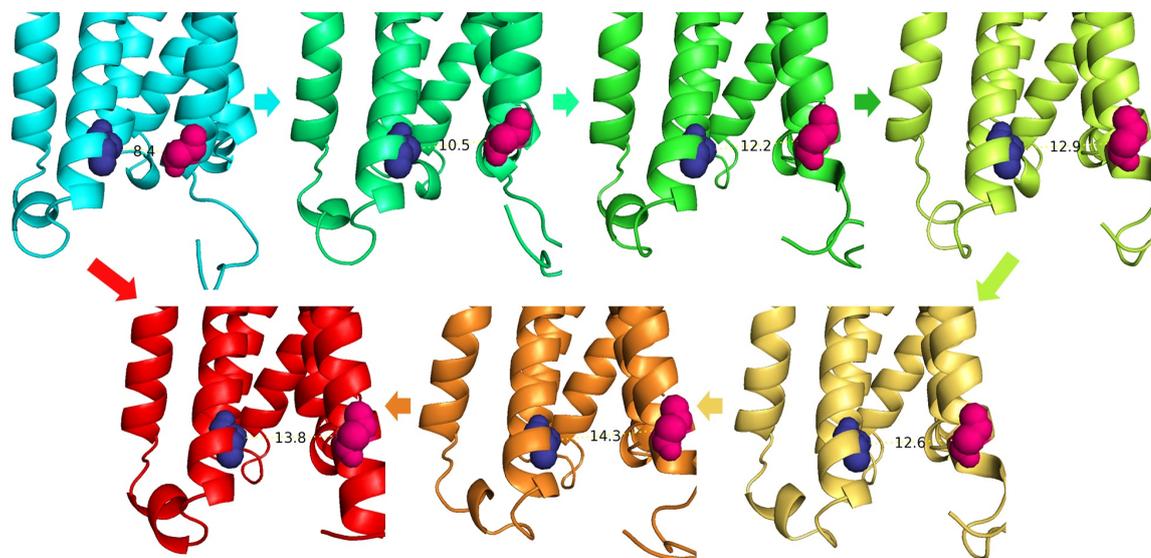


Figure 3.34. The TM3-TM6 distance between R121 (dark blue and sphere) and T386 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQS.

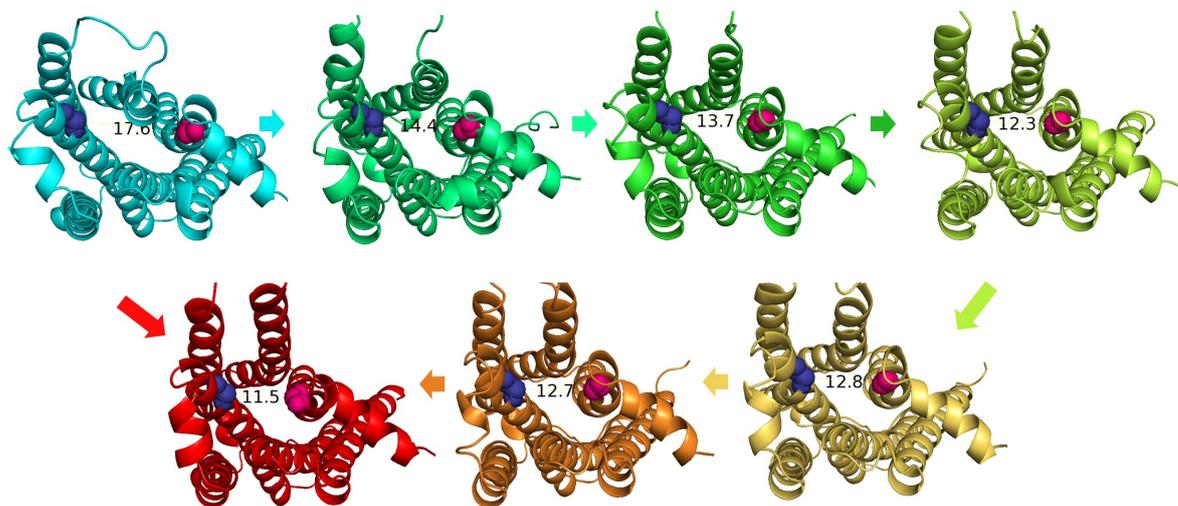


Figure 3.35. The TM3-TM7 distance between R121 (dark blue and sphere) and Y440 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQS.

The active structure of M2 receptor 4MQT bounded to the agonist iperoxo and the positive allosteric modulator, which increases the activation ability of the receptor is used to determine alternative transition pathways of the activation process. The important residues on the snapshots, TM3-TM6 and TM3-TM7 distances are demonstrated in Figure 3.36, 3.37 and 3.38. Intermediate structure proposed in the literature [36] is obtained in the 30<sup>th</sup> cycle. When the snapshot of the 30<sup>th</sup> cycle is investigated, TM3-TM6 distance is determined as 10.3 Å, ionic lock between the residues R121<sup>3.50</sup> and E382<sup>6.30</sup> is broken, and TM5 remains unchanged. The communication between Y206 (G-protein binding site) and T386 at the end of the TM6 is formed in the 75<sup>th</sup> cycle unlike the first case in which it is observed in the 40<sup>th</sup> cycle. W422 in TM7 which bounds to the allosteric modulator, has almost same conformation at the 75<sup>th</sup> cycle with the active structure.

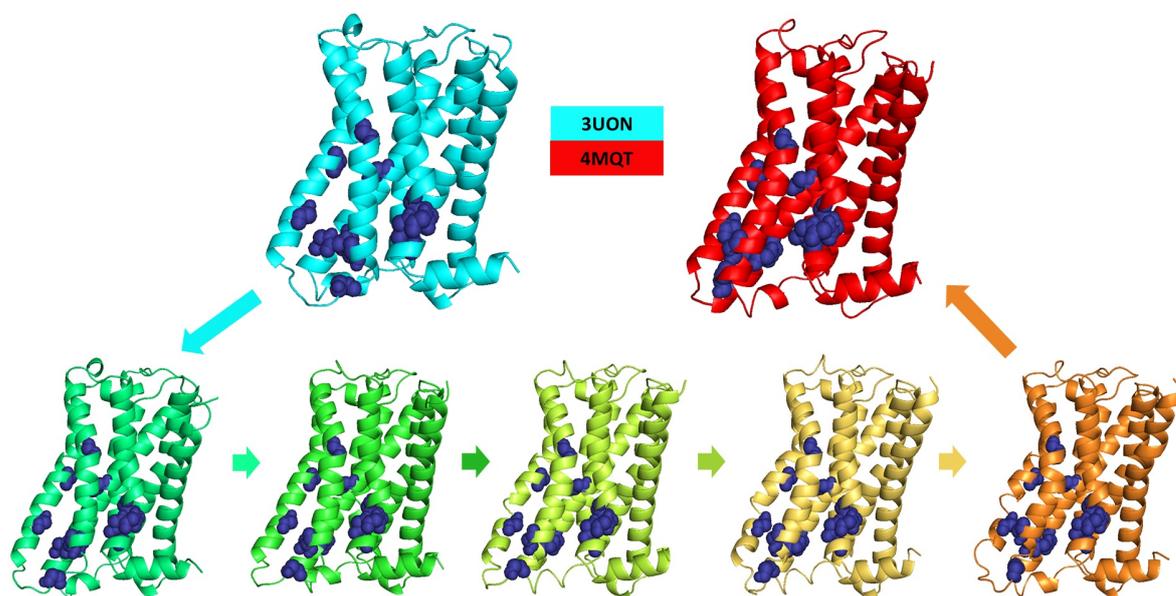


Figure 3.36. Important residues (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, T386, and E382) (dark blue and sphere) demonstrated on the snapshots of simulations from 3UON to 4MQT.

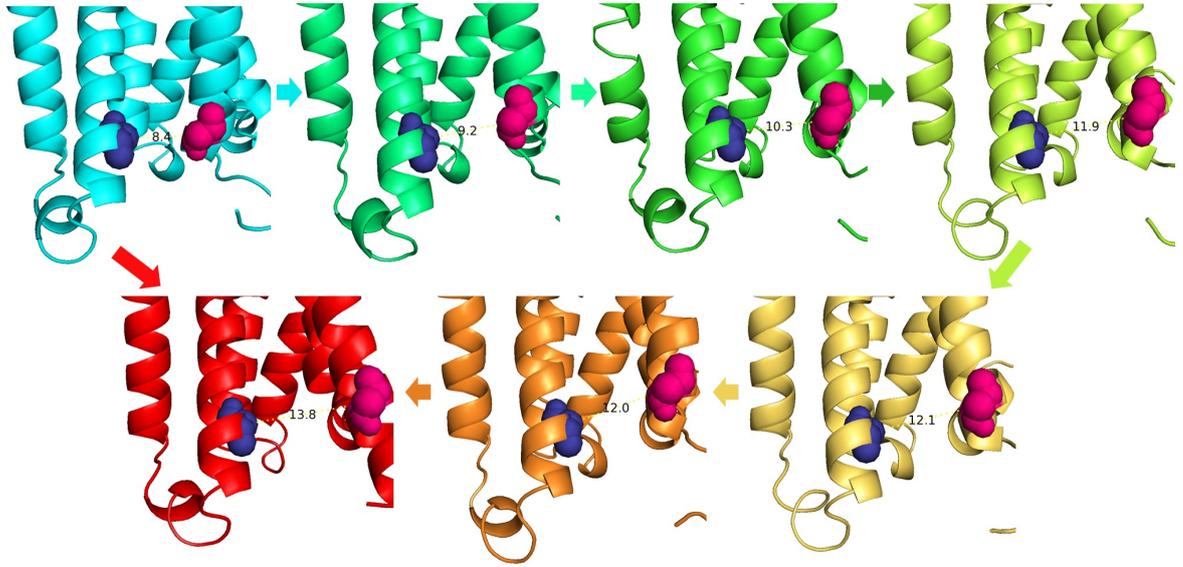


Figure 3.37. TM3-TM6 distance between R121 (dark blue and sphere) and T386 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQT.

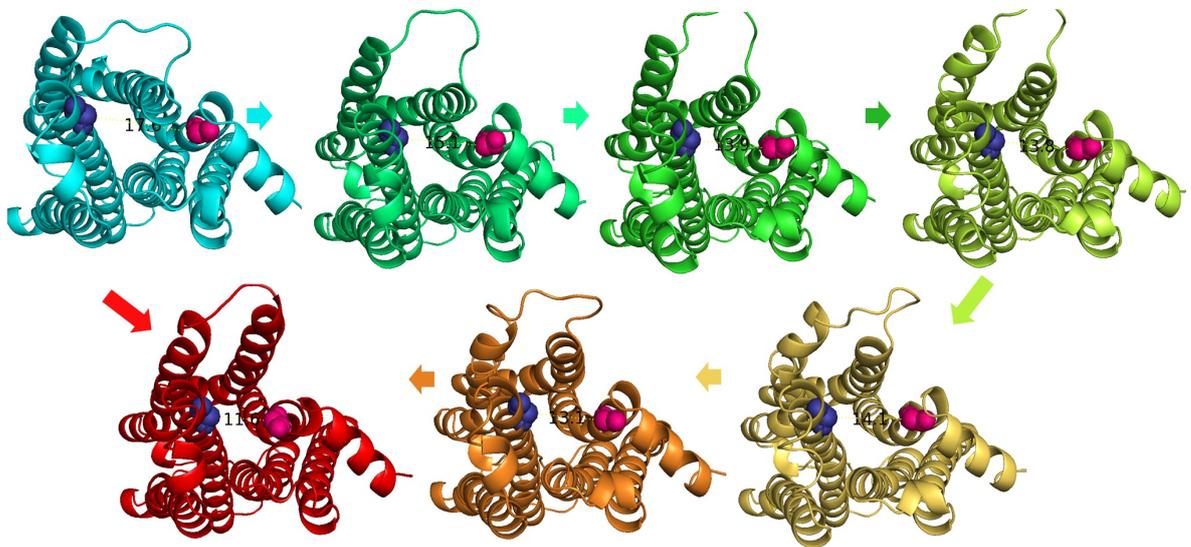


Figure 3.38. TM3-TM7 distance between R121 (dark blue and sphere) and Y440 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3UON to 4MQS.

Intermediate structures obtained from ANM-LD simulations of both cases is given in Figure 3.39. The RMSD of intermediate structure of the simulation from 3UON to 4MQS with respect to the intermediate structure of the simulation from 3UON to 4MQT is calculated as 1.4 Å. The main difference between two structures observed in the extracellular sides (ECL1-ECL2-ECL3), extracellular side of TM1 and TM2, and the intracellular sides (ICL2-ICL3). The distinctness in the extracellular side is caused from the allosteric modulator bounded to the active structure (4MQT). It can be deduced that, the attachment of this modulator to the extracellular side of the structure contributes to the changes in the intracellular side, although there is no difference in the intracellular side of the active structure [40].

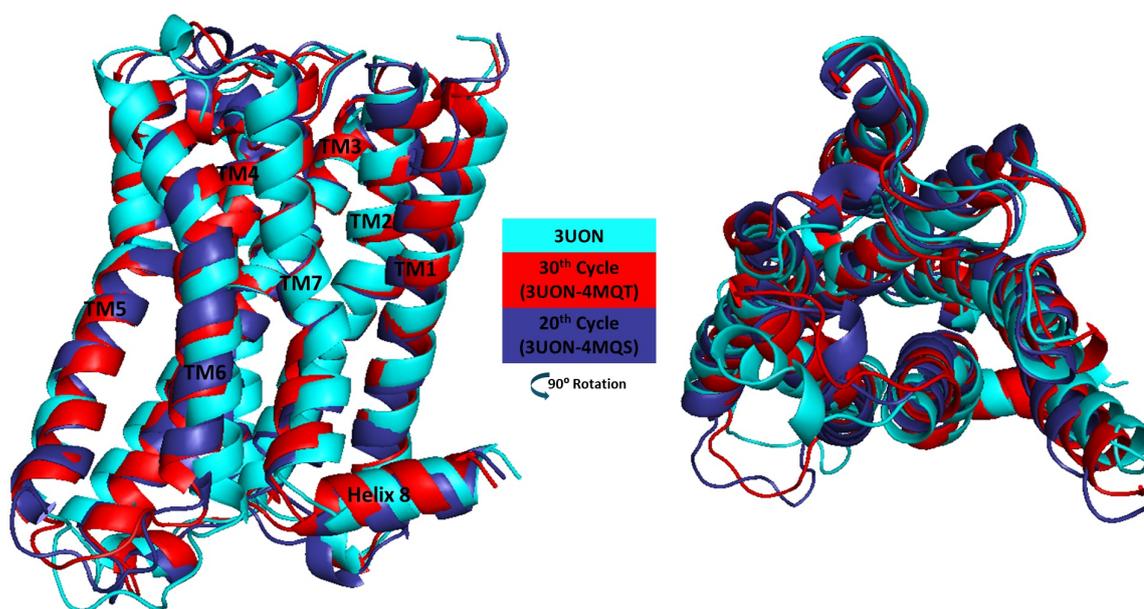


Figure 3.39. Alignment of intermediate structures obtained by ANM-LD simulations of both cases.

The final structures obtained from ANM-LD simulations of both cases is given in Figure 3.40. The RMSD of intermediate structure of the simulation from 3UON to 4MQS with respect to the intermediate structure of the simulation from 3UON to 4MQT is calculated as 1.7 Å. The main difference between two structures detected in

the extracellular sides (ECL1-ECL2-ECL3), extracellular side of TM2, the intracellular sides (ICL2-ICL3), NPxxY motif, and E382.

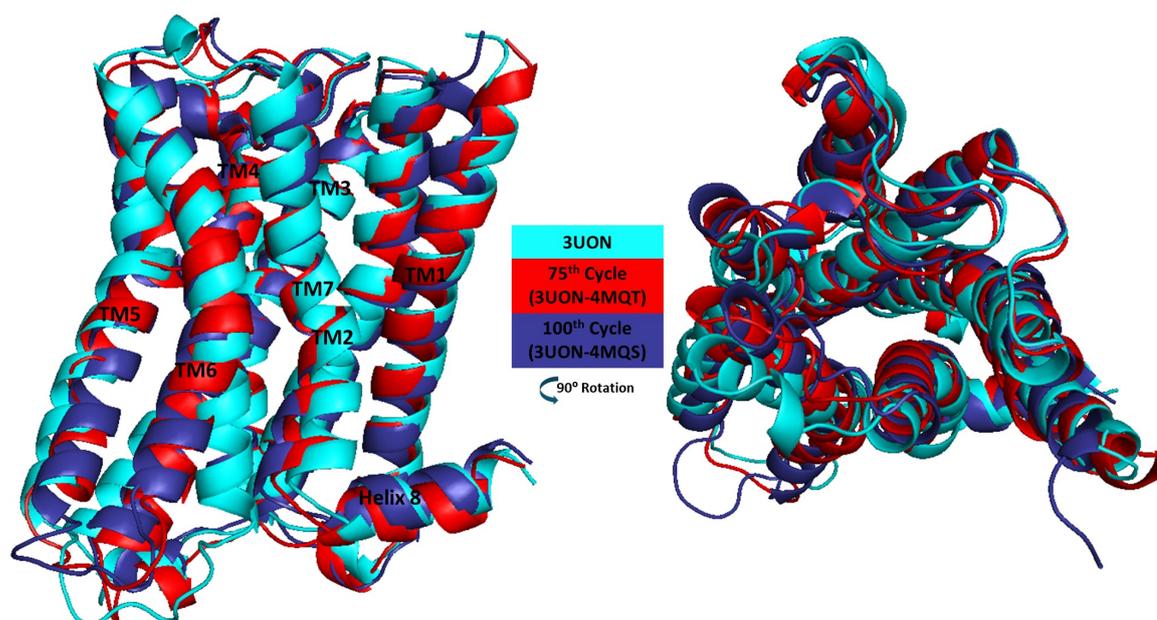


Figure 3.40. Alignment of final structures obtained by ANM-LD simulations of both cases.

### 3.1.3. Allosteric Interactions in M2 Receptor

Allostery is an intricate dynamic phenomenon shown by many proteins, which is defined as changing a property in a distant part of the protein by attaching a small molecule to another place. While ligands usually bind to the orthosteric domain of GPCR, there are also ligands that act as allosteric modulators or bind to another site. These allosteric modulators can affect the structure and the dynamics of the structure. Many allosteric modulators appear to bind to the extracellular or intracellular surface of the receptor and affect the orthosteric binding site [44]. The allosteric agonist can affect the receptor even in the absence of an orthosteric agonist. The effect of an allosteric drug may directly modulate the signal in the GPCR or change the affinity, efficacy, and selectivity of the orthosteric ligand. The knowledge of conformational dynamics

is also important for identifying the differences of associated receptor subtypes, which can enable the design of drugs that select for one subtype over another. Through cross-correlation and ANM mode analysis of the transition pathways revealed from ANM-LD simulations, residues in allosteric sites are determined through the selected ANM modes and cross-correlation analysis for the two cases. As a result of the study conducted in [45] to find druggable binding sites, they brought seven different allosteric regions and important residues located in these regions to literature. They mapped the surface of low-energy structures obtained by accelerated MD simulation using fragment-based mapping algorithm (FTMAP) with 16 different organic solvent molecules [45].

As a result of this study, a list of important residues found in allosteric regions of inactive, active, and intermediate structures is given. The residues in this list are compared in Table 3.5 with the residues determined to be significant as a result of the hinges of selected ANM modes and cross correlation analysis, and intersecting residues are underlined.

According to the study conducted by Miao et al., it was observed that there are seven allosteric sites for the M2 receptor. When the simulations from 3UON to 4MQT is analyzed by ANM mode analysis, Y177, which is determined as a hinge residue of ANM mode 3, is coincided with the key interacting residue in the allosteric site located in ECL2 region [45]. This site is determined as conserved allosteric site in different subtypes of M2 receptor and targeted to design allosteric modulator. It has been observed that the inactive conformation of the ECL2 region is in the form of an “open mouth”, while the conformation of active conformation is called “closed-mouth” [45]. For this reason, this region is an important target for allosteric drug design. In the simulations from 3UON to 4MQS and 4MQT the conformational change from open mouth to closed mouth is observed and given in Figure 3.41. In Figure 3.41, conformational change in ECL2 is demonstrated in inactive structure (cyan), intermediate structure (green), final structure generated by ANM-LD simulation (orange), and the active structure (red). Second allosteric side according to research conducted by Miao et al. is located at the interface between TM5 and TM6. L393 and I397 proposed as

Table 3.5. Key collaborating residues in allosteric sites based on ANM mode and cross correlation analysis.

	<b>3UON-4MQS</b>	<b>3UON-4MQT</b>
<b>TM1</b>	V24, S32-L35	F25, I38,G40
<b>ICL1</b>	I48, Q55	<u>N58</u> , N59
<b>TM2</b>	<u>S76</u>	<u>L62</u> , F63-A68, <u>D69</u> , I72-G73
<b>ECL1</b>	L100- <u>Y104</u>	<u>D103</u> , <u>Y104</u>
<b>TM3</b>	-	-
<b>ICL2</b>	<u>D120-Y122</u> , K134- <u>R135</u>	<u>D120-Y122</u> , <u>R135</u>
<b>TM4</b>	A140, L154, <u>I159</u>	V149-S151
<b>ECL2</b>	D173-C176, F181-A185	V171, <u>Y177</u>
<b>TM5</b>	A191, Y196	I200, H208
<b>ICL3</b>	<u>I209</u> , V385, <u>T386</u>	V385, <u>T386</u> , T388
<b>TM6</b>	I392, <u>L393</u> , N404, <u>V407</u>	I392- <u>I397</u>
<b>ECL3</b>	-	-
<b>TM7</b>	<u>T423</u> , <u>Y426</u> , <u>Y430</u>	<u>Y430</u> , <u>P437-Y440</u>
<b>Helix-8</b>	F447, T450	-

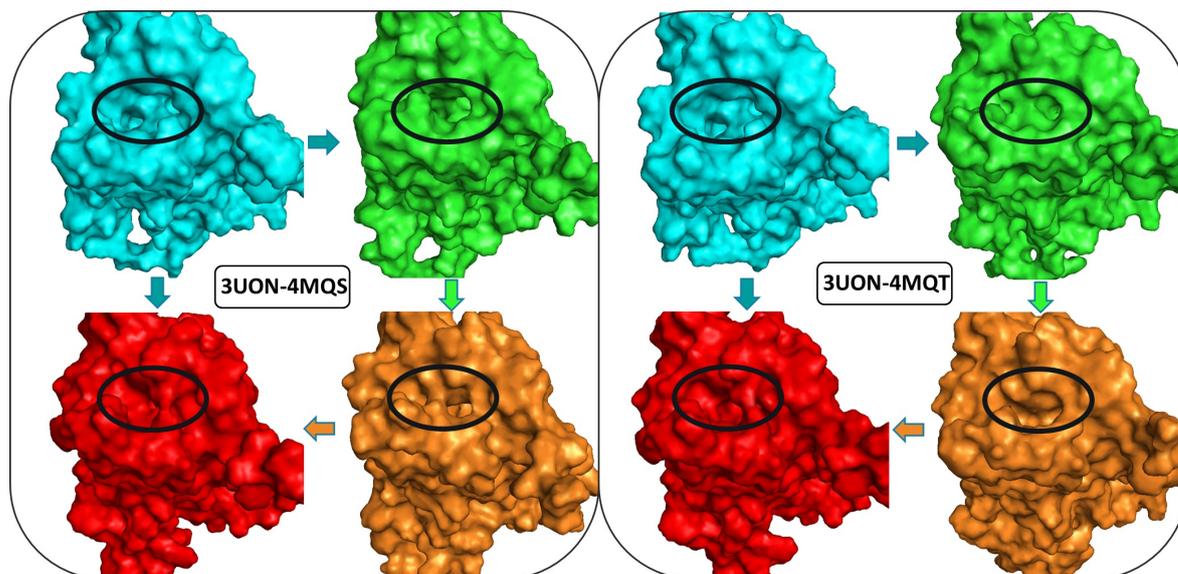


Figure 3.41. Conformational changes in ECL2 allosteric site through the transition pathways generated by ANM-LD simulations.

interacting with probe molecules are identified as hinge residues in ANM mode analysis in the simulations from 3UON to 4MQS and 3UON to 4MQT, respectively [45]. Another proposed allosteric site contains the DRY, which is preserved in the GPCR family. Therefore, it has been observed as an allosteric site in most of the GPCR family members, and the DRY motif coincide with the hinge residues of mostly selected ANM modes. It is observed that the 5<sup>th</sup> and 6<sup>th</sup> allosteric sites TM1, TM2 and TM7 are in the intracellular and extracellular parts. The 5<sup>th</sup> allosteric site could not be detected in this study [45]. The 6<sup>th</sup> allosteric site affects the movement of the NPxxY motif during activation. Because the NPxxY motif is a common feature in the GPCR family, this site is also referred to as the common allosteric site for the GPCR family. Finally, the region where the G-protein is bound is determined as the allosteric site. During activation, the intracellular side of TM6 opens outward, creating a cage for G-protein binding. During this process, the DRY motif, NPxxY motif, Y206 and T386 play an important role. When the allosteric sites given in this article are examined, it is observed that the 4<sup>th</sup> and 5<sup>th</sup> allosteric regions are not found in our analysis, which are located between TM3 and TM4 in intermediate structure; and between TM1, TM2, and TM7 helices, respectively.

### 3.2. $\beta$ 2-Adrenergic Receptor

$\beta$ 2-Adrenergic receptor is the most studied protein which belongs to the family of G-protein coupled receptor (GPCR). This receptor is activated by the isoprenaline and adrenaline and  $\beta$ 2-Adrenergic agonist are used in bronchodilator therapy, treatment in hypertension, chronic obstructive pulmonary disease, preterm labor, and asthma [46, 47].

$\beta$ 2-Adrenergic receptor has an important role in understanding the activation of G proteins (G $\alpha$ s and G $\alpha$ i). The activation of  $\beta$ 2AR regulates adenylate cyclase, which leads to the formation of cyclic AMP (cAMP). Cyclic AMP takes a role in the activation of protein kinase A (PKA) that is responsible for the regulation of various cellular proteins such as L-type  $Ca^{+2}$  channel and  $\beta$ 2AR. Additionally, the activation of  $\beta$ 2AR causes the phosphorylation of G-protein-coupled-receptor kinase (GRK) and activated  $\beta$ 2AR bounds to arrestin [48].

$\beta$ 2-Adrenergic receptor has seven transmembrane helices, three intercellular loops (ICL1-ICL2-ICL3) and three extracellular loops (ECL1-ECL2-ECL3) like all other receptors in GPCR family. Additionally, Helix 8 is situated in  $\beta$ 2AR structure, which is prevalent in Class A GPCRs [47]. The structure and corresponding regions of  $\beta$ 2AR is given in Figure 3.42.

The residue number of  $\beta$ 2-Adrenergic receptor starts with 29 and ends with 341. There are 23 missing residues between 240 and 262, which locates in ICL3. The corresponding regions of residues is given in Figure 3.43.

GPCRs with crystal structures both in inactive and active states appear to show similar conformational changes between different subspecies. However, the activation of different GPCRs varies in several ways. Agonist binding pocket interactions differ significantly between GPCRs, indicating that different GPCRs have evolved to recognize very different ligands. Additionally, although there are similar conformational

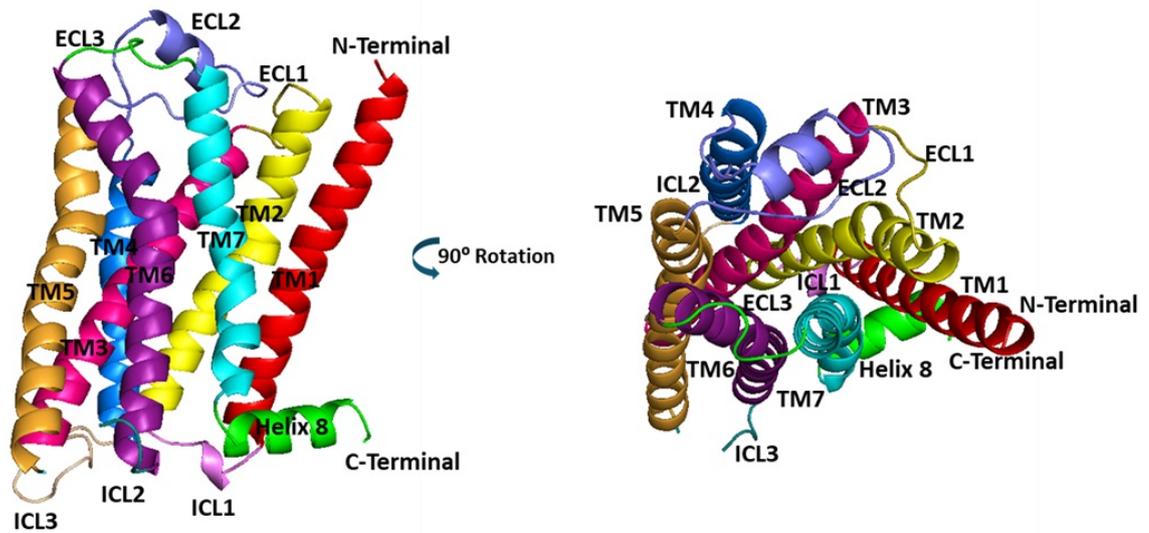


Figure 3.42. The structure and corresponding regions of  $\beta_2$ -Adrenergic receptor.

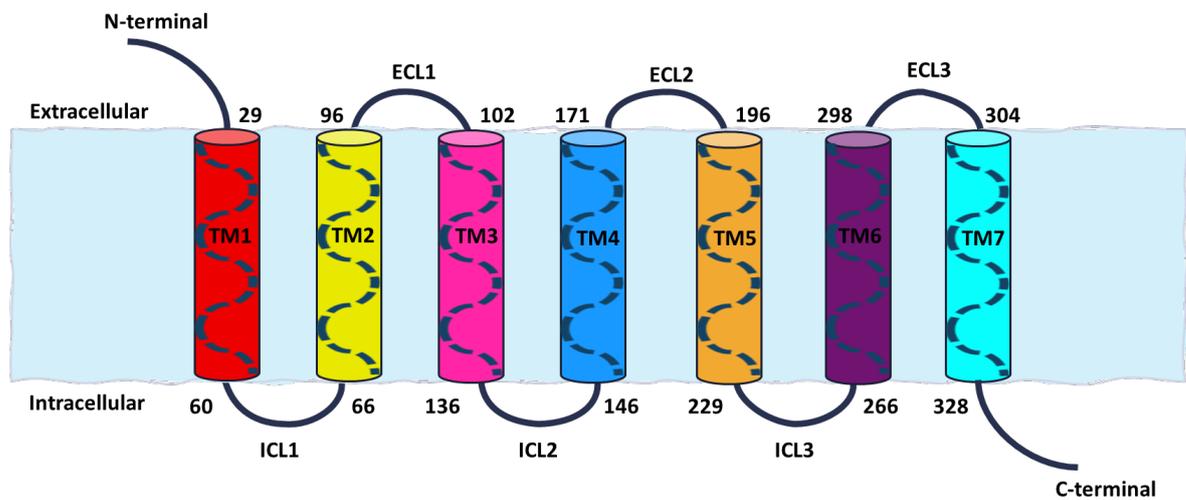


Figure 3.43. The wire diagram of corresponding regions and residue numbers of  $\beta_2$ -Adrenergic receptor.

changes on the intracellular surfaces during the activation and binding to intracellular protein sets, there are differences in the interaction between the binding pocket and the intracellular interface [1].  $\beta$ 2-Adrenergic receptor is used as model structure to understand the GPCR activation mechanism. During the activation, conformational changes occur especially in three different regions, which are called as ligand-binding site, Connector, and G protein binding site. The most important conformational change is seen in the intracellular parts of TM5, TM6, and TM7. The distance between R131<sup>3.50</sup> (DRY Motif) in TM3 and Leu272<sup>6.34</sup> in TM6 is increases while the distance between R131<sup>3.50</sup> (DRY Motif) in TM3 and Y326<sup>7.53</sup> (NPxxY Motif) in TM7 decreases (Ballesteros–Weinstein residue numbering [41] [15, 49]. Additionally, the salt bridge breakage through the activation of  $\beta$ 2AR in one of the key characteristics of GPCR family. The residues forming salt bridge are R131<sup>3.50</sup> and E268<sup>6.30</sup> [50]. The change in the ligand-binding site through activation of  $\beta$ 2AR is the movement of TM5, which is caused from the hydrogen bond between S207<sup>5.46</sup> and p-hydroxyl group of agonists. I121<sup>3.40</sup> enables the intracellular end of TM6 to move from TM3 by shifting toward TM6. G-protein binding sites are determined as Y219<sup>5.58</sup> in TM5 and Y326<sup>7.53</sup> in TM7 [15]. Functionally important residues and corresponding regions of  $\beta$ 2AR are given in Table 3.6 and Figure 3.44.

Table 3.6. Residues and corresponding regions of functionally important sites of  $\beta$ 2-Adrenergic receptor.

Important Sites	Residue Names	Region
TM3 – TM6 Distance	R131 – L272	TM3 – TM6
TM3 – TM7 Distance	R131 – Y326	TM3 – TM7
Salt Bridge	R131 – E268	TM3 – TM6
DRY Motif	D130 – R131 – Y132	TM3
NPxxY Motif	N322 – P323 – Y326	TM7
Connector	I121	TM3
Ligand Binding Site	S207	TM5
GP Binding Site	Y219	TM5

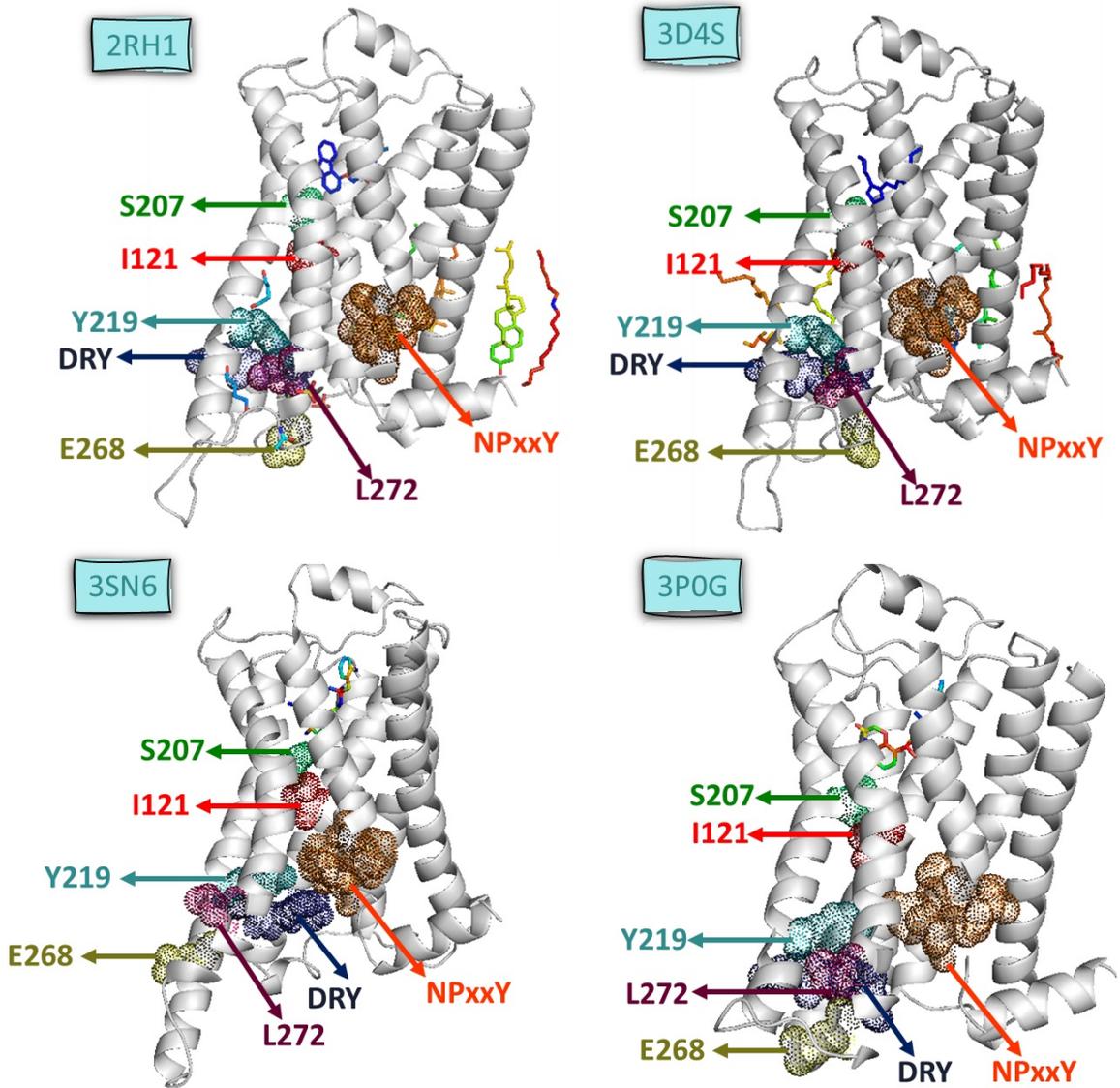


Figure 3.44. Functionally important sites of  $\beta_2$ -Adrenergic receptor.

### 3.2.1. ANM-LD Simulations of $\beta$ 2-Adrenergic Receptor

In ANM-LD simulations of  $\beta$ 2-Adrenergic receptor, two inactive and two active structures are used, which are obtained from Protein Data Bank with PDB IDs 3D4S, 2RH1, 3SN6, and 3P0G [47, 49, 51, 52]. 2RH1 and 3D4S are the inactive structures of  $\beta$ 2-Adrenergic receptor and used as initial structures in the ANM-LD simulations. Target structures used in the simulations are nanobody-stabilized active state of the  $\beta$ 2-Adrenergic receptor (3P0G) and the  $\beta$ 2-Adrenergic receptor-Gs protein complex (3SN6). Properties of  $\beta$ 2-Adrenergic receptor structures are given in Table 3.7.

Table 3.7. Summary of used structure in ANM-LD simulations of  $\beta$ 2-Adrenergic receptor.

PDB ID	State	Complex	Organism	Mutation	Structure
3D4S	Inactive	Timolol Bound	<i>Homo Sapiens</i>	W122E E187N	Shortened length of C-terminus
2RH1	Inactive	Carazolol Bound	<i>Homo Sapiens</i>	E187N	Minor shift in ICL2 and ECL3
3SN6	Active	Gs-protein Complex	<i>Homo Sapiens</i>	T96M T98M E187N	3 Å outward movement at the end of TM6
3P0G	Active	Nanobody Stabilized	<i>Homo Sapiens</i>	E187N	ARG131 has different positions, because it reacts with nanobody

In order to examine the activation process of the  $\beta$ 2-Adrenergic Receptor, for different cases are studied. These cases are from inactive structure (3D4S) to active structures (3SN6-3P0G), and from inactive structure (2RH1) to active structures (3SN6-3P0G), which are given in Figure 3.45.

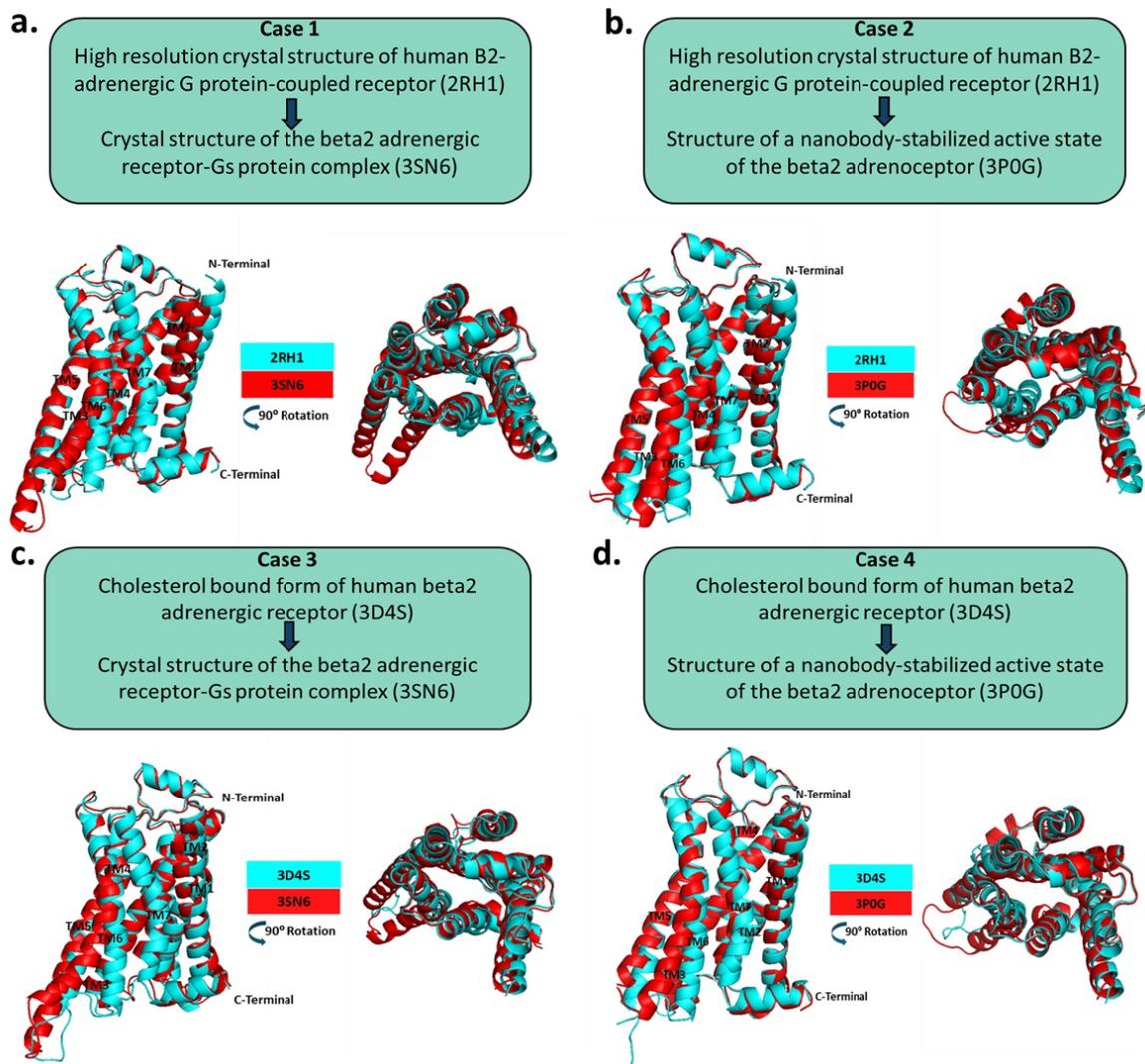


Figure 3.45. Investigated cases in ANM-LD simulations of  $\beta$ 2-Adrenergic receptor.  
 a) 3SN6 to 2RH1 (RMSD: 2.93 Å). b) 3P0G to 2RH1 (RMSD: 2.61 Å). c) 3SN6 to 3D4S (RMSD: 3.22 Å). d) 3P0G to 3D4S (RMSD: 2.71 Å).

3.2.1.1. ANM-LD Simulations from 2RH1 to 3SN6. In the first case of the investigation  $\beta$ 2-Adrenergic receptor activation process, simulations are performed from the inactive structure 2RH1 to the active structure 3SN6. For the preparation of ANM-LD simulations that use unmodified structures, D29, A176-H178, K236-F264, and L342 from the initial structure; Q231-R239 from the final/target structure are deleted. Secondly, A176-H178, K236-F264, and Q231-R239 residues located in ECL2, and ICL3 are modeled by MOD-LOOP [53]. Thirdly, mutations of the initial structure (E187N) and final structure (T96M, T98M, and E187N) are withdrawn by PyMOL [54] and back-mutated structures are used in the ANM-LD simulations.

Parallel ANM-LD simulations are prepared to examine the activation process of  $\beta$ 2-Adrenergic receptor and the result summaries are given in Table 3.8.

Table 3.8. ANM-LD Result Summaries.

Simulation Name	Cycle Number	Initial RMSD (Å) to Target Structure	Final RMSD (Å) to Target Structure	The Most Selected ANM Modes
$D_F = 0.60\text{Å}$ Modemax = 30 $R_{cut} = 8\text{Å}$	40	3.16	1.76	7, 9, 11
$D_F = 0.40\text{Å}$ Modemax = 30 $R_{cut} = 8\text{Å}$	67	3.16	2.07	6, 5, 13
$D_F = 0.60\text{Å}$ Modemax = 30 $R_{cut} = 8\text{Å}$	52	3.16	2.03	13
$D_F = 0.10 - 0.40\text{Å}$ All Modes $R_{cut} = 8\text{Å}$	65	3.16	2.28	5, 13, 20

Root mean square deviations (RMSD) with respect to target structure through the transition/activation pathway in ANM-LD simulations followed and given in Figure 3.46. The RMSD of the initial structure used in the simulation with respect to

the target structure is calculated as 3.16 Å for the structure of which ICL2 and ECL3 regions are modeled, and 2.93 Å for the unmodified structure. RMSD of the generated structures from ANM-LD simulations are converged around 50 cycles and scatter between 2.28 Å and 1.70 Å. The RMSD results of four runs with best convergence are given in Figure 3.47.

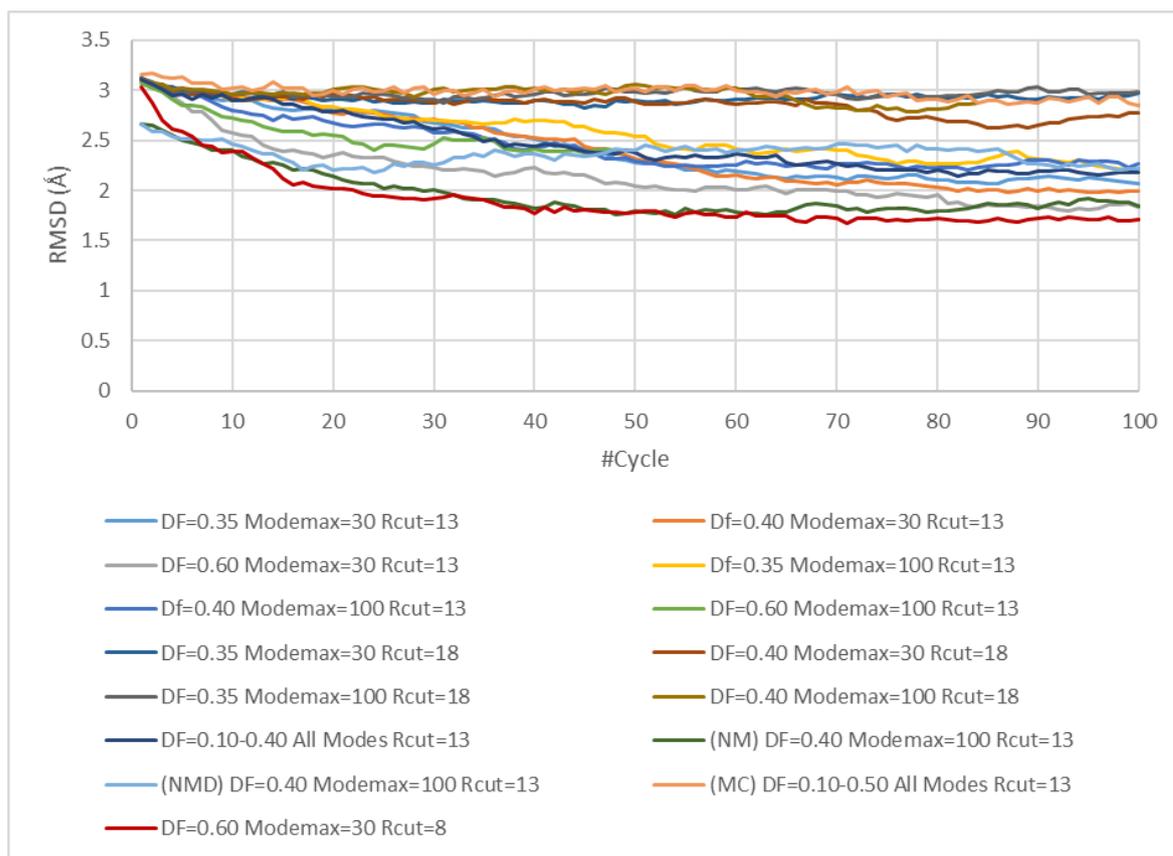


Figure 3.46. The RMSD of  $\beta 2$  adrenergic receptor from 2RH1 to 3SN6 in the ANM-LD simulations.

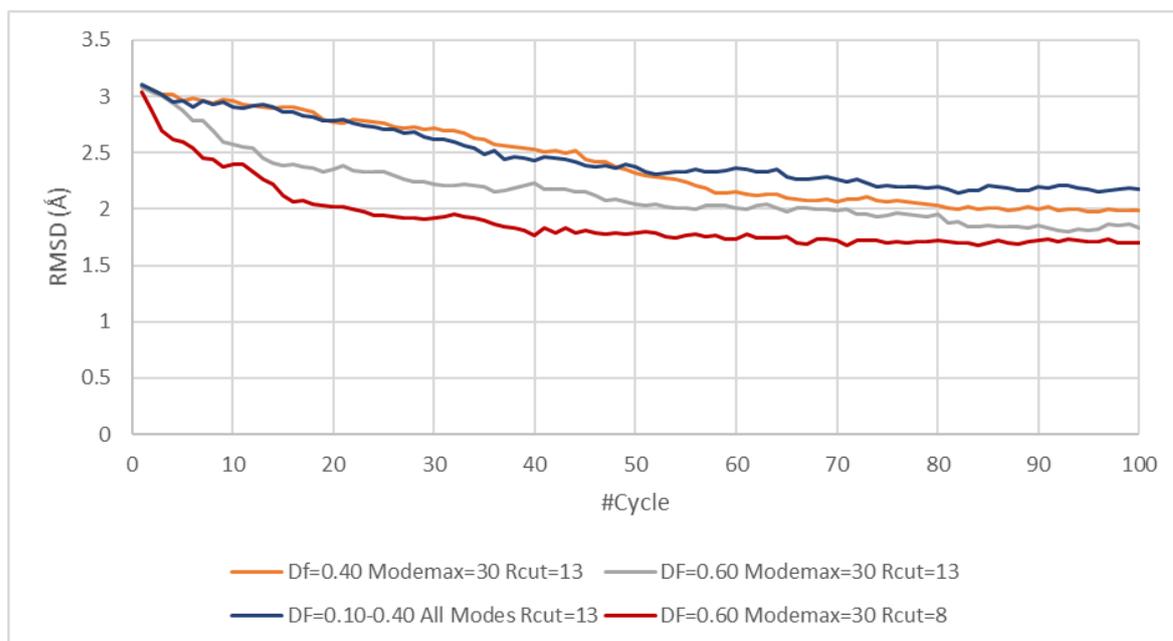


Figure 3.47. The RMSD of selected simulations of  $\beta 2$  adrenergic receptor from 2RH1 to 3SN6 in the ANM-LD simulations.

During the activation process, the changes of the locations of the important sites of the receptor are examined.

- TM3-TM6 and TM3-TM7 Distances and Salt Bridge Breakage.

Firstly, TM3-TM6 and TM3-TM7 distances, and the distance between the residues forming salt bridge in the inactive case are calculated to examine the movement of residues that are important for the activation of the  $\beta 2$  adrenergic receptor. In the calculation process,  $C\alpha$  atoms of R131 (DRY Motif) and L272 at the end of TM6 are used to determine the distance between TM3 and TM6. TM3-TM7 distance is calculated by using  $C\alpha$  atoms of R131 (DRY Motif) and Y326 (NPxxY Motif). Salt bridge breakage through the activation process is investigated by examining the distances of side chains of the residues (R131-E268) which are formed salt bridge in the inactive state [15, 49, 52]. The TM3-TM6 residue distances are 8.6 Å and 15.4 Å for the initial

and final structures. The TM3-TM7 residue distances are calculated as 16.6 Å and 11.6 Å for initial and final structures. Additionally, the distance between atoms forming the salt bridge is determined as 11.5 Å and 23.0 Å for initial and final structures, respectively. The distance profile of TM3-TM6, TM3-TM7, and between the residues formed salt bridge in the inactive structure of  $\beta 2$  Adrenergic receptor shown in Figure 3.48

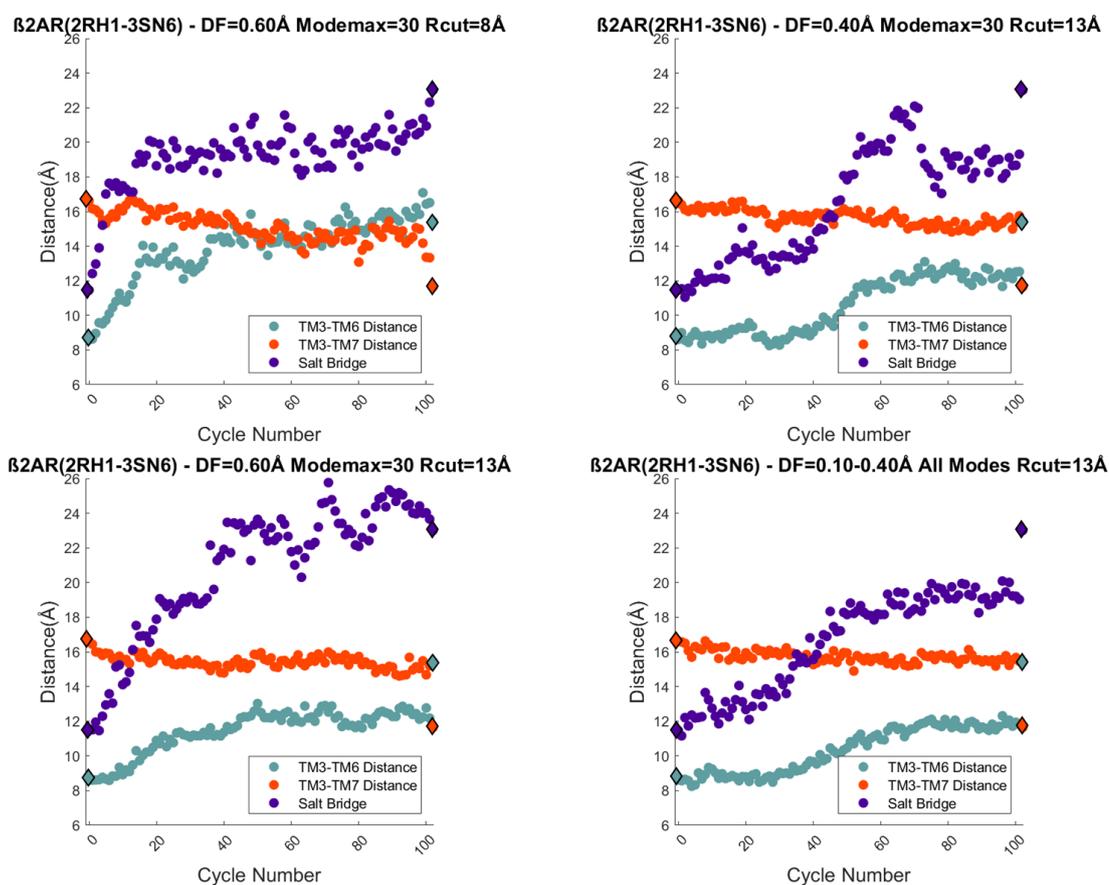


Figure 3.48. Functionally important distances through the transition pathway obtained from ANM-LD results of 2RH1 to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the last five snapshots of the ANM-LD simulation that most accurately reflects the activation process from fifteen parallel simulations are considered, the TM3-TM6 distance is 16.4 Å; the TM3-TM7 distance is 14.2 Å; the distance between the

residues R131 and E268 is increased to 21.1 Å caused by the breakage of the salt bridge. When the distance profile of  $\beta 2$  Adrenergic receptor is analyzed, the simulation with parameters with  $DF=0.60$  Å  $Modemax=30$   $R_{cut}=8$  Å shows the activation pathway better than the other simulations. In the activation process, the first structural change is the breakage of salt bridge between the residues R131 at the end of TM3 and E268 at the end of TM6, which affects the TM3-TM6 distance to increase. However, the distance between TM3 and TM7 do not reach the final structure at the end of the simulation.

- DRY and NPxxY Motifs.

The RMSD of functionally important motifs DRY (D130- R131- Y132) and NPxxY (N322- P323- Y326) [15] with respect to the initial structure is calculated by using  $C\alpha$  atoms through the activation process. The RMSD of DRY and NPxxY motifs of the inactive structure are used as reference value when analyzing the RMSD profile of important motifs and are determined as 1.6 Å and 2.8 Å, respectively. The RMSD of functionally important motifs DRY and NPxxY given in Figure 3.49. When the last five snapshots of two best converged ANM-LD simulations are considered, the RMSD of DRY motif is calculated as 1.7 Å; the RMSD of NPxxY motif is calculated as 2.1 Å. When the RMSD profile of DRY motif and NPxxY motif is analyzed, it is seen that the RMSD of DRY motif converges to the target structure; however, the RMSD of NPxxY motif reaches the target structure in a few cases.

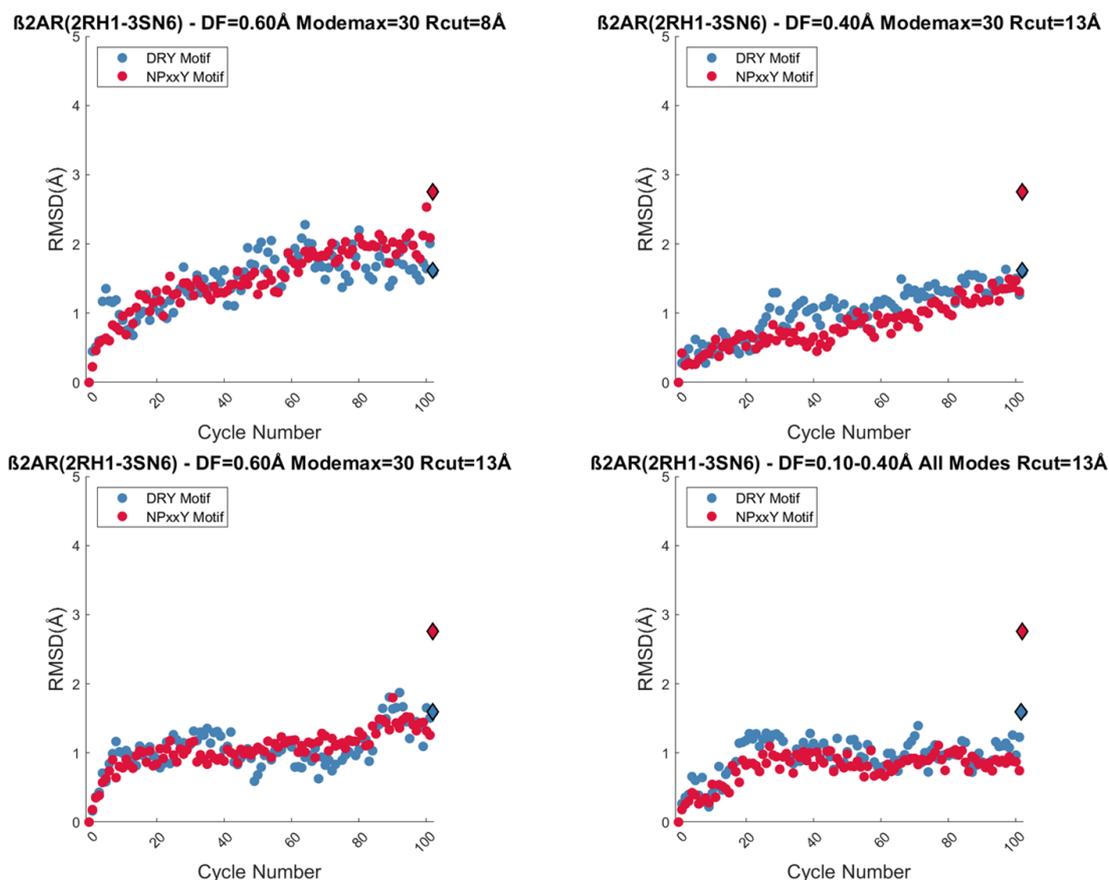


Figure 3.49. The RMSD of important motifs through the activation of M2 receptor from 3UON to 4MQS (initial and final/target structures shown as diamond, snapshots shown as spheres).

- Connector Region, Ligand Binding Site, and G-protein Binding Site.

The RMSD of connector (I121), ligand binding site (S207), and G-protein binding site (Y219) [15] according to initial structure is determined by using  $C\alpha$  atoms through the activation pathway. The RMSD between the initial and active/target structures is 1.4 Å, 1.7 Å, and 2.6 Å for the connector, the ligand binding site, and G-protein binding site, respectively. The RMSD of connector, ligand binding site, and G-protein binding site are shown in Figure 3.50.

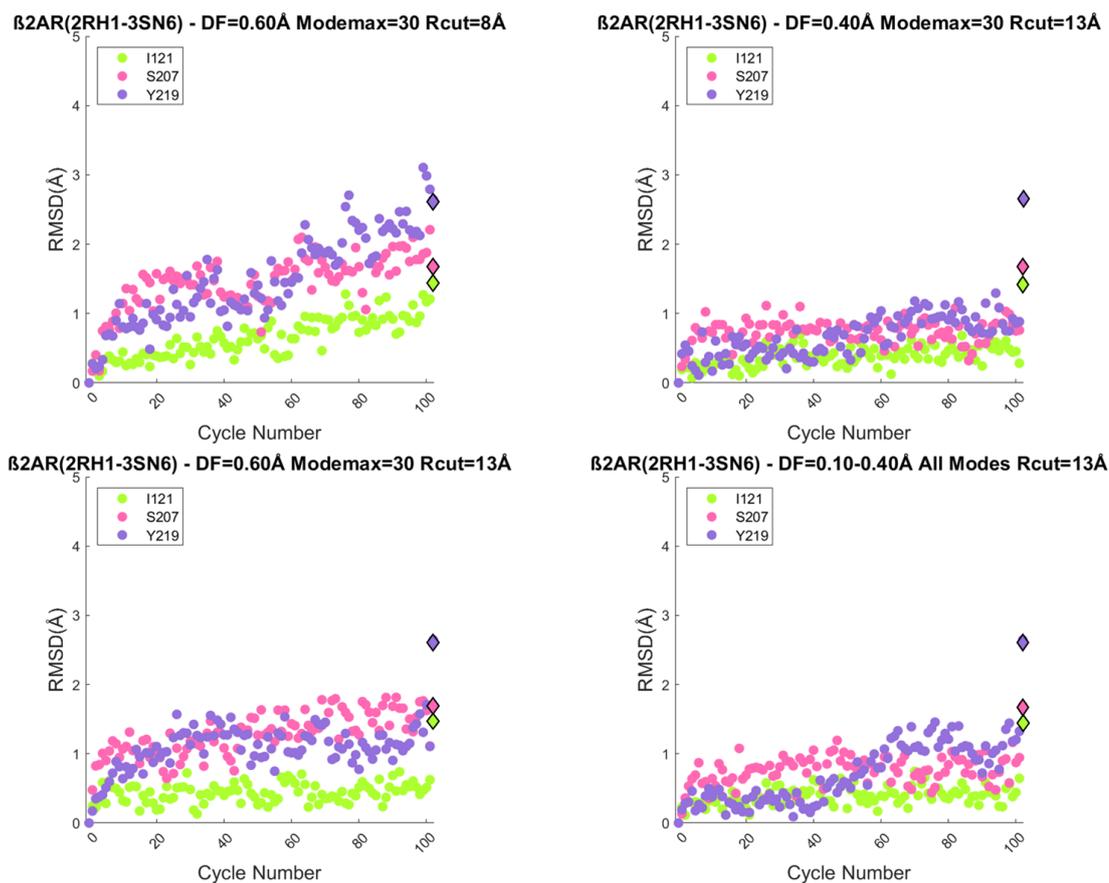


Figure 3.50. The RMSD of important residues through the activation of  $\beta 2$  adrenergic receptor from 2RH1 to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the last five snapshots of two best converged ANM-LD simulations are considered, the RMSD is calculated as 1.1 Å, 1.9 Å, and 2.6 Å for the connector, the ligand binding site, and G-protein binding site, respectively. When the RMSD profile of residues, which played an important role during activation, is examined, the ligand binding site and G-protein binding site show noticeable changes in the early stages of the transition pathway. In contrast, the connector site approaches the target structure with a uniform grade.

- Key Modes of Motion.

The selected ANM modes through the transition pathway from inactive to active structure give a comprehension about the utilization of intrinsic dynamics of the structure. In order to achieve this goal, the most selected ANM modes are resolved to unveil the dynamic determinants underlying the activation process. The analysis of ANM modes are performed according to procedure applied in the Section 3.1.1.1. Selected modes through the transition pathway from 2RH1 to 3SN6 are given in Figure 3.51.

The ANM modes selected in the first fifty cycles of the parallel simulations from 2RH1 to 3SN6 are examined and shown in Figure 3.52. In the parallel ANM-LD simulation ( $DF = 0.60 \text{ \AA}$   $Modemax = 30$   $R_{cut} = 8 \text{ \AA}$ ), where the activation process is best observed, the ANM modes with the highest overlap value are determined as the ANM mode 7 selected four times, the ANM mode 11 selected eight times, and the ANM mode 24 selected seven times.

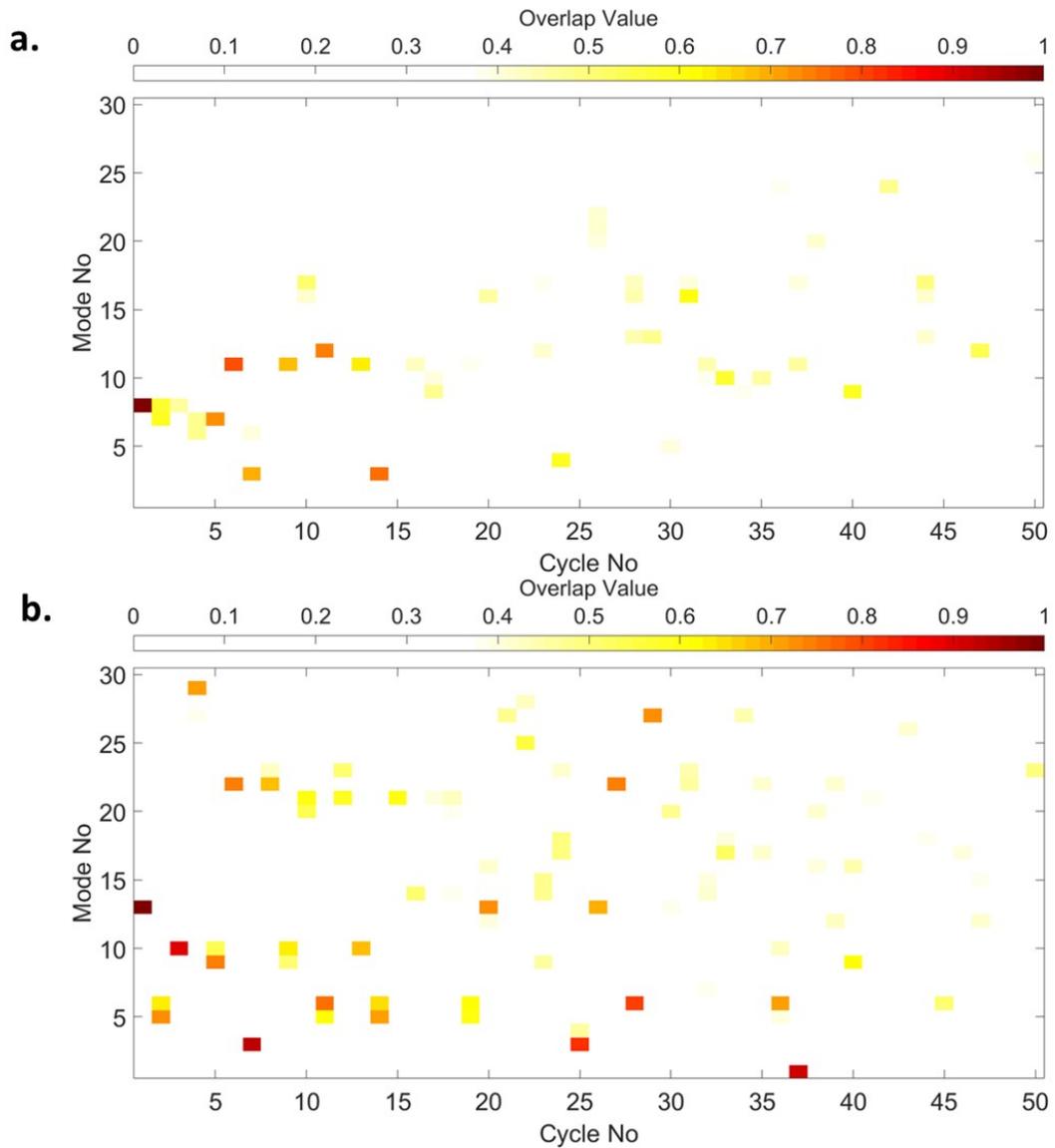


Figure 3.51. Selected Mode Mapping of  $\beta$ 2AR simulations from 2RH1 to 3SN6 a) DF=0.60 Å Modemax=30  $R_{cut}$ =8. b) DF=0.40 Å Modemax=30  $R_{cut}$ =13. c) DF=0.60 Å Modemax=30  $R_{cut}$ =13. d) DF=0.10-0.40 Å All Modes  $R_{cut}$ =13

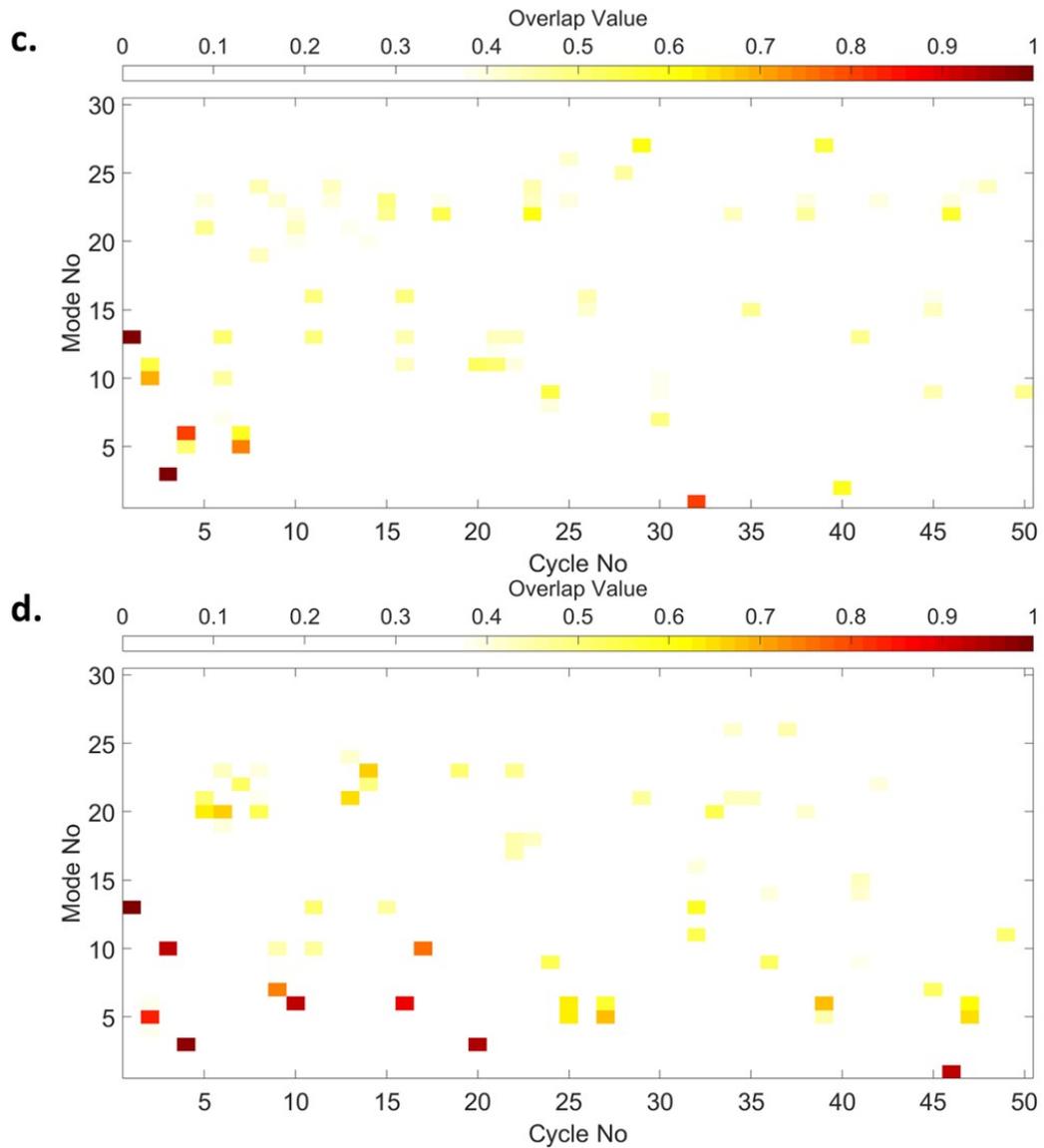


Figure 3.51. Selected Mode Mapping of  $\beta$ 2AR simulations from 2RH1 to 3SN6 a)  $DF=0.60 \text{ \AA}$   $Modemax=30$   $R_{cut}=8$ . b)  $DF=0.40 \text{ \AA}$   $Modemax=30$   $R_{cut}=13$ . c)  $DF=0.60 \text{ \AA}$   $Modemax=30$   $R_{cut}=13$ . d)  $DF=0.10-0.40 \text{ \AA}$  All Modes  $R_{cut}=13$ . (cont.)

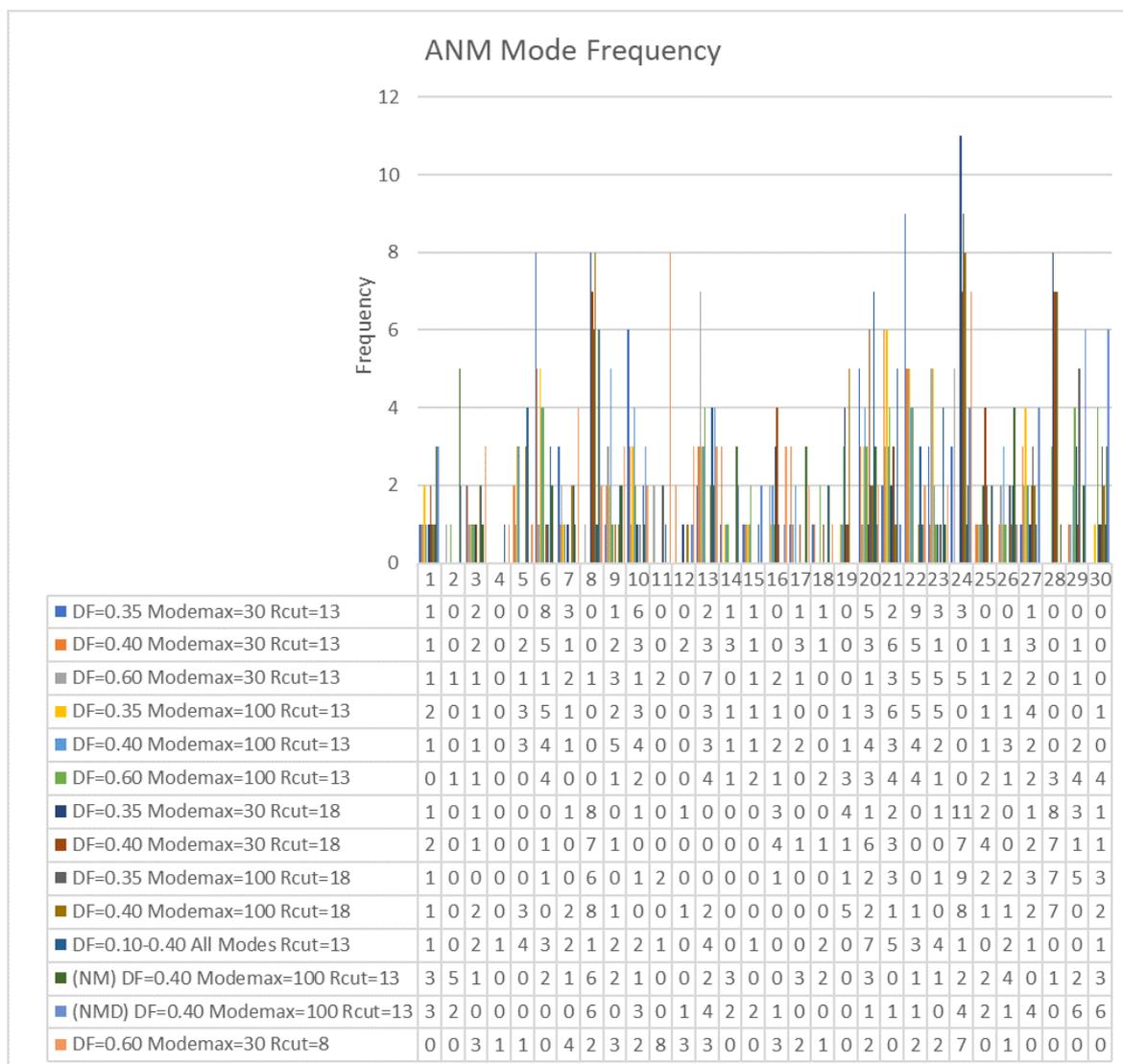


Figure 3.52. ANM mode frequency of parallel simulations from 2RH1 to 3SN6.

MSF of difference vector between the initial and final/target structure is given in Figure 3.53. When this graph is compared with the MSF of selected ANM modes in Figure 3.54, it is observed that MSF of ANM modes 7, 11, and 24 are similar to this difference vector. One reason that the activation process is not fully achieved in other simulations is that ANM mode 11 is not preferred. The main difference of this simulation than others is using  $8 \text{ \AA}$  as  $R_{cut}=13$  value, which indicates that this mode is more effective in short distance between the atoms.

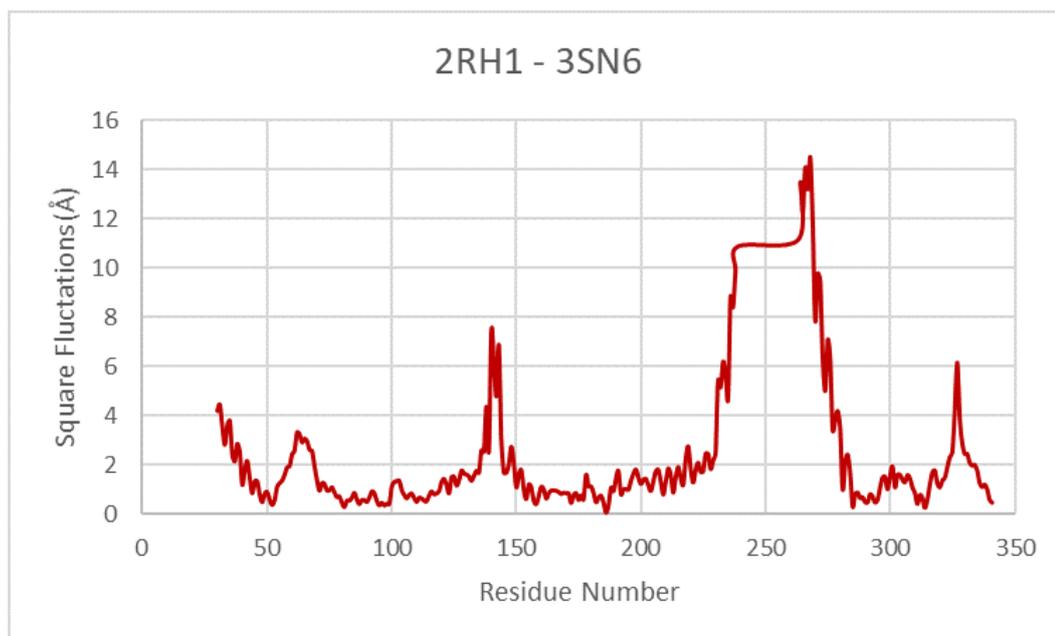


Figure 3.53. Difference vector between initial (2RH1) and final (3SN6) structure.

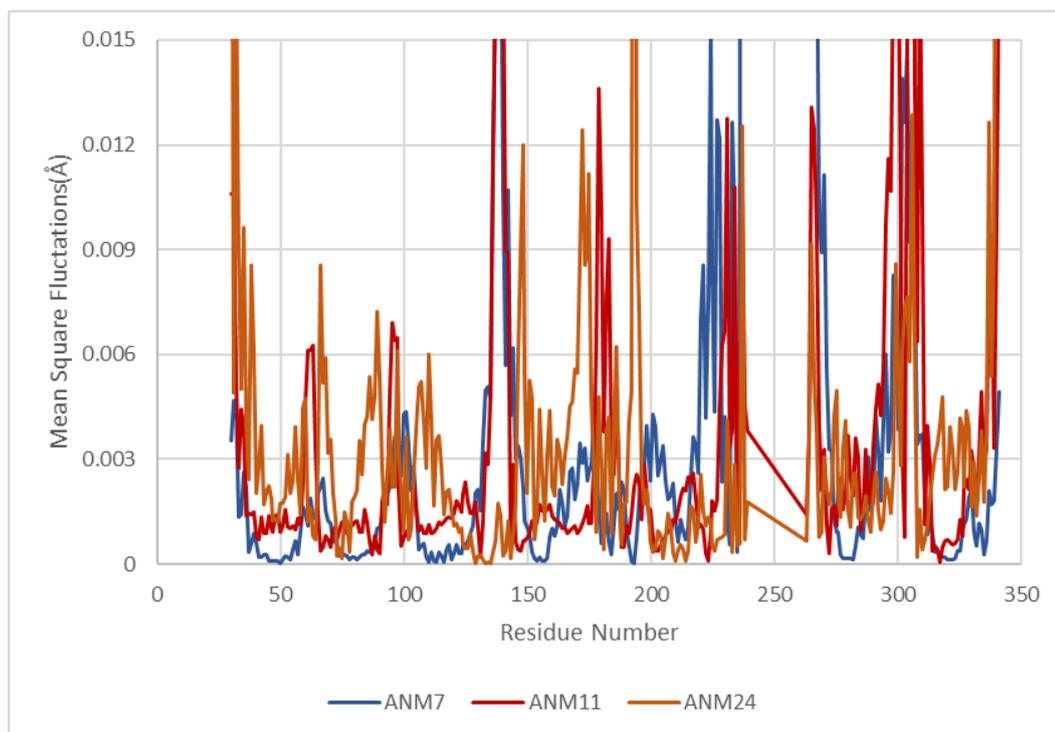


Figure 3.54. Mean square fluctuations of selected ANM modes of  $\beta$ 2AR simulations from 2RH1 to 3SN6.

Minimum points in the mean square fluctuations profile of ANM mode are considered as the hinge residues of this mode, which presumably regulate the motion. The hinge residues of ANM slow mode 11 are identified as S41, T66, V87, G90, W99, R131, K147, T189, S203, F223, K235, L272, C285, I303, and V317. When the hinge residues of ANM mode 11 are investigated, it is seen that functionally important residues (R131 (DRY Motif) and L272) are coincide with the hinge residues of ANM mode 11. Additionally, V87, C285, and S203 which are the proposed as allosteric residues in literature also overlap with the hinge residues [55, 56]. The functionally important regions are matched with the hinge residues indicates that this mode movement plays an important role during activation of the  $\beta 2$  adrenergic receptor. Mean square fluctuation profile of ANM mode 11 is given in Figure 3.55.

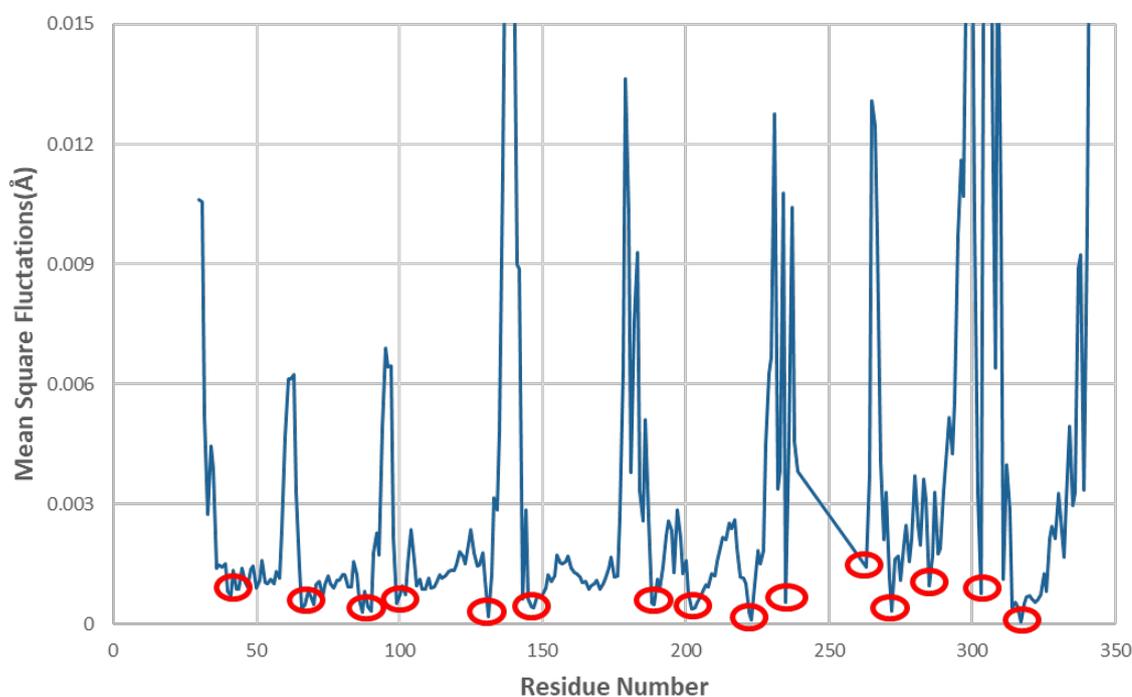


Figure 3.55. ANM mode 11 shape of the initial state of  $\beta 2$  adrenergic receptor.

Hinge residues of ANM slow mode 7 are identified as A46-G50, A78-V81, T110, D113, C116, I154-W158, C184, F193, P211, K232, K235, M279-F282, N318-N322, and A335. F193 located in the extracellular loop (ECL2) determined as the most important

allosteric region is defined as the hinge residue [56]. Additionally, hinge residues of ANM mode 7 overlap with T110 in orthosteric binding site, F282 in connector region, N322 in NPxxY motif, I154, and M279 are matched with the residues in allosteric sites in literature [56]. Mean square fluctuation profile of ANM mode 7 is given in Figure 3.56.

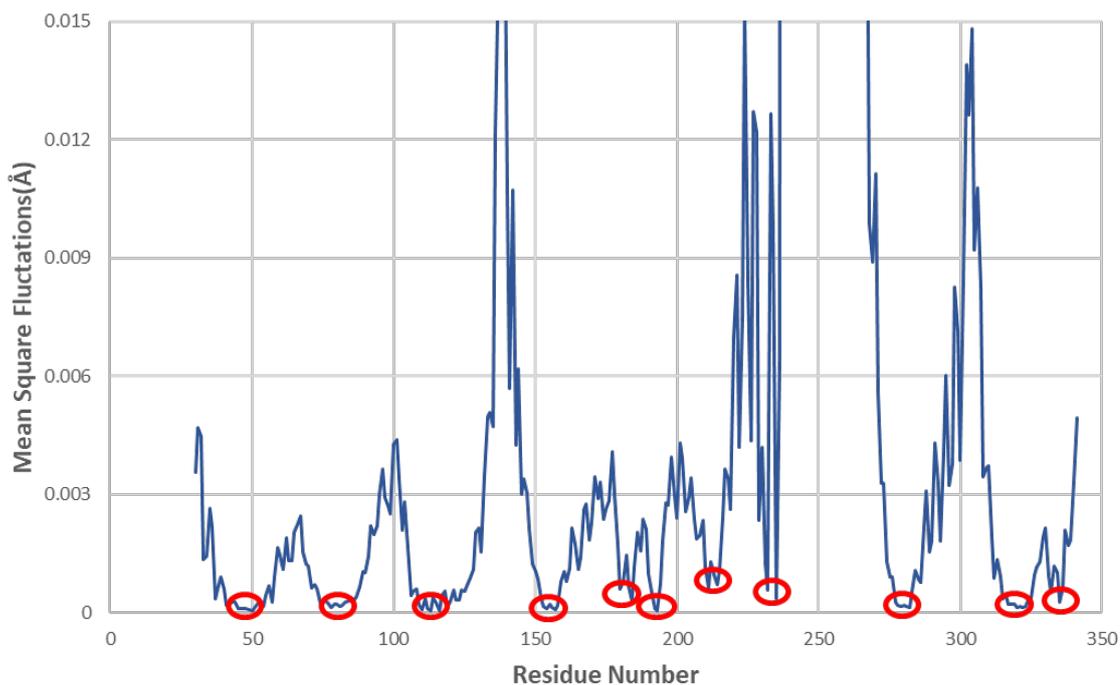


Figure 3.56. ANM mode 7 shape of the initial state of  $\beta 2$  adrenergic receptor.

Hinge residues of ANM mode 24 are identified as I47-F49, L64, L75, A78, A92, T100, G102, V129, F133-I135, S143, I153, H178, A181, C184, E188, I201, I214, E188, I201, I214, A226, I233, E268, Y308, and F332. When the hinge residues of this mode are examined, the most important finding is that H178, which is determined as the residue to which the allosteric modulator can be connected, has been determined as the hinge residue [55]. Additionally, the residues between F133-I135 which plays an important role in the motion of DRY motif are identified as hinge residues. Y308 in TM7 which belongs to the allosteric site in extracellular loop is also determined as hinge residue of this mode [56]. MSF profile of ANM mode 7 is given in Figure 3.57.

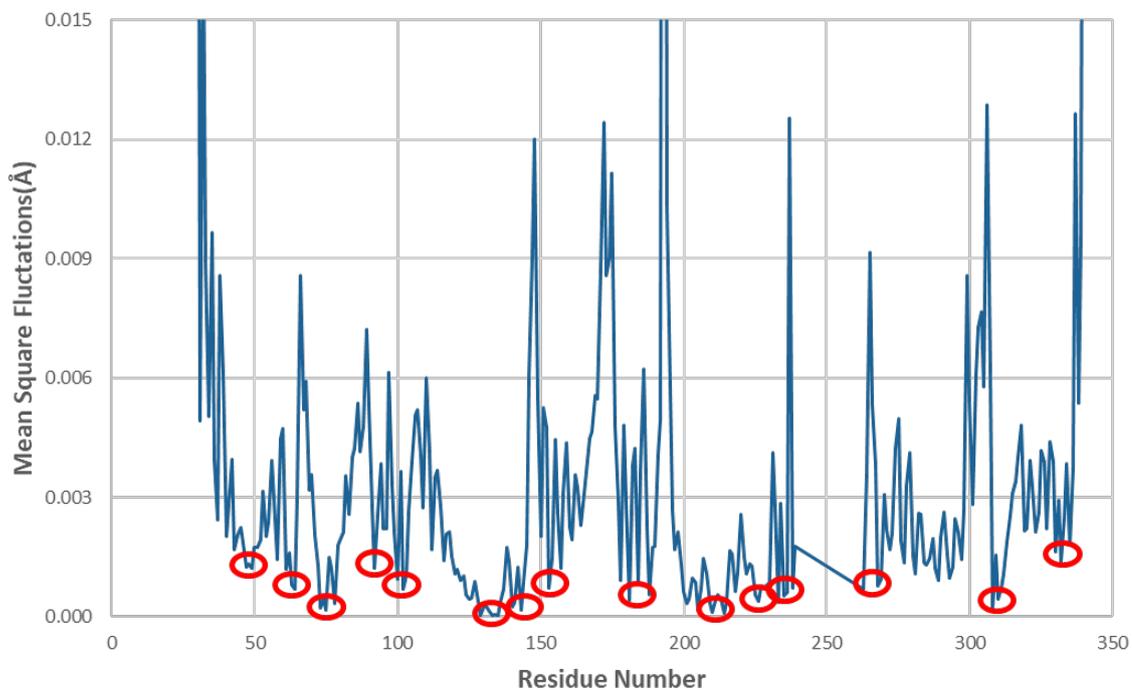


Figure 3.57. ANM mode 24 shape of the initial state of  $\beta 2$  adrenergic receptor.

Hinge residues of ANM mode 7, 11, and 24 are demonstrated on the inactive structure of  $\beta 2$  adrenergic receptor in Figure 3.58.

Functionally important regions (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, L272, and E268) (red and sphere) and the hinge residues of ANM mode 7, 11, and 24 (deep blue) are demonstrated on the initial structure in Figure 3.59.

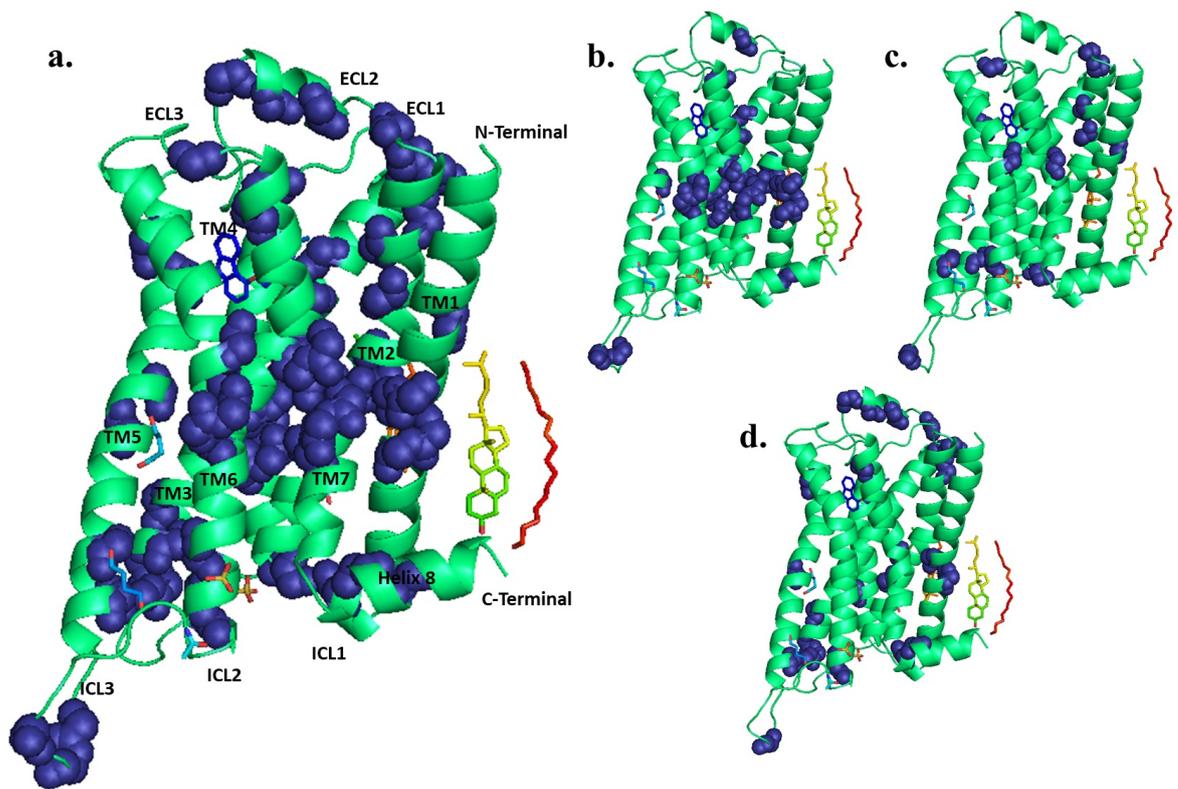


Figure 3.58. Demonstration of hinge residues of ANM mode 7, 11, and 24 (dark blue and sphere) on the structure a) Hinges of ANM mode 7, 11, and 24. b) Hinges of ANM mode 7 c) Hinges of ANM mode 11. d) Hinges of ANM mode 24.

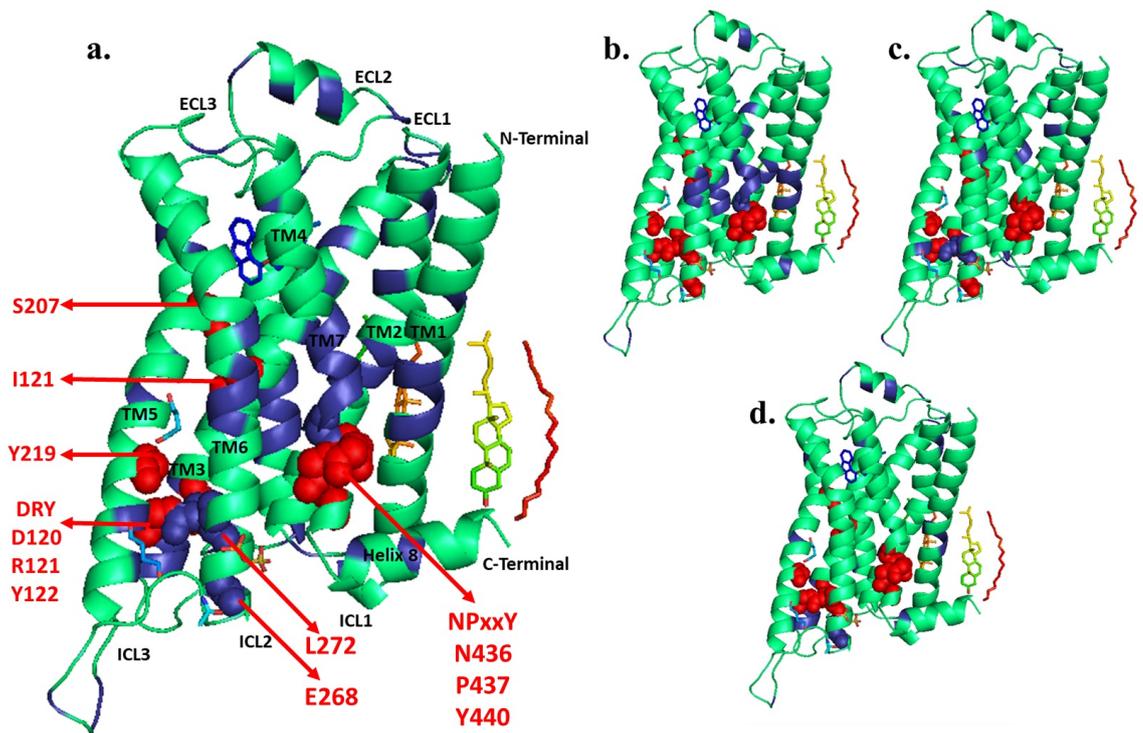


Figure 3.59. Demonstration of ANM mode 7, 11, and 24 hinges (dark blue) and important regions (red and sphere), and the intersection (dark blue and sphere).  
 a) ANM modes 7, 11, and 24 b) ANM mode 7 c) ANM mode 11 d) ANM mode 24.

- Cross-correlation Analysis.

Normalized cross-correlation analysis is used to identify the correlated residues. The analysis of cross-correlation is performed according to procedure applied in the Section 3.1.1.1. Normalized cross-correlation analysis is done through the activation pathway of  $\beta_2$  adrenergic receptor and demonstrated in Figure 3.60.

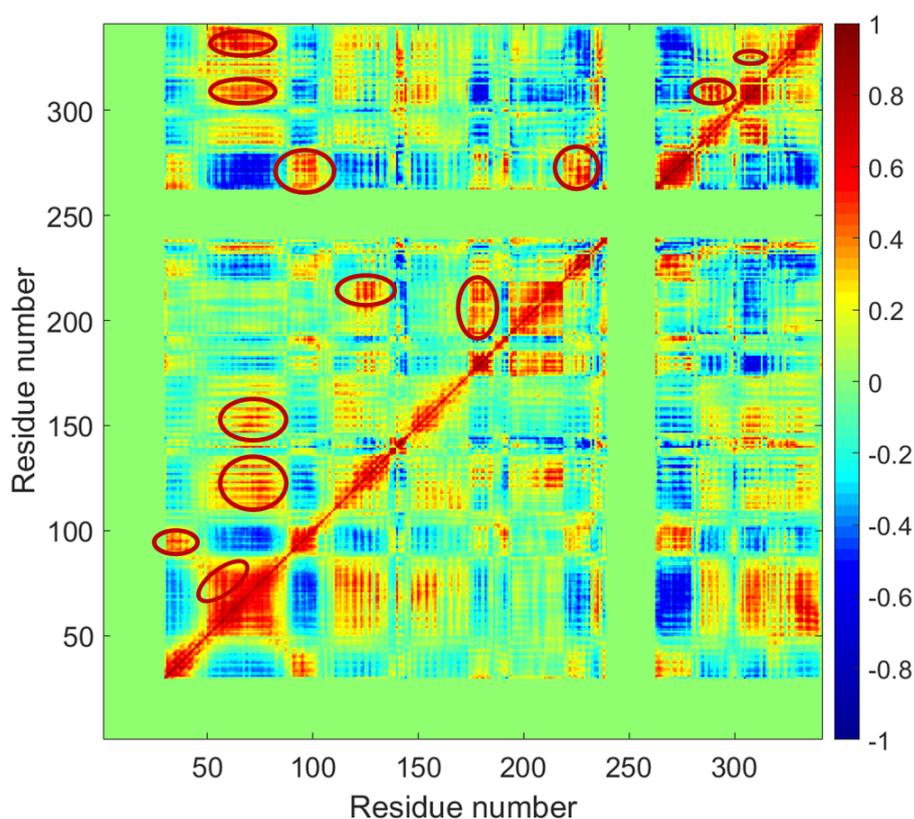


Figure 3.60. Cross correlation map of the activation pathway of  $\beta_2$  adrenergic receptor from 2RH1 to 3SN6.

The highly correlated residues are determined as A63-L66 in ICL1 & D331-F332 in C-terminal, and P288-F289 in TM6 & L311 in TM7. Moreover, the NPxxY motif and Y308 are correlated residues, which are given as allosteric sites in the literature [56].

3.2.1.2. ANM-LD Simulations from 2RH1 to 3P0G. In this section, ANM-LD simulations are executed to investigate the activation mechanism of  $\beta$ 2-Adrenergic receptor from the inactive structure (2RH1) to a different active structure (3P0G). The active structure used in this case (3P0G) is aligned to the active structure used in first case (3SN6) and shown in Figure 3.61. The RMSD of the active structure (3P0G) with respect to the active structure used in the first case (3SN6) is calculated as 1.55 Å.

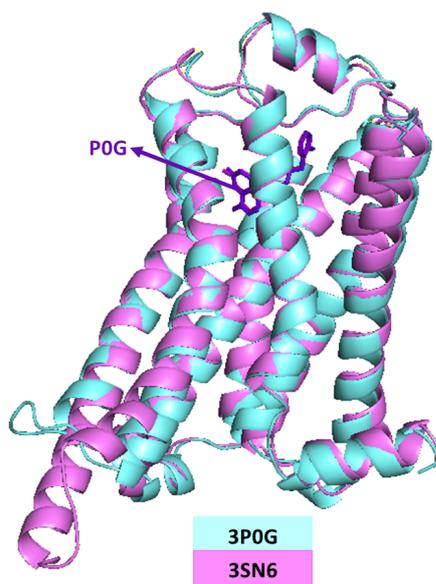


Figure 3.61. Alignment of active structure (3P0G) used in this case with the active structure (3SN6) used in the first case (RMSD: 1.55 Å).

The comparison of the two active structures is given in Table 3.7 in Section 3.2.1. Although the organism of these two active structures is homo sapiens, there are some differences. Crystal structure of the  $\beta$ 2 adrenergic receptor-Gs protein complex (3SN6) has two additional mutations (T96M & T98M) in ECL1 to increase the expression level of T4 lysozyme which is substituted to amino terminus. Additionally, the largest distinctness between two active structure is the outward movement of 3 Å at the end

of TM6 in the  $\beta$ 2AR-Gs. There is also a difference in DRY motif, because R131 in  $\beta$ 2AR-Nb80 reacts with the nanobody [49]. Although there are some differences in two active structures, they have same affinity for the agonist isoproterenol [52].

In the preparation step of the simulations, R228-L230 and K263-C265 from the inactive structure; D23-R28 and R343-R344 from the active structure are deleted for the ANM-LD simulations with unmodified structures. Since there are missing residues at the end of TM5 and in ICL3, R228-L230 and K263-C265 of the final/target structure are modeled by MOD-LOOP server to obtain more accurate results [53]. Lastly, mutations (E187N) of the initial and final/target structure are withdrawn by PyMOL [54] and used in the simulations of back-mutated structures.

Parallel ANM-LD simulations are prepared to examine the activation process of  $\beta$ 2-Adrenergic receptor and the result summaries are given in Table 3.9.

Table 3.9. ANM-LD Result Summaries.

Simulation Name	Cycle Number	Initial RMSD ( $\text{\AA}$ ) to Target Structure	Final RMSD ( $\text{\AA}$ ) to Target Structure	The Most Selected ANM Modes
$D_F = 0.60\text{\AA}$ Modemax = 30 $R_{cut} = 13\text{\AA}$	56	2.43	1.72	2, 5, 7
$D_F = 0.10 - 0.40\text{\AA}$ All Modes $R_{cut} = 13\text{\AA}$	61	2.43	1.98	2, 19
$D_F = 0.60\text{\AA}$ Modemax = 30 $R_{cut} = 8\text{\AA}$	45	2.43	1.55	2, 5, 7, 9
$D_F = 0.10 - 0.40\text{\AA}$ All Modes $R_{cut} = 8\text{\AA}$	42	2.43	1.52	1, 7, 9, 14

The RMSD difference through the activation pathway obtained from ANM-LD simulations is examined and shown in Figure 3.62. The RMSD of the initial and final/target structures is calculated as 2.43 Å for modeled structures and 2.05 Å for unmodified structures. The RMSD of generated structures from ANM-LD simulations are converged around 50 cycles. The lowest RMSD achieved is 1.58 Å, while it remains at the most as high as 2.36 Å. The RMSD results of four ANM-LD runs, which reflects the activation process in the best way according to positions of functionally important residues are given in Figure 3.63.

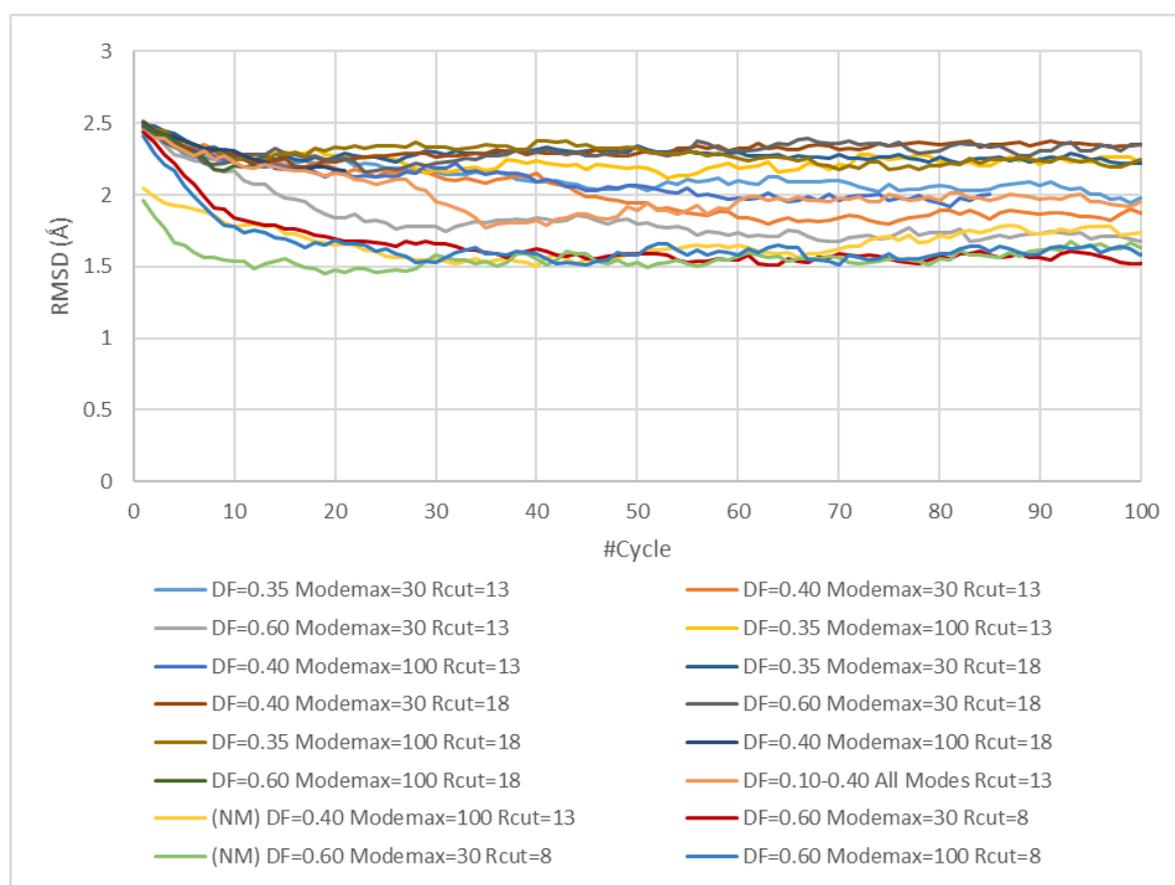


Figure 3.62. RMSD difference through the simulations of  $\beta 2$  adrenergic receptor from 2RH1 to 3P0G.

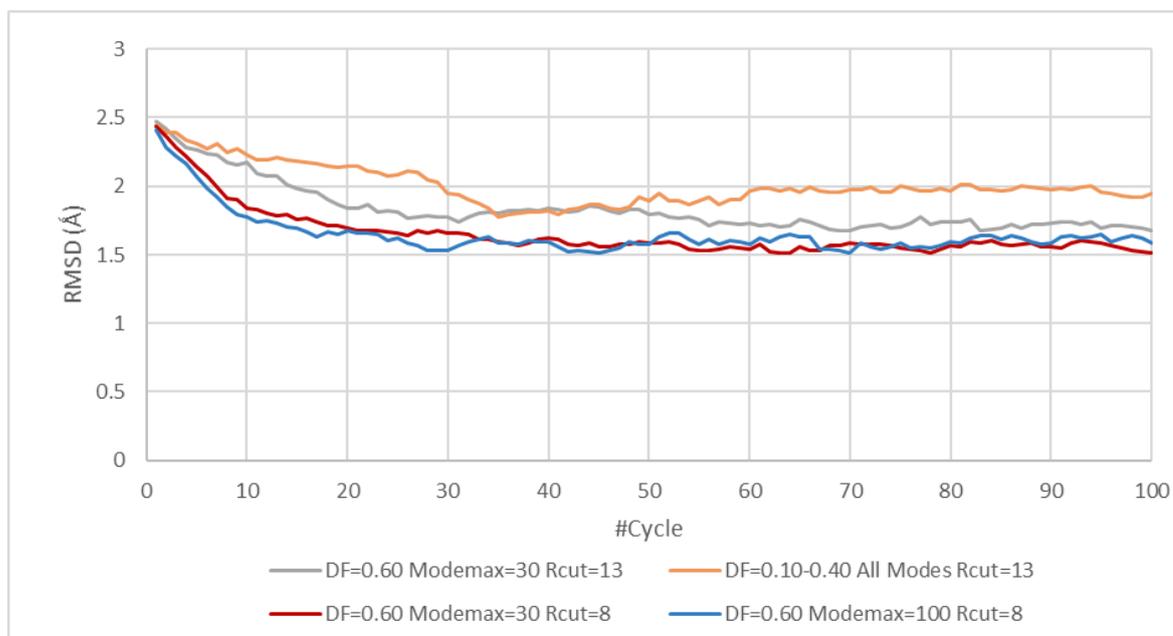


Figure 3.63. RMSD difference of selected runs through the simulations of  $\beta 2$  adrenergic receptor from 2RH1 to 3P0G.

- TM3-TM6 and TM3-TM7 Distances and Salt Bridge Breakage.

Firstly, the TM3-TM6 and TM3-TM7 distances are calculated to examine the movement of residues that are important for the activation of the  $\beta 2$  adrenergic receptor. In the calculation process, C $\alpha$  atoms of R131-L272 are used to determine the distance between TM3 and TM6. The distance between TM3-TM6 is calculated as 8.3 Å and 14.1 Å for the initial and final/target structures, while the TM3-TM6 distance is determined as 15.4 Å for the final/target structure in the simulations from the inactive (2RH1) to active (3SN6) structures, because there is a difference of 3 Å outward movement at the end of TM6 helix [49]. The TM3-TM7 residue distances are calculated as 16.6 Å and 11.3 Å for initial and final structures. Salt bridge breakage through the activation process is investigated by examining the distances of side chains of the residues (R131-E268) which are formed salt bridge in the inactive state [15]. The distance between atoms forming the salt bridge is determined as 11.9 Å and 20.4 Å for initial and final structures, respectively. The TM3-TM6 and the TM3-TM7 distances,

and salt bridge breakage of R131 and E268 are investigated through the transition pathway in Figure 3.64.

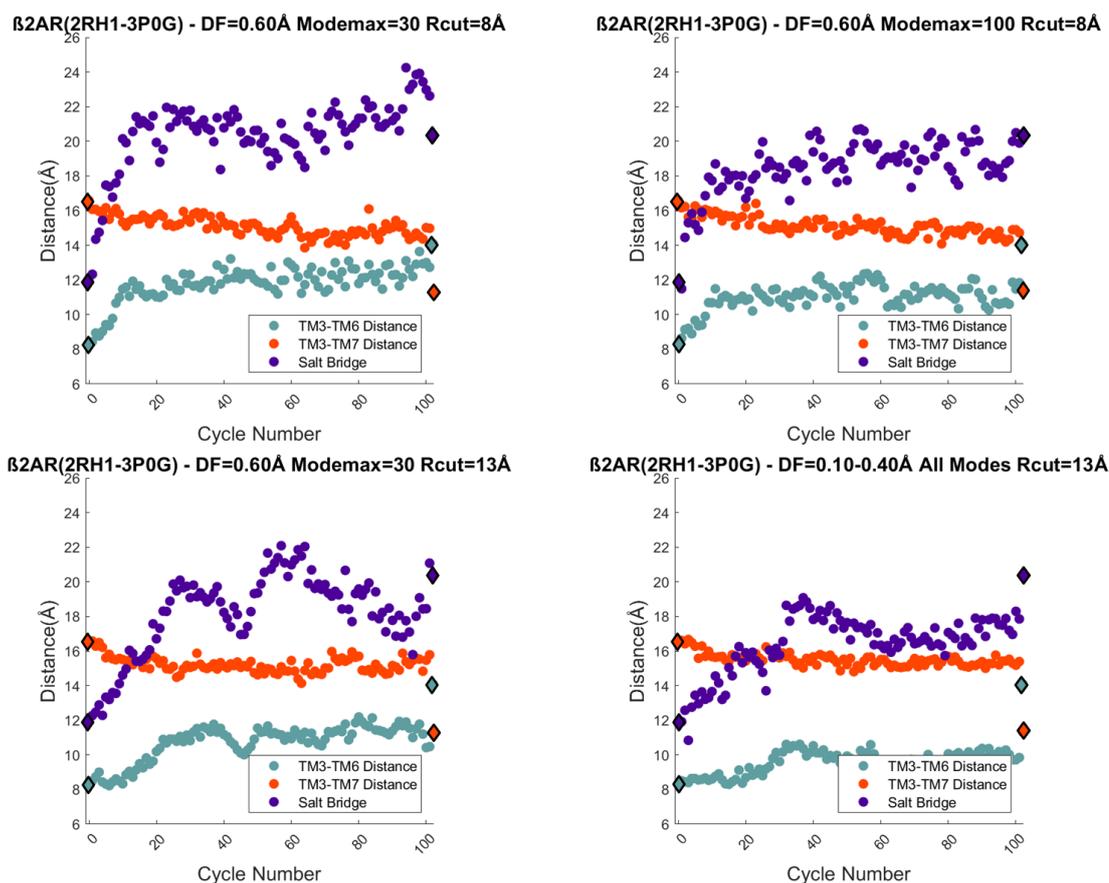


Figure 3.64. Functionally important distances through the transition pathway obtained from ANM-LD results of 2RH1 to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the last five snapshots of the ANM-LD simulation that most precisely mirrors the activation procedure from fifteen parallel simulations are considered, the TM3-TM6 distance is 13.0 Å; the TM3-TM7 distance is 14.7 Å; the distance between the residues R131 and E268 is increased to 23.3 Å caused by the breakage of the salt bridge. As with the distance profiles in the first case, it is observed that the first major change at the start of the activation is the breaking of the salt bridge. Additionally, the TM3-TM6 distance and salt bridge show similar profiles because the salt bridge

breakage plays an important role in the separation of TM3 and TM6. In some of the ANM-LD simulations, while the TM3-TM6 and TM3-TM7 do not converge the distances to the final/target structure, it is observed that the distance between the residues forming the salt bridge converge.

- DRY and NPxxY Motifs.

Although the distinctness between the two active structures (3P0G and 3SN6) examined here is very small for highly conserved motifs DRY (D130- R131- Y132) and NPxxY (N322- P323- Y326), R131 of DRY motif shows differences between two structures, because R131 interacts with Nb80 in (3P0G) and Y391 in  $\beta$ 2AR-Gs (3SN6) [52]. The RMSDs of these two important motifs (DRY and NPxxY) are calculated by using  $C\alpha$  atoms through the activation process. The RMSD of DRY and NPxxY motifs of the inactive structure are used as reference value when analyzing the RMSD profile of important motifs and are determined as 2.2 Å and 2.3 Å, respectively. The RMSDs of these two important motifs (DRY and NPxxY) are shown in Figure 3.65. When the last five snapshots of two best converged ANM-LD simulations are considered, the RMSD of DRY motif is calculated as 1.3Å; the RMSD of NPxxY motif is calculated as 1.2 Å.

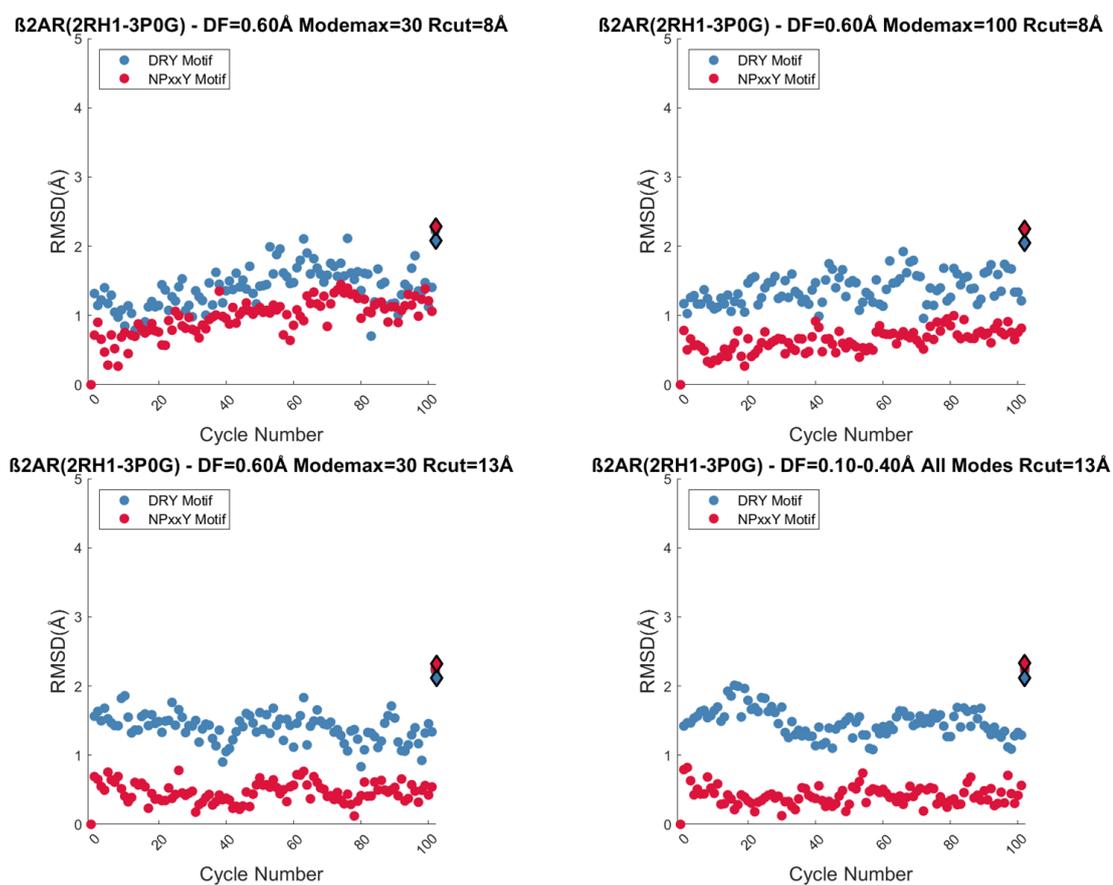


Figure 3.65. RMSD of important motifs through the activation of  $\beta 2$  adrenergic receptor from 2RH1 to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres).

- Connector Region, Ligand Binding Site, and G-protein Binding Site.

The RMSD of connector (I121), ligand binding site (S207), and G-protein binding site (Y219) [15] according to initial structure is determined by using  $C\alpha$  atoms through the activation pathway. The RMSD between the initial and active/target structures is 1.4 Å, 1.7 Å, and 2.6 Å for the connector, the ligand binding site, and G-protein binding site, respectively. The RMSD of connector (I121), ligand binding site (S207), and G-protein binding site given in Figure 3.66. When the last five snapshots of the best converged ANM-LD simulation are evaluated, the RMSD is calculated as 0.8 Å, 1.8 Å, and 2.7 Å for the connector, the ligand binding site, and G-protein binding site, respectively.

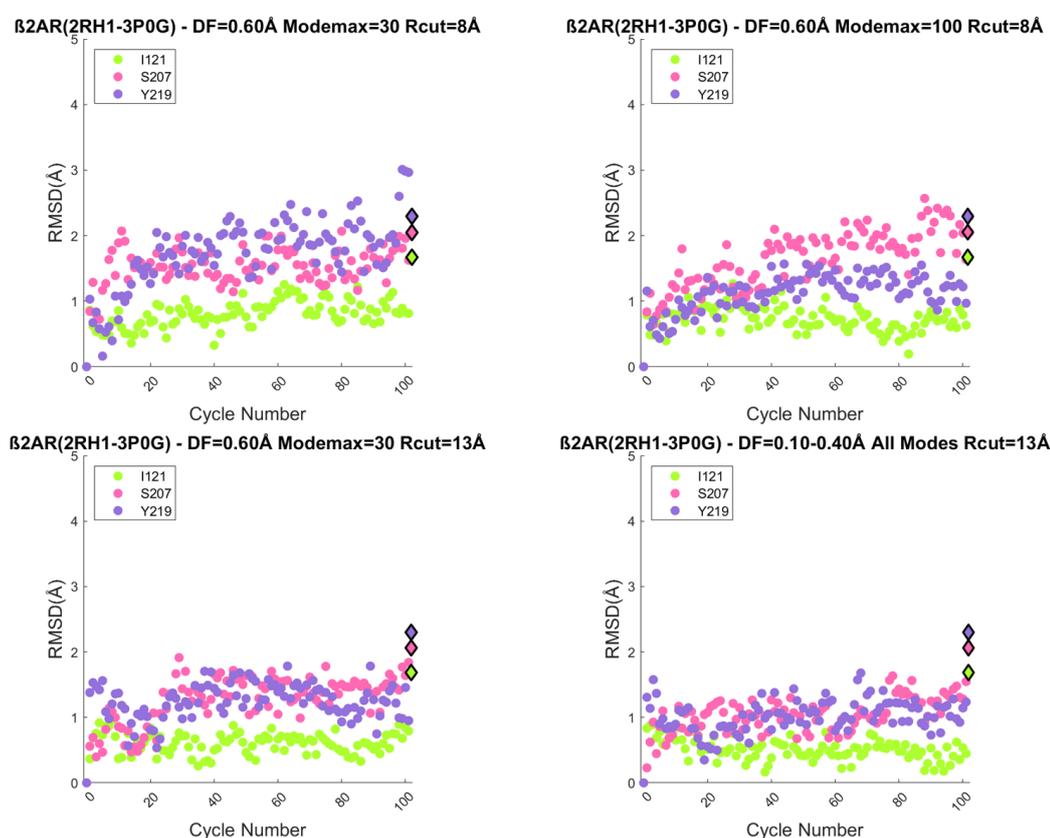


Figure 3.66. RMSD of important residues through the activation of  $\beta_2$  adrenergic receptor from 2RH1 to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres).

- Key Modes of Motion.

The most selected ANM modes are explored to understand the mechanism of  $\beta 2$ -Adrenergic receptor activation, since they have the dynamic determining role. Chosen modes through the progression of the transition pathway from 2RH1 to 3P0G are given in Figure 3.67.

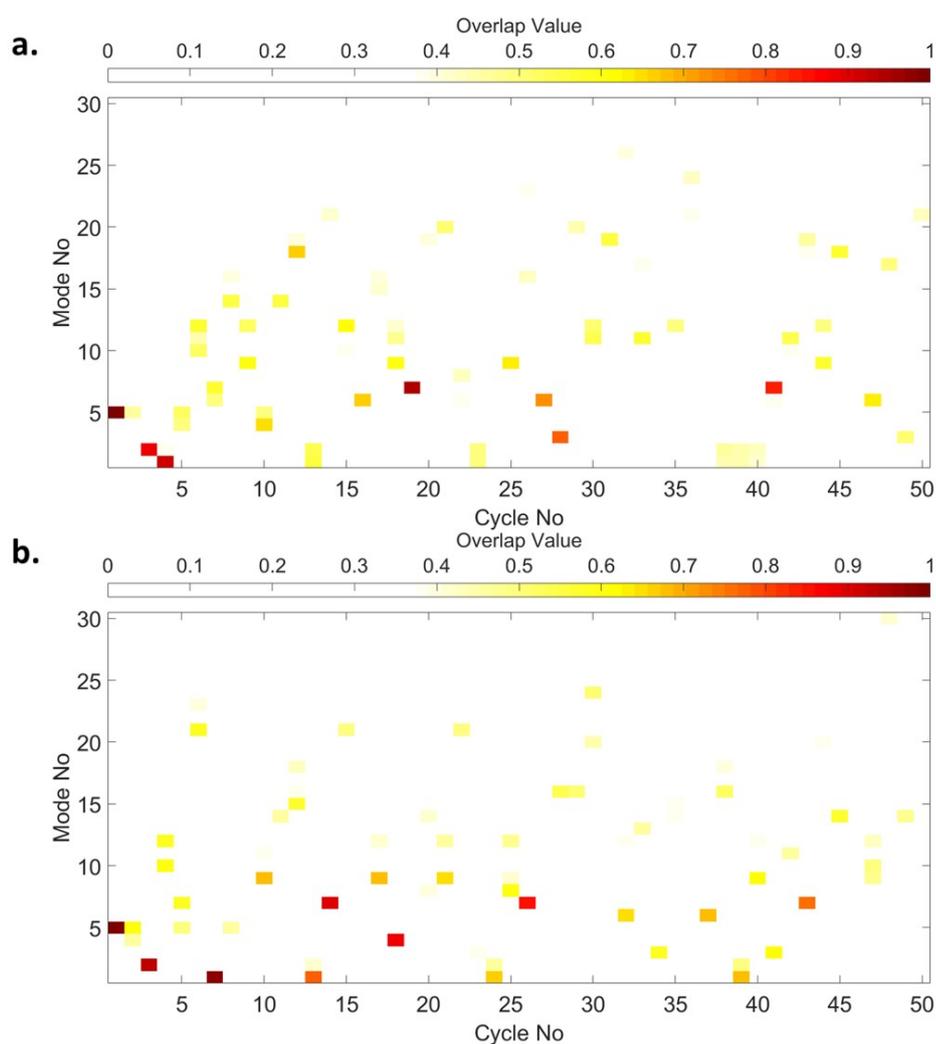


Figure 3.67. Selected Mode Mapping of  $\beta 2$ AR simulations a)DF=0.60 Å Modemax=30  $R_{cut}$ =8 Å. b)DF=0.60 Å Modemax=100  $R_{cut}$ =8 Å. c)DF=0.60 Å Modemax=30  $R_{cut}$ =13 Å. d)DF=0.10-0.40 Å All Modes  $R_{cut}$ =13 Å.

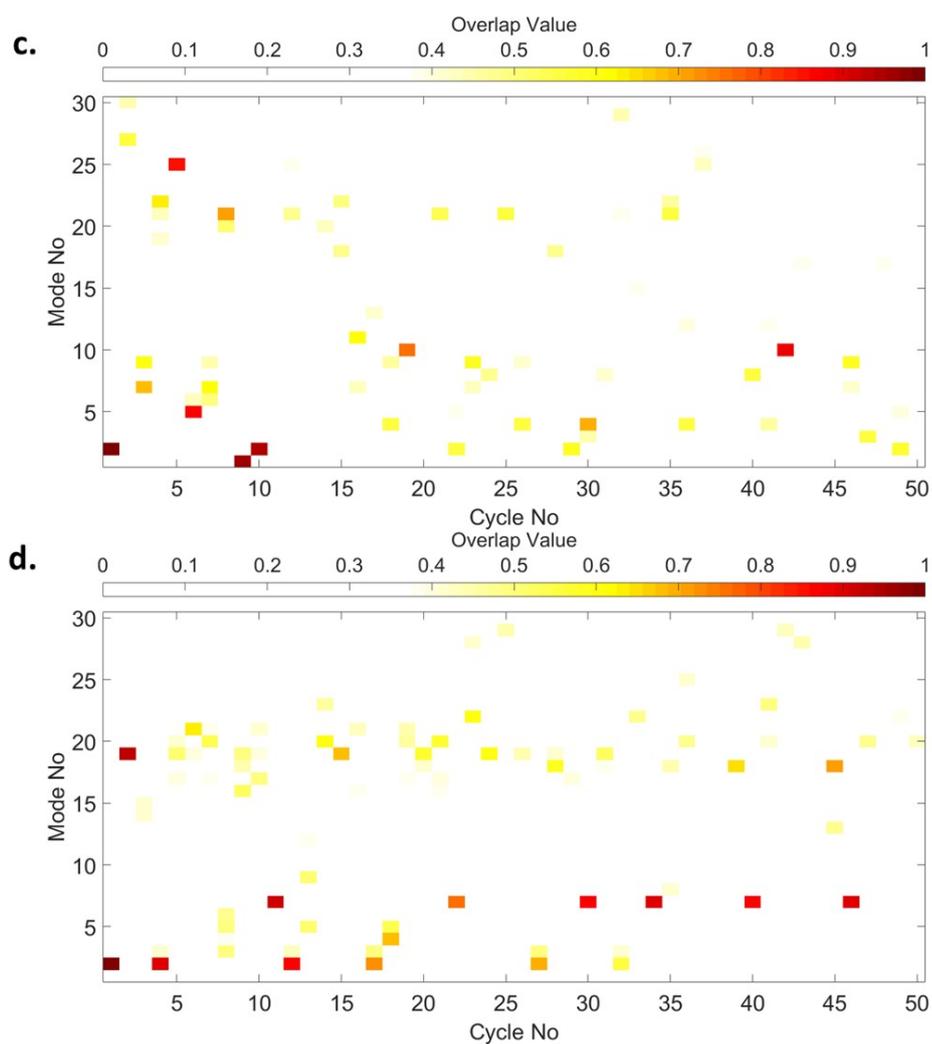


Figure 3.67. Selected Mode Mapping of  $\beta$ 2AR simulations a)  $DF=0.60 \text{ \AA}$   $Modemax=30$   $R_{cut}=8 \text{ \AA}$ . b)  $DF=0.60 \text{ \AA}$   $Modemax=100$   $R_{cut}=8 \text{ \AA}$ . c)  $DF=0.60 \text{ \AA}$   $Modemax=30$   $R_{cut}=13 \text{ \AA}$ . d)  $DF=0.10-0.40 \text{ \AA}$  All Modes  $R_{cut}=13 \text{ \AA}$ . (cont.)

The selected ANM modes in the first fifty cycles from 2RH1 to 3P0G are investigated and appeared in Figure 3.68. The ANM-LD simulation with  $DF=0.60 \text{ \AA}$ ,  $Modemax=30$ ,  $R_{cut}=8 \text{ \AA}$  where the activation pathway is best observed, the most selected mode is determined as ANM mode 9.

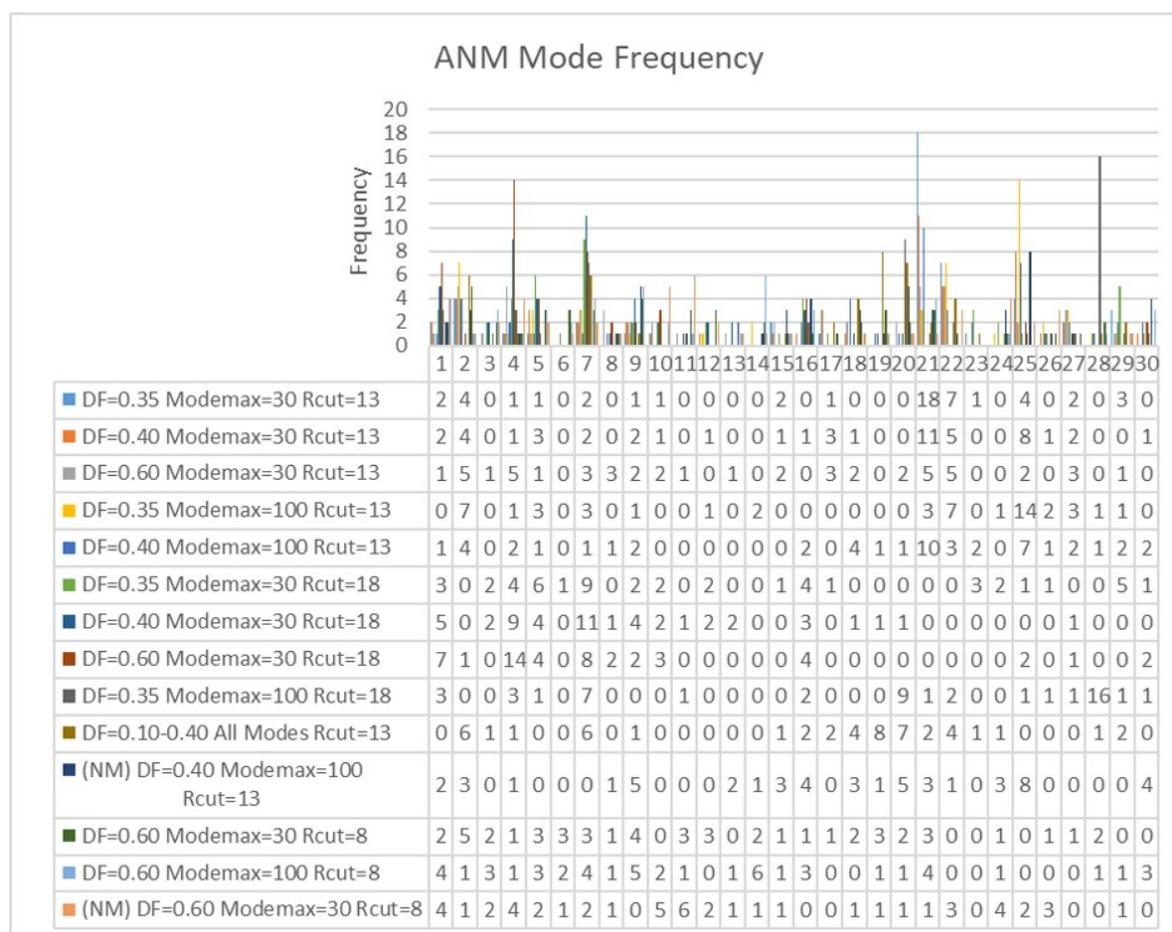


Figure 3.68. ANM mode frequency of parallel simulations from 2RH1 to 3P0G.

The difference vector calculated between the initial (2RH1) and final/target structures (3P0G-3SN6) given in Figure 3.69 to examine the distinctness between this case and the simulations performed from inactive structure (2RH1) to a different active structure (3SN6). It is observed that, the minimum and maximum points of mean square fluctuation profile overlap with each other in two cases except the residues between V52-A78 in intracellular side of TM1 and ICL1 and T177-E187 in the

ECL2. There are nine more missing residues between Q231-R239 in the active structure used in this case (3P0G) than in the active structure (3SN6) used in the first case in this section.

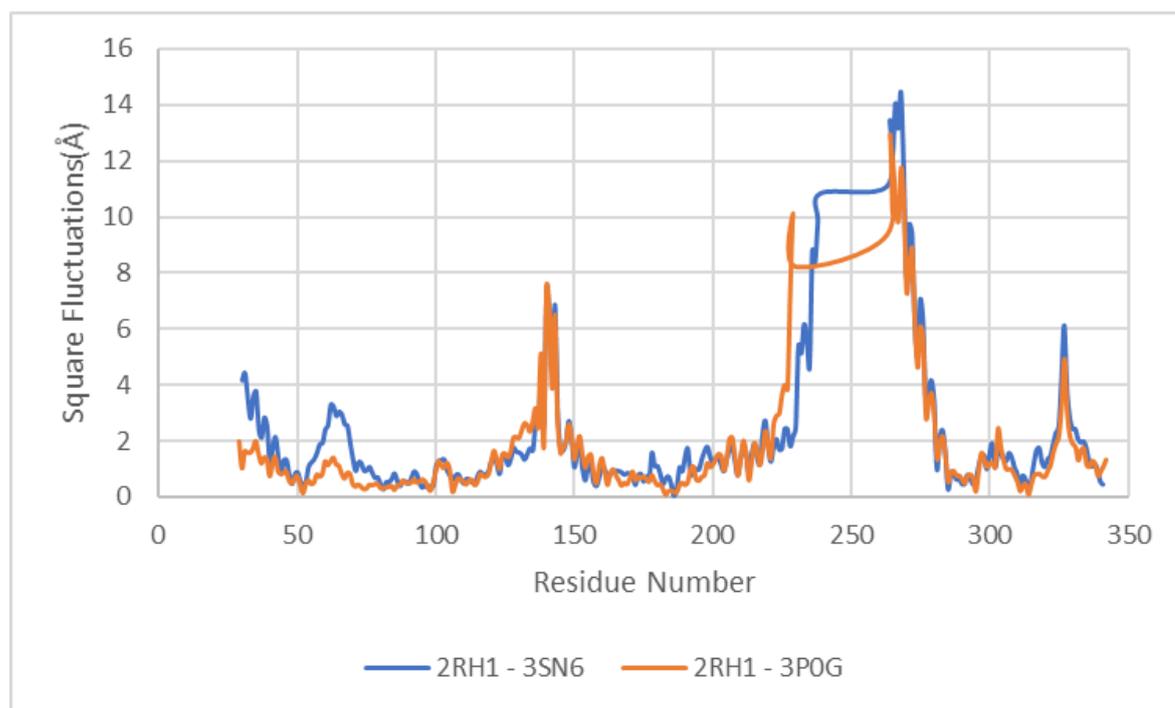


Figure 3.69. The difference vector between initial (2RH1) and final (3SN6 – 3P0G) structures.

Mean square fluctuation profile of ANM mode 9 mostly overlap with the difference vector. Hinge residues of ANM mode 9 is determined from the mean square fluctuations profile and shown in Figure 3.70. V33, G37, M40, I47, L64, T68, A92-I94, F101, D130-A134, L144-L145, I154, C190, S204-S207, V218-V222, A271-K273, F289-I291, Y308-L310, N322-Y326, and L339 are determined as hinge residues. When hinge residues of ANM mode 9 are investigated, it is understood that functionally important residues DRY motif, NPxxY motif, orthosteric ligand binding site (S207), and G-protein binding site (Y219) overlap with the hinge residues. Additionally, Y308 in TM7 and I154 in TM4 which are the proposed as allosteric residues in literature [55, 56] also overlap with the hinge residues coincide with the hinge residues of ANM 9. When hinge

residues determined in this case are compared with the hinge residues obtained from ANM-LD simulations from inactive structure (2RH1) to active structure (3SN6), the residues in the proposed allosteric sites in literature such as intracellular loop (ICL3) and connector region unable to identify [56]. The major reason for this situation is the missing residues in the ICL3 loop of the structure. However, orthosteric ligand binding site (S207), and G-protein binding site (Y219) are the hinge residues of ANM mode 9, while these residues are not identified as hinge residues obtained from ANM-LD simulations from inactive structure (2RH1) to active structure (3SN6).

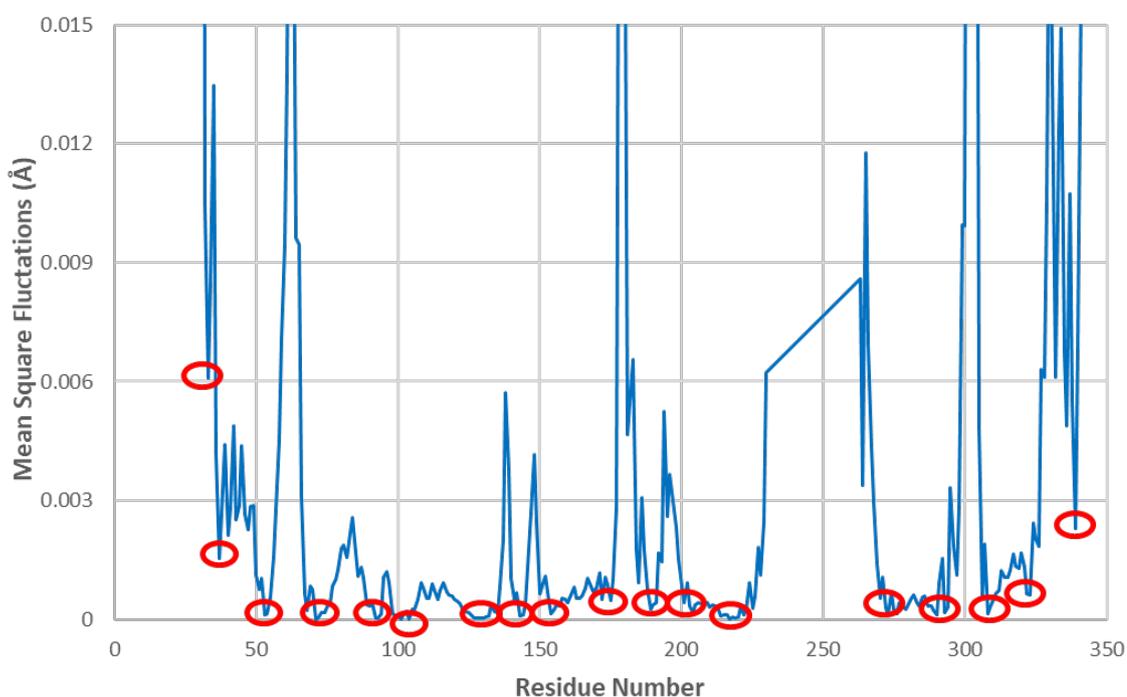


Figure 3.70. ANM mode 9 shape of the initial state of  $\beta 2$  adrenergic receptor.

Hinge residues of ANM mode 9 are displayed on the inactive state of  $\beta 2$  adrenergic receptor in Figure 3.71.

Functionally important regions (red and circle) and the hinge residues of ANM mode 9 (deep blue) are demonstrated on the initial structure in Figure 3.72.

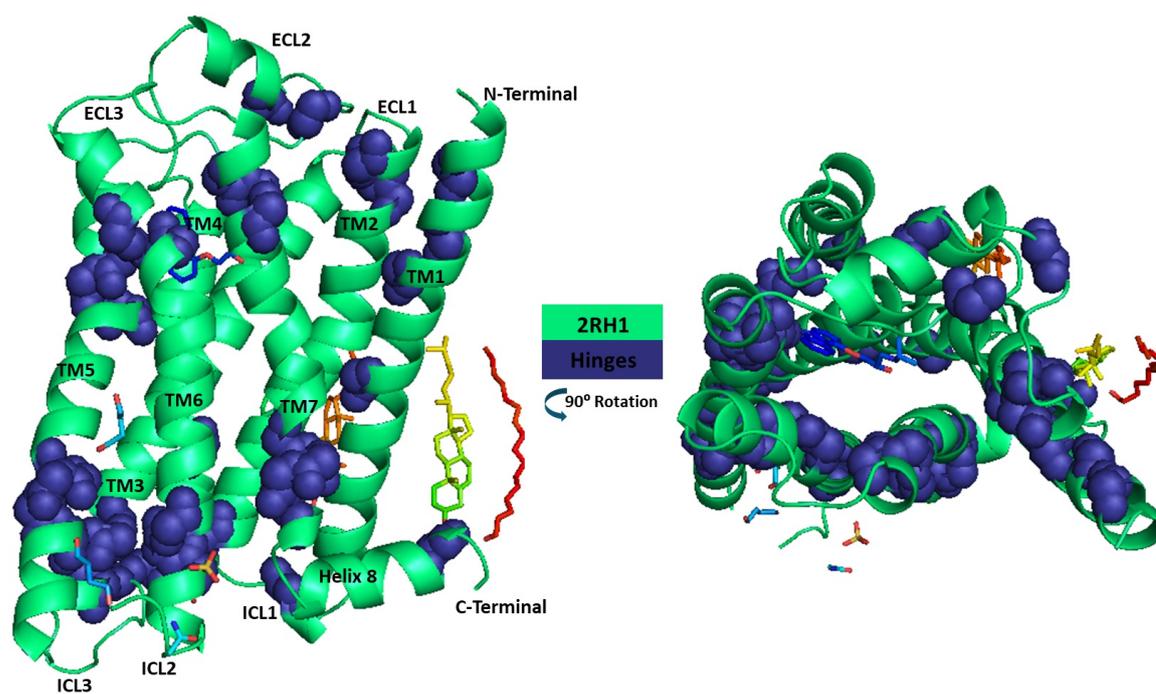


Figure 3.71. Demonstration of hinge residues of ANM mode 9 (dark blue and sphere) on the structure.

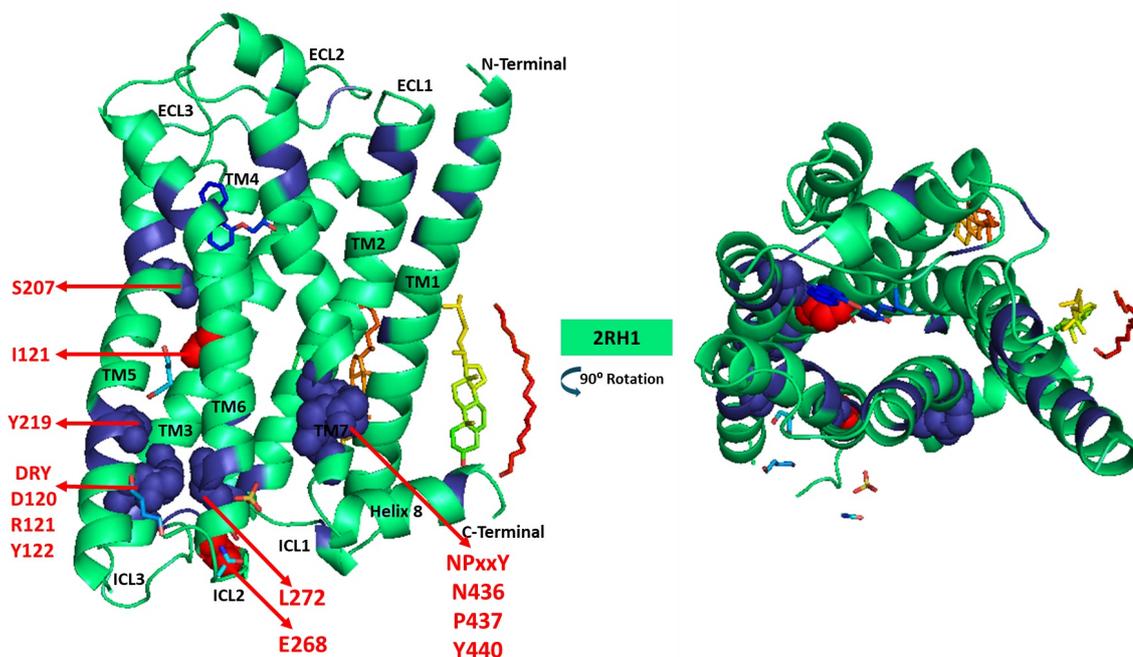


Figure 3.72. Demonstration of ANM mode 9 (dark blue) and functionally important regions (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, L272, and E268) (red and sphere) on the structure.

- Cross-correlation Analysis

Normalized cross-correlation analysis is done through the activation pathway of  $\beta_2$  adrenergic receptor and shown in Figure 3.73.

The highly correlated residues are determined as K60-F61 in ICL1 & E338-L339 in C-terminal, Y141 at the end of TM4 & L305 at the beginning of TM7, V222 in TM5 & A271-L272 in TM6, and A226 in TM5 & L263-L272 in TM6. When the cross correlation analysis performed for this case is compared with the results of the analysis of the simulations from inactive structure (2RH1) to active structure (3SN6), it is seen that the correlation between ICL1 loop and C-terminal is observed for both cases. However, the correlation of L272, the residue, which forms ionic lock, with R131 in inactive case, is observed in this case. Additionally, DRY motif is affected by T68 at

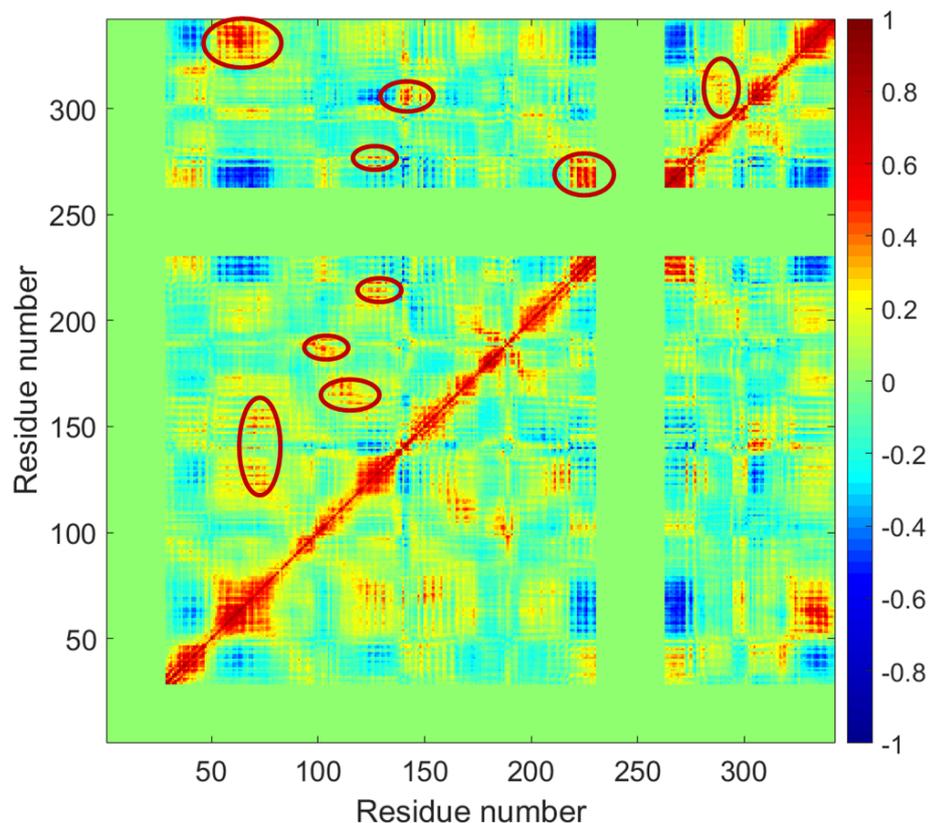


Figure 3.73. Cross correlation map of the activation pathway of  $\beta_2$  adrenergic receptor from 2RH1 to 3P0G.

the end of TM2 and NPxxY motif is correlated with F61-L64 in ICL2, which implies that NPxxY motif is effective in the changes in intracellular side of the structure.

3.2.1.3. ANM-LD Simulations from 3D4S to 3SN6. ANM-LD simulations of  $\beta_2$  adrenergic receptor between the inactive state (3D4S) and active state (3SN6) are performed to investigate the transition pathway of the activation mechanism of  $\beta_2$ -Adrenergic receptor. The inactive structure used in this case (3D4S) is aligned to the inactive structure used in the first two cases (2RH1) and shown in Figure 3.74. The RMSD of the inactive structure (3D4S) with respect to the inactive structure used in the first case (2RH1) is calculated as 0.53 Å.

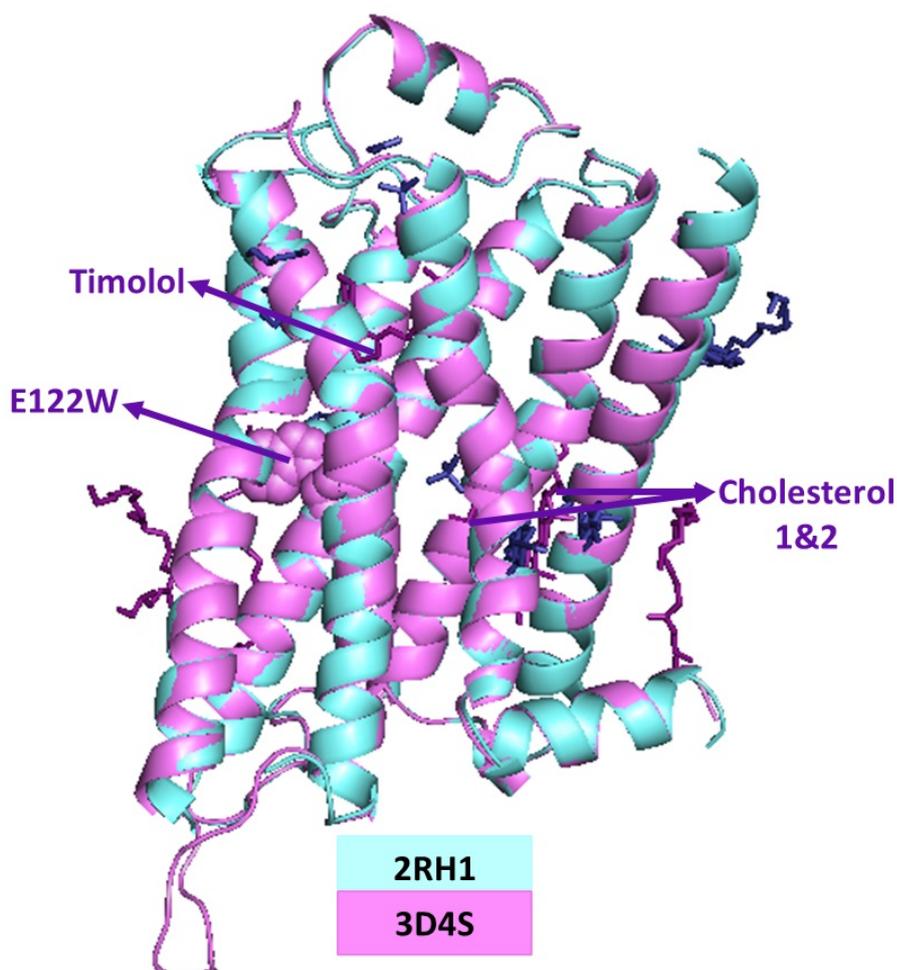


Figure 3.74. Alignment of inactive structure (3D4S) used in this case with the inactive structure (2RH1) used in the first case (RMSD: 0.53 Å).

The comparison of the two inactive structures is given in Table 3.7 in Section 3.2.1. Although the organism of these two inactive structures is homo sapiens, there are some distinctness between them. The inactive structure used in this case has one extra mutation (W122E), which increases the fraction of receptor to fold properly and decreases the affinity to bind the antagonist [57]. Moreover, inactive structure used in this case (3D4S) bound partial inverse agonist timolol at the orthosteric binding site, while inactive structure used in first two cases (2RH1) bound carazolol. For this

reason, there are some differences in the orthosteric binding site. The partial inverse agonist timolol forms hydrogen bond with Y308, N293, and S204. Additionally, T118 is in communication with the thiadiazole group of the partial inverse agonist timolol [51]. Additionally, there is a minor shift in ICL2 and ECL3 in the inactive structure 2RH1, and inactive structure 3D4S has shortened length of C-terminus [51].

Parallel ANM-LD runs are prepared according to different DF values (0.35 Å, 0.40 Å, 0.60 Å, adaptive between 0.4 Å - 0.5 Å - 0.6 Å and 0.1 Å),  $R_{cut}$  values (8 Å, 10 Å, 13 Å, and 18 Å), maximum mode number (30, 100, all modes) and with restricted ANM modes for modeled, unmodified, and back-mutated structures.

In the ANM-LD simulations for the unmodified structures, the structures are prepared by deleting the residues A176-H178, K263-F264, and L342 from the initial structure and E30-V31, and Q231-R239 from the final structure. Since there are missing residues in ECL2 and ICL3, Q231-R239 of the initial structure and A176-H178, K263-F264 of the target structure are modeled by MOD-LOOP server to obtain more accurate results [53]. Thirdly, the mutations of initial structure (W122E, and E187N) and final/target structure (T96M, T98M, and E187N) are withdrawn by PyMOL [54] and used in the parallel runs.

The result summaries of ANM-LD simulations from 3D4S to 3SN6 are given in Table 3.10.

Table 3.10. ANM-LD Result Summaries.

Simulation Name	Cycle Number	Initial RMSD (Å) to Target Structure	Final RMSD (Å) to Target Structure	The Most Selected ANM Modes
$D_F = 0.40\text{Å}$ Modemax = 100 $R_{cut} = 13\text{Å}$	47	2.70	2.01	1, 2, 6, 7
$D_F = 0.60\text{Å}$ Modemax = 30 $R_{cut} = 8\text{Å}$	31	2.70	1.62	1, 4, 7, 9, 10
$D_F = 0.35\text{Å}$ Modemax = 30 $R_{cut} = 8\text{Å}$	60	3.14	2.03	2, 3, 10
$D_F = 0.40\text{Å}$ odemax = 100 $R_{cut} = 13\text{Å}$	37	3.14	2.52	6, 9, 29

The RMSD difference through the activation pathway obtained from ANM-LD simulations is examined and shown in Figure 3.75. The RMSD of the final/target structures according to initial structure is calculated as 3.14 Å for modeled structures and 2.70 Å for unmodified structures. The RMSD of generated structures from ANM-LD simulations are converged around 50 cycles. The lowest RMSD achieved is 1.62 Å, while it remains at the most as high as 2.52 Å. The RMSD results of four ANM-LD runs, which reflects the activation process in the best way with respect to positions of functionally important residues are given in Figure 3.76.

In order to explain the activation process, the movements of important residues are examined along the transition pathway.

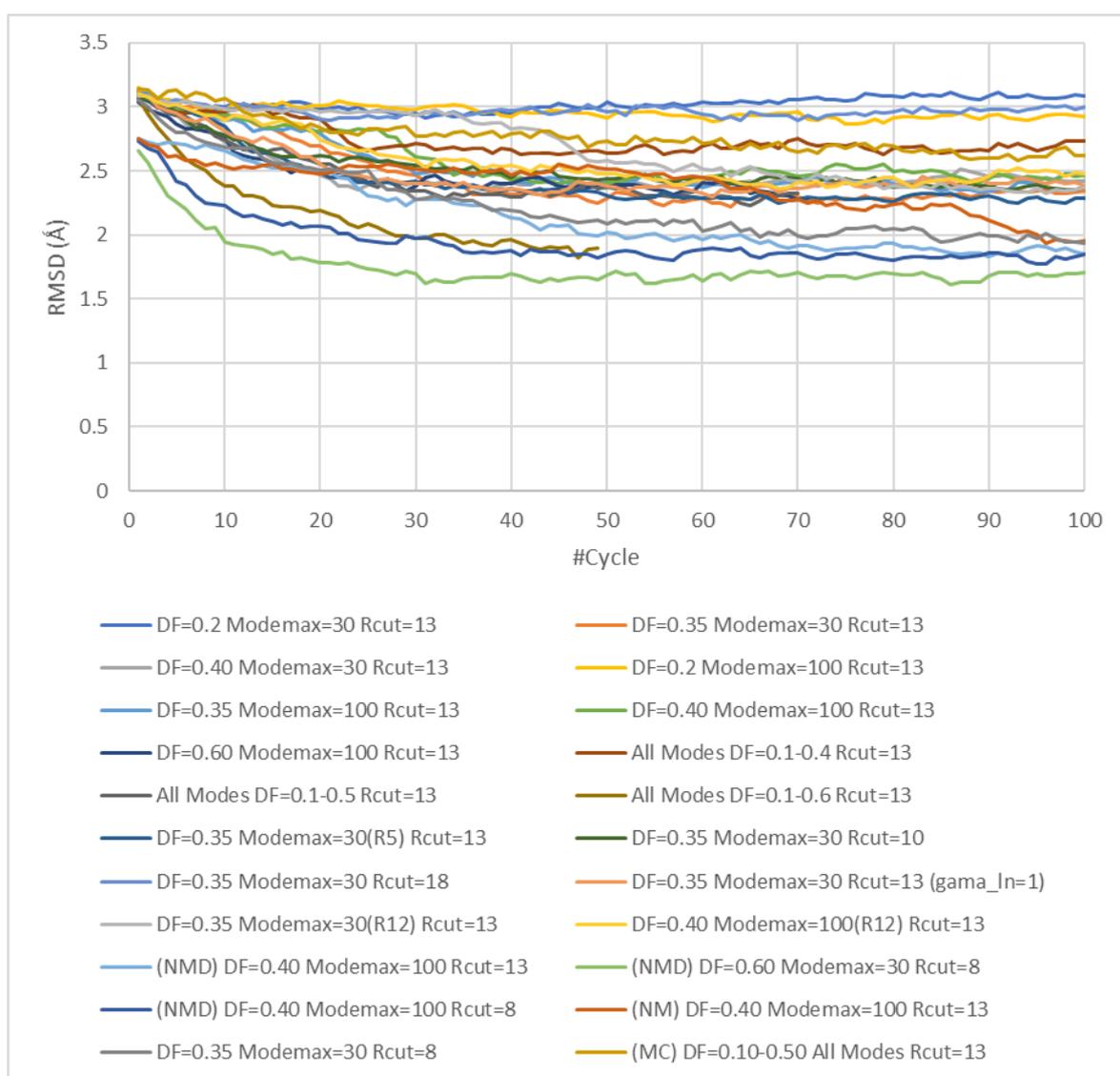


Figure 3.75. The RMSD difference through the simulations of  $\beta_2$  adrenergic receptor from 3D4S to 3SN6.

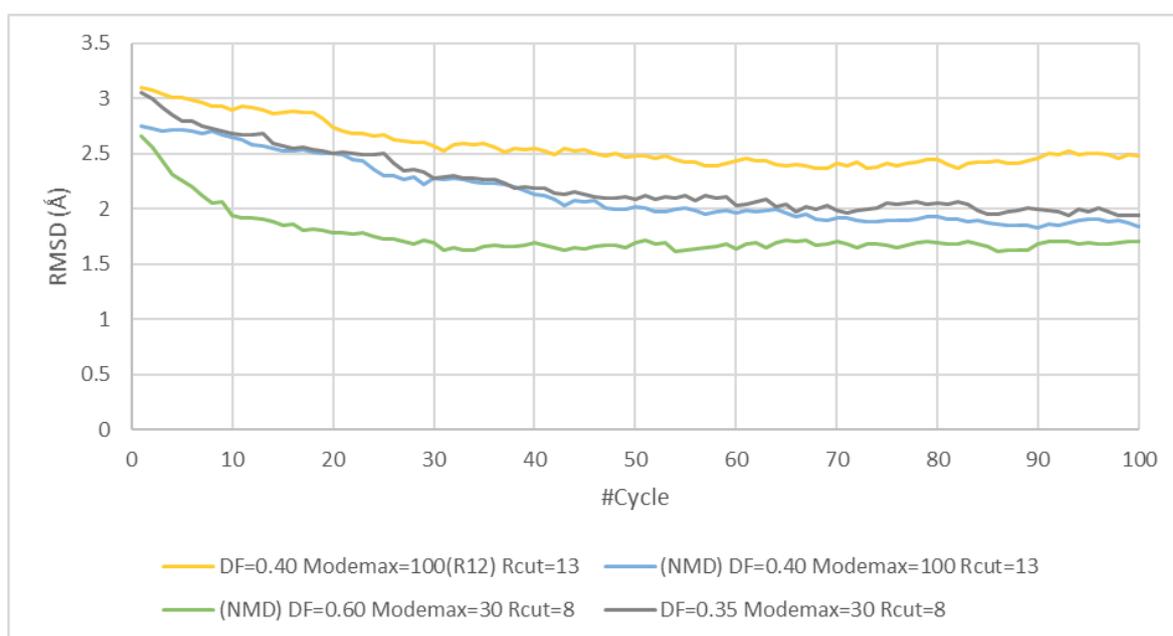


Figure 3.76. The RMSD difference of selected runs through the simulations of  $\beta 2$  adrenergic receptor from 3D4S to 3SN6.

- The TM3-TM6 and TM3-TM7 Distances and Salt Bridge Breakage.

First of all, the distances between TM3-TM6 and TM3-TM7 is examined by using C $\alpha$  atoms of residues R131-L272 and R131-Y326, respectively [52]. Salt bridge breakage through the activation process is investigated by examining the distances of side chains of the residues (R131-E268) which are formed salt bridge in the inactive state [15]. The TM3-TM6 distance is calculated 8.2 Å for the inactive structure and 16.0 Å for the active structure; and the TM3-TM7 distance decreases from 16.5 Å to 11.8 Å through the activation. It is observed that TM3-TM6 and TM3-TM7 distances found approximately equal with the cases in which 2RH1 is used as the inactive structure. The distance between atoms forming the salt bridge in the inactive case is determined as 20.2 Å and 22.2 Å for initial and final structures, respectively. The TM3-TM6 and the TM3-TM7 distances, and salt bridge breakage of R131 and E268 are investigated through the transition pathway in Figure 3.77.

When the last five snapshots of the ANM-LD simulation with DF=0.60 Å Mode-max=30  $R_{cut}$ =8 Å where the unmodified structures used are analyzed, the TM3-TM6 distance is 11.4 Å; the TM3-TM7 distance is 13.3 Å; the distance between the residues R131 and E268 is increased to 25.4 Å caused by the breakage of the salt bridge. When the distance profile is compared with the other cases, it is seen that salt bridge breakage shows similar pathway through the activation pathway. The first event that occurs in the activation process is the breaking of the salt bridge, which affects the TM3-TM6 distance.

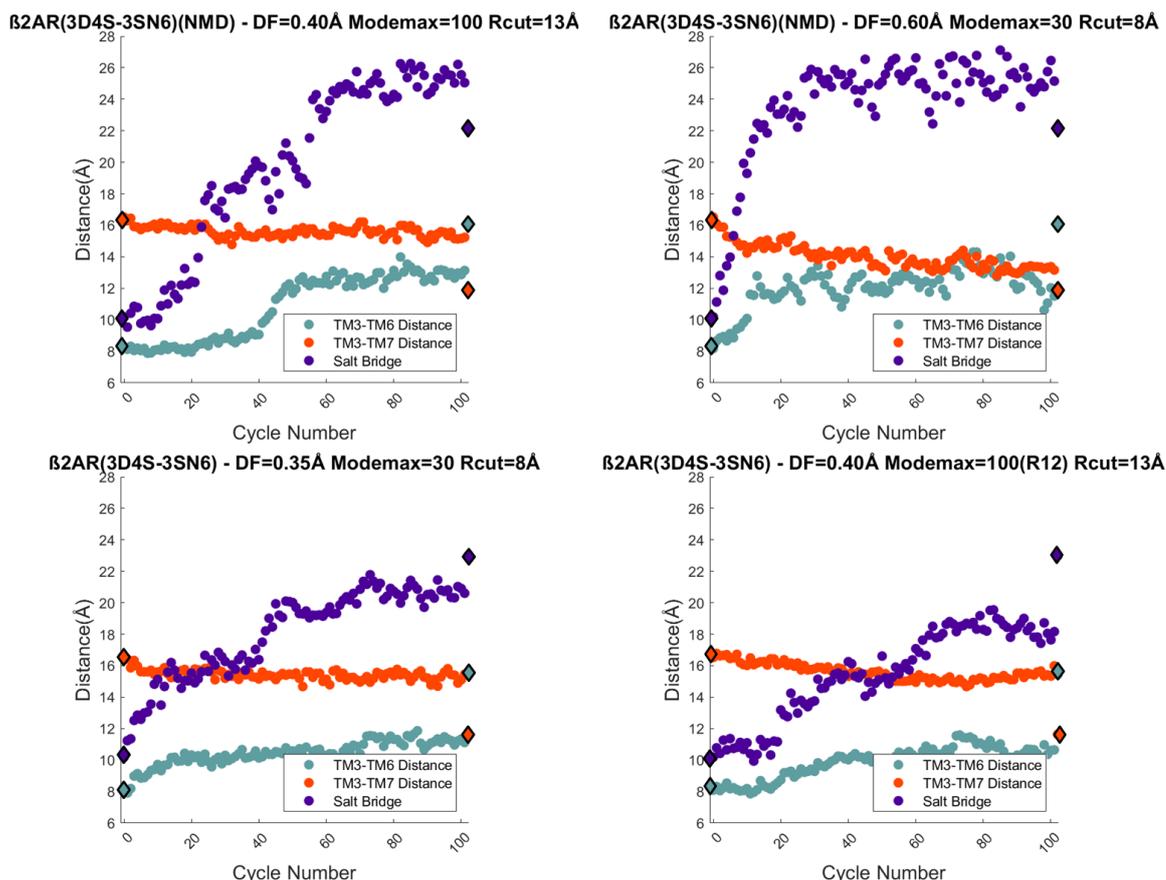


Figure 3.77. Functionally important distances through the transition pathway obtained from ANM-LD results of 3D4S to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres).

- DRY and NPxxY Motifs.

The RMSD of functionally important motifs DRY (D130- R131- Y132) and NPxxY (N322- P323- Y326) [15] with respect to the initial structure is calculated by using  $C\alpha$  atoms through the activation process and given in Figure 3.78. The RMSDs of DRY and NPxxY motifs of the active structure with respect to inactive structure are 1.4 Å and 2.7 Å, respectively.

When the last five snapshots of the ANM-LD simulation with DF=0.60 Å Modemax=30  $R_{cut}$ =8 Å where the unmodified structures used are analyzed, the RMSD of

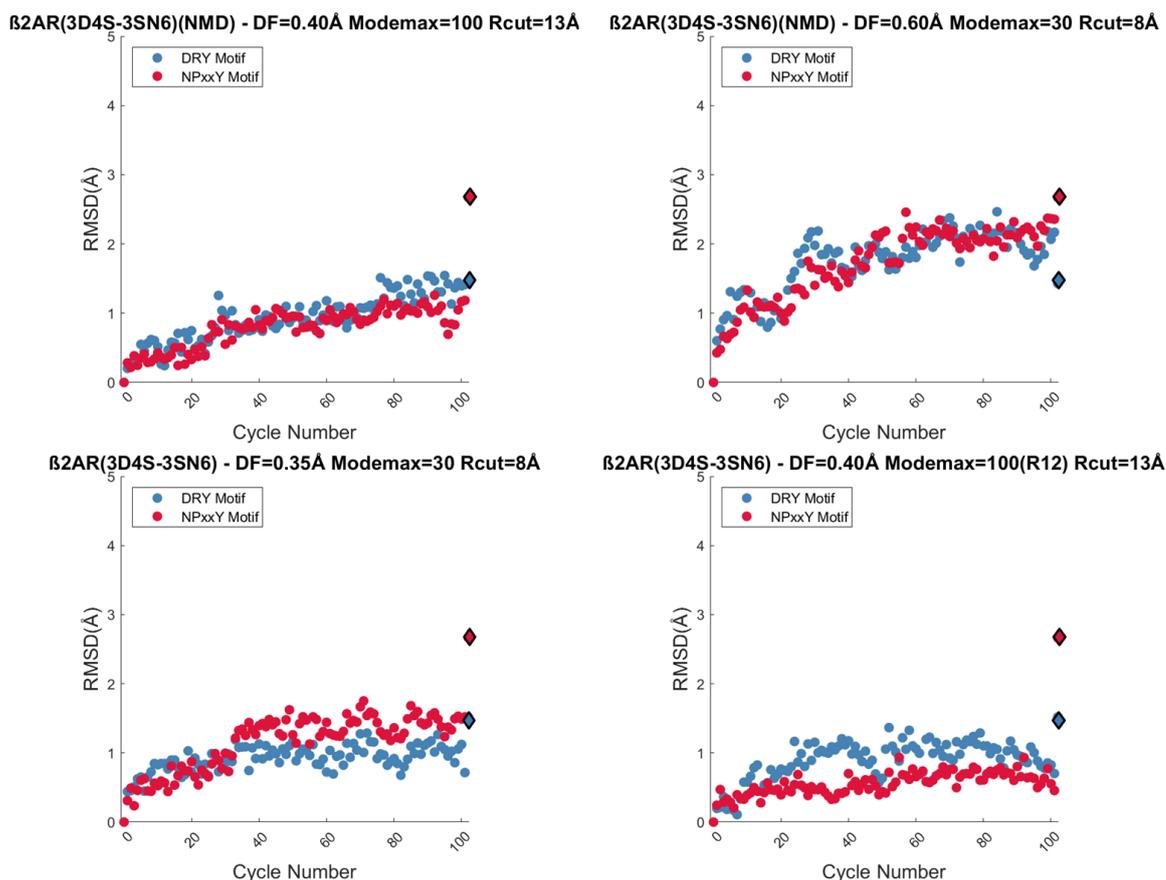


Figure 3.78. RMSD of important motifs through the activation of  $\beta 2$  adrenergic receptor from 3D4S to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres).

DRY motif is calculated as 2.0 Å; the RMSD of NPxxY motif is calculated as 2.3 Å. When the RMSD profile of important motifs is investigated, it is seen that DRY motif unable to converge to the target structure.

- Connector Region, Ligand Binding Site, and G-protein Binding Site.

The RMSD of connector (I121), ligand binding site (S207), and G-protein binding site (Y219) [15] with respect to initial structure is determined by using  $C\alpha$  atoms through the activation pathway. The RMSD between the initial and active/target structures is 1.6 Å, 1.5 Å, and 2.3 Å for the connector, the ligand binding site, and

G-protein binding site, respectively. The RMSDs of connector, ligand binding site, and G-protein binding site through the transition pathway are given in Figure 3.79.

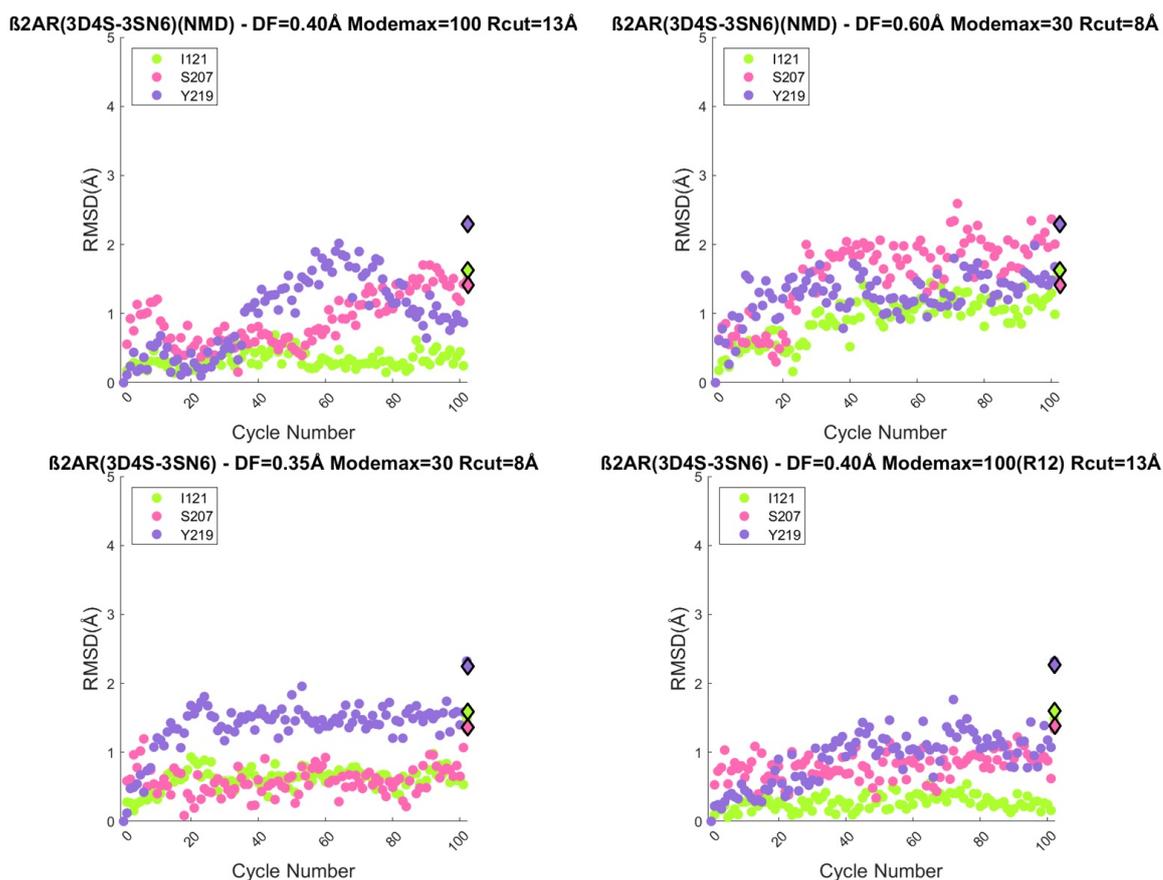


Figure 3.79. RMSD of important residues through the activation of  $\beta_2$  adrenergic receptor from 3D4S to 3SN6 (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the last five snapshots of the ANM-LD simulation with DF=0.60 Å Modemax=30  $R_{cut}$ =8 Å where the unmodified structures used are analyzed, the RMSD is calculated as 1.2 Å, 2.1 Å, and 1.5 Å for the connector, the ligand binding site, and G-protein binding site, respectively. When the RMSD profile of G-protein binding site is investigated, it is seen that the RMSD of Y219 unable to converge to the target structure. Although there are differences in orthosteric ligand binding site between inactive case (3D4S) used in this case and inactive case (2RH1) used in the first two

cases, there is a difference of 0.2 Å in the RMSD of S207. Nevertheless, the RMSD of G-protein binding site in first two cases converges to the active structure better than this case.

- The Key Modes of Motion.

ANM modes selected in transition from the inactive structure to the active structure are investigated to understand the underlying mechanism in the activation process. The analysis of ANM modes are performed according to procedure applied in the Section 3.1.1.1. Selected modes through the transition pathway from 3D4S to 3SN6 are given in Figure 3.80.

First fifty cycles of parallel ANM-LD simulations are investigated to determine ANM mode frequency, which is shown in Figure 3.81. Considering two different ANM-LD simulations that show the activation process in the best way with respect to conformational changes through the activation, ANM mode 6 and ANM mode 1 are defined as the most selected modes with highest overlap values.

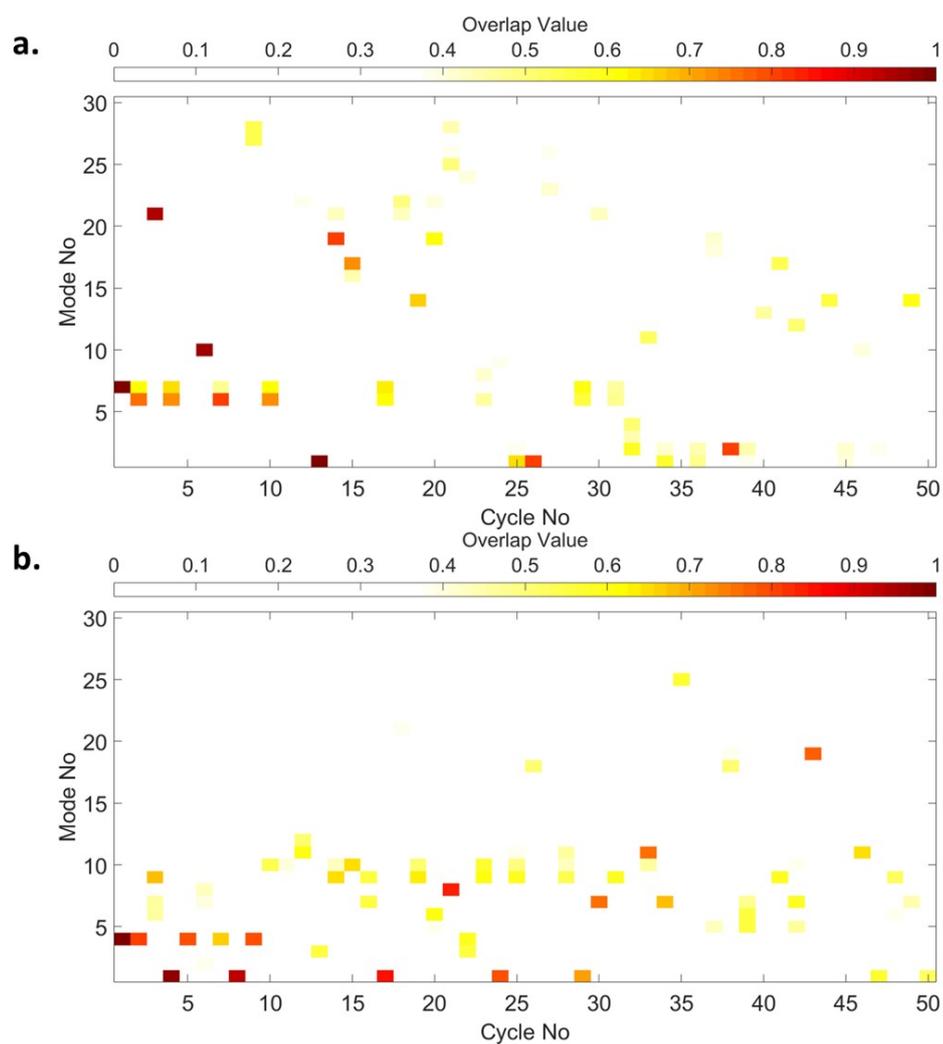


Figure 3.80. Selected Mode Mapping a)(NMD) DF=0.40 Å Modemax=100  $R_{cut}$ =13

b)(NMD) DF=0.60 Å Modemax=30  $R_{cut}$ =8 c)DF=0.35 Å Modemax=30  $R_{cut}$ =8

d)DF=0.40 Å Modemax=100(R12)  $R_{cut}$ =13.

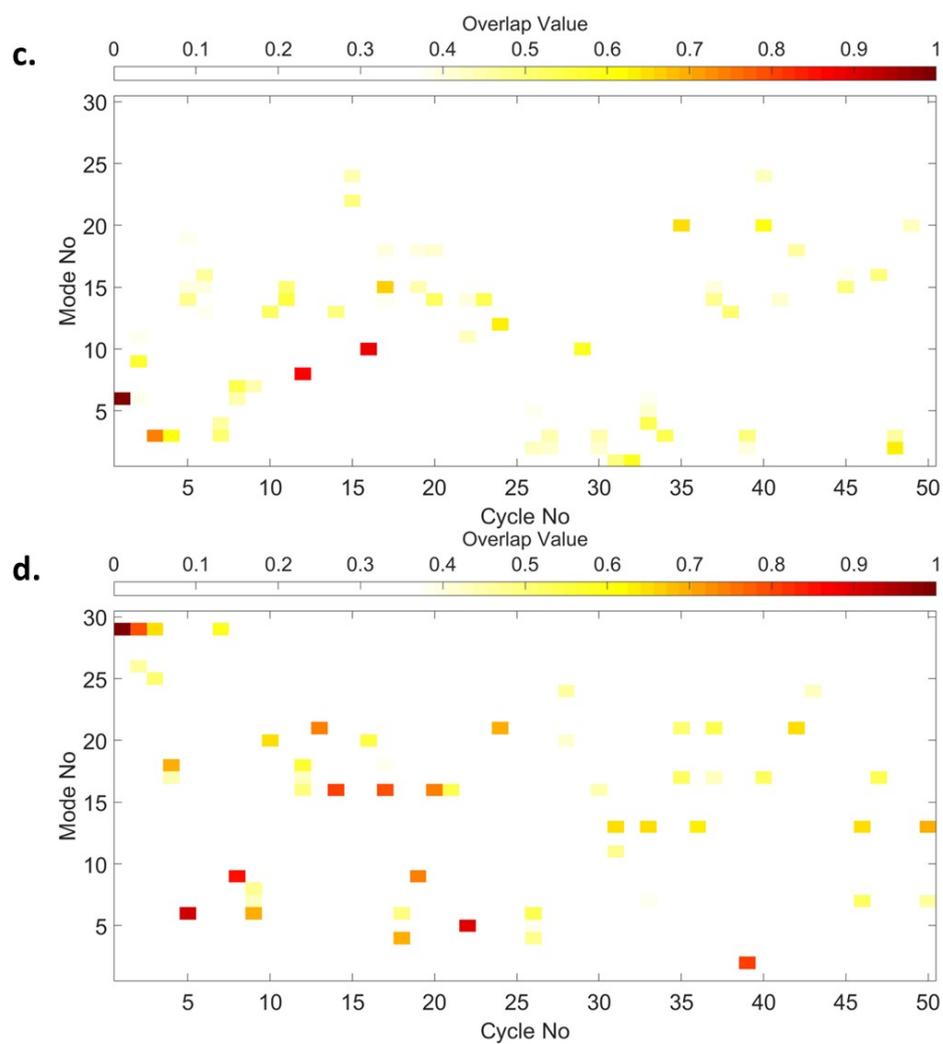


Figure 3.80. Selected Mode Mapping a)(NMD) DF=0.40 Å Modemax=100  $R_{cut}$ =13

b)(NMD) DF=0.60 Å Modemax=30  $R_{cut}$ =8 c)DF=0.35 Å Modemax=30  $R_{cut}$ =8

d)DF=0.40 Å Modemax=100(R12)  $R_{cut}$ =13. (cont.)

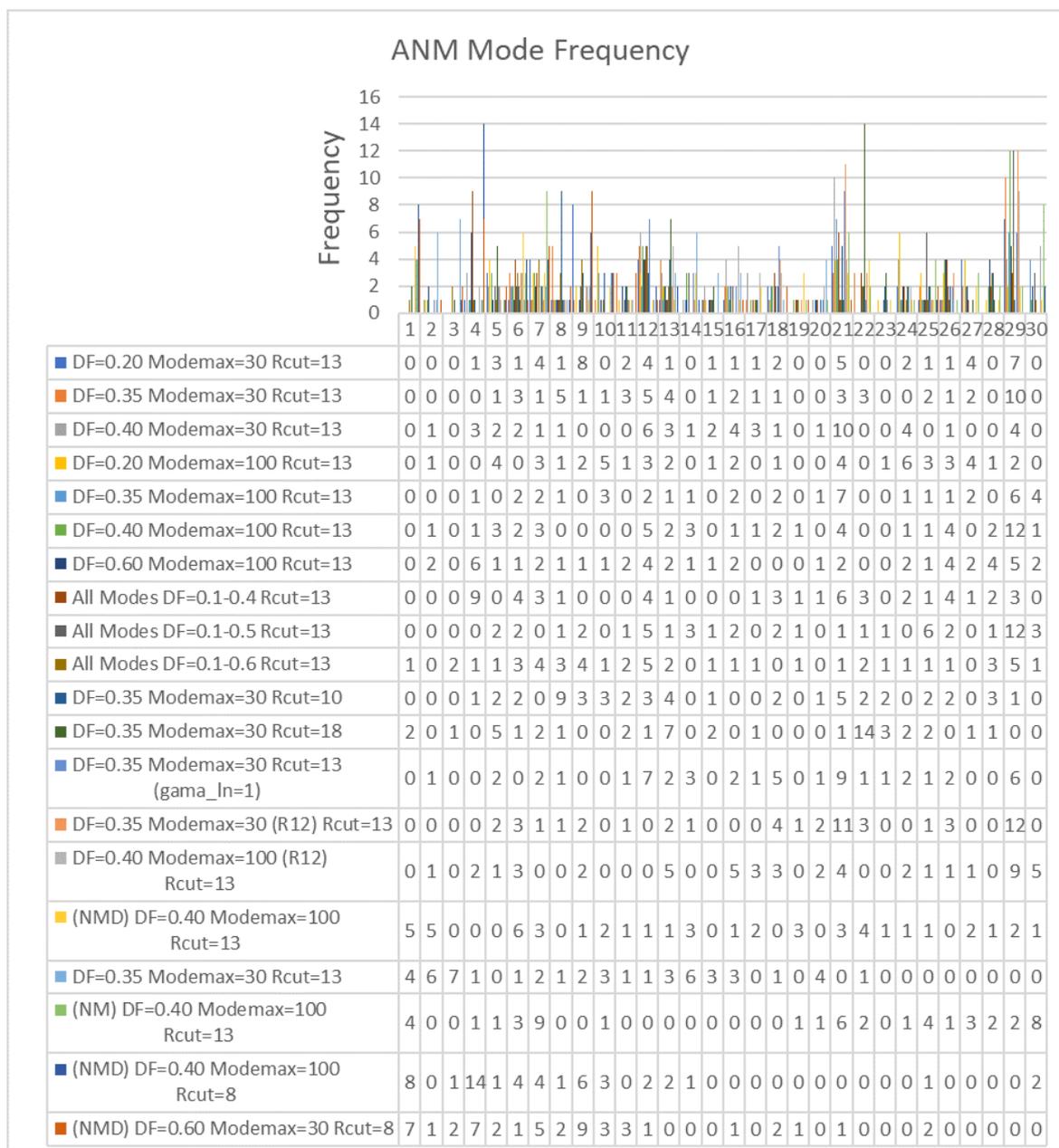


Figure 3.81. ANM mode frequency of parallel simulations from 3D4S to 3SN6.

MS of the difference vector calculated between the initial structures (2RH1 – 3D4S) and the final active structure (3SN6) is shown in Figure 3.82. It can be deduced from the profiles of both cases that the minimum and maximum points overlap with each other. There are only a few differences in MS profile between the residues H172-

D192 in ECL2, and L302-I314 in ECL3 and extracellular side of TM7, because there is a minor shift in ECL3 in the inactive structure 2RH1 [51].

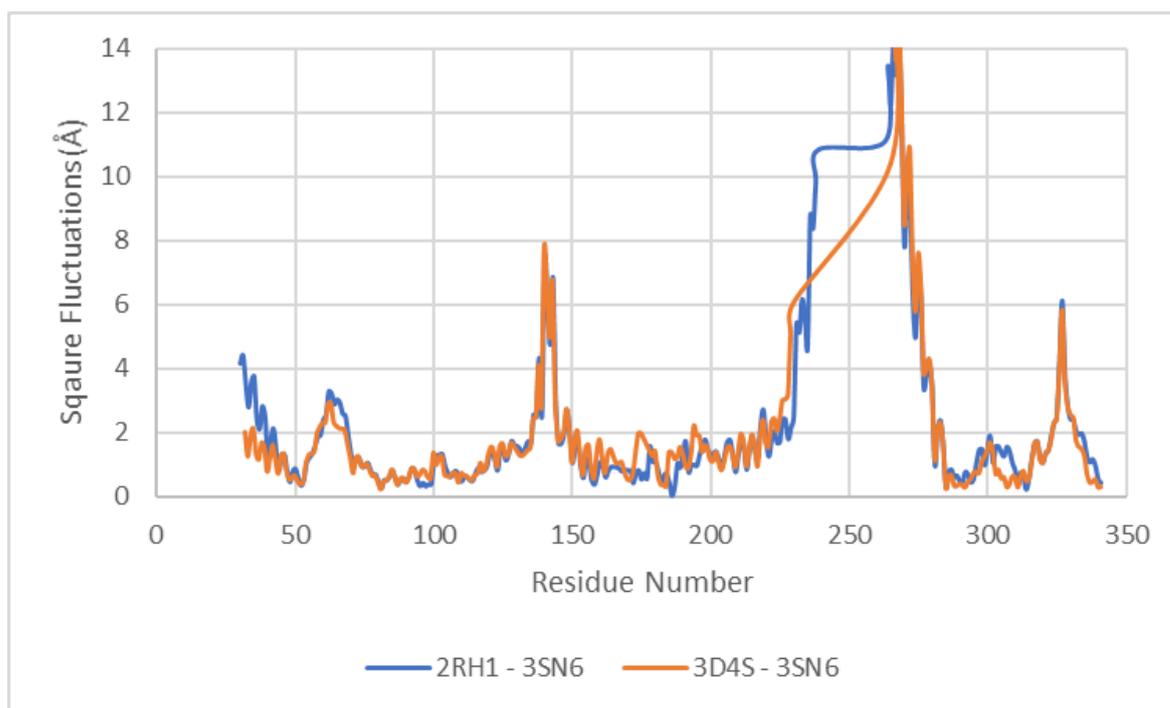


Figure 3.82. Difference vector between initial (2RH1 - 3D4S) and final (3SN6) structure.

When this graph is contrasted with the ANM mode profiles in Figure 3.83, it is observed that the minimum and maximum points of the mean square fluctuations graphs of the selected modes are compatible with the mean square fluctuation graph of the difference vector. In the mean square fluctuations profile of difference vector residues between 184 and 211 show a more stable profile than selected ANM mode shape profiles.

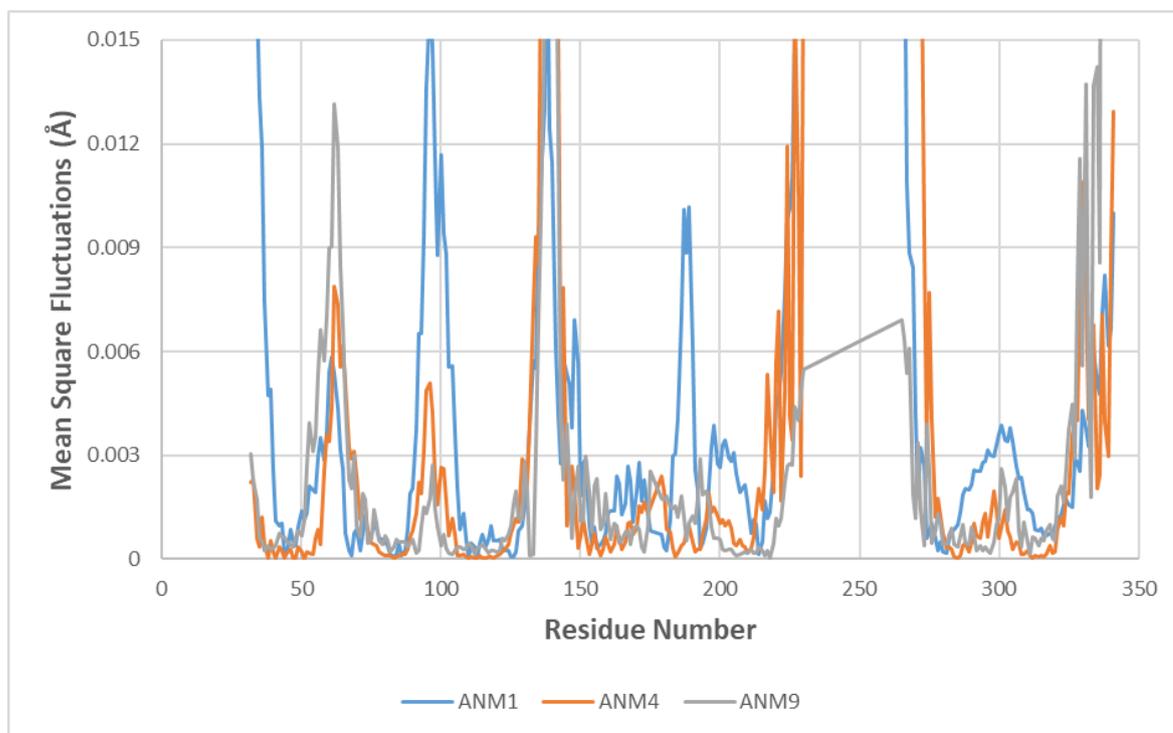


Figure 3.83. Mean square fluctuations of selected ANM modes of  $\beta 2AR$  simulations from 3D4S to 3SN6.

The hinge residues which is responsible of the motion of the structure are identified as minimum points of the mean square fluctuations profile. Mean square fluctuations of ANM mode 1 is shown in Figure 3.84. Hinge residues of ANM mode 1 are identified as L45, T68, G83, V86, T110, V126, M156, A181, F193, I214, T281, and V317-S319. When the hinge residues of ANM mode 1 are investigated, it is seen that hinge residues overlap with F193 and T110 which are proposed as allosteric residues in ECL and orthosteric ligand binding site, respectively [56].

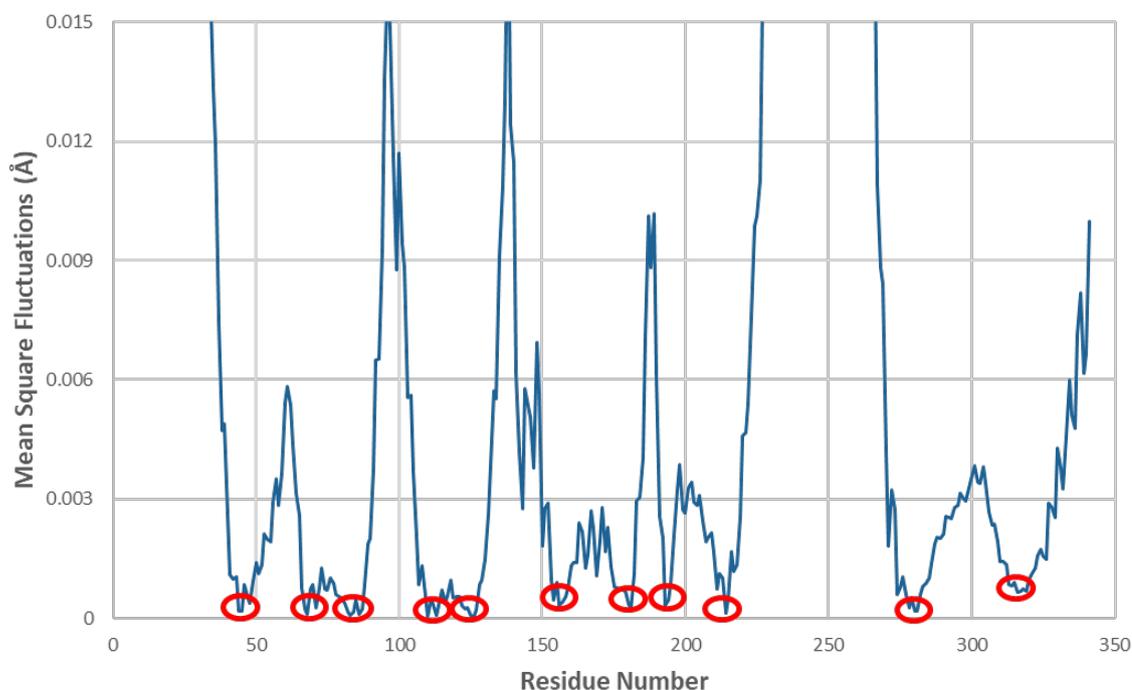


Figure 3.84. ANM mode 1 shape of the initial state of  $\beta 2$  adrenergic receptor.

Mean square fluctuations of ANM mode 4 is shown in Figure 3.85. The hinge residues of ANM mode 4 are I38, N51, L80-A85, E107, A119-I121, K149, C184, C191-F193, V210, C285-W286, and I314-Y316. Hinge residues of ANM mode 4 represent the movement in connector region better than the other selected modes. The residues affect the connector region I121 and C285 overlap with the hinge residues. Moreover, Y316 (orthosteric ligand binding site) which is observed in allosteric sites of intermediate structure of  $\beta 2$ -Adrenergic receptor is identified as hinge residue [56]. It can be concluded that thanks to the ANM modes, we can detect not only allosteric interacting residues in inactive structures, but also allosteric residues in intermediate structures. Additionally, F193 is determined as hinge residue as in other cases of  $\beta 2$ -Adrenergic receptor, which is an important residue belonging to most promising allosteric site in EC loop [56].

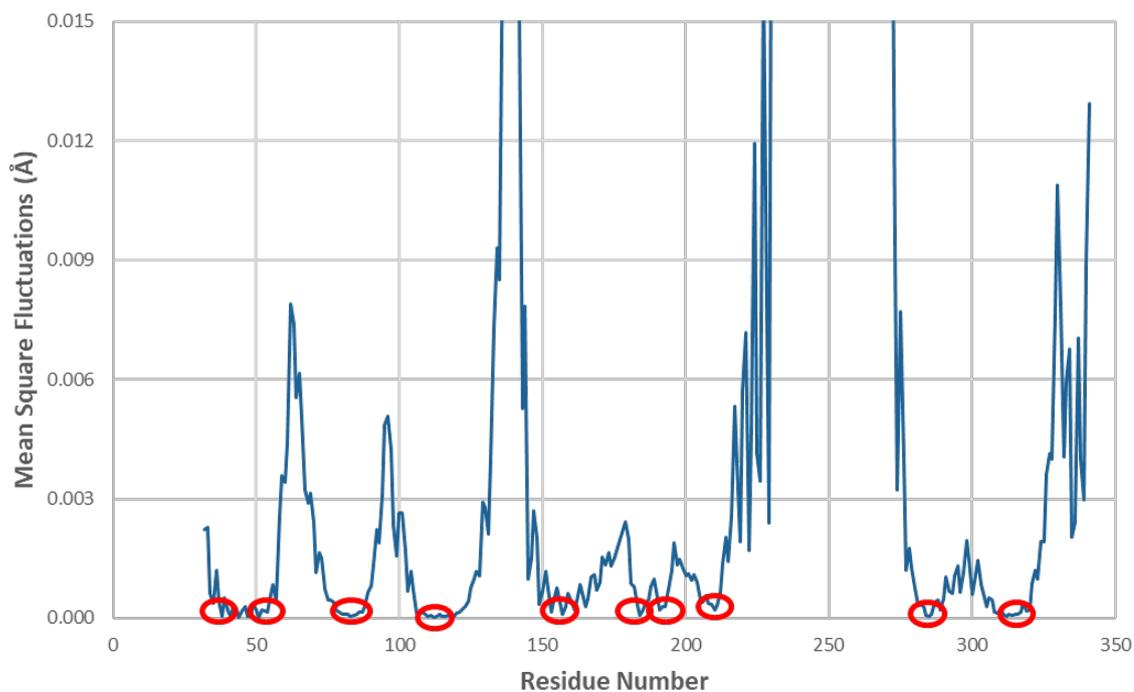


Figure 3.85. ANM mode 4 shape of the initial state of  $\beta 2$  adrenergic receptor.

Mean square fluctuations of ANM mode 9 is shown in Figure 3.86. The hinge residues are G37-M40, A46, F71, A92, W105, I121, F133, K147, W173, E188, S207, V218, K273, V297, L311, and R333. S207 located in orthosteric ligand binding site, I121 in connector region, and F133 located in the area of residues affecting the DRY motif overlap with the hinge residues of ANM mode 9.

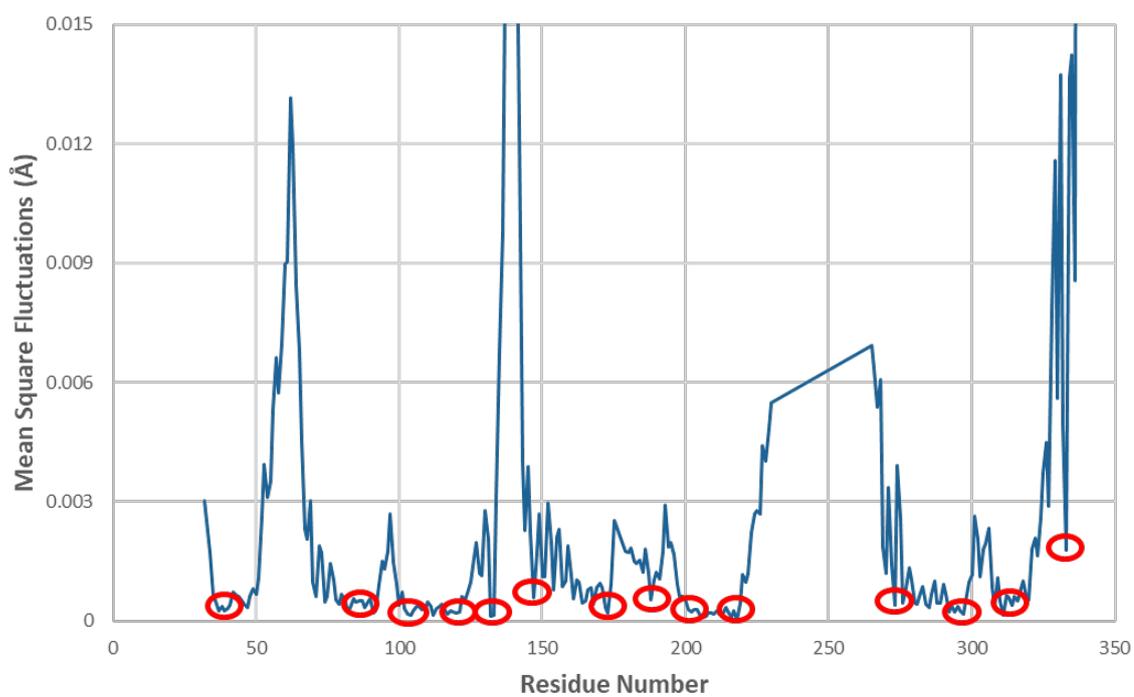


Figure 3.86. ANM mode 9 shape of the initial state of  $\beta 2$  adrenergic receptor.

The hinge residues of ANM mode 1, 4, and 9 are demonstrated on the inactive structure of  $\beta 2$  adrenergic receptor in Figure 3.87.

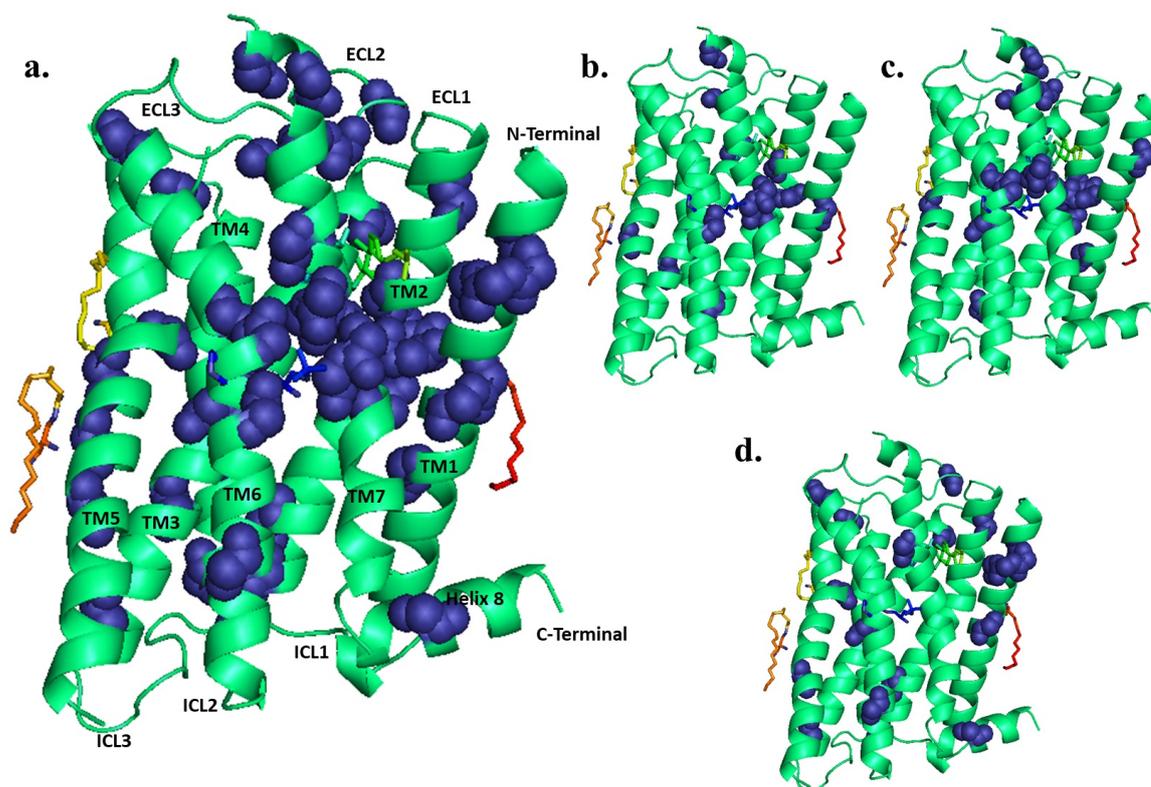


Figure 3.87. Demonstration of hinge residues of ANM mode 1, 4, and 9 (dark blue and sphere) on the structure a) Hinges of ANM mode 1, 4, and 9. b) Hinges of ANM mode 1. c) Hinges of ANM mode 4. d) Hinges of ANM mode 9.

Functionally important regions (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) and the hinge residues of ANM mode 1, 4, and 9 are demonstrated on the initial structure in Figure 3.88.

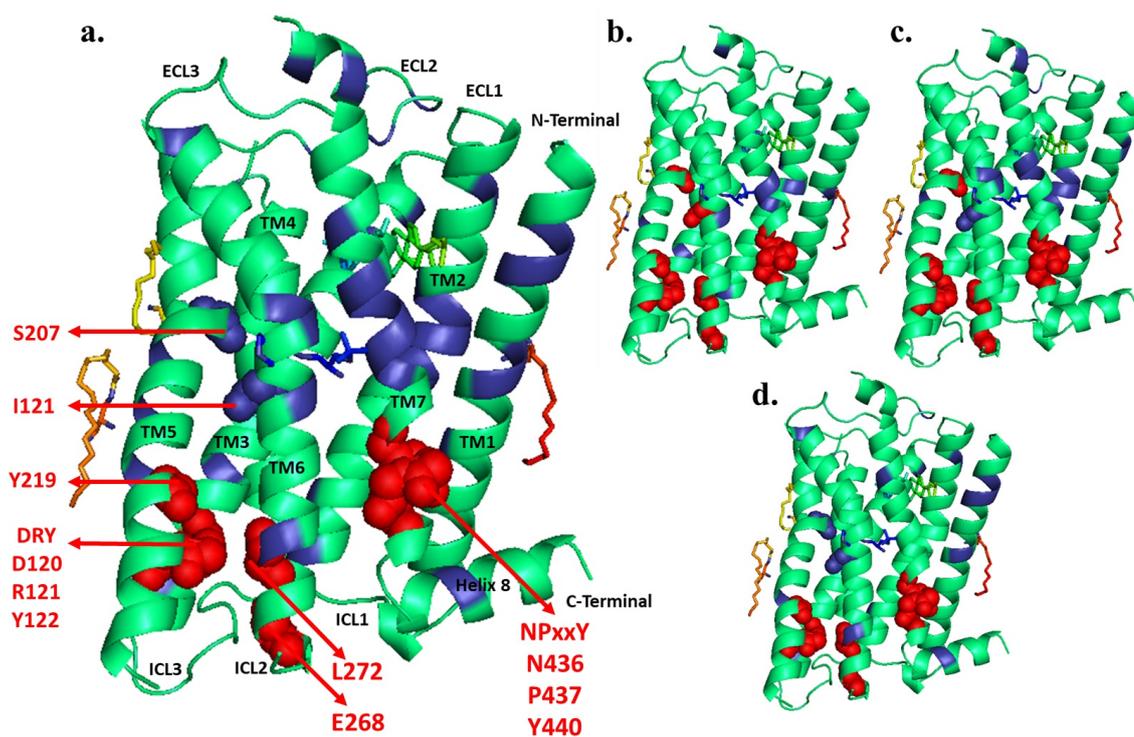


Figure 3.88. Demonstration of selected ANM modes 1, 4, and 9 hinges (dark blue) functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a) ANM mode 1, 4, and 9 with b) ANM mode 1 with c) ANM mode 4 with d) ANM mode 9 with.

- Cross-correlation Analysis.

Normalized cross-correlation analysis is used to identify the correlated residues. The analysis of cross-correlation is performed according to procedure applied in the Section 3.1.1.1. Normalized cross-correlation analysis is done through the activation pathway of  $\beta_2$  adrenergic receptor and demonstrated in Figure 3.89.

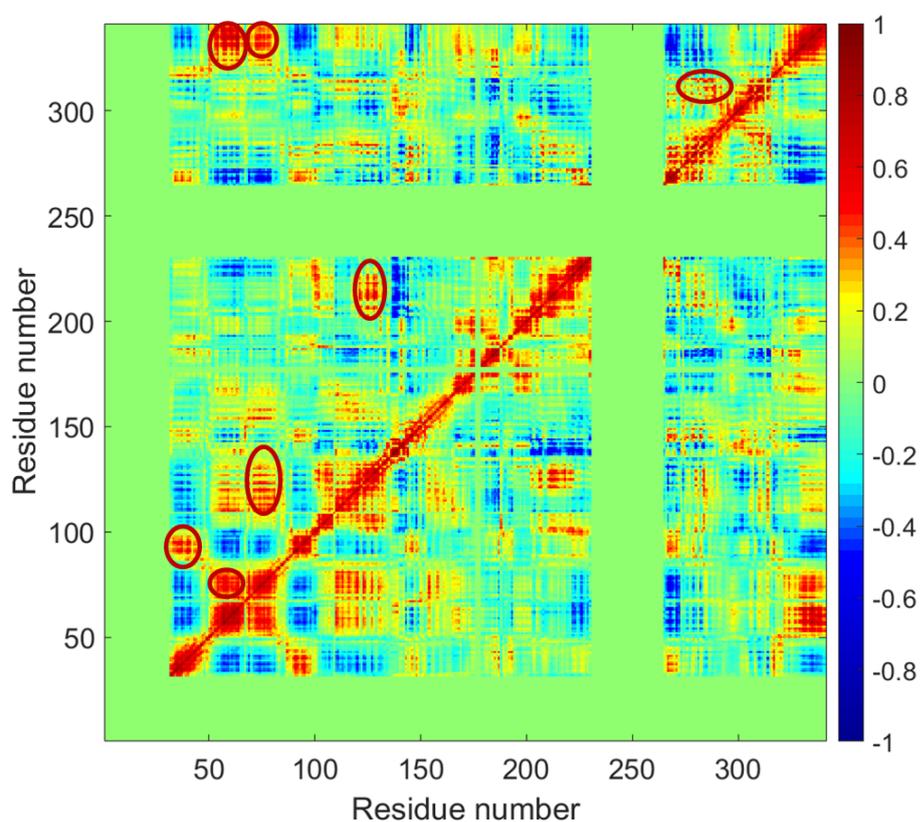


Figure 3.89. Cross correlation map of the activation pathway of  $\beta_2$  adrenergic receptor from 3D4S to 3SN6.

The highly correlated residues are V54-T56 at the end of TM1 / R63-L64 in ICL1, V54 at the end of TM1 / D331-F332 in C-terminal, A57 at the end of TM1 / L339 in C-terminal, R63-L64 in ICL1 / S329-L339 in C-terminal, M40 in TM1 / A91 in TM2, A128 in TM3 / P211-V213 in TM5, C125 in TM3 / P211-V213 in TM5, and L284 in

TM6 / I314 in TM7. When highly cross correlation analysis is investigated, it is seen that the residues at the end of TM1 affects C-terminal. Moreover, the connector region is affected by C125 and A128 in TM3.

3.2.1.4. ANM-LD Simulations from 3D4S to 3P0G. In this section, ANM-LD simulations are executed to investigate the activation mechanism of  $\beta$ 2-Adrenergic receptor from the different inactive structure (3D4S) to a different active structure (3P0G). The comparison of the active and inactive structures is given in given in Table 3.7 in Section 3.2.1.

Parallel runs from inactive structure (3D4S) to a different active structure (3P0G) are prepared to investigate the activation mechanism of  $\beta$ 2-Adrenergic receptor. Structures used in the ANM-LD simulations are prepared in three different ways. Firstly, R228-L230, and K263-C265 are from initial structure; D23-V31 and R343-R344 are deleted from final/target structure for unmodified ANM-LD simulations. Secondly, R228-L230 and K263-C265 residues located at the end of TM5 and in ICL3 are modeled by MOD-LOOP [53]. In the third case, the mutations of the initial structure (W122E, and E187N) and the final structure (E187N) are withdrawn by PyMOL [54].

The result summaries of ANM-LD simulations from 3D4S to 3P0G are given in Table 3.11

Table 3.11. ANM-LD Result Summaries.

Simulation Name	Cycle Number	Initial RMSD ( $\text{\AA}$ ) to Target Structure	Final RMSD ( $\text{\AA}$ ) to Target Structure	The Most Selected ANM Modes
$D_F = 0.10 - 0.60\text{\AA}$ All Modes $R_{cut} = 10\text{\AA}$	34	2.51	1.59	1, 2, 3, 8
$D_F = 0.60\text{\AA}$ Modemax = 30 $R_{cut} = 10\text{\AA}$	17	2.51	1.48	1, 2, 8
$D_F = 0.35\text{\AA}$ Modemax = 100 $R_{cut} = 10\text{\AA}$	40	2.51	1.56	1, 2, 3, 8
$D_F = 0.35\text{\AA}$ Modemax = 100 $R_{cut} = 13\text{\AA}$	81	2.51	1.67	2, 8, 9

RMSD along the activation process obtained by ANM-LD simulations are demonstrated in Figure 3.90. RMSD of initial structure is determined as  $2.51\text{ \AA}$  for modeled structures and  $2.07\text{ \AA}$  for unmodified structures. RMSD profile converges around 30 cycles scatters  $2.32\text{ \AA}$  and  $1.70\text{ \AA}$ . The RMSD values of four runs which shows the activation pathway in the best way are given in Figure 3.91.

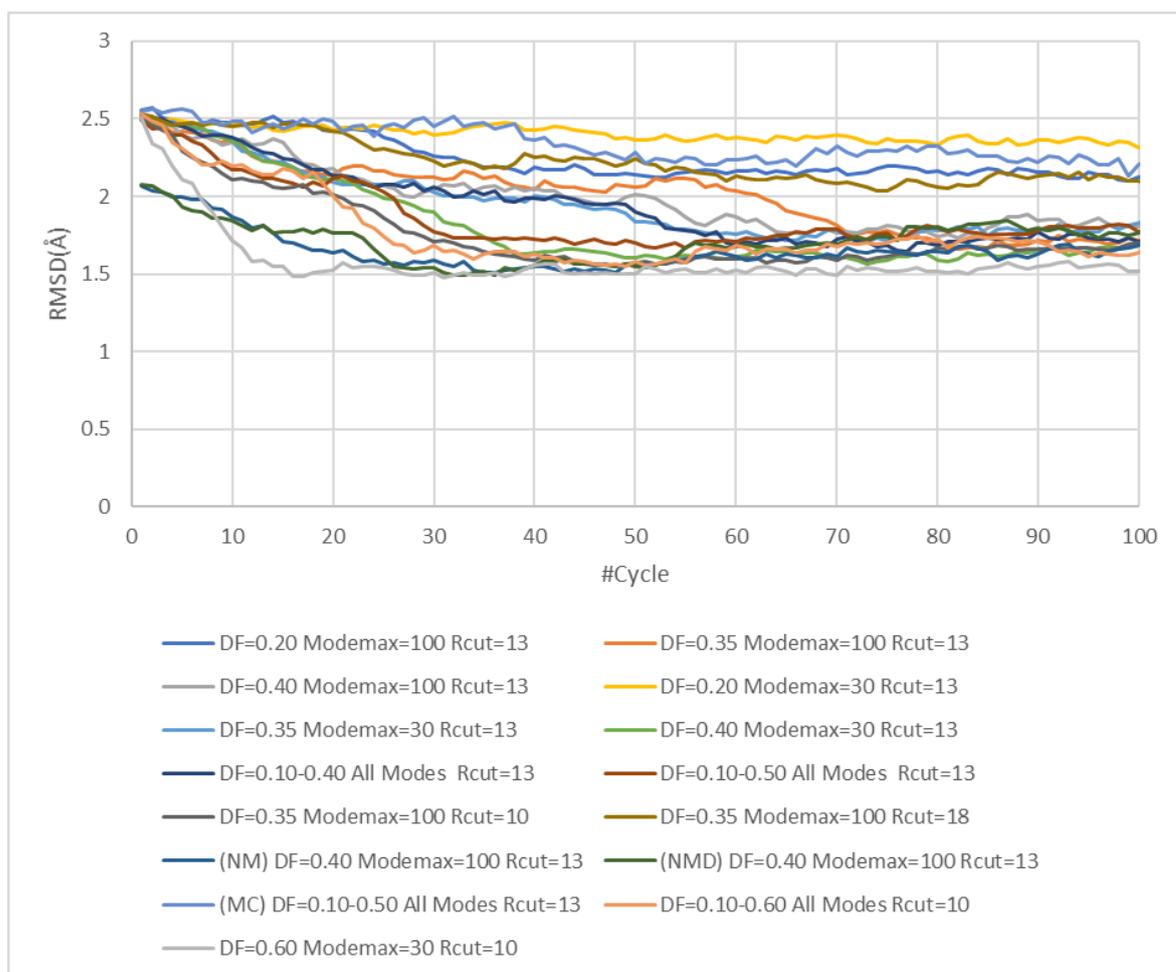


Figure 3.90. RMSD difference through the simulations of  $\beta 2$  adrenergic receptor from 3D4S to 3P0G.

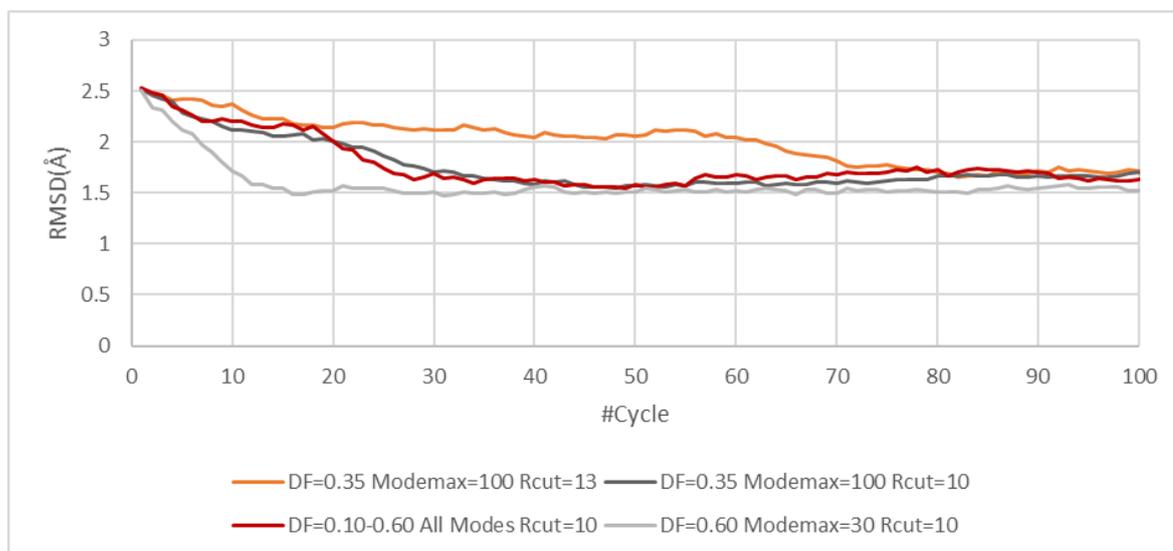


Figure 3.91. RMSD difference of selected runs through the simulations of  $\beta 2$  adrenergic receptor from 3D4S to 3P0G.

In order to clarify the activation process, the conformational changes through the activation are observed.

- TM3-TM6 and TM3-TM7 Distances and Salt Bridge Breakage.

TM3-TM6 and TM3-TM7 distances are calculated to examine the movement of residues that are important for the activation of the  $\beta 2$  adrenergic receptor. In the calculation process,  $C\alpha$  atoms of R131-L272 are used to calculate the distance between TM3 and TM6;  $C\alpha$  atoms of R131 (DRY Motif) and Y326 (NPxxY Motif) are used to calculate the distance between TM3 and TM7. Salt bridge breakage through the activation process is investigated by investigating the distances of side chains of the residues (R131-E268) which are formed salt bridge in the inactive state [15, 49]. The TM3-TM6 residue distances are 7.9 Å and 13.9 Å for the initial and final structures. The TM3-TM7 residue distances are calculated as 16.5 Å and 11.4 Å for initial and final structures. The distance of sidechains between R131 and E268 is 10.0 Å and 21.5 Å for inactive and active structures, respectively. The TM3-TM6 and TM3-TM7

distances and the distance between the residues R131 and E268 which formed salt bridge in the inactive state through the activation pathway of  $\beta$ 2-Adrenergic receptor is demonstrated in Figure 3.92.

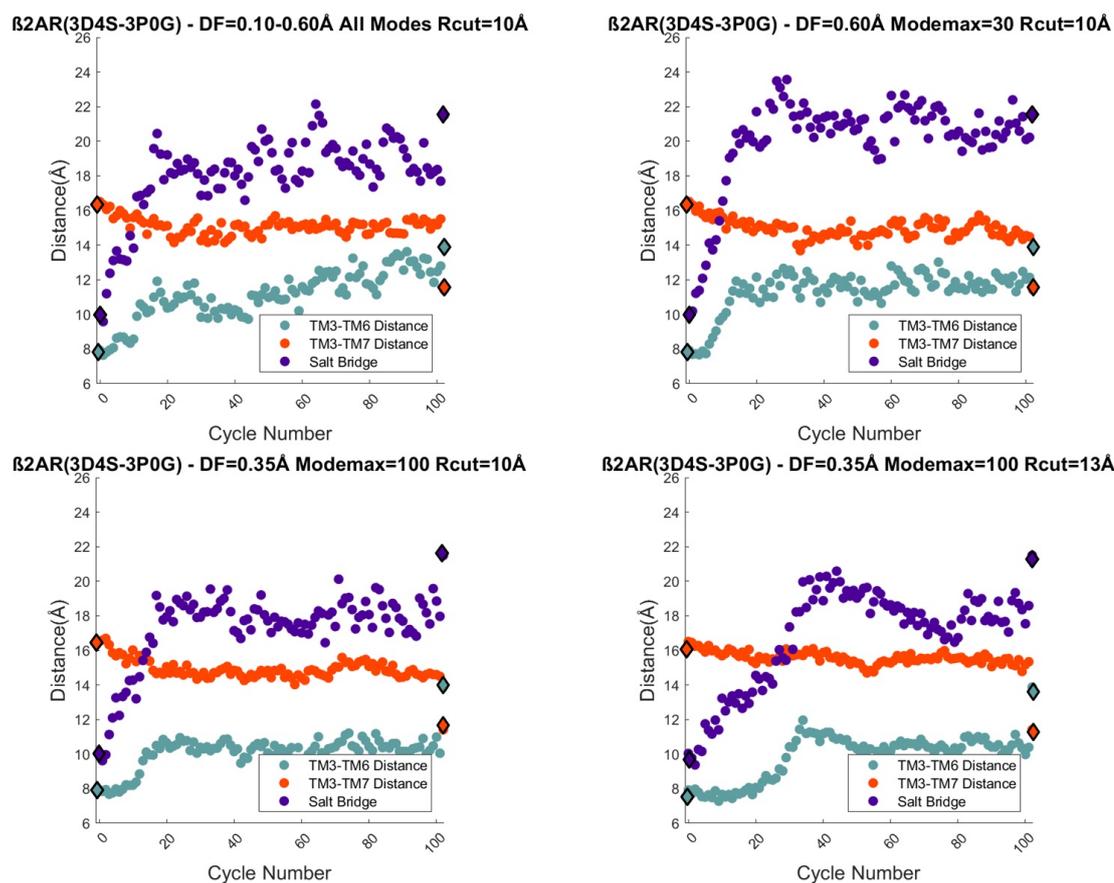


Figure 3.92. Functionally important distances through the transition pathway obtained from ANM-LD results of 3D4S to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the last five snapshots of the ANM-LD simulation that most accurately mirrors the activation pathway from fifteen parallel simulations are considered, the TM3-TM6 distance is 12.5 Å; the TM3-TM7 distance is 15.2 Å; the distance between the residues R131 and E268 is increased to 18.2 Å caused by the breakage of the salt bridge. When the distance profile analyzed, it is seen that the distance profile does not converge to the final structure. Although, the simulations from the same

inactive structure to other active structure (3SN6) give better results than this case, the simulations prepared from other inactive structure (2RH1) to other active structure (3SN6) converge to the target structure in the best way. In the inactive structure (3P0G), R131 in DRY motif has different position and interact with nanobody, which can affect the opening in the intracellular part of TM3 and TM6 and the salt bridge breakage. Normally, the salt bridge breakage causes an increase of 11.5 Å in between R131 and E268. However, in this case there is an increase of 8.2 Å, which can be caused from the interaction between R131 and nanobody.

- DRY and NPxxY Motifs.

The RMSD of the important motifs DRY (D130–R131–Y132) and NPxxY motif (N322–P323–Y326) are explored through the transition pathway. The RMSD of DRY and NPxxY motifs with respect to inactive structure are determined as 1.8 Å and 2.3 Å. The RMSD of DRY and NPxxY motifs through the activation pathway is demonstrated in Figure 3.93. When the last five snapshots of the ANM-LD simulation with parameters  $DF = 0.10-0.60$  Å All Modes and  $R_{cut} = 10$  Å is considered, the RMSD of DRY motif is calculated as 1.9 Å; the RMSD of NPxxY motif is calculated as 0.8 Å. Although the RMSD of DRY motif converges to the target structure, the RMSD of NPxxY motif does not approach to the target value. The results of the case in which 2RH1 used as inactive structure and 3SN6 used as active structure converge to the target structure better than other cases in terms of RMSDs of DRY motif and NPxxY motif.

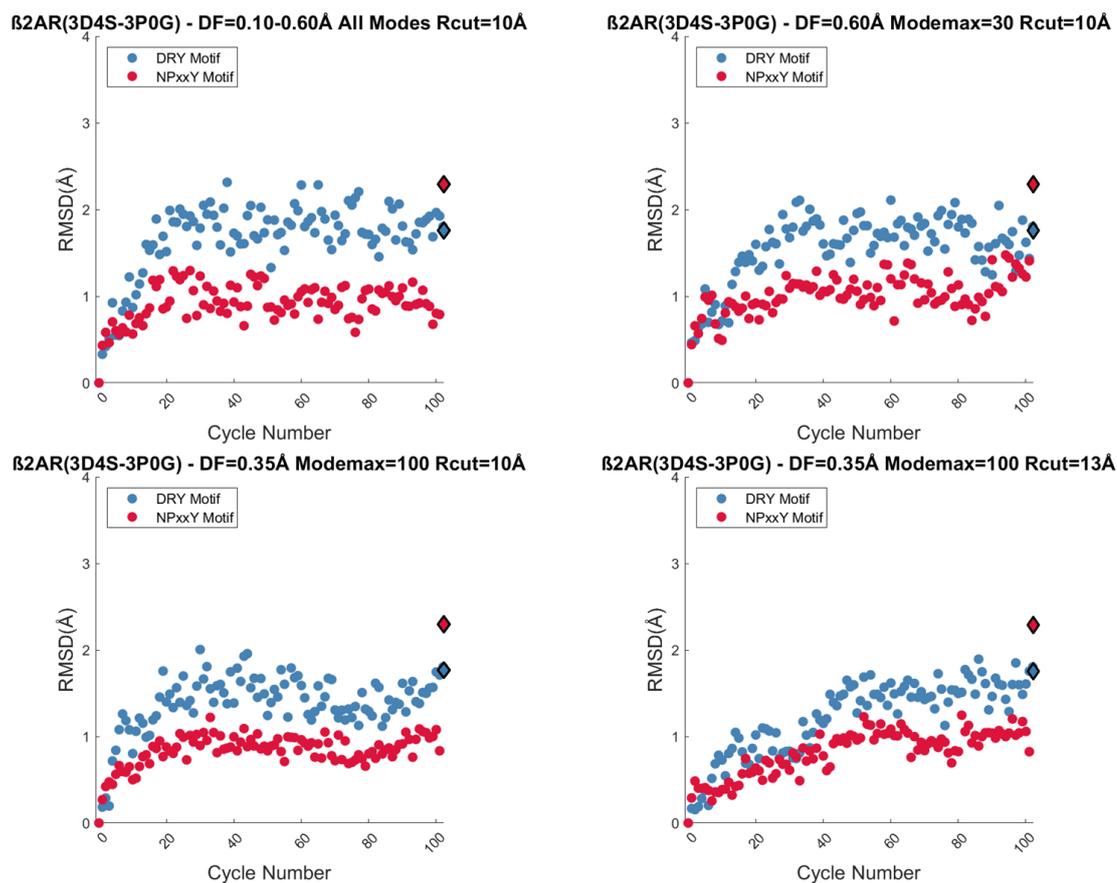


Figure 3.93. RMSD of important motifs through the activation of  $\beta 2$  adrenergic receptor from 3D4S to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres).

### ● Ligand Binding Site, Connector Region, and G-protein Binding Site

The RMSD of connector (I121), ligand binding site (S207), and G-protein binding site (Y219) with respect to initial structure is determined by using  $C\alpha$  atoms through the activation pathway [15]. The RMSD between the initial and active/target structures is 1.7 Å, 1.6 Å, and 2.0 Å for the connector, the ligand binding site, and G-protein binding site, respectively. The RMSD profile of connector, ligand binding site and G-protein binding site is shown in Figure 3.94.

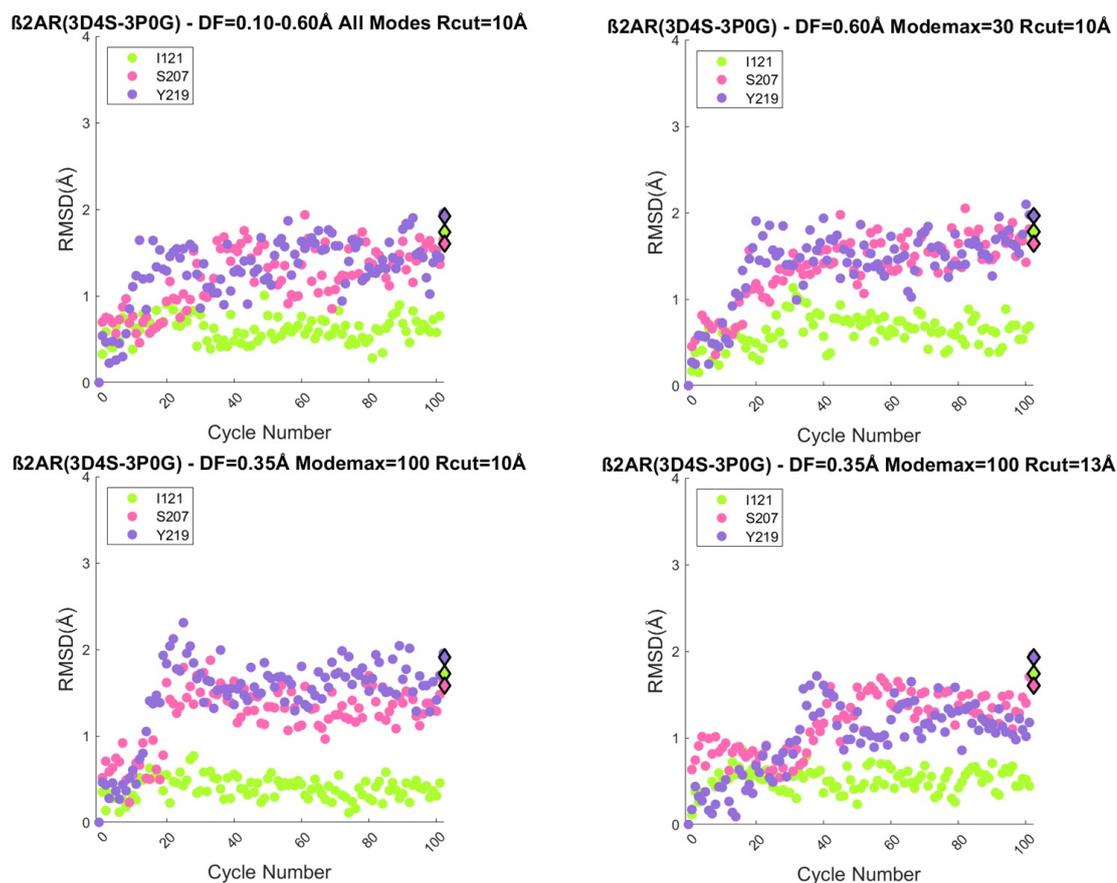


Figure 3.94. RMSD of important residues through the activation of  $\beta_2$  adrenergic receptor from 3D4S to 3P0G (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the last five snapshots of ANM-LD simulation with parameters  $DF=0.10-0.60 \text{ \AA}$  All Modes and  $R_{cut}=10 \text{ \AA}$  is considered, the RMSD is calculated as  $0.6 \text{ \AA}$ ,  $1.5 \text{ \AA}$ , and  $1.3 \text{ \AA}$  for the connector, the ligand binding site, and G-protein binding site, respectively. When RMSDs of important residues are analyzed, it is seen that only ligand binding site (S207) converge to the target structure.

In the simulation prepared with same inactive structure (3D4S) and different active structure (3P0G), it is seen the same results. The convergence of RMSDs of connector region and G-protein binding region does not be achieved. For this reason, the simulations in which 3D4S is used as inactive structure are not able to reflect the

activation pathway. In these two cases, while the ligand binding region reaches the target structure, the reason why the other regions cannot reach the target structure can be thought as cholesterol assumes the task of stabilizing the structural parts of  $\beta$ 2AR. In the study conducted by Zocher et al., the kinetic and mechanical properties of  $\beta$ 2AR in the presence and absence of cholesterol were examined. As a result, they concluded that the core segment where the ligand binding site is located is not affected by the presence of cholesterol [58]. Since there is a difference in ICL2 region which is one of the cholesterol binding sites of inactive structure (2RH1), it can affect the allosteric interactions of  $\beta$ 2-AR [59].

- Key Modes of Motion.

The most selected ANM modes with highest overlap values are examined to comprehend the basic system in the activation procedure. The analysis of ANM modes are performed according to procedure applied in the Section 3.1.1.1. The frequency of selected ANM modes are given in Figure 3.95.

Initial fifty cycles of parallel runs are researched to decide ANM mode recurrence, which is appeared in Figure 3.96. Considering the movements of important residues that show the activation procedure in the most ideal manner, ANM mode 1 and 2 are characterized as the most chosen modes with most noteworthy overlap values.

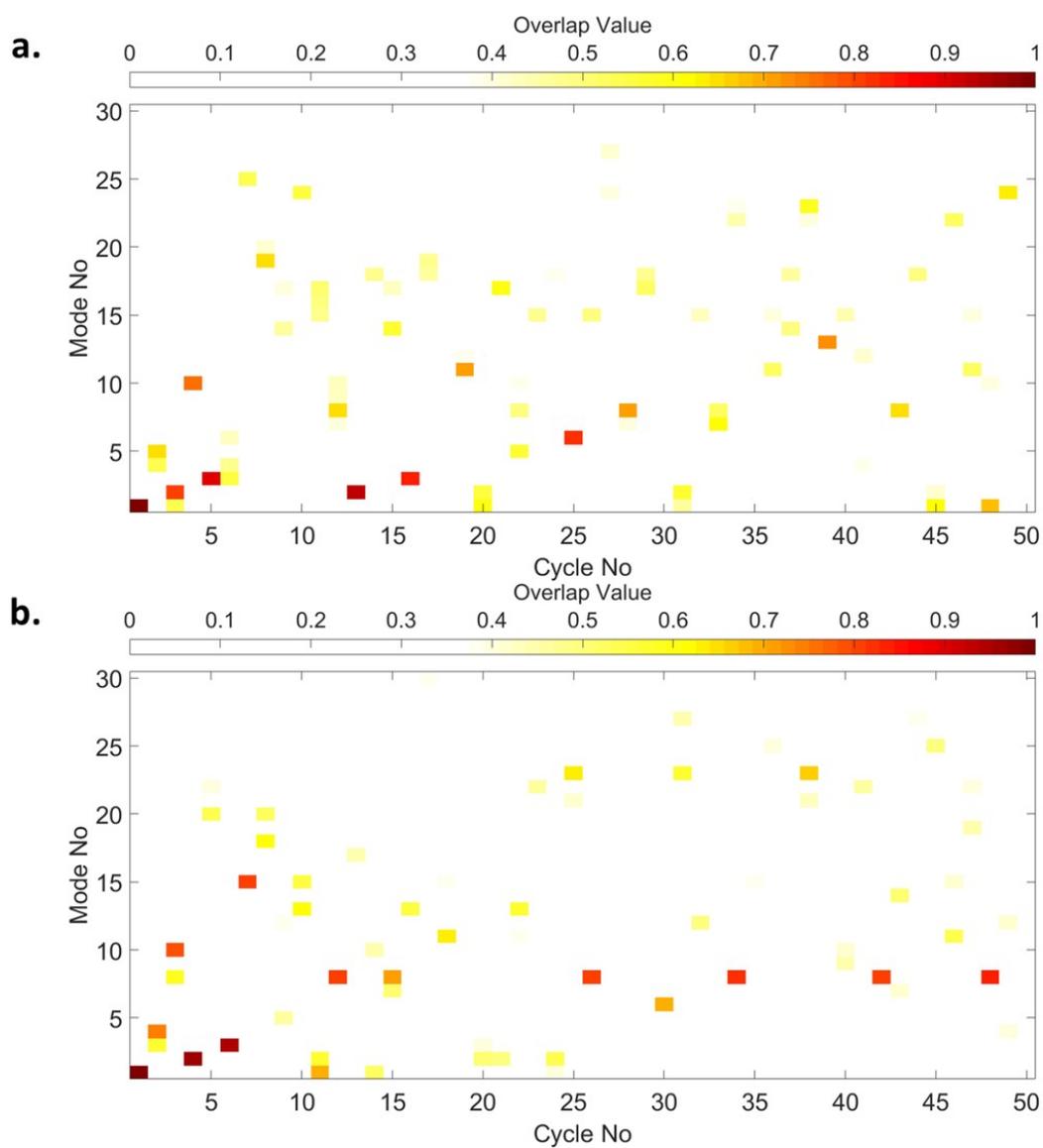


Figure 3.95. Selected Mode Mapping of  $\beta$ 2AR simulations a) DF=0.10-0.60 Å All Modes  $R_{cut}=10$  Å. b) DF=0.60 Å Modemax=30  $R_{cut}=10$ . c) DF=0.35 Å Modemax=100  $R_{cut}=10$  Å. d) DF=0.35 Å Modemax=100  $R_{cut}=13$  Å.

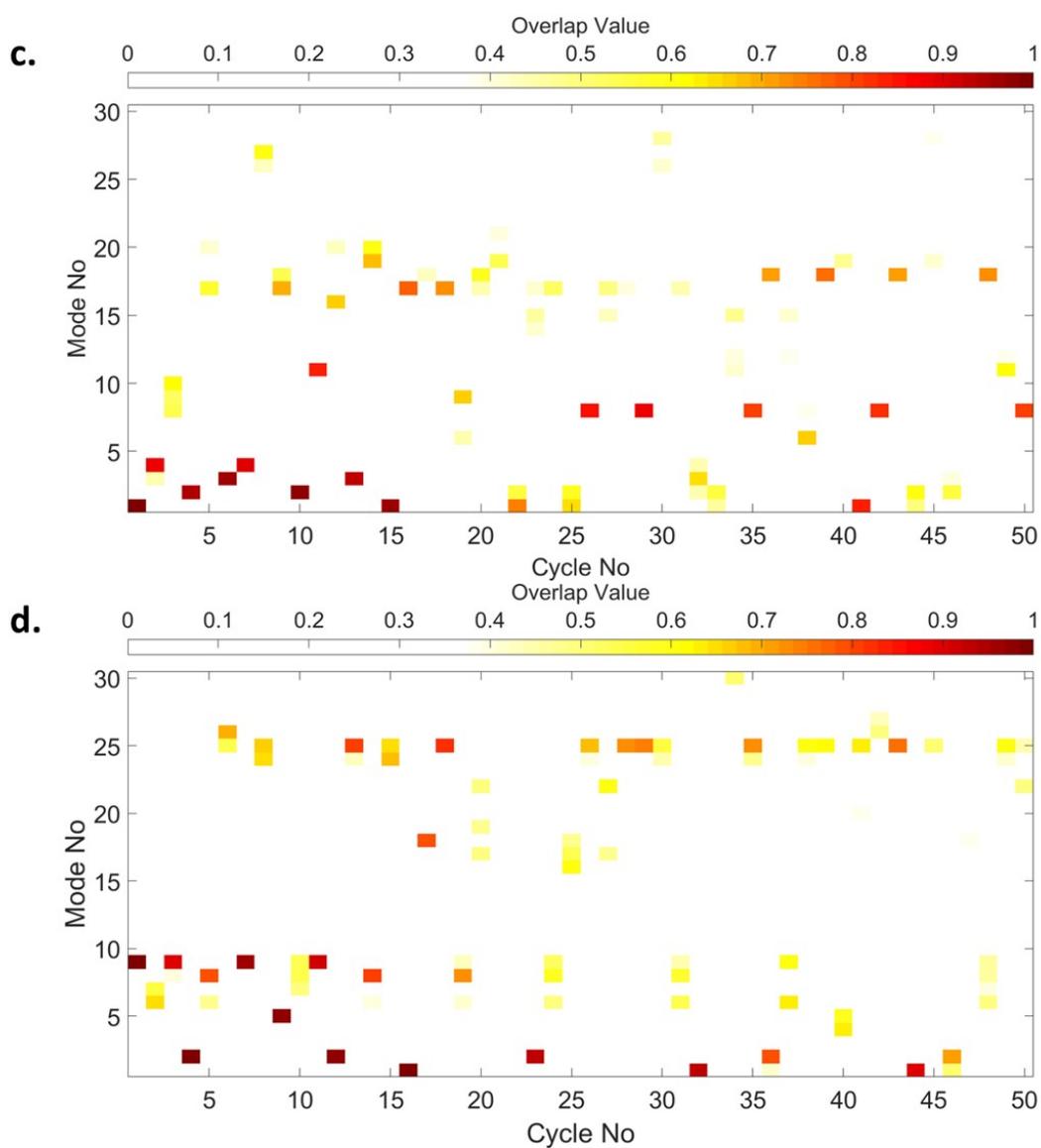


Figure 3.95. Selected Mode Mapping of  $\beta$ 2AR simulations a) DF=0.10-0.60 Å All Modes  $R_{cut}$ =10 Å. b) DF=0.60 Å Modemax=30  $R_{cut}$ =10. c) DF=0.35 Å Modemax=100  $R_{cut}$ =10 Å. d) DF=0.35 Å Modemax=100  $R_{cut}$ =13 Å. (cont.)

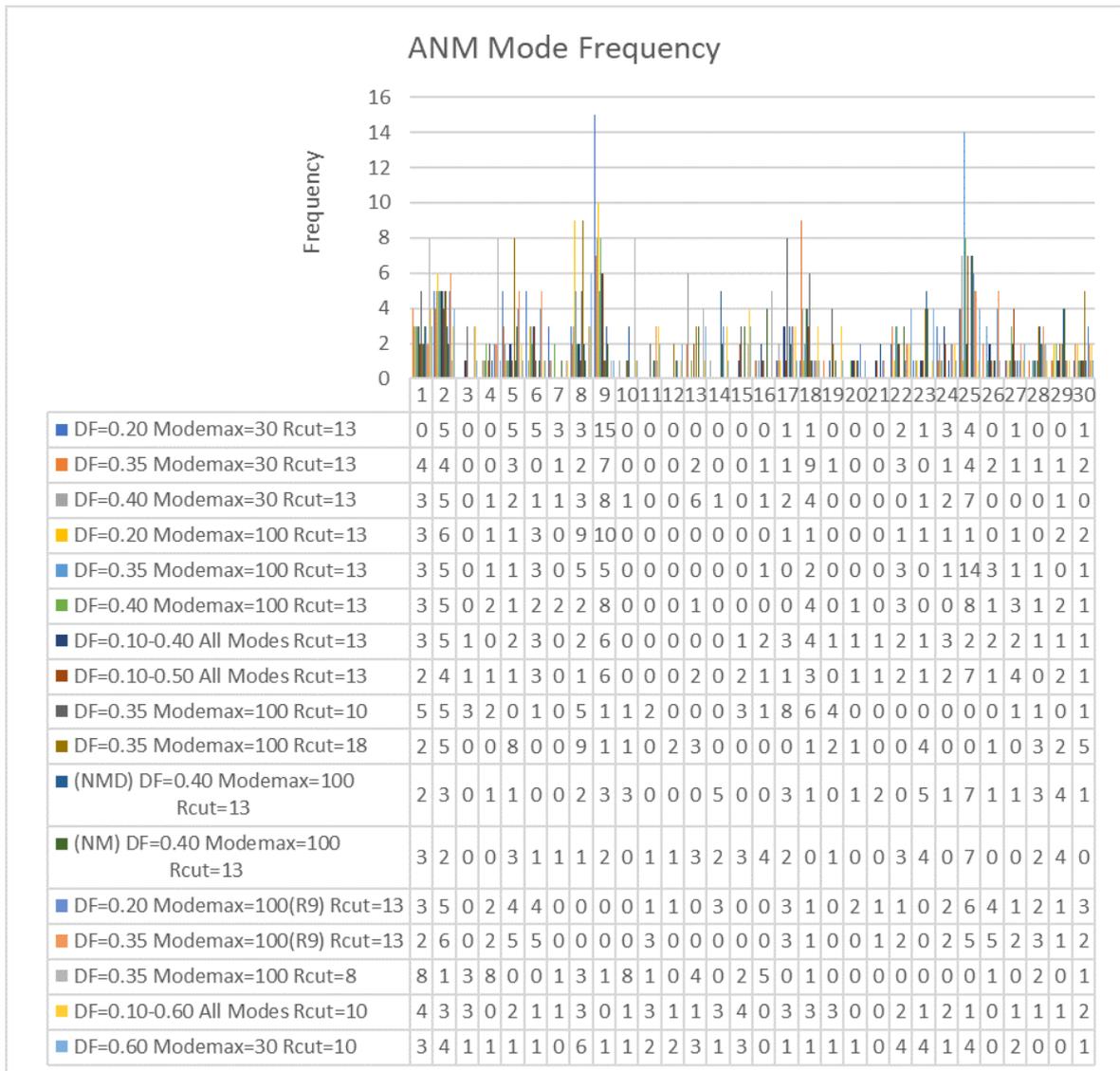


Figure 3.96. ANM mode frequency of parallel simulations from 3D4S to 3P0G.

Difference vector of initial (3D4S) and final (3P0G) is compared with the other cases (2RH1-3P0G and 3D4S-3SN6) and given in Figure 3.97 and Figure 3.98.

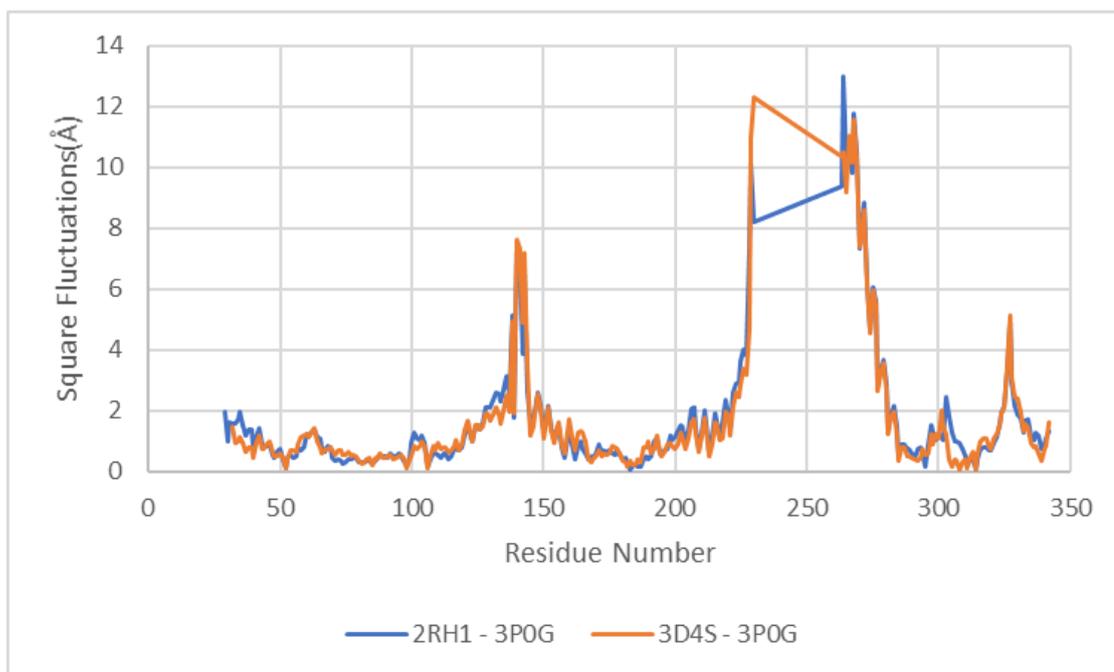


Figure 3.97. Difference in square vector distance between initial (2RH1-3D4S) and final (3P0G) structure.

When SF profile of the simulations from inactive structures (2RH1 and 3D4S) to active structure (3P0G) is compared, it can be understood from the profiles that the minimum and maximum points coincide with each other. There is a few distinctness in the SF profile between the residues H172-D192 in ECL2, and L302-I314 in ECL3 and extracellular side of TM7, since there is a minor shift in ECL3 in the inactive structure (2RH1) [51]. Secondly, MS profile of this case is compared with the case where the inactive structure (3D4S) used is the same but the active structure (3SN6) is different. It is observed that, the minimum and maximum points of MSF profile coincide with each other except the residues between V52-A78 at the end of TM1 and ICL1 and T177-E187 in the ECL2. Additionally, there are nine more missing residues in the active structure (3P0G) than other active structure (3SN6). MS of ANM modes 1 and 2 are shown in Figure 3.99.

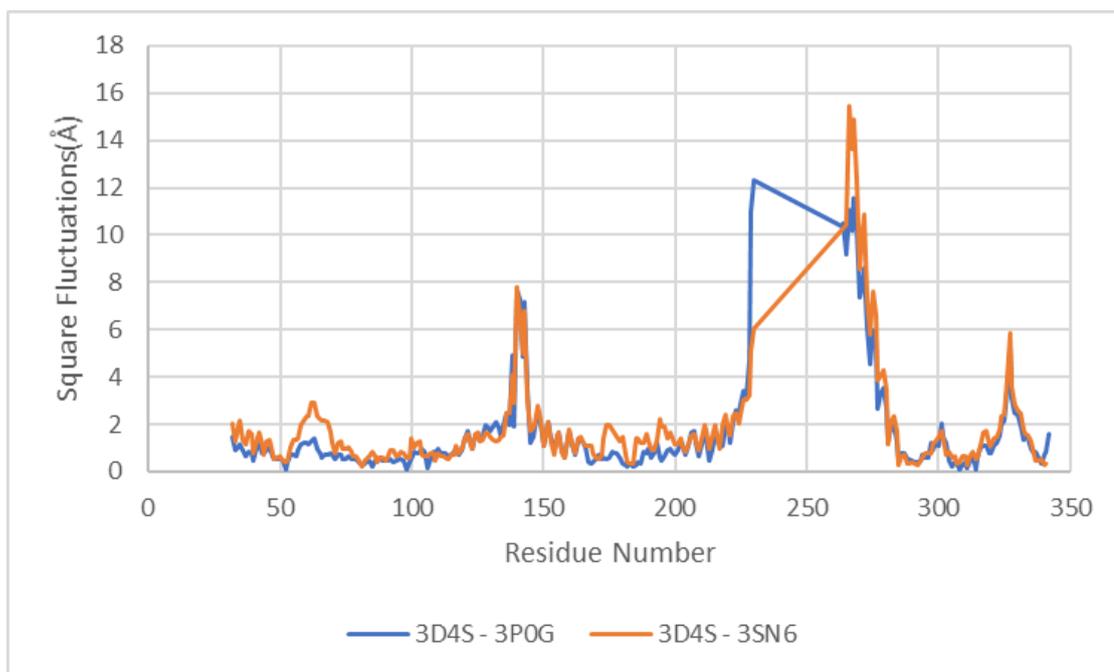


Figure 3.98. Difference vector between initial (3D4S) and final (3SN6 - 3P0G) structure.

Minimum points in the MSF profile of ANM mode are considered as hinge residues of this mode, which presumably regulate the motion. Hinge residues of ANM mode 1 are demonstrated in Figure 3.100 and determined as L45, T68, V86-V87, T110, I127, W158-I159, E180, D192, P211-I214, L284, K305, and N312-Y316. V87 in TM2 and L212 in TM5 are located in connector region and identified as allosteric residues in literature [56]. Additionally, Y316 located in orthosteric ligand binding site is determined as allosteric residue in intermediate conformation [56].

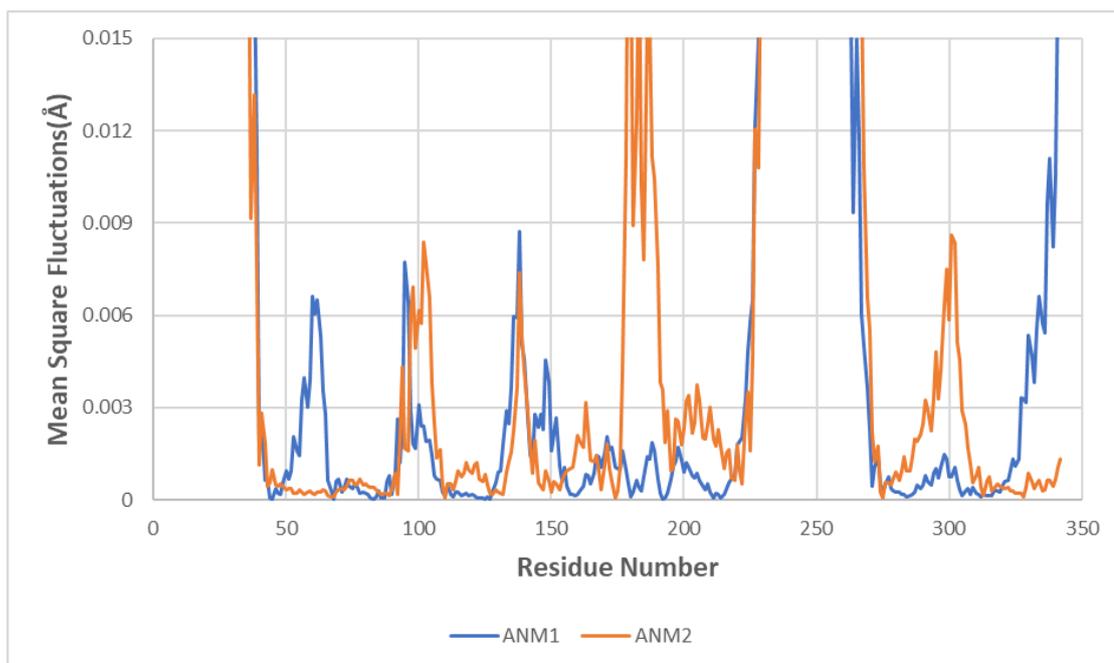


Figure 3.99. Mean square fluctuations of selected ANM modes of  $\beta$ 2AR simulations from 3D4S to 3P0G.

The hinge residues of ANM modes 2 is determined I43, V86-F89, A92, T110, A128, Y132, L147, I153, Y174, T195, V222, L275, W313, and R328 as demonstrated in Figure 3.101. Hinge residues are overlap with DRY motif orthosteric ligand binding site (T110) and connector region (V87 and F89). Additionally, T195 identified as hinge residue given as allosteric residue in extracellular loop in the literature [56].

Hinge residues of ANM mode 1 and 2 are demonstrated on the inactive structure of  $\beta$ 2 adrenergic receptor in Figure 3.102.

As in the distance and RMSD calculations of important residues, the hinges of the selected modes correspond to only the allosteric regions in the ligand binding site, which is not affected by presence and absence of cholesterol. This suggests that RMSD profile not reaching the target structure is due to a small difference in the cholesterol binding site between the two inactive structures [58, 59].

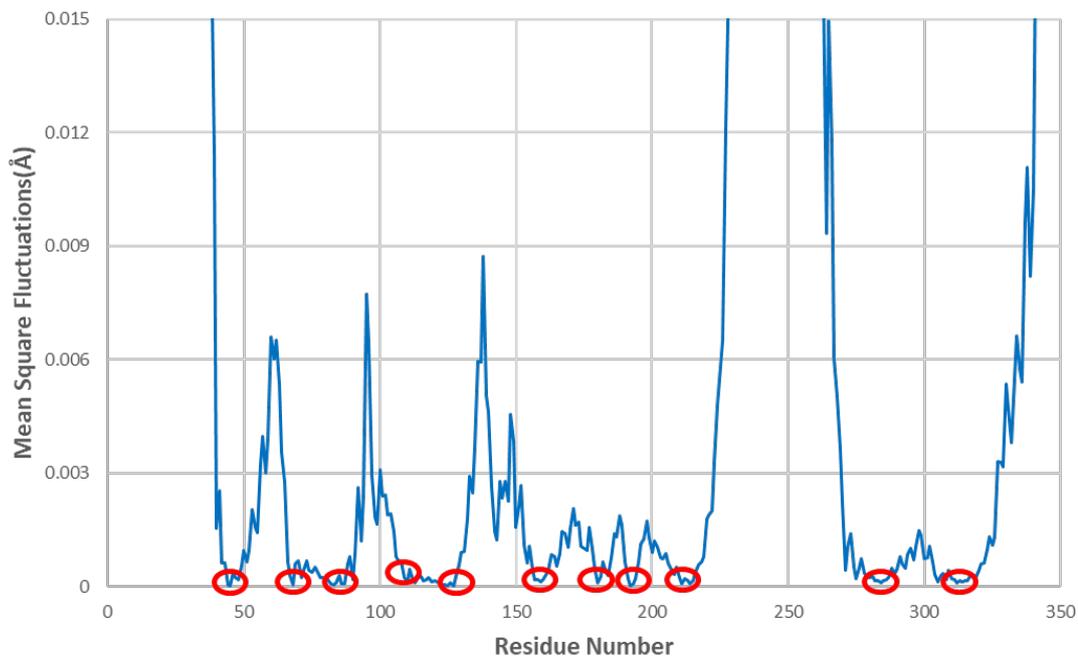


Figure 3.100. ANM mode 1 shape of the initial state of  $\beta 2$  adrenergic receptor.

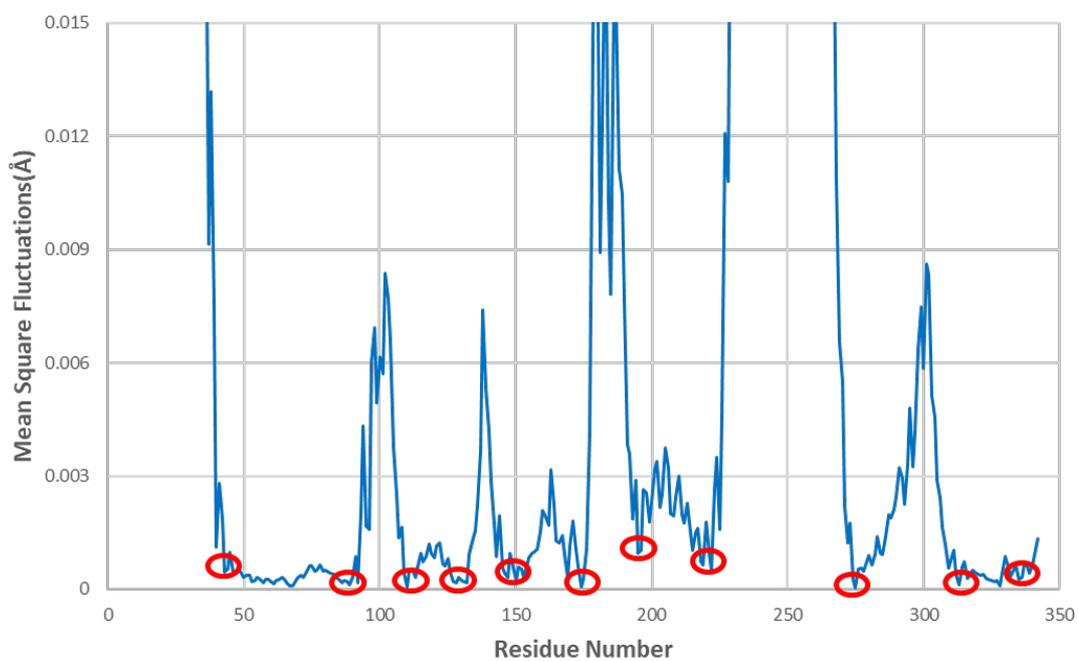


Figure 3.101. ANM mode 2 shape of the initial state of  $\beta 2$  adrenergic receptor.

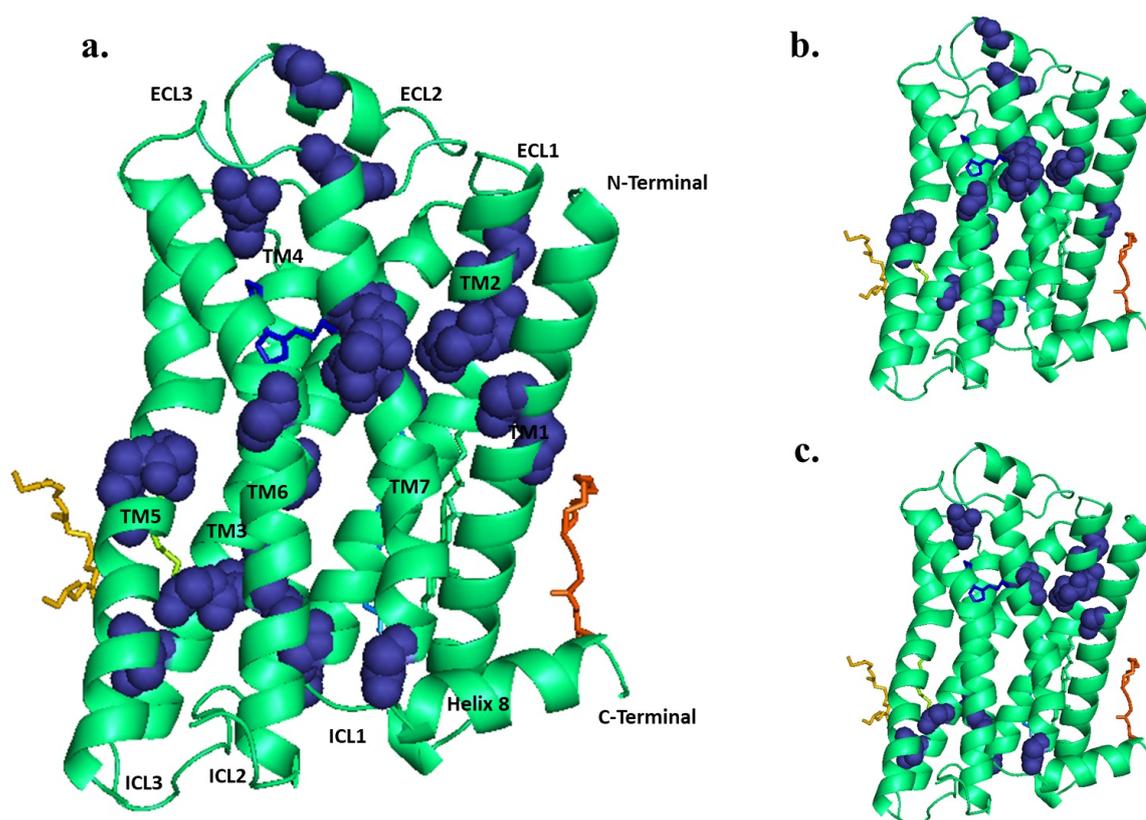


Figure 3.102. Demonstration of hinge residues of ANM mode 1 and 2 (dark blue and sphere) on the structure. a) Hinges of ANM mode 1 and 2. b) Hinges of ANM mode 1. c) Hinges of ANM mode 2.

The hinge residues of ANM mode 1 and 2 with functionally important residues (DRY Motif, NPxxY Motif, Connector, Ligand binding site, G-protein binding site, L272, and E268) are demonstrated on the inactive structure of  $\beta 2$  adrenergic receptor in Figure 3.103.

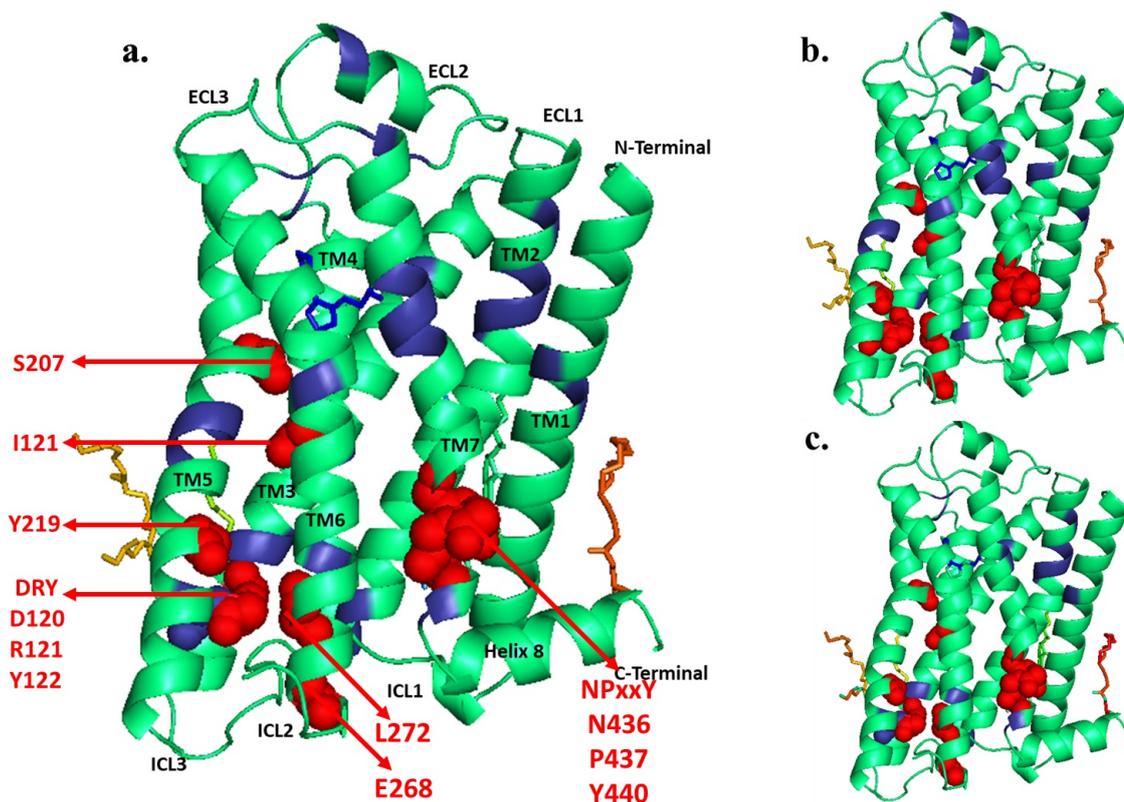


Figure 3.103. Demonstration of ANM mode 1 and 2 hinges (dark blue), functionally important regions (red and sphere), and the intersection (dark blue and sphere) on the structure. a) ANM mode 1 and 2 b) ANM mode 1 c) ANM mode 2.

- Cross-correlation Analysis.

Normalized cross-correlation analysis is used to identify the correlated residues. The analysis of cross-correlation is performed according to procedure applied in the Section 3.1.1.1. Normalized cross-correlation analysis is performed through the activation pathway of  $\beta_2$  adrenergic receptor and shown in Figure 3.104.

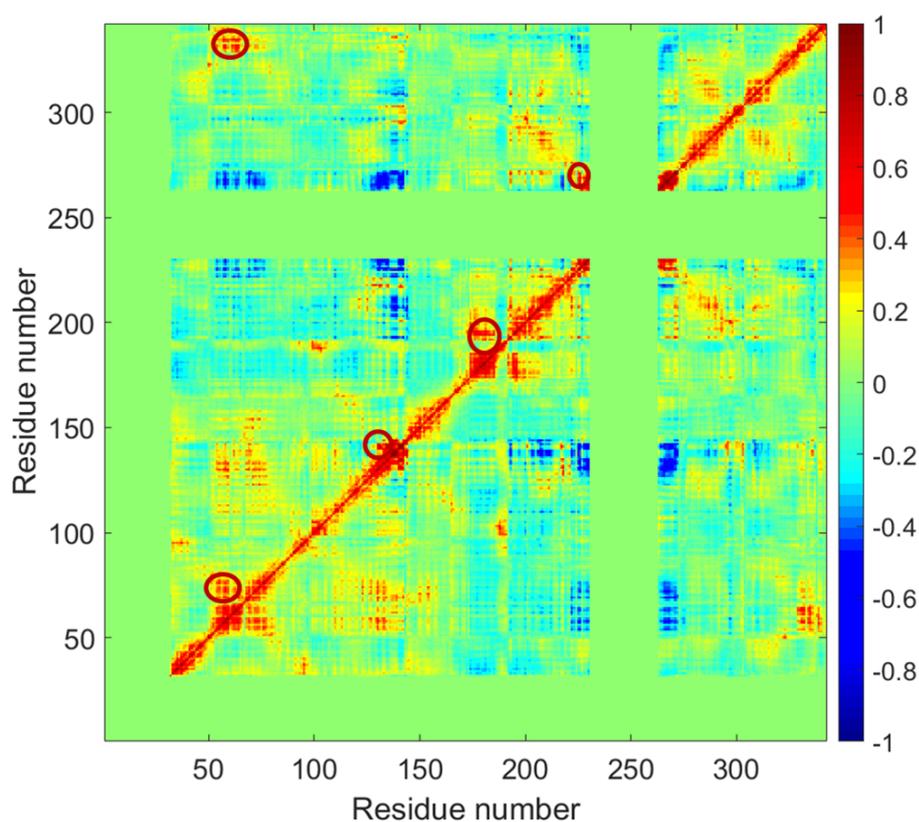


Figure 3.104. Cross correlation map of the activation pathway of  $\beta_2$  adrenergic receptor from 3D4S to 3SN6.

The highly correlated residues are determined as A176 in ECL2 & N196 at the beginning of TM5, A181 in ECL2 & F194 in ECL2, A226 in TM5 & H269 in TM6, and R131-A134 in TM3 & Q142 in ICL2. According to cross-correlation analysis, it is understood that DRY motif has effect on ICL2 region.

### 3.2.2. Activation Pathway of $\beta 2$ Adrenergic Receptor

The conformational changes in GPCR through the activation pathway permit sending extracellular signal across the cell membrane to create an intracellular reaction. For this reason, conformational changes are essential for GPCR activation. In this part of thesis,  $\beta 2$  adrenergic receptor ( $\beta 22AR$ ) shows the greatest conformational changes during activation, on the intracellular side (TM5-TM7) through rearrangements in the ligand binding pocket (TM5-TM6). An allosteric residue network located between TM3, TM5, TM6 and TM7 regions serves as a hub for detecting and responding to conformational change in  $\beta 22AR$  [15], Huang et al., 2015). Some snapshots generated during the ANM-LD simulation are examined and shown in Figures 3.105, 3.106, 3.107, and 3.108 to be able to understand the transition pathway from inactive structures (2RH1-3D4S) to active structures (3SN6-3P0G).

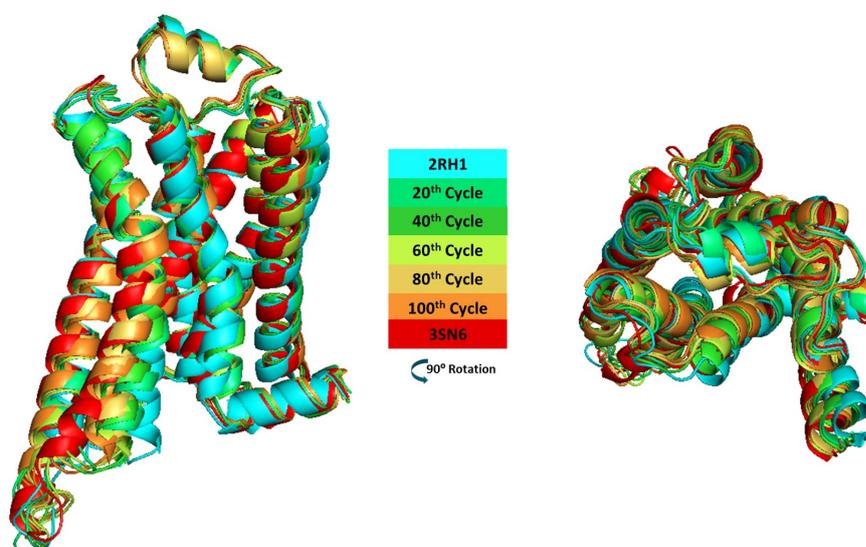


Figure 3.105. Snapshots generated by ANM-LD simulations from inactive (2RH1) to active (3SN6) structure.

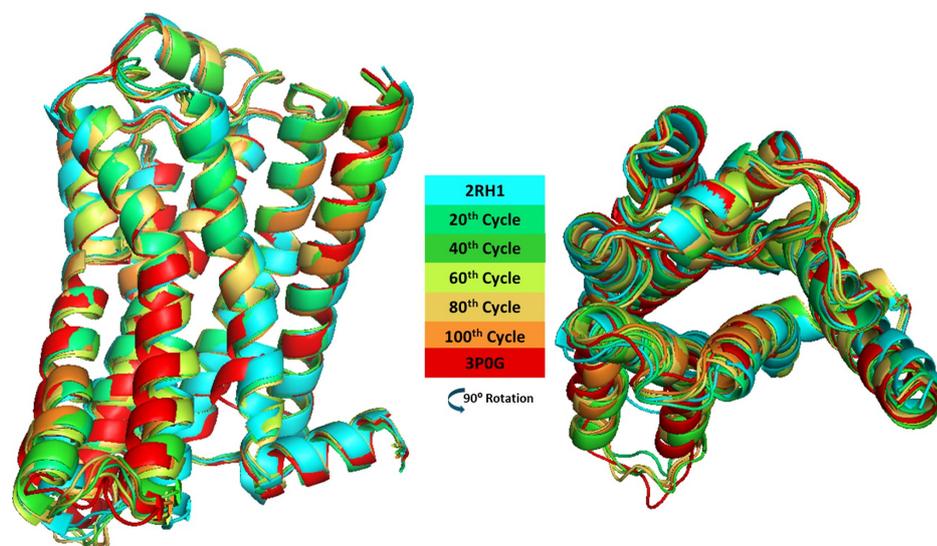


Figure 3.106. Snapshots generated by ANM-LD simulations from inactive (2RH1) to active (3P0G) structure.

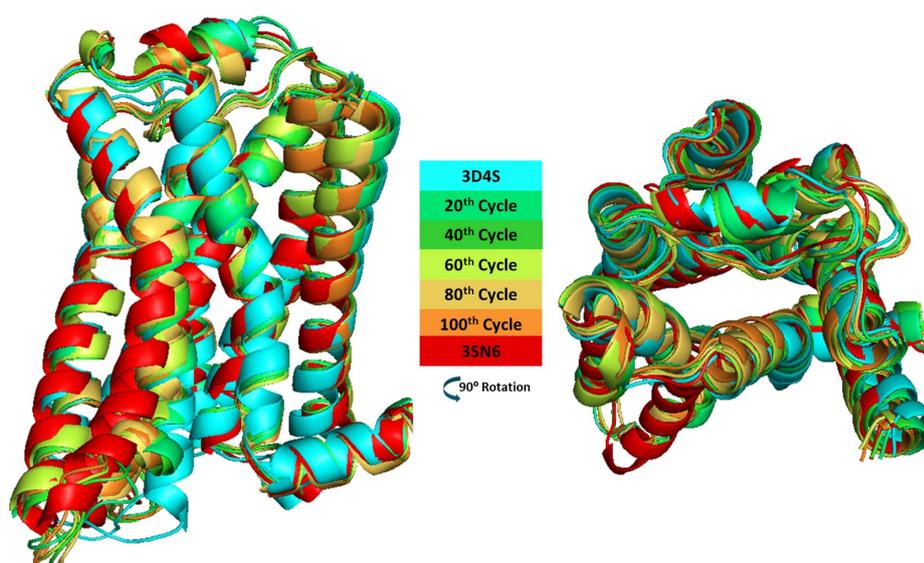


Figure 3.107. Snapshots generated by ANM-LD simulations from inactive (3D4S) to active (3SN6) structure.

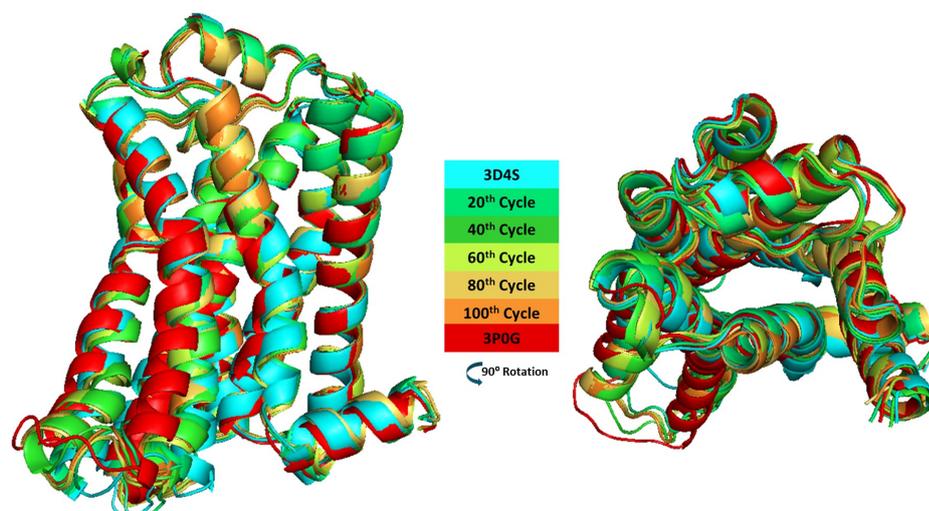


Figure 3.108. Snapshots generated by ANM-LD simulations from inactive (3D4S) to active (3P0G) structure.

The intermediate conformations in between initial and target/final structures obtained by ANM-LD simulations are investigated through the activation pathway. In the intermediate structure proposed in the literature by Dror et al. via accelerated MD simulations, TM3, TM5, and TM6 helices show similar conformation with the active structure, however TM7 motion closes the gap between intracellular ends of TM2, TM3, TM6, and TM7 [15]. When the TM3-TM6 distances calculated between R131 and L272 using  $C\alpha$  atoms of intermediate structure is analyzed, the distance scatters between 12 Å and 16 Å. Additionally, the RMSD profile of NPxxY motif with respect to inactive structure is analyzed and it is found that the NPxxY region RMSD between 0.2 Å and 0.7 Å [15]. The important residues on the snapshots obtained from the simulations from inactive structure (2RH1) to active structure (3SN6), TM3-TM6 and TM3-TM7 distances are demonstrated in Figures 3.109, 3.110 and 3.111 to analyze if the intermediate structures observed in the study of Dror et al., are obtained through the ANM-LD simulation [15].

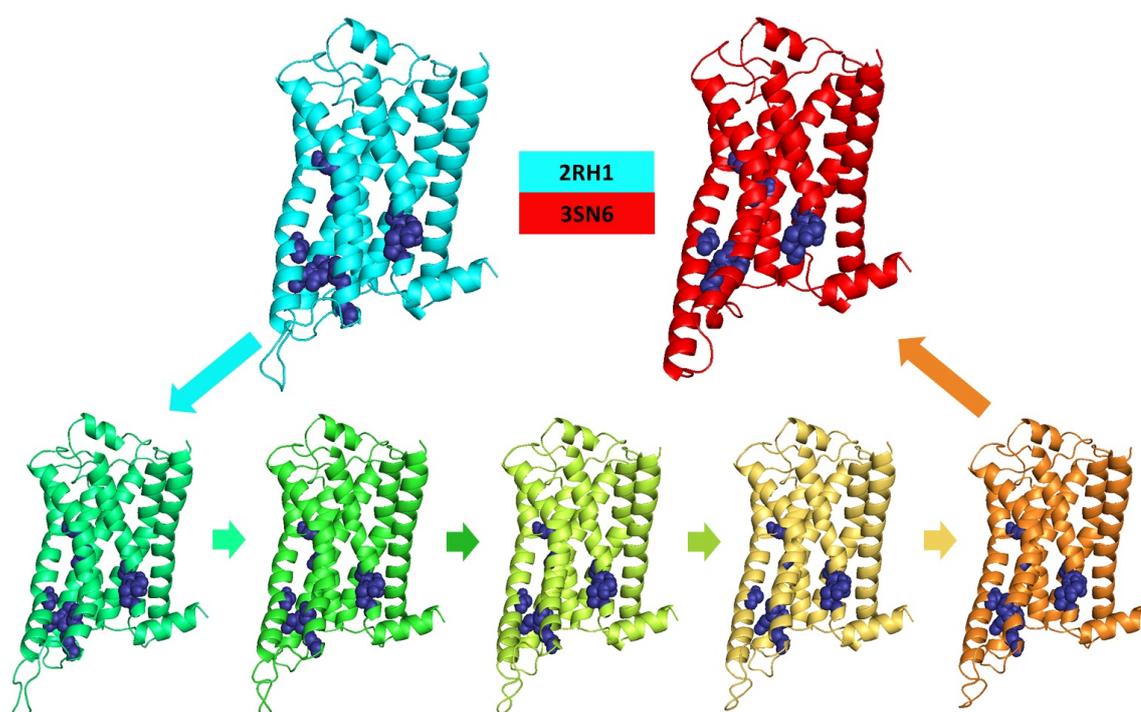


Figure 3.109. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3SN6.

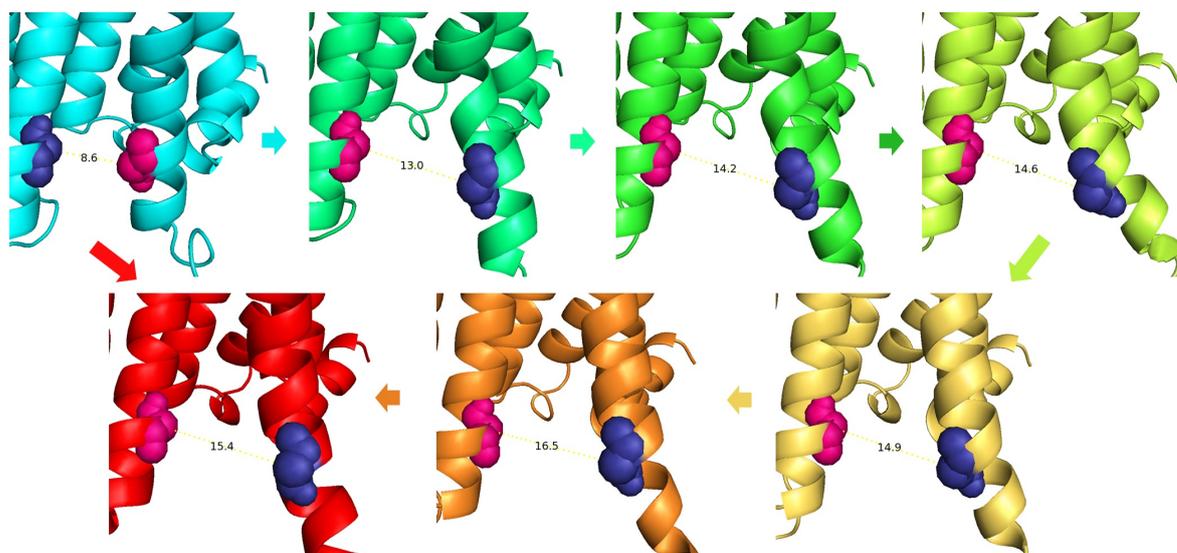


Figure 3.110. The TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3SN6

When the collective variables are examined, it is observed that the intermediate structure appears in the first 20 snapshots. According to the RMSD of important motif calculations performed in Section 3.2.1.1, the RMSD of the NPxxY motif is between 0.2 Å and 0.7 Å in the first 20 cycles. In the 20<sup>th</sup> cycle, the TM3-TM6 distance is calculated as 13.0 Å. In the 60<sup>th</sup> and 80<sup>th</sup> cycles of the simulation, The TM3-TM6 and TM3-TM7 distances remain approximately constant, while the TM5 continues to move outward. In the 100<sup>th</sup> cycle, The TM3-TM6 distance is 16.5 Å and the TM3-TM7 distance is 13.3 Å.

The ANM-LD simulation of  $\beta$ 2AR from inactive structure (2RH1) to other active structure (3P0G) is investigated to observe if there is an alternative transition pathway. The important residues on the snapshots, TM3-TM6 and TM3-TM7 distances are demonstrated in Figures 3.112, 3.113 and 3.114.

Unlike the simulation result from inactive structure (2RH1) to active structure (3SN6), the intermediate structure is observed here in the 60<sup>th</sup> cycle. For this reason, it has been observed that the transition pathways of the two cases differ. In the first

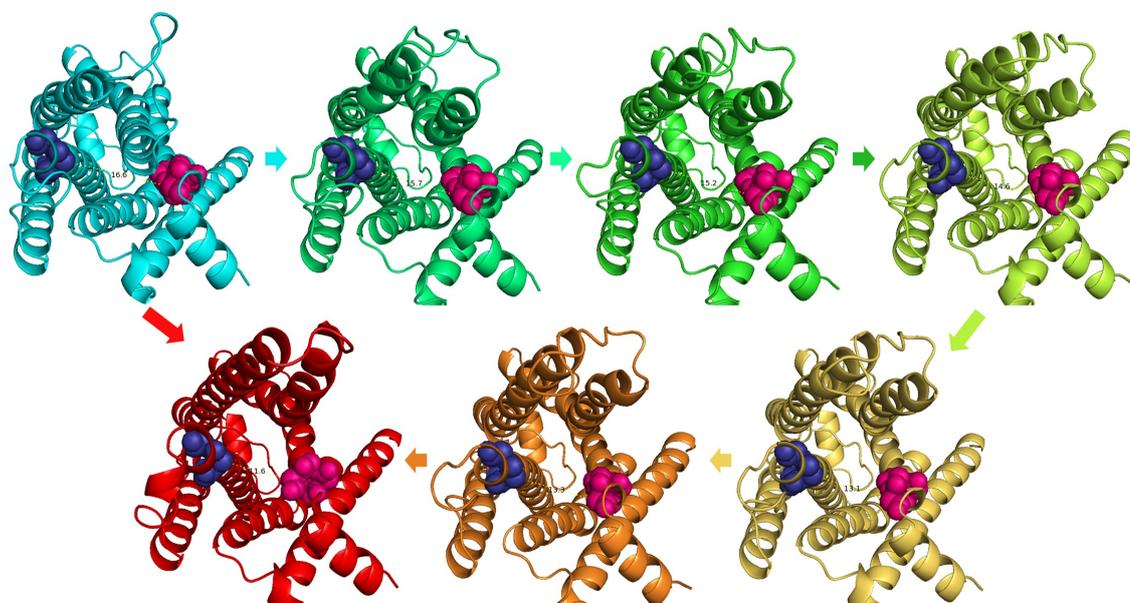


Figure 3.111. The TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3SN6

case (2RH1-3SN6), the TM3-TM6 distance until 40<sup>th</sup> cycle, then the distance between the intracellular sides of TM3 and TM6 stabilizes between the 40<sup>th</sup> and 80<sup>th</sup> cycles. In this case, the TM3-TM6 distance is achieved to 12.9 Å in the 60<sup>th</sup> cycle. These intermediate structures show similarities with the intermediate structure proposed by Dror et al. according to collective variable analysis [15].

The ANM-LD simulation of  $\beta$ 2AR from the different inactive structure (3D4S) to active structures (3D4S-3P0G) is investigated to observe the transition pathway along the activation process. The important residues on the snapshots, TM3-TM6 and TM3-TM7 distances are demonstrated in in Figures 3.115, 3.116, 3.117, 3.118, 3.119 and 3.120.

When the snapshots obtained from ANM-LD simulations from inactive structure (3D4S) to active structure (3SN6-3P0G) are investigated in the manner of collective variables, it is seen that the activation process is not achieved. While it is observed that the intermediate structure is reached in the simulation using 3SN6 as the active

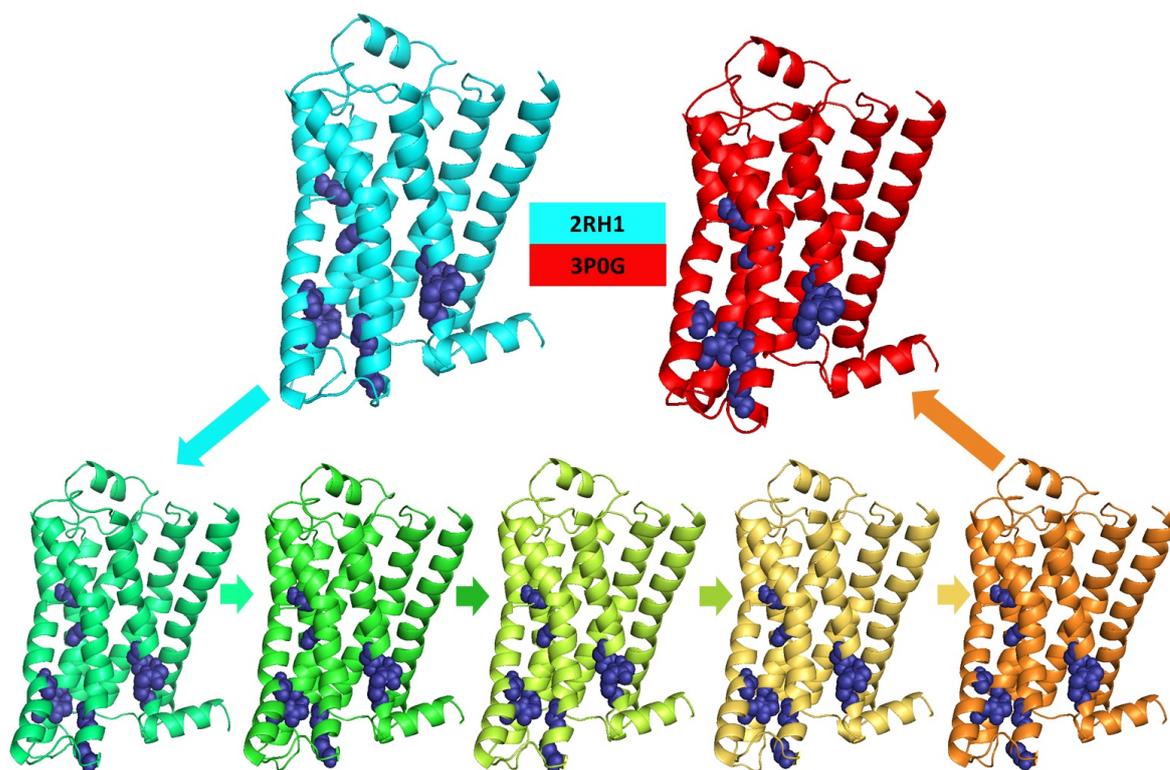


Figure 3.112. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3P0G.

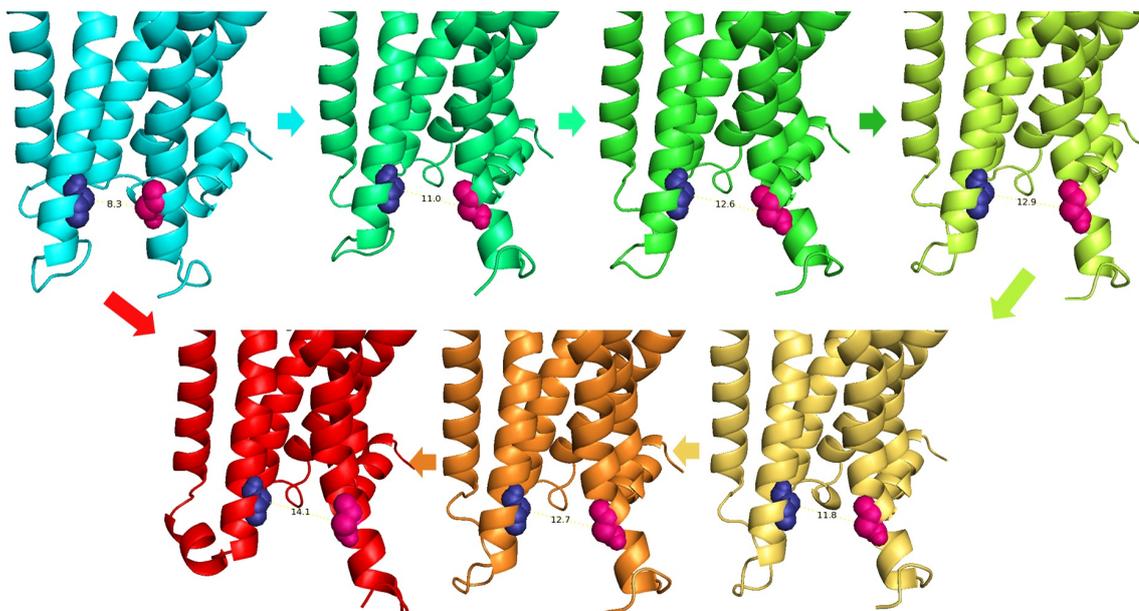


Figure 3.113. TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3P0G.

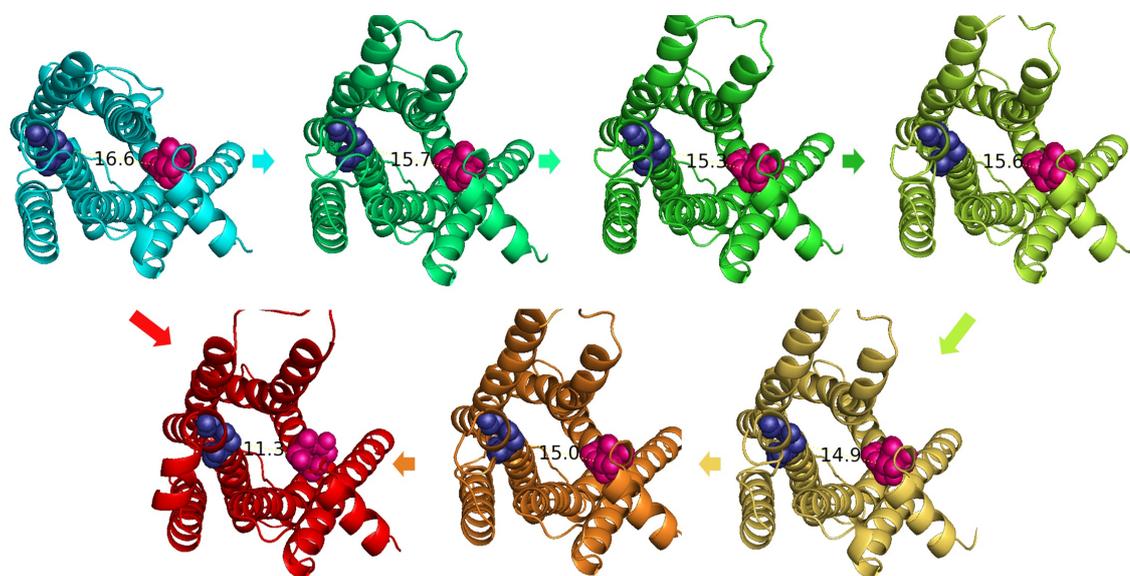


Figure 3.114. TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 2RH1 to 3P0G.

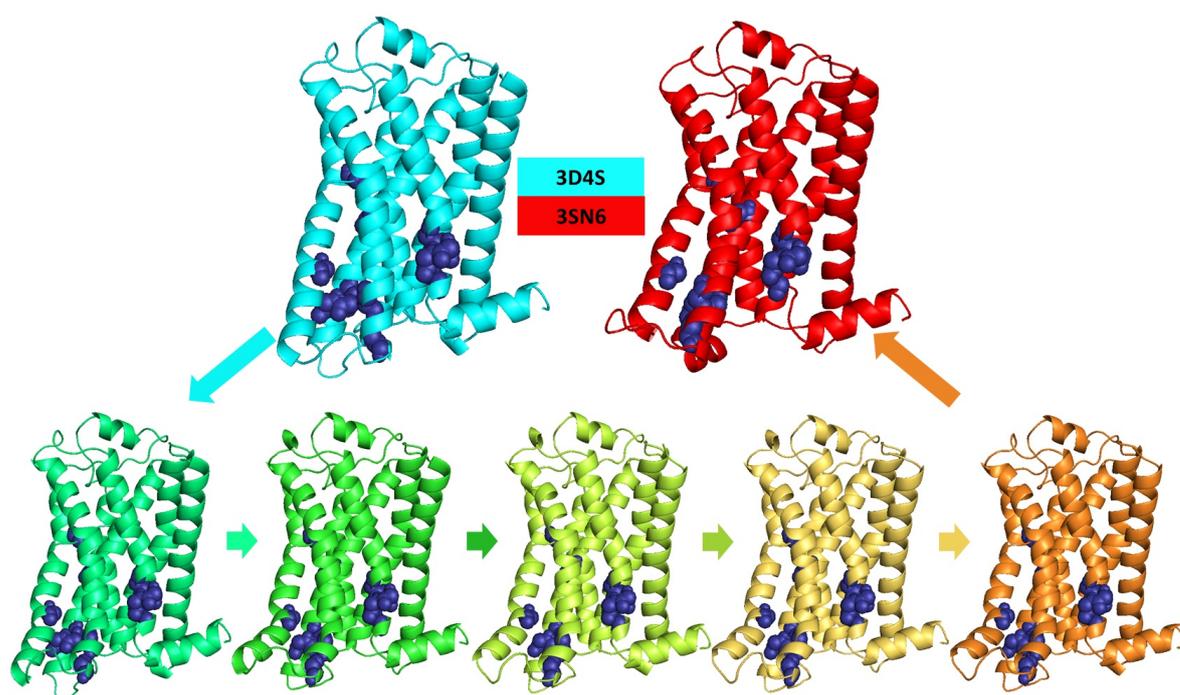


Figure 3.115. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6.

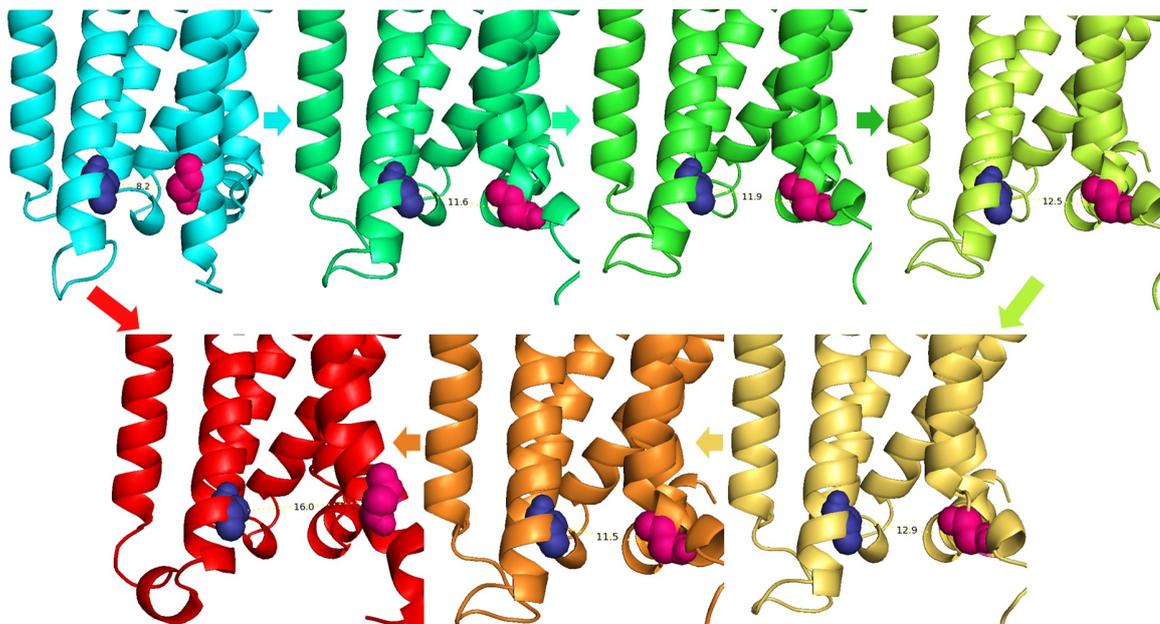


Figure 3.116. TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6.

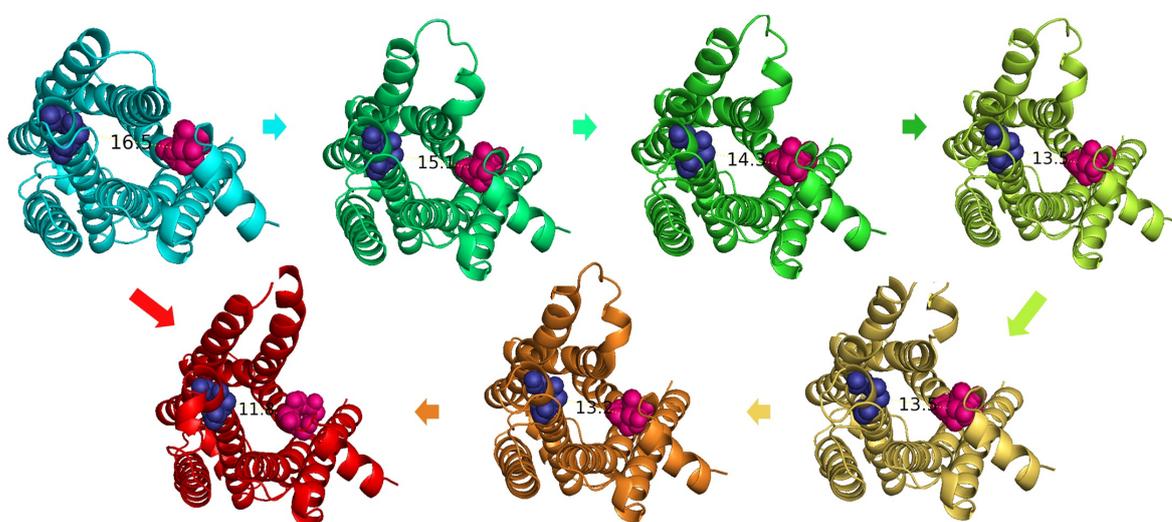


Figure 3.117. TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6.

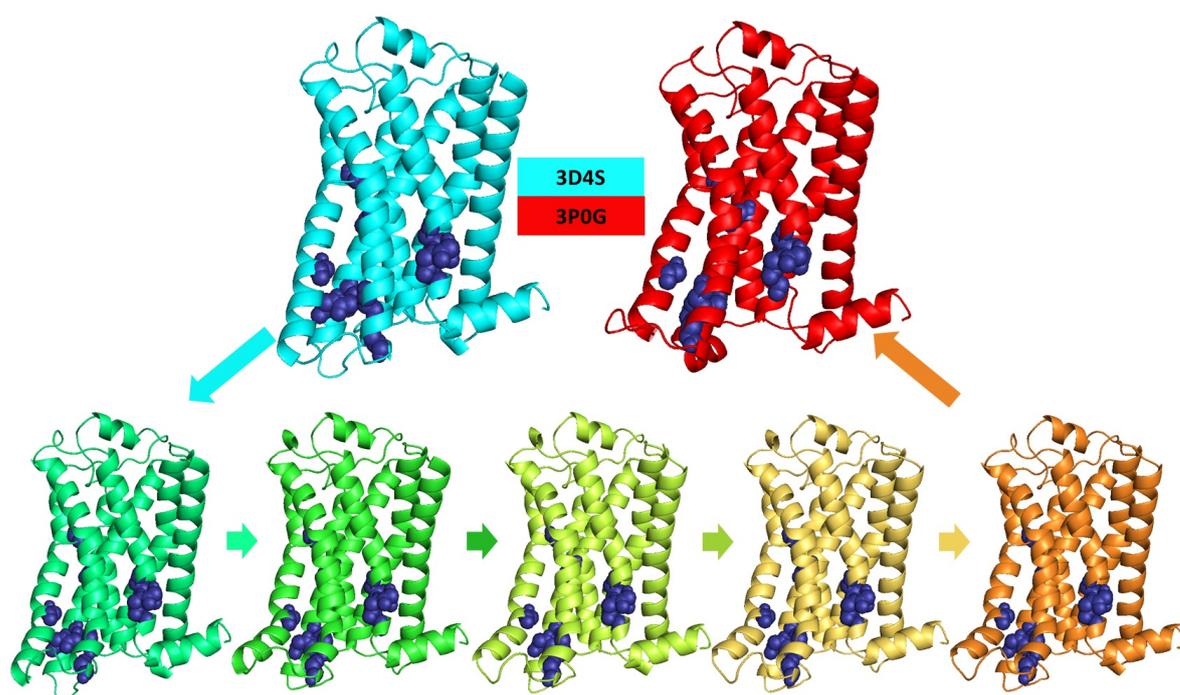


Figure 3.118. Important residues (DRY motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L272 at the end of TM6) (dark blue and sphere) demonstrated on the snapshots of simulations from 3D4S to 3P0G.

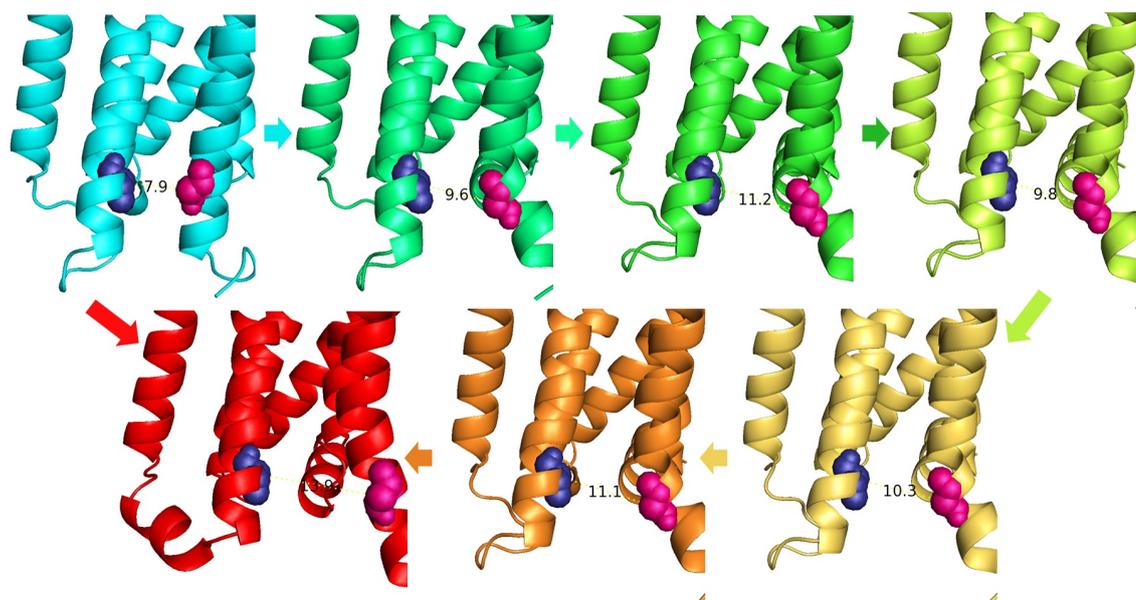


Figure 3.119. TM3-TM6 distance between R131 (dark blue and sphere) and L272 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3SN6.

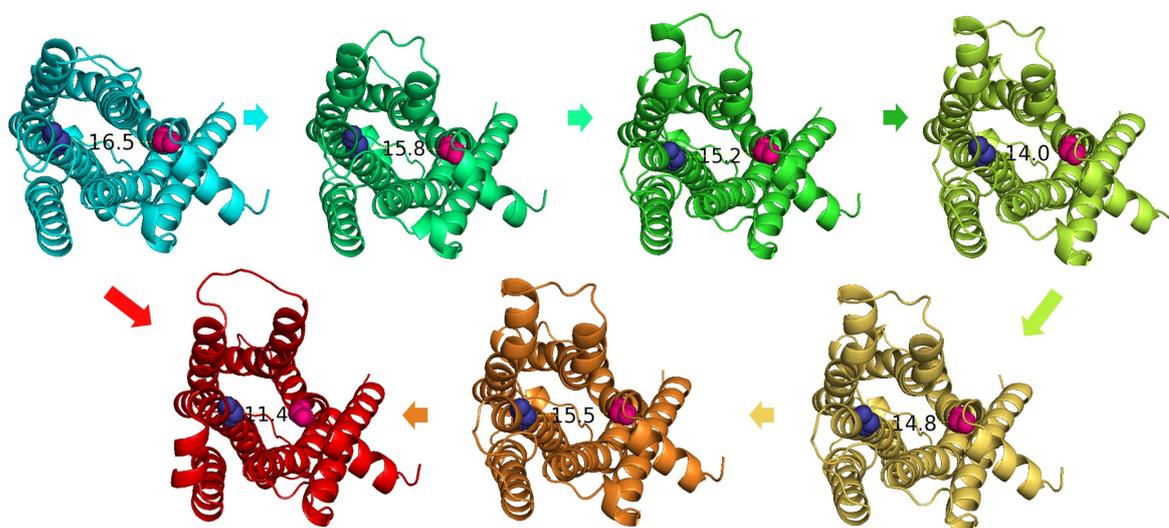


Figure 3.120. TM3-TM7 distance between R131 (dark blue and sphere) and Y326 (pink and sphere) demonstrated on the snapshots of simulations from 3D4S to 3P0G.

structure, even the intermediate structure is not reached in the simulations where the active structure is used as 3P0G. As explained in Section 3.2.1.4, it is thought that the difference in the ICL2 loop with respect to other inactive structure (2RH1), where the cholesterol binding site in 3D4S is located, has a great role in the conformational movements of the structure [58,59].

The intermediate structure obtained from the ANM-LD simulations of  $\beta$ 2AR of two cases (2RH1-3SN6 and 2RH1-3P0G) in which the activation process is fully achieved according to collective variables are demonstrated in Figure 3.121.

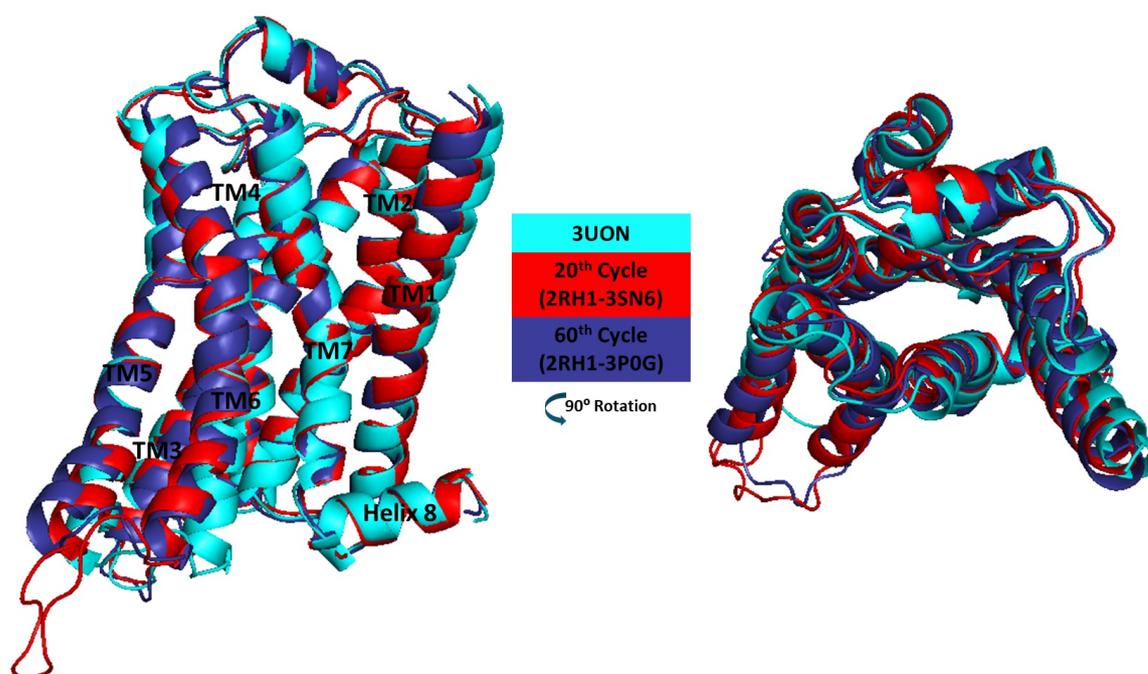


Figure 3.121. Alignment of intermediate structures obtained by ANM-LD simulations of the simulations in which 2RH1 is used as inactive structure.

The RMSD of intermediate structure of the simulation from 2RH1 to 3SN6 with respect to the intermediate structure of the simulation from 2RH1 to 3P0G is calculated as 1.1 Å. The main distinctness between two intermediate structure is shown in the

intracellular side of TM5. In the 60<sup>th</sup> cycle of the simulation from 2RH1 to 3P0G, the outward shifting of intracellular side of TM5 is more obvious. While ECL2 loop of the snapshot in 20<sup>th</sup> cycle of the simulation from 2RH1 to 3SN6 shows similar conformational position with the active structure, it is observed that the snapshot in 60<sup>th</sup> cycle is exactly same as the inactive structure according to conformational position of ECL2 loop.

The final structures obtained from the ANM-LD simulations of  $\beta$ 2AR of two cases (2RH1-3SN6 and 2RH1-3P0G) in which the activation process is fully achieved according to collective variables are demonstrated in Figure 3.122.

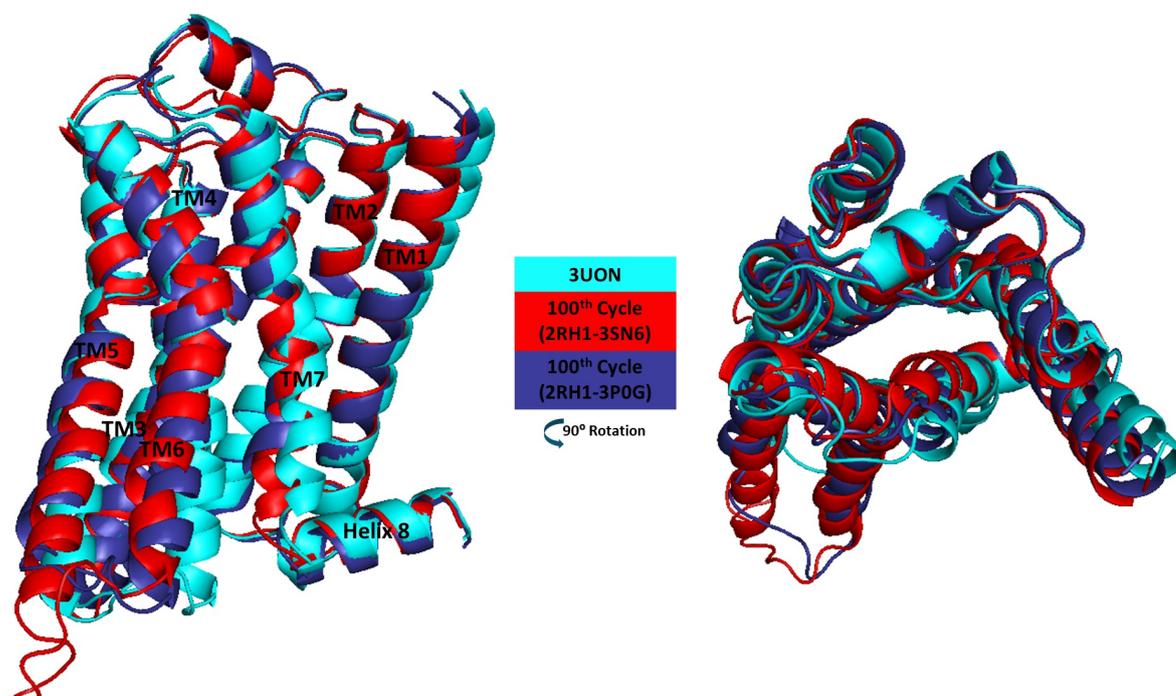


Figure 3.122. Alignment of final structures obtained by ANM-LD simulations of the simulations in which 2RH1 is used as inactive structure.

The RMSD of the final structure of the simulation from 2RH1 to 3SN6 with respect to the final structure of the simulation from 2RH1 to 3P0G is calculated as 1.6 Å. The main difference between two structures detected in the intracellular side of

TM5, DRY motif, intracellular side of TM1 and TM7. The difference in conformational position of DRY motif is expected, since R131 in  $\beta$ 2AR-Nb80 reacts with the nanobody [49].

### 3.2.3. Allosteric Interactions in $\beta$ 2 Adrenergic Receptor

An allosteric residue network located between TM3, TM5, TM6 and TM7 regions serves as a hub for detecting and responding to conformational change in  $\beta$ 2AR [15], [16]). The binding of an agonist to the ligand binding pocket triggers the binding of the intracellular G protein. Conversely, G protein binding will also enhance agonist binding; that is, an agonist actively binds to a GPCR with higher affinity. On the other hand, MD simulations show that the intracellular pairing interface is in equilibrium between active and inactive conformations even in the absence of the agonist. The presence of a bonded agonist helps stabilize the active conformation of the binding pocket, thereby affecting the intracellular junction interface to adopt active conformations. In order to understand the activation mechanisms of GPCRs and the mechanical bases of allosteric changes / interactions, the hinge residues of the ANM modes selected in this section of the thesis are examined, and then residues with high correlation with functionally important sites are determined by cross correlation analysis.

In the investigation of allosteric interactions of the ANM-LD simulation from 2RH1 to 3SN6, it is observed that F193 located in the extracellular loop (ECL2) determined as the most important allosteric region is defined as the hinge residue [56]. T110 in orthosteric binding site, F282 in connector region, N322 in NPxxY motif, the residues between F133-I135 which plays an important role in the motion of DRY motif, an Y308 in TM7 which belongs to the allosteric site in extracellular loop are determined as hinge residues of selected ANM modes [55,56]. According to analysis of cross correlation, NPxxY motif and Y308 are correlated residues, which are given as allosteric sites in the literature [56].

When allosteric interactions of the ANM-LD simulation from 2RH1 to 3P0G are investigated, DRY motif, NPxxY motif, orthosteric ligand binding site (S207), and G-protein binding site (Y219), Y308 in TM7 and I154 in TM4 which are the proposed as allosteric residues in literature also overlap with the hinge residues. At the point when the hinge residues for this case are contrasted with the hinge residues obtained from ANM-LD simulations from inactive structure (2RH1) to active structure (3SN6), the residues in the proposed allosteric sites in literature (intracellular loop (ICL3) and connector region) can not be identified [55, 56].

The ANM-LD simulations from inactive structure (3D4S) to active structures (3SN6-3P0G) unable to mirror the activation pathway of  $\beta$ 2AR. When the selected ANM modes are investigated to identify the hinge residues, it is seen that the allosteric residues (V87-F89, T110, I121, C285, Y316, T195) located in allosteric sites in connector region and ligand binding site overlap with the hinge residues [55, 56]. In the simulation from 3D4S to 3SN6, F193 identified as hinge residue overlap with the proposed allosteric site in extracellular loop. However, in the simulation from 3D4S to 3P0G, hinge residues that can be found with the allosteric side in the extracellular part are not detected. The reason for this situation, the allosteric regions in ligand binding site that is not affected by presence and absence of cholesterol. Although there are cholesterol molecules in the crystal structures of two inactive structures, 3D4S differs from the cholesterol binding site 2RH1 in its inactive structure. This suggests that ANM modes not detecting important allosteric regions for this case is due to a small difference in the cholesterol binding site between the two inactive structures [58, 59].

### 3.3. Neurotensin Receptor 1

Neurotensin (NTS), a short peptide consisted of 13-aminoacid, performs as a modulator in the nervous system and surrounding tissues [60]. NTS modulators have been proposed to use for curation of drug misuse, Parkinson's disease, schizophrenia, dietary issues, and pain [60–64]. Additionally, NTS-related ligands are promising drugs to fight with tumors, since NTS is used as growth factor on some cancer cells [60, 63–66].

There are three different subtypes of NTS receptor (NTSR1, NTSR2, and NTSR3) [60, 62]. While NTSR1 and NTSR2 are member of class A of GPCR family, NTSR3 belongs to sortilin family [60]. A large portion of the known impacts of NTS are intervened through NTSR1, which acts as a modulator in signaling to preferential tissue via G-protein. Many of the impacts of neurotensin are intervened through NTR1, activation of adenylyl cyclase and phospholipase C by providing GDP/GTP exchange in heterotrimeric G proteins [67]. Although, researches about ligand binding to NTSR1 focus on the mutagenesis and modelling, the subtleties of ligand restricting remain ineffectively comprehended at the sub-atomic level [60]. For this reason, the activation process of NTS1 receptors is investigated in this part of thesis to reveal the conformational changes underlying the activation mechanism of NTS1R.

NTSR1 is a member of the peptide receptors in the B-group of class A GPCRs [68]. NTSR1 has seven transmembrane helices (TM1-TM2-TM3-TM4-TM5-TM6-TM7), three extracellular loops (ECL1-ECL2-ECL3), and three intercellular loops (ICL1-ICL2-ICL3) like other receptors in GPCR family. The structure and corresponding regions of NTSR1 is given in Figure 3.123.

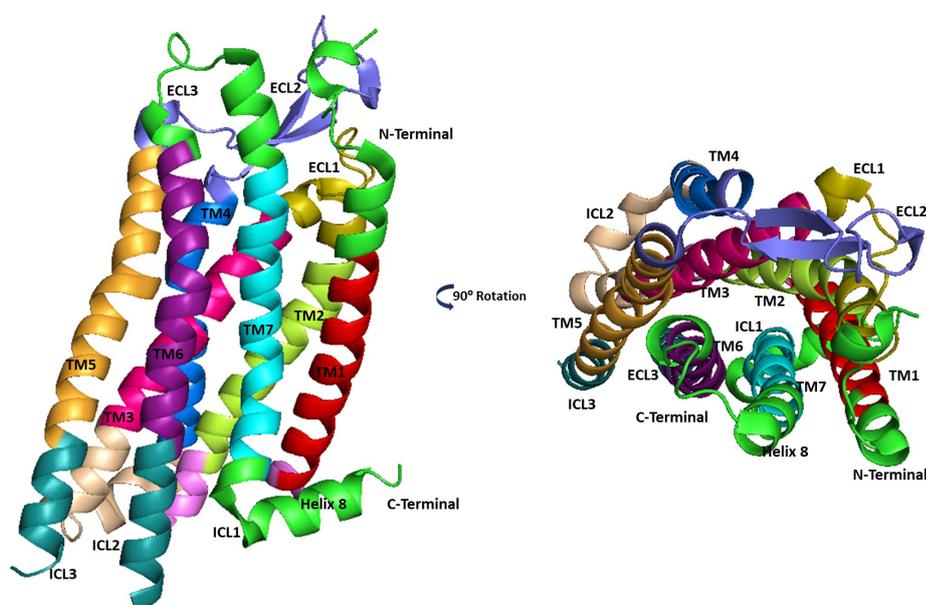


Figure 3.123. The structure and corresponding regions of NTSR1.

The residue number of NTS1 receptor starts with 51 and ends with 386. Since there are 29 missing residues, residue number of M2 receptor is 307. The corresponding regions of residues is given in Figure 3.124.

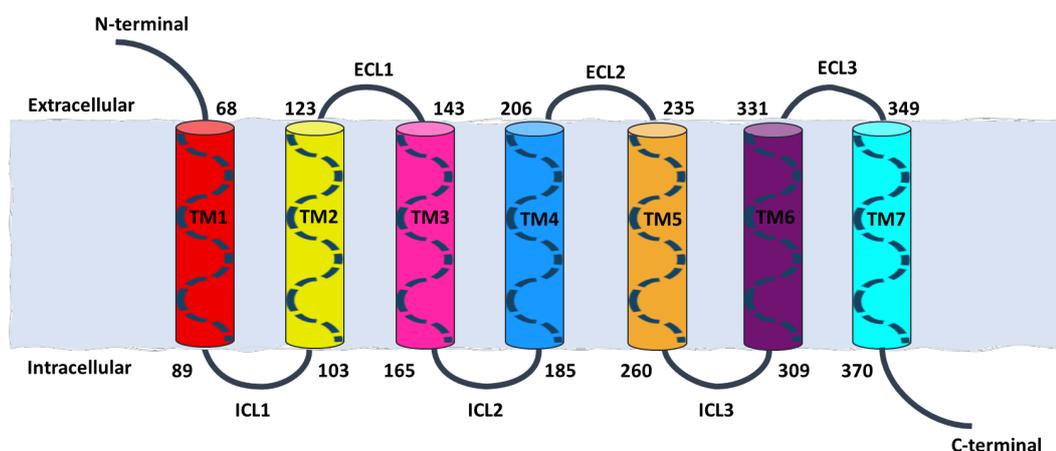


Figure 3.124. The wire diagram of NTS1 receptor.

Inactive and active states of GPCRs appear to show correlative conformational changes between different subspecies, while the activation process of different GPCRs varies in a couple of various ways. In this section, NTSR1 is utilized to explore the activation mechanism of GPCRs.

During the activation process, conformational changes happen especially in three novel districts that are called as ligand-binding site, connector, and G-protein binding site. The distance between R167<sup>3.50</sup> (E/DRY Motif) in ICL2 and L303<sup>6.30</sup> in TM6 is increases while the distance between R167<sup>3.50</sup> (E/DRY Motif) in ICL2 and Y369<sup>7.53</sup> (NPxxY Motif) in TM7 decreases (Ballesteros–Weinstein residue numbering [41]) [68]. The agonist binds to the binding pocket (Y146<sup>3.29</sup>-R327<sup>6.54</sup>) and causes some conformational changes in the structure, and the propagation of these conformational changes to the cytoplasmic surface occurs thanks to polar interactions in the connector region (A157<sup>3.40</sup>-P149<sup>5.50</sup>-F317<sup>6.44</sup>). While strong packing interactions in this region are also seen at the  $\beta$ 2-Adrenergic receptors and  $\mu$ -opioid receptors, it is determined that the

interactions at the muscarinic M2 receptor are weaker. N257<sup>5.58</sup> and Y369<sup>7.53</sup> in the G-protein binding site exhibit large conformational changes during the activation process and enable the stability of active state via the interaction between each other [68].

Functionally important residues and corresponding regions of NTSR1 are given in Table 3.12 and Figure 3.125.

Table 3.12. Residues and corresponding regions of functionally important sites of NTS1R.

<b>Important Sites</b>	<b>Residue Names</b>	<b>Region</b>
TM3 – TM6 Distance	R167 – L303	ICL2 – ICL3
TM3 – TM7 Distance	R167 – Y369	ICL2 – TM7
DRY Motif	E166 – R167 – Y168	ICL2
NPxxY Motif	N365 – P366 – Y369	TM7
Connector	A157-P249-F317	TM3-TM5-TM6
Ligand Binding Site	Y146-R327	TM-TM63
GP Binding Site	N257-Y369	TM5 – TM7

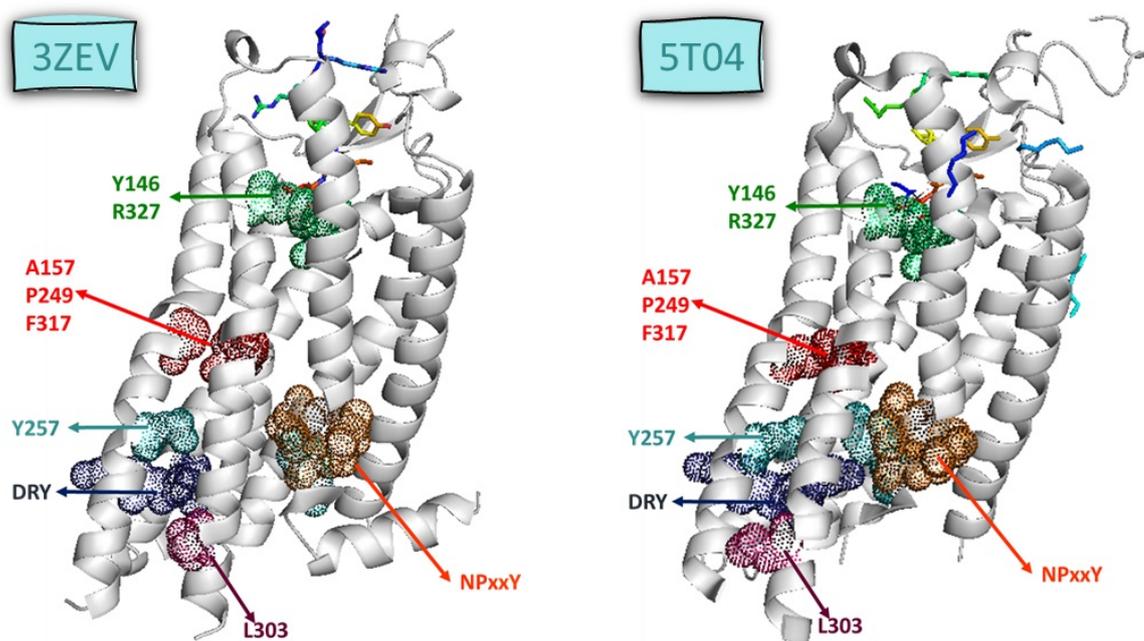


Figure 3.125. Functionally important sites of NTSR1.

### 3.3.1. ANM-LD Simulations of NTS1 Receptor

In ANM-LD simulations of NTSR1, the inactive and active structures which are obtained from Protein Data Bank with PDB IDs 3ZEV and 5T04, respectively. 3ZEV is the inactive state of NTS1 receptor, respectively. 3ZEV is the structure of thermostable agonist bound NTSR1 mutant without lysozyme fusion [67] and 5T04 is the structure constitutively active neurotensin receptor [68]. Properties of NTS1R structures are given in Table 3.13 [67,68].

**3.3.1.1. ANM-LD Simulations from 3ZEV to 5T04.** To explore the activation mechanism of NTS1R, ANM-LD simulations from inactive structure (3ZEV) to active structure (5T04) are examined. The alignment of active structure to inactive structure is demonstrated in Figure 3.126.

Table 3.13. Summary of used structure in ANM-LD simulations of NTS1 receptor.

PDB ID	Ligand & Complex	Organism	Mutation	Structure
3ZEV	GLY	<i>Rattus norvegicus</i>	11	Sidechains of connector region
5T04	TCE PEG GOL	<i>Rattus norvegicus</i>	10	Lack of Helix 8

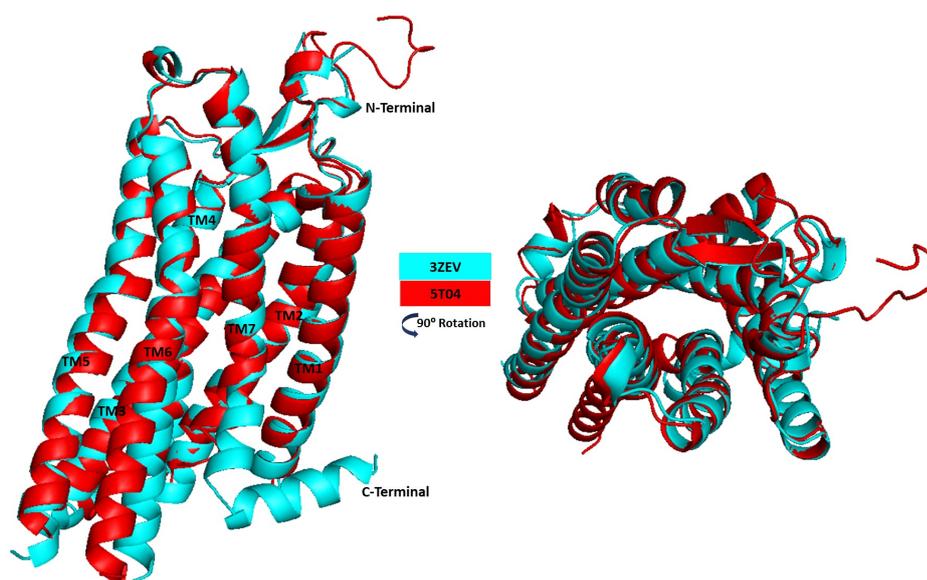


Figure 3.126. Alignment of active structure (5T04) to inactive structure (3ZEV) (RMSD: 2.28 Å).

Different simulations are prepared with different  $R_{cut}$  values of 10 Å, 13 Å, and 18 Å; maximum mod number of 30, 100, and all modes; DF values of 0.35 Å, 0.40 Å, 0.60 Å, adoptive between 0.1 Å and 0.4 Å - 1.0 Å. In the preparation step of initial and final structures for the simulation, S373-C386, G1387-G1391, chain B, chain C, and chain Dare deleted from inactive structure; T43-G50, N1002-S1165, and chain B are deleted from active structure. Additionally, L92-L93 in ICL1, S97-Q99 in ICL1,

and H269-Q270 in ICL3 are modeled by MOD-LOOP [53]. ANM-LD result summary is given in Table 3.14.

Table 3.14. ANM-LD Result Summaries.

Simulation Name	Cycle Number	Initial RMSD ( $\text{\AA}$ ) to Target Structure	Final RMSD ( $\text{\AA}$ ) to Target Structure	The Most Selected ANM Modes
$D_F = 0.10 - 1.00\text{\AA}$ All Modes $R_{cut} = 10\text{\AA}$	38	2.28	1.62	1, 2, 5, 8
$D_F = 0.60\text{\AA}$ Modemax = 30 $R_{cut} = 13\text{\AA}$	40	2.28	1.64	1, 12

The RMSD through the transition pathway obtained from ANM-LD simulations from inactive to active structure is investigated and given in Figure 3.127. RMSD of final generated structures of selected simulations are  $1.62 \text{ \AA}$  and  $1.64 \text{ \AA}$ . The RMSD profile through the transition pathway with best convergence is shown in Figure 3.128.

In spite of the fact that the RMSD difference isn't noteworthy, some key changes happen in the structure during the actuation pathway of NTSR1. These key changes are inspected to decide how the activation process occurs.

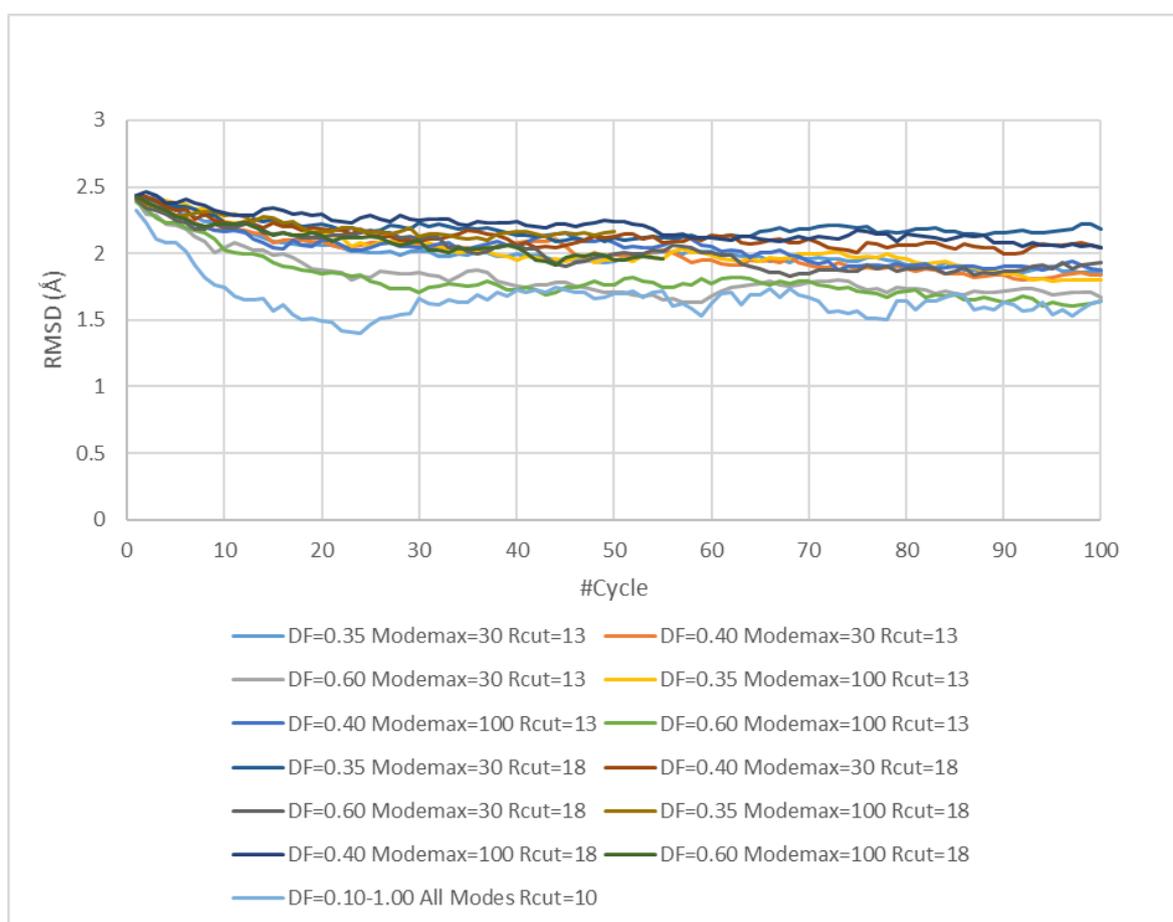


Figure 3.127. The RMSD difference through the simulations of NTS1R from 3ZEV to 5T04.

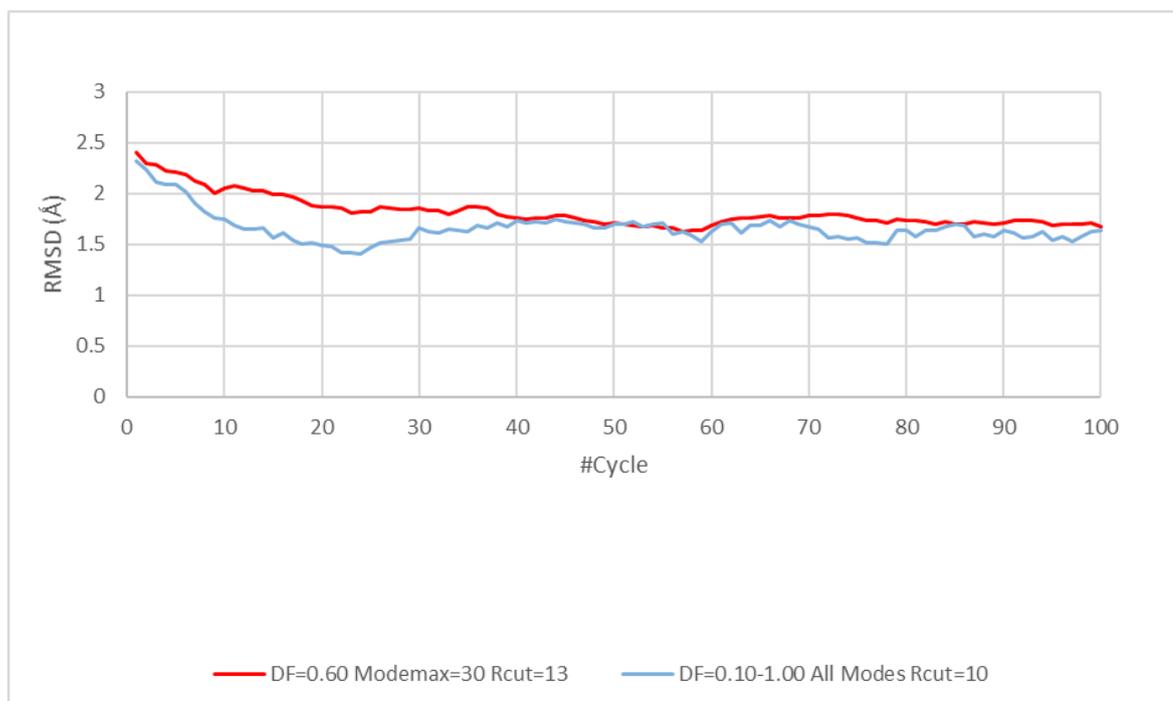


Figure 3.128. The RMSD difference of selected runs through the simulations of NTS1R from 3ZEV to 5T04.

- The TM3-TM6 and TM3-TM7 Distances.

First, the TM3-TM6 and TM3-TM7 distances, are investigated through the transition pathway, which are one of the indications of the activation process [68]. In the calculation of the TM3-TM6 and TM3-TM7 distances,  $C\alpha$  atoms of residues R167-L303 and R167-Y369 are used, respectively. The TM3-TM6 distance is 7.4 Å and 14.9 Å respectively, for the initial and final/target structures; and the TM3-TM7 distance is 15.7 Å and 11.4 Å, respectively, for the initial and final/target structures. The TM3-TM6 and TM3-TM7 distances through the transition pathway are demonstrated in Figure 3.129.

When the last five snapshots of the ANM-LD simulation with parameters of  $DF=0.10-1.00$  Å All Modes  $R_{cut}=10$  Å from thirteen parallel simulations are consid-

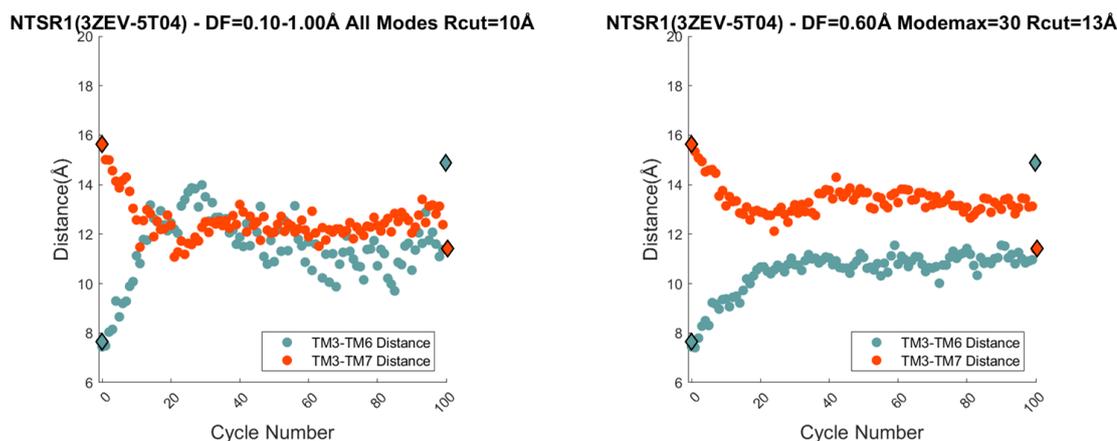


Figure 3.129. Functionally important distances through the transition pathway by ANM-LD simulations of 3ZEV to 5T04 (initial and final/target structures shown as diamond, snapshots shown as spheres).

ered, the TM3-TM6 distance is 11.6 Å; the TM3-TM7 distance is 12.9 Å. When the distance profile is investigated, it is seen that the TM3-TM6 distance of the snapshot in 30<sup>th</sup> cycle is closer to the distance value of the active structure. After 30<sup>th</sup> cycle, the TM3-TM6 distance decreases.

- DRY and NPxxY Motifs.

The RMSD of the important motifs E/DRY (E166 – R167 – Y168) and NPxxY (N365– P366– Y369) with respect to the initial structure is calculated by using C $\alpha$  atoms through the activation process is investigated [68]. The RMSD of DRY and NPxxY motifs of the active structure are used as reference value when investigated the RMSD profile of important motifs and are 1.2 Å and 2.8 Å, respectively. The RMSD profile of DRY and NPxxY motifs through the activation process are shown in Figure 3.130.

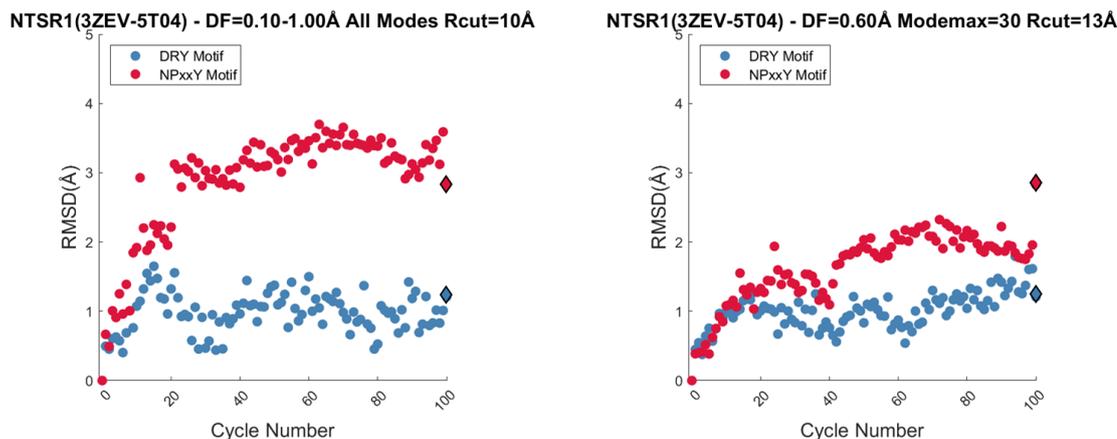


Figure 3.130. The RMSD of important motifs through the activation of NTSR1 from 3ZEV to 5T04 (initial and final/target structures shown as diamond, snapshots shown as spheres).

When the last five snapshots of the ANM-LD simulation with parameter  $DF=0.10-0.10$  Å All Modes  $R_{cut}=10$  Å are considered, the RMSD of DRY motif is calculated as 0.9 Å; the RMSD of NPxxY motif is calculated as 3.4 Å.

- Connector Region, Ligand Binding Site, and G-protein Binding Site.

The RMSD profiles of the connector site (A157-P249-F317), the ligand (GLY) binding site (Y146-R327), and G-protein binding site (N257) from their initial positions with respect to initial structure are investigated by using  $C\alpha$  atoms through the activation process. The RMSD between the initial and active/target structures is 2.0 Å, 1.7 Å, 1.1 Å, 1.3 Å, 0.5 Å, and 0.6 Å for A157, P249, F317, Y146, R327, and N257, respectively. The RMSD profile of connector region, ligand binding site, and G-protein binding site is given in Figure 3.131.

When the last five snapshots of ANM-LD simulation with parameter  $DF=0.10-0.10$  Å All Modes  $R_{cut}=10$  Å are considered, the RMSD of A157 is 0.8 Å; the RMSD of P249 is 1.1 Å; the RMSD of Y317 is 0.7 Å; the RMSD of Y146 is 0.4 Å, the RMSD of R327 is 0.8 Å; the RMSD of N257 is 0.5.

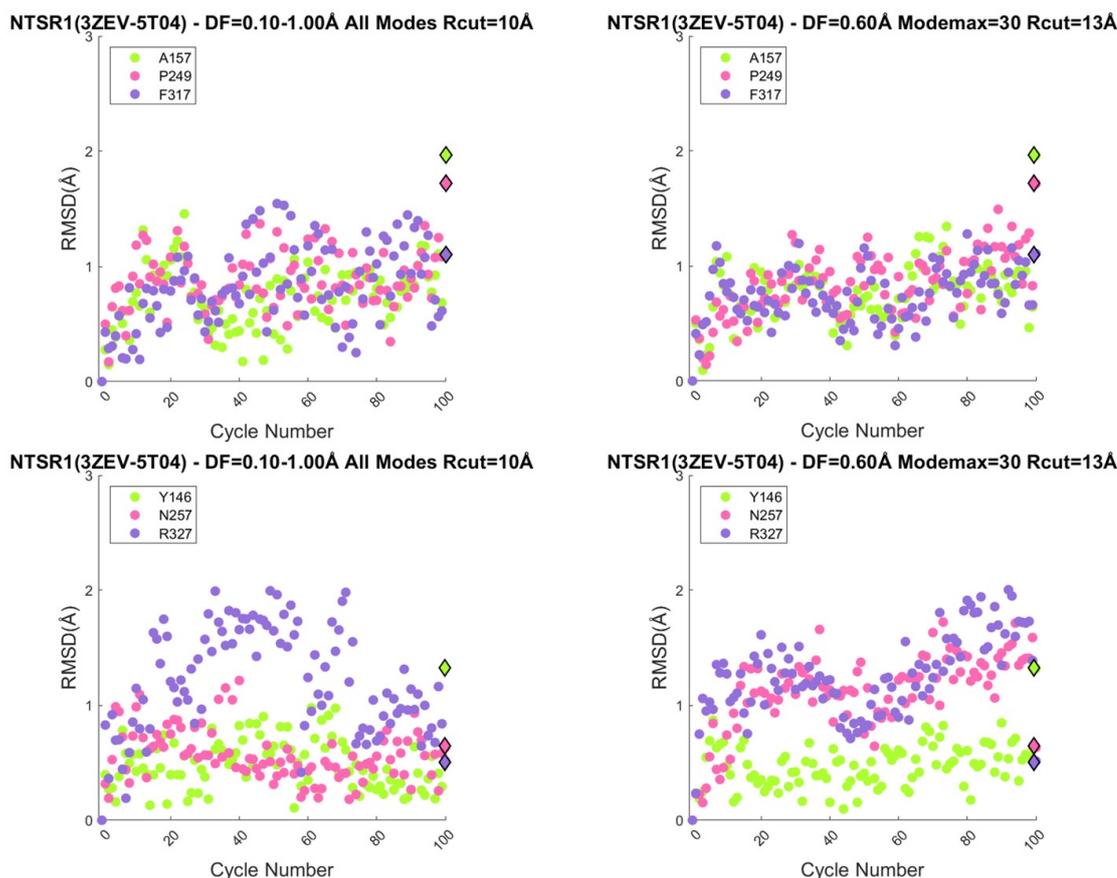


Figure 3.131. RMSD of important residues through the activation of NTSR1 receptor from 3ZEV to 5T04 (initial and final/target structures shown as diamond, snapshots shown as spheres).

- Key Modes of Motion.

The selected ANM modes through the transition pathway from inactive to active structure give an insight about the usage of characteristic elements of the structure. To accomplish this objective, the most selected ANM modes are investigated to reveal the dynamic determinants hidden the activation pathway. The examination of ANM modes are performed by system applied in the Section 3.1.1.1. The selected ANM modes through the transition pathway from 3ZEV to 5T04 are given in Figure 3.132.

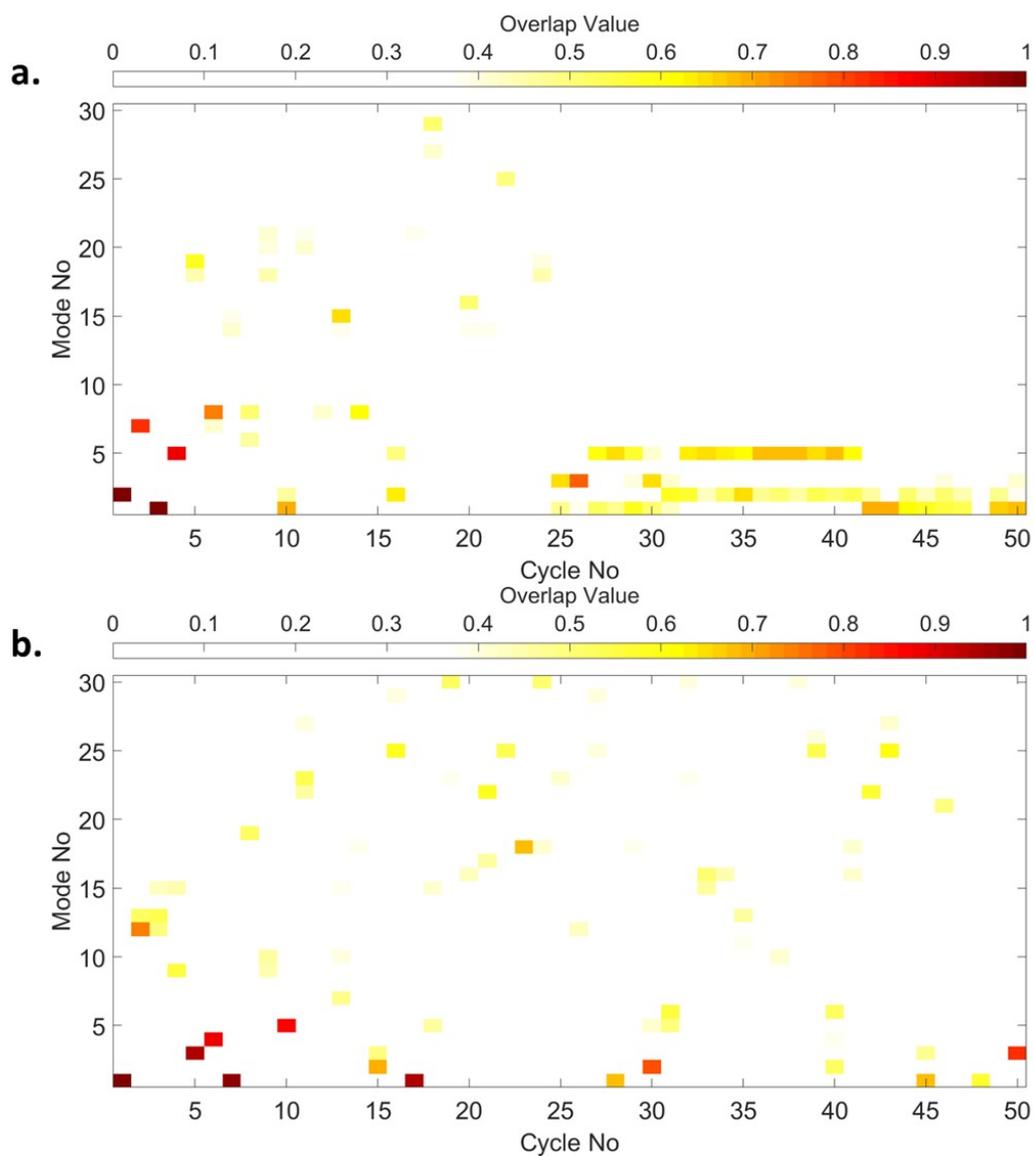


Figure 3.132. Selected Mode Mapping of NTS1R simulations from 3ZEV to 5T04 a) DF=0.10-1.00 Å All Modes  $R_{cut}=10$ . b) DF=0.60 Å Modemax=30  $R_{cut}=13$ .

The selected ANM modes in fifty cycles from 3ZEV to 5T04 are investigated and demonstrated in Figure 3.133. ANM-LD simulation with parameter  $DF=0.10-0.10 \text{ \AA}$  All Modes  $R_{cut}=10$ , where the activation process is best observed in the manner of analysis of collective variables, the most selected ANM modes with highest overlap value are determined as ANM mode 1, 2, 5, and 8.

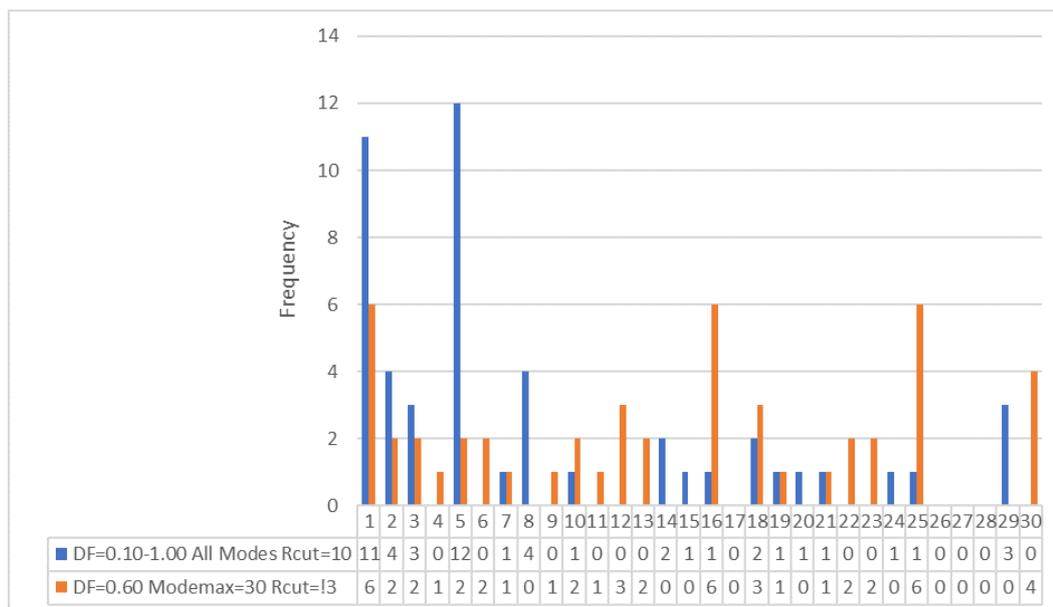


Figure 3.133. ANM mode frequency of parallel simulations from 3ZEV to 5T04.

Square fluctuations (SF) of difference vector between the active and inactive structure is shown in Figure 3.134. When this graph is compared with the SF of the selected ANM modes in Figure 3.135, it is understood that square fluctuations of ANM modes 1, 2, 5, and 8 similar to this difference vector.

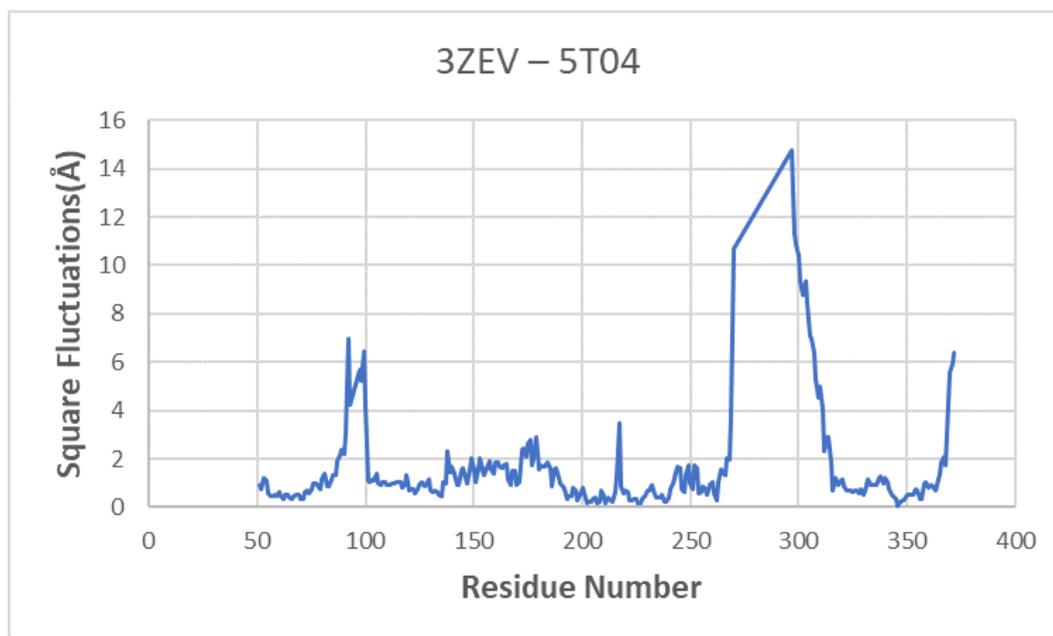


Figure 3.134. Difference vector between initial (3ZEV) and final (5T04) structure.

Minimum points in the MSF profile of ANM mode are identified as hinge residues of this mode, which presumably regulate the motion. The hinge residues of ANM slow mode 1 are identified as K64, T68, T88, V102, Y126-I129, C142, L163-V165, F189, G209, C225, I228, A253, L256, G306, L310, and Y351 and shown in Figure 3.136. When the hinge residues of ANM1 mode are investigated, it is seen that residues around the E/DRY motif and G-protein binding side overlap with the hinge residues.

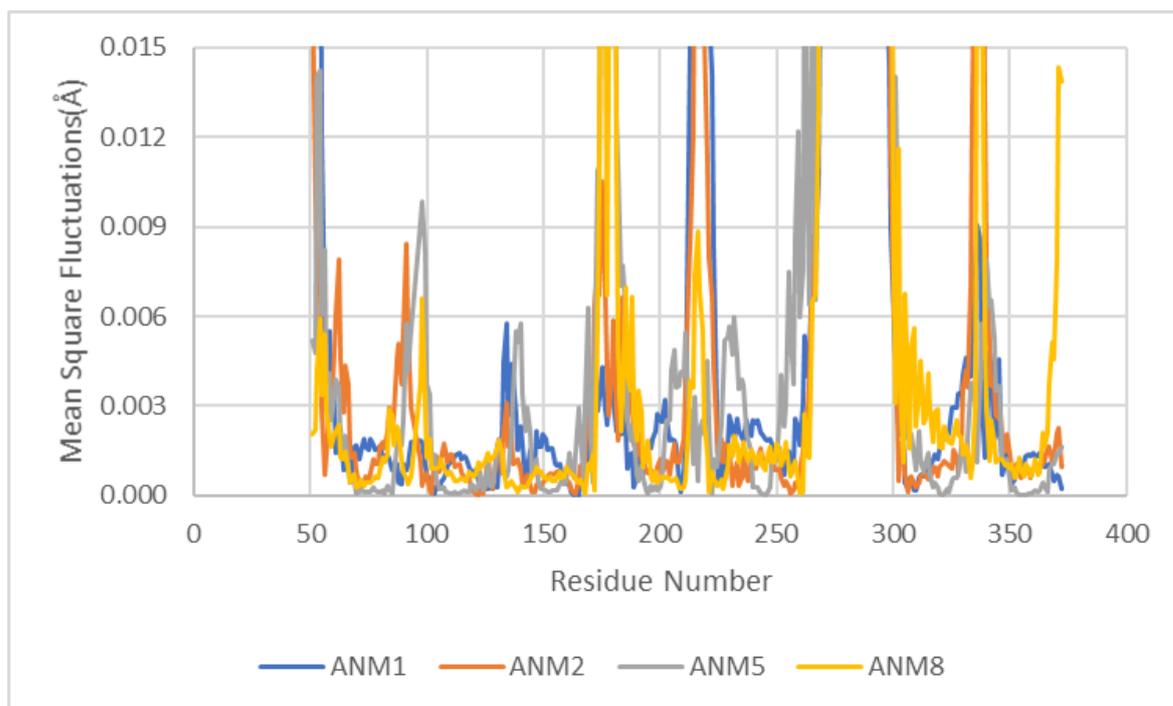


Figure 3.135. Mean square fluctuations of selected ANM modes of NTS1R simulations from 3ZEV to 5T04.

The hinge residues of ANM slow mode 2 are identified as D56, L72, V102, M121-E124, Y146, L163, A198, T231, L256, G306, and N355 and shown in Figure 3.137. Hinge residues of ANM mode 2 overlap with the ligand binding site and the residues around the G-protein binding.

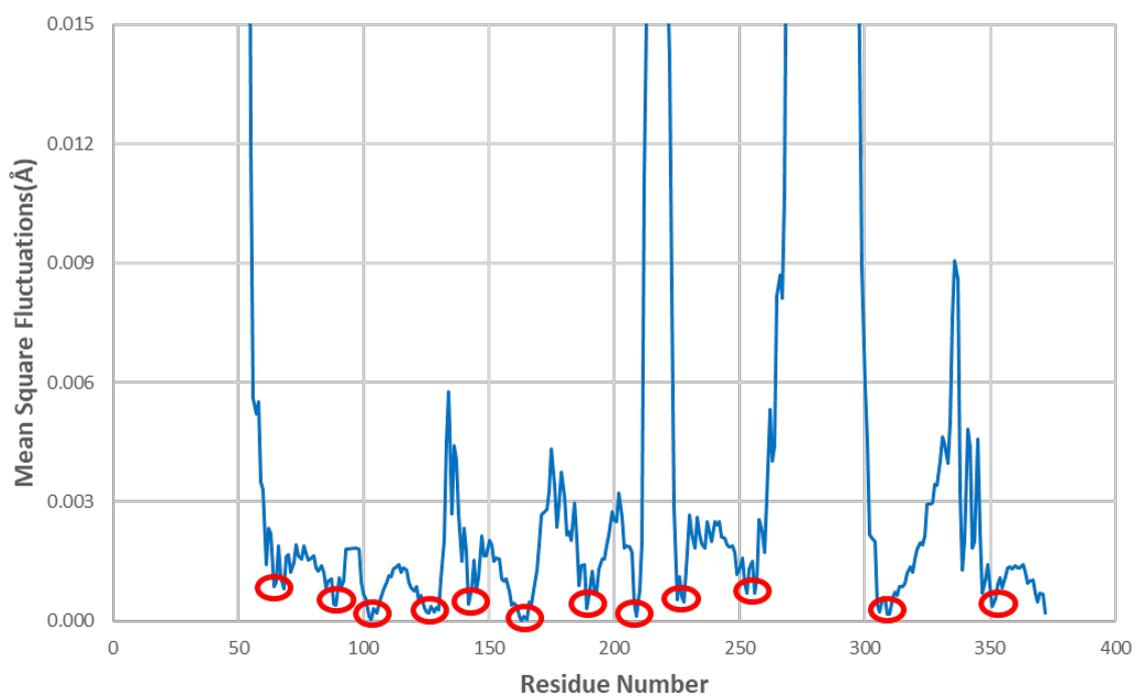


Figure 3.136. ANM mode 1 shape of the inactive state of NTS1R.

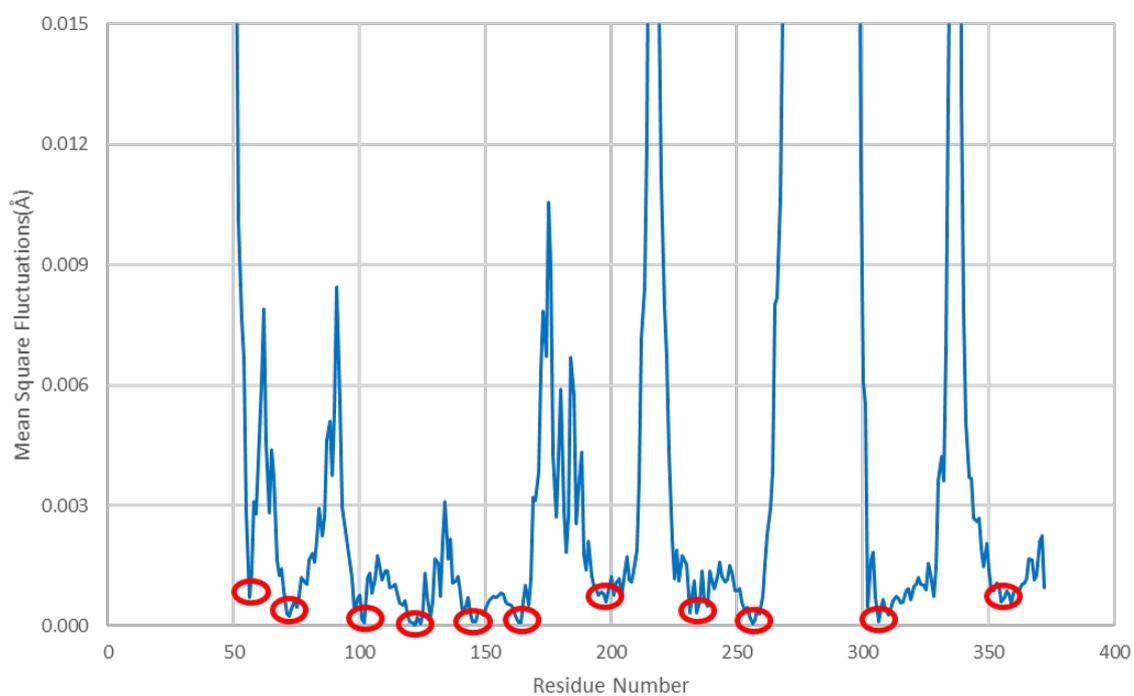


Figure 3.137. ANM mode 2 shape of the inactive state of NTS1R.

The hinge residues of ANM slow mode 5 are identified as L72-A73, T85, L106, Y154, A196-L200, D219, G222, F246, L322, and T354-F358 and shown in Figure 3.138. F358 located in the hydrophobic cascade which is predicted to be responsible for providing communication between the ligand binding site and the connector [68] is determined as a hinge residue of ANM mode 5.

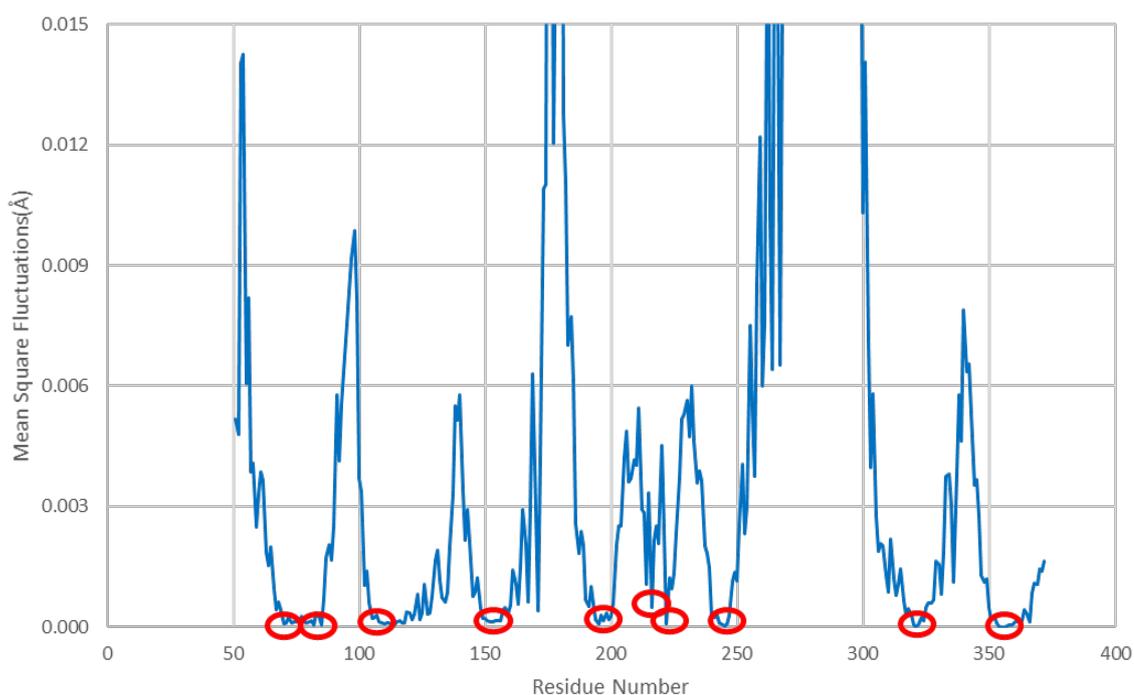


Figure 3.138. ANM mode 5 shape of the inactive state of NTS1R.

The hinge residues of ANM slow mode 8 are identified as I70-L74, L92, D139, Y168, C172, L210, G221, C225, A261, R304, Y333, M352, and Y359 and shown in Figure 3.139. When the hinge residues of ANM mode 8 are investigated, it is observed that E/DRY motif and Y359 which bonds with the antagonist in the antagonist docking studies in the literature [68] overlap with the hinge residues.

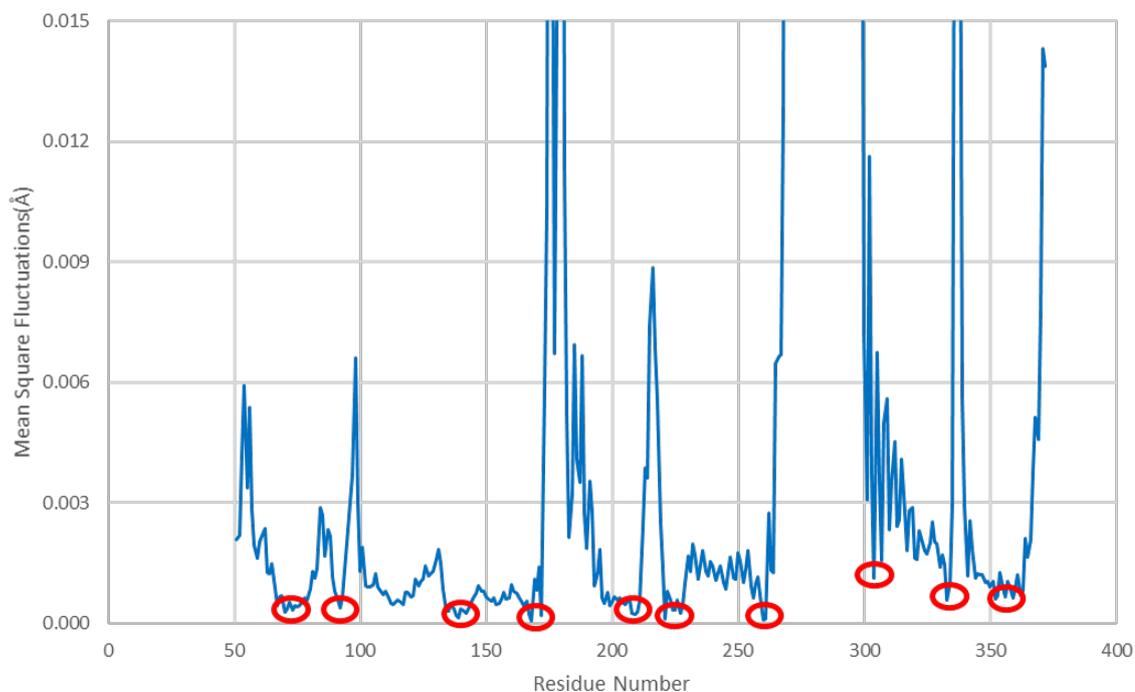


Figure 3.139. ANM mode 8 shape of the inactive state of NTS1R.

Hinge residues of ANM mode 1, 2, 5, and 8 are demonstrated on the inactive structure of NTS1 receptor in Figure 3.140.

Functionally important regions (DRY Motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L303) (red and sphere) and the hinge residues of ANM mode 1, 2, 5, and 8 (deep blue) are demonstrated on the initial structure in Figure 3.141.

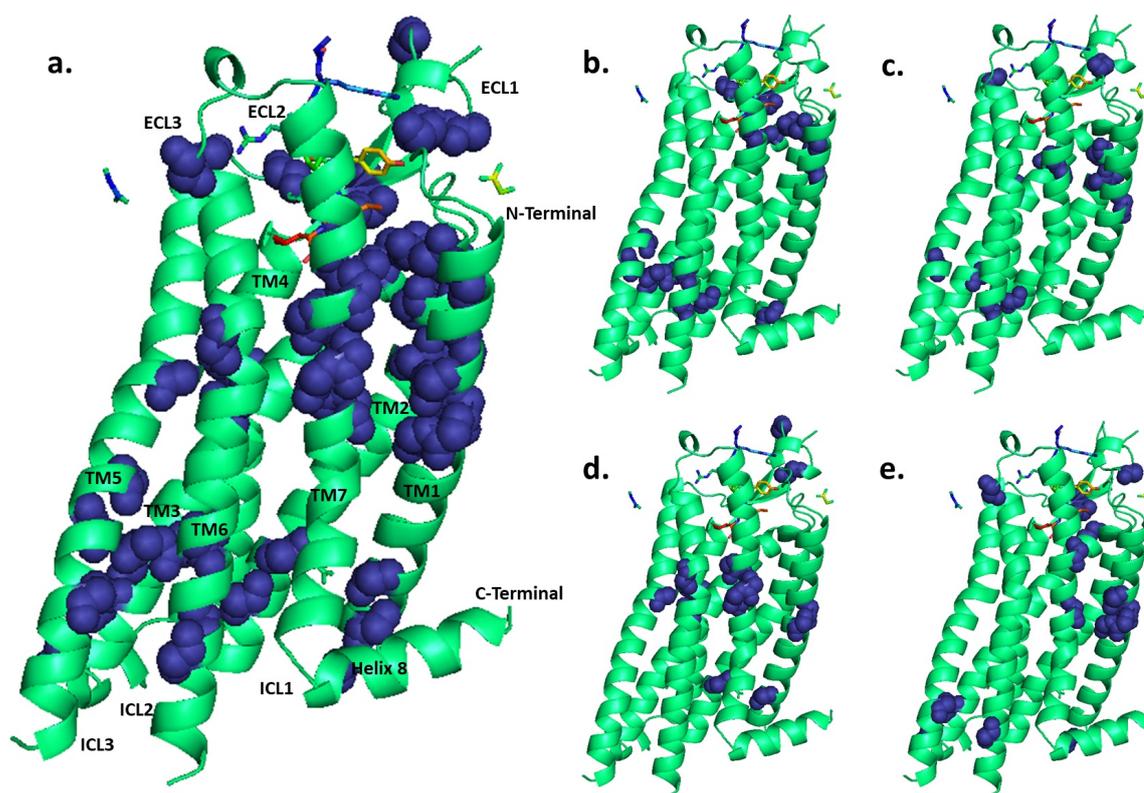


Figure 3.140. Demonstration of hinge residues of ANM mode 1, 2, 5, and 8 (dark blue and sphere) on the structure a)ANM modes 1, 2, 5, and 8 b)ANM mode 1. c)ANM mode 2 d)ANM mode 5 e)ANM mode 8.

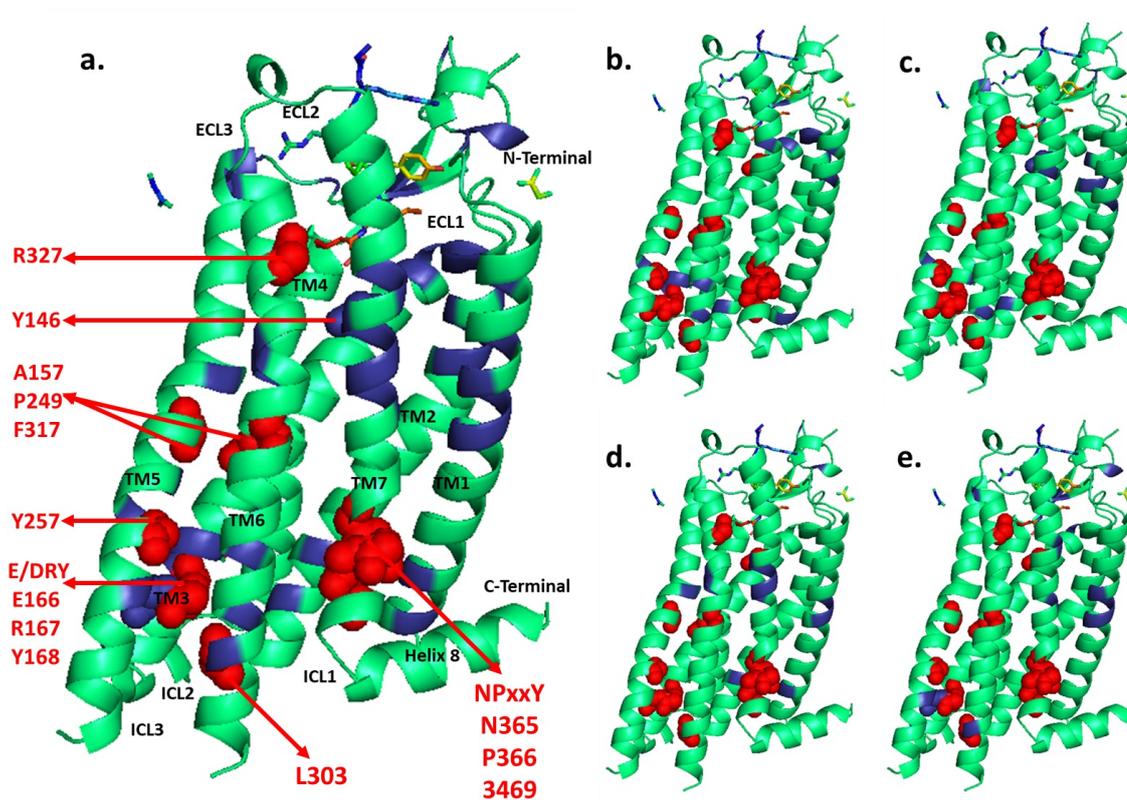


Figure 3.141. Demonstration of ANM modes 1, 2, 5, and 8 hinges (dark blue), important regions (red and sphere), the intersection (dark blue and sphere)  
 a) Modes 1, 2, 5, and 8 b) Mode 1 c) Mode 2 d) Mode 5 e) Mode 8.

- Cross-correlation Analysis.

Normalized cross-connection examination is utilized to distinguish the correlated residues. The investigation of cross-correlation is performed by method applied in the Section 3.1.1.1. Normalized cross-relationship examination is done through the actuation pathway of NTSR1 and showed in Figure 3.142.

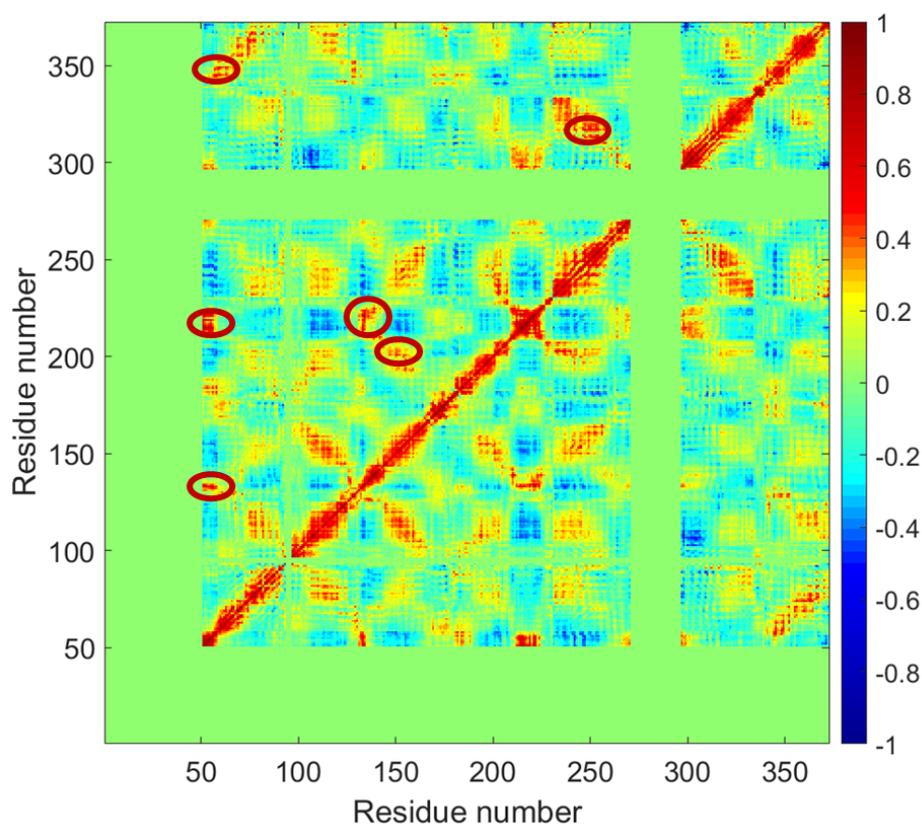


Figure 3.142. Cross correlation map of the activation pathway of NTS1R from 3ZEV to 5T04.

The highly correlated residues are determined as P51 & H153, P51 & D216, P51 & G221-G222, N52 & H133, N52 & S214-D216, N58 & D345, D139 & L210, and M250 & F317. When cross-correlation map is investigated, it is seen that the residues in N-terminal are highly correlated with TM3, TM4, and EC loop (ECL1-ECL2-ECL3).

Additionally, ECL1 loop and ECL2 loop are found as highly correlated regions. The correlation between TM5 and TM6 are found between M250 and F317, which have an important role in agonist- induced signal [68].

### 3.3.2. Activation Pathway of NTS1 Receptor

The conformational change in GPCR permits sending an extracellular sign through the cell membrane to produce an intracellular reaction. This conformational change is essential for GPCR activation process, where the G-protein binds to GPCR. In this part, NTS1R shows important conformational changes on the intracellular side of TM5, TM6, TM7 during the activation pathway via rearrangements in the ligand binding site. During the activation pathway of NTSR1, intracellular side of TM5 and TM6 moves outward and there is rearrangement in the NPxxY motif located in intracellular side of TM7 [68]. To be able to comprehend the dynamic behavior of the activation process and determine intermediate structures from the transition pathway obtained by ANM-LD simulation from 3ZEV to 5T04, some snapshots generated by the simulation are investigated and demonstrated in Figure 3.143.

The intermediate conformations in between the inactive and active structures obtained by ANM-LD simulations are analyzed through the activation pathway of NTSR1. The important residues on the snapshots obtained from the simulations from inactive structure (3ZEV) to active structure (5T04), TM3-TM6 and TM3-TM7 distances are demonstrated in Figures 3.144, 3.145 and 3.146. For the intermediate conformation (4GRV) of NTS1R available in the literature [69], the TM3-TM6 distance is calculated as 14.0 Å and the TM3-TM7 distance is calculated as 11.9 Å. The snapshot obtained from 60th cycle may be representing an alternative intermediate structure, since the snapshot obtained from 60th cycle has TM3-TM7 distance of 11.9 Å, while the TM3-TM6 distance is 11.7 Å. For this reason, it can be said that obtained transition pathway may be an alternative transition of the activation process.

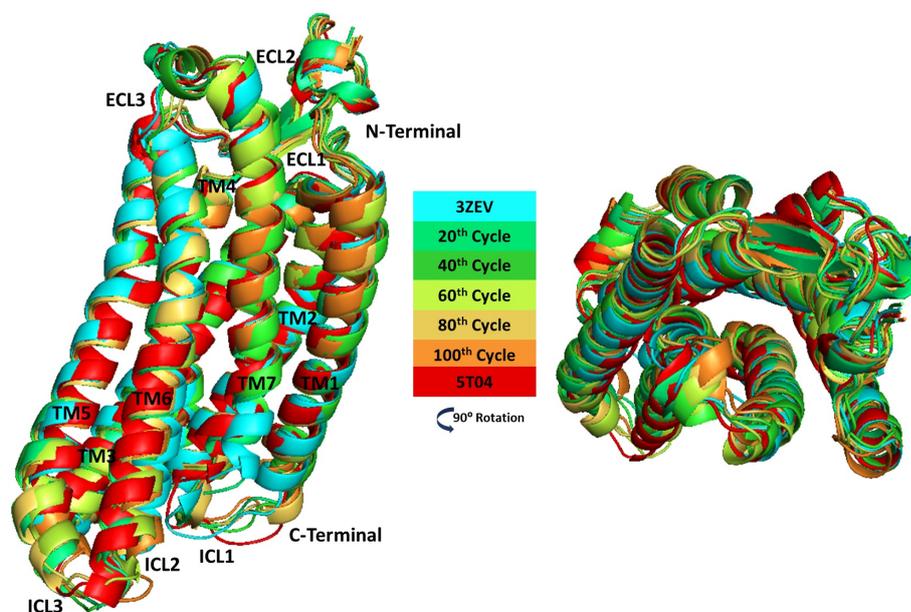


Figure 3.143. Snapshots generated by ANM-LD simulations from inactive (3ZEV) to active (5T04) structure.

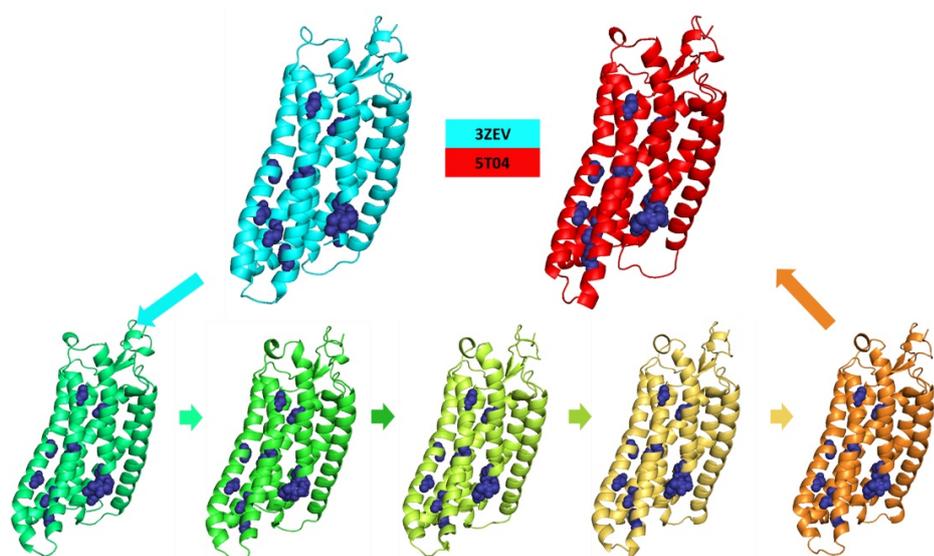


Figure 3.144. Important residues (DRY Motif, NPxxY motif, connector, ligand binding site, G-protein binding site, L303) (dark blue and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3ZEV to 5T04.

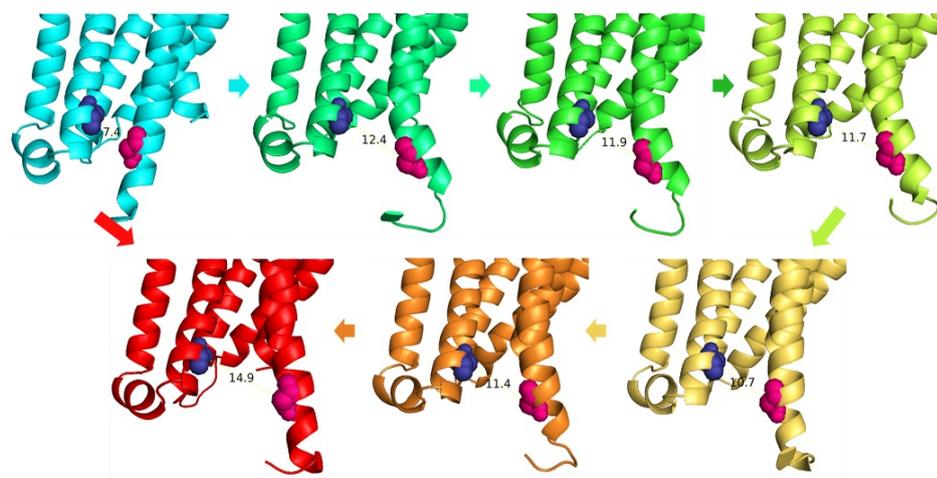


Figure 3.145. TM3-TM6 distance between R167 (dark blue and sphere) and L303 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3ZEV to 5T04.

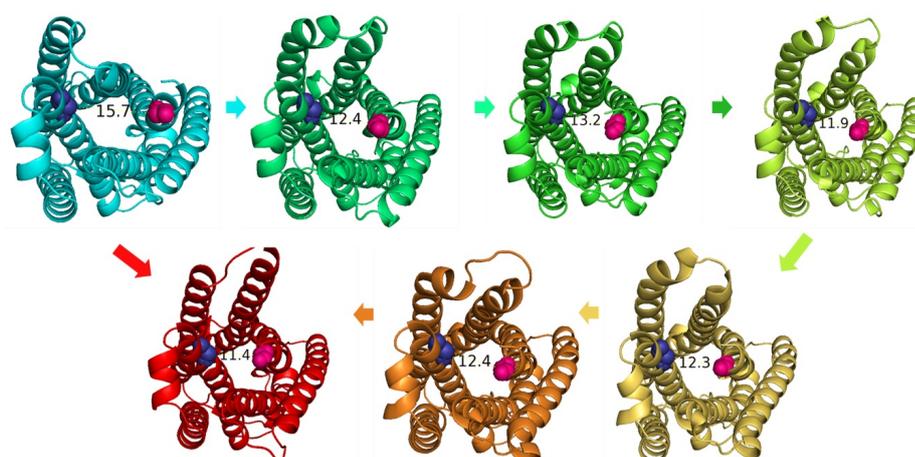


Figure 3.146. TM3-TM7 distance between R167 (dark blue and sphere) and Y369 (pink and sphere) demonstrated on the snapshots generated by ANM-LD simulations from 3ZEV to 5T04.

### 3.4. Comparison of Muscarinic Acetylcholine Receptor, $\beta$ 2-Adrenergic Receptor, and Neurotensin 1 Receptor

In this thesis, transition pathways, and dynamic factors of three different receptors, muscarinic acetylcholine receptor,  $\beta$ 2-Adrenergic receptor, and neurotensin 1 receptor, are investigated. M2R and  $\beta$ 2AR belong to homo sapiens species, while NTSR1 belongs to rattus norvegicus species [15, 36, 68].

In this part of thesis, transition pathways obtained by ANM-LD simulation of these receptors are examined. Alignment of inactive structures of M2R (3UON),  $\beta$ 2AR (2RH1), and NTSR1 (3ZEV) are demonstrated in Figure 3.147.

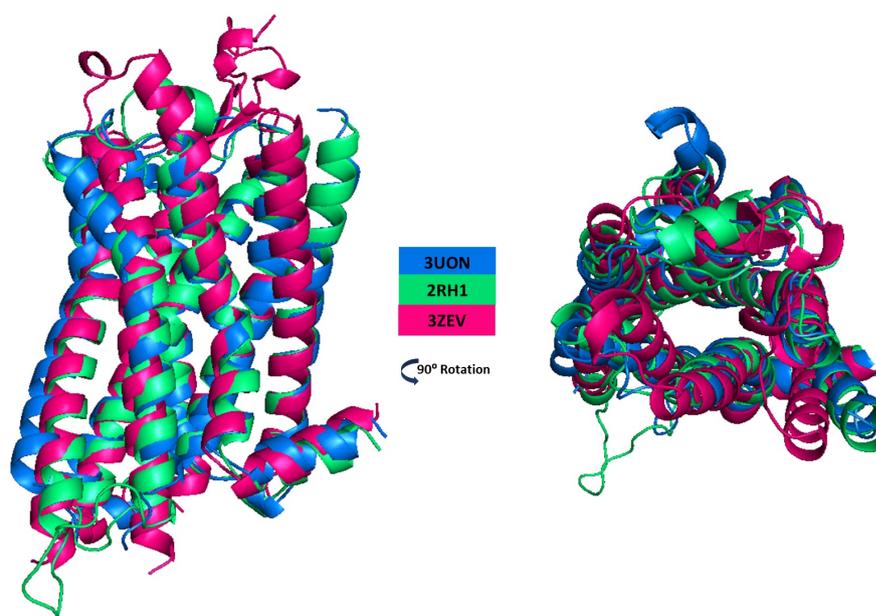


Figure 3.147. Alignment of inactive structures of M2R,  $\beta$ 2AR, and NTSR1.

The evolutionary conservation of the inactive structures of M2R,  $\beta$ 2AR, and NTSR1 from the family of GPCR receptors are examined with pairwise sequence alignment to obtain the similarity and identity index via EMBL-EBI EMBOSS Needle [70]. The similarity matrix is given in Table 3.15. According to the similarity matrix, inactive structures of M2R and  $\beta$ 2AR show 55.5 percent similarity, while inactive structure

of NTS1R show 48.8 percent and 48.7 percent similarity with the inactive structures of M2R and  $\beta$ 2AR, respectively.

Table 3.15. Similarity Matrix (in percent) of inactive structures of M2R,  $\beta$ 2AR, and NTSR1.

	2RH1	3UON	3ZEV
2RH1	100	55.5	48.8
3UON	55.5	100	48.7
3ZEV	48.8	48.7	100

The identity matrix created by EMBOSS Needle [70] is given in Table 3.16. When identity matrix is investigated it is seen that inactive structures of M2R and  $\beta$ 2AR show 28.8 percent identity, while inactive structure of NTS1R show 20.2 percent and 22.5 percent identity with the inactive structures of M2R and  $\beta$ 2AR, respectively.

Table 3.16. Identity Matrix (in percent) of inactive structures of M2R,  $\beta$ 2AR, and NTSR1.

	2RH1	3UON	3ZEV
2RH1	100	28.8	20.2
3UON	28.8	100	22.5
3ZEV	20.2	22.5	100

Multiple sequence alignment (MSA) is used to align three inactive structures of M2R,  $\beta$ 2AR, and NTSR1. This analysis is performed in EMBL-EBI Clustal Omega [70] application. Multiple sequence alignment of these inactive structures is given in Figure 3.148.



According to multiple sequence alignment analysis, NPxxY motif,  $P^{5.50}$ - $F^{6.44}$  in connector region, and W6.48 in ligand binding site are determined as identical. Additionally,  $Y^{5.58}$  located in G-protein binding site is determined as identical for 3UON and 2RH1 inactive structures.

Collective variable analysis performed on the transition pathways generated by ANM-LD simulations are compared to analyze the differences and similarities in the transition pathways [15, 40, 49, 68]. The TM3-TM6 and TM3-TM7 distances, and the RMSD of DRY and NPxxY motif through the transition pathway of M2R,  $\beta$ 2AR, and NTS1R shown in Figure 3.149.

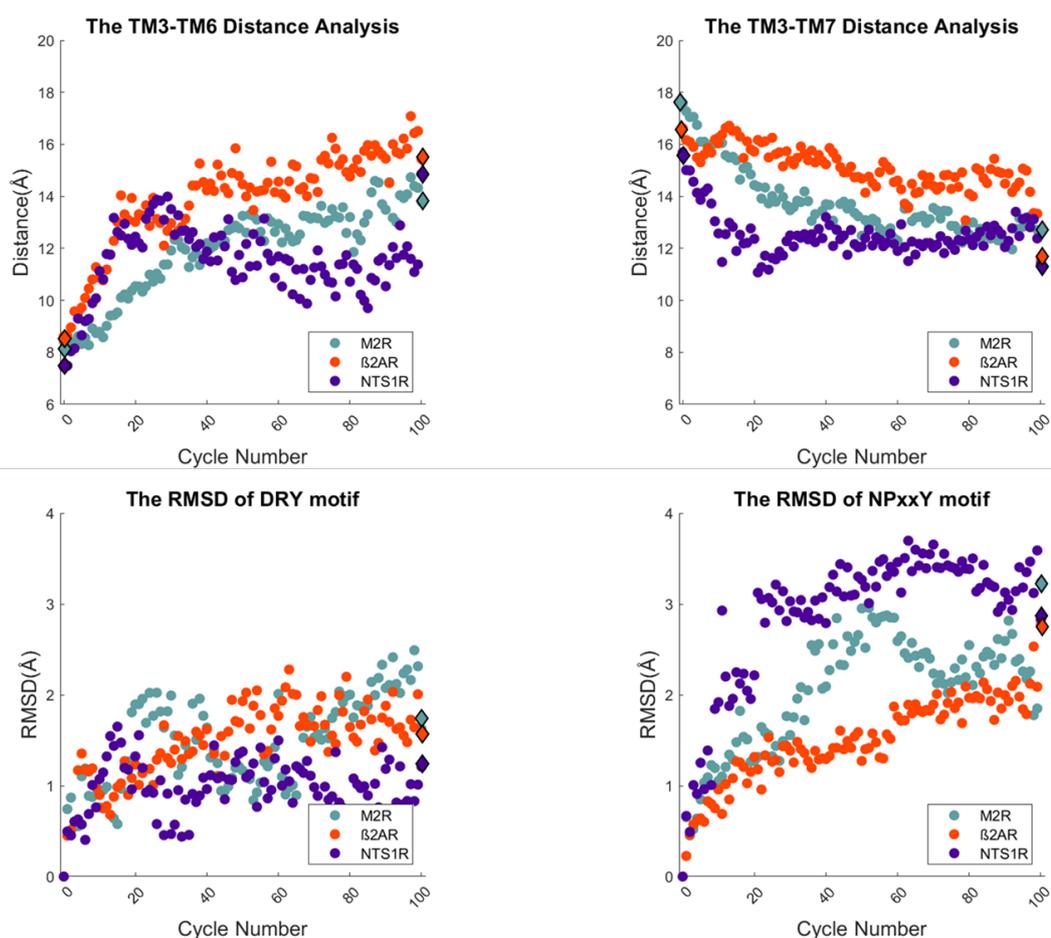


Figure 3.149. Comparison of collective variable analysis through the transition pathway of M2R,  $\beta$ 2AR, and NTS1R.

When collective variable analysis of M2R,  $\beta$ 2AR, and NTS1R compared with each other, it is seen that the calculated distances and RMSD values of inactive and active structure are very close to each other, however there is differences in transition pathway. In TM3-TM6 distance analysis, it is observed that  $\beta$ 2AR and NTS1R show almost identical profile until 30<sup>th</sup> cycle, while the TM3-TM6 distance of M2R increases at a slower rate. The TM3-TM7 distance profile of NTS1R reaches the target structure very quickly compared to other receptors. While the distance of TM3-TM7 of the  $\beta$ 2AR decreases more slowly, it decreases more rapidly at the end of the simulation. When the RMSD values of the DRY motif are examined, the RMSD value of the target structure differs slightly for NTS1R, which may be due to the mutation ( $R167L^{3.50}$ ) in the DRY motif of the NTS1R [68]. The RMSDs of NPxxY motif of active structures of  $\beta$ 2AR and NTS1R are found very close. As in the TM3-TM7 distance profile, the RMSD value of the NTS1R is converged to the active structure in the first 30 cycles. In the simulation using  $\beta$ 2AR, the RMSD of NPxxY motif increased slowly, and this increase accelerated at the end of the simulation. In the calculation for M2R, the RMSD of NPxxY motif approaches the target structure around the 50<sup>th</sup> cycle and then starts to show fluctuations.

When the results of cross-correlation analysis are compared for three receptors, while the NPxxY motif correlated with the extracellular side of TM7 in  $\beta$ 2AR, it is correlated with the TM6 helix at the M2 receptor. No residues in the NTSR1 with a highly correlated NPxxY motif are able to be detected. A correlation is noticed between P157 in TM4 and the residue at the end of TM6, which was used to calculate the distance of TM3-TM6 in M2R, however, this correlation is found in calculations only for M2R. The comparison of cross correlation analysis indicates that interactions differ markedly for different subspecies, which are reflected in the differences of the dynamic infrastructure of allosteric interactions.

## 4. CONCLUSION AND RECOMMENDATIONS

### 4.1. Conclusion

The aim of the thesis is to comparatively study the structural dynamics of homologous proteins, and to determine the dynamic factors underlying activation in the members of GPCR family. For this purpose, it is aimed to generate inactive-active conformational transition pathways of muscarinic acetylcholine,  $\beta$ 2-Adrenergic receptor, and neurotensin 1 receptor with the ANM-LD methodology to determine the most dominant dynamic modes underlying transition pathways that enable the activation process and how they all compare between different GPCR members. Common and specific traits in the functional dynamic behavior of different GPCR members are requested.

Conformational changes in GPCR enable the extracellular signal to be communicated over the cell membrane to create an intracellular reaction. Activation of GPCR depends on this conformational change to form G-protein binding conformational state. During activation, conformational changes happen particularly in three unique regions that are called as ligand-binding site, connector region, and G-protein binding site. Intracellular side of TM6 and TM5 moves outward and NPxxY motif in TM7 rearranges through a cage formation [40]. The broken ionic lock between the side chains  $R^{3.50}$  in DRY motif and  $E^{6.30}$  enables the increase in the distances between  $R^{3.50}$  in DRY motif in the intracellular side of TM3 and  $T^{6.34}$  at the end of TM6. Along, the distance between  $R^{3.50}$  in DRY motif and  $Y^{7.53}$  in NPxxY Motif in TM7 decreases [42, 43, 45]. The predicted transition pathways by ANM-LD simulations are requested if adhere to these latter structural changes for the physical and biological relevance of the activation.

To explore the activation process of the M2 receptor, two different cases (Case 1 - Case 2) are studied. In the first case, the simulations from inactive structure bound to an antagonist (3UON) to active structure bound to agonist iperoxo (4MQS) are investigated. In this case, the TM3-TM6 distance increases to 14.3 Å, the TM3-

TM7 distance decreases to 12.9 Å, and the distance between the residues forming salt bridge increases 19.2 Å. The RMSD of the important motifs DRY (D120 – R121 – Y122) and NPxxY (N436 – P437 – Y440) with respect to the initial structure is calculated by using C $\alpha$  atoms through the activation process. Additionally, The RMSD profiles of the connector site (V111 and P198), the ligand (QNB) binding site (W400) and the G-protein binding site (Y206) from their initial positions are investigated by using C $\alpha$  atoms through the activation process. Based on the results of these collective variable analyses, it is concluded that the transition pathway mirrors the activation process of M2 receptor. The most dominant dynamic modes that underly the transition pathways in the activation disclose mechanistically important amino acids that control the activation dynamics. Accordingly, the most frequently selected modes; ANM modes 1 and 5, yields that the most effective regions during the activation process are intracellular loops (ICL1-ICL2), extracellular loop (ECL2), and residues located around the ligand (QNB) binding site (W400). Moreover, the overlap of the hinge residues of selected ANM modes with previously proposed allosteric sites [36,40,43,45] support the notion that the hinge residues constitute to a dynamic allosteric network. As the second, the simulations from inactive structure bound to an antagonist (3UON) to active structure bound to agonist iperoxo and allosteric modulator (4MQT) are investigated. In this case, the results of the distance between TM3-TM7 and the RMSD of NPxxY with respect to inactive structure are achieved closer to the target structure than the first case, which implies that allosteric modulator in the active structure affects the transition pathway in a functional favorable way. The most frequently selected modes where activation takes place for both cases (Case 1 - Case2) of M2 receptor are compared, a correspondence are observed as such that major profile of residue fluctuations are similar while the differences mainly lies in the orthosteric ligand-binding site. When the allosteric regions are examined, it is seen that the key collaborating residues are better predicted in this case with the activator. Additionally, the agreement of the predicted intermediate conformation by ANM-LD simulations the intermediate structures proposed via accelerated MD simulations [36] supports the physical relevance of the predictions of ANM-LD for M2.

The activation process of  $\beta$ 2AR investigated through four different cases (Case 1 - Case 2 - Case 3 - Case 4); two inactive structures of  $\beta$ 2AR (2RH1-3D4S) and two active structures (3SN6-3P0G) yields as follows. In the first case (ANM-LD simulations from 2RH1 to 3SN6), the behavior of structural motifs discloses the transition pathways that reflects the activation process successfully. The hinges of the mostly selected ANM modes 7, 11, and 24 coincide with DRY Motif, L272 at the end of TM6 helix, T110 in orthosteric binding site, F282 in the connector region. Additionally, F193 is identified as hinge residue overlap with the most promising allosteric site in extracellular loop [55, 56]. Moreover, ICL1 being highly correlated with the C-terminal and intracellular sides of TM6 and TM7 are also determined as highly correlated residues is likely an allosteric side. In the second case, the conformational transition pathway from inactive structure (2RH1) to active structure (3P0G) by ANM-LD simulations represents a physically plausible activation pathway. The hinges of the mostly selected ANM mode 9 overlap with functionally important DRY motif, NPxxY motif, orthosteric ligand binding site (S207), and G-protein binding site (Y219). Although F193 is a hinge residue in ICL3 loop in first case of  $\beta$ 2AR but not in this case. The major reason for this situation is the missing residues in the ICL3 loop of the structure, which affects the selected ANM modes. Furthermore, DRY motif is affected by T68 at the end of TM2 and NPxxY motif is correlated with F61-L64 in ICL2. The intermediate structures proposed previously is observed in the 20<sup>th</sup> snapshot and the 60<sup>th</sup> snapshot generated in the first and second cases, respectively [15], which implies that the intermediate structures of  $\beta$ 2AR are obtained by ANM-LD simulations from the inactive structure (2RH1) to active structures (3SN6-3P0G). When the snapshots obtained from ANM-LD simulations from inactive structure (3D4S) to active structure (3SN6-3P0G) are examined, it is seen that activation process is not fully accomplished. The difference between two inactive structure located in ICL2 loop may be the reason for this situation, since ICL2 where the cholesterol binding site in 3D4S is located, has a great role in the conformational movements of the structure [58, 59]. Therefore, it is understood that the change in the ICL2 region blocks the transition pathway.

In the investigation of activation process of neurotensin 1 receptor, inactive (3ZEV) and active (5T04) structures are used. In activation process, the mostly selected ANM modes are 1, 2, 5, and 8. Hinge residues of mostly selected ANM modes coincide with E/DRY motif, residues around the G-protein binding site, F358 which is responsible of the communication between ligand binding site and the connector. When cross-correlation map is investigated, it is seen that the residues in N-terminal are highly correlated with TM3, TM4, and EC loop (ECL1-ECL2-ECL3), which implies that there is a possible allosteric region located in N-terminal. Additionally, the correlation between TM5 and TM6 is determined between the residues M250 and F317, which have an important role in agonist- induced signal [68]. Intermediate structure (4GRV) of NTS1R available in the literature [69] is compared according to collective variable analysis with the snapshots obtained by ANM-LD simulation. There is a 3 Å difference in the TM3-TM6 distance analysis compared with the intermediate structure obtained by ANM-LD simulation, while the TM3-TM7 distance approximately equal for two structures. For this reason, it is possible that obtained transition pathway may be an alternative transition of the activation process or the transition pathway generated by ANM-LD simulation.

In the last part of the thesis, transition pathways obtained for the three receptors (M2R,  $\beta$ 2AR, NTS1R) mentioned above are compared with each other. M2R and  $\beta$ 2AR belong to homo sapiens species, while NTSR1 belongs to rattus norvegicus species [15, 36, 68].

The evolutionary conservation of these three receptors, their inactive structures, via EMBOSS Needle [70] show that according to the similarity matrix, inactive structures of M2R and  $\beta$ 2AR show 55.5% similarity, while inactive structure of NTS1R show 48.8% and 48.7% similarity with the inactive structures of M2R and  $\beta$ 2AR, respectively. The TM3-TM6 and TM3-TM7 distances and the way of a few functionally imperative motifs depict the pathway that appear commonality and contrasts in three GPCRs. The agonist binding pocket interactions contrast particularly for distinctive subspecies, which reflected within the differences of chosen dynamic modes of movement.

Additionally, multiple sequence alignment is performed to find similarities in sequence between these receptors. The results of multiple sequence alignment indicate that NPxxY motif,  $P^{5.50}$ - $F^{6.44}$  in connector region, and  $W^{6.48}$  in ligand binding site are determined as identical in three receptors and  $Y^{5.58}$  located in G-protein binding site is determined as identical for M2R and  $\beta$ 2AR. When M2R is compared with NTSR1 and  $\beta$ 2AR, the weaker packing interactions result in more adaptable fragments within the connector of NTSR1 and  $\beta$ 2AR resulting within the TM3-TM6 distance decays slower. In spite of the fact that a common agonist binding site does not exist, the preserved DRY, and NPxxY motifs are the common drivers of the G-protein binding site. These receptors are compared with respect to cross-correlation analysis. In M2R and  $\beta$ 2AR, NPxxY motif is highly correlated with TM7 helix and TM6 helix, respectively. There are no residues in NTS1R with a highly correlated NPxxY motif. Additionally, at the end of TM6 helix and TM4 helix are determined as highly correlated residues in M2R. Considering the cross-correlation results, it is implied that there are differences in the dynamic infrastructure of allosteric interactions in different members of GPCR family.

## 4.2. Recommendation

This study can be improved by the determination of transition pathways of different sub-members of GPCR family with physical and biological reality, comparison of water pathway formation in intermediate structures along transition pathways with different homologues. Additionally, to determine alternative transition pathways by restricted ANM-LD simulations enables us to determine evolutionarily conserved and changed dynamic factors. In order to obtain more realistic transition pathway, ANM-LD simulations containing ligands can be performed or all-atom ANM-LD simulations can be performed to evaluate the side chains of amino acids.

## REFERENCES

1. Latorraca, N. R., A. Venkatakrishnan and R. O. Dror, “GPCR dynamics: structures in motion”, *Chemical Reviews*, Vol. 117, No. 1, pp. 139–155, 2017.
2. Geppetti, P., N. A. Veldhuis, T. Lieu and N. W. Bunnett, “G protein-coupled receptors: dynamic machines for signaling pain and itch”, *Neuron*, Vol. 88, No. 4, pp. 635–649, 2015.
3. Park, S. H., B. B. Das, F. Casagrande, Y. Tian, H. J. Nothnagel, M. Chu, H. Kiefer, K. Maier, A. A. De Angelis, F. M. Marassi *et al.*, “Structure of the chemokine receptor CXCR1 in phospholipid bilayers”, *Nature*, Vol. 491, No. 7426, pp. 779–783, 2012.
4. Henzler-Wildman, K. and D. Kern, “Dynamic personalities of proteins”, *Nature*, Vol. 450, No. 7172, pp. 964–972, 2007.
5. Rodríguez-Espigares, I., M. Torrens-Fontanals, J. K. Tiemann, D. Aranda-García, J. M. Ramírez-Anguita, T. M. Stepniewski, N. Worp, A. Varela-Rial, A. Morales-Pastor, B. Medel-Lacruz *et al.*, “GPCRmd uncovers the dynamics of the 3D-GPCRome”, *Nature Methods*, Vol. 17, pp. 1–11, 2020.
6. Selent, J., F. Sanz, M. Pastor and G. De Fabritiis, “Induced effects of sodium ions on dopaminergic G-protein coupled receptors”, *PLoS Comput Biol*, Vol. 6, No. 8, p. e1000884, 2010.
7. Chen, K.-Y. M., D. Keri and P. Barth, “Computational design of G Protein-Coupled Receptor allosteric signal transductions”, *Nature Chemical Biology*, Vol. 16, No. 1, pp. 77–86, 2020.
8. Fas Aykaç, B., *Computational and Experimental Investigations into the Mechanism of Functional Motion in Biomolecular Systems*, Ph.D. Thesis, Bogazici University,

2016.

9. Attwood, T. and J. Findlay, “Fingerprinting G-protein-coupled receptors”, *Protein Engineering, Design and Selection*, Vol. 7, No. 2, pp. 195–203, 1994.
10. Pándy-Szekeres, G., C. Munk, T. M. Tsonkov, S. Mordalski, K. Harpsøe, A. S. Hauser, A. J. Bojarski and D. E. Gloriam, “GPCRdb in 2018: adding GPCR structure models and ligands”, *Nucleic Acids Research*, Vol. 46, No. D1, pp. D440–D446, 2018.
11. Jeschke, G., “DEER distance measurements on proteins”, *Annual Review of Physical Chemistry*, Vol. 63, pp. 419–446, 2012.
12. Lohse, M. J., S. Nuber and C. Hoffmann, “Fluorescence/bioluminescence resonance energy transfer techniques to study G-protein-coupled receptor activation and signaling”, *Pharmacological Reviews*, Vol. 64, No. 2, pp. 299–336, 2012.
13. Xue, L., X. Rovira, P. Scholler, H. Zhao, J. Liu, J.-P. Pin and P. Rondard, “Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer”, *Nature Chemical Biology*, Vol. 11, No. 2, pp. 134–140, 2015.
14. Tobin, A., “G-protein-coupled receptor phosphorylation: where, when and by whom”, *British Journal of Pharmacology*, Vol. 153, No. S1, pp. S167–S176, 2008.
15. Dror, R. O., D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani and D. E. Shaw, “Activation mechanism of the  $\beta$ 2-adrenergic receptor”, *Proceedings of the National Academy of Sciences*, Vol. 108, No. 46, pp. 18684–18689, 2011.
16. Huang, X.-P., J. Karpiak, W. K. Kroeze, H. Zhu, X. Chen, S. S. Moy, K. A. Saddoris, V. D. Nikolova, M. S. Farrell, S. Wang *et al.*, “Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65”, *Nature*, Vol. 527, No. 7579, pp. 477–483, 2015.

17. Zou, Y., J. Ewalt and H.-L. Ng, “Recent Insights from Molecular Dynamics Simulations for G Protein-Coupled Receptor Drug Discovery”, *International Journal of Molecular Sciences*, Vol. 20, No. 17, p. 4237, 2019.
18. Zhou, Q., D. Yang, M. Wu, Y. Guo, W. Guo, L. Zhong, X. Cai, A. Dai, W. Jang, E. I. Shakhnovich *et al.*, “Common activation mechanism of class A GPCRs”, *Elife*, Vol. 8, p. e50279, 2019.
19. Katritch, V., V. Cherezov and R. C. Stevens, “Diversity and modularity of G protein-coupled receptor structures”, *Trends in Pharmacological Sciences*, Vol. 33, No. 1, pp. 17–27, 2012.
20. Venkatakrisnan, A., X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler and M. M. Babu, “Molecular signatures of G-protein-coupled receptors”, *Nature*, Vol. 494, No. 7436, pp. 185–194, 2013.
21. Ngo, T., A. V. Ilatovskiy, A. G. Stewart, J. L. Coleman, F. M. McRobb, R. P. Riek, R. M. Graham, R. Abagyan, I. Kufareva and N. J. Smith, “Orphan receptor ligand discovery by pickpocketing pharmacological neighbors”, *Nature Chemical Biology*, Vol. 13, No. 2, pp. 235–242, 2017.
22. Bahar, I. and A. Rader, “Coarse-grained normal mode analysis in structural biology”, *Current Opinion in Structural Biology*, Vol. 15, No. 5, pp. 586–592, 2005.
23. Bahar, I., “On the functional significance of soft modes predicted by coarse-grained models for membrane proteins”, *Journal of General Physiology*, Vol. 135, No. 6, pp. 563–573, 2010.
24. Zheng, W., B. R. Brooks and G. Hummer, “Protein conformational transitions explored by mixed elastic network models”, *Proteins: Structure, Function, and Bioinformatics*, Vol. 69, No. 1, pp. 43–57, 2007.
25. Ma, J., “Usefulness and limitations of normal mode analysis in modeling dynamics

- of biomolecular complexes”, *Structure*, Vol. 13, No. 3, pp. 373–380, 2005.
26. Atilgan, A. R., S. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, “Anisotropy of fluctuation dynamics of proteins with an elastic network model”, *Biophysical Journal*, Vol. 80, No. 1, pp. 505–515, 2001.
  27. Doruker, P., A. R. Atilgan and I. Bahar, “Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to  $\alpha$ -amylase inhibitor”, *Proteins: Structure, Function, and Bioinformatics*, Vol. 40, No. 3, pp. 512–524, 2000.
  28. Kuriyan, J., G. A. Petsko, R. M. Levy and M. Karplus, “Effect of anisotropy and anharmonicity on protein crystallographic refinement: an evaluation by molecular dynamics”, *Journal of Molecular Biology*, Vol. 190, No. 2, pp. 227–254, 1986.
  29. Haliloglu, T. and I. Bahar, “Adaptability of protein structures to enable functional interactions and evolutionary implications”, *Current Opinion in Structural Biology*, Vol. 35, pp. 17–23, 2015.
  30. Karplus, M. and G. A. Petsko, “Molecular dynamics simulations in biology”, *Nature*, Vol. 347, No. 6294, pp. 631–639, 1990.
  31. Salsbury Jr, F. R., “Molecular dynamics simulations of protein dynamics and their relevance to drug discovery”, *Current Opinion in Pharmacology*, Vol. 10, No. 6, pp. 738–744, 2010.
  32. Adcock, S. A. and J. A. McCammon, “Molecular dynamics: survey of methods for simulating the activity of proteins”, *Chemical Reviews*, Vol. 106, No. 5, pp. 1589–1615, 2006.
  33. Dolinsky, T. J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, “PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations”, *Nucleic Acids Research*,

Vol. 35, No. suppl2, pp. W522–W525, 2007.

34. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee *et al.*, “A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations”, *Journal of Computational Chemistry*, Vol. 24, No. 16, pp. 1999–2012, 2003.
35. Salomon-Ferrer, R., D. A. Case and R. C. Walker, “An overview of the Amber biomolecular simulation package”, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Vol. 3, No. 2, pp. 198–210, 2013.
36. Miao, Y., S. E. Nichols, P. M. Gasper, V. T. Metzger and J. A. McCammon, “Activation and dynamic network of the M2 muscarinic receptor”, *Proceedings of the National Academy of Sciences*, Vol. 110, No. 27, pp. 10982–10987, 2013.
37. Maeda, S., Q. Qu, M. J. Robertson, G. Skiniotis and B. K. Kobilka, “Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes”, *Science*, Vol. 364, No. 6440, pp. 552–557, 2019.
38. Harvey, R. D., “Muscarinic receptor agonists and antagonists: effects on cardiovascular function”, *Muscarinic Receptors*, pp. 299–316, Springer, 2012.
39. Ockenga, W., S. Kühne, S. Bocksberger, A. Banning and R. Tikkanen, “Non-neuronal functions of the m2 muscarinic acetylcholine receptor”, *Genes*, Vol. 4, No. 2, pp. 171–197, 2013.
40. Kruse, A. C., A. M. Ring, A. Manglik, J. Hu, K. Hu, K. Eitel, H. Hübner, E. Pardon, C. Valant, P. M. Sexton *et al.*, “Activation and allosteric modulation of a muscarinic acetylcholine receptor”, *Nature*, Vol. 504, No. 7478, pp. 101–106, 2013.
41. Ballesteros, J. A. and H. Weinstein, “[19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function re-

- lations in G protein-coupled receptors”, *Methods in Neurosciences*, Vol. 25, pp. 366–428, Elsevier, 1995.
42. Liu, W., *Exploring Functional Motions of Muscarinic Acetylcholine M2 Receptor*, M.Sc. Thesis, Bogazici University, 2018.
43. Haga, K., A. C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. I. Weis, T. Okada, B. K. Kobilka, T. Haga *et al.*, “Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist”, *Nature*, Vol. 482, No. 7386, pp. 547–551, 2012.
44. Jazayeri, A., A. S. Doré, D. Lamb, H. Krishnamurthy, S. M. Southall, A. H. Baig, A. Bortolato, M. Koglin, N. J. Robertson, J. C. Errey *et al.*, “Extra-helical binding site of a glucagon receptor antagonist”, *Nature*, Vol. 533, No. 7602, pp. 274–277, 2016.
45. Miao, Y., S. E. Nichols and J. A. McCammon, “Mapping of Allosteric Druggable Sites in Activation-Associated Conformers of the M2 Muscarinic Receptor”, *Chemical Biology & Drug Design*, Vol. 83, No. 2, pp. 237–246, 2014.
46. Barnes, P., “ $\beta$ -Adrenoceptors on smooth muscle, nerves and inflammatory cells”, *Life Sciences*, Vol. 52, No. 26, pp. 2101–2109, 1993.
47. Cherezov, V., D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka *et al.*, “High-resolution crystal structure of an engineered human  $\beta$ 2-adrenergic G protein-coupled receptor”, *Science*, Vol. 318, No. 5854, pp. 1258–1265, 2007.
48. Rosenbaum, D. M., S. G. Rasmussen and B. K. Kobilka, “The structure and function of G-protein-coupled receptors”, *Nature*, Vol. 459, No. 7245, pp. 356–363, 2009.
49. Rasmussen, S. G., H.-J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae,

- B. T. DeVree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka *et al.*, “Structure of a nanobody-stabilized active state of the  $\beta$  2 adrenoceptor”, *Nature*, Vol. 469, No. 7329, pp. 175–180, 2011.
50. Ballesteros, J. A., A. D. Jensen, G. Liapakis, S. G. Rasmussen, L. Shi, U. Gether and J. A. Javitch, “Activation of the  $\beta$ 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6”, *Journal of Biological Chemistry*, Vol. 276, No. 31, pp. 29171–29177, 2001.
51. Hanson, M. A., V. Cherezov, M. T. Griffith, C. B. Roth, V.-P. Jaakola, E. Y. Chien, J. Velasquez, P. Kuhn and R. C. Stevens, “A specific cholesterol binding site is established by the 2.8 Å structure of the human  $\beta$ 2-adrenergic receptor”, *Structure*, Vol. 16, No. 6, pp. 897–905, 2008.
52. Rasmussen, S. G., B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski *et al.*, “Crystal structure of the  $\beta$  2 adrenergic receptor–Gs protein complex”, *Nature*, Vol. 477, No. 7366, pp. 549–555, 2011.
53. Fiser, A. and A. Sali, “ModLoop: automated modeling of loops in protein structures”, *Bioinformatics*, Vol. 19, No. 18, pp. 2500–2501, 2003.
54. DeLano, W. L., “Pymol: An open-source molecular graphics tool”, *CCP4 Newsletter on Protein Crystallography*, Vol. 40, No. 1, pp. 82–92, 2002.
55. Balabin, I. A., W. Yang and D. N. Beratan, “Coarse-grained modeling of allosteric regulation in protein receptors”, *Proceedings of the National Academy of Sciences*, Vol. 106, No. 34, pp. 14253–14258, 2009.
56. Bhattacharya, S. and N. Vaidehi, “Differences in allosteric communication pipelines in the inactive and active states of a GPCR”, *Biophysical Journal*, Vol. 107, No. 2, pp. 422–434, 2014.

57. Roth, C. B., M. A. Hanson and R. C. Stevens, “Stabilization of the human  $\beta$ 2-adrenergic receptor TM4–TM3–TM5 helix interface by mutagenesis of Glu1223.41, a critical residue in GPCR structure”, *Journal of Molecular Biology*, Vol. 376, No. 5, pp. 1305–1319, 2008.
58. Zocher, M., C. Zhang, S. G. Rasmussen, B. K. Kobilka and D. J. Müller, “Cholesterol increases kinetic, energetic, and mechanical stability of the human  $\beta$ 2-adrenergic receptor”, *Proceedings of the National Academy of Sciences*, Vol. 109, No. 50, pp. E3463–E3472, 2012.
59. Manna, M., M. Niemelä, J. Tynkkynen, M. Javanainen, W. Kulig, D. J. Müller, T. Róg and I. Vattulainen, “Mechanism of allosteric regulation of  $\beta$ 2-adrenergic receptor by cholesterol”, *Elife*, Vol. 5, p. e18432, 2016.
60. Kitabgi, P., “Targeting neurotensin receptors with agonists and antagonists for therapeutic purposes.”, *Current Opinion in Drug Discovery & Development*, Vol. 5, No. 5, p. 764, 2002.
61. Kinkead, B. and C. B. Nemeroff, “Neurotensin: an endogenous antipsychotic?”, *Current Opinion in Pharmacology*, Vol. 2, No. 1, pp. 99–103, 2002.
62. Vincent, J.-P., J. Mazella and P. Kitabgi, “Neurotensin and neurotensin receptors”, *Trends in Pharmacological Sciences*, Vol. 20, No. 7, pp. 302–309, 1999.
63. Bissette, G., C. B. Nemeroff, P. T. Loosen, A. J. Prange and M. A. Lipton, “Hypothermia and intolerance to cold induced by intracisternal administration of the hypothalamic peptide neurotensin”, *Nature*, Vol. 262, No. 5569, pp. 607–609, 1976.
64. Carraway, R. E. and A. M. Plona, “Involvement of neurotensin in cancer growth: evidence, mechanisms and development of diagnostic tools”, *Peptides*, Vol. 27, No. 10, pp. 2445–2460, 2006.
65. Griebel, G. and F. Holsboer, “Neuropeptide receptor ligands as drugs for psychi-

- atric diseases: the end of the beginning?”, *Nature Reviews Drug discovery*, Vol. 11, No. 6, pp. 462–478, 2012.
66. Schimpff, R., C. Avar, G. Fenelon, A. Lhiaubet, L. Tenneze, M. Vidailhet and W. Rostene, “Increased plasma neurotensin concentrations in patients with Parkinson’s disease”, *Journal of Neurology, Neurosurgery & Psychiatry*, Vol. 70, No. 6, pp. 784–786, 2001.
67. Egloff, P., M. Hillenbrand, C. Klenk, A. Batyuk, P. Heine, S. Balada, K. M. Schlinkmann, D. J. Scott, M. Schütz and A. Plückthun, “Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in *Escherichia coli*”, *Proceedings of the National Academy of Sciences*, Vol. 111, No. 6, pp. E655–E662, 2014.
68. Krumm, B. E., S. Lee, S. Bhattacharya, I. Botos, C. F. White, H. Du, N. Vaidehi and R. Grisshammer, “Structure and dynamics of a constitutively active neurotensin receptor”, *Scientific Reports*, Vol. 6, p. 38564, 2016.
69. White, J. F., N. Noinaj, Y. Shibata, J. Love, B. Kloss, F. Xu, J. Gvozdenovic-Jeremic, P. Shah, J. Shiloach, C. G. Tate *et al.*, “Structure of the agonist-bound neurotensin receptor”, *Nature*, Vol. 490, No. 7421, pp. 508–513, 2012.
70. Madeira, F., Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. Tivey, S. C. Potter, R. D. Finn *et al.*, “The EMBL-EBI search and sequence analysis tools APIs in 2019”, *Nucleic Acids Research*, Vol. 47, No. W1, pp. W636–W641, 2019.