



UNIVERSITEIT VAN PRETORIA  
UNIVERSITY OF PRETORIA  
YUNIBESITHI YA PRETORIA

Molecular modelling and crystallographic investigation of novel  
synthetic cathinone  
2-methyl-4'-(methylthio)-2-morpholinopropiophenone  
(MMMP)

by

*Lanzo Jacques van Heerden*

Submitted in fulfilment of the requirements for the degree  
Master of Science in Chemistry

In the Department of Chemistry  
Faculty of Natural and Agricultural Sciences  
University of Pretoria  
Pretoria  
March 2024

Supervisors:  
*Dr Cara Slabbert*  
*Dr Krishna Govender*

**Declaration:**

I, Lanzo Jacques van Heerden, declare that the dissertation, which I hereby submit for the degree Master of Science in Chemistry at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:  \_\_\_\_\_

DATE: 2024-03-18

**Dedication:**

I dedicate this work to my partner Altia Peenz and to my parents, Jacques and Cathmarlene van Heerden.

## **Acknowledgments:**

To have achieved this milestone in my life, I would like to express my sincere gratitude to the following people:

- My supervisor - Dr Cara Slabbert
- My Co-supervisor - Dr Krishna Govender - This work would not have been possible had he not agreed to take over this project at such short notice. I will forever be indebted to you.
- Dr Frederick Malan for his assistance in elucidating the crystal structure of MMMP using single-crystal X-ray crystallography, all his patience and guidance.
- Centre for High Performance Computing (CHPC) for granting the use of the Lengau cluster to run Molecular Dynamics simulations.
- My partner for never doubting me and encouraging me to push through even when I faced obstacles that should have discouraged me from completing this project.
- My family for their unconditional love and support in achieving my goals.
- University of Pretoria for granting me this opportunity to further my education.

## **Abstract:**

2-Methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP) is a highly modified synthetic cathinone derived from natural occurring cathinone, the main psychoactive alkaloid present in *Catha edulis* Forsk. or khat. Synthetic cathinones (SCs) belong to the drug class of new psychoactive substances (NPS) which consists of analogues of commonly known abused drugs designed to mimic the psychoactive properties of illicit drugs to circumvent current drug legislations. From a structural perspective, cathinone is a  $\beta$ -keto analogue of amphetamine and SCs are often referred to as “bk-amphetamines”. SCs are capable of evoking psychoactive effects similar to that of amphetamines and cocaine increasing their popularity amongst users as “legal highs”. MMMP itself is commercially available and readily used in the polymer and printing industry in the fixing of thin films, plastics and inks as a Type 1 fragmenting photoinitiator. It was hypothesised that MMMP might have psychoactive properties due to its cathinone origin and isolation in confiscated drug samples. This hypothesis was investigated through molecular modelling techniques to determine if MMMP was capable of interacting with 25 protein targets consisting of 24 monoamine receptors and a transporter which act as common targets for psychoactive substances. From the initial molecular docking results two protein targets, serotonin-1A (SER1A) and serotonin transporter (SERT), were identified as the most likely to form a complex with MMMP. Further analysis through Molecular Dynamics (MD) simulations to evaluate the stability and movement of these complexes revealed that MMMP has a higher probability of interacting and forming a strong complex with the SERT binding pocket, but a weaker complex is formed with SER1A. The crystal structure of MMMP was successfully elucidated through single crystal X-ray diffraction (SCXRD). Analysis of the crystal structure revealed that MMMP crystallises in the non-centrosymmetric space group, *Pca2<sub>1</sub>*. The unit cell consist of four crystallographically independent conformers with each conformer interacting with itself, with conformers within the same unit cell as well as with conformers in adjacent unit cells through a series of intra- and intermolecular interactions. Three symmetry elements, two glide planes and a screw axis, were also identified.

## **Key terms:**

2-methyl-4'-(methylthio)-2-morpholinopropiophenone, MMMP, Igacure 907, MTMP, MMTMP, Synthetic cathinones, Molecular modelling, Molecular docking, Glide, Molecular Dynamics, MD, Single crystal X-ray diffraction, SCXRD, Monoamine receptors, Monoamine transporters, DAT, SERT, NET

## **List of Abbreviations in order of appearance:**

| Abbreviation     | Expansion   |
|------------------|---|
| NPS              | New psychoactive substances                             |
| SCs              | Synthetic Cathinones                                    |
| UNODC            | United Nations Office on Drugs and Crime                |
| MDMA             | 3,4-methylenedioxymethamphetamine                       |
| <i>C. edulis</i> | <i>Catha edulis</i> Forsk.                              |
| CNS              | Central nervous system                                  |
| EMCDDA           | European Monitoring Centre for Drugs and Drug Addiction |
| DAT              | Dopamine transporter                                    |
| SERT             | Serotonin transporter                                   |
| NET              | Norepinephrine transporter                              |
| DA               | Dopamine  |
| 5-HT or SER      | Serotonin   |
| NE               | Norepinephrine  |
| MD               | Molecular Dynamics                                      |
| DOPA             | Dihydroxyphenylalanine                                  |
| MMMP             | 2-Methyl-4'-(methylthio)-2-morpholinopropiophenone      |
| RMSD             | Root-mean-square deviation                              |
| BBB              | Blood-brain barrier                                     |
| ADHD             | Attention-deficit/hyperactivity disorder                |
| VMAT             | Vesicular monoamine transporter                         |
| VMAT1            | Vesicular monoamine transporter 1                       |
| VMAT2            | Vesicular monoamine transporter 2                       |
| GPCR             | G protein-coupled receptor                              |
| DARs             | Dopamine receptors                                      |
| GRK              | G protein-coupled receptor kinase                       |
| ICL3             | Intracellular loop 3                                    |
| GTP              | Guanosine triphosphate                                  |
| GDP              | Guanosine diphosphate                                   |
| ATP              | Adenosine triphosphate                                  |
| cAMP             | Cyclic adenosine monophosphate                          |
| PKA              | Protein kinase A  |
| MAPK/ERK         | Mitogen-activated protein kinases                       |
| RGS              | Regulator of the G-protein signalling                   |

|                  |  |
|------------------|--|
| 5-HTR            | Serotonin receptors                                |
| PLC              | Phospholipase C                                    |
| PI3 kinase       | Phosphoinositide 3-kinase                          |
| GABA             | $\gamma$ -aminobutyric acid                        |
| PIP <sub>2</sub> | Phosphatidylinositol 4,5-bisphosphate              |
| IP <sub>3</sub>  | Inositol 1,4,5-triphosphate                        |
| DAG              | Diacylglycerol                                     |
| ER               | Endoplasmic reticulum                              |
| PKC              | Protein kinase C                                   |
| PLA2             | Phospholipase A <sub>2</sub>                       |
| PNS              | Peripheral nervous system                          |
| cGMP             | Cyclic guanosine monophosphate                     |
| AR               | Adrenergic receptors                               |
| SLC6             | Solute carrier 6                                   |
| TM               | Transmembrane spanning domains                     |
| S1               | Orthosteric or Primary substrate binding site      |
| DOPAC            | 3,4-dihydroxyphenylacetic acid                     |
| SAPS             | South African Police Services                      |
| RASFF            | Rapid Alert System for Food and Feed               |
| CADD             | Computer aided drug design                         |
| QM               | Quantum Mechanics                                  |
| MM               | Molecular Mechanics                                |
| MD               | Molecular Dynamics                                 |
| OPLS             | Optimised potentials for liquid simulations        |
| PDB              | Protein Data Bank                                  |
| UFF              | Universal force field                              |
| MMFF             | Merck molecular force fields                       |
| RMSD             | Root Mean Square Deviation                         |
| MM-GBSA          | Molecular Mechanics Generalised-Born Surface Area  |
| MM-PBSA          | Molecular Mechanics Poisson-Boltzmann Surface Area |
| Glide XP         | Glide Extra Precision                              |
| SER1E            | Serotonin receptor 1E                              |
| NER2A            | Norepinephrine receptor $\alpha$ -2A               |
| NER2B            | Norepinephrine $\alpha$ -2B                        |
| SER1A            | Serotonin receptor 1A                              |

|                 |                                      |
|-----------------|--------------------------------------|
| SER2B           | Serotonin receptor 2B                |
| NER1A           | Norepinephrine receptor 1A           |
| D1              | Dopamine receptor 1                  |
| D2              | Dopamine receptor 2                  |
| D3              | Dopamine receptor 3                  |
| D4              | Dopamine receptor 4                  |
| D5              | Dopamine receptor 5                  |
| SER1D           | Serotonin receptor 1D                |
| SER2C           | Serotonin receptor 2C                |
| NER2C           | Norepinephrine receptor $\alpha$ -2C |
| SER             | Serotonin                            |
| C $\alpha$      | Alpha-carbon                         |
| NMR             | Nuclear magnetic resonance           |
| IR spectroscopy | Infrared spectroscopy                |
| MS              | Mass spectroscopy                    |
| SCXRD           | Single crystal X-ray diffraction     |
| GooF            | Goodness-of-fit                      |
| F <sub>o</sub>  | Observed electron density            |
| F <sub>c</sub>  | Calculated electron density          |



## List of Figures

|  |    |
|--|----|
| <b>Figure 1.1</b> - Structure of 2-methyl-4'-(methylthio)-2-morpholinopropiophenone.....   | 3  |
| <b>Figure 2.1</b> - Line drawing of <i>Catha edulis</i> Forsk. ....  | 10 |
| <b>Figure 2.2</b> - Chemical structures of <b>a.</b> (–)-cathinone [(2S)-2-amino-1-phenylpropan-1-one], <b>b.</b> (+)-cathine [(1S,2S)-2-amino-1-phenylpropan-1-ol] and <b>c.</b> (–)-norephedrine [(1R,2S)-2-amino-1-phenylpropan-1-ol]. ....   | 12 |
| <b>Figure 2.3</b> - Amphetamines and their related $\beta$ -keto analogues: (1,2) Amphetamine-Cathinone, (3,4) Methamphetamine-Methcathinone and (5,6) Ecstasy (MDMA)-Methylone.....   | 16 |
| <b>Figure 2.4</b> - General structure of synthetic cathinones. R1, R2, R3 and R4 represent position at which substitutions can be made to obtain different cathinone derivatives. R1 represent substitution at the aromatic ring, R2 at the $\alpha$ -carbon, R3 and R4 substitutions at the amino group. .... | 20 |
| <b>Figure 2.5</b> - Dopamine receptor signal transducing pathway. <sup>[188]</sup> .....   | 32 |
| <b>Figure 2.6</b> - General representation of monoamine transporters 12-transmembrane domains, extra- and intracellular loops and the location of the N- and C-termini. <sup>[223]</sup> ....  | 40 |
| <b>Figure 2.7</b> - Alternating access mechanism for the translocation of monoamine substrates and ions. <sup>[225]</sup> ....   | 41 |
| <b>Figure 2.8</b> - Photocleavage of MMMP. ....  | 51 |
| <b>Figure 2.9</b> - Structures of (24) amphetamine, (25) phentermine, (26) 4-MTA and (27) methedrone.....  | 54 |
| <b>Figure 3.1</b> - Illustration of the Lennard-Jones potential plot. ....   | 62 |
| <b>Figure 3.2</b> - Periodic boundary condition illustration. ....   | 66 |
| <b>Figure 3.3</b> - Basic MD simulation algorithm. With: $E_{pot}$ - potential energy; $t$ - simulation time; $dt$ - simulation timestep. For $N$ simulated atoms ( $i$ ): $a$ -acceleration; $v$ – velocity; $m$ – mass of the atom; $x_i$ – atom coordinates; $F_i$ – force component.....                   | 67 |
| <b>Figure 4.1</b> - Illustrations of (a) smallest [Norepinephrine 2B, 0.36 Å] and (b) largest RMSD [Serotonin 1E, .....]   | 74 |
| <b>Figure 4.2</b> - Deprotonated and protonated states of MMMP. ....   | 76 |
| <b>Figure 4.3</b> - Labelled serotonin molecule. ....  | 80 |
| <b>Figure 4.4</b> - (a) Interaction diagrams for SER1A-SER complex. (b) Simplified view of interactions with some residues hidden. ....  | 82 |

|   |     |
|---|-----|
| <b>Figure 4.5</b> - Labelled MMMP molecule. ....  | 83  |
| <b>Figure 4.6</b> - (a) Interaction diagrams for SER1A-MMMP complex. (b) Simplified view of interactions with some residues hidden. ....  | 85  |
| <b>Figure 4.7</b> - (a) Interaction diagrams for SER1A-MMMP complex. (b) Simplified view of interactions with some residues hidden. ....  | 88  |
| <b>Figure 4.8</b> - (a) Interaction diagrams for SER1A-MMMP complex. (b) Simplified view of interactions with some residues hidden. ....  | 91  |
| <br>  |     |
| <b>Figure 5.1</b> - Diagram of Bragg's Law. ....  | 105 |
| <b>Figure 5.2</b> - Summary of the technique used in X-ray crystallography. <sup>[395]</sup> .....  | 106 |
| <b>Figure 5.3</b> - Summary of the procedure followed by SHELXT when solving a crystal structure. The intrinsic phasing structure solution, the space group refinement and isotropic refinement are performed in parallel. $G_0$ and $G_c$ are the modified observed and calculated structure factors and $\varphi c$ is the phase of $G_c$ . FFT is the Fast Fourier transform. <sup>[417]</sup> ..... | 110 |
| <b>Figure 5.4</b> - Overview of the different types of centring. (a) Primitive, (b) Single-sided face centring, (c) Body-centred centring and (d) All-sided face centring. ....   | 113 |
| <b>Figure 5.5</b> - The 14 Bravais lattices arising from centring. ....   | 113 |
| <b>Figure 5.6</b> - Glide plane a orientated perpendicular to the drawing plane with a car motif. The car is reflected, thereafter translated by 1/2 a unit cell in the a direction. ....   | 115 |
| <b>Figure 5.7</b> - $3_1$ Screw axis illustration. ....   | 116 |
| <br>  |     |
| <b>Figure 6.1</b> - Nomenclature of MMMP conformers. ....   | 120 |
| <b>Figure 6.2</b> - Arrangement of four MMMP conformers in a unit cell. ....  | 120 |
| <b>Figure 6.3</b> - Intramolecular hydrogen bonding interactions. ....  | 123 |
| <b>Figure 6.4</b> - Intermolecular interactions in the unit cell. ....  | 125 |
| <b>Figure 6.5</b> - Dipole-dipole intermolecular interactions within and between unit cells. Interactions shown between molecules are within the unit cell, while interactions shown to only a single atom can be considered as interactions occurring with neighbouring unit cells. ....   | 127 |
| <b>Figure 6.6</b> (a) - Dipole-dipole interaction between the centroid of conformer 4 and a hydrogen donor of conformer 2. (b) Both interactions present in packed unit cells. ....   | 130 |
| <b>Figure 6.7</b> - Types of ring interactions. Sandwich (left), T-stacking (centre) and offset stacking (right). ....  | 130 |
| <b>Figure 6.8</b> - Parameters between centroids. ....  | 131 |

|   |     |
|---|-----|
| <b>Figure 6.9 (a)-(f)</b> - Intermolecular aromatic interactions. (a)1x2x1 packing of the unit cell; (b) Isolation of the four conformers shown in the black box in (a); (c)-(d) Conformer 1 and 2 packed along the b-axis and (e)-(f) Conformer 3 and 4 packed along the b-axis..... | 132 |
| <b>Figure 6.10</b> - Intermolecular van der Waals interaction. ....   | 133 |
| <b>Figure 6.11</b> - Illustration of the 2-fold screw axis viewed along b-axis in a 1x1x1 packed cell. ....   | 134 |
| <b>Figure 6.12</b> - Illustration of the glide planes, blue horizontal and vertical lines, viewed along the c-axis in a 1x1x1 packed cell. ....   | 134 |
| <b>Figure 6.13</b> - 2x2x2 Packing of MMMP viewed along the a-axis.....   | 135 |
| <b>Figure 6.14</b> - 2x2x2 Packing of MMMP viewed along the b-axis.....   | 135 |
| <b>Figure 6.15</b> - 2x2x2 Packing of MMMP viewed along the c-axis.....   | 136 |

## **List of Tables**

|  |     |
|--|-----|
| <b>Table 2.1</b> - Structures, common- and IUPAC names of a few common synthetic cathinones grouped into four categories. ....   | 21  |
| <b>Table 2.2</b> - Summary of monoamines receptor families, subtypes and their primary signalling pathways. ....   | 38  |
| <b>Table 2.3</b> - Comparison between transporter substrates and blockers. <sup>[144]</sup> .....  | 43  |
| <b>Table 2.4</b> - Synthetic Cathinones and their monoamine transporter targets. ....  | 48  |
| <b>Table 2.5</b> - Synthetic Cathinones and their monoamine receptor targets.....  | 49  |
| <br>   |     |
| <b>Table 4.1</b> - RMSD, Redocking and MMMP docking scores. ....   | 75  |
| <b>Table 4.2</b> - SER1A-SER interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below.....  | 81  |
| <b>Table 4.3</b> - SER1A-MMMP interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below..... | 84  |
| <b>Table 4.4</b> - SERT-SER interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below. ...   | 87  |
| <b>Table 4.5</b> - SERT-MMMP interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below.....  | 90  |
| <br>   |     |
| <b>Table 5.1</b> - The Seven Primitive Crystal Systems.....  | 111 |
| <br>   |     |
| <b>Table 6.1</b> - Crystal data for MMMP crystal structure. ....   | 118 |
| <b>Table 6.2</b> - The four MMMP conformers. ....  | 122 |
| <b>Table 6.3</b> - Intramolecular hydrogen bonding.....  | 124 |
| <b>Table 6.4</b> - Intermolecular hydrogen bonding interactions. ....  | 126 |
| <b>Table 6.5</b> - Other intermolecular dipole-dipole interactions. ....   | 127 |
| <b>Table 6.6</b> - Intermolecular aromatic interactions. ....  | 131 |
| <b>Table 6.7</b> - Intermolecular van der Waals interactions. ....   | 133 |
| <b>Table 6.8</b> - Symmetry elements present in the crystal structure. ....  | 134 |

## **List of Graphs**

|  |    |
|--|----|
| <b>Graph 4.1</b> – <i>Glide</i> docking scores for redocked, deprotonated- and protonated MMMP. ....   | 77 |
| <b>Graph 4.2</b> - Absolute difference between redocked and best protonated/deprotonated MMMP docking score. <sup>a</sup> .....  | 78 |
| <b>Graph 4.3</b> - A plot of the ligand's (SER) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SER1A during the course of a 200 ns MD simulation at 310 K.....  | 92 |
| <b>Graph 4.4</b> - A plot of the ligand's (MMMP) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SER1A during the course of a 200 ns MD simulation at 310 K..... | 93 |
| <b>Graph 4.5</b> - A plot of the ligand's (SER) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SERT during the course of a 200 ns MD simulation at 310 K.....   | 96 |
| <b>Graph 4.6</b> - A plot of the ligand's (SER) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SERT during the course of a 250 ns MD simulation at 310 K.....   | 97 |
| <b>Graph 4.7</b> - A plot of the ligand's (MMMP) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SERT during the course of a 200 ns MD simulation at 310 K.....  | 98 |

# Table of Contents

|   |             |
|---|-------------|
| <i>Declaration:</i> .....   | <i>i</i>    |
| <i>Dedication:</i> .....  | <i>ii</i>   |
| <i>Acknowledgments:</i> .....                                       | <i>iii</i>  |
| <i>Abstract:</i> .....  | <i>iv</i>   |
| <i>List of Abbreviations in order of appearance:</i> .....          | <i>v</i>    |
| <i>List of Figures</i> .....  | <i>viii</i> |
| <i>List of Tables</i> .....   | <i>xi</i>   |
| <i>List of Graphs</i> .....   | <i>xii</i>  |
| <i>Table of Contents</i> .....                                      | <i>1</i>    |
| <b>Chapter 1: Introduction</b> .....                                | <b>1</b>    |
| <b>1.1 Background information</b> .....                             | <b>1</b>    |
| <b>1.2 Problem statement</b> .....                                  | <b>4</b>    |
| <b>1.3 Dissertation statement</b> .....                             | <b>5</b>    |
| <b>1.4 Research aim and objectives</b> .....                        | <b>5</b>    |
| <b>1.5 Research significance</b> .....                              | <b>6</b>    |
| <b>1.6 Delineations and limitations</b> .....                       | <b>7</b>    |
| <b>1.7 Underlying assumptions</b> .....                             | <b>8</b>    |
| <b>1.8 Definition of terms</b> .....                                | <b>8</b>    |
| <b>1.9 Brief chapters overview</b> .....                            | <b>9</b>    |
| <b>Chapter 2: Literature review and Theory</b> .....                | <b>10</b>   |
| <b>2.1 Introduction and history of Khat</b> .....                   | <b>11</b>   |
| 2.1.1 Colloquial names for Khat .....                               | 11          |
| 2.1.2 Ethnomedicinal uses of <i>C. edulis</i> .....                 | 11          |
| 2.1.3 Chemical composition of khat.....                             | 12          |
| 2.1.4 Discovery of the psychoactive compound in Khat .....          | 13          |
| 2.1.5 Stability of cathinone .....                                  | 13          |
| 2.1.6 Pharmacokinetics .....  | 14          |
| 2.1.7 Pharmacodynamics .....  | 15          |
| <b>2.2 Introduction and history of Synthetic Cathinones</b> .....   | <b>16</b>   |
| 2.2.1 Background history of synthetic cathinones.....               | 16          |
| 2.2.2 Common names for synthetic cathinones .....                   | 19          |
| 2.2.3 Chemistry of synthetic cathinones.....                        | 19          |
| 2.2.4 Structures and names of common synthetic cathinones .....     | 21          |
| 2.2.5 Medicinal uses.....   | 23          |
| 2.2.6 Pharmacokinetics .....  | 24          |
| 2.2.7 Pharmacodynamics .....  | 26          |
| <b>2.3 Neuroscience Overview</b> .....                              | <b>28</b>   |
| 2.3.1 Overview of the neurotransmitter cycle.....                   | 28          |
| 2.3.2 Monoamines (Dopamine, Serotonin and Norepinephrine).....      | 29          |
| 2.3.3 Receptors for Dopamine, Serotonin and Norepinephrine.....     | 30          |
| 2.3.4 Transporters for Dopamine, Serotonin and Norepinephrine ..... | 39          |
| 2.3.5 Transporter Substrates vs Blockers.....                       | 42          |

|  |  |            |
|--|--|------------|
| <b>2.4</b>   | <b>Neuropharmacology of Khat and Synthetic Cathinones</b> .....        | <b>44</b>  |
| 2.4.1  | Neuropharmacology of Khat .....  | 44         |
| 2.4.2  | Neuropharmacology of Synthetic cathinones .....                        | 46         |
| <b>2.5</b>   | <b>2-Methyl-4'-(methylthio)-2-morpholinopropiophenone - MMMP</b> ..... | <b>50</b>  |
| 2.5.1  | Background information of MMMP .....                                   | 50         |
| 2.5.2  | Common names of MMMP .....   | 51         |
| 2.5.3  | Chemistry of MMMP .....  | 51         |
| 2.5.4  | Pharmacokinetics .....   | 51         |
| 2.5.5  | Pharmacodynamics .....   | 53         |
| <b>Chapter 3: Molecular modelling</b> .....  |  | <b>55</b>  |
| <b>3.1</b>   | <b>Introduction</b> .....  | <b>55</b>  |
| <b>3.2</b>   | <b>Docking</b> .....   | <b>56</b>  |
| <b>3.3</b>   | <b>Molecular Mechanics Force Fields</b> .....                          | <b>60</b>  |
| 3.3.1  | Bonded interactions .....  | 60         |
| 3.3.2  | Non-bonded interactions .....  | 61         |
| 3.3.3  | Assumptions in Molecular Mechanics .....                               | 62         |
| <b>3.4</b>   | <b>Docking methods</b> .....   | <b>62</b>  |
| 3.4.1  | Ligand selection and preparation .....                                 | 64         |
| 3.4.2  | Receptor selection and preparation .....                               | 64         |
| <b>3.5</b>   | <b>Molecular Dynamics</b> .....  | <b>65</b>  |
| <b>3.6</b>   | <b>Solvent models</b> .....  | <b>68</b>  |
| <b>3.7</b>   | <b>Conclusion</b> .....  | <b>69</b>  |
| <b>Chapter 4: Molecular modelling of 2-methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP)</b> ..... |  | <b>70</b>  |
| <b>4.1</b>   | <b>Introduction</b> .....  | <b>70</b>  |
| <b>4.2</b>   | <b>Molecular modelling methodology</b> .....                           | <b>71</b>  |
| 4.2.1  | Protein target selection.....  | 71         |
| 4.2.2  | Validation through redocking .....                                     | 71         |
| 4.2.3  | Docking Protocol .....   | 72         |
| 4.2.4  | Molecular Dynamics.....  | 73         |
| <b>4.3</b>   | <b>Validation through redocking results and discussion</b> .....       | <b>73</b>  |
| <b>4.4</b>   | <b>Molecular docking results and discussion</b> .....                  | <b>76</b>  |
| 4.4.1  | SER1A-SER complex.....   | 80         |
| 4.4.2  | SER1A-MMMP complex .....   | 83         |
| 4.4.3  | SERT-SER complex .....   | 86         |
| 4.4.4  | SERT-MMMP complex.....   | 89         |
| <b>4.5</b>   | <b>MD simulations results and discussions</b> .....                    | <b>92</b>  |
| 4.5.1  | MD analysis of SER1A-SER and SER1A-MMMP complexes.....                 | 92         |
| 4.5.2  | MD analysis of SERT-SER and SERT-MMMP complexes .....                  | 96         |
| <b>4.6</b>   | <b>Sub-conclusion</b> .....  | <b>101</b> |
| <b>Chapter 5: Single crystal X-ray analysis</b> .....  |  | <b>103</b> |
| <b>5.1</b>   | <b>Introduction</b> .....  | <b>103</b> |
| <b>5.2</b>   | <b>General overview of X-ray crystallography</b> .....                 | <b>103</b> |
| <b>5.3</b>   | <b>Principles of X-ray crystallography</b> .....                       | <b>106</b> |
| 5.3.1  | Derived equations based on the fundamental facts.....                  | 106        |
| 5.3.2  | The Phase Problem .....  | 107        |

|  |  |                   |
|--|--|-------------------|
| <b>5.4</b>   | <b>The basic procedure of crystallography</b> .....                        | <b>108</b>        |
| 5.4.1  | Crystallisation.....   | 108               |
| 5.4.2  | Data collection.....   | 108               |
| 5.4.3  | Solving a crystal structure.....   | 109               |
| 5.4.4  | Analysis of the final crystal structure.....                               | 110               |
| <b>5.5</b>   | <b>Basic theory of crystallography</b> .....                               | <b>110</b>        |
| 5.5.1  | The unit cell.....   | 111               |
| 5.5.2  | The Seven Crystal Systems.....   | 111               |
| 5.5.3  | The 14 Bravais lattices.....   | 112               |
| 5.5.4  | The 32 Point Groups.....   | 113               |
| 5.5.5  | Symmetry in the plane.....   | 114               |
| 5.5.6  | Space groups.....  | 116               |
| <b>5.6</b>   | <b>Conclusion</b> .....  | <b>116</b>        |
| <b><i>Chapter 6: Crystallographic investigation of 2-methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP)</i></b> ..... |  | <b><i>117</i></b> |
| <b>6.1</b>   | <b>Methodology</b> .....   | <b>117</b>        |
| <b>6.2</b>   | <b>Discussion of the overall parameters of the crystal structure</b> ..... | <b>118</b>        |
| <b>6.3</b>   | <b>CIF check</b> .....   | <b>119</b>        |
| <b>6.4</b>   | <b>Nomenclature</b> .....  | <b>119</b>        |
| 6.4.1  | Nomenclature of MMMP conformers.....                                       | 119               |
| 6.4.2  | Numbering of the atoms in the unit cell.....                               | 120               |
| <b>6.5</b>   | <b>Structural chemistry</b> .....  | <b>122</b>        |
| 6.5.1  | Intramolecular interactions.....   | 123               |
| 6.5.2  | Intermolecular interactions.....   | 124               |
| <b>6.6</b>   | <b>Symmetry elements in 1x1x1 unit cell packing</b> .....                  | <b>133</b>        |
| <b>6.7</b>   | <b>2x2x2 Packing of MMMP unit cell</b> .....                               | <b>135</b>        |
| <b>6.8</b>   | <b>Sub-conclusion</b> .....  | <b>136</b>        |
| <b><i>Chapter 7: Conclusion and future research</i></b> .....  |  | <b><i>137</i></b> |
| <b>7.1</b>   | <b>Summary of results</b> .....  | <b>137</b>        |
| <b>7.2</b>   | <b>Conclusion</b> .....  | <b>138</b>        |
| <b>7.3</b>   | <b>Summary of contributions</b> .....                                      | <b>138</b>        |
| <b>7.4</b>   | <b>Implications of this research</b> .....                                 | <b>139</b>        |
| <b>7.5</b>   | <b>Future research</b> .....   | <b>139</b>        |
| <b><i>References</i></b> .....   |  | <b><i>141</i></b> |
| <b><i>Annexures</i></b> .....  |  | <b><i>166</i></b> |



# **Chapter 1: Introduction**

## ***1.1 Background information***

New psychoactive substances (NPS) is a relatively modern term used to describe analogues of commonly known abused drugs designed to mimic the psychoactive properties of controlled drugs. These include psychostimulants, cannabinoids, depressants (including opioids and benzodiazepines) and hallucinogens or novel synthesised drugs such as synthetic cathinones (SCs). NPS are commonly known as synthetic or designer drugs, or by the more popular misleading colloquial terms of “*legal highs*” or “*research chemicals*”. The definition for the classification of a NPS is not consistently defined or accepted, which leads to multiple ways in which a substance can be classified as an NPS.<sup>[1-3]</sup> The United Nations Office on Drugs and Crime (UNODC) defines NPS as substances of abuse which can be in pure form or part of a preparation, that are not controlled by the 1961 Single Convention on Narcotic Drugs or the 1971 Convention on Psychotropic Substances and pose a threat to public health.<sup>[4]</sup>

SCs are chemical analogues or derivatives of the natural occurring compound cathinone, which is the main psychoactive alkaloid present in *Catha edulis* Forsk. (*C. edulis* or khat) and belong to the drug class of NPS that is capable of evoking psychoactive effects similar to amphetamines (including methamphetamine and 3,4-methylenedioxymethamphetamine or MDMA) and cocaine in users. From a structural perspective cathinone is a  $\beta$ -keto analogue of amphetamine, therefore SCs are often referred to as “*bk-amphetamines*”. Amphetamine, methamphetamine, MDMA, cathinone and SCs all contain the same pharmacophore known as phenethylamine.<sup>[5]</sup>

SCs are commonly branded as “*bath salts*”, “*plant food*”, “*carpet cleaners*” or “*research chemicals*” and labelled “*not for human consumption*” or “*for research purposes only*” in an attempt by illicit drug manufactures to circumvent drug legislation. These drugs, as the common name might suggest, has no chemical or pharmacological relation to Epson “bath” salt, but they are chemically and pharmacologically similar to the phenethylamine containing drugs amphetamine and MDMA which are central nervous system (CNS) active drugs.<sup>[5]</sup> A number of factors have contributed to the popularity and increased abuse of SCs since the mid 2000’s such as its cocaine and amphetamine-like psychostimulant effects it elicits, affordability, easy accessibility and the misconception that SCs use are legal and safe.<sup>[6]</sup> Studies also suggested that the decrease in purity, high price and unavailability of amphetamine, MDMA and cocaine might have also contributed to the popularity of SCs.<sup>[7, 8]</sup> It is due to psychoactive similarities

between amphetamine, MDMA and cocaine that *bath salts* are marketed as alternatives for these common abused drugs.

The fast-evolving NPS market is of great concern to authorities as NPS are proliferating at an unprecedented rate posing a risk to public health and a challenge to drug policies. In 2023, the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) reported the seizure of 8.5 tonnes of NPS for the year 2021 and 5.1 tonnes in the previous year. Out of the 8.5 tonnes seized three cathinones (3-Chloromethcathinone, 4-Chloromethcathinone and 3-Methylmethcathinone) alone accounted for 47% of the total NPS seized. By the end of 2022, a total of 930 NPS were on the radar of the EMCDDA of which 41 were reported for the first time and 400 previously reported NPS were detected again. Since the peak introduction of 101 NPS reported for the first time in 2014 the amount year-on-year has been steadily decreasing and has remained below 60 introductions for the period 2017-2022.<sup>[9, 10]</sup>

The United Nations Office on Drugs and Crime (UNODC) report of 2023 paints a similar grim picture. From 2009-2022 preliminary data shows that about 1,184 NPS have been identified and reported to the UNODC early warning advisory on NPS, of which 618 are available on the global market. In 2020, a total of 77 first time reported NPS were identified, this number increased to 87 in 2021. In 2021, synthetic cannabinoid (324 substances) and SCs (201 substances) accounted for 46.6% of the total monitored NPS. After the rapid expansion of distinct NPS on the global market between 2009 and 2018 it would appear that the market has stabilised at around 550 substances actively available annually, around half of the total amount of NPS identified on the global drug market.<sup>[11, 12]</sup>

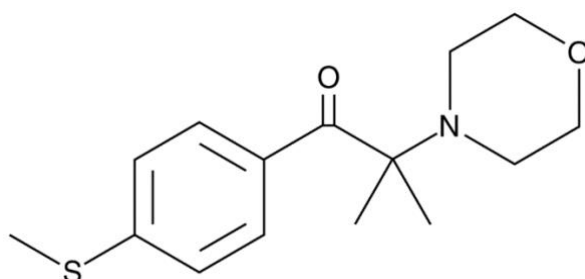
From a basic neurochemistry view SCs have two main routes of eliciting a psychoactive response. Firstly, by interacting with monoamine transporters for dopamine, serotonin and norepinephrine (DAT, SERT and NET respectively) to increase the concentration of monoamine neurotransmitters in the synaptic cleft and secondly by mimicking the monoamine neurotransmitters dopamine (DA), serotonin (5-HT) and norepinephrine (NE) to directly interact with monoamine receptors.<sup>[13]</sup>

In the first route the increase in the synaptic monoamine concentration can occur via two modes of action depending on the characteristics of the SC. SCs can act as either (1) cocaine-like “blockers” (or inhibitors) that interact by binding to transporters and inhibiting the transport of neurotransmitters from the extracellular space to the intracellular space, whereas (2)

amphetamine-like “substrates” (or releasers) interacts by binding to the transporter before getting translocated into the neuronal cytoplasm through the transporter channel. In the cytoplasm they trigger the efflux of intracellular neurotransmitters via a reverse transport mechanism (i.e. transporter-mediated release).<sup>[14, 15]</sup> Both mechanism achieve a common goal of increasing the neurotransmitter concentration in the synaptic cleft, which in turn interacts with their corresponding postsynaptic monoamine receptors which activates their downstream effectors and ultimately prevents the termination of neurotransmitter signalling. Increase in monoamine concentrations of DA, 5-HT and/or NE in the synaptic cleft results in the observed and desired psychoactive effects.

In the second route the SCs directly interact with monoamine receptors. To date little literature is available on SCs interaction with monoamine receptors. Only a few studies have tested SCs direct interaction with some monoamine receptors. The precise mechanism with which some SCs interact with monoamine receptors has not been determined or thoroughly investigated.<sup>[16-22]</sup> In theory the chemical structures of SCs and endogenously produced monoamines such as DA, 5-HT and NE all contain a common phenethylamine pharmacophore. This commonly shared pharmacophore could mimic these monoamines which activate their target receptors and lead to downstream effector activation.<sup>[23-30]</sup>

This study will be on a SC known as 2-methyl-4'-(methylthio)-2-morpholinopropiophenone or MMMP (**Figure 1.1**) which is a highly modified cathinone that is commercially available and readily used in the polymer and printing industry in the fixing of thin films, plastics and inks as a Type 1 fragmenting photoinitiator.<sup>[31]</sup>



*Figure 1.1 - Structure of 2-methyl-4'-(methylthio)-2-morpholinopropiophenone.*

To date MMMP has been detected in the United States of America, Australia<sup>[31]</sup> and in South Africa (personal communication) in mixtures containing known illicit substances. A few biological studies done by various authors have shown that not only has MMMP cytotoxic effects on normal human embryonic lung fibroblasts cells<sup>[32]</sup>, but has also been identified as an

endocrine disrupting compound that interact with estrogen receptors as an agonist.<sup>[33]</sup> In Sprague-Dawley rats acute oral toxicity, possible damage to fertility and the unborn offspring have been reported.<sup>[34]</sup>

NPS, especially SCs, have carved a specific niche for itself in the illicit synthetic drug market, their demand and use driven by providing a substitute for known illicit drugs. These substitutes are often altered variations of the scheduled parent to produce analogues of the illicit drug to circumvent current drug legislation, in turn it also alters pharmacological activity which alters user experience, increase potency and reduces manufacturing costs which makes it more attractive to substance abusers.

Due to the number of structurally diverse cathinones and at the rate at which NPS are introduced into the market, this drug class remains comparatively sparsely documented and insufficient to describe the wide variety of reported and observed clinical effects.<sup>[16]</sup> Compared to traditional amphetamine-type stimulant the synthetic cathinone market is still relatively small, but at the current introduction rate, significant risks arise for current and future public health. Currently there are multiple reports of hospitalisations and fatalities linked to SCs, reasons for these cases range from users being unaware of the contents of the illicit formulations and overdoses.<sup>[35-39]</sup> It becomes apparent that there is a need for rapid throughput methods to identify newly emerged NPS and to characterise these substances to assess their pharmacological properties and determine their public health risks in shorter timeframes. *In silico* modelling based on X-ray crystallographic data becomes more attractive in this regard; however, structures of SCs and receptor targets have only been recently described by X-ray crystallography and to date this data remains fairly limited.<sup>[40-42]</sup>

## **1.2 Problem statement**

At the current unprecedented NPS introduction rate it becomes evident that a move towards a throughput chemical structure-based health and neuropharmacology risk assessment approach is needed to allow the analysis of newly detected NPS in shorter timeframes for faster response times in not only drug overdose cases of unknown substance abuse, but also in law enforcement agencies across the world. *In silico* modelling based on X-ray crystallographic data poses a promising inexpensive method for identifying possible psychoactive substances based on the modelling of available crystallographic data of actual human drug targets and their interactions with newly identified NPS in shorter timeframes.

To date limited literature is available on MMMP, although reports of its misuse alongside other illicit drugs have been reported in South Africa, United States of America and Australia. There is a clear societal need for pharmacological data including complete pharmacokinetics and pharmacodynamics studies of MMMP to determine amongst other the psychoactive and toxicological properties. From this information the need to regulate public exposure to this substance should be assessed as MMMP is readily available to anyone and is extensively used in the printing and polymer industry to produce containers and labels in which, amongst other, foodstuffs are packaged in, to which the public is exposed to daily.

### ***1.3 Dissertation statement***

In light of the limited data available on MMMP; this research study will consist of two parts: The first part will investigate the possible interactions that might occur between MMMP (a known synthetic cathinone) and available human monoamine receptors and a monoamine transporter which act as common targets for illicit substances to elicit psychoactive effects. These possible interactions will contribute to the neuropharmacological data of which there is none, to our knowledge, available at the time of this study. The second part of this study will focus on solving and reporting the crystal structure of MMMP to be added to the library of available crystal structures for future research endeavours.

### ***1.4 Research aim and objectives***

This study consists of two main aims conducted as two independent studies: (1) to evaluate the possibility of MMMP interacting with a human monoamine transporter and receptors, and (2) report the crystal structure of MMMP using single crystal X-ray diffraction technique. To achieve these two aims various objectives must be achieved.

#### *(1) Molecular modelling study*

- Identify and select appropriate human transporters and receptors targets. In some cases, multiple crystal structures for a specific protein target have been reported.
- Establish a pH in which the simulations can be modelled.
- Develop an appropriate docking protocol.
- Validation of docking protocol through comparison with available crystallographic data.
- Obtain docking scores for MMMP ligand.
- Analyse docking results.
- Perform Molecular Dynamics (MD) studies on promising targets.
- Determine the stability of the complexes over the course of the MD simulation.

## (2) *Single crystal X-ray crystallography*

- Establish an appropriate method for the crystallisation of MMMP, this includes finding a solvent to dissolve MMMP, finding optimal temperature for good crystallisation and considering environmental factors such as sunlight, as MMMP is UV sensitive.
- Establishing a protocol to analyse a single MMMP crystal.
- Solve the crystal structure from diffraction data and analyse the results.

### 1.5 *Research significance*

Using *in silico* molecular docking studies based on X-ray crystallographic data it is possible to model interactions between MMMP and possible monoamine transporters and receptors. *In silico* modelling is done with software such as Schrödinger<sup>[43]</sup>, which allows users to evaluate if a ligand such as MMMP interacts with a particular protein target, in this case a specific monoamine receptor or transporter, which in effect may elicit psychoactive effects. Alternatively, if MMMP is found to be inactive it might merely be used as a “bulking agent” to adulterate illicit drugs due to its physical white powder appearance, but in itself has no biological activity. These adulterated drugs get sold to unsuspecting substance abusers which are none the wiser.

The interactions of MMMP with possible monoamine transporters and receptors will contribute to the neuropharmacological knowledge of MMMP of which there is no current data available. If it is found that MMMP, in theory, is capable of interacting with monoamine receptors and/or transporters it has a high possibility of eliciting psychoactive effects. Relevant steps, such as *in vitro* testing (i.e. HEK293 cells expressing a specific human receptor or transporter), can then be taken to confirm the results of the theoretical docking study. If the *in vitro* results are in agreement with the theoretical findings within this study, the information can be used by various law enforcement agencies across the world to evaluate the need to classify MMMP as an illicit substance. By classifying MMMP as an illicit substance helps to regulate public accessibility and exposure to the newly identified psychoactive substance. This study also reports the solved X-ray crystal structure of MMMP to be added to the vast library of available crystal structures which can be used for future research endeavours.

In summary, this study paves the way for a move towards an *in silico* chemical structure-based health and neuropharmacology risk assessment approach based on X-ray crystallographic data of human transporters and receptors and modelling software which would significantly decrease the time required for these assessments and possibly identify substances worth further investigation through *in vitro* studies which can be costly.

## 1.6 Delineations and limitations

### *Delineations*

This study will only be responsible for the following:

- Investigating possible interactions that occur between MMMP and an available human monoamine transporter and various monoamine receptors in the brain region of the CNS. Due to the nature of the modelling, it is not possible to conclusively determine the nature of the interactions between the ligand and the protein. These interactions can be as substrates or blockers in the case of transporters and agonist or antagonists in receptors.
- Using Glide docking extra precision (XP) method<sup>[44, 45]</sup> and MD simulations<sup>[46, 47]</sup> to determine possible interactions occurring between a ligand and protein targets.
- Working with transporters and receptors X-ray crystal structures obtained from the Protein Data Bank (PDB).
- Solving the single X-ray crystal structure of MMMP.

### *Limitation*

- Limitations in the accuracy of the computational software to predict interactions and specific binding poses were overcome by a validation step. MD simulations<sup>[46, 47]</sup> were used to further evaluate the validity of promising Glide docking results.<sup>[44, 45]</sup>
- Limitation in the accuracy of the PDB crystal structures. Crystal structures of receptors and transporters with the best resolution and least modifications were selected for the modelling to support the accuracy of crystal structure representations of the actual human monoamine transporter and receptors.
- Glide docking<sup>[44, 45]</sup> uses a flexible ligand, but rigid receptor approach. Due to the computational and time requirements to model a flexible receptor, which is more accurate, it was opted to use Glide docking. Alternative approaches such as ensemble dockings, induced fit dockings and reducing the van der Waals scaling factor can be used to improve the accuracy of docking results. This will thoroughly be discussed in Chapter 3.
- Modification of the receptors to produce “engineered proteins” could influence binding interactions therefore, the original published papers of a specific protein target were analysed to ensure that the modifications are outside of the conserved binding region to decrease the risk of modifications influencing docking results.
- Limitations associated with docking studies are discussed throughout Chapter 3.

## 1.7 Underlying assumptions

We assume that:

- The brain has a pH of 7.4 therefore, modelling was done at a pH of 7.4 with an appropriate deviation for the docking protocol, this is discussed in Chapter 3.
- The monoamine transporter and receptors used within this study are genetically identical for all humans and that the targets used are an accurate representation of the receptor present in the human population.
- That the crystal structures all contain an accurate representation of the conserved binding regions.
- For engineered proteins (i.e. where extracellular loops have been removed to stabilise the protein structure for elucidation purposes, in the original crystal structure publications, or any other modifications outside of the binding pocket) the modifications outside of the binding pocket do not influence the binding or the interaction of the ligand with the conserved region of the monoamine protein targets.

## 1.8 Definition of terms

**New psychoactive substances (NPS)** is a term used to describe analogues of commonly known abused drugs such as psychostimulants, cannabinoids, depressants (including opioids and benzodiazepines) and hallucinogens or novel synthesised drugs designed to mimic the psychoactive properties of controlled drugs. The definition for the classification of a NPS is not consistently defined or accepted, which leads to multiple ways in which a substance can be classified as an NPS.<sup>[1-3]</sup> The UNODC defines NPS as substances of abuse which can be in pure form or part of a preparation, that are not controlled by the 1961 Single Convention on Narcotic Drugs or the 1971 Convention on Psychotropic Substances and pose a threat to public health.<sup>[4]</sup> Others classify NPS as drugs with the potential of being abused and has only recently gained popularity in the illicit drug market. “New” does not necessarily refer to a completely novel substance, it could refer to an illicit substance that has been available for decades but has only recently been introduced and emerged in popularity.<sup>[48]</sup> The final classification of an NPS refers to these compounds as deliberately designed drugs (not necessarily new) to mimic existing controlled substances of abuse with the intention of bypassing legal regulations.<sup>[49-51]</sup> These classifications are not mutually exclusive, but independent of the basis for their classification, there however exists a general consensus for the drugs that fit within these classifications.<sup>[1]</sup> These drugs include many synthetic cannabinoids, tryptamines, arylcyclohexylamines, aminonindanes, phenethylamines, novel benzodiazepines, novel opioids and plant-based compounds such as SCs, kratom and piperazines.<sup>[50, 52]</sup>



**Agonist** is a substance that mimics the action of a hormone or neurotransmitter to produce a response when it binds to a specific protein target (receptors or transporters) such as activators or releasers. On the other hand, an **antagonist** is a substance that binds to a receptor without activating it; these are typically inhibitors or blockers.

**Conformational state** refers to a specific three-dimensional arrangement of amino acid residues or even secondary structures that can influence the characteristic (active or inactive state) of a tertiary or quaternary protein. Protein **active state** is a state in which the protein is capable of binding an agonist ligand causing a conformational change that trigger a downstream response. An **inactive state** refers to a state where the protein is unable to trigger a downstream response this can be due to an antagonist ligand causing an inactive conformational state. **Apo state** of a protein refers to a protein not bound to any ligands and therefore has an unoccupied binding pocket.

**Orthosteric site/binding pocket/binding site** is a cavity on the surface or within a protein that is lined with specific amino acid residues to which substrates or competitive inhibitors bind to in enzymes or agonists or competitive antagonists bind to in receptors. **Allosteric site or regulatory site** refers to a site, distinctive from the active site to which effectors bind to modulate (increase/decrease or inhibit) the activity of an enzyme (or monoamine transporter) by changing the structure of the orthosteric site.

### 1.9 Brief chapters overview

*Chapter 1* serves as background information and an introduction to the rationale, purpose and significance of this work and the gap in knowledge it will be addressing. In *Chapter 2* literature on cathinone, SCs and MMMP is reviewed. A short explanation on the theory required to understand ligand interactions with monoamine transporters and receptors is also given. *Chapter 3* serves as an introduction to the theory behind the *in silico* studies. In *Chapter 4* the methodology, results and discussion sections are reported for the docking study, which ends with a sub-conclusion based on the results. *Chapter 5* serves as an introduction to the theory and principles that govern single crystal X-ray diffraction, thereafter in *Chapter 6* the methodology, results and discussion of the solved crystal structure are reported and ends with a sub-conclusion based on the findings made in this chapter. The final chapter, *Chapter 7*, serves as a conclusion in which the results of both studies are summarised and the important conclusions from each study are highlighted, a short section on the contribution of this work to the ever-growing knowledge pool and future research will follow thereafter.

**Chapter 2: Literature review and Theory**



*Figure 2.1 - Line drawing of Catha edulis Forsk.*

## 2.1 Introduction and history of Khat

*Catha edulis* shown in **Figure 2.1** is an evergreen shrub that belongs to the moonseed family Celastraceae and was found by Petrus Forskål, a Finnish explorer, on an Arabia Peninsula expedition in 1761. The plant was later renamed *Catha edulis* Forsk. by the mathematician Carsten Niebuhr, the only survivor of the expedition, in honour of his friend that passed away during the expedition. *C. edulis* is an autochthonous plant to East Africa specifically to Ethiopia, formerly known by the exonym Abyssinia and Yemen. Today it is found in several other countries such as Afghanistan, Eastern and Southern Africa and Southwest Arabian Peninsula. Khat refers to the young twigs and leaves of the *C. edulis* shrub which is consumed by millions of people worldwide for its psychostimulant properties.<sup>[53-57]</sup>

### 2.1.1 Colloquial names for Khat

Depending on the region khat is known by different local names. In Yemen it is referred to as qat and gat, in Somalia to qaat or jaad, in Ethiopia as chat or jimma and in Kenya to miraa or veve. Other popular common names still used today include Abyssinian Tea, Flower or paradise, Kaht, Qaad, African salad, Bushman's Tea, Kus-es-salahin miraa, Tchat, Tchaad, Tohai and Tochat.<sup>[55, 57, 58]</sup>

### 2.1.2 Ethnomedicinal uses of *C. edulis*

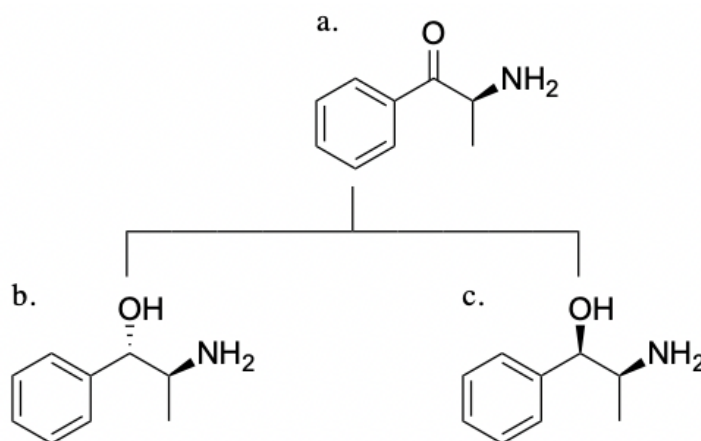
Despite khat containing psychostimulant properties, khat has been used for medicinal purposes as well. In Ethiopia the leaves and roots of *C. edulis* are processed and used for the treatment of coughs, gonorrhoea, influenza, asthma and other chest problems.<sup>[53]</sup> Some of these traditional uses have been investigated and corroborated by research. For example, Freund-Michel *et al.* (2008) found that cathinone present in khat may have beneficial effects in airway diseases displaying a heightened cholinergic such as nocturnal asthma, chronic obstructive pulmonary disease and asthma associated with oesophageal reflux.<sup>[59]</sup> Two compounds extracted from khat, 22 $\beta$ -hydroxytingenone and tingenone, have been shown to have antibacterial properties more potent than streptomycin and isonicotinic acid hydrazide against *Mycobacterium* species and broad cytotoxic activity as well.<sup>[60]</sup> Crude extracts and isolated compounds from khat have been shown to be cytotoxic to multiple leukaemia and prostate cancer cell lines.<sup>[60, 61]</sup> Al-Habori *et al.* (2004) showed that *C. edulis* leaves lowered the plasma cholesterol, glucose and triglyceride levels in New Zealand white rabbits.<sup>[62]</sup> In vitro studies showed that khat extracts have anticariogenic properties which prevents tooth decay.<sup>[63, 64]</sup>

### 2.1.3 Chemical composition of khat

The leaves and roots of *C. edulis* contains several alkaloids which can be categorised into different classes based on their chemical structures. These classes are phenylalkylamines (phenylpropylamines and phenylpentenylamines), cathedulins, monoterpenes, sterols and triterpenes, flavonoids, volatile aromatic compounds and other miscellaneous compounds like amino acids, vitamins and minerals.

The two major classes found in khat are the phenylalkylamines and cathedulins classes. The first major class, phenylalkylamines, can further be divided into the sub-classes phenylpropylamines and phenylpentenylamines<sup>[55]</sup> of which the former consist of (–)-cathinone [(2*S*)-2-amino-1-phenylpropan-1-one], and the two diastereoisomers (–)-norephedrine [(1*R*,2*S*)-2-amino-1-phenylpropan-1-ol] and (+)-cathine [(1*S*,2*S*)-2-amino-1-phenylpropan-1-ol] shown in **Figure 2.2**. These compounds are structurally related to amphetamine and noradrenaline. It should be noted that only the (–)-enantiomer of cathinone (S-(–)-cathinone) occurs in khat which has the same absolute configuration as S-(+)-amphetamine.<sup>[65, 66]</sup> The second sub-class, phenylpentenylamines, consists of three alkaloids found in khat namely pseudomerucathine, merucathine and merucathinone.<sup>[67-69]</sup>

The second major class contains 62 cathedulins characterised by Kite *et al.* in 2003.<sup>[70]</sup> For more information on all the cathedulins alkaloids, which is beyond the scope of this paper, can be found in a paper titled: “Chemical composition of *Catha edulis* (khat): a review” by M. Getasetegn.<sup>[55]</sup>



**Figure 2.2** - Chemical structures of **a.** (–)-cathinone [(2*S*)-2-amino-1-phenylpropan-1-one], **b.** (+)-cathine [(1*S*,2*S*)-2-amino-1-phenylpropan-1-ol] and **c.** (–)-norephedrine [(1*R*,2*S*)-2-amino-1-phenylpropan-1-ol].

#### 2.1.4 *Discovery of the psychoactive compound in Khat*

The composition of khat has puzzled pharmacologists and chemist for over 100 years. The first scientific paper written on *C. edulis* was in the late nineteenth century by Flückiger and Gerock<sup>[71, 72]</sup> in this paper the stimulant effects of khat was primarily attributed to caffeine, which was later reported not to be present in khat by the same workers. However, another alkaloid which they named “katin” was suggested to be responsible for the stimulant effects, but they did not determine its structure. “Katin” was later isolated and identified from dried khat leaves as cathine ((+)-norpseudoephedrine) in 1930 by Wolfes’.<sup>[55, 73]</sup> Cathine was believed to be the main psychoactive compound of khat, however cathine was shown to be a stimulant of low potency and could not account for all the symptoms observed with khat consumption.<sup>[74]</sup> Friebel and Brilla (1963), in an attempt to explain the discrepancy, successfully isolated an unidentified alkaloid from lyophilized fresh khat leaves that was more potent than (+)-norpseudoephedrine in stimulating the motor activity of mice. This unidentified alkaloid could not be isolated from dried khat leaves, therefore it was suggested that the alkaloid was a biosynthetic precursor to (+)-norpseudoephedrine, formed when khat samples are dried.<sup>[75]</sup>

The United Nations Commission on Narcotic Drugs in 1971 directed international attention to the nature and extent of khat use and recommended that the United Nations Laboratory reinvestigate the chemical composition of khat which led to the isolation of (2S)-2-amino-1-phenylpropan-1-one and the name (-)-cathinone was proposed.<sup>[76]</sup> The (-)-enantiomer of cathinone was confirmed by Schorno and Steinegger in 1979.<sup>[77]</sup> **From here on forward, any reference made to cathinone implies the (-)-enantiomer of cathinone which has been proven to be the major psychoactive alkaloid present in khat.**<sup>[77]</sup>

#### 2.1.5 *Stability of cathinone*

Cathinone is a very unstable molecule which readily undergoes enzymatic decomposition by cathinone reductase to (+)-cathine and (-)-norephedrine shown in **Figure 2.2**. Decomposition occurs during the normal aging of the young shoots or directly after harvesting. The freshness of khat is vital to its value, the longer it takes for the fresh khat leaves to reach the end user the lower value it has due to the reduced cathinone content. The freshness of khat is preserved by wrapping it in banana leaves.<sup>[53, 67, 78-80]</sup>

### 2.1.6 Pharmacokinetics

*Administration.* Khat has several ways in which it can be consumed for its desired effects. The most common route of administration is orally, but some users resort to smoking/inhalation. Traditionally khat chewers fill his/her mouth with fresh khat leaves and twigs and then chew it slowly and intermittently to release the alkaloids including the main active ingredient cathinone and to a lesser extent cathine and norephedrine. During the chewing about 80% of the cathinone and cathine, and 90% of the norephedrine is released.<sup>[81]</sup> The juices from the masticated plant material is then swallowed with saliva and the remaining plant material is kept as a ball in the cheek.<sup>[82, 13]</sup> To a lesser extent khat can also be used as follows: (1) fresh khat leaves and twigs can be used to brew teas known as Abyssinian, Arabic or Bushman's tea; (2) dried powdered leaves can be mixed with honey or sugar to produce candies or with other herbal extracts and water to produce a paste; (3) mead (honey wine) can be made by mixing khat extractions/infusions with honey and allowing it to ferment; (4) dried khat leaves can also be rolled up and smoked alone or in combination with tobacco or hashish (this method of consumption is least common).<sup>[54, 13, 83]</sup>

*Distribution.* During khat chewing the psychoactive alkaloids are released into the saliva, which almost entirely gets absorbed through the oral mucous membrane. The remaining alkaloids content is absorbed through the gastrointestinal tract (gastric mucous membrane) once swallowed, reaching the systematic circulation rapidly.<sup>[84, 85]</sup> It has been reported that individuals can consume anything from 100 g to 500 g of khat during a single chewing session lasting three to four hours, binge sessions that can last up to 24 hours has also been reported.<sup>[86, 87]</sup> During these chewing sessions copious amounts of fluids such as cola (sweet beverages), tea and cold water are consumed to counteract the astringent taste and prevent dryness in the mouth.<sup>[82, 87]</sup>

The alkaloid content of khat have been the focus of many studies, factors such as the origin, type, time since harvesting and storage conditions all influence the alkaloid content of khat especially the concentration of cathinone present in fresh khat.<sup>[88]</sup> A study done by Geisshüler *et al.* (1987) determined that fresh khat from Ethiopia, North Yemen, Kenya and Madagascar contained an average of 36mg of cathinone per 100g of fresh khat leaves.<sup>[78]</sup> An average of 77.7 to 342.8 mg cathinone per 100 g of fresh khat leaves was reported for khat from different locations in Yemen.<sup>[89]</sup>

*Metabolism.* In humans, cathinone is rapidly metabolised by phase I metabolic enzymes into (+)-cathine and (-)-norephedrine by reducing the  $\beta$ -ketone moiety to the corresponding alcohols shown in **Figure 2.2**.

*Elimination.* Cathinone mainly gets excreted in the urine in the form of its metabolites with 7% or less of the absorbed parent compound remaining unmetabolised<sup>[65, 79, 81, 90]</sup> and an average elimination half-life of 1.5 to 4.3 hours.<sup>[85, 90]</sup> Interestingly, cathinone shows very low blood–brain barrier (BBB) permeability. It is presumed that cathinone's psychoactive effect is mediated through its lipid solubility property which assists it to cross the BBB into the central nervous system (CNS) more readily than its metabolites, cathine and norephedrine.<sup>[91]</sup>

### 2.1.7 Pharmacodynamics

The effects observed after khat consumption are generally of CNS stimulation and include, logorrhoea, euphoria, hyperthermia, excitation, analgesia, anorexia and increased sensory stimulation, energy, concentration and self-esteem.<sup>[13, 92]</sup> Several clinical trials and animal studies have reported that the effects observed with khat/cathinone usage are similar to those observed with amphetamine usage.<sup>[90, 93-95]</sup> Moreover, khat chewers believe that during their khat chewing sessions they are more alert and think more quickly and clearly, although their concentration and judgment are objectively impaired.<sup>[94]</sup>

These desired stimulant effects are often followed by several adverse effects which include: (1) abuse potential and psychological dependence; (2) prolonged, excessive or chronic use lead to mild withdrawal symptoms such as lack of concentration, insomnia, lethargy, anxiety, depression, irritability, nightmares often paranoid in nature, trembling; and (3) impairment in cognitive function. Other health issues include cardiovascular (i.e. increase in heart and blood rate, myocardial infarction and acute vascular vasospasms), sympathomimetic (i.e. dry mouth and hyperthermia), reproductive (i.e. spermatorrhoea), gastrointestinal (i.e. stomatitis, constipation and periodontal disease) and carcinogenic potential (i.e. neck and head cancer).<sup>[13]</sup>

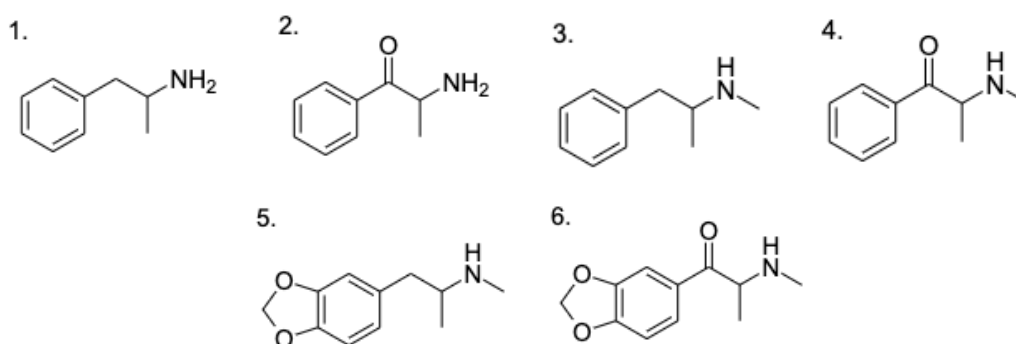
Khat-induced psychostimulation can be assumed to be predominately, or even exclusively due to the cathinone content present in its leaves.<sup>[88]</sup> Although cathinone has a low BBB permeability, its higher lipid solubility facilitates its access into the CNS leading to its psychoactive effects.<sup>[65, 96, 97]</sup> The major metabolites of cathinone, (+)-cathine and (-)-norephedrine, have been shown to possess weaker CNS stimulation due to their less lipophilic properties.<sup>[67]</sup> Other alkaloids such as phenylpentenylamines (pseudomerucathine, merucathine

and merucathinone) are present in low concentrations and were shown to have a weak effect on dopamine release in dopamine prelabelled rat striatal tissue.<sup>[94]</sup> Furthermore, the pharmacology of the 62 cathedulins present in khat have not been well characterised in the CNS and other organs.<sup>[70]</sup>

Current scientific data suggests that cathinone-induced psychostimulation, anorexic effects and dependence-producing potential are mediated primarily via the meso-striatum-corticolimbic dopaminergic pathway, although other pathways cannot be ruled out at this time.<sup>[88, 98]</sup> Cathinone's interaction with monoamine transporters and receptors will be discussed later in this chapter.

## 2.2 Introduction and history of Synthetic Cathinones

Synthetic cathinones (SCs) are chemical analogues of the natural occurring compound cathinone, the main psychoactive alkaloid present in *C. edulis* and belong to the drug class of new psychoactive substances (NPS) that are capable of evoking psychostimulant effects similar to amphetamine in users. From a structural perspective cathinone is a  $\beta$ -keto analog of amphetamine and SCs are often referred as "bk-amphetamines". Examples of amphetamines and their related  $\beta$ -keto analogues include amphetamine-cathinone, methamphetamine-methcathinone and MDMA-methylone shown in **Figure 2.3, Structures 1 to 6**. SCs have become a major problem worldwide with the EMCDDA currently monitoring 930 NPS and UNODC monitoring 1,184 NPS, with novel substances being reported yearly.<sup>[9, 11]</sup>



**Figure 2.3** - Amphetamines and their related  $\beta$ -keto analogues: (1,2) Amphetamine-Cathinone, (3,4) Methamphetamine-Methcathinone and (5,6) Ecstasy (MDMA)-Methylone.

### 2.2.1 Background history of synthetic cathinones

The first SCs were synthesised in the late 1920's<sup>[99, 100]</sup>, attracting attention for their putative medicinal uses such as appetite suppressant and antidepressants in the next decade.<sup>[101-103]</sup> It was only at the beginning of the 21<sup>st</sup> century that SCs have entered the recreational drug market,



gaining popularity and rapid distribution around the globe.<sup>[104]</sup> Other factors that contributed to SCs popularity amongst users is the easy accessibility, affordable prices and their former legal status.<sup>[6]</sup> Over the years several measure have been taken to try and stop this drug pandemic, but drug designers always appear to be one-step ahead of law enforcement agencies.<sup>[105]</sup> When SCs are classified as illegal, the drug market simply gets flooded with new derivatives to replace it. This poses an immense risk to public health as these substances are untested and their purity and composition are questionable.

The first two SCs manufacturing processes were described shortly before the discovery of cathine by Wolfes in 1930<sup>[73]</sup>, in 1928 2-(methylamino)-1-phenylpropan-1-one<sup>[99]</sup>, or methcathinone, and a year later the synthesis of 2-(methylamino)-1-(4-methylphenyl)propan-1-one<sup>[100]</sup>, or mephedrone, was described. The structural similarity between methcathinone and mephedrone to classical amphetamines, in addition to their effects on the CNS sparked interest in synthesising cathinones for medicinal purposes.<sup>[106-109]</sup> SCs used as therapeutic drugs include methcathinone, amfepramone, bupropion, pyrovalerone and methylone (**Table 2.1, Structures 8, 12 to 14 and 17** respectively) of which only the latter three are still in use today. The use of these SCs are seldomly prescribed due to their dependence and potential for abuse, as well as their adverse effects.<sup>[110-113]</sup>

Methylone, or 1-(1,3-benzodioxol-5-yl)-2-(methylamino)propan-1-one, was developed in the late 1990s to be used as an antidepressant and as an anti-Parkinsonism drug<sup>[114]</sup>, but was never marketed for medicinal use due to its potent psychostimulant action similar to that of 3,4-methylenedioxymethamphetamine (MDMA).<sup>[111]</sup> Methylone was among the so-called “first generation/wave” SCs to be marketed as “legal highs” in Japanese and European markets in the early 2000’s, through the internet and “smartshops”.<sup>[104]</sup> At the same time, Israel reported a short period of cathinone abuse which led to a cathinone ban.<sup>[115, 116]</sup> This together with the instability of the MDMA market in the European Union lead to the sale of mephedrone by the Neorganics company (Israel), rapidly spreading throughout Australia and Europe and later to the United States of America.<sup>[116-120]</sup> In the same year, 2-(ethylamino)-1-phenylpropan-1-one (N-ethylcathinone), , 1-(3-fluorophenyl)-2-(methylamino)propan-1-one (3-FMC), 1-(4-fluorophenyl)-2-(methylamino)propan-1-one (flephedrone or 4-FMC), 1-(1,3-benzodioxol-5-yl)-2-(methylamino)butan-1-one (butylone) and 1-(1,3-benzodioxol-5-yl)-2-pyrrolidin-1-ylpentan-1-one (MDPV) (**Table 2.1, Structures 9, 10, 11, 19 and 23**) were reported for the first time through the Early Warning System.<sup>[121]</sup>

Several factors brewed the perfect storm for the increase in demand and popularity of SCs in the mid 2000's, including:

- Cocaine and amphetamine-like psychostimulant effects such as empathy, euphoria, increased openness and sociability, increased libido and sexual performance.
- Affordability
- Easy availability from “smartshops” and the internet
- Eye-catching packaging and appealing names.
- Mistakenly perceived as legal (due to misleading marketing terms such as “legal highs”) and safe (high purity) to consume.
- Lack of field screening tests to detect their abuse.<sup>[6]</sup>

In addition to these reasons some studies also suggested that the decrease in purity, high price and unavailability of MDMA, cocaine and amphetamine might have also contributed to the popularity of SCs.<sup>[7, 8]</sup>

The alarming popularity of SCs between 2009 and 2010, especially mephedrone, led governmental institutions to start acting to try and stop the production and use of SCs. In April 2010 the United Kingdom (UK) Government, upon recommendation of the Advisory on the Misuse of Drugs, scheduled several SCs (including mephedrone) as Class B controlled drugs in the UK Misuse of Drugs Act 1971.<sup>[122, 123]</sup> This scheduling only acted as a catalyst for derivatives to appear on the market to circumvent the law. From 2009 to 2015, 26 new derivatives, which were reported for the first time through the Early Warning System in the European Union (EU), emerged to replace mephedrone.<sup>[121, 124, 125]</sup> This legislative pressure in combination with the high demand for novel and legal derivatives led to 69 new SCs emerging between 2012 and 2015 with a peak of 31 derivatives reported in 2014 alone.<sup>[126-129]</sup>

It became evident that a more serious intervention was required to control the high-speed proliferation of the NPS market after The Commission on Narcotic Drugs Resolution 56/3 of 2013 drew attention to the rate of emergence and the unprecedented number of NPS already on the market and recognised the need to detect, identify and report NPS through the establishment of a global Early Warning System. International cooperation was reinforced by the encouragement of each EU Member State and relevant organisations to collect and share information (i.e. toxicological, forensic, scientific and epidemiological data) relating to these substances, as well as the implementation of restrictions, regulatory measures and legislations in each country for a more swift response.<sup>[130]</sup> Legal approaches implemented by different countries in an attempt to stunt the use and distribution of NPS varies greatly, some

implemented substance-by-substance (specific substance) control, analogue (substances chemically related to a substance already under control), generic (entire group of substances) and others implemented “blanket bans”.<sup>[131]</sup> The implementation of blanket bans forced legal retail outlets (“head- and smartshops” convenience stores) to close their doors and forced NPS, including SCs, to be sold on illegal street markets alongside other common illicit drugs of abuse and to the dark web.<sup>[6]</sup>

Since 2015 the number of SCs reported to the EU Early Warning System has significantly decreased. Although the reasons for this decline is not clear it is believed that legislative control measures adopted by EU Member States, law enforcement operations in countries that supply the precursors required for the manufacturing SCs and increased control measures may have contributed to this phenomenon. In 2017 the Council Decision 2005/387/JHA, which deliberates on the control of NPS, risk assessment and information exchange, was revised through the Directive (EU) 2017/2103 with the aim of establishing a swifter and more effective system. The three step process for responding to NPS remained unchanged (early warning, risk assessment and control measures), however data collection and assessment procedures were accelerated and shorter deadlines were introduced.<sup>[132]</sup>

Although we have come a long way since the peak SCs introduction in 2014 and legislative control measures and law enforcement operations have proven to be vital in controlling the NPS market, chemical designers always seem to be one step ahead of the authorities in this continuous cat and mouse game.

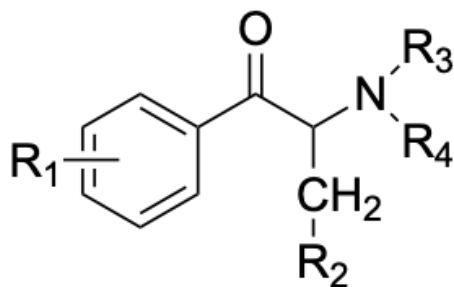
### 2.2.2 *Common names for synthetic cathinones*

SCs were originally marketed as “*bath salts*”, “*plant food*”, “*carpet cleaners*”, “*fertilisers*”, “*research chemicals*” or “*stain removers*” and labelled “*not for human consumption*” or “*for research purposes only*” to avoid regulatory control.<sup>[133]</sup>

### 2.2.3 *Chemistry of synthetic cathinones*

SCs belong to a large family of methylphenethylamine, that structurally resembles classical amphetamines with a key difference of containing an additional  $\beta$ -keto substituent on the amino alkyl side chain. The cathinone backbone works as a template onto which other substituents can be added, to produce a plethora of SCs all with unique psychoactive properties. Modification of the cathinone backbone can occur in four different positions as indicate in

**Figure 2.4**, R1 represents modification at the aromatic ring, R2 the alky side chain and the amino group at positions R3 and R4.<sup>[132, 134-136]</sup>



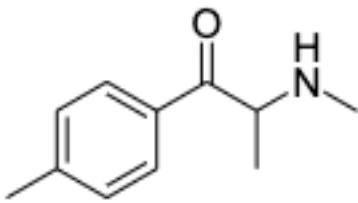
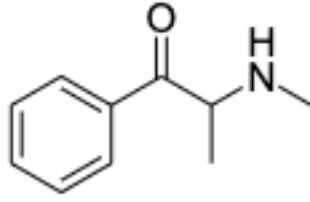
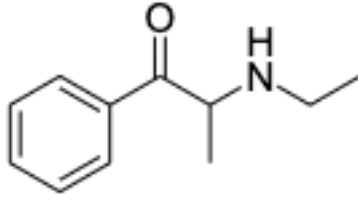
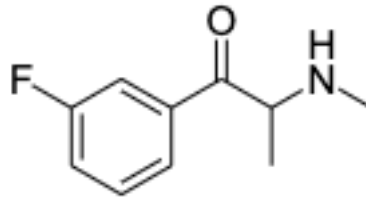
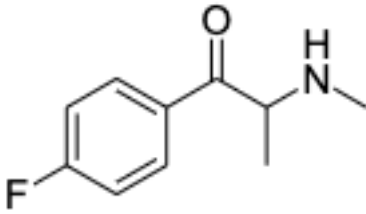
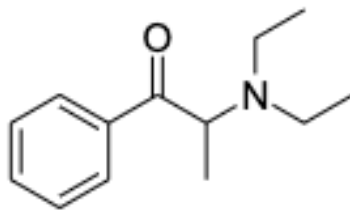
*Figure 2.4 - General structure of synthetic cathinones. R1, R2, R3 and R4 represent position at which substitutions can be made to obtain different cathinone derivatives. R1 represent substitution at the aromatic ring, R2 at the  $\alpha$ -carbon, R3 and R4 substitutions at the amino group.*

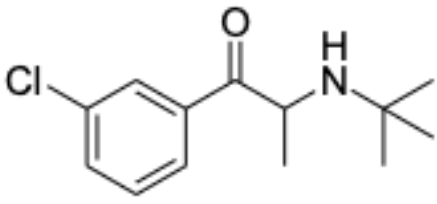
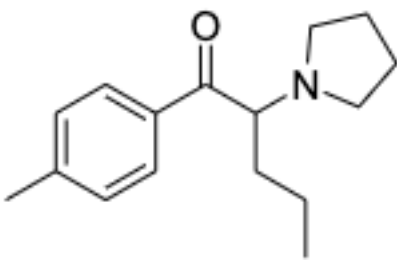
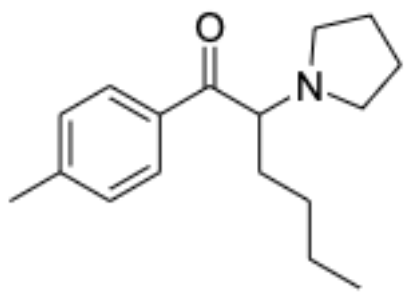
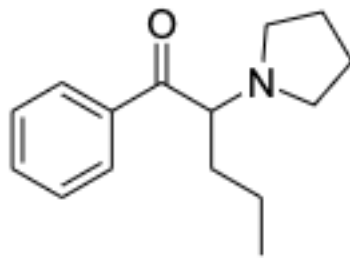
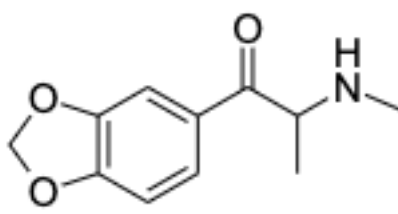
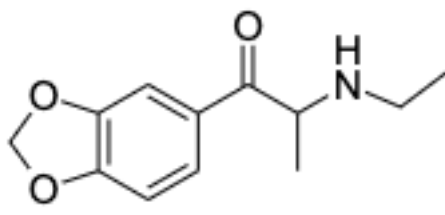
SCs can be classified into four different chemical sub-families depending on where substitutions are made on the cathinone backbone<sup>[13, 134, 136]</sup>:

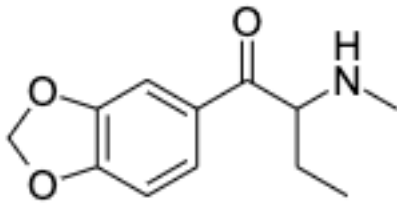
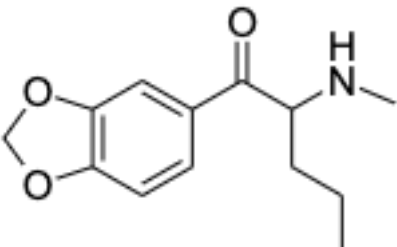
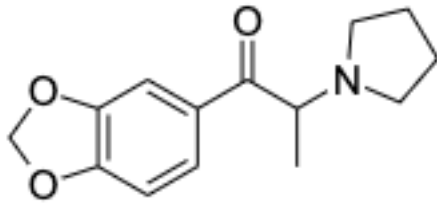
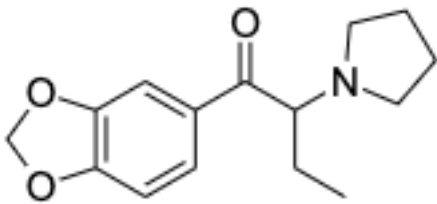
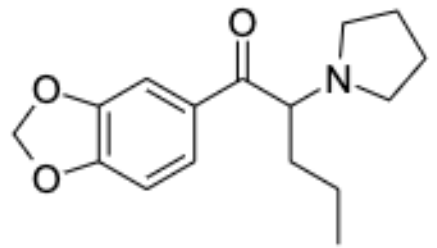
- (1) N-alkyl cathinones, characterised by alkyl substituents on the amino group at positions R3 and/or R4, and possible alkyl or halogen substitutions in the aromatic ring (R1) and/or alkyl substitutions in the  $\alpha$ -carbon of the side chain (R2), i.e. Mephedrone, Methcathinone, Ethcathinone, 3-FMC, 4-FMC, Amfepramone and Bupropion (**Table 2.1, Structures 7 to 13**)
- (2) N-pyrrolidine cathinones, characterised by a pyrrolidinyl substitution in the amino group at R3 and R4 positions, and possible alkyl or halogen substitutions in the aromatic ring (R1) and/or alkyl substitutions in the  $\alpha$ -carbon of the side chain (R2), i.e. Pyrovalerone, MPHP and  $\alpha$ -PVP (**Table 2.1, Structures 14 to 16**)
- (3) 3,4-Methylenedioxy-N-alkyl cathinones, characterised by the addition of a 3,4-methylenedioxy group to the aromatic ring at R1 and alkyl substitutions in the amino group R3 and/or R4, and possible alkyl substitutions at aromatic ring (R1) and the  $\alpha$ -carbon of the side chain (R2) i.e. Methylone, Ethylone, Butylone and Pentylone (**Table 2.1, Structures 17 to 20**)
- (4) 3,4-Methylenedioxy-N-pyrrolidine cathinones, characterised by the addition of a 3,4-methylenedioxy group to the aromatic ring at R1 and a pyrrolidinyl substitution in the amino group (R3 and R4), and possible alkyl substitutions at aromatic ring (R1) and the  $\alpha$ -carbon of the side chain (R2) i.e. MDPPP, MDPBP and MDPV (**Table 2.1, Structures 21 to 23**)
- (5) A miscellaneous chemical sub-family exists into which SCs with unique structures can be grouped in i.e.  $\beta$ -naphyrone in which the phenyl moiety of  $\alpha$ -PVP is replaced by a naphthyl ring.

## 2.2.4 Structures and names of common synthetic cathinones

**Table 2.1** - Structures, common- and IUPAC names of a few common synthetic cathinones grouped into four categories.

| Structure number                        | Common name                               | IUPAC name                                     | Chemical structure  |
|---|---|--|---|
| <b>(1) N-alkyl synthetic cathinones</b> |   |  |   |
| 7.                                      | Mephedrone; 4-Methylmethcathinone; 4-MMC  | 2-(Methylamino)-1-(4-methylphenyl)propan-1-one |    |
| 8.                                      | Methcathinone; Ephedrone                  | 2-(methylamino)-1-phenyl-1-propanone           |    |
| 9.                                      | Ethcathinone; N-Ethylcathinone            | 2-(Ethylamino)-1-phenylpropan-1-one            |   |
| 10.                                     | 3-FMC; 3-Fluoromethcathinone              | 1-(3-Fluorophenyl)-2-methylamino)propan-1-one  |   |
| 11.                                     | 4-FMC; Flephedrone; 4-Fluoromethcathinone | 1-(4-Fluorophenyl)-2-methylamino)propan-1-one  |   |
| 12.                                     | Amfepramone; N,N-Diethylcathinone         | 2-(Diethylamino)-1-phenylpropan-1-one          |  |

|  |   |  |   |
|--|---|--|---|
| 13.  | Bupropion; $\alpha$ -(Tert-butylamino)-mchloropropiophenone; Amfebutamone             | 2-(Tert-butylamino)-1-(3-chlorophenyl)propan-1-one   |     |
| <b>(2) N-pyrrolidine synthetic cathinones</b>              |   |  |   |
| 14.  | Pyrovalerone ;4-Methyl- $\alpha$ -pyrrolidinovalerophenone                            | 1-(4-Methylphenyl)-2-pyrrolidin-1-ylpentan-1-one     |     |
| 15.  | MPHP; 4-Methyl- $\alpha$ -pyrrolidinohexanophenone                                    | 1-(4-Methylphenyl)-2-pyrrolidin-1-ylhexan-1-one      |    |
| 16.  | $\alpha$ -PVP; $\alpha$ -Pyrrolidinopentiophenone; $\alpha$ -Pyrrolidinovalerophenone | 1-Phenyl-2-pyrrolidin-1-ylpentan-1-one               |  |
| <b>(3) 3,4-Methylenedioxy-N-alkyl synthetic cathinones</b> |   |  |   |
| 17.  | Methylone; $\beta$ k-MDMA   | 1-(1,3-Benzodioxol-5-yl)-2-(methylamino)propan-1-one |   |
| 18.  | Ethylone; $\beta$ k-MDEA  | 1-(1,3-Benzodioxol-5-yl)-2-(ethylamino)propan-1-one  |   |

|  |   |  |  |
|--|---|--|--|
| 19.  | Butylone; $\beta$ k-MBDB                                      | 1-(1,3-Benzodioxol-5-yl)-2-(methylamino)butan-1-one    |    |
| 20.  | Pentylone; $\beta$ k-MBDP                                     | 1-(1,3-Benzodioxol-5-yl)-2-(methylamino)pentan-1-one   |    |
| <b>(4) 3,4-Methylenedioxy-N-pyrrolidine cathinones</b> |   |  |  |
| 21.  | MDPPP; 3,4-Methylenedioxy- $\alpha$ -pyrrolidinopropiophenone | 1-(1,3-Benzodioxol-5-yl)-2-pyrrolidin-1-ylpropan-1-one |   |
| 22.  | MDPBP; 3,4-Methylenedioxy- $\alpha$ -pyrrolidinobutiophenone  | 1-(1,3-Benzodioxol-5-yl)-2-pyrrolidin-1-ylbutan-1-one  |  |
| 23.  | MDPV; 3,4-Methylenedioxypropyvalerone                         | 1-(1,3-Benzodioxol-5-yl)-2-pyrrolidin-1-ylpentan-1-one |  |

### 2.2.5 Medicinal uses

The manufacturing of two SCs in 1928 (methcathinone) and 1929 (mephedrone), sparked interest in the development of SCs for clinical purposes due to the structural resemblance of these SCs to classical amphetamines and their effects on the CNS.<sup>[106-109]</sup>

Methcathinone (**Table 2.1, Structure 8**) (also known by the slang names “*cat*”, “*jeff*” or “*mulka*”) was the first SC marketed for medicinal purposes. In the Former Soviet union it was

marketed as an antidepressant in 1930s and 1940s and later in 1957 it was investigated as an analeptic drug in the USA, but was never clinically commercialised in the latter case.<sup>[23, 87, 113, 137, 138]</sup> The first report of methcathinone abuse started to surface in the Soviet Union in the 1970s and 20 years later in the USA leading to its inclusion in the Schedule I of the United Nations Convention on Psychotropic Substances.<sup>[87, 111, 139, 140]</sup>

Amfepramone (**Table 2.1, Structure 12**) was introduced as an appetite suppressant drug in 1958<sup>[101, 110, 141, 142]</sup> and at the same time pyrovalerone (**Table 2.1, Structure 14**) was synthesised to be used as an appetite suppressant drug and anorectic as well as for the clinical treatment of chronic fatigue and lethargy.<sup>[102, 143]</sup> Shortly after the introduction of amfepramone and pyrovalerone reports of their abuse started to emerge leading to their withdrawal from the market and inclusion in the Schedule IV of the United Nations Convention on Psychotropic Substances.<sup>[139]</sup> Amfepramone is still prescribed today as an efficacious adjunct for weight loss under the label Tenuate<sup>®</sup>, and pyrovalerone is an approved medication in the USA, although rarely prescribed.<sup>[144]</sup>

Bupropion (**Table 2.1, Structure 13**) has been used in the treatment of obesity, depression (Wellbutrin<sup>®</sup>) and as a co-adjuvant in smoking cessation (Zyban<sup>®</sup>) and therapy. Additionally it has proven efficacy in patients with seasonal affective disorders in preventing depressive episodes and it has potential pharmacological relevance for the treatment of psychoactive substance dependence and attention-deficit/hyperactivity disorder (ADHD).<sup>[103, 145-151]</sup>

### 2.2.6 Pharmacokinetics

*Administration.* SCs have multiple routes of administration, with oral and nasal insufflation administration being the most common. SCs are consumed orally in the form of tablets or capsules, swallowed after the powder has been wrapped/rolled up in cigarette paper a process known as “bombing”, drunk after SCs have been dissolved in a beverage or nasal insufflation/snorting using a key dipped in powder (“keying”).<sup>[136, 152]</sup> Intravenous (i.v.) injection, also known as “slamming”, has recently been described as another common route of administration.<sup>[1, 153]</sup> Other less common routes of administration include subcutaneous (s.c.) and intramuscular injections, sublingual and gingival delivery, rectal insertion, inhalation (vaporisation/smoking of e-cigarettes and insertion of substances into the eyes).<sup>[1, 153-155]</sup> Multiple simultaneous routes of administration for a single session have been reported<sup>[153, 156]</sup> and SCs being used simultaneously with other drugs of abuse intentionally [i.e. “chemsex” parties where mephedrone in addition to methamphetamine, 4-hydroxybutanoic acid (GHB) or



its precursor oxolan-2-one (GBL) have been reported to be abused together, to sustain, enhance and facilitate the sexual experience]<sup>[157, 158]</sup> or unknowingly (i.e. unknown content of “bath salts” led users to simultaneously abuse mephedrone, caffeine and MDMA)<sup>[125]</sup>. SCs are often abused alongside other drugs such as cannabinoids, classical amphetamines, GHB/GBL, cocaine, ketamine and/or alcohol, or prescription drugs (“Z-drugs” i.e. zolpidem) to expand the psychoactive experience. Other prescription drugs, such as proton pump inhibitors, benzodiazepines and  $\beta$ -blockers are also used in combination with SCs in an attempt to counteract some side-effects relating to SCs usage, these side effects include stomach pain, anxiety and tachycardia respectively, selective phosphodiesterase type 5 inhibitors to increase sexual performance and libido.<sup>[6, 152]</sup>

SCs doses vary largely, from a few milligrams to a few grams. Factors such as the synthetic cathinone derivative used, route of administration and “bath salt” content which have varying concentrations and purity that might be different than the claimed composition that might lead to unwanted effects and in extreme cases to overdosing.<sup>[133-135, 156, 159, 160]</sup>

*Distribution.* Generally SCs, particularly N-alkyl derivatives, are less lipophilic than their corresponding amphetamine analogue due to the  $\beta$ -ketone moiety, N-pyrrolidine derivatives have increased lipophilic properties due to the presence of the pyrrolidine ring which counteracts the decreased lipophilicity attributed to  $\beta$ -ketone moiety.<sup>[134, 136, 161]</sup> Lower lipophilicity of SCs, compared to related amphetamines, may be indicative of lower potency and weaker ability to cross the BBB, therefore to achieve equipotent effects to amphetamine usage, users have to re-dose as well as use greater doses.<sup>[134, 156]</sup> Nonetheless several SCs shows *in vivo* brain-to-serum or brain-to-plasma concentrations ratios greater than one, which supports the fact that SCs freely cross the BBB.<sup>[162]</sup>

*Absorption.* Following the administration of SCs and their absorption into the systemic circulation, SCs are distributed first to the highly irrigated organs such as the brain, lungs, liver and kidneys. SCs distribution into the brain is dependent on their ability to cross the BBB, this first phase is responsible for the initial onset of psychoactive effects. In the second phase substances are distributed to other compartments such as the skin, muscle and fat which is responsible for the redistribution effect observed for some drugs. Several factors influence the distribution of substances in living organisms, specifically the ability of the substance to bind to plasma proteins and to tissues, local pH, blood flow, permeability and cardiac output. Generally SCs have low plasma protein binding<sup>[13, 163]</sup>, this pharmacological parameter may be

linked to the relative low steady-state values and half-life and the rapid elimination of these derivatives such as mephedrone<sup>[164]</sup> and methyldone<sup>[165]</sup>. However, low plasma protein binding does not necessarily equate to low half-life and steady-state values and the rapid elimination of substances as in the case of  $\beta$ -naphyrone and  $\alpha$ -pyrrolidinohexanophenone ( $\alpha$ -PHP) which have relatively long half-lives despite low plasma protein binding.<sup>[166, 167]</sup> In the case of  $\alpha$ -PHP which has a half-life of 150 hours, the SC undergoes rapid phase I distribution to the brain, then redistribution to other organs occurs in phase II from which the derivative is gradually released from the organs into the blood over a period in order to be eliminated.<sup>[167, 168]</sup> Numerous literature supports the redistribution phenomenon.<sup>[39, 169-172]</sup>

*Metabolism.* SCs undergo phase I and phase II metabolism, mainly mediated by cytochrome P450 liver enzymes, although almost all SCs are also excreted in their unchanged form in urine.<sup>[134, 173, 174]</sup> As previously mention SCs can be categorised into five chemical sub-families, each of these categories have characteristic chemical structures which determines their main metabolic pathway, therefore major metabolic pathways are the same intragroup but differ intergroup. Phase I metabolism can involve oxidation, reduction, N-dealkylation, hydroxylation, dehydrogenation, oxidative deamination, demethylation and O-methylation. Phase II metabolism involves glucuronidation and/or sulfation. The specific pathways each of these five sub-families undergo are beyond the scope of this paper and further stepwise metabolism can be obtained from the comprehensive paper by Soares *et al.* (2021) titled: “An updated review on synthetic cathinones”.<sup>[13]</sup>

*Elimination.* Majority of SCs are excreted in urine either in an unchanged form or as metabolites which result from Phase I and II metabolism.<sup>[134]</sup>

### 2.2.7 Pharmacodynamics

The effects observed with SCs, like with cathinone, abuse are generally of CNS stimulation due to their cocaine- and amphetamine-like stimulating effects which enhances sensory and social experiences.<sup>[6]</sup> Desired effects that are stimulating, pleasurable and hallucinogenic in nature are the key drivers for SCs abuse and include the following: (1) euphoria; (2) empathy; (3) increased awareness and alertness; (4) increase energy; (5) increased motivation; (6) greater feeling of well-being and mood lift; (7) increased productivity and work capacity; (8) increased libido and sexual arousal; (9) increased openness; talkativeness; self-confidence and sociability; (10) hallucinations and perceptual disorientation; (11) intensification of sensory experience; (12) reduced appetite; (13) insomnia; (14) motor excitement<sup>[13]</sup>.

As with most drugs of abuse several adverse effects arise after use either in overdosing or acute and chronic intoxication cases. In the case of SCs the most common reported adverse effects are psychiatric, neurological and cardiac in nature generally preceding other effects such as gastrointestinal and hepatic (i.e. nausea, abdominal pain, acute liver failure and increased serum aminotransferase levels), musculoskeletal (i.e. rhabdomyolysis and increased serum creatine kinase levels), haematological (i.e. disseminated intravascular coagulation), pulmonary (i.e. respiratory acidosis and respiratory failure and arrest) and renal (i.e. acute kidney failure and increased serum creatine levels) which may result in multiple organ failure and eventually death.<sup>[1, 152, 153, 175]</sup>

The most reported psychiatric and neurochemical adverse effects in cases of SCs intoxication and overdose related cases are aggressive and erratic behaviour, agitation, cognitive disorders, anxiety, delusions, paranoia, psychosis, seizures and visual and auditory hallucinations. Brain-related adverse effects including encephalopathy, stroke, convulsions and coma have also been reported with SCs abuse which are common clinical features observed for amphetamine and MDMA intoxications.<sup>[172, 176-179]</sup>

The most common reported cardiac symptoms are tachycardia, chest pains and hypertension, typically these adverse effects are related to two toxidromes that are associated with SCs abuse namely sympathomimetic (clinically characterised by anxiety, paranoia, seizures, hypertension and tachycardia)<sup>[180, 181]</sup> and hallucinogenic toxidromes (clinically characterised by disorientation, psychotic episodes and hallucinations)<sup>[180, 182]</sup>, together with the risk of developing excited/agitated delirium (clinically characterised by a state of delirium, haemodynamic imbalance, incoercible psychomotor agitation and aggressiveness)<sup>[183]</sup> and serotonin syndrome (clinically characterised by mental status derangement, hyperactivity of autonomous nervous system and neuromuscular abnormalities).<sup>[13, 184]</sup>

The use of SCs has a high addiction and abuse potential, because of several SCs derivatives' ability to stimulate the dopaminergic system leading to development of not only cravings but also dependence, withdrawal syndrome and the development of tolerance to certain SCs due to repeated high dose usage.<sup>[136, 152, 153]</sup> The interactions of SCs with monoamine transporters and receptors will be discussed later on in this chapter.

## 2.3 Neuroscience Overview

### 2.3.1 Overview of the neurotransmitter cycle

The neurotransmitter cycle describes the entire process of neurotransmission which can be divided into five basic steps, although these steps might sound straightforward they are much more complicated in reality. The cycle starts with the synthesis of neurotransmitters which are stored in specialised vesicles until they are required to be released, once they are released they activate specific receptors until the signal is terminated.

(1) Neurotransmitter synthesis - Monoamine neurotransmitters are synthesised in the nerve terminals of the central nervous system (CNS) by neurotransmitter synthesising enzymes that require neurotransmitter precursors as substrates to produce the required neurotransmitters (some neurotransmitters only undergo final synthesis steps inside the synaptic vesicles). (2) Vesicle storage – Neurotransmitters are taken up into synaptic vesicles where they are stored until a stimulus is received. (3) Release - Upon receiving a stimulus, an action potential invades the terminal and causes a change in the membrane potential. This change in membrane potential leads to the opening of voltage-gated calcium channels in the presynaptic membrane. Due to the steep concentration gradient between the extracellular and intracellular compartments, calcium rapidly diffuses into the presynaptic neuron, resulting in the rise of presynaptic  $\text{Ca}^{2+}$  concentrations. Vesicles containing the neurotransmitters are translocated to the plasma membrane of the presynaptic neuron where they dock and  $\text{Ca}^{2+}$  allows for the fusion of the vesicle with the plasma membrane to release the neurotransmitters (exocytosis) into the synaptic cleft. (4) Activation - These released neurotransmitters diffuse across the synaptic cleft to interact with neurotransmitter receptors located on the postsynaptic neuron. The binding of neurotransmitters to the receptors causes channels on the postsynaptic membrane to open (or sometimes to close), therefore altering the ability of ions to flow into or out of the postsynaptic neuron. The resulting neurotransmitter current flow alters the conductance and (usually) the membrane potential, decreasing or increasing the probability of the neuron firing an action potential. This mechanism allows for the transmission of information from one neuron to another. (5) Termination - Termination of the neurotransmitter signal occurs when the neurotransmitters (DA, 5-HT and NE) are enzymatically degraded by monoamine oxidase (MAO) or catechol-O-methyltransferase (COMT) in the case of DA and NE or internalised back into the presynaptic neuron by their respective transporters. The uptake of the neurotransmitters (or their metabolites) starts a new cycle of synthesis, storage, release, activation and termination.<sup>[185]</sup>

For the scope of this study further understanding of steps (4) Activation of receptors and (5) Termination of the neurotransmitter signals via transporter reuptake are required. The mechanism for receptor activation and transporter reuptake will be discussed in more detail below.

### 2.3.2 *Monoamines (Dopamine, Serotonin and Norepinephrine)*

The monoamine dopamine (DA) is present in several regions of the brain with the major dopamine containing region being the corpus striatum which receives major input from the substantia nigra and plays an essential role in the coordination of body movements and is involved in reinforcement, cognition, emotions, motivation, and reward which makes this a primary target for many drugs of abuse. Dopamine is synthesised from the amino acid tyrosine which gets catalysed by tyrosine hydroxylase and requires oxygen (co-substrate) and tetrahydrobiopterin (cofactor) to synthesise dihydroxyphenylalanine (DOPA) which DOPA decarboxylase then converts to dopamine. Synthesised dopamine gets transported into vesicles by vesicular monoamine transporters (VMAT) where they get stored until an action potential is received which triggers the release of dopamine into the synaptic cleft. Once released, dopamine exclusively interacts with dopamine-specific G protein-coupled receptors (GPCR), which then activates or inhibits adenylate cyclase depending on the G-protein the receptor couples to.<sup>[186]</sup>

Monoamine serotonin (also known as 5-hydroxytryptamine, 5-HT or SER) is primarily found in groups of neurons in the raphe region of the pons and upper brainstem and has widespread projections to the forebrain and spinal cord. 5-HT is not exclusive to only the CNS but is also found in enterochromaffin cells, the gastrointestinal tract and platelets. This monoamine helps to regulate sleep and wakefulness and is produced from the tryptophan amino acid which gets catalysed by the enzyme tryptophan-5-hydroxylase and requires oxygen (co-substrate) to produce 5-hydroxytryptophan before being decarboxylated by the enzyme aromatic L-amino acid decarboxylase to produce serotonin. Synthesised serotonin gets transported into vesicles by VMAT where they get stored until an action potential is received which triggers the release of 5-HT into the synaptic cleft. Once released it interacts with 5-HT specific receptors of which majority of them are GPCR with the exception of the 5-HT<sub>3</sub> receptor which is a ligand-gated ion channel. These channels are non-selective cation channels and therefore mediate excitatory postsynaptic responses. The general structure of these channels are formed by an assembly of multiple subunits, only two types of 5-HT<sub>3</sub> subunits are known which form functional channels by assembling as a heteromultimer.<sup>[186]</sup>

Monoamine norepinephrine (also known as noradrenaline or NE) is a neurotransmitter used in the locus coeruleus, a brainstem nucleus that projects diffusely to a variety of forebrain targets. This neurotransmitter influences attention, sleep and wakefulness, locomotion, eating, thermoregulation, pain, cognition and sexual behaviour and is synthesised from dopamine (which is synthesised from tyrosine as described above) and catalysed by dopamine  $\beta$ -hydroxylase. Synthesised norepinephrine gets transported into vesicles by VMAT where they get stored until an action potential is received which triggers the release of norepinephrine into the synaptic cleft. Once released, norepinephrine interacts with norepinephrine-specific alpha/ $(\alpha)$ - and beta/ $(\beta)$ -adrenergic GPCR.  $\alpha$ -adrenergic receptors are further subdivided into  $\alpha$ -1 and  $\alpha$ -2 adrenergic receptors which are each further divided into three subtypes ( $\alpha$ -1A,  $\alpha$ -1B,  $\alpha$ -1D,  $\alpha$ -2A,  $\alpha$ -2B,  $\alpha$ -2C), activation of  $\alpha$ -1 receptors leads to slow depolarisation linked to inhibition of  $K^+$  channels, while the activation of  $\alpha$ -2 leads to the slow hyperpolarisation due to the activation of a different type of  $K^+$  channel. Three subtypes of  $\beta$ -adrenergic receptors ( $\beta$ -1,  $\beta$ -2,  $\beta$ -3) are also known, two of which are expressed in many types of neurons.<sup>[186]</sup>

### 2.3.3 Receptors for Dopamine, Serotonin and Norepinephrine

#### ***Dopamine Receptors***

Dopamine receptors (DAR) belong to the superfamily of GPCR that can be further divided into two subfamilies, the D1-like (comprised of  $D_1R$ ,  $D_5R$ ) and D2-like (comprised of  $D_2R$ ,  $D_3R$  and  $D_4R$ ) receptor families. The ability of DA to mediate several physiological effects is due to its ability to bind to five distinct receptors which are not only distinguishable by their unique pharmacology but also by their expression pattern, protein structure, gene structure and signal transduction pathway.  $D_1R$  and  $D_2R$  are the most highly expressed DAR and their expression is highest in areas of DA innervation, whereas  $D_3R$ ,  $D_4R$  and  $D_5R$  expression is more specialised.

Dopamine receptor structure consists of seven transmembrane spanning regions, extracellular N-terminus and three loops as well as intracellular C-terminus and three loops. There are significant sequence similarities in the transmembrane domains between members of each group.  $D_1R$  and  $D_5R$  share an overall homology of 80%, but are only 31% identical to D2-like receptors.  $D_2R$  and  $D_3R$  share a 75% homology and  $D_3R$  and  $D_4R$  share a 54% homology in their transmembrane domains. All dopamine receptors subtypes have a similar number of amino acids in their extracellular N-terminus and their loops, both of which contain a variable number of N-glycosylation sites, the second and third loops contain two cysteine residues that form disulphide bridges stabilising the receptor structure. The intracellular loops interact with

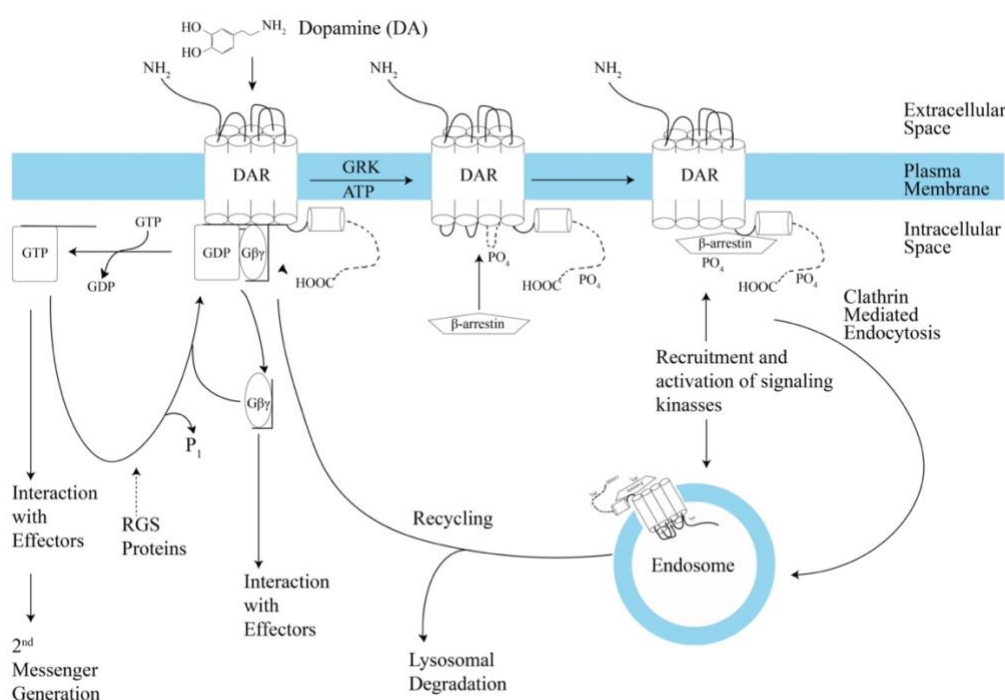
G-proteins, when these loops and C-terminus tails are phosphorylated by G protein-coupled receptor kinases (GRKs) they interact with  $\beta$ -arrestins and other kinases and signalling molecules.

The D1 family of DARs have relatively short intracellular loop 3 (ICL3) and a long C-terminal tail while the D2 family have a long ICL3 and a short C-terminal tail. The G-protein that binds to DARs is a heterotrimeric protein composed of three distinct subunits ( $\alpha$ ,  $\beta$  and  $\gamma$ ), there are many different types of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits which lead to several G-protein permutations. The activated and inhibited states of G-proteins are dependent on whether it is bound to guanosine triphosphate (GTP) or guanosine diphosphate (GDP). Generally the GDP bound protein forms an inactive trimer, once an extracellular signal binds to the GPCR, the GDP is exchanged with GTP which allows for the dissociation of the  $\alpha$  subunit from the  $\beta\gamma$  subunit complex which activates the G-protein. Following activation, both the GTP-bound  $G_\alpha$  subunit and the free  $G_{\beta\gamma}$  subunit complex can bind to downstream effectors that mediate a variety of responses in target cells.

Dopamine receptor general signal transduction pathway starts with the binding of DA to the DAR, which causes the exchange of GDP for GTP in the  $G_\alpha$  subunit and the dissociation from the  $G_{\beta\gamma}$  subunit complex.  $G_\alpha$  then goes on to interact with effectors such as ion channels. D1 family of DARs activates members of the  $G_{\alpha s}$  family (the “s” denotes stimulation) which stimulate adenylylase enzyme that catalyses the conversion of adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP) that acts as a secondary messenger in the transduction pathway that activates cAMP-dependant protein kinase A (PKA) leading to an increase in protein phosphorylation of nuclear, cytoplasmic and membrane proteins which can significantly affect the physiology and cellular regulation.

The D2 family of DAR couples to both pertussis sensitive ( $G_{\alpha i}$ , where the “i” denotes inhibition) and pertussis insensitive ( $G_{\alpha z}$ ) families resulting inhibition of adenylylase that leads to lower concentrations of cAMP and ultimately a decrease in PKA activation. The D2 family is also known to signal through  $G_{\beta\gamma}$ , which activates phospholipase C $\beta$ , G protein-coupled inward rectifying potassium channels and other proteins. Other important effector systems and transduction pathways modulated by dopamine receptor activation include arachidonic acid,  $\text{Na}^+/\text{H}^+$  exchanges, activity of the  $\text{K}^+$  and  $\text{Ca}^{2+}$  ion channels, phospholipases, mitogen-activated protein kinases (MAPK or ERK) and  $\text{Na}^+/\text{K}^+$ -ATPase.

$G_{\alpha}$  activation is terminated by the hydrolysis of GTP to GDP through RGS (Regulator of the G-protein signalling) proteins which accelerates the GTP hydrolysis associated with the  $G_{\alpha}$  subunit of the G-protein to form a  $G_{\alpha}$ -GDP complex. The  $G_{\alpha}$ -GDP complex then re-associates with  $G_{\beta\gamma}$  to form the inactive G-protein trimer. In addition, binding of DA to the DARs causes the phosphorylation of the intracellular residues of the receptor by GRKs, which stabilises the receptor in a conformation that favours  $\beta$ -arrestin binding. This mediates desensitisation of G-protein signalling as well as internalisation through interaction with clathrin-coated pits. Competent DAR that had previously been internalised can be recycled to the plasma membrane or undergo lysosomal degradation. The DAR-arrestin complex can initiate G-protein independent receptor mediated signalling in its own right.<sup>[187-192]</sup> This pathway, as well as others, are shown in **Figure 2.5** and summarised in **Table 2.2** below.



*Figure 2.5 - Dopamine receptor signal transducing pathway.<sup>[188]</sup>*

### ***Serotonin receptors***

Serotonin receptors (5-HT<sub>R</sub>) are characterised and classified by their gene organisation, amino acid sequence, pharmacological properties and second-messenger coupling pathway. All serotonin receptors belong to the GPCR superfamily, except 5-HT<sub>3</sub> which is a ligand-gated ion channel. Serotonin receptors are divided into seven receptor families namely 5-HT<sub>1</sub>, 5-HT<sub>2</sub>, 5-HT<sub>3</sub>, 5-HT<sub>4</sub>, 5-HT<sub>5</sub>, 5-HT<sub>6</sub> and 5-HT<sub>7</sub>, of which 5-HT<sub>1</sub>, 5-HT<sub>2</sub> and 5-HT<sub>5</sub> have subtypes listed in **Table 2.2** bringing the total number of serotonin receptors to 14. The basic protein structure of the serotonin receptors are predicted to contain seven transmembrane domains linked by



three extra- and intracellular loops and an extracellular N-terminus and intracellular C-terminus. These receptors make use of G-proteins to link to their signal transducing pathways.

A very similar sequence of events that occur in the dopamine pathway also occurs in the serotonin receptors, which starts with the binding of 5-HT to the serotonin receptor. The binding of 5-HT causes the bound heterotrimeric G-proteins'  $\alpha$  subunit to exchange GDP for GTP, which leads to the dissociation of the  $\beta\gamma$  subunit complex and the activation of the G-protein. It is important to note that both the GTP-bound  $G_\alpha$  subunit and the free  $G_{\beta\gamma}$  subunit complex can bind to downstream effectors that mediate a variety of responses in the target cells. The activated  $G_\alpha$  subunit interacts the effectors adenylate cyclase or phospholipase C (PLC) which triggers their activation or inhibition. In turn the effector enzymes generate secondary messengers that regulate cellular processes such as  $Ca^{2+}$  release and the activation or inhibition of protein kinases and phosphatases.<sup>[193]</sup> The activated  $G_{\beta\gamma}$  subunit complex is known to interact with several effectors including adenylate cyclase<sup>[194]</sup>, PLC $\beta$ 2 and  $\beta$ 3<sup>[195]</sup>, phosphoinositide 3-kinase (PI3 kinase)<sup>[196]</sup>, components of the MAPK cascade and  $K^+$  and  $Ca^{2+}$  channels.<sup>[197-199]</sup> These 14 receptors can be grouped into four categories according to their primary G protein-secondary-messenger linkage namely receptors that can (1) inhibit or (2) activate adenylate cyclase, (3) activate PLC and (4) ligand gated ion channels.

For simplicity only some of the pathways linked to the  $G_\alpha$  subunit in the CNS are discussed in more detail, but it should be noted that the  $G_{\beta\gamma}$  subunit complex also play a role in specific signal transducing pathways.

#### *(1) Serotonin receptors that inhibit adenylate cyclase*

The 5-HTR that inhibit adenylate cyclase belong to the 5-HT<sub>1</sub> and 5-HT<sub>5</sub> receptor families which are further subdivided into subtypes. The 5-HT<sub>1</sub> family consists of five subtypes, 5-HT<sub>1A</sub>, 5-HT<sub>1B</sub>, 5-HT<sub>1D</sub>, 5-HT<sub>1E</sub> and 5-HT<sub>1F</sub>, while the 5-HT<sub>5</sub> family consists of two subtypes, 5-HT<sub>5A</sub> and 5-HT<sub>5B</sub>. All these receptors are negatively coupled to the adenylate cyclase pathway through their  $G_{ai/o}$  subunit of their G-proteins, which when activated leads to the inhibition of adenylate cyclase leading to a decrease in cAMP secondary messengers which causes a decrease in PKA activation. Many of these receptors are also connected to a plethora of signalling pathways and effectors.<sup>[193, 200-202]</sup>

### (2) *Serotonin receptors that activate adenylate cyclase*

The 5-HTR that activate adenylate cyclase belong to the 5-HT<sub>4</sub>, 5-HT<sub>6</sub> and 5-HT<sub>7</sub> receptor families, which are connected not only to the adenylate cyclase pathway through their G<sub>as</sub> subunit of their G-proteins, but also to a plethora of signalling pathways and effectors. These three receptors activate adenylate cyclase, which directly leads to increase production of the secondary messenger cAMP and an increased activation of PKA. 5-HT<sub>4</sub> are found in the hippocampus and is involved in memory and learning and has been linked to the modulation of the release of DA, 5-HT,  $\gamma$ -aminobutyric acid (GABA) and acetylcholine. 5-HT<sub>6</sub> are found in the nucleus accumbens, striatum, cortex, amygdala and hippocampus. Limited available pharmacological data suggests that 5-HT<sub>6</sub> modulate neurotransmitters such as DA, glutamate and acetylcholine. 5-HT<sub>7</sub> are widely expressed throughout the brain, with the greatest levels of expression occurring in the hippocampus and thalamus. This receptor is believed to play a role in thermoregulation and circadian rhythms.<sup>[193, 202]</sup>

### (3) *Serotonin receptors that activate Phospholipase C*

The 5-HTR that are responsible for the activation of PLC belongs to the 5HT<sub>2</sub> receptor family which has 3 subtypes 5-HT<sub>2A</sub>, 5-HT<sub>2B</sub> and 5-HT<sub>2C</sub>. In the CNS 5-HT<sub>2A</sub> are located in the thalamus, cerebral cortex, basal ganglia and claustrum regions of the brain as well as in the spinal cord where it plays a role in pain. It is also present in the peripheral nervous system (PNS) where it is responsible for smooth muscle contraction (bronchial, tracheal, uterine and urinary), arterial vasoconstriction and platelet aggregation.<sup>[193, 202]</sup>

In the CNS 5-HT<sub>2B</sub> play a vital role in coordinating the proper formation of structures such as the cranial neural crest cells<sup>[203]</sup> and the heart<sup>[204, 205]</sup> and also occur in the PNS where they are expressed in gastro-intestinal and cardiovascular tissues including the gut, stomach fundus, cardiomyocytes, pulmonary smooth muscle as well as in the liver and kidneys.<sup>[193, 202, 206, 207]</sup> 5-HT<sub>2C</sub> are mainly, if not exclusively, expressed in the CNS. In the choroid plexus, it regulates spinal fluid production<sup>[208]</sup> and ion exchange between the brain and the cerebral spinal fluid.<sup>[193]</sup> They are also found in other regions of the brain such as amygdala, basal ganglia, cortex, hippocampus and thalamus to a lesser extent. 5-HT<sub>2C</sub> may be a therapeutic target for treating substance abuse as they can modulate mesolimbic dopamine neurotransmission.<sup>[193, 202]</sup>

These receptors bind 5-HT which leads to the activation of the membrane-bound enzyme PLC, through their coupled G-proteins (G<sub>αq/11</sub>), that catalyses the hydrolyses of phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>) to produce the secondary messengers inositol 1,4,5-triphosphate (IP<sub>3</sub>)

and diacylglycerol (DAG).<sup>[209]</sup> Secondary messenger IP<sub>3</sub> triggers the mobilisation of Ca<sup>2+</sup> from the endoplasmic reticulum (ER), Ca<sup>2+</sup> induces multiple responses such as the activation of calcium/calmodulin-dependent protein kinase enzymes that phosphorylate protein substrates in the cell. DAG on the other hand activates protein kinase C (PKC).<sup>[193, 202]</sup>

#### *(4) Serotonin 3 receptors – Ligand-Gated Ion Channel*

The 5-HT<sub>3</sub> is a ligand-gated ion channel and consists of a pentameric structure surrounding a central ion channel and contains a characteristic conserved cysteine loop in the extracellular N-terminal domain.<sup>[210]</sup>

5-HT<sub>3</sub> are found on CNS neurons located in the nucleus tractus solitarius, area postrema, hippocampus and in the PNS.<sup>[211]</sup> These receptors are located on both the presynaptic terminal where they modulate the release of neurotransmitters and postsynaptic terminal on neuronal cell bodies and dendrites where they trigger fast excitatory responses.<sup>[212]</sup> When 5-HT binds to the receptor it triggers the opening of a nonselective cation channel which causes rapid depolarisation due to the inward current generated from the influx of Na<sup>+</sup>, Ca<sup>2+</sup> and the efflux of K<sup>+</sup>.<sup>[193, 202]</sup> Activation of the receptor also leads to the stimulation of nitric oxide synthase, which in turn induces cyclic guanosine monophosphate (cGMP) production that elicits cGMP-dependant Cl<sup>-</sup> efflux.<sup>[213]</sup>

#### ***Adrenergic receptors - Epinephrine/Norepinephrine receptors***

All adrenergic receptors (AR) receptors belong to the GPCR superfamily. As previously mentioned AR are divided into three receptor families namely  $\alpha$ -1,  $\alpha$ -2 and  $\beta$ . Each of these receptor families have three subtypes ( $\alpha$ -1A,  $\alpha$ -1B,  $\alpha$ -1D,  $\alpha$ -2A,  $\alpha$ -2B,  $\alpha$ -2C,  $\beta$ -1,  $\beta$ -2 and  $\beta$ -3) bringing the total number of AR to nine. The basic protein structure of the AR are a single polypeptide chain containing seven transmembrane domains that form alpha helical structure linked by three extra- and intracellular loops and an extracellular N-terminus and intracellular C-terminus. These receptors make use of G-proteins to link to their signal transducing pathways. Norepinephrine is both a neurotransmitter and a hormone that occur in both the CNS and PNS and epinephrine is a hormone released from the adrenal gland. In the CNS AR are involved in many functions such as alertness, memory and learning.<sup>[214-218]</sup>

The norepinephrine cells in the CNS consists of two sets of neurons which are localised in the most posterior part of the brain. The first set of neurons, the locus coeruleus and adjacent nuclei, are located in the upper pons. The second set, the medullary nuclei, innervate forebrain areas

such as the septum, hypothalamus, piriform cortex and amygdala, via the ventral NE bundle and mainly play a role in the control peripheral functions such as vegetative functions and endocrine regulations.<sup>[219]</sup>

A very similar sequence of events that occur in the dopamine pathway also occur in AR, which starts with the binding of norepinephrine or epinephrine to the adrenergic receptor. The binding of norepinephrine or epinephrine causes the bound heterotrimeric G-proteins'  $\alpha$  subunit to exchange GDP for GTP, which leads to the dissociation of the  $\beta\gamma$  subunit complex and the activation of the G-protein. Similar to the dopamine pathway both the GTP-bound  $G_\alpha$  subunit and the free  $G_{\beta\gamma}$  subunit complex can bind to downstream effectors that mediate a variety of responses in the target cell. The activated  $G_\alpha$  subunit interacts with the effectors PLC and adenylate cyclase which triggers their activation or inhibition. In turn the effector enzymes generate secondary messengers that regulate cellular processes such as  $Ca^{2+}$  release and the activation or inhibition of protein kinases.

#### *Alpha-1 Adrenergic Receptors*

The  $\alpha$ -1 AR are located in both the CNS and PNS, in the CNS they are predominantly located postsynaptically in the thalamus and anterior cerebral cortex where they trigger an excitatory response. The  $\alpha$ -1 AR gets activated by either epinephrine or norepinephrine binding at the extracellular binding site of the receptor. On the intracellular side  $\alpha$ -1 AR interacts with the  $G_{\alpha_q/11}$  family of G-proteins which causes the dissociation of the  $\alpha$  and  $\beta\gamma$  subunits and the subsequent stimulation of PLC. PLC in turn catalyses the hydrolyses of phosphatidylinositol 4,5-bisphosphate ( $PIP_2$ ) to form the secondary messengers  $IP_3$  and DAG.  $IP_3$  mediates intracellular  $Ca^{2+}$  release via the  $IP_3$  receptor while DAG activates PKC. Other signalling pathways that have also been shown to be activated by  $\alpha$ -1 AR include arachidonic acid release,  $Ca^{2+}$  influx via voltage-independent and -dependent calcium channels and activation of MAPK, PLA2 and phospholipase D.<sup>[214, 216, 219-221]</sup>

#### *Alpha-2 Adrenergic Receptors*

The  $\alpha$ -2 AR are located in both the CNS and PNS, located either pre- or postsynaptically. In the CNS the  $\alpha$ -2 AR can regulate the release of neurotransmitter by acting as either an autoreceptor on noradrenergic nerve terminals or as a heteroreceptor on non-noradrenergic nerve terminals. Both  $\alpha$ -2A and  $\alpha$ -2C are involved in the suppression of catecholamines release from CNS neurons. The  $\alpha$ -2A acts as the primary presynaptic autoreceptor and appears to be critical for hypotensive effect (lowering blood pressure), anesthetic sparing, working memory

and sedation and  $\alpha$ -2C plays a critical role in the suppression of epinephrine release from the adrenal chromaffin cells. In contrast to  $\alpha$ -2A,  $\alpha$ -2B causes vascular hypertensive effects (increasing blood pressure).

The  $\alpha$ -2 AR activate the  $G_{\alpha i/o}$  family of G-proteins to alter (classically inhibit) the activity of the enzyme adenylate cyclase, which in turn inhibit the secondary messenger cAMP production. Other signalling pathways that have also been shown to be activated by  $\alpha$ -2 AR include the activation or inhibition of  $Ca^{2+}$  channels, activation of  $K^+$  channels,  $Na^+/H^+$  exchange, PLA2, PLC and MAPK.<sup>[216, 217, 219, 220]</sup>

### *Beta Adrenergic Receptors*

The  $\beta$  AR are located in both the CNS and PNS,  $\beta$ -1 AR predominate in the cerebral cortex and in the heart,  $\beta$ -2 AR predominate in the cerebellum and lungs and  $\beta$ -3 originate in the brain stem, hypothalamus, bladder and brown adipose tissue. Activation of  $\beta$ -1 AR mediates increased cardiac contractability and heart rate,  $\beta$ -2 AR mediate presynaptic norepinephrine release, bronchodilation and vasodilation, while  $\beta$ -3 located in the CNS provide multisynaptic innervation to brown and white adipose depots and in the PNS mediate relaxation of bladder and uterus, attenuated cardiac contractility, thermogenesis and lipolysis.

All  $\beta$  AR activate the  $G_{\alpha s}$  family of G proteins that activates adenylate cyclase, increasing cAMP concentrations that stimulate secondary effectors such as PKA.  $\beta$ -2 and  $\beta$ -3 can couple with  $G_{\alpha i}$  which inhibits adenylate cyclase, decreasing the concentration of cAMP which decrease PKA activity through limited activation by the secondary messengers.  $\beta$  AR have been shown to interact with many other signalling proteins including the phosphoprotein ezrin–radixin–moesin-binding phosphoprotein-50,  $Na^+/H^+$  exchanger regulatory factor and the guanine nucleotide exchange factor (CNrasGEF).<sup>[216, 218-220, 222]</sup>

The G-proteins'  $G_{\alpha}$  subunit pathways discussed here for all three monoamine receptor classes are summarised in **Table 2.2** below. These pathways should not be seen as the complete picture of each receptor class as these receptors trigger a whole plethora of transducing signalling pathways throughout the CNS and PNS. More information on all the pathways in which the monoamine receptors are involved in can be found in the listed references. Two comprehensive papers deserve a special mention: (1) Masson *et al.* (2012) titled: “Serotonergic signalling: multiple effectors and pleiotropic effects”.<sup>[202]</sup> (2) Drouin *et al.* (2017) titled: “Norepinephrine”.<sup>[219]</sup>

**Table 2.2 - Summary of monoamines receptor families, subtypes and their primary signalling pathways.**

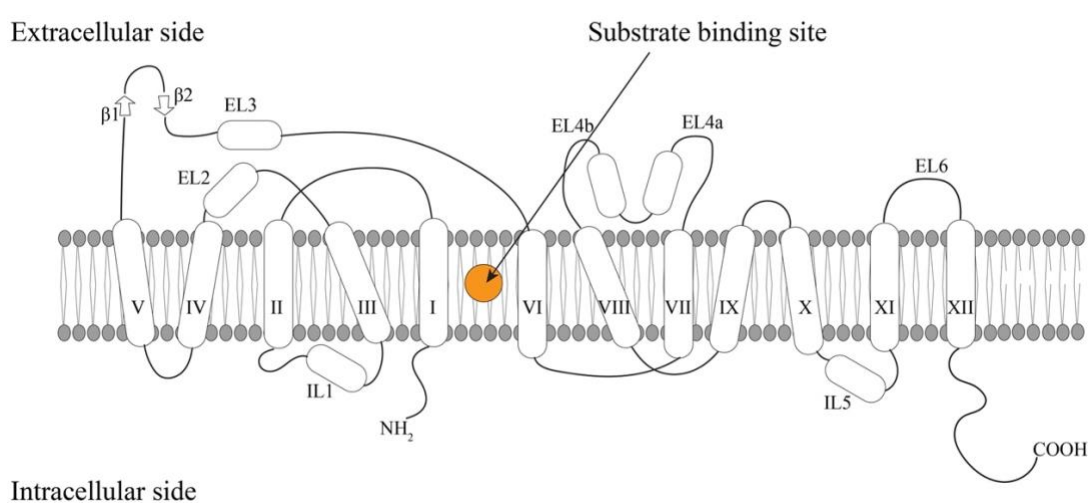
| Receptor Class   | Receptor Family   | Receptor subtype   | Signalling molecule        | Primary transduction pathway    | Secondary messenger system | Final Target   |
|------------------|-------------------|--|----------------------------|---------------------------------|----------------------------|--|
| <b>Dopamine</b>  | D <sub>1</sub>    | N/A  | G $\alpha_s$               | Adenylate cyclase activation    | Increase cAMP              | Increase protein kinase A (PKA) activity   |
|                  | D <sub>2</sub>    | D <sub>2S</sub><br>D <sub>2L</sub> (2 isoforms)  | G $\alpha_i$ /G $\alpha_z$ | Adenylate cyclase inhibition    | Decrease cAMP              | Decrease protein kinase A (PKA) activity   |
|                  | D <sub>3</sub>    | N/A  | G $\alpha_i$ /G $\alpha_z$ |                                 |                            |  |
|                  | D <sub>4</sub>    | N/A  | G $\alpha_i$ /G $\alpha_z$ |                                 |                            |  |
|                  | D <sub>5</sub>    | N/A  | G $\alpha_s$               | Adenylate cyclase activation    | Increase cAMP              | Increase protein kinase A (PKA) activity   |
| <b>Serotonin</b> | 5-HT <sub>1</sub> | 5-HT <sub>1A</sub><br>5-HT <sub>1B</sub><br>5-HT <sub>1D</sub><br>5-HT <sub>1E</sub><br>5-HT <sub>1F</sub> | G $\alpha_{i/o}$           | Inhibition of adenylate cyclase | Decrease cAMP              | Decrease protein kinase A (PKA) activity   |
|                  | 5-HT <sub>2</sub> | 5-HT <sub>2A</sub><br>5-HT <sub>2B</sub><br>5-HT <sub>2C</sub>   | G $\alpha_{q/11}$          | Activation of phospholipase C   | IP <sub>3</sub> and DAG    | IP <sub>3</sub> triggers multiple responses i.e. (calcium/calmodulin-dependent protein kinase enzymes).<br>DAG increase protein kinase C (PKC) activity. |
|                  | 5-HT <sub>3</sub> | N/A  | No G-protein               | Ligand gated ion channel        | N/A                        | Rapid depolarisation through the nonselective inward flux of Na <sup>+</sup> , Ca <sup>2+</sup> and efflux of K <sup>+</sup>                             |
|                  | 5-HT <sub>4</sub> | N/A  | G $\alpha_s$               | Activation of adenylate cyclase | Increase cAMP              | Increase protein kinase A (PKA) activity   |
|                  | 5-HT <sub>5</sub> | 5-HT <sub>5A</sub><br>5-HT <sub>5B</sub>   | G $\alpha_{i/o}$           | Inhibition of adenylate cyclase | Decrease cAMP              | Decrease protein kinase A (PKA) activity   |
|                  | 5-HT <sub>6</sub> | N/A  | G $\alpha_s$               | Activation of adenylate cyclase | Increase cAMP              | Increase protein kinase A (PKA) activity   |
|                  | 5-HT <sub>7</sub> | N/A  | G $\alpha_s$               | Activation of adenylate cyclase | Increase cAMP              |  |

|                                       |                           |              |                       |   |                         |  |
|---------------------------------------|---------------------------|--------------|-----------------------|---|-------------------------|--|
| <b>Epinephrine and Norepinephrine</b> | Alpha-1<br>( $\alpha$ -1) | $\alpha$ -1A | $G\alpha_{q/11}$      | Activation of phospholipase C               | IP <sub>3</sub> and DAG | Increase intracellular Ca <sup>2+</sup> concentration and activation of protein kinase C (PKC)   |
|                                       |                           | $\alpha$ -1B |                       |   |                         |  |
|                                       |                           | $\alpha$ -1D |                       |   |                         |  |
|                                       | Alpha-2<br>( $\alpha$ -2) | $\alpha$ -2A | $G\alpha_{i/o}$       | Inhibition of adenylyate cyclase            | Decrease cAMP           | Primarily negatively coupled to adenylyate cyclase, suppression of voltage-gated Ca <sup>2+</sup> currents, increased Ca <sup>2+</sup> release from intracellular sources, activation of inwardly rectifying K <sup>+</sup> channels, activation of phospholipase A <sub>2</sub> and C, Na <sup>+</sup> /H <sup>+</sup> exchange and MAPK. |
|                                       |                           | $\alpha$ -2B |                       |   |                         |  |
|                                       |                           | $\alpha$ -2C |                       |   |                         |  |
|                                       | Beta<br>( $\beta$ )       | $\beta$ -1   | $G\alpha_s$           | Activation of adenylyate cyclase            | Increase cAMP           | Increase protein kinase A (PKA) activity   |
|                                       |                           | $\beta$ -2   | $G\alpha_s/G\alpha_i$ | Activation/inhibition of adenylyate cyclase | Increase/Decrease cAMP  | Increase/decrease protein kinase A (PKA) activity  |
|                                       |                           | $\beta$ -3   | $G\alpha_s/G\alpha_i$ |   |                         |  |

#### 2.3.4 Transporters for Dopamine, Serotonin and Norepinephrine

Dopamine transporter (DAT, SLC6A3), serotonin transporter (SERT, SLCA4) and norepinephrine transporter (NET, SLC6A2) belongs to the solute carrier 6 (SLC6) transporter family and are imbedded in the membrane of presynaptic neuronal terminals of their respective pathways located in the CNS, among other places.<sup>[223]</sup> In the CNS monoamine transporters are exclusively expressed in their corresponding monoaminergic neuron that have projections to the neocortex, limbic forebrain and basal ganglia; these projections have interactions with and innervate other neurons in the amygdala, cortex hypothalamus and hippocampus regions of the brain.<sup>[224]</sup> Monoamine transporters are primarily responsible for the rapid (1 molecule/second) translocation of monoamine neurotransmitters, dopamine (DA), serotonin (5-HT) and norepinephrine (NE), from the extracellular space (synaptic cleft) into the neuronal cytoplasm of the presynaptic neuron, a process referred to as neurotransmitter “uptake” which is vital for neuronal homeostasis and signal termination.<sup>[225, 226]</sup> These translocated monoamines are sequestered into synaptic vesicles (*via* VMAT) for recycling or are degraded by MOA or COMT.<sup>[223]</sup> As mentioned previously monoamine neurotransmitters are responsible for controlling a number of emotional, physiological and behavioural functions and controlling their available concentrations are key. Many therapeutic drugs target these transporters for monoamine neurotransmitters to treat and/or relief disorders such as anxiety, ADHD, depression, psychostimulant abuse and schizophrenia.<sup>[223]</sup>

The general structure of monoamine transporters includes 12 alpha-helical transmembrane spanning domains (TM) connected to one another by flexible extracellular and intracellular loops (Figure 2.6), the N- and C-termini are located on the intracellular side. The high affinity orthosteric or primary substrate binding site (S1) is located between TM1 and TM6 at the core of the translocation pathways. The S1 site binds the substrate along with one or two Na<sup>+</sup> ions. The S1 pocket has a hydrophobic (non-polar) region which interacts with non-polar regions of the substrate and a hydrophilic (or polar) region which interacts with polar or charged regions of substrates, this hydrophilic region contains a highly conserved aspartate amino acid (Asp, D) [DAT: D79, SERT: D98, NET:D75].<sup>[223]</sup>



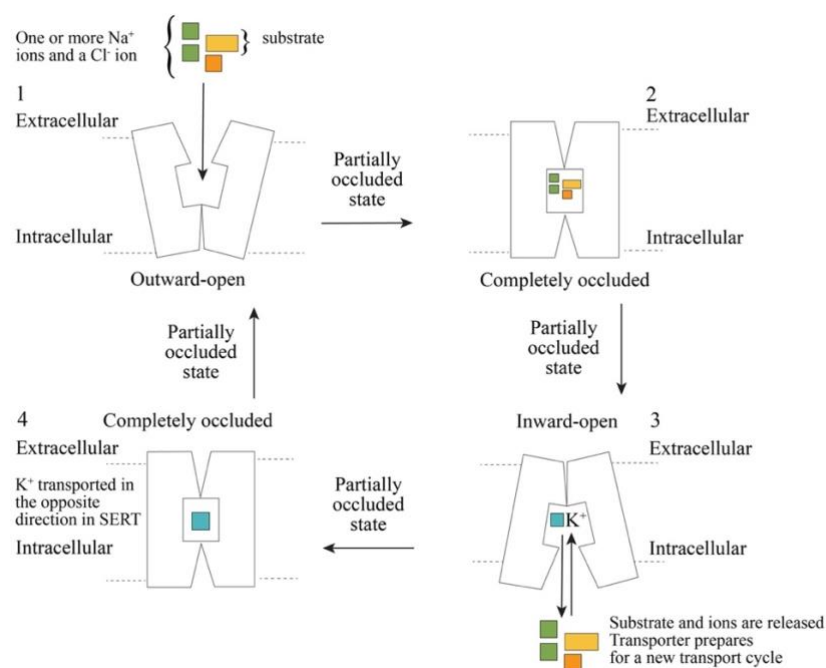
**Figure 2.6** - General representation of monoamine transporters 12-transmembrane domains, extra- and intracellular loops and the location of the N- and C-termini.<sup>[223]</sup>

Translocation of monoamines is an active-transport mechanism which involves the co-transport of ions (Na<sup>+</sup> and Cl<sup>-</sup>) with one molecule of substrate through the transporters. The transport of the monoamine substrates is favoured by the energy gradient created by the movement of Na<sup>+</sup> ions inside the cell, and driven by the gradient created by the Na<sup>+</sup>/K<sup>+</sup> ATPase. DAT and NET transport one DA and one NE molecule respectively with two Na<sup>+</sup> ions and one Cl<sup>-</sup> ion, whereas SERT co-transport one 5-HT along with one Na<sup>+</sup> and one Cl<sup>-</sup> in addition SERT also transports one K<sup>+</sup> ion in the opposite direction. These monoamine transporters are sometimes referred to as Na<sup>+</sup>/Cl<sup>-</sup> symporters.<sup>[223, 144]</sup>

Monoamine transporters are believed to follow a three-state “alternating access” mechanism (Figure 2.7). This mechanism implies that the transporter is capable of adopting three distinct conformations in an alternating manner to limit the access to the S1 binding site from either the intracellular or extracellular side of the membrane. These three states are an outward-open,



occlude and inward-open state, the transporter then returns to an occlude state before starting a new cycle with the outward-open state. The gating network consists of two pairs of charged residues and a few hydrophobic amino acids that are located on opposite sides of the S1 binding site. It has been proposed that an influx of substrates and ions into the intracellular space is accomplished through a series of sequential binding and conformational changes. In State 1 the transporter has an outward-open conformation, exposed to the extracellular side, to which  $\text{Na}^+$  and  $\text{Cl}^-$  ions bind to initiate the sequence, followed by the binding of the substrate to the central S1. State 2 is triggered by the occupation of the substrate in the S1 site triggering conformation changes in the transporter which closes the extracellular gate and forms an occluded state that traps both the ions and the substrate in the middle of the channel through gating networks on either sides. State 3 the occluded state undergoes opening of the intracellular gate which forms the inward-facing conformation, this state releases the ions and the substrate into the cytoplasm via diffusion facilitated by hydration of the site, in the case of SERT a  $\text{K}^+$  ion also binds to the transporter. State 4 the transporter returns to an occluded state similar to state 2 before returning to an outward-open conformation (in the case of SERT the bound  $\text{K}^+$  ion is released into the extracellular space).<sup>[223, 227]</sup> More detailed information and proof of the alternating access mechanism can be found in the review “SLC6 Neurotransmitter Transporters: Structure, Function, and Regulation” by Kristensen *et al.*<sup>[225]</sup>



**Figure 2.7** - Alternating access mechanism for the translocation of monoamine substrates and ions.<sup>[225]</sup>

Monoamine transporters play a vital role in controlling monoamine signal amplitude and duration of neurotransmission through the alteration of the monoamine concentration present

in the synaptic cleft of the CNS. Therefore, any modulation of the transporters can significantly affect the regulation of neuronal activity.<sup>[228]</sup> Many compounds are used to modulate or control monoamine neurotransmission in the brain which are used as therapeutic drugs or pharmacological tools, but monoamine transporters are also the main target for a number of recreational drugs and psychostimulants such as methamphetamine, cocaine, 3,4-methylenedioxy-methamphetamine (MDMA), synthetic cathinones (“bath salts”) and many more which block or reverse the transport of neurotransmitters leading to increased concentration of monoamines leading to stimulatory effects.<sup>[223]</sup>

Multiple therapeutic inhibitors with high affinity and selectivity for either DAT, SERT and NET have been discovered as a result of a number of drug discovery efforts. SERT inhibitors such as tricyclic antidepressants (i.e., amitriptyline and clomipramine) and selective inhibitors of SERT (selective serotonin-reuptake inhibitors) such as fluoxetine, sertraline, escitalopram and paroxetine are commonly prescribed for anxiety, depression and panic disorders. Bupropion (**Table 2.1, Structure 13**) and methylphenidate are both DAT inhibitors, bupropion is used as an anti-depressant and smoking cessation aid and methylphenidate is marketed for ADHD. Reboxetine is a NET inhibitor used in the treatment of depression, ADHD and panic disorders outside of the United States. Synthetic compounds that act as non-selective substrates and trigger monoamine efflux, have also been clinically used for example Adderall<sup>®</sup> which is a racemic mixture of amphetamine isomers prescribed in low doses to treat ADHD and is commonly used off-label to promote wakefulness in patients suffering from narcolepsy and to improve cognition. However, such drugs and other cognition-enhancing drugs that interact with monoamine transporters are strictly regulated and controlled due to their rewarding properties and abuse liability.<sup>[223]</sup>

### 2.3.5 *Transporter Substrates vs Blockers*

Drugs that interact with monoamine transporters can be divided into two categories based on their mode of action: (1) cocaine-like “blockers” (also known as inhibitors) that interact with transporters by binding to the orthosteric site inhibiting the transport of neurotransmitters from the extracellular space into the intracellular space in order to inhibit the signal, whereas (2) amphetamine-like “substrates” interact with transporters by binding to the orthosteric site as well, but gets translocated into the neuronal cytoplasm through the transporter channel. In the cytoplasm they trigger the efflux of intracellular neurotransmitters via reverse transport (i.e. transporter-mediated release). Transporter substrates are sometimes referred to as transporter “releasers” because they trigger the non-exocytotic transporter-mediated neurotransmitter

release from neurons.<sup>[14, 15]</sup> Both of these mechanisms achieve a common goal of dramatically increasing the extracellular monoamine concentration *in vivo* which in turn amplify cell-to-cell chemical signalling throughout the CNS.<sup>[144]</sup>

*Transporter blockers/inhibitors* acts as competitive substrates that bind to the orthosteric site of the respective transporter and competitively prevent monoamines from binding and being transported from the synaptic cleft into the neuronal cytoplasm effectively preventing the uptake of the neurotransmitters and ultimately preventing the termination of the neurotransmitter signalling.

*Transporter substrates/releasers* mechanism of action involves substrates along with sodium ions being translocated into the neuronal cells which induces inward depolarising currents<sup>[229, 230]</sup> and reverse the normal direction of transporter flux from intracellular to extracellular, triggering non-exocytotic release of neurotransmitters (i.e. transporter-mediated release).<sup>[231, 232]</sup> Due to substrate-type drugs being transported into the cytoplasm of neuronal cells, they can interact with neuronal proteins to disrupt vesicular storage by interacting with VMAT2 integrated into the membranes of synaptic vesicles of presynaptic neurons. VMAT is responsible for transporting monoamine neurotransmitters into the vesicles which in turn fuses with the neuronal membrane to release neurotransmitters into the synaptic cleft via exocytosis. Lastly they can interact with neurotransmitter synthesis to inhibit monoamine production. Both interactions with neuronal proteins lead to long-term neurotransmitter deficits.<sup>[233, 234]</sup> **Table 2.3** below summarises the fundamental differences between transporter substrates and blockers.

**Table 2.3** - Comparison between transporter substrates and blockers.<sup>[144]</sup>

| <b>Parameter</b>   | <b>Monoamine transporter substrates</b> | <b>Monoamine transporter blockers</b> |
|--|---|---------------------------------------|
| Inhibit neurotransmitter uptake                                      | Yes                                     | Yes                                   |
| Enter into neurons   | Yes                                     | No                                    |
| Trigger reverse transport (transporter-mediated release)             | Yes                                     | No                                    |
| Induce inward depolarising sodium currents                           | Yes                                     | No                                    |
| Increase extracellular concentrations of monoamine neurotransmitters | Yes                                     | Yes                                   |
| Long-term neurotoxic deficits in monoamine neurons                   | Yes                                     | No                                    |

## 2.4 Neuropharmacology of Khat and Synthetic Cathinones

### 2.4.1 Neuropharmacology of Khat

Cathinone, the main psychoactive constituent of khat, has been the subject of many studies, but very limited research has been done on the neurochemical changes produced by full extracts of khat. It is assumed that the neurochemistry of cathinone would reflect that of khat and therefore, in 1980 a series of studies looked at cathinone's effect on the release of DA, 5-HT and NE via transporters and its interaction with monoamine receptors. The stimulatory effect of cathinone is believed to be mediated by the dopaminergic system and is believed to behave in a similar manner to amphetamine.

#### *Effect of Cathinone on Transporters*

It has been shown that cathinone triggers the release of DA from rabbit caudate nucleus slices<sup>[235]</sup>, rabbit striatal tissue<sup>[236]</sup>, rat caudate putamen, rat nucleus accumbens<sup>[237]</sup>, rat striatal slices<sup>[238]</sup> and synaptosomes prepared from rat neostriatum<sup>[239]</sup> through prelabelled <sup>3</sup>H-DA and microdialysis experiments. Furthermore, it was found that cathinone increases DA concentration in the synaptic cleft in a dose-dependent manner in various tissues.<sup>[237-240]</sup>

Catecholamine reuptake inhibitors such as benztropine, cocaine, mazindol and nomifensine was found to almost completely inhibited the DA releasing effect of cathinone<sup>[240, 241]</sup>, which demonstrated that cathinone has to penetrate intraneuronal sites in order to cause its releasing effects. The penetration of cathinone into intraneuronal sites indicates that cathinone acts by inducing the release of DA rather than inhibiting the reuptake of released DA. It can therefore be said that cathinone acts as a transporter substrate, that interacts with DAT to promote the inward movement of cathinone that accelerates the counter-transport of DA into the synaptic cleft effectively increasing its concentration.<sup>[95]</sup> A neurochemical finding in agreement that cathinone acts as a DAT substrate is that of Mereu *et al.* (1983) that demonstrated that acute administration of cathinone leads to decreased concentration of the DA metabolite 3,4-dihydroxyphenylacetic acid (DOPAC).<sup>[242]</sup> Decreased DOPAC in tissue samples is characteristic of enhanced dopamine release and has been shown to occur in rat caudate nucleus and nucleus accumbens.<sup>[243]</sup>

So far there is no clear-cut evidence on the role of serotonergic, noradrenergic and/or other pathways in the stimulatory effect of cathinone. Conflicting information has been reported by

various investigators, some reporting that repeated administration of cathinone in rat brain samples produced unchanged levels of 5-HT<sup>[239, 243]</sup> and others reporting 5-HT releasing effects in rat caudate nucleus prelabelled with <sup>3</sup>H-5-HT.<sup>[244]</sup> Shortall *et al.* (2013) reported elevated levels of serotonin in rat striatum and hypothalamus.<sup>[245]</sup> Moreover, cathinone was shown to decrease the uptake of <sup>3</sup>H-5-HT by SERT by up to 7% (single dose administration) and 58% (multiple dose administrations) in striatal synaptosomes from drug-treated rats.<sup>[246]</sup>

Cathinone has also been shown to increase the efflux of prelabelled <sup>3</sup>H-NE from rabbit atrium tissue<sup>[247]</sup>, in agreement with this study Rothman *et al.* (2003) reported that cathinone acts as a substrate at both the DAT and NET in cloned human receptors (receptorome), with cathinone being more potent at NET<sup>[248]</sup>. In contrast to these findings Cleary and Docherty (2003) in rat left ventricular slices reported that cathinone possibly acts as a competitive inhibitor of the NET, which also has the ability to passively enter nerve terminals and cause carrier-mediated release of norepinephrine.<sup>[249]</sup>

A fairly recent publication by Simmler *et al.* (2013) using transfected human embryonic kidney HEK 293 cells support some of the results obtained above. This study examined (1) the potency of cathinone to inhibit monoamine transporters DAT, SERT and NET and (2) its binding affinity to monoamine transporters, dopamine receptors, serotonergic receptors and noradrenergic receptors. It found that cathinone potently inhibited NET and acted as a substrate at DAT.<sup>[16]</sup>

#### *Effect of Cathinone on Monoamine Receptors*

Cathinone interaction with monoamine transporters have been extensively studied since the 1980's, but interaction with DA, 5-HT and NE receptors have received little attention due to the limited information available at the time. One of the first studies on the 5-HT receptors was done by Glennon and Liebowitz in 1982 where they determined that cathinone interacts with receptors (specific receptors not specified) in a competitive manner and that it had an affinity four times greater than amphetamine.<sup>[250]</sup> Rothman *et al.* (2003) reported weak binding affinity for cathinone at  $\alpha$ -2 AR (weakest affinity for  $\alpha$ -2A) and 5-HT<sub>7</sub>, but no significant activity at  $\beta$ -AR and  $\alpha$ -1 AR.<sup>[248]</sup> Simmler *et al.* (2013) reported weak binding affinity (<10  $\mu$ M) at  $\alpha$ -1A and  $\alpha$ -2A receptors and no binding affinity for 5-HT<sub>(1A,2A,2C)</sub> receptors.<sup>[16]</sup>

#### 2.4.2 Neuropharmacology of Synthetic cathinones

SCs as previously mentioned resembles not only the natural occurring cathinone, which provides the backbone to which other substituents can be attached to produce a whole spectrum of SCs, but also amphetamines which lack the ketone functional at the beta position.<sup>[251, 152]</sup> It would be a logical conclusion that due to the similarity in the structures between SCs and classical amphetamines that SCs would interact with transporters and receptors within the brain in a similar manner.

##### *Effect of Synthetic Cathinones on Transporters*

SCs interact with monoamine transporters of DA, 5-HT and NE namely DAT, SERT and NET respectively, acting as blockers and/or substrates with the common goal of increasing the monoamine concentration in the synaptic cleft and consequently leading to hyperstimulation of the postsynaptic receptors.<sup>[28-30, 251-253]</sup> Although all SCs share the phenethylamine core their potency, affinity and selectivity for monoamine transporters and receptors vary wildly resulting in different pharmacological effects observed. Therefore, the potency, affinity and selectivity a specific SC has for a monoamine system (dopaminergic, serotonergic and noradrenergic) is of utmost importance as the interaction with a specific system, or multiple simultaneous systems, leads to different clinical and toxic effects observed. A classification system based on the pharmacological action, rather than chemical structures has been proposed, since SCs belonging to the same structural sub-family induces a broad spectrum of pharmacological effects.<sup>[16-18]</sup>

*In vitro*, almost all SCs are potent NET uptake inhibitors, the main difference regarding their pharmacological profiles seems to result from their action on dopaminergic and serotonergic systems.<sup>[20]</sup> These SCs can be grouped into three categories according to their ability to inhibit or act as releasers at monoamine transporters:

(1) MDMA-cocaine like SCs

Selective uptake inhibition of SERT, resembling MDMA

Non-selective uptake inhibition of DAT, SERT and NET similar to cocaine

Also induce transporter-mediated monoamine release of DA, 5-HT and NE

(2) Methamphetamine-like SCs

Selective uptake inhibition of DAT when compared to SERT

Potent uptake inhibitor of NET

Also induce receptor-mediated monoamine release (not only, but specifically DA)

(3) Selective catecholamine uptake inhibitors

Extremely potent inhibitor of DAT and NET with weak to negligible potency at SERT

Lacking receptor-mediated monoamine releasing properties

**Table 2.4** below consists of findings from multiple sources. As previously discussed SCs can be pure inhibitors or substrates. However, all substrates, including endogenous substrates such as DA, 5-HT and NE present uptake inhibition properties due to their competition for the same transporter.<sup>[25]</sup> Therefore, it should be noted that in the table below for a specific monoamine transporter a SCs can act as both an inhibitor or as a substrate and therefore has a tick mark in both columns in cases like these the SCs is considered to act as a substrate. In cases where there is only a tick mark in the inhibitor column but none in the substrate column it can be considered that the SCs is a pure inhibitor. All interactions listed in **Table 2.4** (1) are tests done not only on human models, but on animal models as well (2) to be considered an interactions the IC<sub>50</sub> and EC<sub>50</sub> value had to be  $\leq 10 \mu\text{M}$  taking into account the standard error of the mean, standard deviation and confidence intervals where applicable, if the lower parameter places the value  $\leq 10 \mu\text{M}$  it was included.

*Effect of Synthetic Cathinones on Monoamine Receptors*

SCs not only interact with monoamine transporters but to a lesser extent interact with the receptors for 5-HT and NE. These receptor interactions have not been the focus of many studies, and therefore not much literature is available. **Table 2.5** consists of findings from multiple sources that have been summarised, from this table it is clear that the main receptor targets are the serotonergic and adrenergic receptors, this sentiment is echoed throughout multiple studies.<sup>[16-19]</sup> Activation of 5-HT<sub>2A</sub> is associated with hallucinogenic properties and noradrenergic receptors are involved in the sympathomimetic stimulation leading to vasoconstriction, increased heart rate and blood pressure and hyperthermia.<sup>[20-22]</sup>

All interactions listed in **Table 2.5** (1) are tests done on human receptors (2) are merely binding interactions of SCs with receptors and does not equate to receptor activation, (3) to be considered an interactions the K<sub>i</sub> (inhibitor constant) value had to be  $\leq 10 \mu\text{M}$  taking into account the standard error of the mean and the standard deviation where applicable, if the lower parameter places the value  $\leq 10 \mu\text{M}$  it was included.

**Table 2.4 - Synthetic Cathinones and their monoamine transporter targets.**

| Synthetic cathinone                                  | Dopamine Transporter (DAT) |                             | Serotonin Transporter (SERT) |                             | Norepinephrine Transporter (NET) |                                 | Additional References            |
|--|----------------------------|-----------------------------|------------------------------|-----------------------------|----------------------------------|---------------------------------|----------------------------------|
|  | Inhibitor                  | Substrate                   | Inhibitor                    | Substrate                   | Inhibitor                        | Substrate                       |                                  |
| <b>(1) MDMA-cocaine like SCs</b>                     |                            |                             |                              |                             |                                  |                                 |                                  |
| <b>MDMA-like selective reuptake inhibitor</b>        |                            |                             |                              |                             |                                  |                                 |                                  |
| 2,3-Dimethylmethcathinone                            | ✓ <sup>[254]</sup>         | ✓ <sup>[254]</sup>          | ✓ <sup>[254]</sup>           | ✓ <sup>[254]</sup>          | ✓ <sup>[254]</sup>               | ✓ <sup>[254]</sup>              |                                  |
| 2,4-Dimethylmethcathinone                            | ✓ <sup>[254]</sup>         | ✓ <sup>[254]</sup>          | ✓ <sup>[254]</sup>           | ✓ <sup>[254]</sup>          | ✓ <sup>[254]</sup>               | ✓ <sup>[254]</sup>              |                                  |
| 3,4-Dimethylmethcathinone                            | ✓ <sup>[254]</sup>         | ✓ <sup>[254]</sup>          | ✓ <sup>[254]</sup>           |                             | ✓ <sup>[254]</sup>               | ✓ <sup>[254]</sup>              |                                  |
| 4-Ethylmethcathinone                                 | ✓ <sup>[19]</sup>          | ✓ <sup>[19]</sup>           | ✓ <sup>[19]</sup>            | ✓ <sup>[19]</sup>           |                                  | ✓ <sup>[19]</sup>               |                                  |
| Methedrone   | ✓ <sup>[18]</sup>          |                             | ✓ <sup>[18]</sup>            | ✓ <sup>[18]</sup>           | ✓ <sup>[18]</sup>                | ✓ <sup>[18]</sup>               |                                  |
| <b>Cocaine-like non-selective reuptake inhibitor</b> |                            |                             |                              |                             |                                  |                                 |                                  |
| 4-Methylethcathinone (4-MEC)                         | ✓ <sup>[18]</sup>          |                             | ✓ <sup>[18]</sup>            | ✓ <sup>[18]</sup>           | ✓ <sup>[18]</sup>                | ✓ <sup>[255]</sup>              | [23, 24, 26, 27, 253, 256]       |
| Mephedrone   | ✓ <sup>[29, 254]</sup>     | ✓ <sup>[19, 254, 257]</sup> | ✓ <sup>[29, 254]</sup>       | ✓ <sup>[16, 254, 257]</sup> | ✓ <sup>[29, 254]</sup>           | ✓ <sup>[19, 28, 254, 257]</sup> |                                  |
| Methylone  | ✓ <sup>[16, 29]</sup>      | ✓ <sup>[257]</sup>          | ✓ <sup>[29, 257]</sup>       | ✓ <sup>[16, 257]</sup>      | ✓ <sup>[16, 29]</sup>            | ✓ <sup>[28, 257]</sup>          |                                  |
| Ethylone   | ✓ <sup>[16]</sup>          |                             | ✓ <sup>[16]</sup>            | ✓ <sup>[16]</sup>           | ✓ <sup>[16]</sup>                |                                 |                                  |
| Butylone   | ✓ <sup>[16, 29]</sup>      |                             | ✓ <sup>[16]</sup>            | ✓ <sup>[16]</sup>           | ✓ <sup>[16, 29]</sup>            |                                 |                                  |
| Pentylone  | ✓ <sup>[18]</sup>          |                             | ✓ <sup>[18]</sup>            | ✓ <sup>[18]</sup>           | ✓ <sup>[18]</sup>                |                                 |                                  |
| β-naphyrone  | ✓ <sup>[257]</sup>         |                             | ✓ <sup>[19, 257]</sup>       |                             | ✓ <sup>[257]</sup>               |                                 |                                  |
| <b>(2) Methamphetamine-like SCs</b>                  |                            |                             |                              |                             |                                  |                                 |                                  |
| N,N-Dimethylcathinone                                | ✓ <sup>[18]</sup>          |                             |                              |                             | ✓ <sup>[18]</sup>                |                                 | [23, 88]                         |
| Buphedrone   | ✓ <sup>[18]</sup>          |                             |                              |                             | ✓ <sup>[18]</sup>                | ✓ <sup>[18]</sup>               |                                  |
| Flephedrone  | ✓ <sup>[16]</sup>          | ✓ <sup>[16, 19]</sup>       | ✓ <sup>[16]</sup>            | ✓ <sup>[16]</sup>           | ✓ <sup>[16]</sup>                | ✓ <sup>[16]</sup>               |                                  |
| Cathinone  | ✓ <sup>[16]</sup>          | ✓ <sup>[16]</sup>           |                              |                             | ✓ <sup>[16]</sup>                | ✓ <sup>[248]</sup>              |                                  |
| Methcathinone  | ✓ <sup>[16]</sup>          | ✓ <sup>[16]</sup>           | ✓ <sup>[258]</sup>           | ✓ <sup>[258]</sup>          | ✓ <sup>[16]</sup>                | ✓ <sup>[16]</sup>               |                                  |
| Pentadone  | ✓ <sup>[18]</sup>          | y                           |                              |                             | ✓ <sup>[18]</sup>                |                                 |                                  |
| <b>(3) Selective catecholamine uptake inhibitors</b> |                            |                             |                              |                             |                                  |                                 |                                  |
| Pyrovalerone   | ✓ <sup>[13]</sup>          | X <sup>a</sup>              | Weak to negligible potency   | X <sup>a</sup>              | ✓ <sup>[13]</sup>                | X <sup>a</sup>                  | [18, 19, 253, 255, 257, 259-262] |
| α-PVP  | ✓ <sup>[13]</sup>          | X <sup>a</sup>              |                              | X <sup>a</sup>              | ✓ <sup>[13]</sup>                | X <sup>a</sup>                  |                                  |
| MDPBP  | ✓ <sup>[13]</sup>          | X <sup>a</sup>              |                              | X <sup>a</sup>              | ✓ <sup>[13]</sup>                | X <sup>a</sup>                  |                                  |
| MDPV   | ✓ <sup>[13]</sup>          | X <sup>a</sup>              |                              | X <sup>a</sup>              | ✓ <sup>[13]</sup>                | X <sup>a</sup>                  |                                  |
| MDPPP  | ✓ <sup>[13]</sup>          | X <sup>a</sup>              |                              | X <sup>a</sup>              | ✓ <sup>[13]</sup>                | X <sup>a</sup>                  |                                  |

X<sup>a</sup> - As per definition of selective uptake inhibitors all these SCs do not act as substrates



**Table 2.5 - Synthetic Cathinones and their monoamine receptor targets.**

| Synthetic cathinone                                  | Dopamine receptor | Serotonin receptors        |                             |                         | Norepinephrine receptors |                    |
|--|-------------------|----------------------------|-----------------------------|-------------------------|--------------------------|--------------------|
|  |                   | D3                         | 5-HT <sub>1A</sub>          | 5-HT <sub>2A</sub>      | 5-HT <sub>2C</sub>       | α-1A               |
| <b>(1) MDMA-cocaine like SCs</b>                     |                   |                            |                             |                         |                          |                    |
| MDMA-like <b>selective</b> reuptake inhibitor        |                   |                            |                             |                         |                          |                    |
| 3-Methylmethcathinone                                |                   | √ <sup>[254]</sup>         | √ <sup>[254]</sup>          | √ <sup>[254]</sup>      | √ <sup>[254]</sup>       | √ <sup>[254]</sup> |
| 4-Methylethcathinone                                 |                   |                            | √ <sup>[18]</sup>           | √ <sup>[18]</sup>       |                          |                    |
| 2,3-Dimethylmethcathinone                            |                   |                            | √ <sup>[254]</sup>          | √ <sup>[254]</sup>      | √ <sup>[254]</sup>       | √ <sup>[254]</sup> |
| 2,4-Dimethylmethcathinone                            |                   |                            | √ <sup>[254]</sup>          | √ <sup>[254]</sup>      | √ <sup>[254]</sup>       | √ <sup>[254]</sup> |
| 3,4-Dimethylmethcathinone                            |                   |                            | √ <sup>[254]</sup>          | √ <sup>[254]</sup>      | √ <sup>[254]</sup>       | √ <sup>[254]</sup> |
| 2,3-Dimethylmethcathinone                            |                   |                            | √ <sup>[254]</sup>          | √ <sup>[254]</sup>      | √ <sup>[254]</sup>       | √ <sup>[254]</sup> |
| 3-Fluoromethcathinone                                |                   |                            |                             | √ <sup>[18]</sup>       |                          | √ <sup>[18]</sup>  |
| 4-Fluoromethcathinone                                |                   |                            | √ <sup>[257]</sup>          |                         |                          |                    |
| 4-Bromomethcathinone                                 |                   |                            | √ <sup>[19]</sup>           |                         |                          |                    |
| 4-Ethylmethcathinone                                 |                   |                            | √ <sup>[19]</sup>           | √ <sup>[19]</sup>       |                          |                    |
| Cocaine-like <b>non-selective</b> reuptake inhibitor |                   |                            |                             |                         |                          |                    |
| Mephedrone   |                   |                            | √ <sup>[16, 254, 257]</sup> | √ <sup>[254, 257]</sup> | √ <sup>[16, 254]</sup>   | √ <sup>[254]</sup> |
| Methylone  |                   |                            |                             | √ <sup>[257]</sup>      |                          |                    |
| Butylone   |                   |                            |                             |                         |                          |                    |
| β-naphyrone  |                   | √ <sup>[16, 19, 257]</sup> | √ <sup>[257]</sup>          |                         |                          | √ <sup>[16]</sup>  |
| Ethcathinone   |                   | √ <sup>[18]</sup>          |                             | √ <sup>[18]</sup>       |                          |                    |
| <b>(2) Methamphetamine-like SCs</b>                  |                   |                            |                             |                         |                          |                    |
| N,N-Dimethylcathinone                                |                   |                            | √ <sup>[18]</sup>           | √ <sup>[18]</sup>       |                          |                    |
| Flephedrone  |                   |                            | √ <sup>[16]</sup>           |                         | √ <sup>[16]</sup>        |                    |
| Methcathinone  |                   |                            | √ <sup>[16, 19]</sup>       |                         | √ <sup>[16]</sup>        |                    |
| <b>(3) Selective catecholamine uptake inhibitors</b> |                   |                            |                             |                         |                          |                    |
| Pyrovalerone   | √ <sup>[16]</sup> |                            |                             |                         |                          |                    |
| α-PVP  |                   | √ <sup>[19]</sup>          |                             |                         |                          |                    |
| MDPV   | √ <sup>[16]</sup> | √ <sup>[19]</sup>          |                             |                         |                          |                    |
| MDPPP  |                   | √ <sup>[19]</sup>          | √ <sup>[19]</sup>           |                         |                          |                    |

## 2.5 2-Methyl-4'-(methylthio)-2-morpholinopropiophenone - MMMP

### 2.5.1 Background information of MMMP

As mentioned earlier MMMP (**Figure 1.1**) can be classified as a highly modified cathinone, which is most readily used in the polymer and printing industry in the fixing of thin films, plastics and inks as a Type 1 fragmenting photoinitiator.<sup>[31]</sup> Other photoinitiators have successfully been used for biological applications such as blood vessel adhesion<sup>[263]</sup>, encapsulation of pancreatic islet cells<sup>[264, 265]</sup> and even in the dentistry industry.<sup>[266]</sup>

Photoinitiators are used as catalysers for inks and lacquers that are cured with ultraviolet (UV) light, UV curing inks have major advantages over conventional methods of printing on paper, carton or even plastic packaging as it doesn't contaminate foodstuffs with organic solvent residues, but photoinitiator in these types of ink have been shown to contaminate food through mass transference.<sup>[267, 268]</sup> It has also been detected in intravenous injection bag solutions<sup>[269]</sup> and packaged milk.<sup>[267, 270]</sup>

Reports of MMMP being found in drug seizures have emerged in South Australian and in the United States. A case published by Nash *et al.* (2019) reports one such case in which both MMMP and furanylfentanyl had been found in the post-mortem peripheral blood of a deceased 44 year old male known to be a poly-substance abuser.<sup>[31]</sup> A bit closer to home MMMP had been isolated and identified as part of seized tablets that contained N-ethylpentylone ( $\beta$ -keto-ethylbenzodioxolypentanamine) by the South African Police Service (SAPS) Forensic Science Laboratory in late 2016. MMMP's structure was confirmed but it was not reported as illicit substance due to the presence of the already scheduled N-ethylpentylone. Due to the limited information about the structural activity of MMMP it was labelled as a possible contaminant or adulterant in the tablets (personal communication).

To date limited literature is available on the potential health risks, pharmacokinetics, pharmacodynamics and the crystal structure of MMMP, which is concerning considering that MMMP is classified as an unrestricted synthetic cathinone that is easily obtainable in bulk quantities that gets exposed to foodstuffs, including baby formula and milk cartons on a daily basis.

### 2.5.2 Common names of MMMP

MMMP is also known as by the names 2-methyl-1-(4-methylthiophenyl)-2-morpholinopropan-1-one, Irgacure 907, Caccure 907, Speedcure 97 and Acetocure 97 and by the abbreviations MTMP and MMTMP.

### 2.5.3 Chemistry of MMMP

MMMP has a methylthio ring which is analogous to 4-methylthioamphetamine, a dimethyl on the  $\alpha$ -carbon analogues to phentermine, a morpholino group which contains the nitrogen and a carbonyl group on the  $\beta$ -carbon. Coincidentally it has a common cathinone backbone (**Figure 2.2a**) leading to its classification as a synthetic cathinone. During irradiation of this photoinitiator two types of scissions can occur to produce two stable radicals which initiates polymerisation. In the primary (main) type, scission occurs between the carbonyl carbon and alpha tertiary carbon bond to produce a benzoyl and an alkylamino radical. In the secondary type, scission occurs between the alpha tertiary carbon and the nitrogen containing morpholino group to produce a phenacyl and morpholino radical (**Figure 2.8**).<sup>[271-273]</sup>

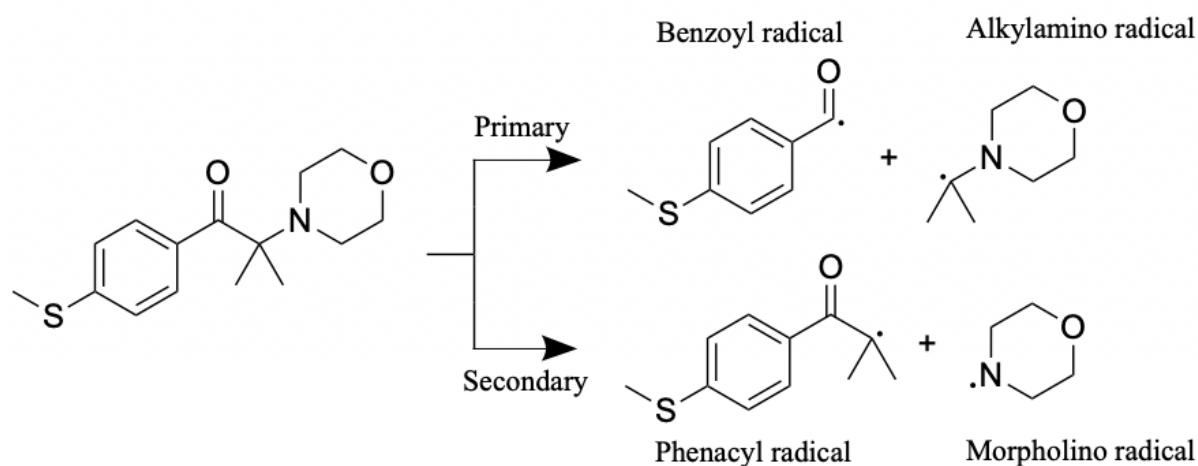


Figure 2.8 - Photocleavage of MMMP.

### 2.5.4 Pharmacokinetics

*Administration.* Routes of administration specifically for MMMP have to our knowledge not been reported, but it could be safe to assume that administration would occur in a similar way as other SCs with oral and nasal insufflation being the main routes. The discovery of N-ethylpentylone and MMMP tablets by the SAPS would support the hypothesis of oral administration, these tablets can be crushed for snorting or “keying” although there has been a report of MMMP powder found at a scene.<sup>[31, 136, 152]</sup> Other routes of administration such as intravenous (i.v.) injection, subcutaneous (s.c.) and intramuscular injections, sublingual and gingival delivery, rectal insertion, inhalation (vaporisation/smoking of e-cigarettes and

insertion of substances into the eyes could all be possible routes of administration for MMMP.<sup>[1, 153-155]</sup> *Distribution.* To our knowledge no information on MMMP distribution has been reported.

*Metabolism.* A wide variety of SCs undergo phase I metabolism which involve N-dealkylation and keto-reduction, however MMMP structurally differs in two respects to other SCs which may influence its metabolism: two methyl groups attached to the  $\alpha$ -carbon and a 4-methylthio group attached to the aromatic ring. It has been hypothesised that the sulphur group will be metabolised by sulfoxidation to both the sulfoxide and sulphone based on other sulphur containing drugs metabolism. *Elimination* of MMMP is most likely through the urine in the form of metabolites (beta-hydroxy-MMMP, beta-hydroxy-MMMP-sulfoxide and beta-hydroxy-MMMP-sulfone had been identified by Nash *et al.* (2019) as possible metabolites) or in an unchanged form.<sup>[31]</sup>

*Toxicology.* Negative health effects directly linked to photoinitiators have been reported in case studies. The photoinitiator benzophenone used in sunscreen has been reported to induce allergic skin reactions<sup>[274]</sup> similar to skin irritants that cause photoallergic reactions, allergic contact dermatitis<sup>[275]</sup> and facial erythema.<sup>[276]</sup> Isopropylthioxanthone (ITX), another photoinitiator, has been the cause of an alert by the Rapid Alert System for Food and Feed (RASFF) after Italian authorities informed the European Commission that 2-ITX had been detected in milk and fruit beverages packaged in UV-printed cartons.<sup>[277-279]</sup> The consequence of this alert led to the recall and destruction of 30 million litres of ready-to-feed infant formula by producers in Italy, France, Portugal and Spain in 2005.<sup>[280]</sup>

MMMP itself has been the cause of several notification from the European Commission's RASFF in combination with other packaging related substances in printed plastic drinking cups and packaged foodstuffs.<sup>[31]</sup> To date there has been no published/reported toxic or psychoactive effects for MMMP in humans, although it has been shown to be cytotoxic to human peripheral blood mononuclear cells<sup>[269]</sup> and that cell death occurs via apoptosis in both a caspase independent and dependant pathway, it is still unclear which pathway is more critical in the MMMP-mediated apoptosis.<sup>[281]</sup>

A study done by Kawasaki *et al.* (2015) showed that MMMP in combination with 1,2-dichloropropane produced a more potent cytotoxic effect and induced apoptosis in MRC-5, normal human embryonic lung fibroblasts cells, which may lead to acute cytotoxicity and increase the risk of respiratory diseases.<sup>[32]</sup>

MMMP along with five other photoinitiators have also been shown to significantly increase MCF-7, estrogen-sensitive human breast cancer cell line, cells and have been identified as endocrine-disrupting compounds (EDC) which interacts with estrogen receptors as agonists.<sup>[33]</sup> Exposure to EDC's may lead to serious abnormalities, including formation of several hormone-dependent cancers such as ovarian and testicular and impaired reproductive function.<sup>[282, 283]</sup> Furthermore, Takai *et al.* (2018) reported that UV-irradiated MMMP is capable of producing frameshift mutations in *Salmonella typhimurium* strain TA 97, a highly sensitive frameshift-type strain.<sup>[284]</sup> Lastly, tests done on animal models (Sprague-Dawley rats) found that MMMP has acute oral toxicity and may damage fertility and the unborn offspring.<sup>[34]</sup>

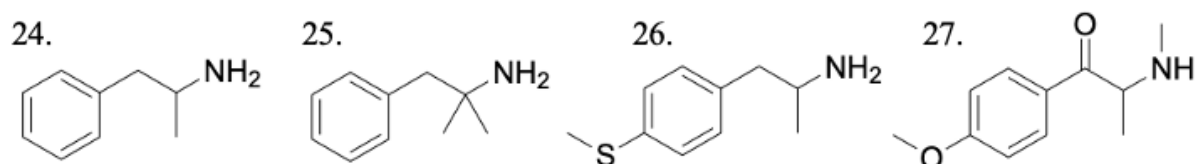
#### 2.5.5 Pharmacodynamics

To our knowledge there is no published data on MMMP and its possible psychoactive effects or its pharmacological potential as a drug of abuse. No SC have been reported, to our knowledge, containing a morpholino, methylthio or di-alpha methyl modification, making pharmacological activity predictions based on structural relationships very difficult.

The most interesting modification of MMMP is the morpholino group, even though modifying the nitrogen to be included into a cyclic structure such as a pyrrolidine is not uncommon in some SC derivatives for example pyrovalerone,  $\alpha$ -PVP and MDPV (**Table 2.1, Structures 14, 16 and 23 respectively**).<sup>[31]</sup> In the case of MDPV, the pyrrolidine structure has been shown to contribute to the molecule's function as a dopamine and norepinephrine re-uptake inhibitor which enhances its psychoactive effects, however the significance of the morpholino ring is unknown in MMMP.<sup>[144, 285]</sup>

The second methyl at the  $\alpha$ -carbon might be analogues to the difference between amphetamine (**Figure 2.9, Structure 24**) and phentermine (**Figure 2.9, Structure 25**). Phentermine invokes less CNS stimulation and dopaminergic response and lowers its addiction and abuse potential.<sup>[286, 287]</sup> Substitution at the 4-position of the aromatic ring is a very common modification in amphetamines and SCs, this modification tends to increase serotonin release and lead to entactogenic responses similar to MDMA<sup>[22]</sup>. Examples of 4-substituted amphetamine and SC that resembles MMMP are 4-methylthioamphetamine (4-MTA; **Figure 2.9, Structure 26**) and 4-methoxymethcathinone (methedrone; **Figure 2.9, Structure 27**) which has both been linked to death and hyperthermic effects.

It should be noted that extrapolation of the activity of known psychoactive phenethylamines to new SCs based on their structural relationships to other known SCs should be approached with caution as activity predictions cannot be reliably made without extensive intoxication and pharmacological reports. Taking this into consideration it is clear that the need for pharmacological data including complete pharmacokinetics and pharmacodynamics studies of MMMP to determine amongst other the psychoactive and toxicological properties and possibly regulate human exposure to these substances.



*Figure 2.9 - Structures of (24) amphetamine, (25) phentermine, (26) 4-MTA and (27) methedrone.*

## **Chapter 3: Molecular modelling**

### **3.1 Introduction**

Computer aided molecular docking studies is a technique aimed at predicting the interaction between a protein and a ligand/drug candidate which has most commonly been employed in computer aided drug design (CADD). Molecular docking studies have in the past decade become increasingly popular with the dramatic advancement of computational power.<sup>[288-290]</sup> With these advancements it has become possible to model the interface between biology and chemistry to allow for the development and discovery of pharmaceutical drugs with the aid of computational techniques. The increasing crystal structure library and the development of sophisticated computational software has paved the way for *in silico* modelling and has become an integral part of novel drug discovery and development.

Molecular docking studies has paved the way for rational drug design that spare scarce resources, time and money with several examples echoing this sentiment.<sup>[288-290]</sup> One case of particular interest in which CADD methods had successfully been employed to develop a selective serotonin receptor agonist which reached clinical trials within two years from inception.<sup>[291]</sup> While this is an unusual case it supports the use of *in silico*-based methods for more effective drug development processes.

Computational chemistry typically applies calculations based on quantum mechanics (QM) which accurately model systems without compromise by approximating the Schrödinger equation. However, the resources in terms of time and computational power required are substantial and increase exponentially with the number of atoms in a system. As a result of this, quantum mechanical systems that are being studied are usually very small in comparison to biological systems that consists of thousands of atoms such as receptors and enzymes, and therefore applying first principles to these large systems are not feasible with current computational power.

Due to computational limitations, molecular mechanics (MM) methods are employed which allows for the study of larger systems, such as biological systems, that would otherwise be impossible to study using QM based methods. MM relies on force fields which approximate how molecules twist and interact with one another and dramatically simplifies the mathematical complexity of systems under investigation, but this comes with the downside of sacrificing accuracy to enable faster processing.

This study will employ molecular modelling techniques to analyse if an isolated molecule known to be a derivative of an already scheduled drug is able to interact with identified biological drug targets (monoamine receptors and transporter) to elicit psychoactive effects.

This chapter will serve as a brief summary and introduction to the theoretical background of molecular modelling. In the first part docking programs and how they work will be discussed, thereafter an explanation of MM will follow and how force fields are developed to model large systems. Finally, molecular dynamics (MD) simulations will be discussed and why this method has become increasingly popular to model biological systems.

## 3.2 Docking

Docking involves placing an identified ligand into a binding pocket of a known target and scoring that pose. The docking score can then be used as an indication of the possibility of the ligand forming a ligand-protein complex.

### 3.2.1 Docking Programs

The aim of docking programs are to (i) predict the binding pose of a ligand within the binding pocket of a target protein (or DNA in rare cases<sup>[292, 293]</sup>) and (ii) estimate the binding affinity of a docked ligand.

Docking studies can be done by using commercially available software such as Glide<sup>[44, 45, 294]</sup> and GOLD<sup>[295]</sup> or open-source software such as Qvina<sup>[296]</sup>, AutoDock<sup>[297]</sup> and Vinardo<sup>[298]</sup>. These programs are used to “dock” ligands into the binding pocket of selected target protein in various conformations of the ligand and orientations within the binding pocket. These different conformations are evaluated with a scoring function to identify the best binding pose and approximate the binding affinity. Poses with the best scores are reported, in theory the best binding score is the ligand-protein complex with the best pose, greatest binding affinity and therefore the lowest docking score. However it has been shown that the best score is not always the best pose that corresponds with the crystallographic binding orientation in self-docking and cross-docking studies.<sup>[299]</sup>

Docking programs each use a different approach to generate binding poses and different criteria for the scoring of each pose. In this study we will be focussing on Glide software as it was used throughout this study for docking calculations. Glide is a reliable docking program, which is able to analyse each small molecule with an exhaustive conformational search algorithm in order to identify the rigid cores it then rotates each “rotamer group” to produce different ligand



conformations. High energy conformations are removed and the remaining conformations are docked into the rigid binding pocket.

Each pose is evaluated in a hierarchical scoring scheme. The receptor-ligand interactions are evaluated with a “greedy” scoring algorithm that scores the ligand atom’s position and estimates the best score that the atom could achieve by moving  $\pm 1$  Å in any direction to estimate the score that could be obtained after refinement. From this step the top ~5000 poses are refined by allowing the ligand to rigidly move  $\pm 1$  Å in any direction from which the top ~400 poses are selected for further minimisation on the electrostatic and van der Waals grids of the optimised potentials for liquid simulations (OPLS) force fields.<sup>[44]</sup> On the best three to six lowest-energy poses the Monte Carlo procedure is used to help orientate peripheral groups on the molecule. These final poses are then scored using GlideScore - which will be discussed in the next subsection.<sup>[300]</sup> The best scoring poses are ranked by a score that combines the GlideScore, internal strain energy and the energy-grid score.

DOCK is another docking program that utilises a more unorthodox approach to docking. Hypothetical spheres are placed in the binding pocket of a protein which represent locations which ligands normally occupy, a ligand of interest is then fragmented and reconstructed in the active site with the goal of matching as many spheres as possible, before being analysed and scored using an empirical scoring function. Case studies have shown that the programs Glide and GOLD are more accurate than DOCK.<sup>[294, 301]</sup>

### 3.2.2 Scoring methods

Poses generated for ligands within the binding pocket of selected target proteins are evaluated using scoring functions. These scoring functions can be loosely categorised as (i) MM-based or force-field, (ii) knowledge-based, (iii) consensus-based, (iv) empirical-based and (v) hybrid-based scoring functions.<sup>[302]</sup>

**Force-field based scoring functions** makes use of a force-field (discussed in **section 3.3** below) to determine the binding energy. Earlier versions of AutoDock<sup>[303]</sup> and DOCK made use of the AMBER force field<sup>[304, 305]</sup> based energy functions for their scoring.

**Knowledge-based scoring functions** sum pairwise statistical potentials between the ligand and the protein:

$$A = \sum_i^{ligand} \sum_j^{protein} \omega_{ij}(r) \quad (1)$$

The distance-dependant potential between atom pair  $ij$ ,  $\omega_{ij}$ , is derived from an inverse Boltzmann analysis:

$$\omega_{ij} = -k_B T \ln \left[ \frac{\rho_{ij}(r)}{\rho_{ij}^*} \right] \quad (2)$$

where the term  $\rho_{ij}(r)$  is the numeric density of the atom pair  $ij$  at a distance of  $r$ . The term  $\rho_{ij}^*$  refers to the numeric density of the same atom pair in the reference state where the interatomic interaction is assumed to be equal to zero. With this approach the occurrence frequency of a pairwise contact is assumed to be a measure of its energetic contribution to ligand binding. This means that if a specific pairwise contact occurs more frequently than in the reference state, i.e. a random distribution, it is indicative of a interaction between the given atom pair that is energetically favourable. The opposite is also true, less frequent interaction between the atom pair compared to the reference state is indicative of an unfavourable interaction. Thus the “knowledge base” is created by using the ligand-protein complex data contained within the Protein Data Bank (PDB) structure as its training set. From this set, distance-dependant potentials for each possible atom pair are derived from the occurrence frequency. Knowledge-based approaches are attractive to users due to their conceptual and computational simplicity. There are a few docking programs that utilise knowledge-based scoring functions such as DrugScore<sup>[306, 307]</sup>, KECSA<sup>[308]</sup> and ITScore<sup>[309, 310]</sup>.

**Consensus-based scoring functions** have multiple scoring functions that are used simultaneously for virtual screening. This improves scoring as other functions can compensate for another’s deficiencies.<sup>[311, 312]</sup>

**Empirical based scoring function** calculates the fitness of protein-ligand binding by summing several terms of which each represent a single component of the protein-ligand binding. The first empirical scoring function was developed by Böhm in 1994<sup>[313]</sup>. ChemScore<sup>[300]</sup> is an empirical based scoring function which is utilised in several docking programs, the function can be written as:

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{lipo} \sum f(r_{lr}) + \Delta G_{hbond} \sum g(\Delta r)h(\Delta \alpha) + \Delta G_{metal} \sum f(r_{lm}) + \Delta G_{rotb} H_{rotb} \quad (3)$$

The *first term* describes the standard change in Gibbs free energy. The *second term* describes the lipophilic interaction between the ligand-atom/receptor-atom pair as defined in ChemScore, with  $r_{lr}$  being the distance between the defined pair. The *third term* describes all hydrogen-bonding interactions between the ligand and the protein. The same weight is given to ionic and non-ionic hydrogen bonds.  $g, h$  and  $f$  are functions that are awarded a full score of 1.0 for angles and distances that lie within specified limits and a partial score between 0.0 – 1.0 for those that lie outside of these limits but inside a threshold value.  $\Delta r$  is the deviation of the hydrogen bond length from 1.85 Å and  $\Delta\alpha$  is the deviation of the hydrogen bond angle from its ideal value of 180°. [300]

The popular docking program, Glide, uses a modified ChemScore scoring function known as GlideScore. [44, 300] The GlideScore function can be written as:

$$\begin{aligned} \Delta G_{bind} = & \Delta G_{lipo} \sum f(r_{lr}) + \Delta G_{hbond-neut-neut} \sum g(\Delta r)h(\Delta\alpha) \\ & + \Delta G_{hbond-neut-charged} \sum g(\Delta r)h(\Delta\alpha) + \Delta G_{hbond-charged-charged} \sum g(\Delta r)h(\Delta\alpha) \\ & + \Delta G_{metal} \sum f(r_{lm}) + \Delta G_{rotb}H_{rotb} + \Delta G_{polar-phob}V_{polar-phob} + \Delta G_{coul}V_{coul} \\ & + \Delta G_{vdW}V_{vdW} + solvation \end{aligned} \quad (4)$$

Some of the changes observed in the GlideScore scoring function include:

- (i) Weighting of hydrogen bonding interactions based on whether the partners are charged or neutral.
- (ii) Addition of favourable energy for polar non-hydrogen bonding atoms which are located in hydrophobic environments.
- (iii) Terms for electrostatic, solvation and van der Waals interactions were also added.
- (iv) Weighting for all the terms were also optimised in an effort to provide the best enrichment and to improve the match between experimental and predicted affinities.

**Hybrid-based scoring functions** uses a combination of scoring functions classes in an attempt to improve the results. Popular hybrid scoring functions such as AutoDock Vina [314] and Vinardo [298] are based on X-Score [315]. X-Score uses a combination of knowledge and empirical-based scoring functions, which are calibrated by using the structures of hundreds of protein-ligand complexes. Knowledge-based hybrids in which entropy and solvation terms are incorporated are also worth noting. [316, 317]

### 3.3 Molecular Mechanics Force Fields

To date numerous force fields have been developed, each developed through different means and designed for a particular system. The most common force field is the Universal force field (UFF)<sup>[318, 319]</sup> which is able to accommodate nearly all atom types, but at a cost of reduced accuracy compared to alternative specialised force fields such as the Merck molecular force fields (MMFF)<sup>[320, 321]</sup> which is capable of simulating organic molecules more accurately.

The OPLS force field used within the Schrödinger software suite was initially developed by Jorgensen and co-workers<sup>[322, 323]</sup> and was further improved by a research group within Schrödinger to produce OPLS\_2005<sup>[324]</sup> and later OPLS2.1, OPLS3<sup>[325]</sup> and OPLS3e<sup>[326]</sup>. Today OPLS4<sup>[322]</sup> is used within the suite and is designed to be suitable for most organic molecules relevant to drug discovery or any other similar applications. The most notable improvements of OPLS4 is the improvement of model accuracy in hydration representation, improved treatment of molecular ions and improved sulphur interaction accuracy.

MM programs sum individual energy functions to calculate the energy of the system according to the equation below:<sup>[325]</sup>

$$E = E_{bond} + E_{angle} + E_{dihedrals} + E_{non-bond} \quad (5)$$

Below follows a discussion of how the OPLS3/3e/4 force fields calculate the individual energy contributions.

#### 3.3.1 Bonded interactions

The terms  $E_{bond}$ ,  $E_{angle}$  and  $E_{dihedrals}$  describes the energy force fields of atoms which are covalently bonded to one another.  $E_{bond}$  describes covalent bonds as springs deviating from Hooke's law, where the further apart an atom is from the equilibrium,  $r_{eq}$ , the higher the energy. The coefficient  $K_r$  increases with the strength of the bond.

$$E_{bond} = \sum_{bonds} K_r (r - r_{eq})^2 \quad (6)$$

The potential energy of an angle, is described using the same approach.

$$E_{angle} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \quad (7)$$

where,  $\theta_{eq}$  represents the angle at equilibrium and  $K_\theta$  a coefficient of how easily it can deviate.

The potential energy of the dihedral angles are calculated with the sum of a series of cos function.

$$E_{dihedrals} = \sum_{dihedrals} \left[ \frac{V_1}{2}(1 + \cos \varphi) + \frac{V_2}{2}(1 - \cos 2\varphi) + \frac{V_3}{2}(1 + \cos 3\varphi) + \frac{V_4}{2}(1 - \cos 4\varphi) \right] \quad (8)$$

where,  $\varphi_i$  is the dihedral angle;  $V_1$ ,  $V_2$ ,  $V_3$  and  $V_4$  are the coefficients in the truncated Fourier series.

### 3.3.2 Non-bonded interactions

Non-bonded interactions between atoms in close proximity can be separated into electrostatic ( $E_{elec}$ ) and van der Waals ( $E_{vdW}$ ) interactions.

$$E_{non-bond} = E_{elec} + E_{vdW} \quad (9)$$

Ligands that bond reversibly only have non-bonded interactions with the binding pocket. The electrostatic interactions are simply based on the distance between charged atoms, while the van der Waals interactions are modelled with a Lennard-Jones potential. **Figure 3.1** below illustrates the Lennard-Jones potential. From this figure we observe that when the distance  $\sigma_{ij}$  is small, the first term is dominant and the repulsion between atoms  $i$  and  $j$  increases with a decrease in distance. As  $\sigma_{ij}$  approaches  $r_{ij}$ , the second term becomes more significant, meaning that the repulsion decreases and attraction between the atoms  $i$  and  $j$  increases until it reaches the bottom of the well. The depth of the well is determined by  $\epsilon_{ij}$ . As  $\sigma_{ij}$  increases further, the attractive forces experienced between the two atoms becomes negligible.

$$E_{non-bond} = \sum_{i>j} f_{ij} \left[ \frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right] \quad (10)$$

As the number of atoms in a system increases the number of non-bonded interactions increase exponentially, this results in the bulk of the computational load for an MM based system. This load is reduced by making an assumption that non-bonded interactions which are sufficiently far away from each other (typically 14 Å) are assumed to be equal to zero.

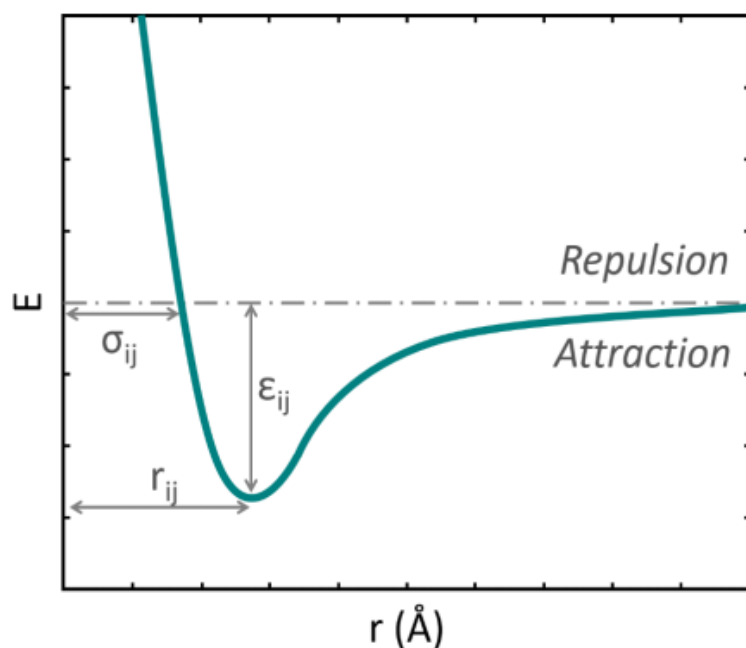


Figure 3.1 - Illustration of the Lennard-Jones potential plot.

### 3.3.3 Assumptions in Molecular Mechanics

The following assumptions made by the OPLS force fields, as discussed in the original OLPS3 publication<sup>[325]</sup> and directly quoted here, are worth noting:

- (i) *“The classical treatment of nuclear motion does not systematically distort results as compared to experimental.”*
- (ii) *“The use of a fixed charge force field provides an adequate description of electrostatic interactions in the condensed phase.”*
- (iii) *“Torsional terms for to the gas phase can be transferred to the condensed phase with minimal errors.”*
- (iv) *“The contribution of fundamentally nonclassical quantum mechanical nonbonded interactions, such as oxygen-sulphur interactions,  $\pi$ - $\pi$  interactions and non-linear van der Waals coupling, is negligible.”*

### 3.4 Docking methods

In all docking approaches the ligand is seen as flexible and can occur in many different conformations due to rotatable bonds and chirality within a ligand, docking approaches are broadly categorised according to how the protein is viewed. There are two common docking approaches, in the first approach the protein is seen as flexible and residues in the binding pocket can mould around the ligand for a better fit (thus greater affinity between the ligand and the protein), termed an “induced-fit model”. In the second approach the receptor is seen as rigid and residues in the binding pocket are fixed. The latter is by far the most popular approach as

it requires the least computational power and therefore time; this model is commonly referred to as a “lock-and-key” model.

This lock-and-key model assumes that due to the selectivity of protein targets that the binding pocket only takes on a single or a certain number of specific binding pocket poses, in the cases of protein targets that get activated by only a single ligand for the former or more than one ligand for the latter. However, the downside to this approach is that different docked ligands require shifts in binding pocket residues in order to form a stable ligand-protein complex through the formation of interactions that would not be possible with a single rigid binding pocket pose. This may lead to suboptimal dockings scores, possibly due to clashes, indicative of an unfavourable interaction between a ligand and a protein, producing false negative results. To overcome this one simple practice is to reduce the van der Waals size for the atoms of the binding pocket and/or the ligands. By following this approach the docking score becomes more forgiving to small clashes which would not have been present if the binding pocket was considered to be flexible as well. While this approach is effective for small conformational shifts that may be present in the pocket, it does not account for complete rotations of sidechains. In such cases it would be better to use an induced-fit docking protocol to accommodate ligands that bind to different binding pocket conformations.<sup>[327]</sup>

Induced-fit protocols unfortunately increases the possibility of false positive hits, as these protocols are actively designed to accommodate the largest range of ligands as possible without taking the strain residues experience in order to adopt a particular conformation into consideration. Another downside of induced-fit docking is the increased computational resources required which increases exponentially with an increase in the number of binding pocket residues that are considered to be flexible.

The last approach called ensemble docking has the advantage of induced-fit docking without the need for a significant increase in computational resources. In this approach a ligand is docked into an ensemble of rigid receptor structures, therefore several conformations of the binding pocket is obtained and taken into considerations while the computational resources increase linearly with each receptor structure that is added into the ensemble.

For protein targets (receptors) that are not available an alternative approach known as homology modelling is utilised but this requires a protein structure that has been elucidated (usually of another specie) and a sequence of the related protein. The sequence has to have at

least 50% similarity to be considered reliable enough with only minor errors in the conformations of the side chains<sup>[328]</sup>; however minor changes can have a significant influence on the obtained docking results especially if the errors occur in a conserved region of the protein targets.

### 3.4.1 *Ligand selection and preparation*

As mention in all docking approaches the ligand is considered to be flexible, different conformations for a ligand is generated as part of the ligand preparation steps. These conformations can be stereoisomers and tautomers of the original input ligand, during the refinement process high energy conformations are removed as these are unlikely to occur.

Ligand preparation can be done using commercially available LigPrep<sup>[329]</sup> and open-source programs such as OpenBabel<sup>[330]</sup>, Balloon<sup>[331, 332]</sup> and RDKit<sup>[333]</sup>. LigPrep for example uses a set of programs bound together into a coherent package for its protocol. Some compounds from libraries can be imported into chosen ligand preparation programs using simplified molecular input line entry system (SMILES) format or a 2D structure that can be drawn within some of the programs for example Schrödinger's 2D sketcher that automatically generates a 3D structure from the input. Once the 3D structure is obtained the ligand preparation can commence which generates different tautomers and stereoisomers, a specific pH range can also be specified to produce ligands with different protonation states.

### 3.4.2 *Receptor selection and preparation*

With the exponential increase in protein crystal structures complexed to different ligands, it becomes an attractive approach to use molecular docking studies as an initial step for drug design. More and more protein targets such as receptors, enzymes and transporters become available which can be used for other research purposes such as ligand screening and method of action determination for known pharmaceutical and illicit drugs.

Protein crystal structures can be evaluated to determine the binding pose as different ligands docked to the same protein can produce different binding pocket poses. These different poses can be useful when trying to dock a single ligand to obtain a good docking score. Furthermore, water molecules within crystal protein structure needs to be taken into consideration. In some cases water molecules can occur in the binding pocket which form strong hydrogen bonds with protein residues, while other waters can be in hydrophobic pockets<sup>[334]</sup>; in other cases water molecules within the binding pocket can mediate an interaction, between the ligand and an



amino acid within the binding pocket known as water bridges. During docking these water molecules are kept in the binding pocket to allow for these type of interactions to be accounted for. However, water molecules in the binding pocket should be visually inspected to determine if they are of any importance during ligand-protein docking, all non-essential solvent molecules should be removed in order to avoid steric clashes from occurring that would prevent larger ligands from being accommodated in the binding pocket. Currently molecular docking approaches are being developed to take these effects into account.<sup>[335, 336]</sup>

Protein preparation is an essential step in docking as successful structure-based modelling projects not only demands accurate software, but also accurate starting materials. Protein preparation software are designed to ensure structural correctness and equipping users with high-confidence structures to be used in a wide variety of modelling applications. Common problems observed with crystallographic structures include missing hydrogen atoms, ambiguous protonation states, incomplete loops and side chains and flipped residues all of which can take considerable time and effort to correct. Programs such as the Protein Preparation Wizard from Schrödinger automates, aggregates and integrates the most frequently used tools and techniques to correct a protein structure.<sup>[337, 338]</sup>

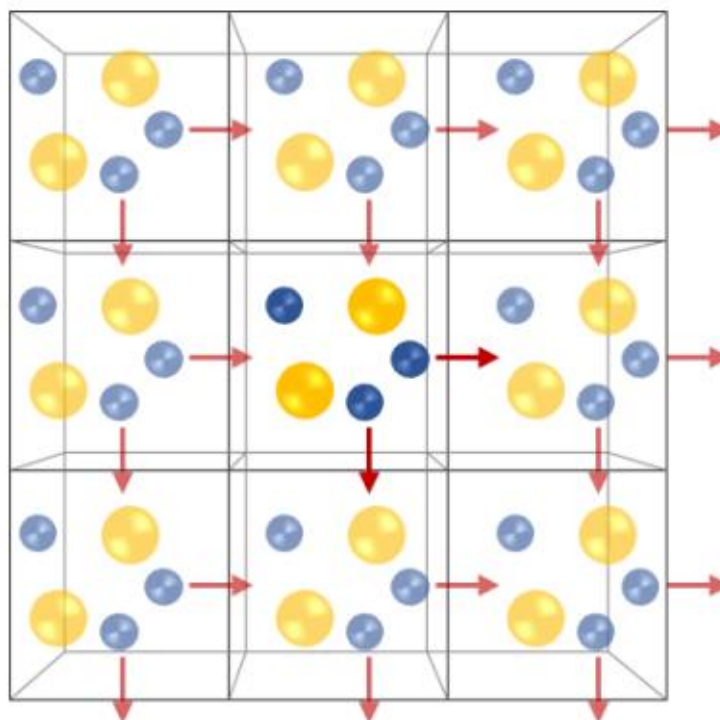
### 3.5 *Molecular Dynamics*

MD was first developed in the 1970's and attempts to model the movement of atoms over a period of time.<sup>[339, 340]</sup> Several factors need to be considered such as which force field to use, type of ensemble, the time length of the simulation, time steps and solvation. Instead of using atomistic representations, coarse-grained simulations can also be used; however this is used for long simulations or extremely large systems.<sup>[341]</sup>

Initial configuration of the system usually utilises structures obtained from theoretical models (docking or homology modelling) or experimental data (Cryo-EM, X-ray crystallography or Nuclear magnetic resonance). The solvation method and boundary condition need to be defined, interactions at the boundary can influence the energy calculation and therefore it is required to be defined or taken into consideration. A periodic boundary condition is most commonly used for biological systems; which involves placing the system in a cell and then surrounding the cell with mirror cells containing replicas as illustrated in **Figure 3.2**. Across cell boundaries interaction energies can be determined resulting in a larger system being simulated and overcoming boundary effects with only a reasonable increase in computational requirements. An important requirement is to ensure that the size of the cell is large enough to ensure that the biological system, i.e. protein, does not interact with itself. Therefore the size

of the box generated around the protein or ligand of interest has cell walls that are typically 10 Å away. Solvation of the box occurs with an *in vacuo* approach, implicit- or explicit solvent models which will be discussed in **section 3.6** below. All the atoms in the cell are typed for the partial atomic charges and force field allocated, before being submitted for a series of minimisation steps.

Once the complete MD system has been set up, it must then be minimised in order to reduce any significant forces experienced by any atom within the system which may lead to the final simulation failing. Minimisation typically occurs in a multi-step fashion in which the solvent is minimised and the solute is constrained. A dynamic run of the solvent is then performed for a period of time, usually less than 100 ps. The system then gets minimised while the constraints on the solute are reduced.



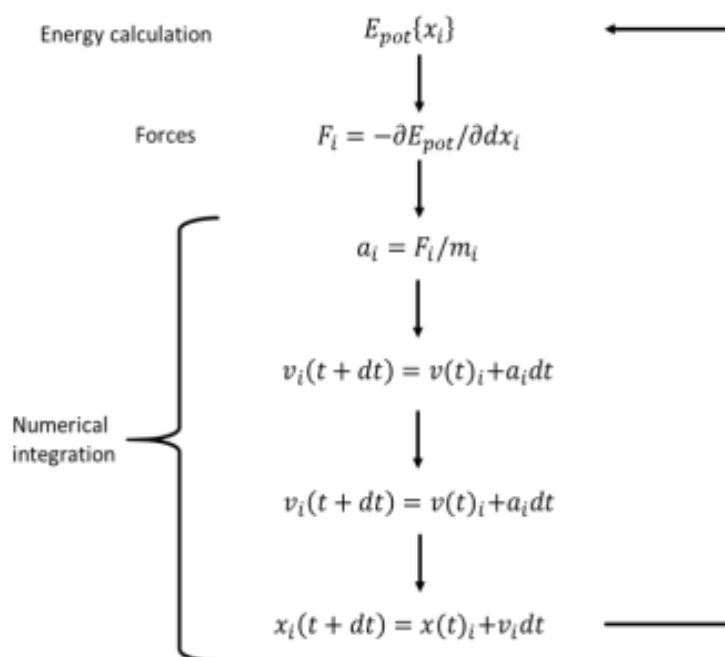
*Figure 3.2 - Periodic boundary condition illustration.*

The ensemble class then needs to be taken into consideration, with MD simulations frequently using NVE, which holds the number of particles (N), volume (V) and the energy (E) constant. Monte Carlo based methods utilise canonical ensemble, NVT, in which N, V and temperature (T) are kept constant. It is more practical for biological systems to use the isothermal-isobaric ensemble, NPT, which keeps the N, T and pressure (P) constant.

Before the full MD simulation can run the length of the simulation first needs to be decided. The length of the simulation is affected by several factors that should be considered before deciding on a simulation length these factors include:

- (i) The process being studied (i.e. a rare conformational change being studied).
- (ii) The nature of the system (a large system will require a lot of computational resources due to the number of atoms present).
- (iii) Computational resources available. Only data obtained from the production stage is typically used for further energy calculations and analysis.

In MD simulations (described in **Figure 3.3** below) the force field is used to obtain the forces acting on every atom in the system as well as the potential energy. Using the classical Newtonian laws of motion and the forces on each atom the velocities and acceleration and thus the position of the atoms are updated. The integration of movement is done numerically and a timestep which is shorter than the fastest atomic movement is typically used. For example in a 1 ns simulation where a 1 fs timestep is chosen the calculations are iterated a million times. In course-grained simulations a larger timestep can be used which allows for a dramatic increase in the simulation time length.



**Figure 3.3** - Basic MD simulation algorithm. With:  $E_{pot}$  - potential energy;  $t$  - simulation time;  $dt$  - simulation timestep. For  $N$  simulated atoms ( $i$ ):  $a$  - acceleration;  $v$  - velocity;  $m$  - mass of the atom;  $x_i$  - atom coordinates;  $F_i$  - force component.

Rare events such as ligand binding, protein folding and loop movement require much longer simulations, in the order of micro- to milliseconds; making these simulations extremely

computationally expensive. Around the 2000's publications reported MD simulations with a length of about a nanosecond<sup>[342-344]</sup>, while more recently much longer simulations are reported directly attributed to the advancement in computational resources.<sup>[345-349]</sup> These advancements include the increase in size and efficiencies of computer clusters, but the development of graphics processing units (GPU's) have made the largest impact. Modern MD software packages now include the use of general-purpose computing on graphics processing units (GPGPU) which allows for many arithmetic units to work in parallel within GPU's to significantly improve computational time.<sup>[350-353]</sup>

The movement of atoms are studied using the Root Mean Square Deviation (RMSD), which measures the change in displacement for a selection of atoms in a given frame compared to a reference frame containing the same atoms. The RMSD for frame  $x$ :

$$\text{RMSD}_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i'(t_x) - r_i(t_{\text{ref}}))^2} \quad (11)$$

where  $N$  represents the number of atoms in the atoms selection;  $r_i'$  is the position of the selected atoms in frame  $x$  after superimposing on the reference frame, where frame  $x$  is recorded at time  $t_x$  and  $t_{\text{ref}}$  is the reference time (typically the first frame). The procedure is repeated for every frame in the simulation trajectory.

### 3.6 Solvent models

In order to study molecules as accurately as possible, the environment that surrounds them needs to be taken into consideration. Cases in which the environment (i.e. solvent) is not considered are said to be modelled *in vacuo*. With *in vacuo* cases the dielectric constant is set to one, but this results in unrealistically large electrostatic energies.<sup>[354]</sup>

There are two main solvent models namely implicit and explicit solvent models. Implicit solvent models provides a more accurate representation of a molecule in a solvent, but requires greater computational resources. This approach is most useful when large areas of chemical space needs to be considered (i.e. conformational searching for ligands and proteins) to estimate the electrostatic energies. Implicit solvent models utilises special energy terms that represent the solvent as a continuous medium. The first commonly used method is known as the Poisson-Boltzmann equation and the second uses a linear approximation of the Poisson-Boltzmann equation known as the Generalised Born equation which requires less computational resources to compute. The Generalised Born equation in combination with a hydrophobic solvent accessible surface area term forms the Molecular Mechanics Generalised-

Born Surface Area (MM-GBSA) continuum solvent model. If Poisson-Boltzmann equation is used instead of Generalised Born equation the solvation model is known as the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) solvation model. Some shortfalls of the implicit solvent models includes entropic effects which are not taken into account which can be a major factor in ligand binding, protein folding and loop movement. Furthermore, hydrogen bonding is taken into account, but the directionality thereof is not.

Lastly explicit solvent models provide the most accurate results, but requires significantly more computational resources. With this approach solvent molecules surround the system of interest, a dielectric constant of one is used (similar to modelling *in vacuo*) and electrostatic energies are applied to all individual atoms. Today several explicit water models are in existence of which TIP4P is the most commonly used model. TIP4P is known as a “4-site” model which models water as a rigid molecule with four interaction sites.<sup>[355, 356]</sup> 3-, 5- and 6-site models are in existence as well, but due to the increased computational resource requirements associated with 5- and 6-site water models they are typically used for simulating water dynamics.<sup>[357-359]</sup> Work done by Schrodinger in 2010 showed that the computation of aqueous hydration free energies was relatively insensitive to the water model used.<sup>[360]</sup>

### 3.7 Conclusion

Computer aided molecular docking studies have become so advanced that it has become a vital approach in studying the interface between molecular biology and chemistry. It is of vital importance that users of these methods are aware and understand the limitations and assumptions. These computational modelling methods almost always generate a result and large amount of data can easily be generated in relatively short periods of time, but one continuously need to consider how accurate the generated results are.

The docking of MMMP to various monoamine receptors and a transporter in this study will utilises MM based methods, as QM based methods such as DFT require significantly more computational resources. The downside to this is that MM based methods unfortunately has approximations which do not appropriately model QM based interactions such as cation- $\pi$  and  $\pi$ - $\pi$  interactions efficiently.

The theory discussed in this chapter provides a mean of understanding the methods applied and the results obtained within this study. However, the computational results obtained in this study are only theoretical and these theoretical predictions can only be confirmed once biological studies have been performed which is beyond the scope of this study.

## Chapter 4: Molecular modelling of 2-methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP)

### 4.1 Introduction

In Chapter 3 the theoretical basis of the Molecular Mechanics (MM) approach to study large systems such as ligand-protein complexes were discussed. This chapter will cover the methodology, results and discussion of the computational study to analyse the interaction of a known ligand with identified biological protein targets.

One of the most critical components of drug discovery is to determine if a ligand is capable of interacting with a specific target or targets; without an interaction a target cannot elicit a response. As extensively discussed in Chapter 2, a response with regards to monoamine receptors can be activating (agonist) or inhibitory (antagonist) in nature and is dependent on the structure of the ligand, while in the case of monoamine transporters a ligand can either act as a substrate or a blocker. If a ligand is capable of binding to a target, it takes on a specific binding pose. The most conclusive approach to determine if a ligand is capable of interacting with a specific protein is through the use of *in vitro* bioassay methods such as ligand-binding, activity, inhibition and release assays; the most conclusive way to determine a specific binding pose of a ligand is through co-crystallisation of the active compound with the target protein. Both *in vitro* and crystallographic approaches are very time-consuming, expensive and requires a great deal of technical expertise and specialised equipment, hence the move to *in silico* based methods.

Although *in silico* based methods such as docking are becoming a very attractive approach, caution needs to be taken with the generated results as these programs use an exhaustive approach to generate a result. These results are quantified with a docking score, i.e. GlideScore, by using a docking scoring function, discussed in Chapter 3.<sup>[44, 45, 294, 324]</sup> A review by Y. Chen titled "*Beware of docking!*" highlights many issues to consider in docking studies, even though they are becoming increasingly popular in routine drug development endeavours. The review stresses that despite a pose scoring the highest docking score it can be the incorrect pose, meaning that it doesn't accurately represent the actual pose of an active ligand in the binding pocket which can lead to failed Molecular Dynamics (MD) simulations. Furthermore, docking is incapable of determining if a ligand is an agonist, antagonist, substrate or blocker as it is only capable of determining binding affinities, bioassays are required to confirm the type of activity.<sup>[361]</sup> Some of these sentiments are echoed in a paper published by D. Ramírez and J. Caballero.<sup>[299]</sup>

To help address the inaccuracies associated with docking, additional computational approaches such as MD simulations can be employed, which can support results to add to its credibility. These simulations can show if a ligand is stable in the binding pocket with only minor movements over a specified period. The frequency of particular interactions between a protein and a ligand can also be analysed and compared to other known active ligands. With MD simulations an adequate simulation length needs to be considered in order to determine the stability of a ligand within the binding pocket and any significant conformational changes that occur in the protein can also be observed, noted and rationalised.

## 4.2 *Molecular modelling methodology*

### 4.2.1 *Protein target selection*

Protein targets which consists of human monoamine receptors and transporters were selected from the Protein Data Bank (PDB) and listed in **Table 4.1**. For some of the targets multiple crystal structures were available. Targets were selected based on a combination of the following criteria:

- (i) Best resolution.
- (ii) Active state conformations were given preference over apo or inactive states, to ensure that binding pocket amino acid residues are as close to active state conformations as possible.
- (iii) Least modified protein structures. Modifications of some protein structures are unavoidable and necessary to stabilise the protein during the crystallisation and X-ray crystallography/NMR analysis.
- (iv) After redocking, discussed below, Root Mean Square Deviations (RMSD) were calculated from the crystallised pose and the Glide predicted pose. Where RMSD values were greater than three an alternative PDB structures of the same receptor was used to try and obtain an improved RMSD.

### 4.2.2 *Validation through redocking*

Redocking is a common method used to evaluate the accuracy of a chosen docking method, in this study the Glide<sup>[45]</sup> docking protocol with the OPLS4 force field within the Schrödinger release 2023-3 suite was utilised.

For each ligand-protein complex the process involved preparing the ligand-protein complex with Schrödinger's Protein Preparation Wizard.<sup>[338]</sup> The crystal ligand was prepared using LigPrep<sup>[329]</sup> at a pH of  $7.4 \pm 2.0$  from which different protonated states, tautomers and stereoisomers of the ligand were generated. The protein together with the crystal ligand, used

to indicate the active site, was used to generate receptor grids. Glide Extra Precision (XP)<sup>[45]</sup> docking was used to dock the generated LigPrep ligands back into the active site. The ligands were treated as flexible and only the best pose for each ligand was generated. The ligands were effectively redocked back into the protein to predict a binding pose for the ligand. Each pose was then scored by the scoring function, GlideScore<sup>[44, 300]</sup>, to obtain a docking score.

Out of the generated poses the pose with the greatest docking score together with the crystal ligand pose was used to calculate an RMSD value. The RMSD (Chapter 3, **Eq. 11**) calculation is used to measure the change in displacement of the atoms in the redocked ligand pose compared to the reference crystal ligand pose which contain the same atoms. The RMSD command line was run from the Centre of High Performance Computing (CHPC) terminal. The smaller the RMSD value the greater the agreement between the crystal ligand pose and the predicted redocked ligand pose, which is used to evaluate how well the docking protocol performs.

#### 4.2.3 Docking Protocol

The 2-methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP) ligand was prepared using LigPrep<sup>[329]</sup> within the Schrödinger release 2023-3 suite (visualised by Maestro graphical user interface<sup>[43]</sup>) with the OPLS4 force field and the protonation states were predicted with Epik<sup>[362-364]</sup> in a pH range of  $7.4 \pm 2.0$ . Protein preparation was done with Protein Preparation Wizard<sup>[338]</sup> in a pH range of  $7.4 \pm 2.0$  for most part the default settings were utilised, missing loops and side chains were filled in using Prime.<sup>[365-367]</sup> Water further than 5 Å from any part of the ligand were deleted and an energy minimisation step was performed to relieve unfavourable constraints. Default settings were utilised for the receptor grid generation and Glide docking as well. Receptor grids were generated using the centroid of the selected ligand to automatically generate a 10 x 10 x 10 Å box around the ligands in the binding pocket of each protein, visual inspection was done to ensure that the ligands fit into the virtual box. Molecular docking was performed using Glide XP scoring function<sup>[44, 45, 294]</sup> and ligand structures were treated as flexible, which allowed for sampling of ring conformations and nitrogen inversion, to obtain 5 poses for each docked MMMP molecule. The complexes with the smallest difference between the redocked and the MMMP docking score for a protein were selected and submitted for MD simulation.



#### 4.2.4 Molecular Dynamics

Selected complexes were submitted for MD simulations using Desmond<sup>[46, 47]</sup> from the Schrödinger Release 2023-3 suite with the OPLS4 forcefield. Systems were built for the complexes using an orthorhombic box shape whose edges were 10 Å away from the protein structure and the box was solvated using the TIP4P water model.<sup>[356]</sup> The systems were neutralised with sodium and chloride ions and had ion concentrations of 0.15 M NaCl, to mimic biological conditions. These minimised systems were submitted for 200 ns production runs with the default relaxation protocol (discussed below). The simulations were performed under NPT (constant number of particles, pressure and temperature) conditions using the Nose-Hoover thermostat of 310 K (1.0 ps relaxation time), Martyna-Tobias-Klein barostat of 1.01325 bar (2.0 ps relaxation time) and particle-mesh Ewald electrostatics with a cut-off of 9 Å. Frames were written every 100 ps and time-step calculations were performed every 2 fs. Simulation interaction diagram module within the Schrödinger suite was used to analyse the simulations. The calculation of the RMSD to describe atom movement during a simulation is discussed in Chapter 3.

The default Desmond relaxation protocol<sup>[47]</sup> was utilised prior to the MD simulation production run to equilibrate the system, the steps for this protocol are as follow:

*Stage 1:* Task - recognizing the simulation setup parameters

*Stage 2:* Simulate - Brownian Dynamics NVT, T = 10 K, small timesteps and restraints on solute heavy atoms, 100 ps.

*Stage 3:* Simulate - NVT, T = 10 K, small timesteps and restraints on solute heavy atoms, 12 ps.

*Stage 4:* Simulate - NPT, T = 10 K, and restraints on solute heavy atoms, 12 ps.

*Stage 5:* Simulate - NPT and restraints on solute heavy atoms, 12 ps.

*Stage 6:* Simulate - NPT and no restraints, 24 ps.

*Stage 7:* Simulate - Production run.

## Results and Discussion

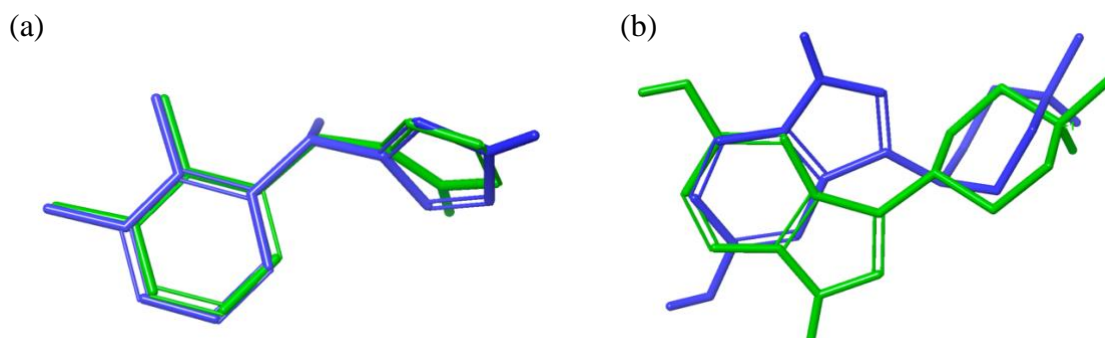
### 4.3 Validation through redocking results and discussion

As described above for each protein-ligand complex the crystal ligand was redocked back into the protein structure to obtain binding poses which were then scored by the scoring function GlideScore<sup>[44, 300]</sup> to obtain docking scores. Only the best docking score pose was compared to the original pose of the crystal ligand to calculate the RMSD.

The results from the RMSD calculations are tabulated in **Table 4.1** along with the PDB accession code for the crystal structure used for each protein target. From the results two of the redockings, Serotonin receptor 1E (SER1E) and Norepinephrine receptor  $\alpha$ -2A (NER2A), yielded a RMSD greater than 2 Å while 23 scored below 2 Å. Out of these 23, 15 scored a RMSD below 1 Å. The best RMSD that was obtained was for norepinephrine  $\alpha$ -2B (NER2B, PDB: 6K41) with a score of 0.36 Å. For the two structures with an RMSD greater than 2 Å, no alternative structures were available at the time of this study for SER1E (RMSD of 2.37 Å) and the only alternative structure (PDB: 7EJ8) for NER2A yielded a 6.03 Å RMSD. In general an RMSD  $\leq$  2 Å is accepted as a good pose but 1 Å is preferred for smaller ligands.<sup>[368, 369]</sup>

A good RMSD indicates a good agreement between the crystal ligand pose and the predicted ligand pose generated from the docking procedure, indicative of an adequate docking protocol. **Figure 4.1 (a) - (b)** below illustrates the difference in overlap between the smallest and largest RMSD obtained from two different crystal structure. It becomes evident that already at 2.37 Å a large difference in a pose exists which can cause unfavourable interactions with the amino acids in the binding pocket, which in effect can alter the activity of a ligand in the binding pocket. It is vital for a docking program to predict the pose as close to the crystal ligand as possible as any other poses might be unfavourable.

The redocking scores range from -5.751 (Serotonin receptor 1A, SER1A) to -14.259 kcal/mol (Serotonin receptor 2B, SER2B), both the highest and lowest docking score is obtained in the serotonin receptor class. From the RMSD and redocking scores it can be concluded that the docking protocol was capable of predicting the correct binding pose, for 23 out of the 25 docking targets, and that an acceptable docking protocol has been established that is capable of predicting the correct binding pose.



**Figure 4.1** - Illustrations of (a) smallest [Norepinephrine 2B, 0.36 Å] and (b) largest RMSD [Serotonin 1E, 2.37 Å] of two crystal ligands. The green ligand represents the crystal ligand and the blue the redocked ligand.

Table 4.1 - RMSD, Redocking and MMMP docking scores.

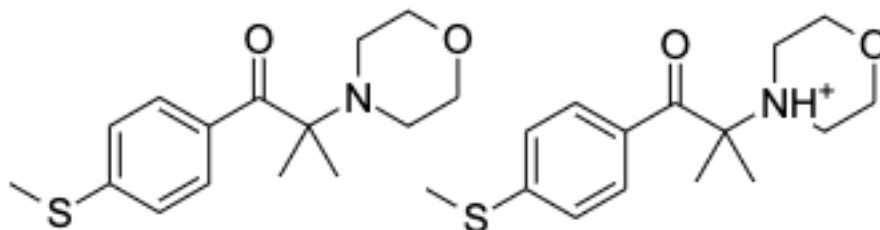
| Receptors/<br>Transporter             | Protein<br>PDB ID     | RMSD<br>(Å) | Glide Docking scores (kcal/mol) |               |               |                                     |
|---------------------------------------|-----------------------|-------------|---------------------------------|---------------|---------------|-------------------------------------|
|                                       |                       |             | Redocking<br>Score              | Deprotonated  | Protonated    | Absolute<br>Difference <sup>a</sup> |
| <b>Dopamine</b>                       |                       |             |                                 |               |               |                                     |
| <b>D1</b>                             | 7JOZ <sup>[370]</sup> | 0.94        | -6.032                          | <b>-4.625</b> | -3.470        | 1.407                               |
| <b>D2</b>                             | 6CM4 <sup>[371]</sup> | 0.39        | -12.009                         | -4.667        | <b>-5.604</b> | 6.405                               |
| <b>D3</b>                             | 7CMV <sup>[372]</sup> | 0.52        | -7.648                          | <b>-4.732</b> | -3.253        | 2.916                               |
| <b>D4</b>                             | 5WIU <sup>[373]</sup> | 1.43        | -11.048                         | -4.260        | <b>-4.727</b> | 6.321                               |
| <b>D5</b>                             | 8IRV <sup>[374]</sup> | 0.55        | -11.456                         | <b>-4.689</b> | -2.722        | 6.767                               |
| <b>Serotonin</b>                      |                       |             |                                 |               |               |                                     |
| <b>SER1A</b>                          | 7E2Y <sup>[375]</sup> | 1.14        | -5.751                          | <b>-4.976</b> | -0.405        | 0.775                               |
| <b>SER1B</b>                          | 4IAR <sup>[376]</sup> | 0.57        | -11.395                         | <b>-5.247</b> | -4.418        | 6.148                               |
| <b>SER1D</b>                          | 7E32 <sup>[375]</sup> | 0.81        | -7.309                          | <b>-5.573</b> | -3.772        | 1.736                               |
| <b>SER1E</b>                          | 7E33 <sup>[375]</sup> | 2.37        | -6.024                          | <b>-4.831</b> | -3.278        | 1.193                               |
| <b>SER1F</b>                          | 7EXD <sup>[377]</sup> | 1.32        | -8.735                          | <b>-4.247</b> | -2.354        | 4.488                               |
| <b>SER2A</b>                          | 6A94 <sup>[378]</sup> | 0.94        | -10.362                         | <b>-6.766</b> | -4.907        | 3.596                               |
| <b>SER2B</b>                          | 4IB4 <sup>[379]</sup> | 0.62        | -14.259                         | <b>-6.120</b> | -3.431        | 8.139                               |
| <b>SER2C</b>                          | 6BQH <sup>[380]</sup> | 0.41        | -13.837                         | <b>-6.704</b> | -5.406        | 7.133                               |
| <b>SER4</b>                           | 7XTA <sup>[381]</sup> | 1.20        | -7.073                          | <b>-4.845</b> | -2.655        | 2.228                               |
| <b>SER5A</b>                          | 7UM4 <sup>[382]</sup> | 0.57        | -10.357                         | <b>-4.969</b> | -3.653        | 5.388                               |
| <b>SER6</b>                           | 7XTB <sup>[381]</sup> | 0.41        | -7.839                          | <b>-5.745</b> | -1.225        | 2.094                               |
| <b>SER7</b>                           | 7XTC <sup>[381]</sup> | 0.60        | -8.916                          | <b>-4.030</b> | -3.541        | 4.886                               |
| <b>Norepinephrine</b>                 |                       |             |                                 |               |               |                                     |
| <b>NER1A</b>                          | 8THK <sup>[383]</sup> | 1.01        | -10.826                         | <b>-7.335</b> | -3.103        | 3.491                               |
| <b>NER2A</b>                          | 7EJ0 <sup>[384]</sup> | 2.34        | -9.065                          | <b>-4.450</b> | -3.950        | 4.615                               |
| <b>NER2B</b>                          | 6K41 <sup>[385]</sup> | 0.36        | -9.624                          | <b>-6.452</b> | 0.500         | 3.172                               |
| <b>NER2C</b>                          | 6KUW <sup>b</sup>     | 1.75        | -8.970                          | <b>-6.395</b> | -5.204        | 2.575                               |
| <b>NERB1</b>                          | 7BU7 <sup>[386]</sup> | 0.93        | -13.038                         | <b>-5.586</b> | -4.768        | 7.452                               |
| <b>NERB2</b>                          | 3SN6 <sup>[387]</sup> | 1.42        | -10.608                         | -5.209        | <b>-5.618</b> | 4.990                               |
| <b>NERB3</b>                          | 7XJH <sup>[388]</sup> | 0.71        | -10.846                         | <b>-4.443</b> | -2.188        | 6.403                               |
| <b>Serotonin<br/>Transporter</b>      |                       |             |                                 |               |               |                                     |
| <b>Orthosteric<br/>binding pocket</b> | 7LIA <sup>[227]</sup> | 1.54        | -6.577                          | -5.316        | <b>-6.058</b> | 0.519                               |

<sup>a</sup>Calculated using Eq. 12.

<sup>b</sup>Paper not yet published. No alternative structures were available at the time of this study.

#### 4.4 Molecular docking results and discussion

MMMP has no chiral centres and therefore no stereoisomers were generated, however two different protonated states for MMMP within the specified pH range  $7.4 \pm 2.0$  were generated. Protonation occurred at the nitrogen of the morpholine group as illustrated in **Figure 4.2**. These ligands were docked into the different binding pockets and the results are reported in **Table 4.1**. A docking score for both the protonated and deprotonated MMMP ligands are reported.

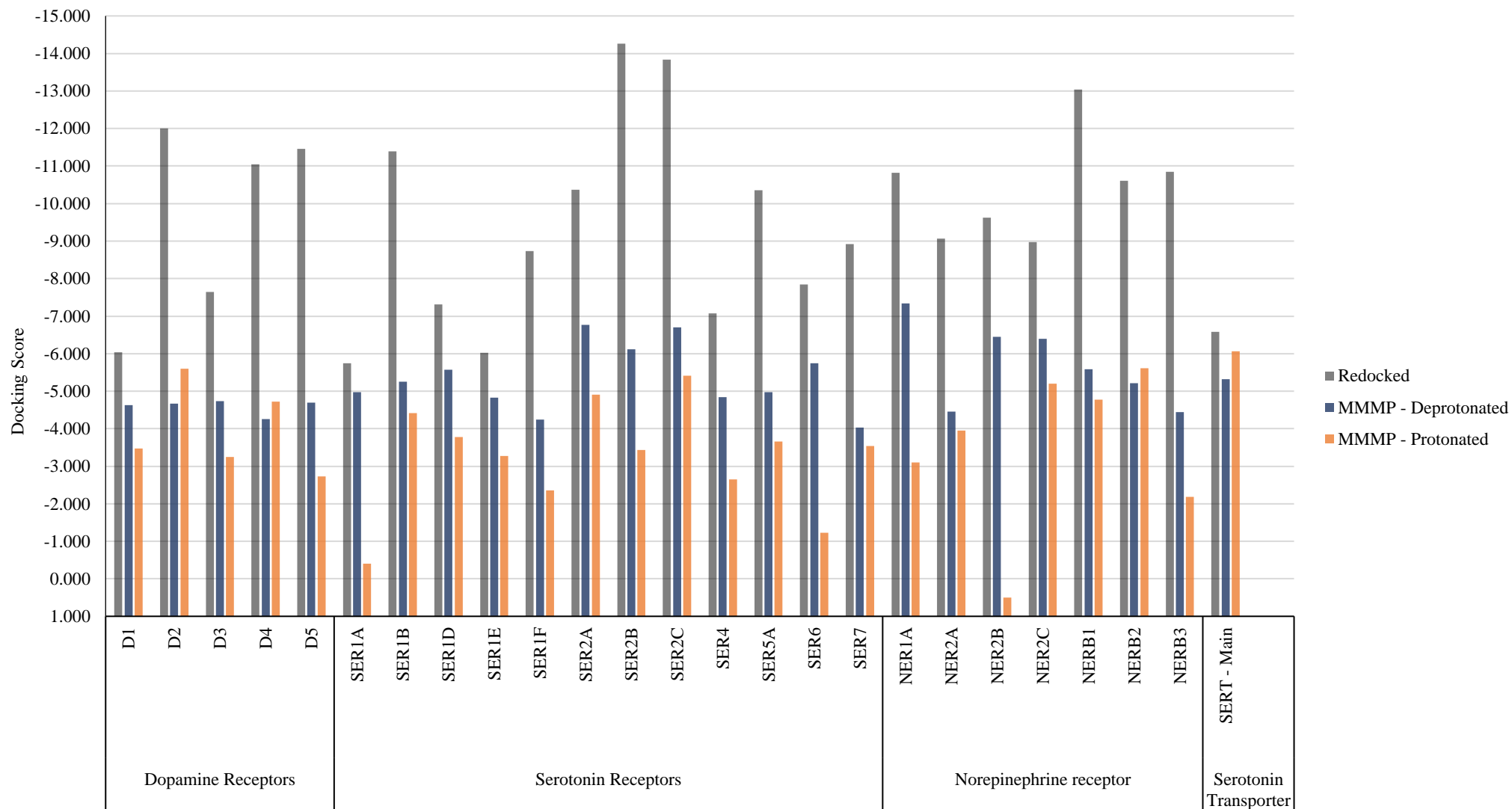


*Figure 4.2 - Deprotonated and protonated states of MMMP.*

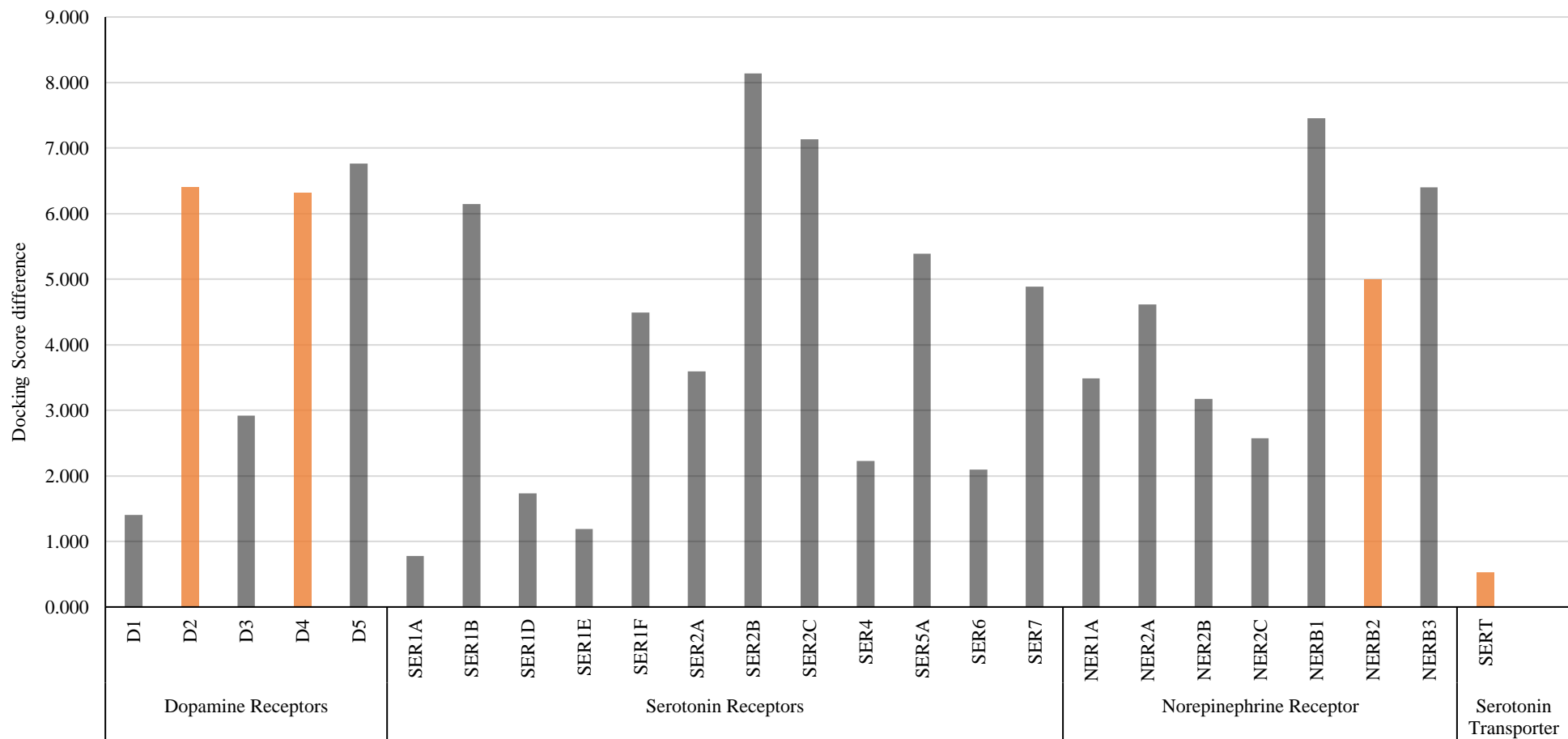
The bar graph below, **Graph 4.1**, visually represents the three obtained docking scores, for each receptor the redocked (grey bars), deprotonated MMMP (blue bars) and finally the protonated MMMP (orange bars) docking scores are plotted on the graph. The docking scores for MMMP range from 0.500 (NER2B) to -7.335 kcal/mol (Norepinephrine receptor 1A, NER1A). Ideally in docking analysis a score better (more negative) or equal to the crystal ligand score is desirable for newly docked ligands, however it is not the case with any of the MMMP dockings. None of the MMMP-protein complexes scored a docking score better or equal to the redocked score, indicating that none of the MMMP ligands were able to form a better interaction with the protein target compared to the crystal ligand. Furthermore, in the majority of the cases the deprotonated MMMP ligand was more favourable (bolded in **Table 4.1**), resulting in better docking scores and all the serotonin receptors have a preference for the deprotonated over the protonated MMMP ligand.

From these docking results the difference between the redocked ligand score and the best docking score for the MMMP ligand were calculated, with **Eq. 12** below. For the purpose of this discussion and to eliminate any possible confusion that might arise, the absolute values were used in the calculation. The best docking score could be from either the deprotonated or protonated MMMP ligand. These calculated differences are shown in **Table 4.1** and visually represented in the bar graph, **Graph 4.2** below.

**Graph 4.1** – *Glide* docking scores for redocked, deprotonated- and protonated MMMP.



**Graph 4.2** - Absolute difference between redocked and best protonated/deprotonated MMMP docking score. <sup>a</sup>



<sup>a</sup> Difference between docking scores were calculated as follow:

Redocking score – Best deprotonated/protonated MMMP docking score

<sup>b</sup> Grey bars are calculated using the deprotonated MMMP absolute docking score and the orange bars are calculated using the protonated MMMP absolute docking scores.

It goes without saying that the smaller the difference between the redocked and MMMP docking score the more likely it would be that MMMP interacts with a similar probability as the crystal ligand which is known to interact with the protein target from co-crystallisation studies.

$$|\text{Redocking score}| - |\text{Best deprotonated/protonated MMMP docking score}| \quad (12)$$

For the **dopamine receptors**, the smallest difference was calculated for dopamine receptor 1 (D1) and dopamine receptor 3 (D3) with values 1.407 kcal/mol and 2.916 kcal/mol respectively, with the remaining receptors, dopamine receptor 2 (D2, 6.405 kcal/mol), dopamine receptor 4 (D4, 6.321 kcal/mol) and dopamine receptor 5 (D5, 6.767 kcal/mol) having calculated differences greater than 6.

For the **serotonin receptors**, SER1A (0.775 kcal/mol) is the only receptor with a difference below one, two receptors serotonin receptor 1D (SER1D, 1.736 kcal/mol) and SER1E (1.193 kcal/mol) have a calculated difference smaller than two. The rest of the receptors had differences greater than two with the largest difference being calculated for serotonin receptor 2C (SER2C, 7.133 kcal/mol).

All **norepinephrine receptors** have calculated differences greater than two, with the smallest difference being calculated for norepinephrine receptor  $\alpha$ -2C (NER2C, 2.575 kcal/mol) and the largest for norepinephrine receptor  $\beta$ -1 (NERB1, 7.452 kcal/mol). The **serotonin transporter (SERT)** orthosteric binding pocket had the smallest difference out of the 25 protein targets with a value of 0.519 kcal/mol being calculated.

For the purpose of this study it was decided to further test promising protein targets by using the absolute difference calculated. The targets that scored a difference of less than one were selected for further evaluation through MD simulations to evaluate their stability over a 200 ns time frame with a recording interval of 100 ps. Therefore, SER1A-redocked ligand, SER1A-MMMP (deprotonated), SERT-redocked ligand and SERT-MMMP (protonated) protein ligand complexes were selected.

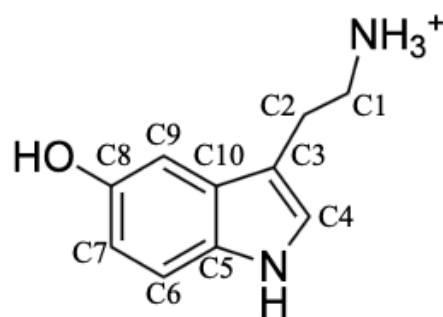
In the next section SER1A and SERT binding pockets and their interactions with their ligands (SER/MMMP) are examined for their general interactions with the binding pocket and possible interaction similarities that exist within each of the protein targets. It should be noted that all the interactions shown in the tables below (**Tables 4.2 to 4.5**) do not appear on the interaction diagrams for each of the complexes as the Schrödinger program only shows the most prominent

interactions that are present. These interactions shown in the interaction diagrams are for simplicity bolded in the tables.

Hydrogen bond interactions were considered significant when two conditions were met: (1) distance between hydrogen and acceptor is not greater than 3.5 Å and (2) the angle between the donor atom-hydrogen...acceptor is not smaller than 120°. [389, 390] Aromatic ring interactions were considered significant (1) if the distances between the ring centroids (Cg) were less than 7 Å (2) and the dihedral angles between the two planes drawn lengthwise through each ring fall between 30° to 90°. [391, 392] Ion-ion interactions (salt bridge), ion-dipole and ion-induced dipole interactions were considered significant if the interaction lengths were less than 3.5 Å. [393, 394]

#### 4.4.1 SER1A-SER complex

Interaction diagram (**Figure 4.4**) for SER1A-serotonin (SER) complex reveals a strong salt bridge interaction that forms between ASP<sup>116</sup>-COO<sup>-</sup>...NH<sub>3</sub><sup>+</sup>-SER with a interaction length of 3.22 Å. Six possible hydrogen bond interactions occur between SER1A binding pocket and serotonin, these hydrogen bond interaction lengths range between 1.59 to 3.22 Å and have angles greater than 120°. The short bond length of 1.59 Å in the ASP<sup>116</sup>-COO<sup>-</sup>...H<sub>3</sub>N<sup>+</sup>-SER interaction can be explained by the ion-ion interaction already occurring between ASP<sup>116</sup> and the protonated amine residue in serotonin.  $\pi - \pi$  Interaction are also observed between PHE<sup>361</sup> and PHE<sup>362</sup> and the indole ring of serotonin. All these interactions are summarised **Table 4.2** and graphically shown in **Figure 4.4 (a) - (b)** below.



*Figure 4.3 - Labelled serotonin molecule.*

Two distinct regions exist within the binding pocket of SER1A, these regions accommodate the different polarity regions of SER. A hydrophilic region is created by amino acid residues SER<sup>199</sup> and THR<sup>121, 200</sup> which accommodate the electron rich atoms/regions and a hydrophobic region created by amino acid residues ALA<sup>203</sup>, CYS<sup>120</sup>, ILE<sup>124, 167</sup>, PHE<sup>361, 362</sup>, TRP<sup>358</sup>, TYR<sup>390</sup> and VAL<sup>117</sup> in which the rest of SER is accommodated.

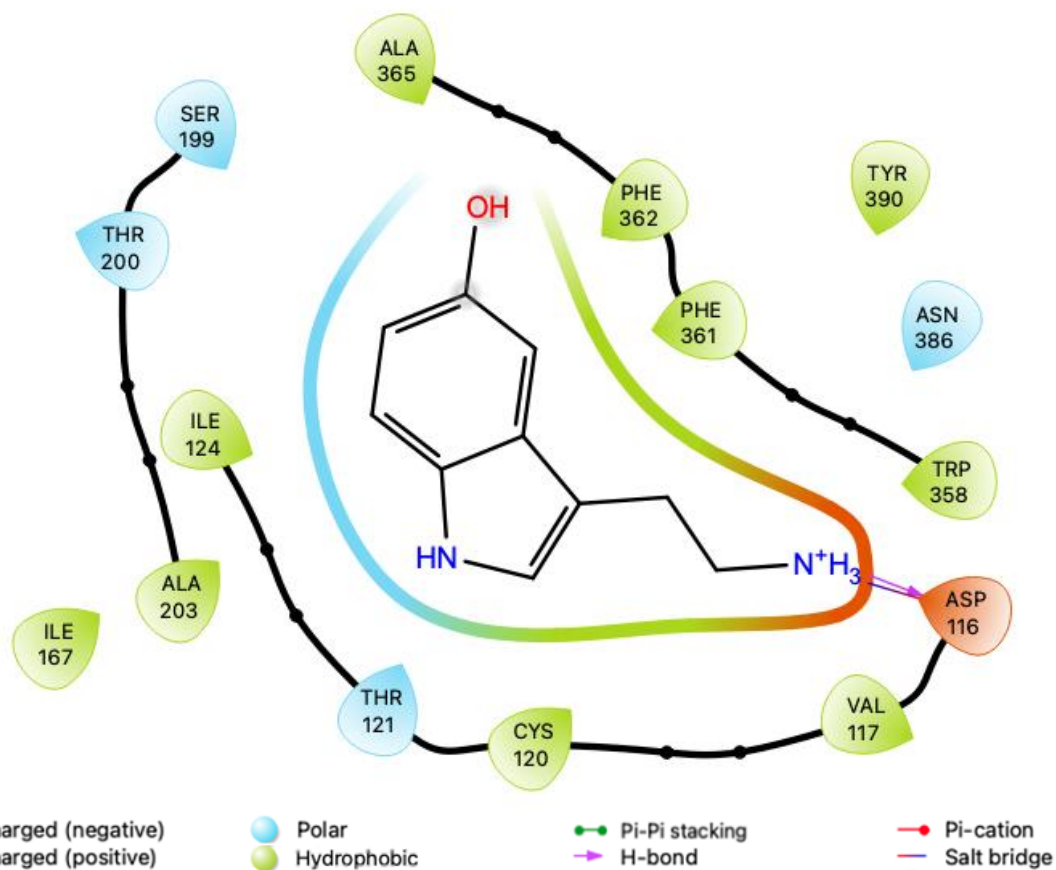


**Table 4.2** - SERIA-SER interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below.

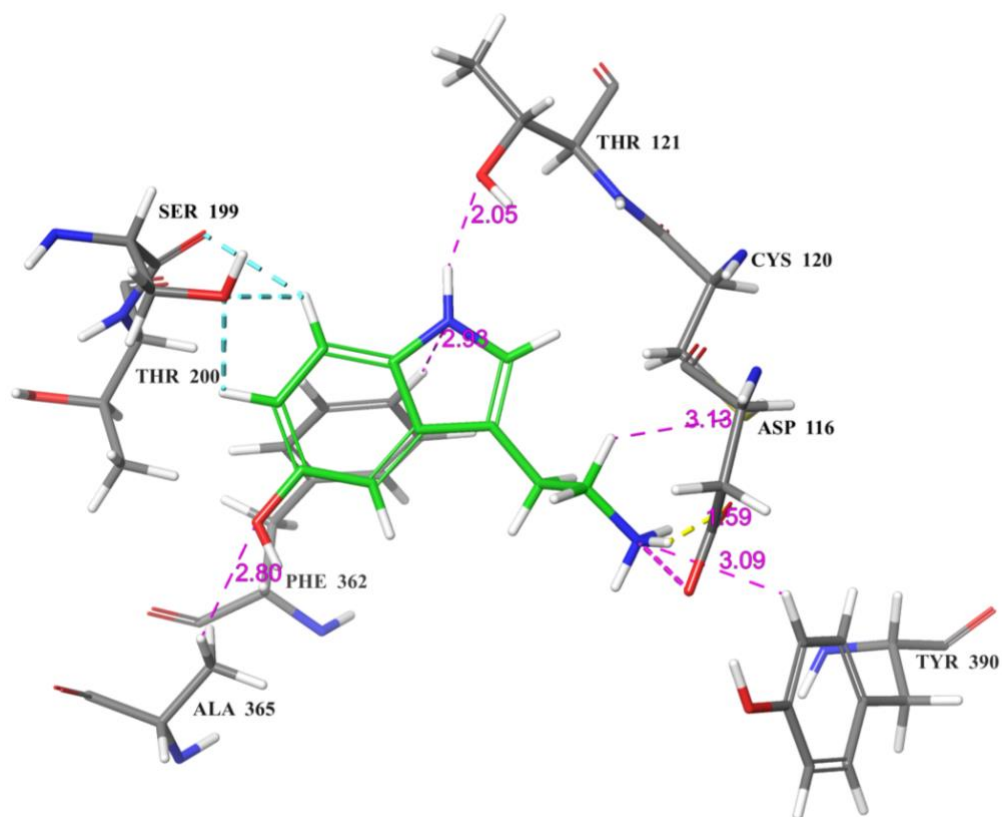
| Interaction number                | Protein residue...Ligand<br>[Description of amino acid residue that forms the interaction]                    | Interaction length (Å)  | ∠ Angles (°)      |
|-----------------------------------|---|---|-------------------|
| <b>Ion-Ion interaction</b>        |   |   |                   |
| 1.                                | <b>ASP<sup>116</sup>-COO<sup>-</sup>...NH<sub>3</sub><sup>+</sup>-SER</b><br>[Charged oxygen side chain]      | 3.22  | N/A               |
| <b>Hydrogen Bond Interactions</b> |   |   |                   |
| 2.                                | <b>ASP<sup>116</sup>-COO<sup>-</sup>... H<sub>3</sub>N<sup>+</sup>-SER</b><br>[Carbonyl oxygen of side chain] | 1.59  | 145.5             |
| 3.                                | TYR <sup>390</sup> -CH... H <sub>3</sub> N <sup>+</sup> -SER<br>[Aromatic ring hydrogen]                      | 3.09  | 124.9             |
| 4.                                | THR <sup>121</sup> -HO...HN-SER<br>[Hydroxyl oxygen in side chain]  | 2.05  | 150.4             |
| 5.                                | PHE <sup>362</sup> -CH...NH-SER<br>[Aromatic ring hydrogen]   | 2.93  | 147.3             |
| 6.                                | ALA <sup>365</sup> -CH <sub>3</sub> ...OH-SER   | 2.80  | 133.5             |
| 7.                                | Serine <sup>199</sup> =O...H(C6)-SER<br>[Carbonyl oxygen of backbone]   | 2.56  | 136.9             |
| <b>π – π Interactions</b>         |   |   |                   |
|                                   |   | Interaction length (Å)<br>between centroids                                   | Dihedral angle(°) |
| 8.                                | PHE <sup>361</sup> -Cg(phenyl)...Cg(Indole ring)-SER <sup>a</sup>   | 6.08/6.21<br>Distance to centre of the benzene and pyrrole ring respectively. | 82.6              |
| 9.                                | PHE <sup>362</sup> -Cg(phenyl)...Cg(Indole ring)-SER <sup>a</sup>   | 4.53/4.84<br>Distance to centre of the benzene and pyrrole ring respectively. | 56.1              |

<sup>a</sup>Cg refers to the aromatic ring centroid.

(a)



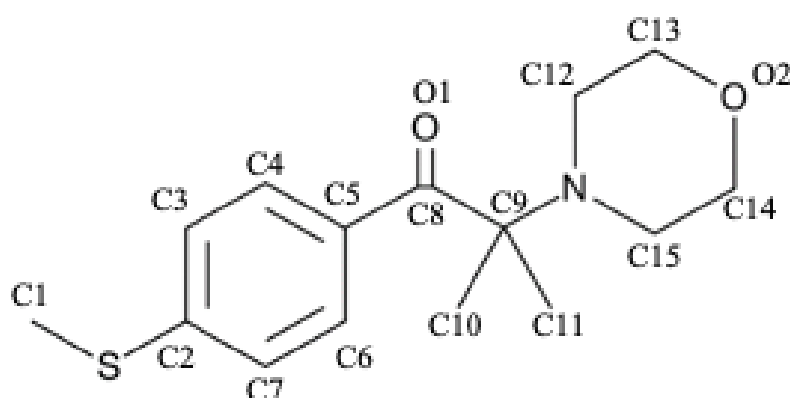
(b)



*Figure 4.4 - (a) Interaction diagrams for SERIA-SER complex. (b) Simplified view of interactions with some residues hidden.*

#### 4.4.2 SER1A-MMMP complex

The interactions observed between SER1A receptor and MMMP are very similar to those observed with SER. The strong ion-ion interaction is replaced by a weaker ion-dipole interaction between the protonated amine of LYS<sup>191</sup> and the oxygen of the morpholino moiety, (O2, **Figure 4.5**). Seven possible hydrogen bond interactions are also observed between the ligand and the receptor, these hydrogen bond interaction lengths range between 2.43 - 3.05 Å, with angles greater than 120°. Three of the hydrogen bond interactions are formed with the same amino acid residues (PHE<sup>362</sup>, Serine<sup>199</sup> and ALA<sup>365</sup>, **Interactions 2, 4 and 8 listed in Table 4.3**) as observed in SER1A-SER complex.



*Figure 4.5 - Labelled MMMP molecule.*

$\pi - \pi$  Interaction, similar to SER1A-SER complex, are also observed between PHE<sup>361</sup> and PHE<sup>362</sup> and the methylthiophenyl moiety of MMMP. All these interactions are summarised in **Table 4.3** and graphically presented in **Figure 4.6 (a) - (b)** below.

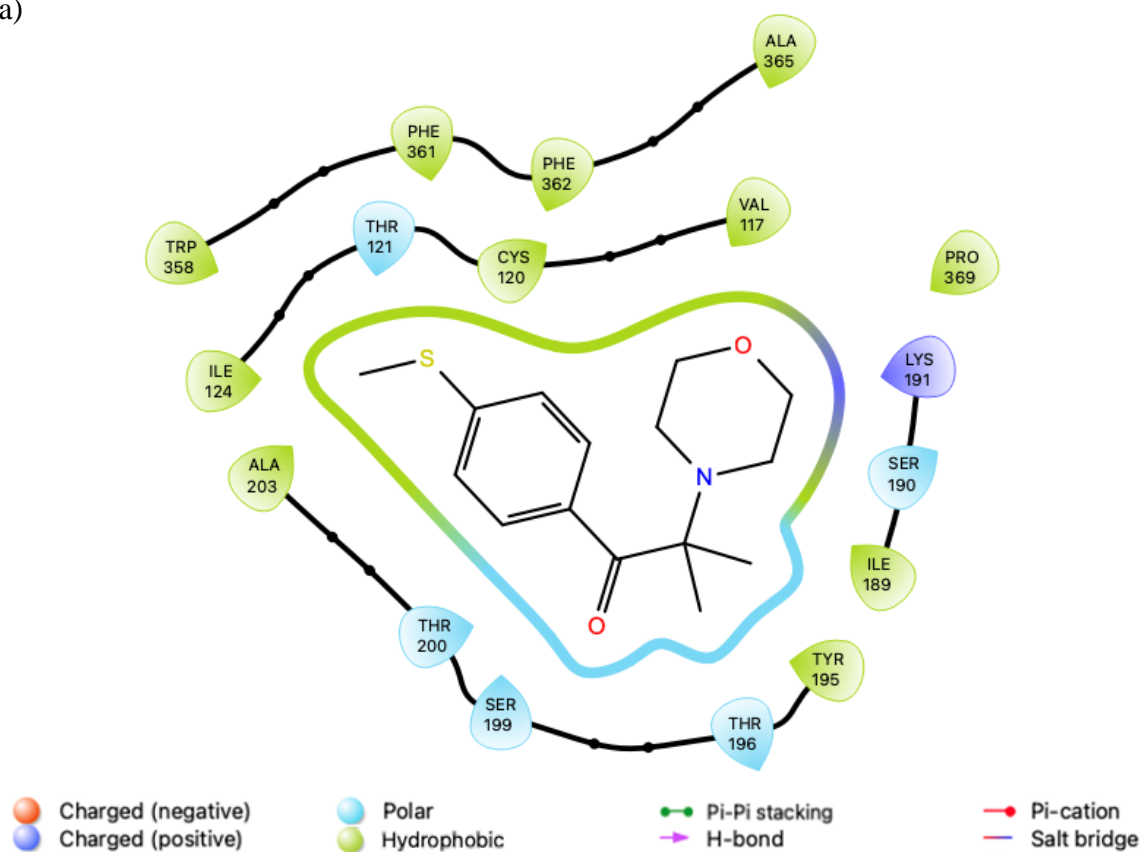
Similarly to the SER1A-SER complex two distinct regions exist within the binding pocket of SER1A, these regions accommodate the different polarity regions of MMMP. A hydrophilic region is created by amino acid residues SER<sup>190, 199</sup> and THR<sup>200, 196</sup> which accommodate the electron rich atoms/regions and a hydrophobic region created by amino acid residues ALA<sup>203, 365</sup>, CYS<sup>120</sup>, ILE<sup>124, 189</sup>, PHE<sup>361, 362</sup>, TRP<sup>358</sup>, TYR<sup>195</sup> and VAL<sup>117</sup> in which the rest of MMMP is accommodated.

**Table 4.3** - SERIA-MMMP interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below.

| Interaction number                | Protein residue...Ligand<br>[Description of amino acid residue that forms the interaction] | Interaction length (Å)                      | ∠ Angles (°)      |
|-----------------------------------|--|---|-------------------|
| <b>Ion-Dipole Interaction</b>     |  |   |                   |
| 1.                                | <b>LYS<sup>191</sup>-H<sub>3</sub>N<sup>+</sup>... (O2)-MMMP</b>                           | 3.36  | N/A               |
| <b>Hydrogen Bond Interactions</b> |  |   |                   |
| 2.                                | <b>PHE<sup>362</sup>-CH...S(C1)H<sub>3</sub>-MMMP</b>                                      | 2.90  | 127.1             |
| 3.                                | <b>CYS<sup>120</sup>=O...H<sub>3</sub>(C1)S-MMMP</b><br>[Carbonyl oxygen of backbone]      | 3.05  | 154.0             |
| 4.                                | <b>Serine<sup>199</sup>-HO...H(C4)-MMMP</b><br>[Side chain hydroxyl group]                 | 2.42  | 125.9             |
| 5.                                | <b>TYR<sup>195</sup>-CH... (O1)-MMMP</b><br>[Aromatic ring H-bond]                         | 2.76  | 117.7             |
| 6.                                | <b>THR<sup>196</sup>-HO...H<sub>3</sub>(C11)-MMMP</b><br>[Side chain hydroxyl group]       | 2.42  | 147.9             |
| 7.                                | <b>Serine<sup>190</sup>=O...H<sub>3</sub>(C11)-MMMP</b><br>[Carbonyl oxygen of backbone]   | 2.56  | 152.4             |
| 8.                                | <b>ALA<sup>365</sup>-O...H(C15)-MMMP</b><br>[Carbonyl oxygen of backbone]                  | 2.86  | 154.7             |
| <b>π – π Interactions</b>         |  |   |                   |
|                                   |  | Interaction length (Å)<br>between centroids | Dihedral angle(°) |
| 9.                                | <b>PHE<sup>361</sup>-Cg(phenyl)...<br/>Cg(methylthiophenyl)-MMMP<sup>a</sup></b>           | 5.85  | 80.6              |
| 10.                               | <b>PHE<sup>362</sup>-Cg(phenyl)...<br/>Cg(methylthiophenyl)-MMMP<sup>a</sup></b>           | 5.55  | 63.9              |

<sup>a</sup>Cg refers to the aromatic ring centroid.

(a)



(b)

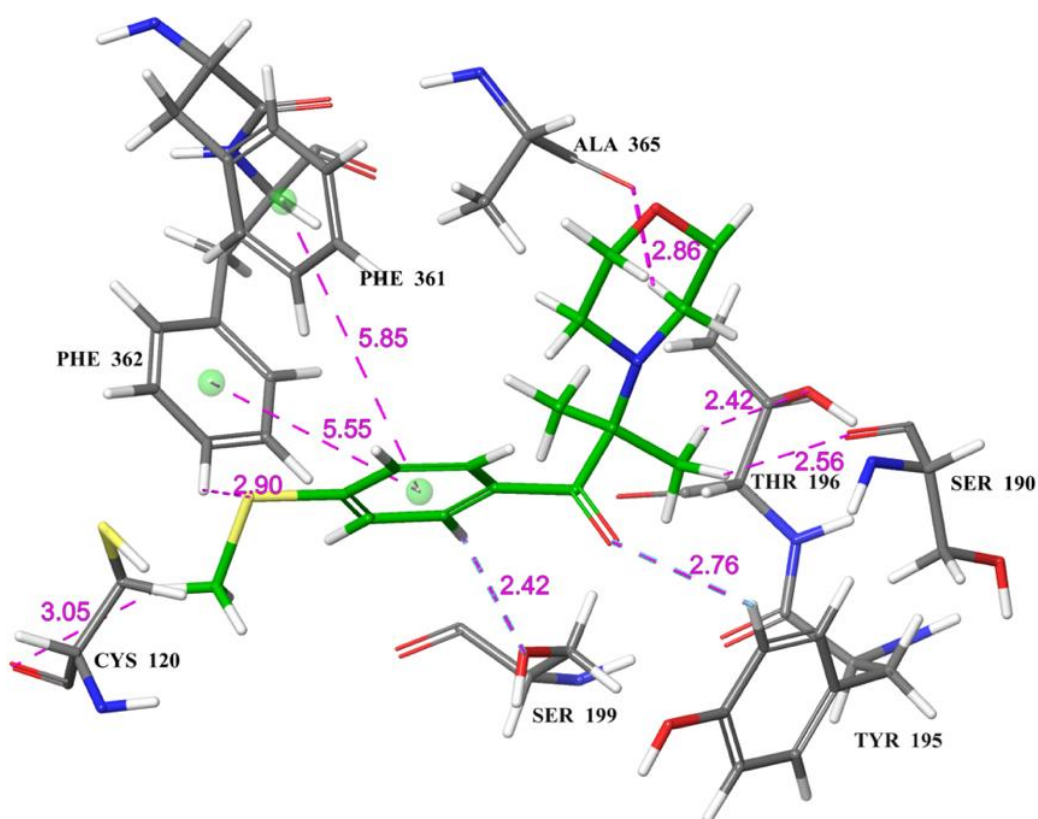


Figure 4.6 - (a) Interaction diagrams for SERIA-MMMP complex. (b) Simplified view of interactions with some residues hidden.

#### 4.4.3 SERT-SER complex

Analysis of the SERT-SER complex reveals a plethora of interactions between the ligand and the protein. Five different types of interactions can be observed: (1) ion-ion, (2) ion-dipole, (3) hydrogen bonding, (4)  $\pi$ -cation and (5)  $\pi - \pi$  interactions.

Both the ion-ion and ion-dipole interaction occur between carboxylic acid side chain of ASP<sup>98</sup> and the protonated amine residue of SER, with the difference being the atom with which ASP<sup>98</sup> interacts with. ASP<sup>98</sup> is a highly conserved amino acid in SERT, interactions with it is of vital importance to stabilise the complex.<sup>[223]</sup> As expected ASP<sup>98</sup> in SERT forms an interaction with the charged amine nitrogen of SER yielding a strong ion-ion interaction with an interaction length of 2.72 Å, whilst the interaction with its hydrogen yields the ion-dipole interaction with a interaction length of 1.78 Å, it should be noted that this interaction can only occur between one of the amine hydrogens as the other two hydrogens have unfavourable angles of less than 70°. Another ion-dipole interaction is observed between the backbone carbonyl oxygen of PHE<sup>335</sup> and the charged amine residue with an interaction length of 2.75 Å.

Seven hydrogen bond interactions are observed with interaction lengths ranging from 1.75 to 3.47 Å of which three of these hydrogen interactions are between TYR<sup>95</sup> and SER. The very short interaction length observed between PHE<sup>335</sup> and the charged amine residue of SER can be attributed to the dominant and stronger ion-dipole interaction also present between the charged residues. A single  $\pi$ -cation interaction with the phenol side chain of TYR<sup>95</sup> and the protonated amine residue of SER with an interaction bond length of 5.69 Å. Three  $\pi - \pi$  interactions are also seen, of particular interest is once again the aromatic interaction with the amino acid residue TYR<sup>95</sup>. Based on the number of interactions with this amino acid it might be concluded that TYR<sup>95</sup> plays a critical role in the binding of possible ligands in the orthosteric binding pocket of SERT. All interactions mentioned here are summarised in **Table 4.4** below as well as graphically presented in **Figure 4.7 (a) - (b)** below.

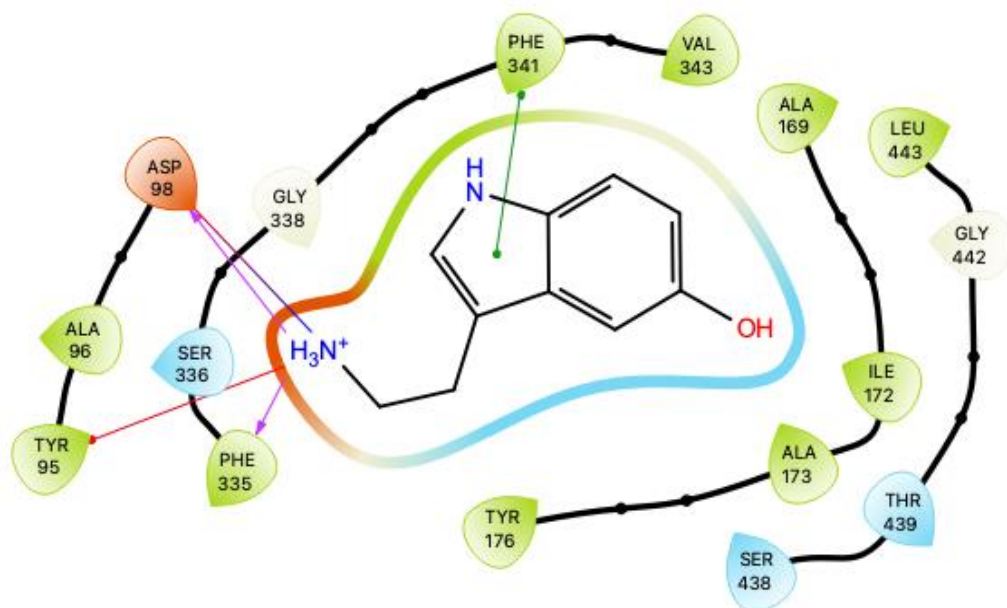
Similarly to the SER1A complexes two distinct regions exists within the binding pocket of SERT, these regions accommodate the different polarity regions of SER. A hydrophilic region is created by amino acids residues SER<sup>336, 438</sup> and THR<sup>439</sup> which accommodate the electron rich atoms/regions and a hydrophobic region created by amino acid residues ALA<sup>96, 169, 173</sup>, ILE<sup>172</sup>, LEU<sup>443</sup>, PHE<sup>335, 341</sup>, TYR<sup>95, 176</sup> and VAL<sup>343</sup> in which the rest of MMMP is accommodated.

**Table 4.4** - SERT-SER interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below.

| Interaction number                | Protein residue...Ligand<br>[Description of amino acid residue that forms the interaction]                                      | Interaction length (Å)                      | ∠ Angles (°)            |
|-----------------------------------|---|---|-------------------------|
| <b>Ion-Ion Interaction</b>        |   |   |                         |
| 1.                                | <b>ASP<sup>98</sup>-COO<sup>-</sup>...NH<sub>3</sub><sup>+</sup>-SER</b><br>[Charged oxygen of the carboxylic acid side chain]  | 2.72  | N/A                     |
| <b>Ion-dipole Interactions</b>    |   |   |                         |
| 2.                                | ASP <sup>98</sup> -COO <sup>-</sup> ...H <sub>3</sub> N <sup>+</sup> -SER<br>[Charged oxygen of the carboxylic acid side chain] | 1.78  | N/A                     |
| 3.                                | PHE <sup>335</sup> =O...NH <sub>3</sub> <sup>+</sup> -SER<br>[Carbonyl oxygen of backbone]                                      | 2.75  | N/A                     |
| <b>Hydrogen Bond Interactions</b> |   |   |                         |
| 4.                                | ILE <sup>172</sup> -CH <sub>3</sub> ...OH-SER   | 3.26  | 152.7                   |
| 6.                                | PHE <sup>341</sup> -CH <sub>2</sub> ...NH-SER   | 3.13  | 125.1                   |
| 7.                                | TYR <sup>95</sup> -HO...H(C4)-SER<br>[Hydroxyl of side chain]   | 2.74  | 127.1                   |
| 8.                                | TYR <sup>95</sup> =O...H <sub>2</sub> (C2)-SER<br>[Carbonyl oxygen of backbone]   | 3.47  | 154.6                   |
| 9.                                | <b>ASP<sup>98</sup>-COO<sup>-</sup>...H<sub>3</sub>N<sup>+</sup>-SER</b><br>[Carbonyl carbon of the carboxylic acid side chain] | 3.47  | 127.0                   |
| 10.                               | TYR <sup>95</sup> =O...H <sub>3</sub> N <sup>+</sup> -SER<br>[Carbonyl oxygen of backbone]                                      | 3.39  | 155.9                   |
| 11.                               | <b>PHE<sup>335</sup>=O...H<sub>3</sub>N<sup>+</sup>-SER</b><br>[Carbonyl oxygen of backbone]                                    | 1.75  | 166.4                   |
| <b>π-cation Interaction</b>       |   |   |                         |
| 12.                               | <b>TYR<sup>95</sup>-Cg(phenol ring)...</b><br><b>SER</b><br>[Carbonyl carbon of the carboxylic acid side chain]                 | 5.69  | Could not be determined |
| <b>π – π Interactions</b>         |   |   |                         |
|                                   |   | Interaction length (Å)<br>between centroids | Dihedral angle(°)       |
| 13.                               | <b>PHE<sup>341</sup>-Cg(phenyl)...</b><br><b>Cg(pyrrole ring)-SER<sup>a</sup></b>   | 4.26  | 64.2                    |
| 14.                               | TYR <sup>95</sup> -Cg(phenol ring)...   | 5.38  | 57.1                    |
| 15.                               | TYR <sup>176</sup> -Cg(phenol ring)...  | 5.57  | 81.4                    |
|                                   | Cg(benzene ring)-SER <sup>a</sup>   |   |                         |

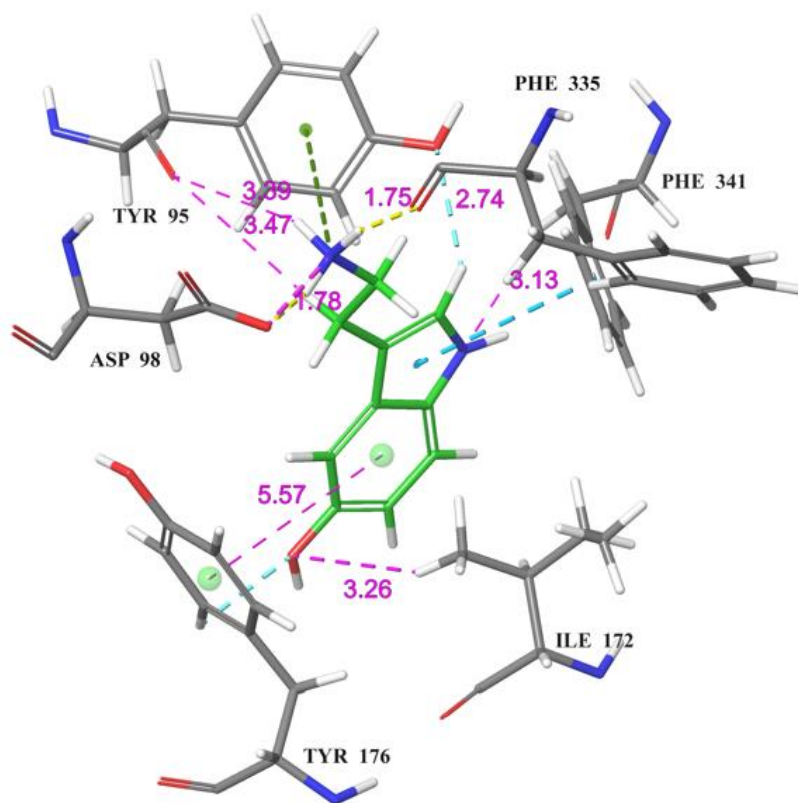
<sup>a</sup>Cg refers to the aromatic ring centroid.

(a)



- Charged (negative)
- Charged (positive)
- Polar
- Hydrophobic
- Pi-Pi stacking
- H-bond
- Pi-cation
- Salt bridge

(b)



**Figure 4.7** - (a) Interaction diagrams for SERIA-MMMP complex. (b) Simplified view of interactions with some residues hidden.



#### 4.4.4 SERT-MMMP complex

Analysis of the SERT-MMMP complex reveals that four interactions types that are present within the SERT-SER also occur between SERT and MMMP with the ion-dipole being replaced by ion-induced dipole interactions. Similarly to the SERT-SER complex the ion-ion interaction occurs with the charged carboxylic acid of ASP<sup>98</sup> and a charged protonated nitrogen of the morpholino moiety with an interaction length of 3.47 Å. ASP<sup>98</sup> is a highly conserved amino acid in SERT, interactions with it is of vital importance to stabilise the complex. The ion-induced dipole interactions that occur are between the side chain charged carboxylic acid of ASP<sup>98</sup> and the hydrogens of the C12 and C13 of the morpholino moiety.

Six hydrogen bond interactions with interaction lengths ranging from 2.34 to 3.28 Å occur. Similarly observed with SER two of these interactions are between the backbone carbonyl oxygen of TYR<sup>95</sup> and alkyl substituents and one interaction between the backbone carbonyl carbon of PHE<sup>335</sup> and a hydrogen of a nitrogen donor atom. A  $\pi$ -cation interaction is also observed, in the case of the MMMP ligand this interaction occurs between PHE<sup>341</sup>, instead of TYR<sup>95</sup> in the case of SER, and the charged amine residue with an interaction length of 4.95 Å. All three  $\pi - \pi$  interactions between amino acid residues TYR<sup>95,176</sup> and PHE<sup>341</sup> are also observed with the MMMP ligand. All interactions mentioned here are summarised in **Table 4.5** as well as graphically presented in **Figure 4.8 (a) - (b)** below.

Similarly to the SER1A complexes two distinct regions exists within the binding pocket of SERT, these regions accommodate the different polarity regions of SER. A hydrophilic region is created by amino acids residues ASN<sup>177</sup>, SER<sup>336, 438</sup> and THR<sup>439, 497</sup> which accommodate the electron rich atoms/regions and a hydrophobic region created by amino acid residues ALA<sup>169, 173</sup>, ILE<sup>172</sup>, LEU<sup>443</sup>, PHE<sup>335, 341</sup>, TYR<sup>95, 175, 176</sup> and VAL<sup>343</sup> in which the rest of MMMP is accommodated.

The analysis of SER1A-SER and SER1A-MMMP/SERT-SER and SERT-MMMP has shed some light on the similar interactions that occur between the crystal ligands (SER) and the docked ligands (MMMP). These interactions have echoed and enforced the validity of the small differences in docking scores observed for each protein target (**Table 4.1** and **Graph 4.2**). From the above mentioned interactions it was seen that the number of similarities between SERT and the ligands (SER and MMMP) were greater than those observed between SER1A and the same ligands. The highly conserved ASP<sup>98</sup> in SERT was observed to form salt bridges in both SERT-SER and SERT-MMMP complexes. The strongest ion-ion interaction (ASP<sup>116</sup>-

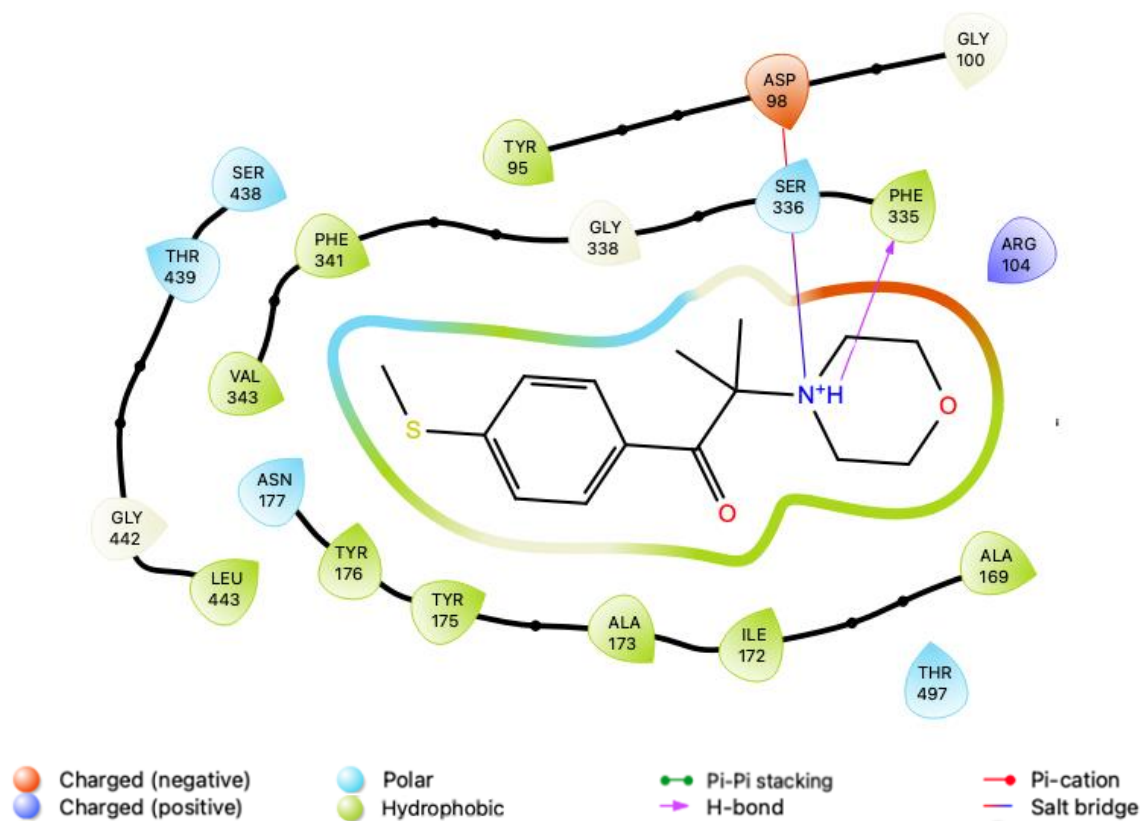
COO<sup>-</sup>...NH<sub>3</sub><sup>+</sup>-SER) between SER1A and SER was absent for the SER1A-MMMP, but possibly replaced by a weaker ion-dipole interaction (LYS<sup>191</sup>- H<sub>3</sub>N<sup>+</sup>...(O<sub>2</sub>)-MMMP).

**Table 4.5** - SERT-MMMP interactions through analysis of the binding pocket. Bolded interactions are the most prominent interactions as indicated on the ligand interaction diagram below.

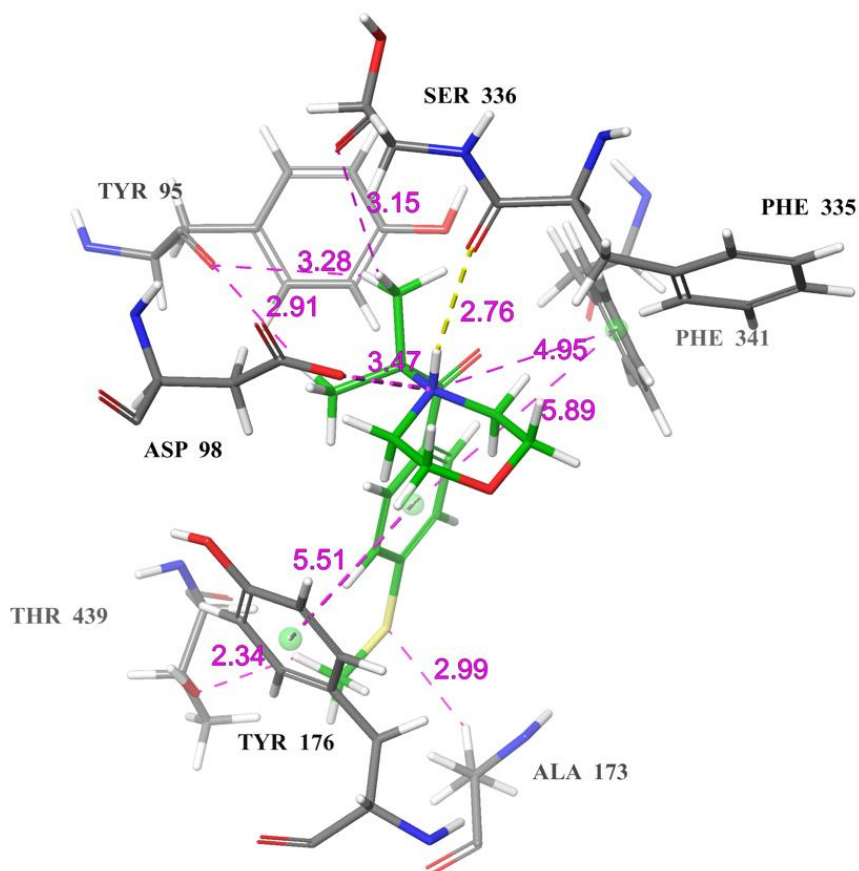
| Interaction number                     | Protein residue...Ligand<br>[Description of amino acid residue that forms the interaction]                             | Interaction length (Å)                                   | ∠ Angles (°)            |
|--|--|--|-------------------------|
| <b>Ion-Ion Interaction</b>             |  |  |                         |
| 1.                                     | <b>ASP<sup>98</sup>-COO<sup>-</sup>...N<sup>+</sup>H-MMMP</b><br>[Charged oxygen of the carboxylic acid side chain]    | 3.47   | N/A                     |
| <b>Ion-induced dipole interactions</b> |  |  |                         |
| 2.                                     | ASP <sup>98</sup> -COO <sup>-</sup> ...H <sub>2</sub> (C12)-MMMP<br>[Charged oxygen of the carboxylic acid side chain] | 2.16   | N/A                     |
| 3.                                     | ASP <sup>98</sup> -COO <sup>-</sup> ...H <sub>2</sub> (C13)-MMMP<br>[Charged oxygen of the carboxylic acid side chain] | 2.17, 2.97<br>[Possible to interact with both hydrogens] | N/A                     |
| <b>Hydrogen Bond Interactions</b>      |  |  |                         |
| 4.                                     | THR <sup>439</sup> -HO...H <sub>3</sub> (C1)S-MMMP   | 2.34   | 132.5                   |
| 5.                                     | ALA <sup>173</sup> -CH <sub>3</sub> ...S(C1)H <sub>3</sub> -MMMP   | 2.99   | 130.2                   |
| 6.                                     | TYR <sup>95</sup> =O...H <sub>3</sub> (C10)-MMMP<br>[Carbonyl oxygen of backbone]                                      | 2.91   | 131.0                   |
| 7.                                     | TYR <sup>95</sup> =O...H <sub>3</sub> (C11)-MMMP<br>[Carbonyl oxygen of backbone]                                      | 3.28   | 121.8                   |
| 8.                                     | Serine <sup>336</sup> =O...H <sub>3</sub> (C11)-MMMP<br>[Carbonyl oxygen of backbone]                                  | 3.15   | 143.1                   |
| 9.                                     | <b>PHE<sup>335</sup>=O...HN<sup>+</sup>-MMMP</b><br>[Carbonyl oxygen of backbone]                                      | 2.76   | 170.5                   |
| <b>π-cation Interaction</b>            |  |  |                         |
| 10.                                    | PHE <sup>341</sup> -Cg(phenyl)...N <sup>+</sup> H-MMMP   | 4.95   | Could not be determined |
| <b>π - π Interactions</b>              |  |  |                         |
|  |  | Interaction length (Å)<br>between centroids              | Dihedral angle(°)       |
| 11.                                    | TYR <sup>95</sup> -Cg(phenol ring)...<br>Cg(methylthiophenyl)-MMMP <sup>a</sup>  | 6.40   | 70.3                    |
| 12.                                    | TYR <sup>176</sup> -Cg(phenol ring)...<br>Cg(methylthiophenyl)-MMMP <sup>a</sup>                                       | 5.51   | 89.2                    |
| 13.                                    | PHE <sup>341</sup> -Cg(phenyl)...<br>Cg(methylthiophenyl)-MMMP <sup>a</sup>  | 5.89   | 56.7                    |

<sup>a</sup>Cg refers to the aromatic ring centroid.

(a)



(b)



**Figure 4.8** - (a) Interaction diagrams for SERIA-MMMP complex. (b) Simplified view of interactions with some residues hidden.

In the case of SER1A and SERT it can be concluded that the number of similarities in terms of interactions observed between the protein targets and the ligands are directly proportional to obtaining a similar docking score and inversely proportional to the absolute difference calculated between the crystal and docked ligand for a specific protein target. Thus, a greater similarity in terms of interactions that occur within the binding pocket and other ligands the more similar the docking score becomes and therefore, a smaller absolute difference in docking scores are observed.

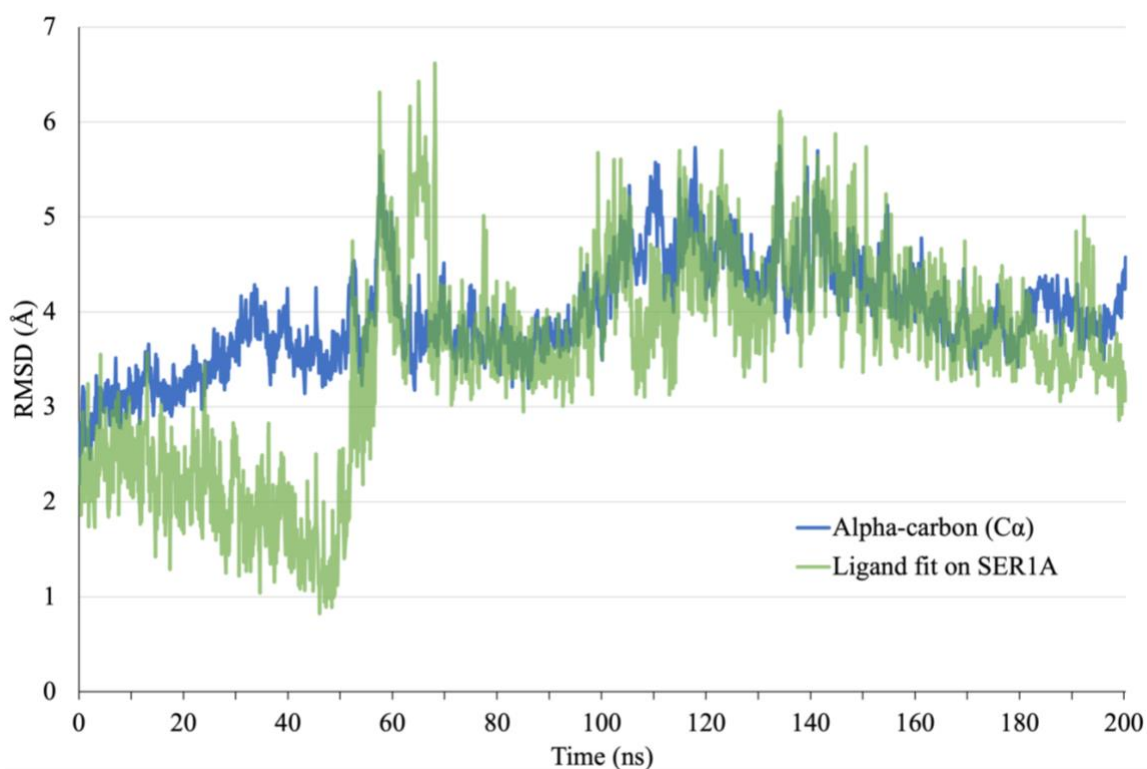
#### 4.5 MD simulations results and discussions

MD simulations were performed on the most promising docking targets, SER1A and SERT, over a 200 ns period to access the stability of the ligands.

##### 4.5.1 MD analysis of SER1A-SER and SER1A-MMMP complexes

In **Graph 4.3** below the RMSD of the alpha-carbon ( $C\alpha$ ), which is the carbon to which the amino acids' amine- and carboxylic acid functional groups and the side chain is attached to, ranges between 3 - 5 Å for majority of the frames measured. At around 100 to 140 ns the largest  $C\alpha$  movement is observed with RMSD values greater than 5 Å, the largest shift of 5.642 Å was measured at 141.4 ns.

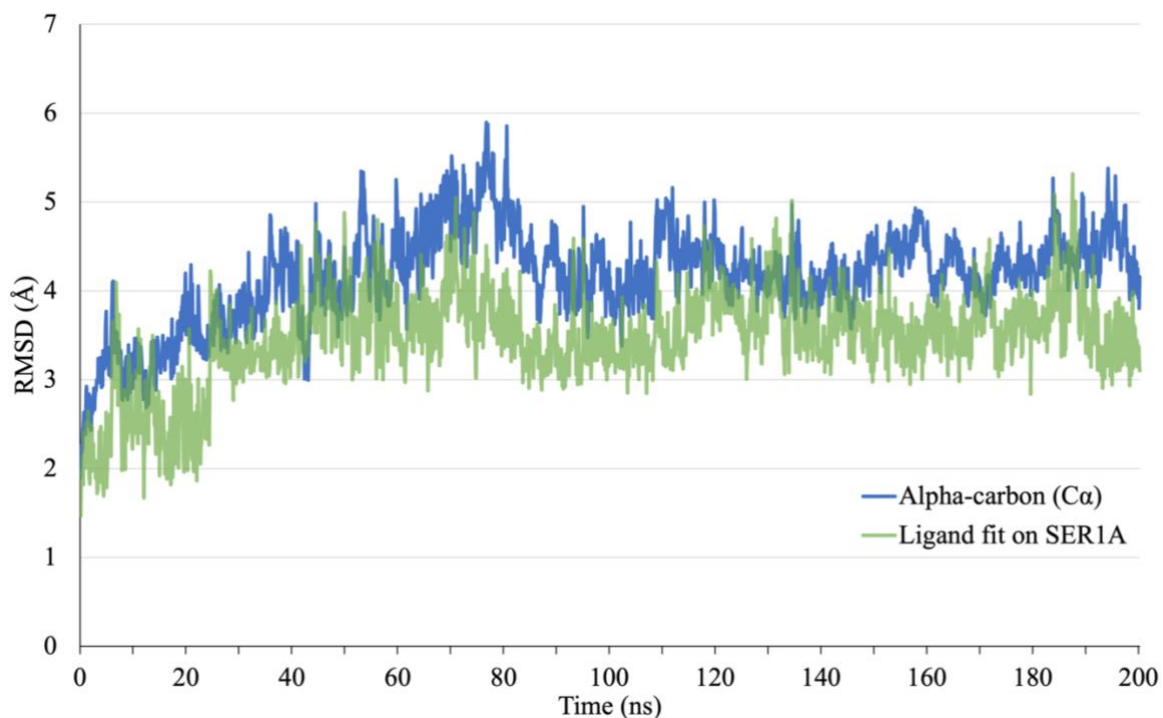
**Graph 4.3** - A plot of the ligand's (SER) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon ( $C\alpha$ ) of SER1A during the course of a 200 ns MD simulation at 310 K.



For the first 50 ns of **the ligand's RMSD** the greatest number of ligand-protein interactions were recorded which lead to a decrease in RMSD from a local maximum of 3.552 Å at 4.20 ns to a local minimum of 0.824 at 46.1 ns. This decrease in RMSD, as observed in the trajectories, can be attributed to the increased interactions that stabilises the ligand in the binding site of SER1A. From 50 to 70 ns the sharpest increase in the ligand's RMSD is observed with the highest RMSD of 6.619 being reported at 68.2 ns, thereafter the ligand's RMSD ranges between 3 to 6 Å with the majority of the measurements falling in the 3 to 5 Å range from about 70 to 200 ns. Overall both the C $\alpha$  and the ligand have moved quite a bit when compared to the initial reference frame, but can be seen converging at around 4 Å from 70 ns onward. The results of the MD simulation of the SER1A-SER complex is attached in **Annexure 1**.

The RMSD graph of the SER1A-MMMP complex (**Graph 4.4**) shows that after about 25 ns both the C $\alpha$  and the ligand RMSD ranges between 3 to 6 Å with majority of the measurements falling in the 3 to 5 Å range. Both the ligand and the C $\alpha$  appears to converge and equilibrate after about 80 ns with no major spikes in the RMSD recorded for the rest of the period, the C $\alpha$  movement fluctuates around 4.3 Å and the ligand around 3.5 Å.

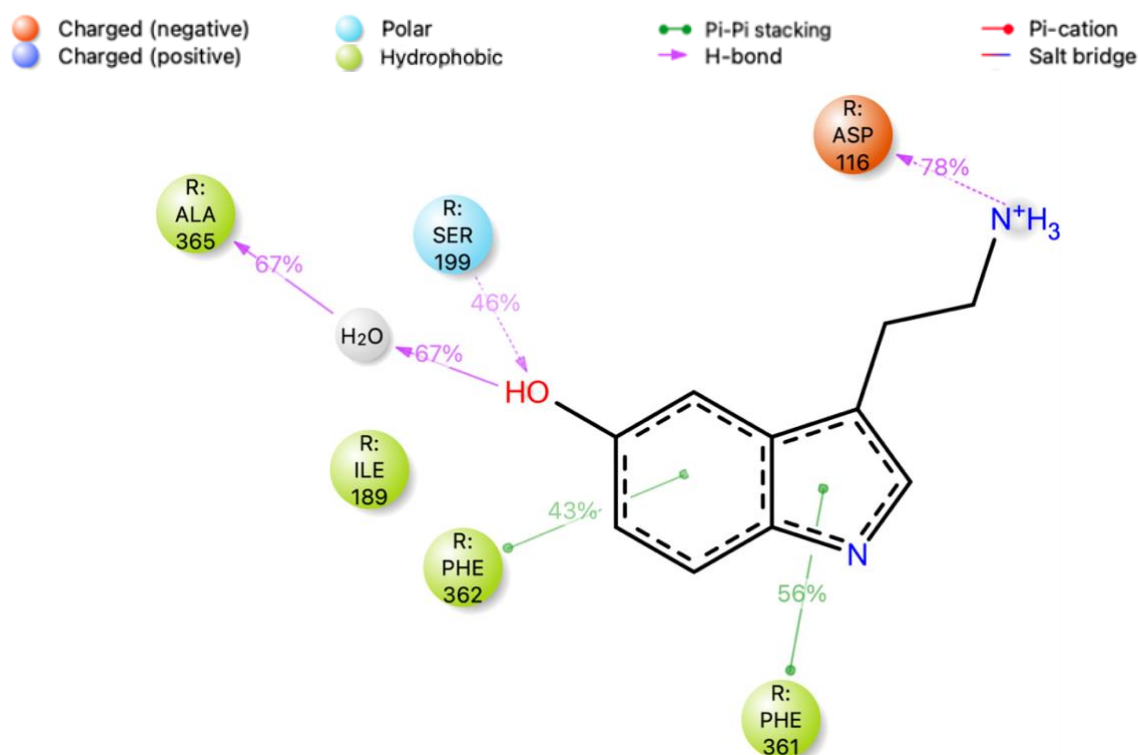
**Graph 4.4** - A plot of the ligand's (MMMP) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SER1A during the course of a 200 ns MD simulation at 310 K.



Examination of the trajectories within Maestro revealed that MMMP is in a very rigid pose with the majority of the rotations occurring around two bonds, the first being the bond between

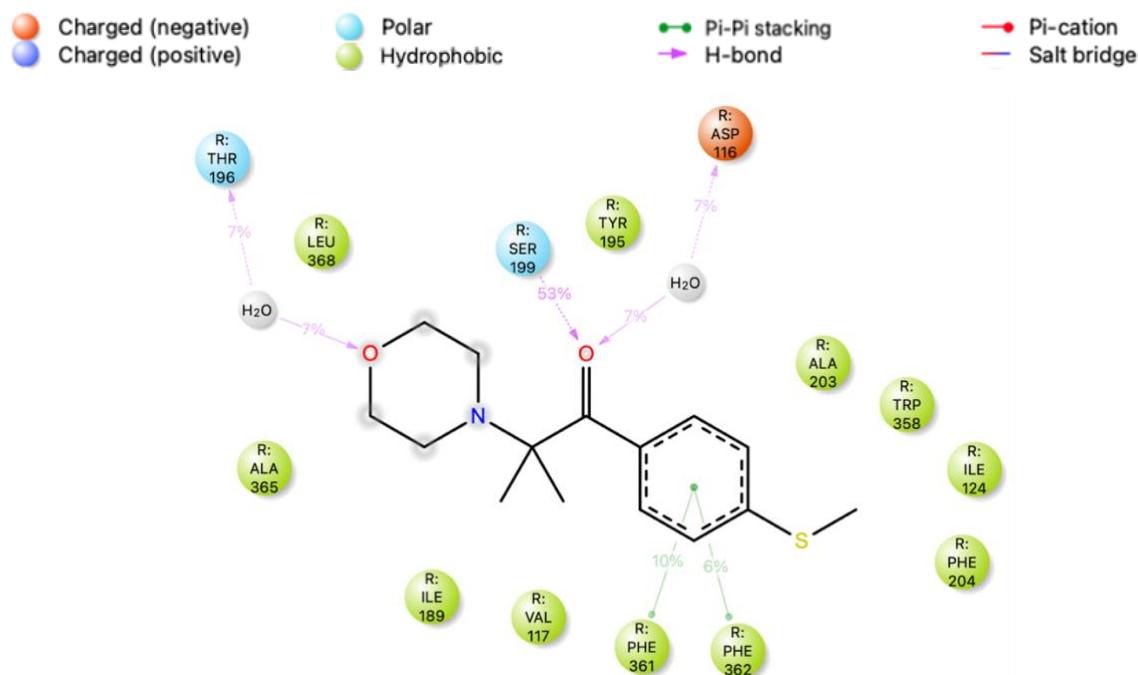
the aromatic ring and the methylthio moiety and the second the bond between the  $\beta$ -carbon and the aromatic ring. These torsion angles are confirmed by the ligand torsion profile in **Annexure 2**, from these diagrams it can be seen that the rotation of the methylthio bond causes the methyl to have torsion angles in the entire  $360^\circ$  some angles occurring more frequent than other. The  $\beta$ -carbon-aromatic ring bond only has discrete torsion angles between  $0$  to  $90^\circ$  and  $135$  to  $260^\circ$ . Furthermore, the bonds between the  $\alpha$ - and  $\beta$ -carbon and  $\alpha$ -carbon and morpholino nitrogen, only allow torsion angles between approximately  $10$  to  $110^\circ$  for the former and  $10$  to  $80^\circ$  for the latter. These limited torsion angles can be attributed to (i) steric hindrance from the bulky dimethyl substituents and morpholino moiety, (ii) repulsion between the electron dense carbonyl oxygen and morpholino moiety or (iii) possible clashes between the amino acid residues and MMMP in the binding pocket.

**Figure 4.9** shows the interactions that occur for more than **30%** of the simulation time, 78% of the time the protonated amine of SER interacted with ASP<sup>116</sup>, this interaction is primarily hydrogen bonding but contains an ionic interaction component between the charged residues as well. Serine<sup>199</sup> forms a hydrogen bond interactions with the hydroxyl of SER 46% of the time. Moreover, a water bridge is formed between the hydroxyl of SER and ALA<sup>365</sup> (67%), along with  $\pi - \pi$  interactions between the indole ring and the amino acid residues PHE<sup>361</sup> (56%) and PHE<sup>362</sup> (43%).



**Figure 4.9** - Interaction diagram for SERIA-SER for interactions that occur in more than 30% of the simulation are shown.

**Figure 4.10** shows the interactions that occur for more than 5% of the simulation time. The only interaction that has an interaction for more than 30% of the simulation is a hydrogen bond interaction between Serine<sup>199</sup> and the carbonyl oxygen of MMMP, present for 53% of the time. Two interactions between MMMP and binding pocket amino acid residues (ASP<sup>116</sup> and THR<sup>196</sup>) are mediated through a water bridge for 7% of the simulation time. The  $\pi - \pi$  interactions observed are between the phenol residue of MMMP and the amino acids residues PHE<sup>361</sup> and PHE<sup>362</sup>, for 10% and 6% of the time respectively.



*Figure 4.10 - Interaction diagram for SER1A-MMMP for interactions that occur in more than 5% of the simulation are shown.*

Looking at the number of interactions in the SER1A-SER complex, with SER as a known ligand of SER1A, especially the strong interaction occurring between ASP<sup>116</sup> and the protonated amine of SER with a 78% frequency. It would be expected that for an unknown ligand to be able to interact with the binding pocket with similar affinity it would have to form not only the same type of interactions with amino acid residue in the binding pocket but also with similar frequencies. In the case of MMMP there are four interactions that occur with the same amino acid residues in the binding pocket of SER1A when compared to SER of which only the interaction with Serine<sup>199</sup> occurs for more than 50% of the simulation while the other three occur for 10% or less.

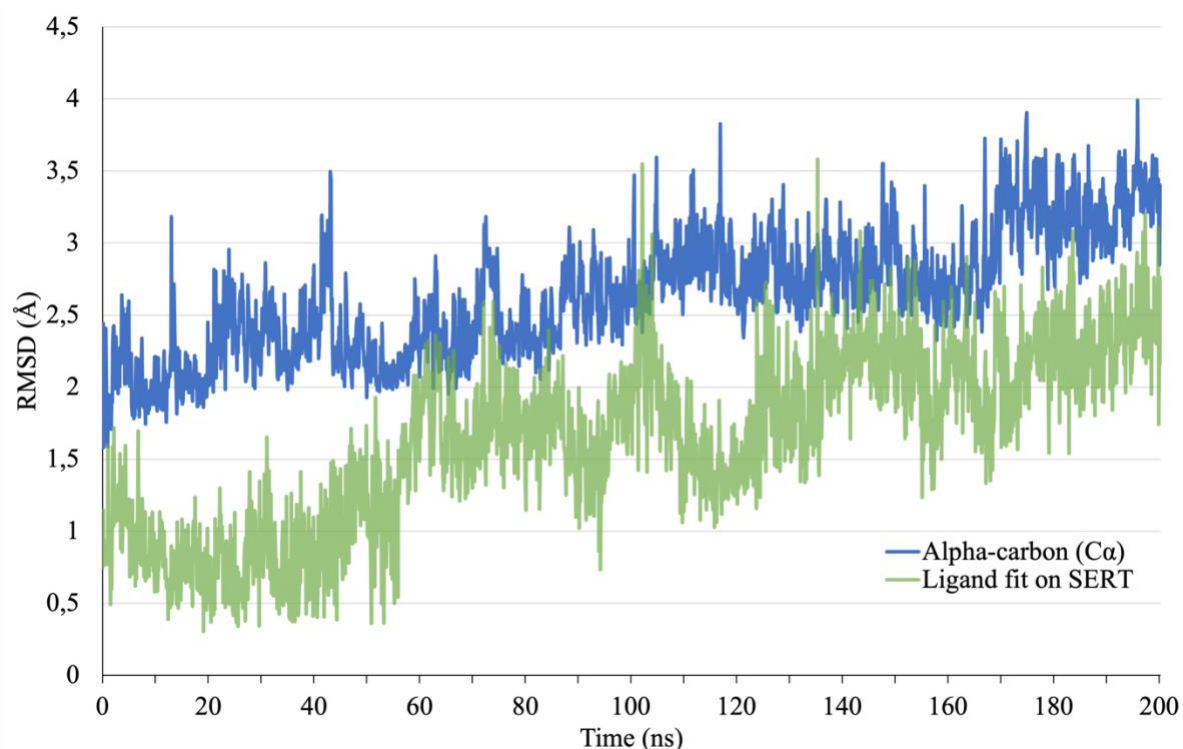
Taking into consideration the number and frequency of interactions in both the SER1A-MMMP complex in comparison to the SER1A-SER complex, and that no large spikes in the RMSD of the ligand is recorded in **Graph 4.4** which indicates that the ligand doesn't diffuse

out of the binding pocket but forms a stable complex around 80 ns. Based on these findings it would be reasonable to conclude that MMMP forms a weak complex with SER1A. Possible reason for the limited interactions between MMMP and SER1A can be attributed to the previously mentioned (i) steric hindrance, (ii) electron dense regions causing electron-electron repulsion or (iii) clashes between amino acid residues and MMMP.

#### 4.5.2 MD analysis of SERT-SER and SERT-MMMP complexes

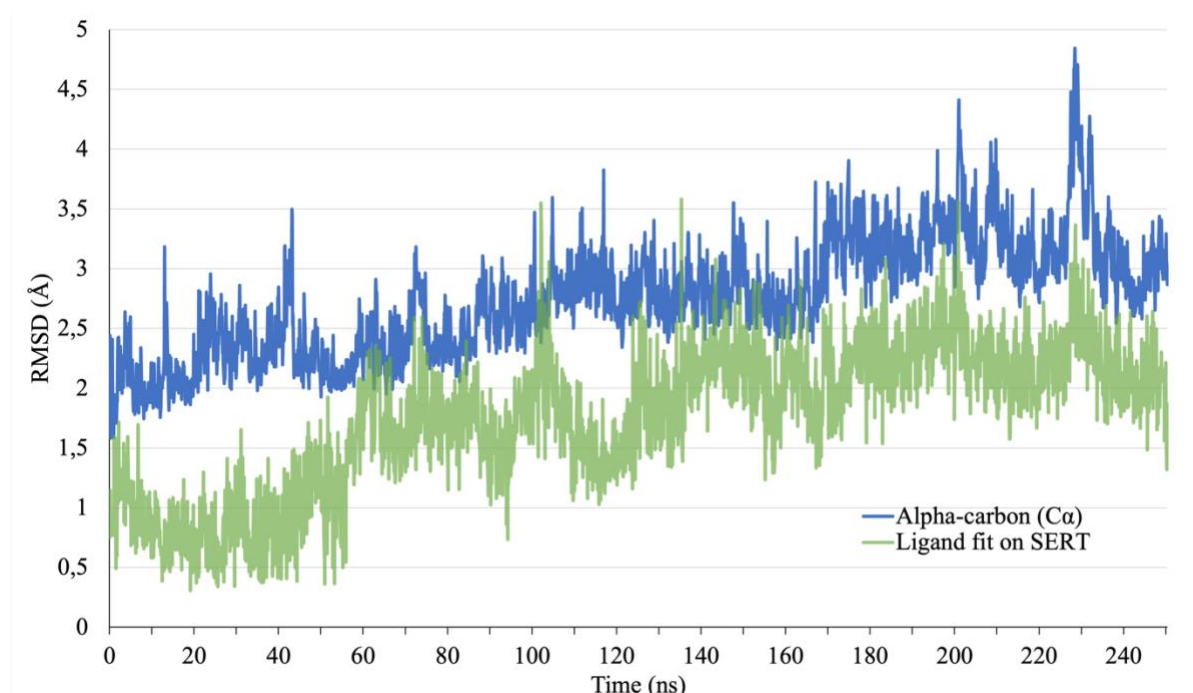
The 200 ns MD simulation plot of the SERT-SER complex shown in **Graph 4.5** reveals that both the C $\alpha$  and the ligand steadily undergoes a change causing the RMSD of both to increase. It was decided that due to the steady increase in RMSD and no noticeable equilibration to run the MD simulation for an additional 50 ns to produce a 250 ns plot shown in **Graph 4.6**. It is only from this plot that we observe equilibration at around 160 ns where both graphs fluctuate between certain values, however a large spike in the protein RMSD is observed around 230 ns indicative of a conformational change in the protein. From 160 ns onward the C $\alpha$  fluctuates between 2.4 to 4.5 Å (excluding the large spike), while the ligand fluctuates between 1.5 to 3.5 Å. Given the relatively small fluctuation it would appear that SERT-SER complex has formed a stable complex. The results of the 200 ns MD simulation of the SERT-SER complex is attached in **Annexure 3**.

**Graph 4.5** - A plot of the ligand's (SER) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SERT during the course of a 200 ns MD simulation at 310 K.



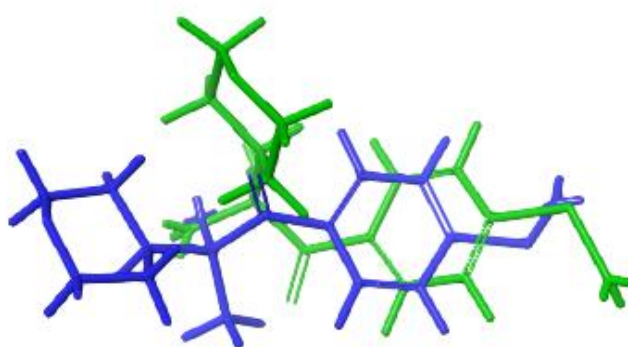


**Graph 4.6** - A plot of the ligand's (SER) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon (C $\alpha$ ) of SERT during the course of a 250 ns MD simulation at 310 K.



The 200 ns MD simulation plot of the SERT-MMMP complex shown in **Graph 4.7** reveals that the C $\alpha$  doesn't undergo a major conformational change and its RMSD ranges between 2 to 3.5 Å while the ligand ranges between 1 to 2.5 Å for the first 47 ns thereafter a jump in RMSD is observed reaching a local top at 50.2 ns with a value of 3.769 Å. For the rest of the simulation the ligand RMSD ranges between 2.5 and 3.5 Å. Converging and equilibration of the C $\alpha$  and the ligand starts to occur at around 100 ns with both fluctuating between 2.5 to 3.5 Å. The results of the 200 ns MD simulation of the SERT-MMMP complex is attached in **Annexure 4**.

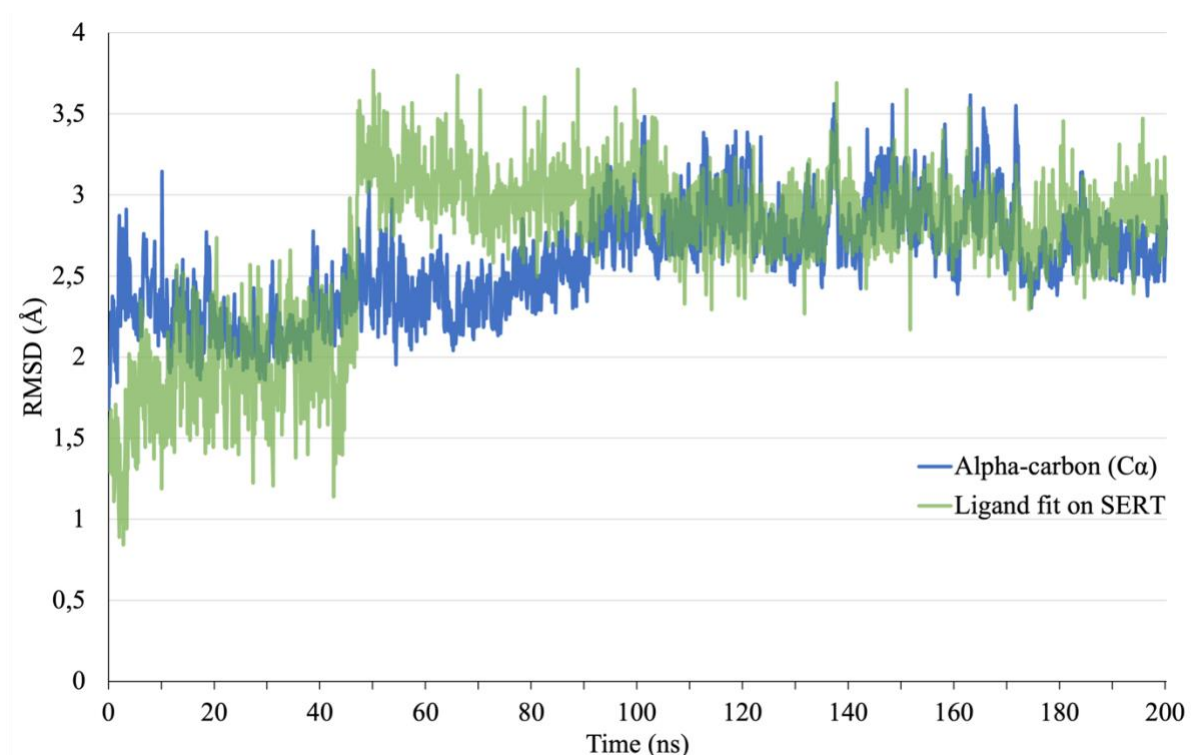
Examination of the MD simulation frames reveals that the large jump in the MMMP RMSD observed from 47 ns onwards is due to the movement of the morpholino moiety to an almost perpendicular position to the methylthiophenyl moiety, considering electron densities and steric hindrance it would be expected that this pose would not be favourable. However, this



**Figure 4.11** - Overlaid image of the reference frame ligand pose (blue) and the ligand pose at 140.0 ns (green). Note the almost perpendicular angle between the morpholino and methylthiophenyl moieties and the pose at 140.0 ns.

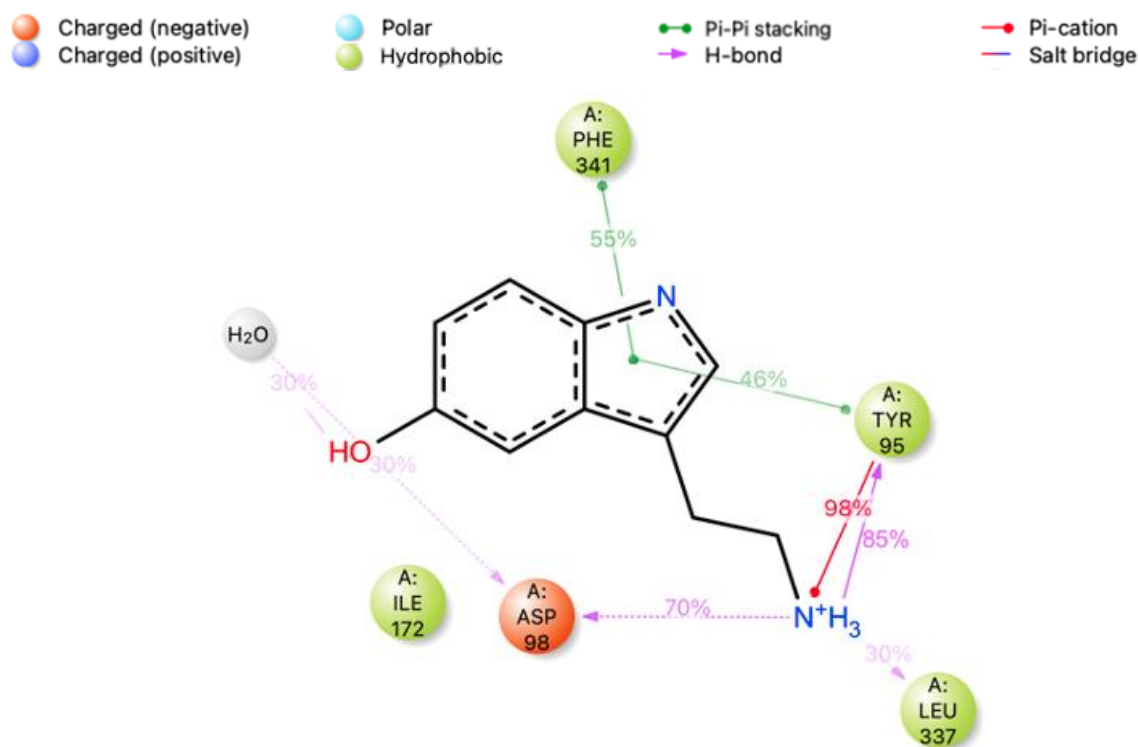
perpendicular ligand pose is stabilised for the rest of the simulation by a  $\pi$ -cation interaction, a water bridge between protonated nitrogen of the morpholino moiety and TYR<sup>95</sup> as well as a  $\pi - \pi$  interaction with PHE<sup>341</sup>. Two  $\pi - \pi$  interaction between the methylthiophenyl moiety and TYR<sup>176</sup> and PHE<sup>341</sup> stabilises the methylthiophenyl moiety. These interactions are shown in the interaction diagram of **Figure 4.13** and the perpendicular pose (Green) overlaid with the original pose (Blue) at the start of the simulation is shown in **Figure 4.11**.

**Graph 4.7** - A plot of the ligand's (MMMP) RMSD with respect to the protein backbone and the RMSD of the alpha-carbon ( $C\alpha$ ) of SERT during the course of a 200 ns MD simulation at 310 K.



The interaction diagram (**Figure 4.12**) of SERT-SER complex shows three prominent interactions that have a frequency of 70% or more. These three interactions are between the protonated amine and two amino acid residues in the binding pocket of SERT, two of which are hydrogen bond interaction and a single  $\pi$ -cation interaction. The first hydrogen bond interaction is with ASP<sup>98</sup> (70%) the second is with TYR<sup>95</sup> (85%) while the  $\pi$ -cation interaction is with TYR<sup>95</sup> (98%) as well. Both TYR<sup>95</sup> and PHE<sup>341</sup> are involved in a  $\pi - \pi$  interactions with the pyrrole ring for 46% and 55% of the simulation respectively. Lastly, a water bridge is observed between the hydroxyl of SER and ASP<sup>98</sup> with a frequency of 30%. Three interactions not shown on the interaction diagram worth mentioning is a hydrogen bond interaction between the protonated amine and Serine<sup>336</sup> (26%), a salt bridge between the protonated amine and ASP<sup>98</sup> (25%) and a hydrogen bond interaction between THR<sup>439</sup> and the hydroxyl of SER

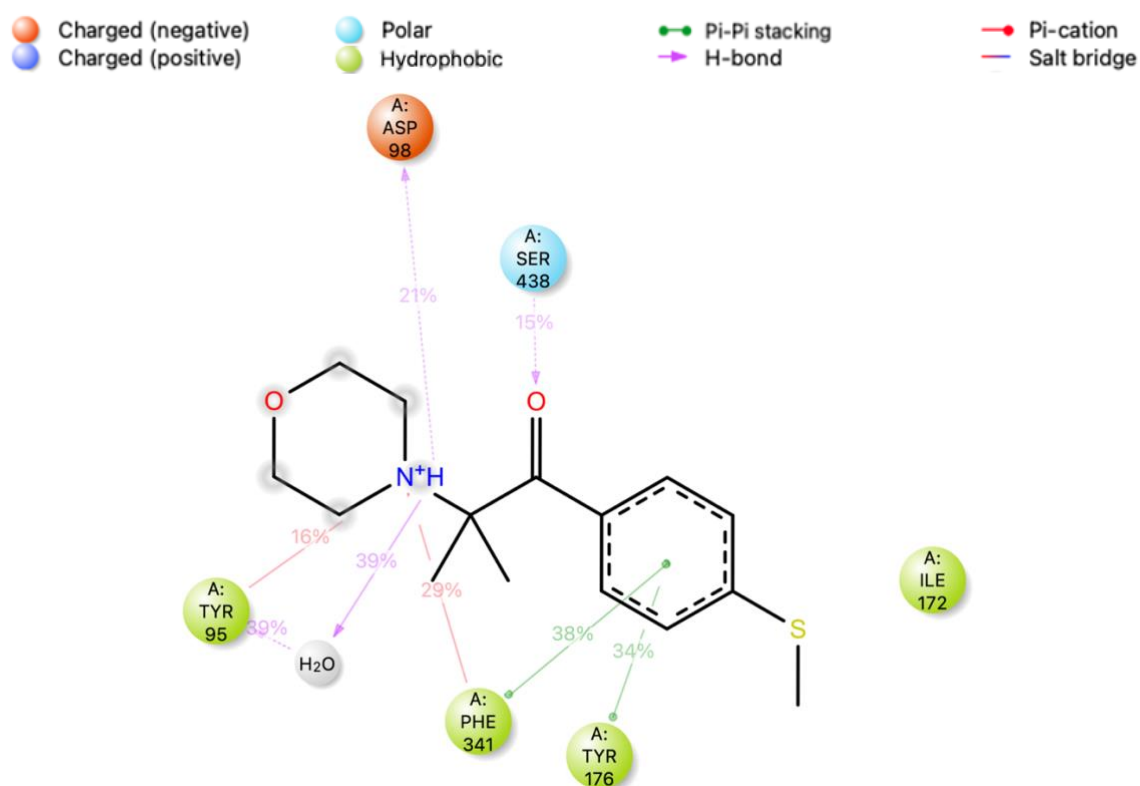
(23%). It would appear that not only ASP<sup>98</sup>, which has been shown to be a highly conserved amino acid residue<sup>[223]</sup>, but also TYR<sup>95</sup> is an important amino acid residue. A fairly recent publication by Yang and Gouaux (2021) mentions two hydrogen bond interactions in particular the first is the ASP<sup>98</sup> bond to the amine of SER and a bond between THR<sup>439</sup> and the hydroxyl group.<sup>[227]</sup> Both of these bonds are present within the simulation.<sup>[227]</sup>



**Figure 4.12** - Interaction diagram for SERT-SER for interactions that occur in more than 30% of the simulation are shown.

The interaction diagram (**Figure 4.13**) for the SERT-MMMP complex reveals quite a lot of interactions occurring within the binding pocket, although none of them occur for more than 40% of the simulation time. ASP<sup>98</sup> and THR<sup>493</sup> have been identified in literature through crystallographic studies as important residues<sup>[223, 227]</sup>, while from the SERT-SER MD simulation TYR<sup>95</sup> was also identified as a possible important amino acid residue. From the SERT-MMMP interaction diagram ASP<sup>98</sup> forms a hydrogen bond interaction with the protonated nitrogen of the morpholino moiety with an interaction frequency of 21%. THR<sup>497</sup> forms a bond with the oxygen of the morpholino moiety but has an interaction frequency of 2% (not shown in **Figure 4.13**). TYR<sup>95</sup> interacts with the protonated nitrogen of the morpholino moiety through two types of interactions, the first is a water bridge and the second a  $\pi$ -cation with interaction frequencies of 39% and 16% respectively. Both PHE<sup>341</sup> and TYR<sup>176</sup> form  $\pi - \pi$  interaction with benzene ring of the methylthiophenyl moiety with interaction frequencies of 38% and 34% respectively, additionally PHE<sup>341</sup> forms a strong  $\pi$ -cation interaction with the

protonated nitrogen of the morpholino moiety with an interaction frequency of 29%. Finally, Serine<sup>438</sup> forms a hydrogen bond interactions with the carbonyl oxygen of MMMP with an interaction frequency of 15%.



*Figure 4.13 - Interaction diagram for SERT-MMMP for interactions that occur in more than 15% of the simulation are shown.*

Based on the two interaction diagrams (**Figure 4.12 and 4.13**) there are quite a few similar interactions that occur in the SERT-MMMP complex when compared to SERT-SER complex. These interactions include a (i) hydrogen bond interactions between ASP<sup>98</sup> and a protonated hydrogen, (ii) hydrogen bond interaction (through a water bridge) and  $\pi$ -cation interaction between TYR<sup>95</sup> and a protonated nitrogen and (iii) the  $\pi - \pi$  interaction between the pyrrole ring of SER and TYR<sup>95</sup> is replaced by a similar  $\pi - \pi$  interaction between TYR<sup>176</sup> and the benzene ring of the methylthiophenyl moiety of MMMP and lastly (iv) a  $\pi - \pi$  interaction between PHE<sup>341</sup> and a ring structure, in SER the pyrrole ring and in MMMP a benzene ring. Given the number of interactions in the SERT-MMMP complex, the number of similar interactions and the stable SERT-MMMP complex formed based on the RMSD fluctuating between 2.5 to 3.5 Å from 100 ns onward in **Graph 4.7**, it would be reasonable to conclude that there is a high probability that MMMP would interact with SERT similar to SER.

#### 4.6 Sub-conclusion

Molecular docking results for 25 protein targets consisting of 24 monoamine receptors and a single monoamine transporter were successfully determined. Out of these 25 protein targets SER1A and SERT were identified as being the most likely to interact with MMMP. These two protein targets scored an absolute difference in docking scores of  $\leq 1$  and were submitted for further evaluation through MD simulations to assess their stability in the binding pockets over the course of a 200 ns simulation.

From the MD simulations it was concluded that MMMP forms a weak complex with SER1A, but strong enough interactions occur to prevent the ligand from diffusing out of the binding pocket for the duration of the simulation. MMMP's interaction with SER1A is not a revelation in terms of a synthetic cathinone (SC) interacting with this receptor as literature have reported SCs such as 3-methylmethcathinone<sup>[254]</sup>,  $\beta$ -naphyrone<sup>[16, 19, 257]</sup>, ethcathinone<sup>[18]</sup>,  $\alpha$ -PVP, MDPV and MDPPP<sup>[19]</sup> also being able to interact, albeit weak affinity or activity, with SER1A.

The MD simulations of the SERT-MMMP have revealed that MMMP has a high probability of interacting with the SERT binding pocket due to the amount of interactions and the frequency with which they occur. From the RMSD graph (**Graph 4.7**) it was confirmed that a stable SERT-MMMP complex is formed that fluctuates between 2 to 3.5 Å. Monoamine transporters are the main target of many SCs with multiple sources reporting on the interaction of SCs with the three transporters, as seen in Chapter 2, **Table 2.4**.<sup>[21]</sup> Unfortunately, no crystal structures for the human dopamine (DAT) or norepinephrine transporter (NET) were available at the time of this study.

It is interesting to note that previous publications that have tested monoamine receptors as possible targets for SCs have reported similar weak to no affinity or activity with these targets, but strong activity at a combination of transporters.<sup>[16, 18, 19, 21, 254, 255, 257]</sup> It is therefore no surprise that SER1A has been identified as a possible weak interaction, but SERT as a more likely interaction.

One shortfall of these docking results, although being positive and indicating possible interactions is that it cannot determine with certainty the nature of these ligand interactions. In the case of the monoamine receptor SER1A the ligand can interact as an agonist or antagonist. Further studies would be required to investigate the nature of this interaction to determine if this possible interaction is activating or inhibiting in nature, as it is not possible to determine this from the docking and MD results. Nonetheless, if MMMP acts as an agonist it has the

ability to activate SER1A which in turn decrease protein kinase A activation through their coupled G-protein  $G_{\alpha i/o}$  subunit and an antagonist prevent the activation of SER1A by its monoamine. In the case of transporter SERT the ligand can act as a blocker preventing the transport of serotonin back into the presynaptic neuron or as a substrate that reverse the normal direction of transporter flux from intracellular to extracellular (transporter-mediated release), disrupt the normal vesicular storage by interacting with vesicular monoamine transporter 2 (VMAT2) integrated into the membranes of synaptic vesicles of presynaptic neurons or influence neurotransmitter synthesis. All these mechanisms of action increases the monoamine concentration of SER in the synaptic cleft which is able to interact with monoamine receptors on the postsynaptic neurons.

As mentioned and discussed in Chapter 3, any docking results should not be taken as a final conclusive result and merely serve as a stepping stone that can aid in identifying and narrowing down targets for further biological studies. Therefore, the work contained in this chapter are purely theoretical and predicts possible interactions that can occur with selected protein targets.

## **Chapter 5: Single crystal X-ray analysis**

### **5.1 Introduction**

From the early works of John Dalton and his atomic theory in the 1800's we know that all substances around us consist of atoms. The arrangement of different types of atoms and their three dimensional orientations forms molecules that defines the overall structure of materials and therefore their nature. Since the properties, characteristics and functions of a material can be directly linked to their molecular structure extensive research over the years have attempted to peak into the atomic level of nature to elucidate molecular structures.

Molecular structures cannot be elucidated with the naked eye due to two limitations, the first being the limited wavelengths observable with the human eye, hence the term visible light (visible to the human unaided eye) and secondly the limited resolution possible with electromagnetic radiation which falls in the visible light spectrum with wavelength of 380 nm to 700 nm which is too long to distinguish between two atoms. Various techniques have been developed over time to elucidate molecular structures including nuclear magnetic resonance (NMR), infrared spectroscopy (IR) and mass spectroscopy (MS). However, all three these techniques have one major disadvantage of producing a list of partial structures which needs to be “stitched” together to produce a complete molecular structure which can sometimes be very difficult. Single crystal X-ray diffraction (SCXRD) analysis on the other hand is capable of producing a complete three-dimensional structure of a molecule.<sup>[395]</sup>

It is for this reason we opted for SCXRD analysis of a crystallised sample of pure 2-methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP) to elucidate the crystal structure, which to our knowledge, has never been reported before. This chapter will serve as a brief introduction to single crystal X-ray crystallography and the principles that makes it possible to elucidate the molecular structure through the use of X-rays.

### **5.2 General overview of X-ray crystallography**

The history of crystallography starts with the discovery of X-rays in 1895 by Wilhelm Conrad Röntgen.<sup>[396, 397]</sup> The first X-ray diffraction pattern was observed by Max von Laue in 1912 from copper sulphate crystals. For his work on X-ray diffraction by crystals he was awarded the Nobel Prize in 1914.<sup>[398-400]</sup> The following year William Lawrence Bragg together with his father William Henry Bragg discovered Bragg's law which makes it possible to calculate the positions of atoms.<sup>[401]</sup> Over the next couple of decades various materials such as inorganic, organic and biological materials were subjected to X-ray crystallography<sup>[402-406]</sup> of which the major discoveries include the discovery of the double helical structure of nucleic acids in DNA

molecules in 1953 by James D. Watson and Francis Crick<sup>[407]</sup> and the elucidation of the insulin molecular structure in 1969 by Dorothy Hodgkin<sup>[408]</sup>.

In order to view three-dimensional structure of molecules quantitatively the following three conditions need to be satisfied:

- (i) Different types of atoms need to be distinguishable.
- (ii) The position of individual atoms inside large molecules can be determined even if the molecule is as large as a protein.
- (iii) The view of molecules can be obtained from many different directions and from these different views the relative position of the molecules can be determined precisely.

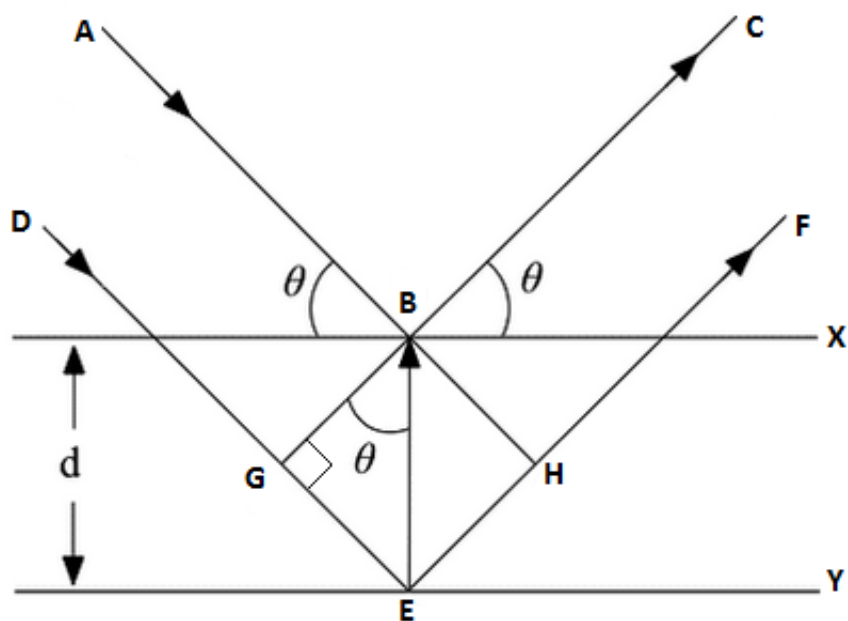
All three of these conditions are satisfied by X-ray crystallography making it the most popular analytical tool to solve the three-dimensional structure of a molecule.

Why are X-rays used? As mentioned earlier visible light has wavelengths ranging between 380 and 700 nm which is visible to the human eye, however sub-microscopic objects such as atoms require shorter wavelengths in order to be observed with adequate resolution. The dimensions of most molecules range between a few to several dozen Å's which is far too small to be observed with visible light. Typical 1Å (0.1 nm) X-rays are used which is in the same order as the radius of an atom and covalent bonds.<sup>[395]</sup>

The basic principle of X-ray crystallography is as follows, parallel X-rays irradiate an object. A shadow image is projected on the detector perpendicular to the X-rays, while at the same time most of the X-rays are reflected similar to visible light in a light microscope. These X-rays are known as scattered X-rays. Scattered X-rays induce the phenomenon of coherent interference, because X-rays can be thought of as waves which can interact with one another constructively or destructively resulting in X-ray diffraction.<sup>[395]</sup> Here Bragg's law, which makes it possible to determine the position of atoms comes into play, states the following: *"When the X-ray is incident onto a crystal surface, its angle of incidence,  $\theta$ , will reflect with the same angle of scattering,  $\theta$ , and when the path difference,  $d$ , is equal to a whole number,  $n$ , of wavelength,  $\lambda$ , constructive interference will occur."*<sup>[409]</sup> **Equation (Eq.) (13)** below is Bragg's law and is illustrated in **Figure 5.1**.

$$n\lambda = 2d \sin\theta \quad (13)$$





*Figure 5.1 - Diagram of Bragg's Law.*

When crystals are irradiated, the detector observes diffraction spots to a specific direction. These diffraction spots carry contributions from all the atoms in the original crystal and therefore it is necessary to collect all diffraction spots from various angles in order to recreate the real image. The information such as the intensities and the location of the diffraction spots are recorded by the detector.

In previous years a scintillation counter was used to measure the X-ray diffraction but could only collect one diffraction at a time, the development of two-dimensional detectors allowed for multiple diffraction spots to be collected at a time which significantly reduced measurement times. Based on the intensities and positions of the collected diffraction patterns, electron density distributions in the crystal can be calculated, even though these calculations can be very difficult recent development in software has made this process effortless. From the electron density distribution we assume that atoms are present at the electron density maxima and therefore the atomic coordinates or atomic positions of atoms are determined which results in an elucidated molecular model.<sup>[395]</sup> This process is summarised and shown in **Figure 5.2**.

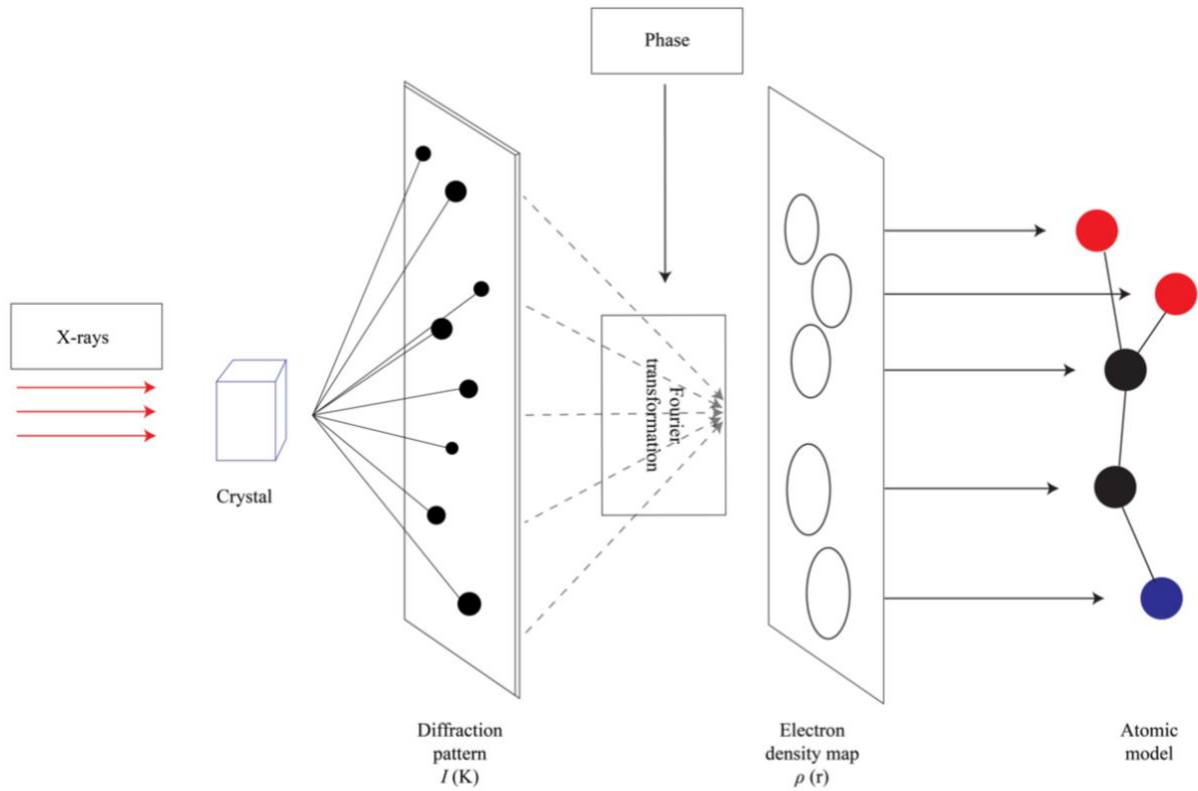


Figure 5.2 - Summary of the technique used in X-ray crystallography.<sup>[395]</sup>

### 5.3 Principles of X-ray crystallography

There are three fundamental facts that need to be understood in order to understand the principles of X-ray crystallography: Firstly, electron density,  $\rho$ , within materials (such as crystals) are responsible for X-ray scattering. The scattering amplitude is equal to the summation of the scattering wave from each electron and the scattering intensity is proportional to the square of the scattering amplitude. Secondly, the electron density  $\rho$  within a crystal has a periodicity, which is virtually an infinite repetition in three independent directions. Thirdly, all materials consist of atoms, these atoms each have a constant electron density even in different materials.

#### 5.3.1 Derived equations based on the fundamental facts

From these three facts, the following three simplified equations can be derived:

$$|F(K)|^2 = I(K) \quad (14)$$

$$F(K) = \int_V \rho(r) \exp[2\pi i(K \cdot r)] dv_r \quad (15)$$

$$\rho(r) = \frac{1}{V} \int_{V_k} F(K) \exp[-2\pi i(K \cdot r)] dv_K \quad (16)$$

Where  $I(K)$  is the diffraction pattern and  $r$  is the position of all atoms contained in a unit cell. In **Eq. (14)**  $I(K)$  can be obtained directly from the diffraction measurements and  $F(K)$  is known as the structure factor. The structure factor is directly involved in the atomic positions. **Eq. (15)** shows that the structure factor can be determined from the electron density ( $\rho$ ) and **Eq. (16)** shows that the electron density  $\rho$  can be obtained from the structure factors. This electron density is used to derive the atomic position within molecular structures.<sup>[395]</sup>

### 5.3.2 The Phase Problem

X-rays, like with all waves, have an amplitude and phase associated with them, however during a diffraction analysis of a molecule we are only able to measure the amplitude of a reflection, but not its phase. It is of vital importance to determine the phase of each reflection in order to sum the waves together correctly.

In **Eq. (14)** only the absolute value of  $F(K)$  is given, but in reality  $F(K)$  is defined as a complex number.  $F(K) = |F(K)| e^{i\varphi(K)}$ , where  $\varphi(K)$  is the phase as represented in **Eq. (15)**. However, in **Eq. (16)** the electron density cannot be determined readily due to the loss of phase information in the diffraction experiment. This is known as the “phase problem” and a large part of X-ray crystallography is spent finding the correct phases.<sup>[395, 410]</sup>

Currently a few methods exist to solve the phase problem in order to determine the phases; the method applied is dependent on the characteristics of the crystal. These methods include Direct method<sup>[411]</sup>, Patterson function<sup>[412, 413]</sup>, Patterson search method<sup>[414]</sup>, Dual-space method<sup>[415, 416]</sup>, Intrinsic phasing<sup>[417]</sup> and Isomorphous replacement method<sup>[418, 419]</sup>.

**Direct methods** statistically estimate the phases by using the amplitudes of the normalised structure factor and are uniquely employed to solve the phase problem for small molecule crystals. The **Patterson function method** uses the square of the structure factor, that are intensities, as Fourier coefficients to eliminate phases as the phase problem arises from using the structure factors in the first place. This method is typically used to solve structures that contain heavy atoms. The **Patterson search method** uses the knowledge of similar structures to solve the current structure. A suitable model is obtained through searching databases of solved crystal and NMR structures. This search model which is used as a template to solve the current crystal structure is first translated and then rotated to obtain a maximum fit between the calculated and observed diffraction data. If the correct orientation and position is found, the template can then serve as a starting point for the current model rebuilding and refinement.

With **dual-space method** the phases of the structure factors are retrieved by iterative modifications of the current model in direct space (electron density) and reciprocal space (structure factors). **Intrinsic phasing** plays to the strengths of direct methods, which can best solve structures that belong to the *P1* space group. With the Laue group known, equivalent intensities are averaged before the data is expanded to *P1*. Intrinsic phasing departs from classical direct methods by starting the structure solution process from a Patterson superposition minimum function and not from random phases. Lastly, the heavy atom **isomorphous replacement method** is generally used to solve the phase problem with crystal proteins. The key to successful phasing with any of the above-mentioned methods is to collect accurate diffraction data with the greatest resolution possible.

## 5.4 The basic procedure of crystallography

### 5.4.1 Crystallisation

This is the first and often the most complicated step in crystallography, all results obtained depends on the quality of the crystal. Obtaining a good quality crystal, diffracts well and assures good results. This process usually involves a trial and error approach until a satisfactory crystal is obtained. In previous years a large crystal was required for diffraction analysis, but due to the advancements in the field small crystals with the dimensions 0.05 mm<sup>3</sup> or smaller are becoming measurable with in-house systems. In addition, cube shaped crystals are preferred, but due to the development of robust absorption correction software plate-like or needle-like crystals can also be analysed.

### 5.4.2 Data collection

In the second step the best crystals are selected usually with the aid of a microscope before being a single crystal is mounted and placed in the X-ray diffractometer to collect diffraction data. The crystal is mounted in the path of the X-ray beam and the intensity of all the diffraction spots are recorded for every angle of the crystal. Low quality crystals produce low quality results which makes the next steps of solving the crystal structure particularly difficult. In cases where poor diffraction data is collected, problematic crystals should not be used. Alternatives are to use a different crystal from the original crystal sample, regrowing crystals in different solvents, changing conditions and/or environments with the goal of obtaining better crystals for analysis. Nearly 80% of X-ray structure analysis can be considered finished if good quality crystals are used and good diffraction data is obtained from them.<sup>[395, 420]</sup>

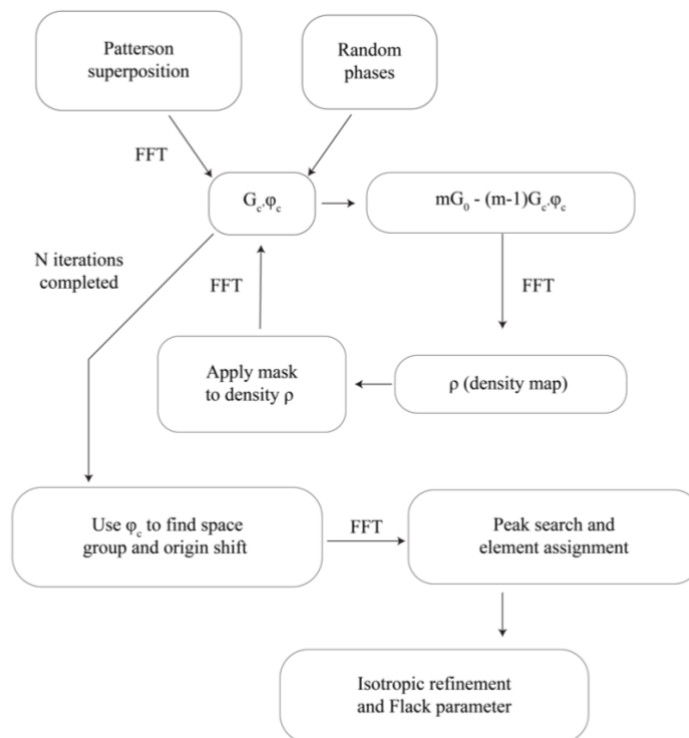
### 5.4.3 Solving a crystal structure

After the diffraction data is collected from the crystal the information is passed onto the structural analysis program to finally obtain the atomic level structure. There is a common misconception that only the intensity data is required for the structural analysis, but the information such as the chemical composition of the molecule is also required. For molecules where the chemical composition is unknown it becomes difficult to conclude the crystal structure. Therefore, X-ray crystallography is better at solving the crystal structure than identifying the type of atoms within the structure as sometimes atoms are misidentified which needs to be corrected at a later stage during crystal structure refinement.

Software packages such as WinGX<sup>[421]</sup> and OLEX2<sup>[422]</sup> which employ SHELXT<sup>[417]</sup> to solve and refine crystal structures against the X-ray diffraction data can be used. These programs perform the difficult calculations discussed in **section 5.3** allowing users to solve the structure like a puzzle. The collected diffraction data is combined computationally along with the complimentary chemical information, such as the chemical composition, to produce an electron density map to which atoms are assigned to. Repeated phase refinement and model fitting produces a final refined atomic model.

SHELXT<sup>[417, 423]</sup> is a computer program that uses intrinsic phasing to solve the phase problem for single-crystal reflection data in order to determine the space group and the molecular structure. **Figure 5.3** below summarises the steps followed by SHELXT to solve and determine the crystal structure.

As previously mentioned intrinsic phasing plays to the strengths of direct methods, which can best solve structures that belong to the *PI* space group. With the Laue group known, equivalent intensities are averaged before the data is expanded to *PI*. The initial phases from the structure solution is then used to determine the appropriate space group. The space group provides the symmetry information for averaging phases to calculate improved electron densities; furthermore, dual-space recycling is then performed to improve the quality of the electron density model. Dual-space recycling uses random omit-maps which entails randomly omitting a certain percentage of the peaks and then calculating the phases based on the remaining atoms. Intrinsic phasing also implements the free-lunch algorithm that uses the density modification to calculate phases for reflections that have not been measured, thus completing the data to a given resolution. Next atoms are assigned to the density's maxima.<sup>[417]</sup> The intrinsic phasing method of solving the phase problem has been shown to be less demanding on data quality and completeness compared to other direct methods.



**Figure 5.3** - Summary of the procedure followed by SHELXT when solving a crystal structure. The intrinsic phasing structure solution, the space group refinement and isotropic refinement are performed in parallel.  $G_0$  and  $G_c$  are the modified observed and calculated structure factors and  $\varphi_c$  is the phase of  $G_c$ . FFT is the Fast Fourier transform.<sup>[417]</sup>

#### 5.4.4 Analysis of the final crystal structure

The final model from the refinement provides information on the atomic coordinates and temperature factors for each atom. From this information bond angles, torsion angles and distances can easily be calculated and the molecular model and Ortep diagrams can be drawn. The temperature factors describes the anisotropic vibrations of atoms and is also known as the thermal vibration parameter. These final models can be analysed and viewed through software such as Mercury.<sup>[424]</sup>

### 5.5 Basic theory of crystallography

In this section the basic theory of crystallography used to describe crystal structures will be discussed. This section was written through the extensive use of the book “Introduction to crystallography” by Frank Hoffmann.<sup>[425]</sup>

### 5.5.1 The unit cell

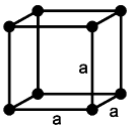

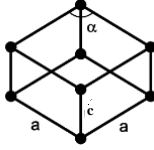
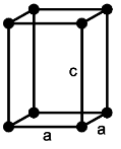
The unit cell is described as the smallest building block of any crystal which consists of identically placed atoms or molecules such that when the unit cells are stacked together in all three spatial directions the macro crystal structure is observed.

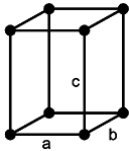
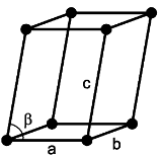
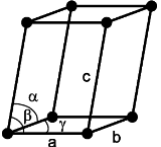
Characterisation of the unit cell firstly occurs through its lattice parameters or metrics. Six parameters are required to unequivocally describe the unit cell, these six parameters consists of three edge lengths  $a$ ,  $b$  and  $c$  which per definition run along the  $x$ ,  $y$  and  $z$  axis respectively and three angles  $\alpha$ ,  $\beta$  and  $\gamma$ . Additionally a unit cell should contain all the symmetry elements (discussed later) of the crystal, therefore when a unit cell is translated along all three spatial directions and joined together no additional symmetry elements should appear.

### 5.5.2 The Seven Crystal Systems

From the six lattice parameters the number of principally different unit cell shapes are very limited arising to only seven unique crystal systems into which all known crystals can be classified into. These seven crystal systems are known as cubic, hexagonal, trigonal (rhombohedral), tetragonal, orthorhombic, monoclinic and triclinic which differ with respect to their lattice parameters. This classification scheme is based on their symmetry and not the metric itself leading to certain values for the parameters. The parameters for each crystal system is summarised in **Table 5.1** below.

**Table 5.1** - The Seven Primitive Crystal Systems.

| Crystal system          | Crystal system figure   | Restrictions   |   |
|-------------------------|---|----------------|---|
|                         |   | Length of axes | Angles of the cell                              |
| Cubic                   |  | $a = b = c$    | $\alpha = \beta = \gamma = 90^\circ$            |
| Hexagonal               |  | $a = b$        | $\alpha = \beta = 90^\circ, \gamma = 120^\circ$ |
| Trigonal (Rhombohedral) |  | $a = b$        | $\alpha = \beta = 90^\circ, \gamma = 120^\circ$ |
| Tetragonal              |  | $a = b$        | $\alpha = \beta = \gamma = 90^\circ$            |

|              |   |                   |                                      |
|--------------|---|-------------------|--------------------------------------|
| Orthorhombic |  | None <sup>a</sup> | $\alpha = \beta = \gamma = 90^\circ$ |
| Monoclinic   |  | None <sup>a</sup> | $\alpha = \gamma = 90^\circ$         |
| Triclinic    |  | None <sup>a</sup> | None <sup>a</sup>                    |

<sup>a</sup>Parameter can have any conceivable value

### 5.5.3 The 14 Bravais lattices

From the seven crystal systems we only get primitive unit cells (lattice types), which only have lattice points at the corners of the unit cell, but none in the centre of the unit cell nor on the edges or in the centre of the faces. Through the addition of additional lattice points (known as centring) in the centre or the centre of faces of unit cells gives rise to seven additional lattice types. The seven primitive lattice types and the seven additional lattice types through centring creates a total of 14 lattice types. These 14 lattice types are known as the 14 Bravais lattices.

The seven additional lattices are obtained through four types of centring, in the first type of centring known as *single-sided face centring* (C) there is a single lattice point at the centre of one face of the unit cell. The second type of centring is known as *body-centred centring* (I) in which an additional lattice point is observed exactly in the centre of the unit cell. The third type of centring additional lattice points are observed at all faces but not inside the unit cell this is known as *all-sided face centring* (F). The final centring type is known as *rhomboidal centring* (R) which is a special case of body centring which have lattice points on opposing faces, on one face the lattice point is located at the centre  $\frac{1}{3}$  from one of the edges and on the opposing face the lattice point is located at the centre  $\frac{2}{3}$  from the equivalent edge. The primitive and the first three types of centring are shown in **Figure 5.4** below while the 14 Bravais lattices are summarised in **Figure 5.5**.



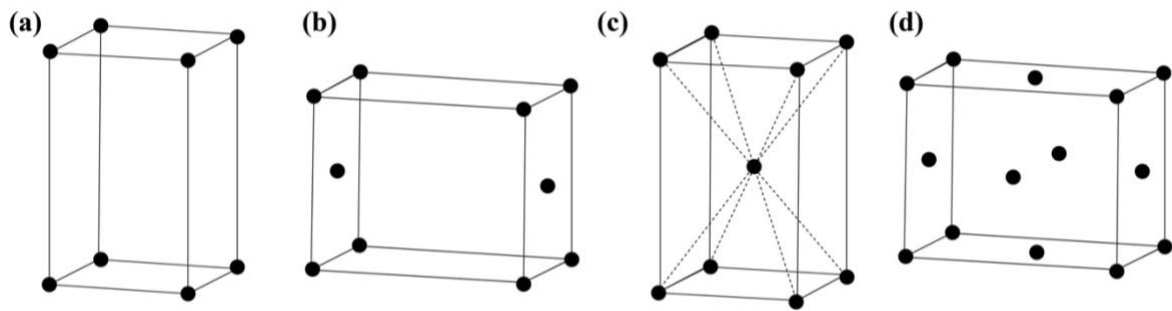


Figure 5.4 - Overview of the different types of centring. (a) Primitive, (b) Single-sided face centring, (c) Body-centred centring and (d) All-sided face centring.

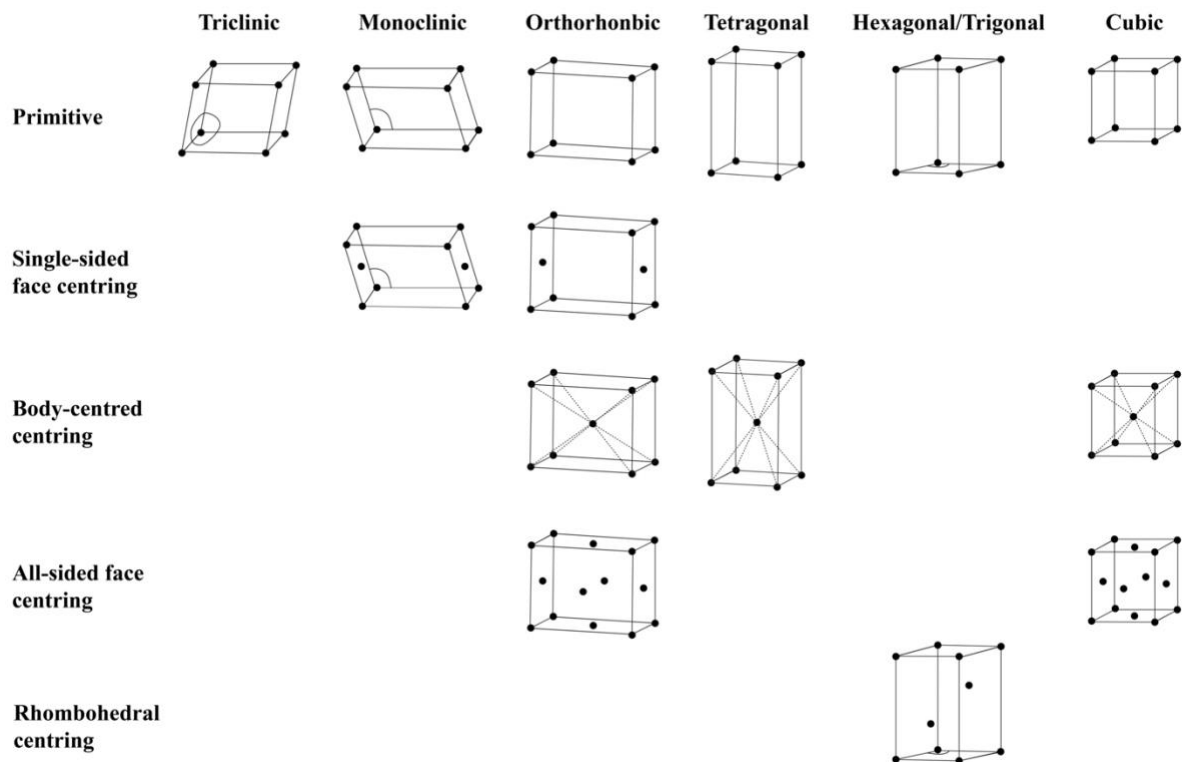


Figure 5.5 - The 14 Bravais lattices arising from centring.

#### 5.5.4 The 32 Point Groups

Symmetry operation is a geometric transformation that maps an object onto itself, that may lead to an indistinguishable configuration. It is important to note that this operation is carried out imaginarily. While a symmetry element is a geometric object on which the operation is carried out which can be a point, plane or a line. Point symmetry refers to a symmetry operation being carried out in which at least one point of the object is not moving. If we solely consider the external symmetry of macroscopic objects there are only five point symmetry elements on which symmetry operations can be performed. These are known as identity, mirror plane, axis of rotation, centre or inversion and rotoinversion axis.

**Identity symmetry ( $E$ ).** All objects have at least one element of symmetry, even very asymmetric ones, this is also known as a one-fold axis of rotation. To better understand this if a  $360^\circ$  rotation (symmetry operation) is performed on any object the final state of the object will be indistinguishable from the original state, therefore has an identity symmetry element.

**Mirror symmetry ( $\sigma$ )** in an object means that a mirror line/plane is present and that reflection around the line/plane leaves the object unchanged.

**Rotational symmetry ( $C_n$ ,  $n = 1, 2, 3, \dots, \infty$ )** around an axis implies that a rotation operation leaves the final state indistinguishable from the original, the amount of degrees with which an object can be rotated depends on the orders or rotational symmetry. Thus, a 2-fold axis of rotation means that a  $180^\circ$  will produce an indistinguishable final state, while a 3-fold axis of rotation means that every  $120^\circ$  rotation will yield a state similar to the original state. There are examples of object that have an infinite fold axis of rotation (i.e. a wine glass).

**Centrosymmetry ( $i$ )** is a symmetry element that occur in an object that contains a centre of inversion, they are constructed in such a way that there are two corresponding parts which are equidistant from the symmetry centre, but are located in exactly opposite directions of space.

**Rotoinversions axis ( $S_n$ )** is a symmetry element in which two operations are performed one after the other, the first operation is a rotation by  $360^\circ/n$ , where  $n$  is the order of the rotoinversion axis and the second operation is an inversion through a point.

From the five point symmetry elements we are able to classify crystals with respect to their outer shape through their point symmetry. 32 Unique point groups arises from these symmetry elements of which all crystals can only be assigned to a single one of the 32 possible point groups. The specifics of the 32 point groups are beyond the scope of this study, but can be found in almost all crystallography literature including textbooks.

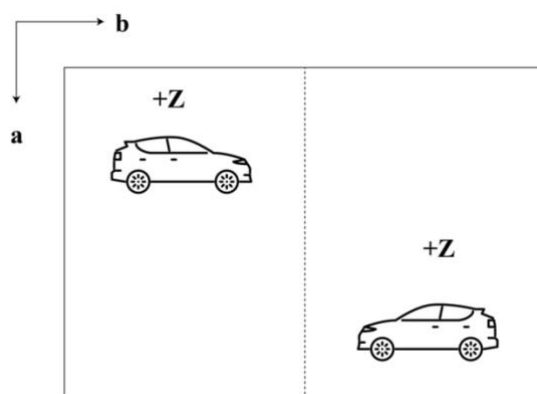
#### 5.5.5 Symmetry in the plane

There are three types of symmetry that is required to completely describe the symmetry of patterns in a plane of which all contain a translational component.

- (i) Pure translations
- (ii) Glide planes
- (iii) Screw axis

Pure translations are translations that occur by whole units of a primitive unit cell, in other words from lattice point to an equivalent lattice point in an adjacent unit cell. Glide planes and screw axes on the other hand describe translations that have translational components smaller than an entire unit cell; however, the symmetry operations do not exclusively consist of translations. Glide planes can occur in both a two-dimensional and three-dimensional periodicity while screw axes require a three-dimensional periodicity.

A glide plane is a symmetry element on which a glide reflection operation is performed, glide reflections are coupled symmetry operations in which two operations are carried out one after the other. The first operation is a reflection thereafter a translation usually by half of the unit cell dimension (translations of a quarter unit cell among other are possible as well). There is a total of six glide planes, namely *a*, *b*, *c*, *n*, *d* and *e* where the translation along *a*, *b* and *c* takes place along the x-, y- and z-axis respectively. Glide planes *n*, *d* and *e* are more complicated and will not be dealt with in this study. **Figure 5.6** below illustrates glide plane *a* along the x-axis.



*Figure 5.6 - Glide plane a orientated perpendicular to the drawing plane with a car motif. The car is reflected, thereafter translated by  $\frac{1}{2}a$  unit cell in the a direction.*

Screw axis is a symmetry element on which a screw rotation operation is performed, like with glide reflection operation, screw rotation operation is a coupled symmetry operation as well. The first operation is a rotation of a certain amount of degrees around an axis followed by a translation parallel to the axis. A screw axis has the following format:  $n_m$ , where *m* and *n* are always integer values and *n* is always larger than *m*. This format describes the corresponding operation in which an object first gets rotated by  $360^\circ/n$  around an axis, thereafter it gets translated by a value  $\frac{m}{n}$  units of an entire unit cell length. Per definition the rotation should always be carried out in the sense of a right-handed coordinate system in which the x-axis gets rotated towards the y-axis so that the translation is always parallel to the z-axis (i.e. perpendicular to the xy plane). A  $3_2$  screw axis means that the rotation operation of a  $120^\circ$  ( $360^\circ/3$ ) occurs before a translation of  $\frac{2}{3}$  perpendicular to the screw axis occur. The screw axes that can occur in crystal structures are two- ( $2_1$ ), three- ( $3_1$ ), four- ( $4_1$ ,  $4_2$ ) and six-fold screw axes ( $6_1$ ,  $6_2$ ,  $6_3$ ). Below **Figure 5.7** illustrates a  $3_1$  screw axis.

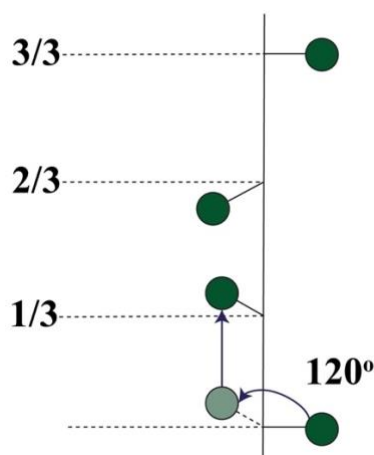


Figure 5.7 -  $3_1$  Screw axis illustration.

### 5.5.6 Space groups

Space groups are the final piece of the proverbial puzzle that ties all the discussed elements in this section together. The combination of the 14 Bravais lattices, 32 point groups (arising from mirroring, rotation, inversion and rotoinversion) used to describe the outer shapes of crystals, pure translational symmetry together with the symmetry elements with translational components (glide planes and screw axes) produces 230 different arrangements known as space groups to which crystals can be assigned to. The term space group can be defined as a set of symmetry elements together with their respective symmetry operations which completely describes the spatial arrangement of a three dimensional periodic pattern.

## 5.6 Conclusion

Single crystal X-ray analysis uses the phenomena of interference and diffraction of X-rays which arises from scattering by the electrons within atoms. Furthermore, the diffraction intensity,  $I$ , is used to calculate the electron density,  $\rho$ , which reveals the three-dimensional arrangement of atoms or molecules (within a unit cell).

This chapter serves as a brief introduction to the theory and techniques used in SCXRD technique. It is without a doubt that single crystal X-ray analysis is a very powerful tool that in past century has developed into a very useful technique. Not only does this technique solve the molecular structures of an unknown molecules but can also determine the specific chirality and the molecular arrangements of molecules within a regular repeating unit know as a unit cell. Developments made in hardware and software of X-ray crystallography has made this technique easy to use by even the most inexperienced users with very little knowledge on crystallisation, data collection and solving the crystal structure.

## Chapter 6: Crystallographic investigation of 2-methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP)

### 6.1 Methodology

Crystals for single crystal X-ray diffraction (SCXRD) analysis were grown from a 0.164 M solution containing 98% pure 2-methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP) and 99.8% HPLC grade ethanol both obtained from Sigma-Aldrich®. The solution was stored in a glass vial sealed with perforated parafilm®, evaporation of the solvent occurred at a constant temperature-controlled laboratory of 21 °C until the crystals started to form. Once most of the solvent evaporated the vial was completely sealed to prevent further evaporation and complete drying of the crystals.

A single blade shaped colourless crystal was selected and mounted on suitable support. Data were collected at the University of Pretoria, South Africa, using a Rigaku XtaLAB Synergy R, DW system with a HyPix diffractometer operating at 152.0(2) K.

Data were measured by employing  $\omega$  scans of 0.5° per frame for 0.1 s using monochromatic Cu K $\alpha$  radiation with a wavelength ( $\lambda$ ) of 1.54184 Å. The diffraction pattern was indexed and CrysAlisPro<sup>[426]</sup> was used to determine the strategy for the number of runs and images. A maximum resolution of 72.11° (0.81 Å) was achieved. The unit cell was refined using CrysAlisPro on 23055 (71%) of the total 32547 observed reflections. Data reduction, absorption and scaling corrections were performed using CrysAlisPro. A multi-scan absorption correction was performed using spherical harmonics implemented in SCALE3 ABSPACK scaling algorithm.<sup>[427]</sup>

The structure was solved by using direct methods with SHELXT<sup>[417]</sup>, as part of the OLEX2<sup>[422]</sup> suite. Structure refinement was carried out using SHELXT in OLEX2. Non-hydrogen atoms were all refined anisotropically and all hydrogens were placed in idealised positions based on their respective carbon atoms. The structure was evaluated for possible twinning. A twin refinement was done which yielded a BASF factor of 0, indicating that no twinning was present. None of the data pointed to the possible presence of non-merohedral twinning and merohedral twinning is only possible in trigonal, tetragonal, hexagonal and cubic crystal systems.<sup>[428]</sup> No bad reflections were observed or omitted from this dataset. All the crystal data, intensity data collection and structure refinement details are summarised in **Table 6.1** below. The final .cif file of the crystal structure was viewed and analysed using Mercury version 4.2.0 software.<sup>[424]</sup>

## Results and Discussion

### 6.2 Discussion of the overall parameters of the crystal structure

MMMP crystallises in the orthorhombic non-centrosymmetric space group  $Pca2_1$ . The asymmetric unit contains four crystallographically independent molecules ( $Z = 4$  and  $Z' = 1$ ) of MMMP as shown in **Figure 6.1** below. The unit cell has the following dimensions  $a = 37.0231(3) \text{ \AA}$ ,  $b = 6.5766(1) \text{ \AA}$  and  $c = 24.1761(3) \text{ \AA}$  with angles  $\alpha = \beta = \gamma = 90^\circ$  and a total volume of  $5886.54(12) \text{ \AA}^3$ .

The calculated Flack parameter of  $0.302(6)$  is high due to the achiral space group,  $Pca2_1$ , in which MMMP crystallises in. The Flack parameter is solely calculated for centrosymmetric space groups, this parameter is therefore not applicable as MMMP crystallises in a non-centrosymmetric space group. The  $R_{\text{int}}$  value which serves as a measurement of the precision generally should be less than  $0.10$ , a value of  $0.0405$  were reported indicative of correct integration strategy and unit cell assignment.

Furthermore, the goodness-of-fit (GooF) parameter was calculated to be  $1.0475$  very close to converging to a value of  $1.0$ . The final  $wR2$  (all data) was calculated to be  $0.1261$  and a  $R$ -factor ( $R1$ ) of  $0.0466$  (or  $4.66\%$ ). These three parameters indicate an agreement between the observed ( $F_o$ ) and calculated ( $F_c$ ) electron density from the solved model, therefore a valid solution has been found.

**Table 6.1** - Crystal data for MMMP crystal structure.

| Parameter                                 | Value                   | Parameter                                 | Value         |
|---|-------------------------|---|---------------|
| Formula of unit cell                      | $C_{60}H_{84}N_4O_8S_4$ | $V (\text{\AA}^3)$                        | $5886.54(12)$ |
| Formula Weight ( $\text{g.mol}^{-1}$ )    | 1117.622                | $Z$                                       | 4             |
| Formula of single molecule of MMMP        | $C_{15}H_{21}NO_2S$     | $Z'$                                      | 1             |
| Formula Weight ( $\text{g.mol}^{-1}$ )    | 279.39                  | Wavelength ( $\text{\AA}$ )               | 1.54184       |
| $D_{\text{calc.}}$ ( $\text{g.cm}^{-3}$ ) | 1.261                   | Radiation type                            | Cu $K\alpha$  |
| $\mu$ ( $\text{mm}^{-1}$ )                | 1.934                   | $\theta$ range ( $^\circ$ )               | 3.01 to 72.11 |
| Colour                                    | Colourless              | Measured Reflections                      | 32547         |
| Shape                                     | Blade                   | Independent Reflections                   | 9767          |
| Size ( $\text{mm}^3$ )                    | 0.22 x 0.18 x 0.08      | Reflections with $I \geq \sigma(I)$       | 9312          |
| T (K)                                     | 152.0(2)                | Completeness to theta = $72.11^\circ$ (%) | 99.89         |
| Crystal System                            | Orthorhombic            | $R_{\text{int}}$                          | 0.0405        |

|                 |                         |   |                    |
|-----------------|-------------------------|---|--------------------|
| Flack Parameter | 0.302(6)                | Parameters                                | 697                |
| Hooft Parameter | 0.302(6)                | Restraints                                | 1                  |
| Space Group     | <i>Pca2<sub>1</sub></i> | Largest diffraction peak and deepest hole | 0.7245 and -0.2797 |
| a (Å)           | 37.0231(3)              | Goodness-of-fit (GooF)                    | 1.0475             |
| b (Å)           | 6.5766(1)               | wR2 (all data)                            | 0.1261             |
| c (Å)           | 24.1761(3)              | wR2                                       | 0.1233             |
| α (°)           | 90                      | R1 (all data)                             | 0.0490             |
| β (°)           | 90                      | R1  | 0.0466             |
| γ (°)           | 90                      |   |                    |

Note: Crystal data report is attached in **Annexure 5**.

### 6.3 CIF check

Analysis of the final crystal structure was done using *checkCIF* available directly from the International Union of Crystallography (IUCr) website. A few problems were identified of various importance. The most serious problem PLAT029 (Alert level A) to “check the reported diffraction measured fraction theta full” can be explained by the data being collected on a relatively small and weakly diffracting crystal. The value should be close to 1.0 but the recorded value was 0.845. Other possible reasons for this error could be that reflections at high angles may not be detected at all since a Cu K $\alpha$  source was utilised on a weak diffracting crystal or due to the missing reflections obscured by the beamstop and therefore, these reflections were not observed. Alerts that fall in the C and G category were considered for possible improvements, but none of the alerts could be improved. The *checkCIF* document can be found in **Annexure 6**.

### 6.4 Nomenclature

#### 6.4.1 Nomenclature of MMMP conformers

As mentioned above the unit cell consists of four crystallographically independent conformers shown in **Figure 6.2**. To describe the four conformers, the thioether- and carbonyl- moieties, as circled in **Figure 6.1**, middle, are seen as essentially planar. These groups are *anti* from one another in molecules 1 and 3, and *syn* from one another in molecules 2 and 4. When considering the C<sub>1</sub>–C<sub>2</sub> bond as central and viewing the molecule down the central bond as indicated in **Figure 6.1**, top right, further conformational differences can be described. The C<sub>1</sub>-carbonyl group and the C<sub>2</sub>-morpholino group are anticlinal from one another, with the O–C<sub>1</sub>–C<sub>2</sub>–N dihedral angle roughly 120° in molecules 1 and 2, and roughly 240° in molecules 3 and 4 shown in **Figure 6.1**, top and bottom right.

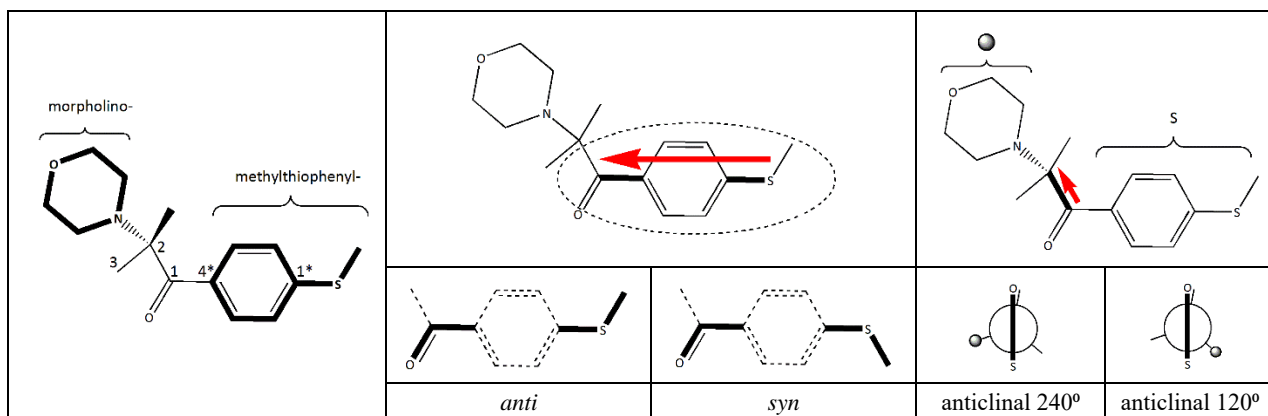


Figure 6.1 - Nomenclature of MMMP conformers.

#### 6.4.2 Numbering of the atoms in the unit cell

Each of the four MMMP conformers consists of C, H, N, S and O atoms, a single conformer was picked at random and assigned as conformer 1, from the selected conformer 1 the other conformers were assigned a number in a clockwise direction. Therefore, conformer 2 would be to the right of conformer 1, conformer 3 would be below 2 and 4 would be to the left of 3 and below 1 as shown is **Figure 6.2**.

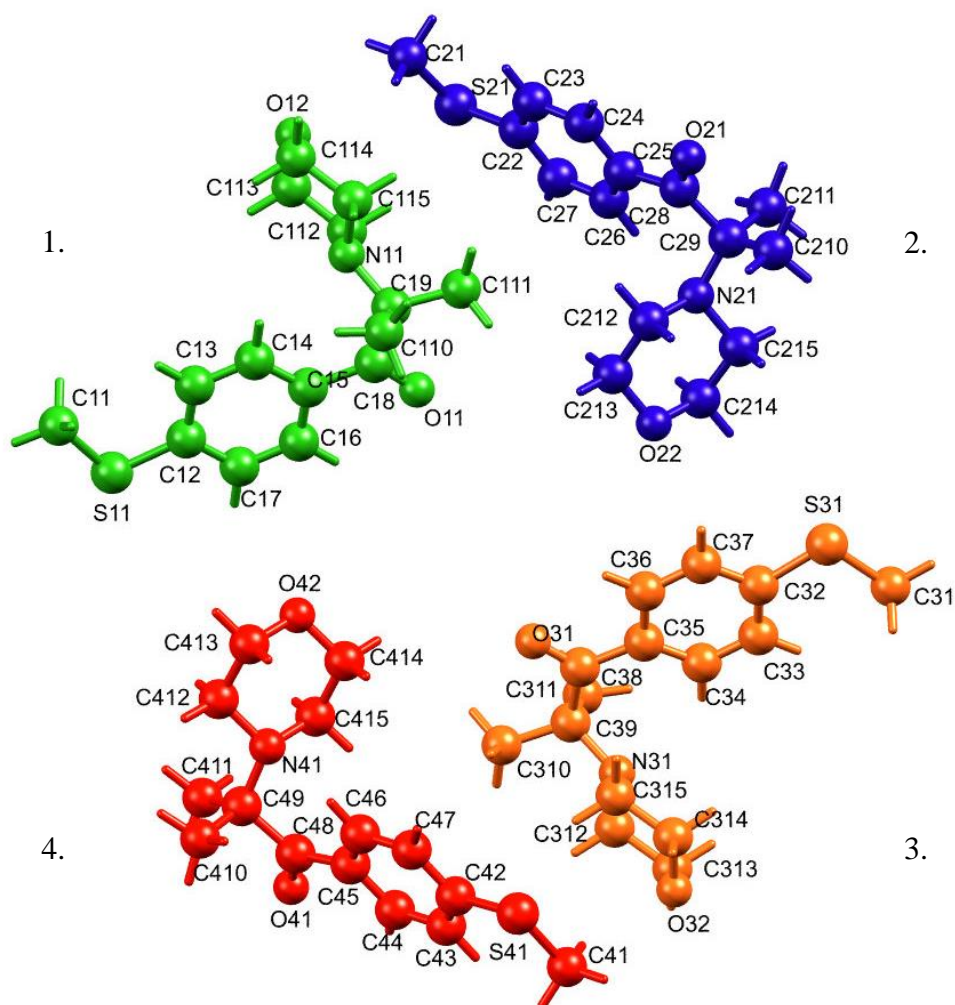


Figure 6.2 - Arrangement of four MMMP conformers in a unit cell.




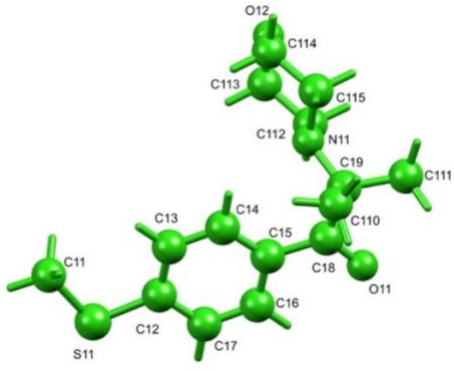

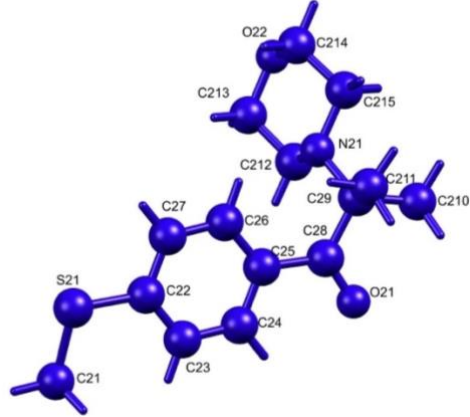

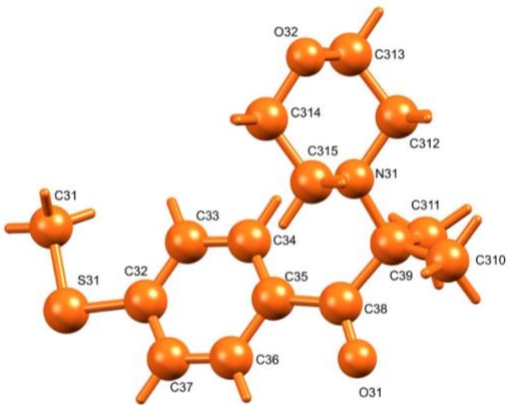

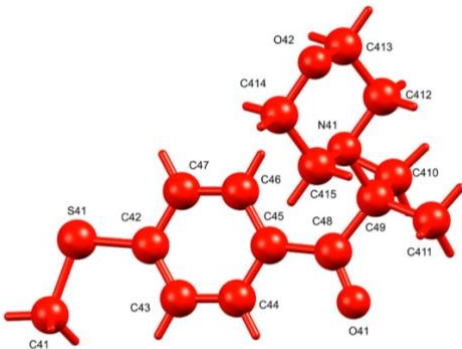
For simplicity the numbering of the atoms within conformer 1 will be discussed, the same principles are applied to conformers 2, 3 and 4. The carbons are numbered starting at the methyl attached to the sulphur, continuing to the first carbon of the phenyl ring. The carbons of the phenyl ring are numbered in a clockwise direction until the last carbon of the ring is numbered, thereafter the carbonyl carbon is numbered followed by the  $\alpha$ -carbon and its methyl substituents. The carbons of the morpholino moiety are numbered clockwise as well. There is only a single sulphur and nitrogen atom which are assigned a value of 1. The two oxygen atoms present within a MMMP molecule, are assigned the values 1 and 2 with 1 being assigned to the carbonyl oxygen and 2 to the morpholino oxygen. For Conformer 1, a value of 1 is assigned after the atomic symbol, thereafter another number is assigned according to the number assigned to that atom as described above, thus C(11) refers to the first carbon atom in conformer 1. Similarly, C(215) refers to the 15<sup>th</sup> carbon atom in conformer 2 and S31 refers to the first sulphur atom in conformer 3.

Due to the vast number of hydrogens present the hydrogens were numbered automatically with an alphabetical number. The first 2 numbers of the carbon to which the hydrogens are attached to are kept and all the hydrogens are counted alphabetically i.e. H(11a), in effect having the smallest number and alphabetical order. To rule out any confusion when referring to a specific hydrogen either a diagram will be provided or the carbon to which the hydrogen is attached (in the case of a  $sp^2$  carbon) will be referred to.

## 6.5 Structural chemistry

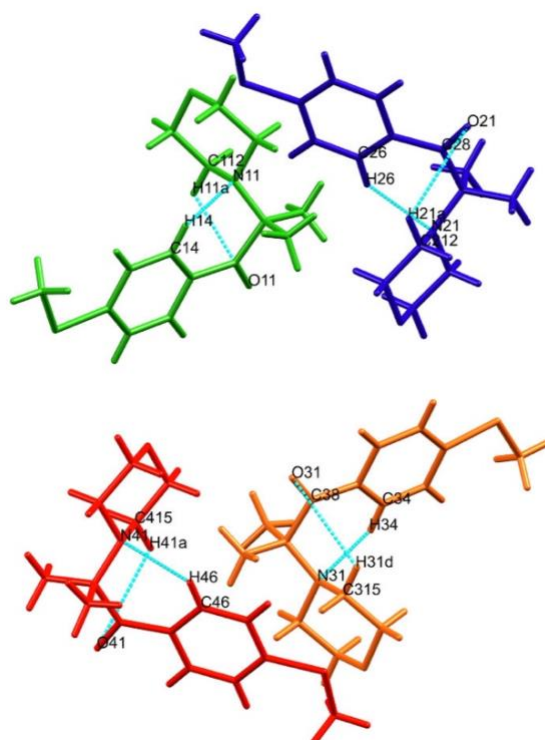
Looking at the asymmetric unit, conformers 1 to 4 arrange themselves in a circular manner, with the *anti*-conformers (1 and 3) and the *syn*-conformers (2 and 4) respectively opposing each other. **Table 6.2** describes and graphically represent the orientation of the substituents in each of the conformers observed in **Figure 6.2** above.

**Table 6.2** - The four MMMP conformers.

|  |   |   |  |
|--|---|---|--|
|    |   |    |   |
| 1 ( <i>anti</i> anticlinal 120°)   |   | 2 ( <i>syn</i> anticlinal 120°)   |  |
|  |  |  |  |
| 3 ( <i>anti</i> anticlinal 240°)   |   | 4 ( <i>syn</i> anticlinal 240°)   |  |

### 6.5.1 Intramolecular interactions

Apart from the intramolecular covalent bonds between the atoms there are interactions between regions of the same conformer that hold each conformer in its specific conformation to form a specific conformer; these interactions occur between hydrogens, nitrogen and oxygen atoms. There are two strong hydrogen bonding interactions within each conformer, the first interaction is between the carbon-hydrogen and morpholino nitrogen and the second between carbon-hydrogen and the carbonyl oxygen in all of these interactions the donor atom is carbon. These interactions are summarised in **Table 6.3** and graphically shown in **Figure 6.3**.



*Figure 6.3 - Intramolecular hydrogen bonding interactions.*

It is interesting to note the lengths of the phenyl C-H bonds are all 0.950 Å in length while the C-H bonds in the morpholino moiety are 0.990 Å. The reason for the shorter bond could be attributed to the increased stability of the phenyl ring due to resonance which makes it more stable than the morpholino ring. Better delocalisation of the electron density causes greater attraction between carbon and hydrogen atoms. The morpholino moiety is also a saturated  $sp^3$ -C heteroatom ring while the phenyl ring contains unsaturated  $sp^2$ -C,  $sp^2$ -C have characteristically shorter bonds compared to  $sp^3$ -C. Distances of  $H\cdots N$  range from 2.376 Å to 2.421 Å and of  $H\cdots O$  range from 3.098 Å to 3.273 Å of which both are five bond lengths away from one another. The shorter distance can be explained by the stronger bond between hydrogen and nitrogen due to the nitrogen being a stronger nitrogen bond acceptor.<sup>[429]</sup> From the angles we observe a smaller angle ( $112.77^\circ$  and  $112.38^\circ$ ) between  $C-H\cdots O$  for the *syn*-

conformers 2 and 4 and larger angles (117.56° and 115.48°) for the *anti*-conformers 1 and 3. Similar angles between C-H...N are observed for all conformers ranging from 118.28° to 119.92°.

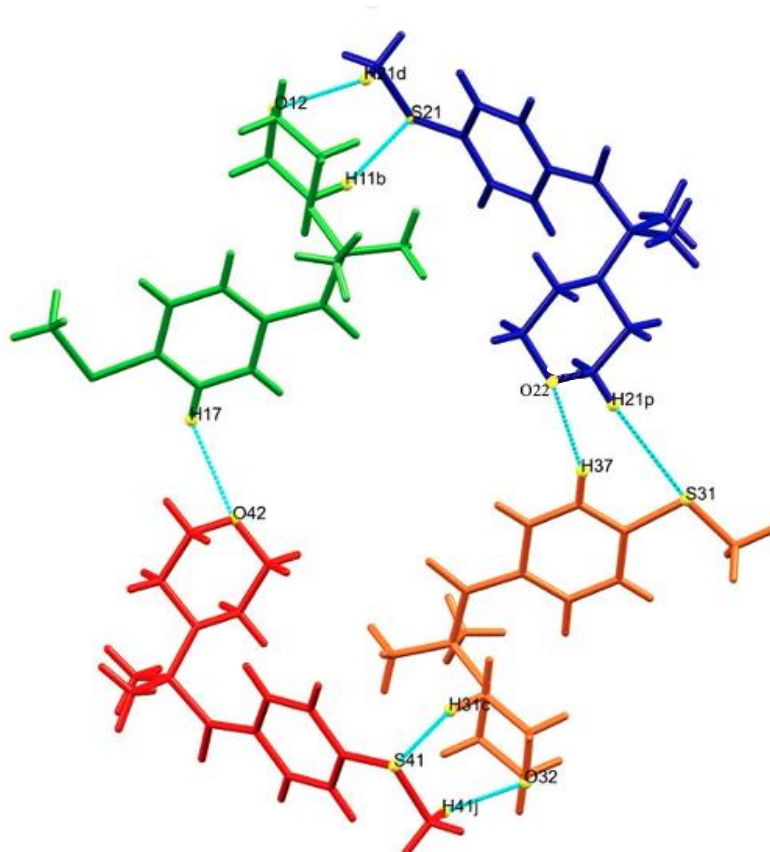
**Table 6.3 - Intramolecular hydrogen bonding.**

| Conformer                       | Donor-H...Acceptor    | Donor-H Length (Å) | H...Acceptor (Å) | ∠(D-H...A) (°) |
|---------------------------------|-----------------------|--------------------|------------------|----------------|
| (1) <i>Anti</i> anticlinal 120° | C(14)-H(14)...N(11)   | 0.950              | 2.376            | 119.92         |
|                                 | C(112)-H(11a)...O(11) | 0.990              | 3.098            | 117.59         |
| (2) <i>Syn</i> anticlinal 120°  | C(26)-H(26)...N(21)   | 0.950              | 2.421            | 119.12         |
|                                 | C(212)-H(21a)...O(21) | 0.990              | 3.187            | 113.28         |
| (3) <i>Anti</i> anticlinal 240° | C(34)-H(34)...N(31)   | 0.950              | 2.414            | 118.28         |
|                                 | C(315)-H(31d)...O(31) | 0.990              | 3.174            | 115.48         |
| (4) <i>Syn</i> anticlinal 240°  | C(46)-H(46)...N(41)   | 0.950              | 2.419            | 119.68         |
|                                 | C(415)-H(41a)...O(41) | 0.990              | 3.273            | 112.77         |

### 6.5.2 Intermolecular interactions

Evaluation of the unit cell reveals 37 intermolecular interactions of different types. These interactions can be grouped into dipole-dipole forces (I) within and (II) between unit cells, (III) aromatic interactions and (IV) induce dipole-induced dipole forces between unit cells. Dipole-dipole interactions and more specifically hydrogen bond interactions were considered significant when two conditions were met: (1) distance between hydrogen and acceptor is not greater than 3.5 Å and (2) the angle between the donor atom-hydrogen...acceptor is not smaller than 120°. [389, 390] **Table 6.4 and 6.5** below contains interactions for both the interactions within and between unit cells.

**(I) Intermolecular interactions within the unit cell.** Within a unit cell the four conformers interact with one another to form seven unique interactions. These interactions are dipole-dipole type interactions of which four are hydrogen bond interactions and the other interactions are between hydrogen and sulphur as listed in **Tables 6.4 and 6.5** and illustrated in **Figure 6.4** below.



*Figure 6.4 - Intermolecular interactions in the unit cell.*

In all four conformers the oxygen of the morpholino moiety interacts with a hydrogen of an adjacent conformer. In *anti*-conformers 1 and 3 the morpholino oxygen interacts with a hydrogen on the methylthio moiety of conformers 2 and 4 (**Table 6.4, Interaction no. 5 and 14**); in the *syn*-conformers 2 and 4 the oxygen of the morpholino moiety interacts with a hydrogen in the aromatic ring of conformers 1 and 3 (**Table 6.4, Interaction no. 11 and 22**). Sulphur atoms of conformers 2 to 4 interact with hydrogens of the morpholino moiety in conformers 1 to 3 (**Table 6.5, Interaction no. 25, 28 and 30**). However, the sulphur of conformer 1 does not interact with any of the hydrogens within the unit cell but do however interact with two other hydrogens of two separate conformer 4 molecules in different unit cells as listed in **Table 6.5, Interaction no. 23 and 24**, and shown in **Figure 6.5**. In both types of intermolecular dipole-dipole interaction cases the hydrogens that form these interactions are not more than three bond lengths away from a very electronegative atom (N, O or S) which assists in drawing electron density away from the hydrogens. Hydrogens become slightly more electropositive and the electronegative atoms slightly more negative assisting in the forming of these dipole-dipole interactions.

**Table 6.4** - Intermolecular hydrogen bonding interactions.

| Interaction number            | Donor-H... Acceptor               | Donor-H Length (Å) | H... Acceptor (Å) | ∠(D-H...A) (°) |
|-------------------------------|-----------------------------------|--------------------|-------------------|----------------|
| Interactions with oxygen (11) |                                   |                    |                   |                |
| 1.                            | C(311)-H(31i)...O(11)             | 0.979              | 2.627             | 149.33         |
| 2.                            | C(115)-H(11d)...O(11)             | 0.990              | 2.516             | 137.78         |
| 3.                            | C(110)-H(11m)...O(11)             | 0.979              | 2.918             | 125.70         |
| Interactions with oxygen (12) |                                   |                    |                   |                |
| 4.                            | C(31)-H(31p)...O(12)              | 0.980              | 2.679             | 130.80         |
| 5.                            | C(21)-H(21d)...O(12) <sup>a</sup> | 0.979              | 2.538             | 132.61         |
| 6.                            | C(210)-H(21l)...O(12)             | 0.981              | 3.056             | 122.23         |
| Interactions with oxygen (21) |                                   |                    |                   |                |
| 7.                            | C(11)-H(11k)...O(21)              | 0.980              | 2.801             | 166.54         |
| 8.                            | C(211)-H(21h)...O(21)             | 0.979              | 2.684             | 134.06         |
| 9.                            | C(215)-H(21j)...O(21)             | 0.990              | 2.426             | 163.19         |
| Interactions with oxygen (22) |                                   |                    |                   |                |
| 10.                           | C(110)-H(11l)...O(22)             | 0.980              | 2.903             | 152.74         |
| 11.                           | C(37)-H(37)...O(22) <sup>a</sup>  | 0.950              | 2.494             | 125.95         |
| Interactions with oxygen (31) |                                   |                    |                   |                |
| 12.                           | C(312)-H(31g)...O(31)             | 0.990              | 2.396             | 147.80         |
| 13.                           | C(110)-H(11n)...O(31)             | 0.980              | 2.634             | 150.36         |
| Interactions with oxygen (32) |                                   |                    |                   |                |
| 14.                           | C(41)-H(41j)...O(32) <sup>a</sup> | 0.979              | 2.557             | 145.47         |
| 15.                           | C(11)-H(11i)...O(32)              | 0.979              | 2.448             | 151.27         |
| 16.                           | C(412)-H(41e)...O(32)             | 0.990              | 3.233             | 146.89         |
| 17.                           | C(411)-H(41m)...O(32)             | 0.980              | 3.178             | 156.13         |
| Interactions with oxygen (41) |                                   |                    |                   |                |
| 18.                           | C(412)-H(41f)...O(41)             | 0.990              | 2.425             | 145.62         |
| 19.                           | C(31)-H(31q)...O(41)              | 0.980              | 2.907             | 146.60         |
| Interactions with oxygen (42) |                                   |                    |                   |                |
| 20.                           | C(11)-H(11j)...O(42)              | 0.979              | 2.931             | 151.90         |
| 21.                           | C(311)-H(31k)...O(42)             | 0.980              | 3.193             | 141.05         |
| 22.                           | C(17)-H(17)...O(42) <sup>a</sup>  | 0.950              | 2.713             | 132.87         |

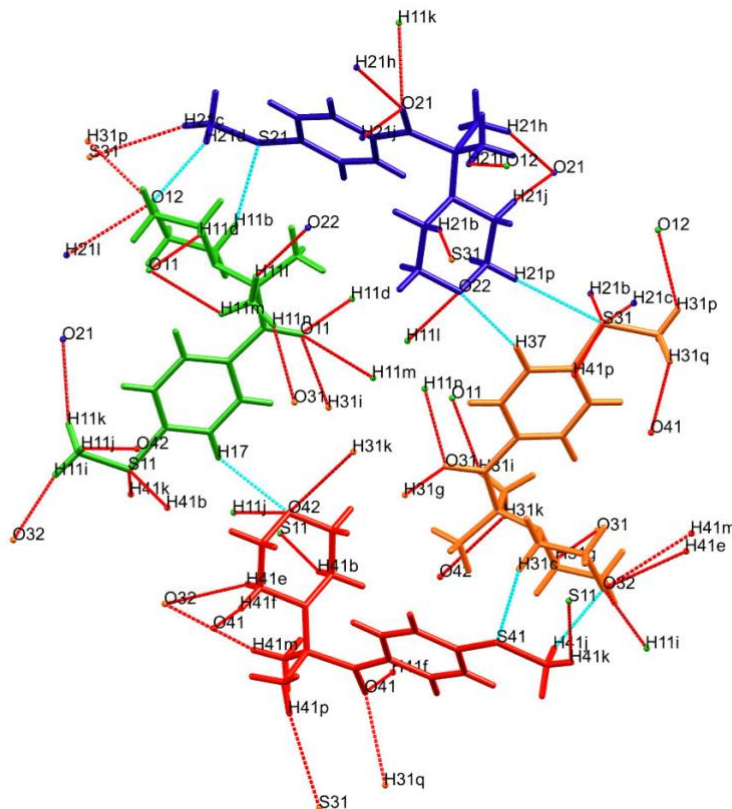
<sup>a</sup> Intermolecular interactions within the unit cell

**Table 6.5** - Other intermolecular dipole-dipole interactions.

| Interaction number                             | Donor-H...Acceptor                 | Donor-H Length (Å) | H...Acceptor (Å) | $\angle(\text{D-H}\cdots\text{A})$ (°) |
|--|------------------------------------|--------------------|------------------|--|
| Dipole-dipole interactions with sulphur atoms  |                                    |                    |                  |  |
| 23.  | C(415)-H(41b)···S(11)              | 0.990              | 2.873            | 161.62                                 |
| 24.  | C(41)-H(41k)···S(11)               | 0.982              | 3.172            | 124.19                                 |
| 25.  | C(112)-H(11b)···S(21) <sup>a</sup> | 0.990              | 3.099            | 147.08                                 |
| 26.  | C(212)-H(21b)···S(31)              | 0.990              | 3.025            | 158.55                                 |
| 27.  | C(21)-H(21c)···S(31)               | 0.981              | 3.073            | 149.85                                 |
| 28.  | C(214)-H(21p)···S(31) <sup>a</sup> | 0.990              | 3.248            | 147.56                                 |
| 29.  | C(410)-H(41p)···S(31)              | 0.980              | 3.039            | 170.21                                 |
| 30.  | C(315)-H(31c)···S(41) <sup>a</sup> | 0.990              | 3.121            | 140.28                                 |
| Dipole-dipole interactions with aromatic rings |                                    |                    |                  |  |
| 31.  | C(21)-H(21e)···Cg(4) <sup>b</sup>  | 0.981              | 2.601            | 157.25                                 |
| 32.  | C(41)-H(41i)···Cg(2) <sup>b</sup>  | 0.981              | 2.731            | 161.32                                 |

<sup>a</sup> Intermolecular interactions within the unit cell.

<sup>b</sup> Cg refers to the centroid of the aromatic ring in conformer 2 or 4.



**Figure 6.5** - Dipole-dipole intermolecular interactions within and between unit cells. Interactions shown between molecules are within the unit cell, while interactions shown to only a single atom can be considered as interactions occurring with neighbouring unit cells.

**(II) Intermolecular interactions between the unit cells.** Unit cells are linked together with a series of dipole-dipole, aromatic and van der Waals interactions. These three interactions will be discussed in the sections below.

**Intermolecular dipole-dipole interactions** - Eighteen hydrogen bond interactions between oxygens and donor hydrogens form most of the interactions between unit cells and are shown in **Table 6.4**. It is interesting to note that none of the intermolecular hydrogen bond interactions are formed with nitrogen. Possible reason for this could be the steric hindrance caused by the dimethyl one bond length away and the bulky and electron rich morpholino moiety repelling interactions. Both oxygens are located in more accessible parts of the MMMP molecule, with the carbonyl oxygen being the least accessible and the morpholino oxygen the most. Seven dipole-dipole interactions occur between the unit cells of which five are between hydrogen and sulphur atoms and two between the aromatic ring of methylthiophenyl moiety of one conformer and the methylthio of another conformer. These 7 interactions are listed in **Table 6.5**.

**General discussion for the dipole-dipole interactions (I) within and (II) between unit cells.**

For all 22 unique hydrogen bond interactions (within a unit cell and between unit cells) the average length between the donor atom and its hydrogen is 0.979 Å. The hydrogen bond interactions between the hydrogen and acceptor oxygen have an average length of 2.739 Å. The shortest hydrogen interaction measured 2.396 Å and is between two conformer 3 molecules offset stacked parallel to one another allowing the bulky and electron rich substituents to optimally pack leading to the least amount of repulsion allowing for the carbonyl oxygen of one conformer to interact with C(312)-H(31g), in the morpholino moiety, of the other. The longest bond measured 3.233 Å, steric and electronic clash between two morpholino moieties causes greater repulsion between the two morpholino rings. Therefore, an interaction is still possible but at a greater distance to reduce the repulsion and steric clashes between the two identical moieties. All the lengths and angles of the hydrogen bond interactions are reported in **Table 6.4**.

Considering all eight sulphur interactions within and between unit cells, S(31) forms four interactions to different hydrogens, S(11) forms two interactions, S(21) and S(41) both only have one interaction. These interactions are made possible by the accessible positions of sulphur atoms within the unit cell, with all four conformers the sulphurs are located on the outer edge of the unit cell making them readily available to form dipole-dipole interactions with little steric hindrance from nearby bulky groups. It should be noted that the average

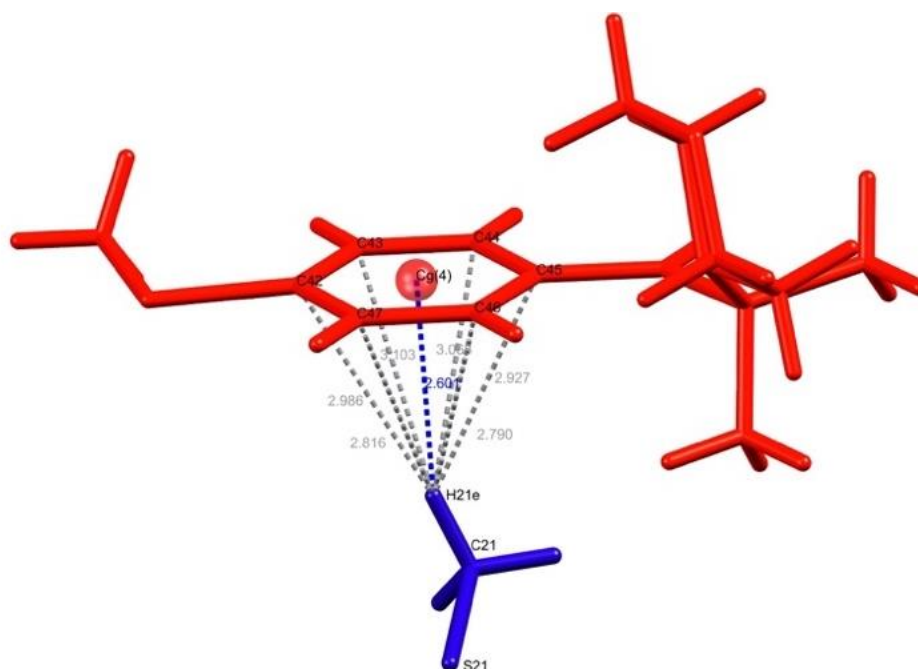


distance between the donor atom and its hydrogen are very similar (0.987 Å) to those involved in interactions with oxygen; hydrogens and acceptor sulphur have an average length of 3.081 Å somewhat longer than the average hydrogen and oxygen interaction length of 2.739 Å. These longer lengths correspond to a weaker type of dipole-dipole interaction.

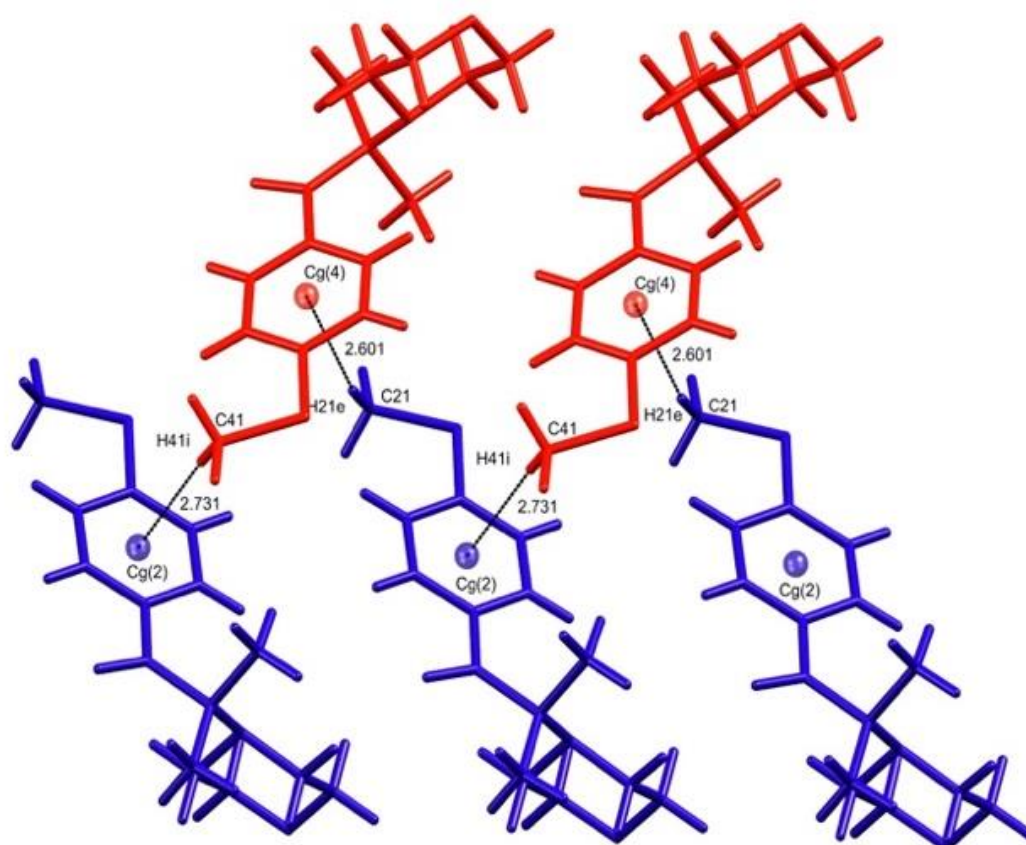
The final dipole-dipole interaction occurs between the centroids of the aromatic ring in conformer 2 and 4 and a hydrogen on the methylthio moiety of the other conformer (**Table 6.5**). These interactions occur due to the small partial negative charges centred on the ring carbon atoms creating a partial negative charge within the centre of the ring this allows the aromatic ring to interact with a partially positive charged hydrogen atom. These dipole-dipole interactions are about half as strong as normal hydrogen bond interactions.<sup>[430, 431]</sup> **Figure 6.6 (a)** below illustrates one of these interaction between the aromatic ring centroid of conformer 4 and the hydrogen atom on conformer 2, an identical interaction occurs between the centroid of conformer 2 and the hydrogen atom in conformer 4. **Figure 6.6 (b)** shows both interactions in packed unit cells. It is important to note that the centroid is closer to the hydrogen than any of the carbon or hydrogen atoms.

Similarly, to the seven dipole-dipole interactions **within** the unit cell, these 25 interactions **between** unit cells have hydrogens that are not more than three bond lengths away from a very electronegative atom (N, O or S) which assists in drawing electron density away from the hydrogens causing them to become slightly more electropositive, which aid in interaction formations.

(a)

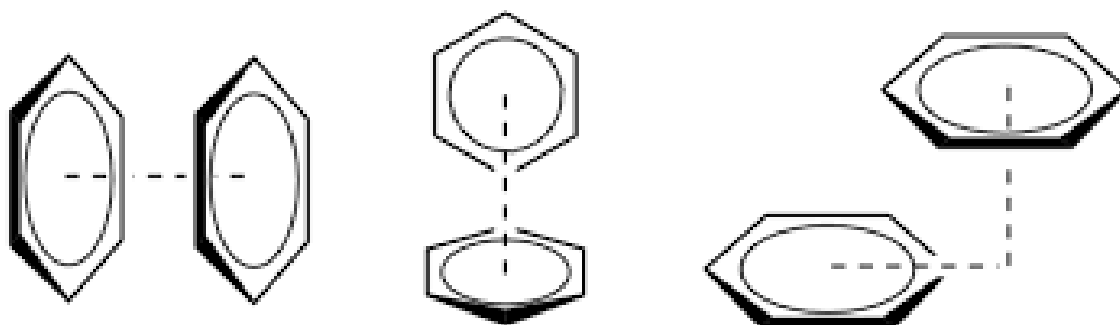


(b)



*Figure 6.6 (a) - Dipole-dipole interaction between the centroid of conformer 4 and a hydrogen donor of conformer 2. (b) Both interactions present in packed unit cells.*

**(III) Intermolecular aromatic interactions.** Another type of interaction between unit cells are facilitated by the electron rich aromatic rings of the morpholino moiety of all four conformers. The different conformers interact with the equivalent conformer at symmetry position  $(x, y-1, z)$ , (See **Figure 6.9 (a)-(b)**). Three types of ring interactions exist the first and least common is a sandwich configuration (face-to-face) in which rings are parallel and opposite one another; the other two are T-stacking (edge-to-face), where the second ring is perpendicular to the first ring, and offset stacking (parallel displaced). These ring interactions are shown in **Figure 6.7** below.



*Figure 6.7 - Types of ring interactions. Sandwich (left), T-stacking (centre) and offset stacking (right).*

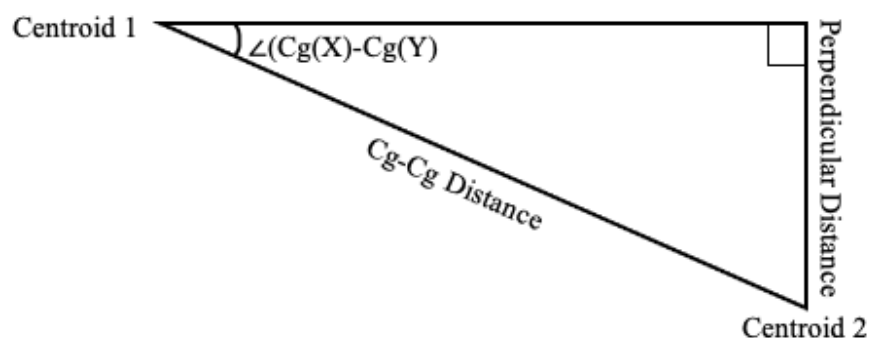
Generally, the electrostatic forces or dispersive forces contribute to repulsion of the rings favouring either the T-stacking or the offset configuration.<sup>[391, 432]</sup> However, it has been found that multiple substituents on the phenyl ring (such as the methylthio substituent seen here) lower the repulsive forces. In this crystal structure offset stacking is observed between identical conformers stacked along the b-axis, these interactions are summarised in **Table 6.6** and shown in **Figure 6.9 (a)-(f)**.

For all four conformers the vector distances drawn between the two centroids are 6.577 Å. The perpendicular distance is defined as the distance between the two planes drawn lengthwise through each of the phenyl rings parallel to one another, all vary from one conformer pair interaction to the next which directly influences the angle,  $\angle(\text{Cg}(\text{X})-\text{Cg}(\text{Y}))$ , between the centroids. Aromatic ring interactions were considered significant if the distances between the ring centroids (Cg) were less than 7 Å.<sup>[392]</sup>

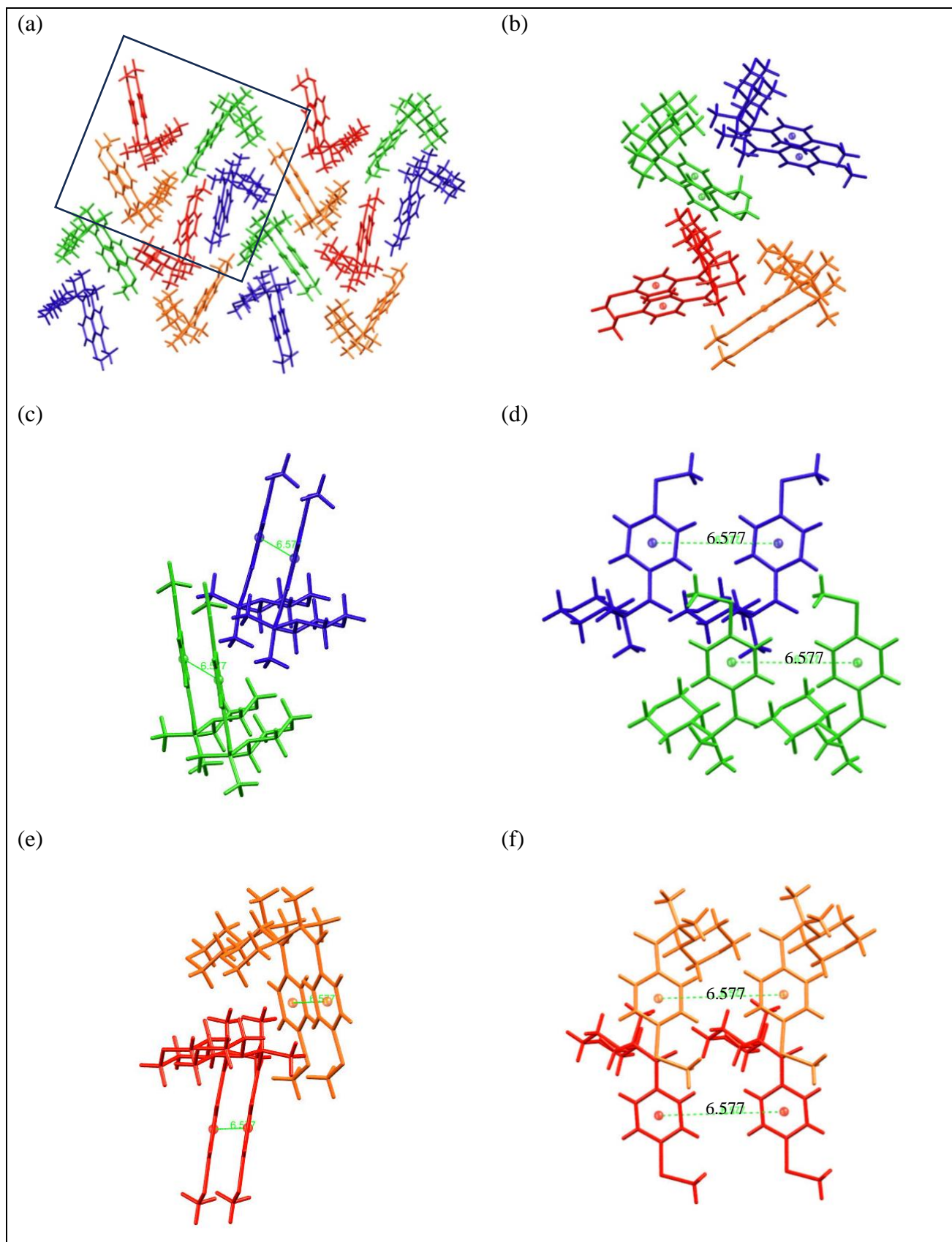
**Table 6.6** - Intermolecular aromatic interactions.

| Interaction number | <sup>a</sup> Cg(X)-Cg(Y) | <sup>b</sup> Cg-Cg (Å) | <sup>c</sup> Perpendicular distance (Å) | <sup>d</sup> $\angle(\text{Cg}(\text{X})-\text{Cg}(\text{Y}))$ (°) |
|--------------------|--------------------------|------------------------|---|--|
| 33.                | Cg(1)-Cg(1)              | 6.577                  | 1.789                                   | 15.78  |
| 34.                | Cg(2)-Cg(2)              | 6.577                  | 1.864                                   | 16.46  |
| 35.                | Cg(3)-Cg(3)              | 6.577                  | 1.427                                   | 12.53  |
| 36.                | Cg(4)-Cg(4)              | 6.577                  | 1.637                                   | 14.41  |

<sup>a</sup> Cg is the centroid of the six-membered aromatic ring of each MMMP conformer, 1-4 refers to the conformer in which the centroid originates in. <sup>b</sup> Cg-Cg is the vector distance between the two parallel displaced aromatic ring centroids of the same conformer. <sup>c</sup> Perpendicular distance is the distance between the two planes, drawn lengthwise through the phenyl rings, parallel to one another. <sup>d</sup>  $\angle(\text{Cg}(\text{X})-\text{Cg}(\text{Y}))$  is the angle between the centroids.



**Figure 6.8** - Parameters between centroids.



**Figure 6.9** (a)-(f) - Intermolecular aromatic interactions. (a)  $1 \times 2 \times 1$  packing of the unit cell; (b) Isolation of the four conformers shown in the black box in (a); (c)-(d) Conformer 1 and 2 packed along the *b*-axis and (e)-(f) Conformer 3 and 4 packed along the *b*-axis.

(IV) **Intermolecular induced dipole-induced dipole interactions.** These are the most common interactions to occur and are the weakest type of intermolecular forces. For the purpose of this study only the sulphur-sulphur interaction will be mentioned (**Table 6.7**). The sulphur atom of conformer 1 interacts with the sulphur atom of conformer 4 in an adjacent unit cell this interaction has a length of 3.534 Å

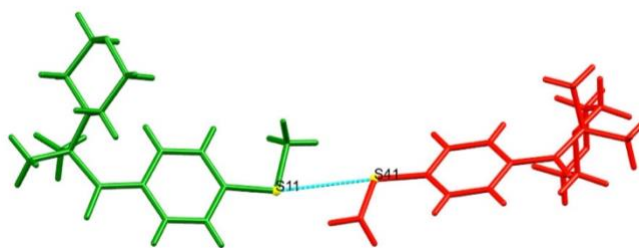


Figure 6.10 - Intermolecular van der Waals interaction.

(**Figure 6.10**). Comparison of the average dipole-dipole interaction length of O···H (2.739 Å) and S···H (3.081 Å), confirms this van der Waals force is weaker compared to dipole-dipole forces.

**Table 6.7** - Intermolecular van der Waals interactions.

| Interaction number | Atom···Atom   | S(X)···S(Y) (Å) |
|--------------------|---------------|-----------------|
| 37.                | S(11)···S(41) | 3.534           |

## 6.6 Symmetry elements in 1x1x1 unit cell packing

Within the MMMP crystal structure three symmetry elements exist, the first is a 2-fold ( $2_1$ ) screw axis located at  $[\frac{1}{2}, 0, 0]$  and represented as a horizontal light green line in **Figure 6.11**. The other two elements are glide planes the first is located at  $[\frac{1}{4}, 0, 0]$  and equivalently at  $[\frac{3}{4}, 0, 0]$  which forms a plane perpendicular to the  $[1, 0, 0]$  plane and has glide component  $[0, 0, \frac{1}{2}]$ . The second is located at  $[0, \frac{1}{2}, 0]$  which forms the glide plane perpendicular to the  $[0, 1, 0]$  plane, with a glide component  $[\frac{1}{2}, 0, 0]$ . The first glide component is shown in **Figure 6.12** as vertical blue lines the second as a horizontal blue line. In both figures lattice point  $[0, 0, 0]$  is located in the top left corner. All these symmetry elements are summarised in **Table 6.8** below and is also mentioned in the space group  $Pca2_1$  in which MMMP crystallises in,  $c$  refers to glide plane  $c$  perpendicular to the  $a$ -axis,  $a$  refers to glide plane  $a$  perpendicular to the  $b$ -axis and  $2_1$  refers to a two-fold screw axis.

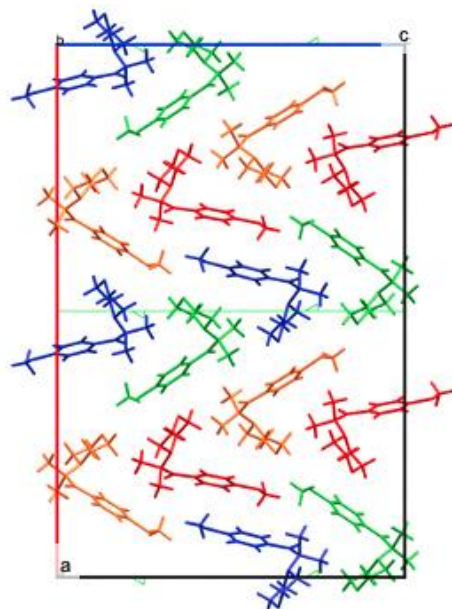


Figure 6.11 - Illustration of the 2-fold screw axis viewed along *b*-axis in a  $1 \times 1 \times 1$  packed cell.

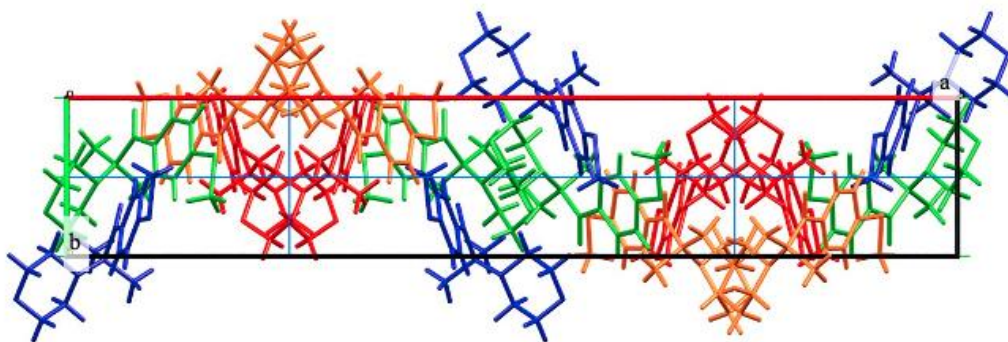


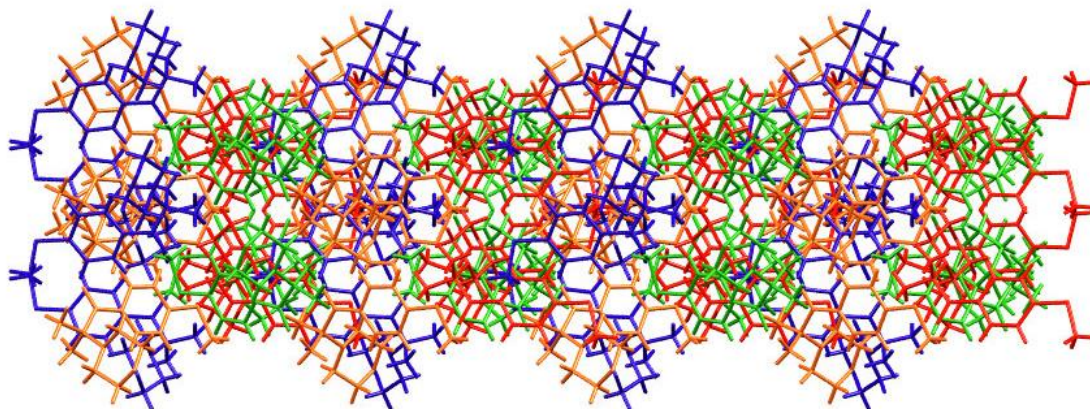
Figure 6.12 - Illustration of the glide planes, blue horizontal and vertical lines, viewed along the *c*-axis in a  $1 \times 1 \times 1$  packed cell.

**Table 6.8** - Symmetry elements present in the crystal structure.

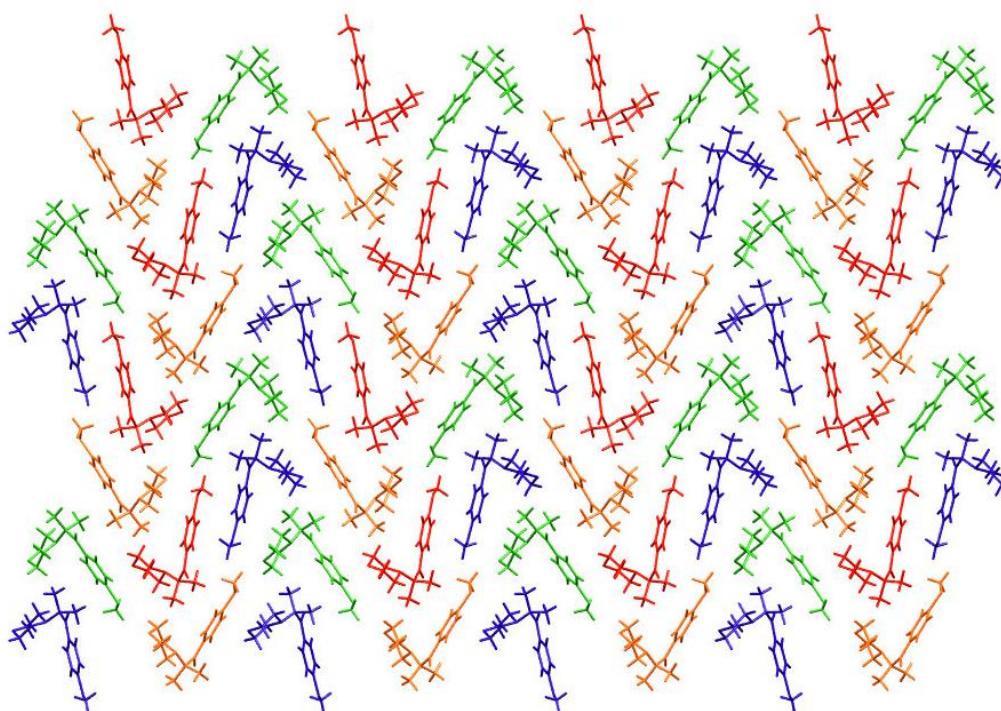
| Number | Description         | Detailed Description   | Symmetry Operator                     |
|--------|---------------------|--|---------------------------------------|
| 1.     | Screw axis (2-fold) | 2-fold screw axis located at $\frac{1}{2}, 0, 0$ , with direction $[0, 0, 1]$ with screw component $[0, 0, \frac{1}{2}]$ | $-x, -y, \frac{1}{2} + z$             |
| 2.     | Glide plane         | Glide plane perpendicular to $[1, 0, 0]$ with glide component $[0, 0, \frac{1}{2}]$                                      | $\frac{1}{2} - x, y, \frac{1}{2} + z$ |
| 3.     | Glide plane         | Glide plane perpendicular to $[0, 1, 0]$ with glide component $[\frac{1}{2}, 0, 0]$                                      | $\frac{1}{2} + x, -y, z$              |

## 6.7 2x2x2 Packing of MMMP unit cell

**Figure 6.13 to 6.15** illustrates a 2x2x2 packing of unit cells along the a, b and c-axis. These figures are included for illustration purposes to observe the effects of the symmetry elements on a larger scale. These crude unit cells truly form a masterpiece not only at the macroscopic level but also at the atomic level.



*Figure 6.13 - 2x2x2 Packing of MMMP viewed along the a-axis.*



*Figure 6.14 - 2x2x2 Packing of MMMP viewed along the b-axis.*

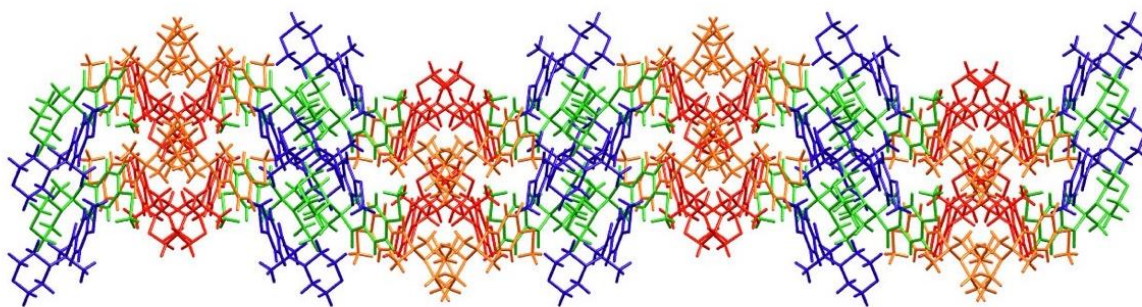


Figure 6.15 - 2x2x2 Packing of MMMP viewed along the c-axis.

### 6.8 Sub-conclusion

The crystal structure of MMMP has successfully been elucidated using SCXRD techniques and crystallographic software to solve the collected diffraction data. From this solved data it was observed that MMMP crystallises in the non-centrosymmetric space group,  $Pca2_1$ . Four crystallographically independent neutral conformers form the unit cell with each conformer interacting with itself, with conformers within the same unit cell as well as with conformers in adjacent unit cells to form the macrostructure.

Within each unit cell the four conformers are kept in their specific conformation through dipole-dipole intramolecular forces between hydrogen, nitrogen and oxygen atoms forming strong hydrogen bond interactions. The four conformers within a unit cell interact with one another through a series of dipole-dipole forces as well as with conformers in adjacent unit cells through a series of different interactions including dipole-dipole, aromatic and induced dipole-induced dipole forces.

Dipole-dipole forces are facilitated through hydrogen, oxygen and sulphur atoms, it had been noted that no intermolecular interactions occur with the nitrogen atoms, due to its inaccessible nature caused by steric hindrance and repulsion from the electron rich morpholino ring moiety. Aromatic interactions occur between the aromatic phenyl rings of equivalent conformers in adjacent unit cells in packing arrangements. Lastly weak van der Waals forces are the most common intermolecular interaction that occur between atoms, of note is the interaction between the two sulphur atoms of conformers 1 and 4. Together these intra- and intermolecular interactions play an integral part in allowing the formation of the MMMP macrocrystal structure in its lowest energy state.



## **Chapter 7: Conclusion and future research**

### **7.1 Summary of results**

For the first part of this study molecular docking results for 25 protein targets consisting of 24 monoamine receptors and a single monoamine transporter were successfully obtained through Glide extra precision (XP) dockings.<sup>[44, 45, 294]</sup> Out of the initial 25 protein targets, serotonin-1A (SER1A) and the serotonin transporter (SERT) were identified as being the most likely to interact with the ligand 2-Methyl-4'-(methylthio)-2-morpholinopropiophenone (MMMP).

The two protein targets were submitted for further evaluation through Molecular Dynamics (MD) simulations to evaluate the stability and movement of these complexes over a 200 ns period. These simulations revealed that MMMP most likely forms a weak complex with SER1A with strong enough interactions occurring to prevent the ligand from diffusing out of the binding pocket for the duration of the simulation. The MD simulation of the SERT-MMMP complex has revealed that MMMP has a high probability of interacting with the SERT binding pocket forming a stronger interaction (compared to the SER1A-MMMP complex) due to the amount and frequency of interactions that occur between the binding pocket of SERT and MMMP. Whether these interactions are capable of triggering downstream psychoactive responses are yet to be determined through biological testing.

For the second part of this study the crystal structure of MMMP was successfully elucidated using single crystal X-ray diffraction (SCXRD) and crystallographic software to solve the collected diffraction data. From this solved data, it was observed that MMMP crystallises in the non-centrosymmetric space group, *Pca2<sub>1</sub>*. A single unit cell consisted of four crystallographically independent conformers in a neutral state with each conformer interacting with itself, with conformers within the same unit cell and with adjacent unit cells to form the macrocrystal structure. The symmetry elements within the crystal structure were also identified.

Within each unit cell, the four conformers are kept in their specific conformation through intramolecular dipole-dipole forces between hydrogen, nitrogen and oxygen atoms forming strong hydrogen bond interactions. The four conformers within a unit cell interact with one another through a series of intermolecular dipole-dipole forces as well as with conformers in adjacent unit cells through different intermolecular interactions including dipole-dipole, aromatic and induced dipole-induced dipole forces. Three symmetry elements were also identified namely an *a* glide plane perpendicular to the b-axis, a *c* glide plane perpendicular to

the a-axis and  $2_1$  screw axis. These symmetry elements are responsible for the symmetry observed in the packing arrangements of unit cells.

## 7.2 Conclusion

This study had set out to investigate the possible interactions that MMMP might have with monoamine protein targets. MMMP was shown to have the ability to form complexes with two of the protein targets, SER1A and SERT, of which the SERT-MMMP complex was identified as having the greatest possibility of forming which might elicit a downstream psychoactive response. The crystal structure of MMMP was successfully elucidated and thoroughly investigated to determine all the interactions that occur within and between the unit cells as well as all the symmetry elements that together gives rise to the observed macrocrystal structure.

Based on these results we hypothesise that MMMP isolated in tablets seized by the South African Police Services (SAPS) back in 2016 (personal communication) and MMMP found in drug seizures in South Australia and in the United States<sup>[31]</sup> was not added as merely a bulking agent, but rather to influence the psychoactive effects experienced by the users. The findings made in this study are purely theoretical and will require further biological studies to confirm these results.

## 7.3 Summary of contributions

This study has shown that MMMP theoretically has the ability to interact with two protein targets SER1A and SERT. These theoretical results can now be confirmed through an *in vitro* study to confirm that (i) the interactions do occur and (ii) are capable of eliciting a response from the protein targets. This work has significantly reduced the time and cost associated with an *in vitro* study by only having to test two out of the 25 initial protein targets.

Additionally, this study paves the way for a move towards an *in silico* chemical structure-based health and neuropharmacology risk assessment approach based on X-ray crystallographic data of human protein targets (i.e. monoamine transporters and receptors) and modelling software which significantly decrease the time required for these assessments. By identifying a few promising targets out of a large pool of targets can significantly reduce the time required for risk assessments and the determination of the pharmacological properties which allows for shorter periods between the identification of a possible illicit substance and scheduling them as illicit. This would particularly be helpful in the case of new psychoactive substances (NPS) which have an unprecedented introduction rate year-on-year. These shorter periods would significantly assist law enforcement agencies to timeously limit the exposure of these

substances to the public as well as assist healthcare workers in identifying overdose cases and administering the correct treatment.

The successfully elucidated crystal structure of commercially available MMMP will be added to the vast library of available crystal structures which can be used for future research endeavours. Analysis of the crystal structure has determined the unit cell, the space group in which MMMP crystallises in, the intra- and intermolecular forces that are responsible for the formation of the stable crystal structure as well as the symmetry elements responsible for the arrangement of the conformers. All of these elements together forms the observed macrocrystal structure.

#### **7.4 Implications of this research**

This work has shown that the highly substituted synthetic cathinone MMMP has the potential to interact with both a serotonin receptor and a serotonin transporter (SER1A and SERT respectively), if however these results are confirmed through biological studies it would warrant intervention from various law enforcement agencies across the globe to have MMMP classified as a controlled substance that has the potential of eliciting psychoactive effects. Classification of MMMP would significantly reduce the availability and ease of accessibility by the public, as well as impact commercial businesses where MMMP is utilised as a Type 1 fragmenting photoinitiator such as in the printing and polymer industries.

#### **7.5 Future research**

It is the dream of many researchers that their work would evoke interest, motivate and/or inspire other researchers to further the research done in a specific field. This work in particular is no different, topics that can be further investigated in future research endeavours include:

- A biological study. Due to the theoretical nature of this work, the next logical step would be to confirm these findings through *in vitro* studies on human models such as those outlined by Simmler and Liechti (2017) titled: “Interactions of Cathinone NPS with Human Transporters and Receptors in Transfected Cells”.<sup>[21]</sup> Other papers published by Simmler *et al.* (2013, 2014, 2018)<sup>[16, 18, 20]</sup> and Eshleman *et al.* (2013, 2017)<sup>[255, 257]</sup> can also be used for biological study methods.
- Performing molecular modelling analysis on the missing protein targets as their crystal structures become available. Molecular modelling techniques were performed on protein targets that were available at the time of this study, future molecular docking studies can be done on the missing monoamine receptors (SER3, SER5B, NER1B and NER1D) and especially the two transporters for dopamine and norepinephrine (DAT

and NET), which have been shown to be the main targets for synthetic cathinones, as their PDB data becomes available.

- Using homology modelling to produce a human DAT from the *Drosophila* dopamine transporter crystal structure and the human DAT amino acid sequence.
- Determination of the nature of the MMMP ligand when interacting with SER1A and SERT. For the monoamine receptor it should be determine if MMMP acts as an agonist or antagonist ligand and in the case of the transporter if MMMP acts as a blocker or a substrate.
- Pharmacokinetics and pharmacodynamics studies can be done on MMMP to determine among other MMMP's distribution of which there is no current data available as well as a full toxicology study.

## References

- [1] RILEY, A. L., NELSON, K. H., TO, P., LÓPEZ-ARNAU, R., XU, P., WANG, D., WANG, Y., SHEN, H. W., KUHN, D. M., ANGOA-PEREZ, M., ANNEKEN, J. H., MUSKIEWICZ, D. & HALL, F. S. 2020. Abuse potential and toxicity of the synthetic cathinones (i.e., "Bath salts"). *Neurosci Biobehav Rev*, 110, 150-173.
- [2] SHAFI, A., BERRY, A. J., SUMNALL, H., WOOD, D. M. & TRACY, D. K. 2020. New psychoactive substances: a review and updates. *Ther Adv Psychopharmacol*, 10, 2045125320967197.
- [3] PEACOCK, A., BRUNO, R., GISEV, N., DEGENHARDT, L., HALL, W., SEDEFOV, R., WHITE, J., THOMAS, K. V., FARRELL, M. & GRIFFITHS, P. 2019. New psychoactive substances: challenges for drug surveillance, control, and public health responses. *The Lancet*, 394, 1668-1684.
- [4] UNODC World Drug Report 2022. United Nations Publication 2021.
- [5] BANKS, M. L., WORST, T. J., RUSYNIAK, D. E. & SPRAGUE, J. E. 2014. Synthetic cathinones ("bath salts"). *The Journal of emergency medicine*, 46, 632-642.
- [6] CORKERY, J., GUIRGUIS, A., PAPANTI, D., ORSOLINI, L. & SCHIFANO, F. 2018. Synthetic Cathinones—Prevalence and Motivations for Use.
- [7] CARHART-HARRIS, R. L., KING, L. A. & NUTT, D. J. 2011. A web-based survey on mephedrone. *Drug and Alcohol Dependence*, 118, 19-22.
- [8] MEASHAM, F., MOORE, K., NEWCOMBE, R. & Z. 2010. Tweaking, bombing, dabbing and stockpiling: The emergence of mephedrone and the perversity of prohibition. *Drugs and Alcohol Today*, 10, 14-21.
- [9] EMCDDA 2022. European Drug Report 2022: Trends and Developments. *European Centre for Drugs and Drug Addiction*. Luxembourg: Publications Office of the European Union.
- [10] EMCDDA 2023. European Drug Report 2023: Trends and Developments. *European Centre for Drugs and Drug Addiction*. Luxembourg: Publications Office of the European Union.
- [11] UNODC 2022. World Drug Report 2022.
- [12] UNODC 2023. World Drug Report 2022.
- [13] SOARES, J., COSTA, V. M., BASTOS, M. L., CARVALHO, F. & CAPELA, J. P. 2021. An updated review on synthetic cathinones. *Arch Toxicol*, 95, 2895-2940.
- [14] REITH, M. E., BLOUGH, B. E., HONG, W. C., JONES, K. T., SCHMITT, K. C., BAUMANN, M. H., PARTILLA, J. S., ROTHMAN, R. B. & KATZ, J. L. 2015. Behavioral, biological, and chemical perspectives on atypical agents targeting the dopamine transporter. *Drug Alcohol Depend*, 147, 1-19.
- [15] SITTE, H. H. & FREISSMUTH, M. 2015. Amphetamines, new psychoactive drugs and the monoamine transporter cycle. *Trends Pharmacol Sci*, 36, 41-50.
- [16] SIMMLER, L. D., BUSER, T. A., DONZELLI, M., SCHRAMM, Y., DIEU, L. H., HUWYLER, J., CHABOZ, S., HOENER, M. C. & LIECHTI, M. E. 2013. Pharmacological characterization of designer cathinones in vitro. *British Journal of Pharmacology*, 168, 458-470.
- [17] IVERSEN, L., GIBBONS, S., TREBLE, R., SETOLA, V., HUANG, X. P. & ROTH, B. L. 2013. Neurochemical profiles of some novel psychoactive substances. *European Journal of Pharmacology*, 700, 147-151.
- [18] SIMMLER, L. D., RICKLI, A., HOENER, M. C. & LIECHTI, M. E. 2014. Monoamine transporter and receptor interaction profiles of a new series of designer cathinones. *Neuropharmacology*, 79, 152-160.
- [19] RICKLI, A., HOENER, M. C. & LIECHTI, M. E. 2015. Monoamine transporter and receptor interaction profiles of novel psychoactive substances: Para-halogenated amphetamines and pyrovalerone cathinones. *European Neuropsychopharmacology*, 25, 365-376.

- [20]SIMMLER, L. D. 2018. Monoamine Transporter and Receptor Interaction Profiles of Synthetic Cathinones. *In: ZAWILSKA, J. B. (ed.) Synthetic Cathinones: Novel Addictive and Stimulatory Psychoactive Substances*. Cham: Springer International Publishing.
- [21]SIMMLER, L. D. & LIECHTI, M. E. 2017. Interactions of Cathinone NPS with Human Transporters and Receptors in Transfected Cells. *In: BAUMANN, M. H., GLENNON, R. A. & WILEY, J. L. (eds.) Neuropharmacology of New Psychoactive Substances (NPS): The Science Behind the Headlines*. Cham: Springer International Publishing.
- [22]LIECHTI, M. E. 2015. Novel psychoactive substances (designer drugs): Overview and pharmacology of modulators of monoamine signalling. *Swiss Medical Weekly*, 145, w14043.
- [23]COZZI, N. V., SIEVERT, M. K., SHULGIN, A. T., JACOB, P., 3RD & RUOHO, A. E. 1999. Inhibition of plasma membrane monoamine transporters by beta-ketoamphetamines. *Eur J Pharmacol*, 381, 63-69.
- [24]NAGAI, F., NONAKA, R. & SATOH HISASHI KAMIMURA, K. 2007. The effects of non-medically used psychoactive drugs on monoamine neurotransmission in rat brain. *European Journal of Pharmacology*, 559, 132-137.
- [25]ROTHMAN, R. B., BAUMANN, M. H., DERSCH, C. M., ROMERO, D. V., RICE, K. C., CARROLL, F. I. & PARTILLA, J. S. 2001. Amphetamine-type central nervous system stimulants release norepinephrine more potently than they release dopamine and serotonin. *Synapse*, 39, 32-41.
- [26]HADLOCK, G. C., WEBB, K. M., MCFADDEN, L. M., CHU, P. W., ELLIS, J. D., ALLEN, S. C., ANDRENYAK, D. M., VIEIRA-BROCK, P. L., GERMAN, C. L., CONRAD, K. M., HOONAKKER, A. J., GIBB, J. W., WILKINS, D. G., HANSON, G. R. & FLECKENSTEIN, A. E. 2011. 4-Methylmethcathinone (Mephedrone): Neuropharmacological Effects of a Designer Stimulant of Abuse. *Journal of Pharmacology and Experimental Therapeutics*, 339, 530-536.
- [27]KEHR, J., ICHINOSE, F., YOSHITAKE, S., GOINY, M., SIEVERTSSON, T., NYBERG, F. & YOSHITAKE, T. 2011. Mephedrone, compared with MDMA (ecstasy) and amphetamine, rapidly increases both dopamine and 5-HT levels in nucleus accumbens of awake rats. *Br J Pharmacol*, 164, 1949-1958.
- [28]BAUMANN, M. H., AYESTAS, M. A., JR., PARTILLA, J. S., SINK, J. R., SHULGIN, A. T., DALEY, P. F., BRANDT, S. D., ROTHMAN, R. B., RUOHO, A. E. & COZZI, N. V. 2012. The designer methcathinone analogs, mephedrone and methylone, are substrates for monoamine transporters in brain tissue. *Neuropsychopharmacology*, 37, 1192-2203.
- [29]LÓPEZ-ARNAU, R., MARTÍNEZ-CLEMENTE, J., PUBILL, D., ESCUBEDO, E. & CAMARASA, J. 2012. Comparative neuropharmacology of three psychostimulant cathinone derivatives: butylone, mephedrone and methylone. *Br J Pharmacol*, 167, 407-420.
- [30]MARTÍNEZ-CLEMENTE, J., ESCUBEDO, E., PUBILL, D. & CAMARASA, J. 2012. Interaction of mephedrone with dopamine and serotonin targets in rats. *Eur Neuropsychopharmacol*, 22, 231-236.
- [31]NASH, C., BUTZBACH, D., STOCKHAM, P., SCOTT, T., ABROE, G., PAINTER, B., GILBERT, J. & KOSTAKIS, C. 2019. A Fatality Involving Furanylfentanyl and MMMP, with Presumptive Identification of Three MMMP Metabolites in Urine. *Journal of Analytical Toxicology*, 43, 291-298.
- [32]KAWASAKI, Y., TSUBOI, C., YAGI, K., MORIZANE, M., MASAOKA, Y., ESUMI, S., KITAMURA, Y. & SENDO, T. 2015. Photoinitiators enhanced 1,2-dichloropropane-induced cytotoxicity in human normal embryonic lung fibroblasts cells in vitro. *Environmental science and pollution research international*, 22, 4763-70.

- [33] MORIZANE, M., KAWASAKI, Y., MIURA, T., YAGI, K., ESUMI, S., KITAMURA, Y. & SENDO, T. 2015. Photoinitiator-Initiated Estrogenic Activity in Human Breast Cancer Cell Line MCF-7. *Journal of Toxicology and Environmental Health, Part A*, 78, 1450-1460.
- [34] 2014. CLH Report: Substance name: 2-methyl-1-(4-methylthiophenyl)-2-morpholinopropan-1-one. Ludwigshafen, Germany: European Chemical Agency.
- [35] THIRAKUL, P., L, S. H., K, L. B. & J, M. P. 2017. Clinical Presentation, Autopsy Results and Toxicology Findings in an Acute N-Ethylpentylone Fatality. *J Anal Toxicol*, 41, 342-346.
- [36] MURRAY, B. L., MURPHY, C. M. & BEUHLER, M. C. 2012. Death following recreational use of designer drug "bath salts" containing 3,4-Methylenedioxypropylvalerone (MDPV). *J Med Toxicol*, 8, 69-75.
- [37] SCHIFANO, F., CORKERY, J. & GHODSE, A. H. 2012. Suspected and confirmed fatalities associated with mephedrone (4-methylmethcathinone, "meow meow") in the United Kingdom. *J Clin Psychopharmacol*, 32, 710-714.
- [38] HASEGAWA, K., WURITA, A., MINAKATA, K., GONMORI, K., NOZAWA, H., YAMAGISHI, I., WATANABE, K. & SUZUKI, O. 2014. Postmortem distribution of PV9, a new cathinone derivative, in human solid tissues in a fatal poisoning case. *Forensic Toxicology*, 33.
- [39] CAWRSE, B. M., LEVINE, B., JUFER, R. A., FOWLER, D. R., VORCE, S. P., DICKSON, A. J. & HOLLER, J. M. 2012. Distribution of methylone in four postmortem cases. *J Anal Toxicol*, 36, 434-439.
- [40] NYCZ, J. E., MALECKI, G., ZAWIAZALEC, M. & PAZDZIOREK, T. 2011. X-ray structures and computational studies of several cathinones. *Journal of Molecular Structure*, 1002, 10-18.
- [41] TRZYBIŃSKI, D., NIEDZIAŁKOWSKI, P., OSSOWSKI, T., TRYNDĄ, A. & SIKORSKI, A. 2013. Single-crystal X-ray diffraction analysis of designer drugs: Hydrochlorides of metaphedrone and pentedrone. *Forensic Science International*, 232, e28-e32.
- [42] WOOD, M. R., LALANCETTE, R. A. & BERNAL, I. 2015. Crystallographic investigations of select cathinones: emerging illicit street drugs known as 'bath salts'. *Acta Crystallographica Section C*, 71, 32-38.
- [43] 2023. Schrödinger 2023-3: Maestro. New York, NY: Schrödinger, LLC.
- [44] FRIESNER, R. A., BANKS, J. L., MURPHY, R. B., HALGREN, T. A., KLICIC, J. J., MAINZ, D. T., REPASKY, M. P., KNOLL, E. H., SHELLEY, M., PERRY, J. K., SHAW, D. E., FRANCIS, P. & SHENKIN, P. S. 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47, 1739-1749.
- [45] FRIESNER, R. A., MURPHY, R. B., REPASKY, M. P., FRYE, L. L., GREENWOOD, J. R., HALGREN, T. A., SANSCHAGRIN, P. C. & MAINZ, D. T. 2006. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry*, 49, 6177-6196.
- [46] BOWERS, K., CHOW, E., XU, H., DROR, R., EASTWOOD, M., GREGERSEN, B., KLEPEIS, J., KOLOSSVÁRY, I., MORAES, M., SACERDOTI, F., SALMON, J., SHAN, Y. & SHAW, D. Molecular dynamics---Scalable algorithms for molecular dynamics simulations on commodity clusters. 2006. 84.
- [47] 2023. Schrödinger 2023-3: Desmond Molecular Dynamics System, D.E. Shaw Research. New York, NY: Maestro-Desmond Interoperability Tools, Schrödinger.
- [48] BROWNSTEIN, H. H. 2015. The Handbook of Drugs and Society. 1st ed.: Wiley-Blackwell.
- [49] BAUMANN, M. H., SOLIS, E., WATTERSON, L. R., MARUSICH, J. A., FANTEGROSSI, W. E. & WILEY, J. L. 2014a. Baths salts, spice, and related designer drugs: The science behind the headlines. *Journal of Neuroscience*, 34, 15150-15158.

- [50]BRUNT, T. M., ATKINSON, A. M., NEFAU, T., MARTINEZ, M., LAHAIE, E., MALZCEWSKI, A., PAZITNY, M., BELACKOVA, V. & BRANDT, S. D. 2017. Online test purchased new psychoactive substances in 5 different European countries: A snapshot study of chemical composition and price. *International Journal of Drug Policy*, 44, 105-114.
- [51]KARILA, L., MEGARBANE, B., COTTENCIN, O. & LEJOYEUX, M. 2015. Synthetic Cathinones: A New Public Health Problem. *Current Neuropharmacology*, 13, 12-20.
- [52]ASSI, S., GULYAMOVA, N., KNELLER, P. & OSSELTON, D. 2017. The effects and toxicity of cathinones from the users' perspectives: A qualitative study. *Hum Psychopharmacol*, 32.
- [53]AL-HEBSHI, N. N. & SKAUG, N. 2005. Khat (*Catha edulis*)-an updated review. *Addict Biol*, 10, 299-307.
- [54]ALLES, G. A., FAIRCHILD, M. D., JENSEN, M. & ALLES, A. 1961. Chemical Pharmacology of *Catha Edulis*. *Journal of Medicinal and Pharmaceutical Chemistry*, 3, 323-352.
- [55]GETASETEGN, M. 2016. Chemical composition of *Catha edulis* (khat): a review. *Phytochemistry Reviews*, 15, 907-920.
- [56]KRIKORIAN, A. D. 1984. Khat and its use: An historical perspective. *Journal of Ethnopharmacology*, 12, 115-178.
- [57]PATEL, N. B. 2015. "Natural Amphetamine" Khat: A Cultural Tradition or a Drug of Abuse? In: TABA, P., LEES, A. & SIKK, K. (eds.) *International Review of Neurobiology*. Academic Press.
- [58]SIKIRU, L. & BABU, S. M. 2010. Khat (*Catha edulis*): Academic, health and psychosocial effects on "mature" students. *African Journal of Drug and Alcohol Studies*, 8.
- [59]FREUND-MICHEL, V. C., BIRRELL, M. A., PATEL, H. J., MURRAY-LYON, I. M. & BELVISI, M. G. 2008. Modulation of cholinergic contractions of airway smooth muscle by cathinone: potential beneficial effects in airway diseases. *European Respiratory Journal*, 32, 579-584.
- [60]ELHAG, H. M., MOSSA, J. S. & EL-OLEMY, M. M. Antimicrobial and Cytotoxic Activity of the Extracts of Khat Callus Cultures. 1999.
- [61]DIMBA, E., GJERTSEN, B. T., FRANCIS, G. W., JOHANNESSEN, A. C. & VINTERMYR, O. K. 2003. *Catha edulis* (Khat) induces cell death by apoptosis in leukemia cell lines. *Ann N Y Acad Sci*, 1010, 384-388.
- [62]AL-HABORI, M. & AL-MAMARY, M. 2004. Long-term feeding effects of *Catha edulis* leaves on blood constituents in animals. *Phytomedicine*, 11, 639-644.
- [63]AL-HEBSHI, N., AL-HARONI, M. & SKAUG, N. 2006. In vitro antimicrobial and resistance-modifying activities of aqueous crude khat extracts against oral microorganisms. *Arch Oral Biol*, 51, 183-188.
- [64]AL-HEBSHI, N. N., AL-SHARABI, A. K., SHUGA-ALDIN, H. M., AL-HARONI, M. & GHANDOUR, I. 2010. Effect of khat chewing on periodontal pathogens in subgingival biofilm from chronic periodontitis patients. *J Ethnopharmacol*, 132, 564-569.
- [65]KALIX, P. & BRAENDEN, O. 1985. Pharmacological aspects of the chewing of khat leaves. *Pharmacological Reviews*, 37, 149-164.
- [66]KALIX, P. 1984. The pharmacology of khat. *General Pharmacology: The Vascular System*, 15, 179-187.
- [67]NENCINI, P. & AHMED, A. M. 1989. Khat consumption: a pharmacological review. *Drug and Alcohol Dependence*, 23, 19-29.
- [68]KALIX, P., GEISSHÜSLER, S. & BRENNEISEN, R. 1987. The effect of phenylpentenyl-khatamines on the release of radioactivity from rat striatal tissue prelabelled with [3H]dopamine. *J Pharm Pharmacol*, 39, 135-137.
- [69]BRENNEISEN, R., GEISSHÜLSER, S. & SCHORNO, X. 1984. Merucathine, A New Phenylalkylamine from *Catha edulis*. *Planta Med*, 50, 531.



- [70] KITE, G. C., ISMAIL, M., SIMMONDS, M. S. & HOUGHTON, P. J. 2003. Use of doubly protonated molecules in the analysis of cathedulins in crude extracts of khat (*Catha edulis*) by liquid chromatography/serial mass spectrometry. *Rapid Commun Mass Spectrom*, 17, 1553-1564.
- [71] FLÜCKIGER, F. A., GEROCK, J. E., BUTLER & TANNER 1887. *Contributions to the Knowledge of Catha Leaves*, Pharm J Transvaal.
- [72] SZENDREI, K. 1980. The chemistry of khat. *Bull Narc*, 32, 5-35.
- [73] WOLFES, O. Über das Vorkommen von d-nor-iso-Ephedrin in *Catha edulis*. *Archiv der Pharmazie*, 268.
- [74] BRÜCKE, F. 1941. Über die zentralerregende Wirkung des Alkaloides Cathin. *Archiv für Experimentelle Pathologie und Pharmakologie*, 198, 100-106.
- [75] FRIEBEL, H. & BRILLA, R. 1963. Über den zentralerregenden Wirkstoff der frischen Blätter und Zweigspitzen von *Catha edulis* Forskal. *Naturwissenschaften*, 50, 354-355.
- [76] UNDND 1975. Studies on the chemical composition of khat. III. Investigations on the phenylalkylamine fraction. . United Nations Division on Narcotic Drugs
- [77] SCHORNO, X. & STEINEGGER, E. 1979. CNS-active phenylpropylamines of *Catha edulis* FORSK. (Celastraceae) of Kenyan origin. *Experientia*, 35, 572-574.
- [78] GEISSHÜSLER, S. & BRENNEISEN, R. 1987. The content of psychoactive phenylpropyl and phenylpentenyl khatamines in *Catha edulis* Forsk. of different origin. *J Ethnopharmacol*, 19, 269-77.
- [79] BRENNEISEN, R., GEISSHÜSLER, S. & SCHORNO, X. 1986. Metabolism of cathinone to (-)-norephedrine and (-)-norpseudoephedrine. *J Pharm Pharmacol*, 38, 298-300.
- [80] BRENNEISEN, R. & GEISSHÜSLER, S. 1985. Psychotropic drugs. III. Analytical and chemical aspects of *Catha edulis* Forsk. *Pharm Acta Helv*, 60, 290-301.
- [81] TOENNES, S. W. & KAUERT, G. F. 2002. Excretion and detection of cathinone, cathine, and phenylpropanolamine in urine after khat chewing. *Clin Chem*, 48, 1715-9.
- [82] WABE, N. & MOHAMMED, M. 2012. What science says about khat (*Catha edulis* Forsk)? Overview of chemistry, toxicology and pharmacology. *Integrative Cancer Therapies*, 2, 29-37.
- [83] WABE, N. T. 2011. Chemistry, pharmacology, and toxicology of khat (*catha edulis* forsk): a review. *Addict Health*, 3, 137-49.
- [84] DUNNE, F. J., JAFFAR, K. & HASHMI, S. I. Legal Highs - Not so new and still growing in popularity. 2015.
- [85] TOENNES, S. W., HARDER, S., SCHRAMM, M., NIESS, C. & KAUERT, G. F. 2003. Pharmacokinetics of cathinone, cathine and norephedrine after the chewing of khat leaves. *Br J Clin Pharmacol*, 56, 125-30.
- [86] ARUNOTAYANUN, W. & GIBBONS, S. 2012. Natural product 'legal highs'. *Nat Prod Rep*, 29, 1304-1316.
- [87] PATEL, N. 2018. Khat—A Natural Source of Cathinone. In: ZAWILSKA, J. B. (ed.) *Synthetic Cathinones. Current Topics in Neurotoxicity*. Cham: Springer International Publishing.
- [88] KALIX, P. 1990. Pharmacological properties of the stimulant khat. *Pharmacol Ther*, 48, 397-416.
- [89] AL-MOTARREB, A., BAKER, K. & BROADLEY, K. J. 2002. Khat: pharmacological and medical aspects and its social use in Yemen. *Phytotherapy Research*, 16, 403-413.
- [90] WIDLER, P., MATHYS, K., BRENNEISEN, R., KALIX, P. & FISCH, H. U. 1994. Pharmacodynamics and pharmacokinetics of khat: a controlled study. *Clin Pharmacol Ther*, 55, 556-62.
- [91] KALIX, P. 1991. The pharmacology of psychoactive alkaloids from ephedra and catha. *J Ethnopharmacol*, 32, 201-8.
- [92] PATEL, N. B. 2000. Mechanism of action of cathinone: the active ingredient of khat (*Catha edulis*). *East Afr Med J*, 77, 329-32.

- [<sup>93</sup>]BRENNEISEN, R., FISCH, H. U., KOELBING, U., GEISSHÜSLER, S. & KALIX, P. 1990. Amphetamine-like effects in humans of the khat alkaloid cathinone. *Br J Clin Pharmacol*, 30, 825-8.
- [<sup>94</sup>]KALIX, P., GEISSHUSLER, S., BRENNEISEN, R., KOELBING, U. & FISCH, H. U. 1990. Cathinone, a phenylpropylamine alkaloid from khat leaves that has amphetamine effects in humans. *NIDA Res Monogr*, 105, 289-90.
- [<sup>95</sup>]KALIX, P. 1992. Cathinone, a natural amphetamine. *Pharmacol Toxicol*, 70, 77-86.
- [<sup>96</sup>]HASSAN, N. A., GUNAID, A. A. & MURRAY-LYON, I. M. 2007. Khat (*Catha edulis*): health aspects of khat chewing. *East Mediterr Health J*, 13, 706-18.
- [<sup>97</sup>]ZELGER, J. L., SCHORNO, H. X. & CARLINI, E. A. 1980. Behavioural effects of cathinone, an amine obtained from *Catha edulis* Forsk.: comparisons with amphetamine, norpseudoephedrine, apomorphine and nomifensine. *Bull Narc*, 32, 67-81.
- [<sup>98</sup>]GOSNELL, B. A., YRACHETA, J. M., BELL, S. M. & LANE, K. E. 1996. Intravenous self-administration of cathinone by rats. *Behav Pharmacol*, 7, 526-531.
- [<sup>99</sup>]HYDE, J. F., BROWNING, E. & ADAMS, R. 1928. SYNTHETIC HOMOLOGS OF d,l-EPHEDRINE. *Journal of the American Chemical Society*, 50, 2287-2292.
- [<sup>100</sup>]SAEM DE BURNAGA SANCHEZ, J. 1929. Sur un Homologue de L'ephedrine (On an Analogue of Ephedrine). *Bulletin de la Société Chimique de France*, p. 284-286.
- [<sup>101</sup>]DERAMOS, E. C. 1964. THE USE OF DIETHYLPROPION IN THE TREATMENT OF OBESITY. *The British journal of clinical practice*, 18, 210-1.
- [<sup>102</sup>]GARDOS, G. & COLE, J. O. 1971. Evaluation of pyrovalerone in chronically fatigued volunteers. *Curr Ther Res Clin Exp*, 13, 631-5.
- [<sup>103</sup>]SOROKO, F. E., MEHTA, N. B., MAXWELL, R. A., FERRIS, R. M. & SCHROEDER, D. H. 1977. Bupropion hydrochloride ((±)  $\alpha$ -t-butylamino-3-chloropropiophenone HCl): a novel antidepressant agent. *Journal of Pharmacy and Pharmacology*, 29, 767-770.
- [<sup>104</sup>]BOSSONG, M. G., VAN DIJK, J. P. & NIESINK, R. J. 2005. Methylone and mCPP, two new drugs of abuse? *Addict Biol*, 10, 321-3.
- [<sup>105</sup>]BONSON, K. R., DALTON, T. & CHIAPPERINO, D. 2019. Scheduling synthetic cathinone substances under the Controlled Substances Act. *Psychopharmacology (Berl)*, 236, 845-860.
- [<sup>106</sup>]MEHTA, N. B. 1974. *Meta chloro substituted- $\alpha$ -butylamino-propiofenones*. Burroughs Wellcome Co., North Carolina, USA patent application US-3819706-A.
- [<sup>107</sup>]SCHÜTTE, J. 1961. *Anorexigenic propiophenones*. Temmler-Werke, Hamburg-Neugraben, Germany patent application US-3001910-A.
- [<sup>108</sup>]THOMAE, K. 1963.  *$\alpha$ -Pyrrolidino-ketones*. Dr. Karl Thomae GmbH, Biberach an der Riss, Germany patent application GB-933507-A.
- [<sup>109</sup>]WANDER, A. 1963.  *$\alpha$ -Pyrrolidinovalerophenones.  $\alpha$ -Pyrrolidino Valerophenones*.
- [<sup>110</sup>]CLEIN, L. J. & BENADY, D. R. 1962. Case of Diethylpropion addiction. *British Medical Journal*, 2, 456.
- [<sup>111</sup>]DAL CASON, T. A., YOUNG, R. & GLENNON, R. A. 1997. Cathinone: An investigation of several N-alkyl and methylenedioxy-substituted analogs. *Pharmacology Biochemistry and Behavior*, 58, 1109-1116.
- [<sup>112</sup>]DENIKER, P., LÔO, H., CUCHE, H. & ROUX, J. M. 1975. [Abuse of pyrovalerone by drug addicts]. *Ann Med Psychol (Paris)*, 2, 745-8.
- [<sup>113</sup>]DERUITER, J., HAYES, L., VALAER, A., CLARK, C. R. & NOGGLE, F. T. 1994. Methcathinone and Designer Analogues: Synthesis, Stereochemical Analysis, and Analytical Properties. *Journal of Chromatographic Science*, 32, 552-564.
- [<sup>114</sup>]JACOB, P. & SHULGIN, A. T. 1996. *Novel n-substituted-2-amino-3',4'-methylenedioxypropiofenones*. PCT/US96/09603.
- [<sup>115</sup>]BENTUR, Y., BLOOM-KRASIK, A. & RAIKHLIN-EISENKRAFT, B. 2008. Illicit cathinone ("Hagigat") poisoning. *Clinical Toxicology*, 46, 206-210.

- [116]SCHIFANO, F., ALBANESE, A., FERGUS, S., STAIR, J. L., DELUCA, P., CORAZZA, O., DAVEY, Z., CORKERY, J., SIEMANN, H., SCHERBAUM, N., FARRE, M., TORRENS, M., DEMETROVICS, Z., GHODSE, A. H., DI FURIA, L., FLESLAND, L., MANNONEN, M., MAJAVA, A., PAGANI, S., PELTONIEMI, T., PASINETTI, M., PEZZOLESI, C., SKUTLE, A., VAN DER KREEFT, P., ENEA, A., DI MELCHIORRE, G., SHAPIRO, H., SFERRAZZA, E., DRUMMOND, C., PISARSKA, A., MERVÓ, B., MOSKALEWICZ, J., FLORIDI, L. & HAUGEN, L. S. Y. 2011. Mephedrone (4-methylmethcathinone; 'Meow meow'): Chemical, pharmacological and clinical issues. *Psychopharmacology*, 214, 593-602.
- [117]BRUNO, R., MATTHEWS, A. J., DUNN, M., ALATI, R., MCILWRAITH, F., HICKEY, S., BURNS, L. & SINDICICH, N. 2012. Emerging psychoactive substance use among regular ecstasy users in Australia. *Drug and Alcohol Dependence*, 124, 19-25.
- [118]BRUNT, T. M., POORTMAN, A., NIESINK, R. J. M. & VAN DEN BRINK, W. 2011. Instability of the ecstasy market and a new kid on the block: Mephedrone. *Journal of Psychopharmacology*, 25, 1543-1547.
- [119]CAMILLERI, A., JOHNSTON, M. R., BRENNAN, M., DAVIS, S. & CALDICOTT, D. G. E. 2010. Chemical analysis of four capsules containing the controlled substance analogues 4-methylmethcathinone, 2-fluoromethamphetamine,  $\alpha$ -phthalimidopropiophenone and N-ethylcathinone. *Forensic Science International*, 197, 59-66.
- [120]DICKSON, A. J., VORCE, S. P., LEVINE, B. & PAST, M. R. 2010. Multiple-drug toxicity caused by the coadministration of 4-methylmethcathinone (mephedrone) and heroin. *Journal of Analytical Toxicology*, 34, 162-168.
- [121]EMCDDA 2008. Europol 2008 Annual Report on the Implementation of Council Decision 2005/387/JHA *European Centre for Drugs and Drug Addiction*.
- [122]2010. Advisory Council on the Misuse of Drugs: Consideration of the Anabolic Steroids. *Advisory Council on the Misuse of Drugs on Consideration of the Cathinones*.
- [123]MORRIS, K. 2010. UK places generic ban on mephedrone drug family. *Lancet*, 375, 1333-1334.
- [124]EMCDDA 2009. Europol 2009 Annual Report on the Implementation of Council Decision 2005/387/JHA. *European Centre for Drugs and Drug Addiction*.
- [125]EMCDDA 2010. Europol 2010 Annual Report on the Implementation of Council Decision 2005/387/JHA. *European Centre for Drugs and Drug Addiction*.
- [126]EMCDDA 2012. Europol 2012 Annual Report on the Implementation of Council Decision 2005/387/JHA *European Centre for Drugs and Drug Addiction*.
- [127]EMCDDA 2013. Europol 2013 Annual Report on the Implementation of Council Decision 2005/387/JHA *European Centre for Drugs and Drug Addiction*.
- [128]EMCDDA 2014. Europol 2014 Annual Report on the Implementation of Council Decision 2005/387/JHA *European Centre for Drugs and Drug Addiction*.
- [129]EMCDDA 2015. Europol 2015 Annual Report on the Implementation of Council Decision 2005/387/JHA *European Centre for Drugs and Drug Addiction*.
- [130]2013. Resolution 56/4 Enhancing international cooperation in the identification and reporting of new psychoactive substances. *Commission on Narcotic Drugs*.
- [131]UNODC 2013. The Challenge of New Psychoactive Substances. *The Challenge of New Psychoactive Substances*.
- [132]EMCDDA 2019. Europol 2019 Annual Report on the Implementation of Council Decision 2005/387/JHA. *Europol 2019 Annual Report on the Implementation of Council Decision 2005/387/JHA*.
- [133]ZAWILSKA, J. & WOJCIESZAK, J. 2018. Novel Psychoactive Substances: Classification and General Information.
- [134]KELLY, J. P. 2011. Cathinone derivatives: A review of their chemistry, pharmacology and toxicology. *Drug Testing and Analysis*, 3, 439-453.

- [135]PAILLET-LOILIER, M., CESBRON, A., LE BOISSELIER, R., BOURGINE, J. & DEBRUYNE, D. 2014. Emerging drugs of abuse: current perspectives on substituted cathinones. *Subst Abuse Rehabil*, 5, 37-52.
- [136]VALENTE, M. J., GUEDES DE PINHO, P., DE LOURDES BASTOS, M., CARVALHO, F. & CARVALHO, M. 2014. Khat and synthetic cathinones: a review. *Archives of toxicology*, 88, 15-45.
- [137]BARCELOUX, D. G. 2012. Methcathinone, Mephedrone, and Methylone. *Medical Toxicology of Drug Abuse*, 120-125.
- [138]SIKK, K. & TABA, P. 2015. Methcathinone "Kitchen Chemistry" and permanent neurological damage. *International Review of Neurobiology*, 120, 257-271.
- [139]UNODC 1971. United Nations convention on psychotropic substances. *United Nations Conference for the Adoption of a Protocol on Psychotropic Substances*.
- [140]EMERSON, T. S. & CISEK, J. E. 1993. Methcathinone: A russian designer amphetamine infiltrates the rural midwest. *Annals of Emergency Medicine*, 22, 1897-1903.
- [141]BOLDING, O. T. 1974. Diethylpropion hydrochloride: an effective appetite suppressant. *Current Therapeutic Research - Clinical and Experimental*, 16, 40-48.
- [142]SEATON, D. A., DUNCAN, L. J. P., ROSE, K. & SCOTT, A. M. 1961. Diethylpropion in the treatment of "refractory" obesity. *British Medical Journal*, 1, 1009-1011.
- [143]GOLDBERG, J., GARDOS, G. & COLE, J. O. 1973. A controlled evaluation of pyrovalerone in chronically fatigued volunteers. *International pharmacopsychiatry*, 8, 60-69.
- [144]BAUMANN, M. H., WALTERS, H. M., NIELLO, M. & SITTE, H. H. 2018. Neuropharmacology of Synthetic Cathinones. *Handb Exp Pharmacol*, 252, 113-142.
- [145]CARROLL, F. I., BLOUGH, B. E., ABRAHAM, P., MILLS, A. C., HOLLEMAN, J. A., WOLCKENHAUER, S. A., DECKER, A. M., LANDAVAZO, A., MCELROY, K. T., NAVARRO, H. A., GATCH, M. B. & FORSTER, M. J. 2009. Synthesis and biological evaluation of bupropion analogues as potential pharmacotherapies for cocaine addiction. *Journal of Medicinal Chemistry*, 52, 6768-6781.
- [146]HUGHES, J. R., STEAD, L. F., HARTMANN-BOYCE, J., CAHILL, K. & LANCASTER, T. 2014. Antidepressants for smoking cessation. *The Cochrane database of systematic reviews*, 1, CD000031.
- [147]MAGOVERN, M. & CRAWFORD-FAUCHER, A. 2017. Extended-release Bupropion for preventing seasonal affective disorder in adults. *American Family Physician*, 95, 10-11.
- [148]SAUNDERS, K. H., UMASHANKER, D., IGEL, L. I., KUMAR, R. B. & ARONNE, L. J. 2018. Obesity Pharmacotherapy. *Medical Clinics of North America*, 102, 135-148.
- [149]STAHL, S. M., PRADKO, J. F., HAIGHT, B. R., MODELL, J. G., ROCKETT, C. B. & LEARNED-COUGHLIN, S. 2004. A review of the neuropharmacology of bupropion, a dual norepinephrine and dopamine reuptake inhibitor. *Prim Care Companion J Clin Psychiatry*, 6, 159-166.
- [150]VERBEECK, W., BEKKERING, G. E. & VAN DEN NOORTGATE, W. K., C. 2017. Bupropion for Attention Deficit Hyperactivity Disorder (ADHD) in adults. *Cochrane Database of Systematic Reviews*, 10.
- [151]CASTELLS, X., CUNILL, R., PÉREZ-MAÑÁ, C., VIDAL, X. & CAPELLÀ, D. 2016. Psychostimulant drugs for cocaine dependence. *Cochrane Database of Systematic Reviews*.
- [152]ZAWILSKA, J. B. & WOJCIESZAK, J. 2013. Designer cathinones-An emerging class of novel recreational drugs. *Forensic Science International*, 231, 42-53.
- [153]KARILA, L. & BENYAMINA, A. 2018. The effects and risks associated with synthetic cathinone use in humans. *Curr. Topics Neurotox.*, 12, 191-202.
- [154]GONÇALVES, J. L., ALVES, V. L., AGUIAR, J., TEIXEIRA, H. M. & CÂMARA, J. S. 2019. Synthetic cathinones: an evolving class of new psychoactive substances. *Critical Reviews in Toxicology*, 49, 549-566.

- [155]PAPASEIT, E., MOLTÓ, J., MUGA, R., TORRENS, M., DE LA TORRE, R. & FARRÉ, M. 2017. Clinical Pharmacology of the Synthetic Cathinone Mephedrone. *Curr Top Behav Neurosci*, 32, 313-331.
- [156]PROSSER, J. M. & NELSON, L. S. 2012. The Toxicology of Bath Salts: A Review of Synthetic Cathinones. *Journal of Medical Toxicology*, 8, 33-42.
- [157]MCCALL, H., ADAMS, N., MASON, D. & WILLIS, J. 2015. What is chemsex and why does it matter? *BMJ (Online)*, 351.
- [158]STUART, D. 2016. A chemsex crucible: The context and the controversy. *J Fam Plann Reprod Health Care*, 42, 295-296.
- [159]BRANDT, S. D., FREEMAN, S., SUMNALL, H. R., MEASHAM, F. & COLE, J. 2011. Analysis of NRG 'legal highs' in the UK: Identification and formation of novel cathinones. *Drug Testing and Analysis*, 3, 569-575.
- [160]DAVIES, S., WOOD, D. M., SMITH, G., BUTTON, J., RAMSEY, J., ARCHER, R., HOLT, D. W. & DARGAN, P. I. 2010. Purchasing 'legal highs' on the Internet-is there consistency in what you get? *QJM*, 103, 489-493.
- [161]COPPOLA, M., MONDOLA, R., OLIVA, F., PICCI, R. L., ASCHERI, D. & TRIVELLI, F. 2016. Treating the Phenomenon of New Psychoactive Substances: Synthetic Cannabinoids and Synthetic Cathinones. In: PREEDY, V. R. (ed.) *Neuropathology of Drug Addictions and Substance Misuse*. Academic Press.
- [162]HITCHCOCK, S. A. & PENNINGTON, L. D. 2006. Structure-brain exposure relationships. *Journal of Medicinal Chemistry*, 49, 7559-7583.
- [163]PAUL, A. 2019. Drug distribution. *Introduction to Basics of Pharmacology and Toxicology: Volume 1: General and Molecular Pharmacology: Principles of Drug Action*.
- [164]MARTÍNEZ-CLEMENTE, J., LÓPEZ-ARNAU, R., CARBÓ, M., PUBILL, D., CAMARASA, J. & ESCUBEDO, E. 2013. Mephedrone pharmacokinetics after intravenous and oral administration in rats: Relation to pharmacodynamics. *Psychopharmacology*, 229, 295-306.
- [165]LÓPEZ-ARNAU, R., MARTÍNEZ-CLEMENTE, J., CARBÓ, M., PUBILL, D., ESCUBEDO, E. & CAMARASA, J. 2013. An integrated pharmacokinetic and pharmacodynamic study of a new drug of abuse, methylone, a synthetic cathinone sold as "bath salts". *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 45, 64-72.
- [166]DERUNGS, A., SCHIETZEL, S., MEYER, M. R., MAURER, H. H., KRÄHENBÜHL, S. & LIECHTI, M. E. 2011. Sympathomimetic toxicity in a case of analytically confirmed recreational use of naphyrone (naphthylpyrovalerone). *Clinical Toxicology*, 49, 691-693.
- [167]FUJITA, Y., MITA, T., USUI, K., KAMIJO, Y., KIKUCHI, S., ONODERA, M., FUJINO, Y. & INOUE, Y. 2018. Toxicokinetics of the Synthetic Cathinone  $\alpha$ -Pyrrolidinohexanophenone. *Journal of Analytical Toxicology*, 42, e1-e5.
- [168]LAKSHMANAN, M. 2019. Drug metabolism. *Introduction to Basics of Pharmacology and Toxicology: Volume 1: General and Molecular Pharmacology: Principles of Drug Action*.
- [169]HASEGAWA, K., WURITA, A., MINAKATA, K., GONMORI, K., NOZAWA, H., YAMAGISHI, I., WATANABE, K. & SUZUKI, O. 2015. Postmortem distribution of PV9, a new cathinone derivative, in human solid tissues in a fatal poisoning case. *Forensic Toxicology*, 33, 141-147.
- [170]HASEGAWA, K., SUZUKI, O., WURITA, A., MINAKATA, K., YAMAGISHI, I., NOZAWA, H., GONMORI, K. & WATANABE, K. 2014. Postmortem distribution of  $\alpha$ -pyrrolidinovalerophenone and its metabolite in body fluids and solid tissues in a fatal poisoning case measured by LC-MS-MS with the standard addition method. *Forensic Toxicology*, 32, 225-234.

- [171]MARINETTI, L. J. & ANTONIDES, H. M. 2013. Analysis of synthetic cathinones commonly found in bath salts in human performance and postmortem toxicology: Method development, drug distribution and interpretation of results. *Journal of Analytical Toxicology*, 37, 135-146.
- [172]WURITA, A., HASEGAWA, K., MINAKATA, K., GONMORI, K., NOZAWA, H., YAMAGISHI, I., SUZUKI, O. & WATANABE, K. 2014. Postmortem distribution of  $\alpha$ -pyrrolidinobutiophenone in body fluids and solid tissues of a human cadaver. *Legal Medicine*, 16, 241-246.
- [173]TYRKKO, E., ANDERSSON, M. & KRONSTRAND, R. 2016. The toxicology of new psychoactive substances: Synthetic cathinones and phenylethylamines. *Therapeutic Drug Monitoring*, 38, 190-216.
- [174]ZAITSU, K. 2018. Metabolism of synthetic Cathinones. *Synthetic cathinones: novel addictive and stimulatory psychoactive substances*, 71-79.
- [175]WEINSTEIN, A. M., ROSCA, P., FATTORE, L. & LONDON, E. D. 2017. Synthetic cathinone and cannabinoid designer drugs pose a major risk for public health. *Frontiers in Psychiatry*, 8.
- [176]CAPELA, J. P., CARMO, H., REMIÃO, F., BASTOS, M. L., MEISEL, A. & CARVALHO, F. 2009. Molecular and cellular mechanisms of ecstasy-induced neurotoxicity: An overview. *Molecular Neurobiology*, 39, 210-271.
- [177]TSATSAKIS, A., DOCEA, A. O., CALINA, D., TSAROUHAS, K., ZAMFIRA, L. M., MITRUT, R., SHARIFI-RAD, J., KOVATSI, L., SIOKAS, V., DARDIOTIS, E., DRAKOULIS, N., LAZOPOULOS, G., TSITSIMPIKOU, C., MITSIAS, P. & NEAGU, M. 2019. A Mechanistic and pathophysiological approach for stroke associated with drugs of abuse. *Journal of Clinical Medicine*, 8, 1295.
- [178]TURCANT, A., DEGUIGNE, M., FEREC, S., BRUNEAU, C., LEBORGNE, I., LELIEVRE, B., GEGU, C., JEGOU, F., ABBARA, C., LE ROUX, G. & BOELS, D. 2017. A 6-year review of new psychoactive substances at the Centre antipoison Grand-Ouest d'Angers: Clinical and biological data. *Toxicologie Analytique et Clinique*, 29, 18-33.
- [179]WIERGOWSKI, M., ASZYK, J., KALISZAN, M., WILCZEWSKA, K., ANAND, J. S., KOT-WASIK, A. & JANKOWSKI, Z. 2017. Identification of novel psychoactive substances 25B-NBOMe and 4-CMC in biological material using HPLC-Q-TOF-MS and their quantification in blood using UPLC-MS/MS in case of severe intoxications. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 1041-1042, 1-10.
- [180]KRONSTRAND, R., GUERRIERI, D., VIKINGSSON, S., WOHLFARTH, A. & GRÉEN, H. 2018. Fatal poisonings associated with new psychoactive substances. *Handbook of Experimental Pharmacology*.
- [181]KING, A., DIMOVSKA, M. & BISOSKI, L. 2018. Sympathomimetic Toxidromes and Other Pharmacological Causes of Acute Hypertension. *Current Hypertension Reports*, 20.
- [182]PENDERS, T. M. & GESTRING, R. 2011. Hallucinatory delirium following use of MDPV: "Bath Salts". *General Hospital Psychiatry*, 33, 525-526.
- [183]HALL, C. A. 2015. Excited Delirium. *Encyclopedia of Forensic and Legal Medicine: Second Edition*.
- [184]SCOTTON, W. J., HILL, L. J., WILLIAMS, A. C. & BARNES, N. M. 2019. Serotonin Syndrome: Pathophysiology, Clinical Features, Management, and Potential Future Directions. *International Journal of Tryptophan Research*, 12.
- [185]PURVES, D., AUGUSTINE, G. J., FITZPATRICK, D., HALL, W. C., LAMANTIA, A., MCNAMARA, J. O. & WILLIAMS, S. M. 2004. Chapter 5: Synaptic Transmission. *Neuroscience*. 3 ed. Sinauer Associates.

- [186]PURVES, D., AUGUSTINE, G. J., FITZPATRICK, D., HALL, W. C., LAMANTIA, A., MCNAMARA, J. O. & WILLIAMS, S. M. 2004. Chapter 6: Neurotransmitters and Their Receptors. *Neuroscience*. 3 ed. Sinauer Associates.
- [187]PURVES, D., AUGUSTINE, G. J., FITZPATRICK, D., HALL, W. C., LAMANTIA, A., MCNAMARA, J. O. & WILLIAMS, S. M. 2004. Chapter 7: Molecular Signaling within Neurons. *Neuroscience*. 3 ed. Sinauer Associates.
- [188]PETERSON, S. M., URS, N., CARON, M. G. 2012. Chapter 13: Dopamine Receptors. In: URS, N. C., M. G. (ed.) *Primer on the Autonomic Nervous System*. 3rd ed. ed. London: Elsevier.
- [189]HABIBI, M. 2010. Dopamine Receptors. In: KOMPOLITI, K. & METMAN, L. V. (eds.) *Encyclopedia of Movement Disorders*. Oxford: Academic Press.
- [190]NEVE, K. A. 2013. Dopamine Receptors. In: LENNARZ, W. J. & LANE, M. D. (eds.) *Encyclopedia of Biological Chemistry (Second Edition)*. Waltham: Academic Press.
- [191]MARTEL, J. C. & GATTI MCARTHUR, S. 2020. Dopamine Receptor Subtypes, Physiology and Pharmacology: New Ligands and Concepts in Schizophrenia. *Front Pharmacol*, 11, 1003.
- [192]SPRANG, S. R. 2016. Invited review: Activation of G proteins by GTP and the mechanism of G $\alpha$ -catalyzed GTP hydrolysis. *Biopolymers*, 105, 449-62.
- [193]GRESCH, P. J. 2013. Serotonin Receptor Signaling. In: LENNARZ, W. J. & LANE, M. D. (eds.) *Encyclopedia of Biological Chemistry (Second Edition)*. Waltham: Academic Press.
- [194]TANG, W. J. & GILMAN, A. G. 1991. Type-specific regulation of adenylyl cyclase by G protein beta gamma subunits. *Science*, 254, 1500-3.
- [195]KATZ, A., WU, D. & SIMON, M. I. 1992. Subunits  $\beta\gamma$  of heterotrimeric G protein activate  $\beta 2$  isoform of phospholipase C. *Nature*, 360, 686-689.
- [196]STEPHENS, L., SMRCKA, A., COOKE, F. T., JACKSON, T. R., STERNWEIS, P. C. & HAWKINS, P. T. 1994. A novel phosphoinositide 3 kinase activity in myeloid-derived cells is activated by G protein beta gamma subunits. *Cell*, 77, 83-93.
- [197]INGLESE, J., KOCH, W. J., TOUHARA, K. & LEFKOWITZ, R. J. 1995. G beta gamma interactions with PH domains and Ras-MAPK signaling pathways. *Trends Biochem Sci*, 20, 151-6.
- [198]LOGOTHETIS, D. E., KURACHI, Y., GALPER, J., NEER, E. J. & CLAPHAM, D. E. 1987. The beta gamma subunits of GTP-binding proteins activate the muscarinic K<sup>+</sup> channel in heart. *Nature*, 325, 321-6.
- [199]IKEDA, S. R. 1996. Voltage-dependent modulation of N-type calcium channels by G-protein beta gamma subunits. *Nature*, 380, 255-8.
- [200]LANFUMEY, L. & HAMON, M. 2000. Central 5-HT(1A) receptors: regional distribution and functional characteristics. *Nucl Med Biol*, 27, 429-35.
- [201]RAYMOND, J. R., MUKHIN, Y. V., GELASCO, A., TURNER, J., COLLINSWORTH, G., GETTYS, T. W., GREWAL, J. S. & GARNOVSKAYA, M. N. 2001. Multiplicity of mechanisms of serotonin receptor signal transduction. *Pharmacol Ther*, 92, 179-212.
- [202]MASSON, J., EMERIT, M., HAMON, M. D. & DARMON, M. 2012. Serotonergic signaling: Multiple effectors and pleiotropic effects. *Wiley Interdisciplinary Reviews: Membrane Transport and Signaling*, 1, 685-713.
- [203]CHOI, D. S., WARD, S. J., MESSADDEQ, N., LAUNAY, J. M. & MAROTEAUX, L. 1997. 5-HT2B receptor-mediated serotonin morphogenetic functions in mouse cranial neural crest and myocardial cells. *Development*, 124, 1745-55.
- [204]NEBIGIL, C. G., HICKEL, P., MESSADDEQ, N., VONESCH, J. L., DOUCHET, M. P., MONASSIER, L., GYÖRGY, K., MATZ, R., ANDRIANTSITOHAINA, R., MANIVET, P., LAUNAY, J. M. & MAROTEAUX, L. 2001. Ablation of serotonin 5-HT(2B) receptors in mice leads to abnormal cardiac structure and function. *Circulation*, 103, 2973-9.

- [205] NEBIGIL, C. G. & MAROTEAUX, L. 2003. Functional consequence of serotonin/5-HT<sub>2B</sub> receptor signaling in heart: role of mitochondria in transition between hypertrophy and heart failure? *Circulation*, 108, 902-8.
- [206] FOGUET, M., HOYER, D., PARDO, L. A., PAREKH, A., KLUXEN, F. W., KALKMAN, H. O., STÜHMER, W. & LÜBBERT, H. 1992. Cloning and functional characterization of the rat stomach fundus serotonin receptor. *Embo j*, 11, 3481-7.
- [207] KURSAR, J. D., NELSON, D. L., WAINSCOTT, D. B., COHEN, M. L. & BAEZ, M. 1992. Molecular cloning, functional expression, and pharmacological characterization of a novel serotonin receptor (5-hydroxytryptamine<sub>2F</sub>) from rat stomach fundus. *Mol Pharmacol*, 42, 549-57.
- [208] SANDERS-BUSH, E. & BREEDING, M. 1988. Putative selective 5-HT-2 antagonists block serotonin 5-HT-1c receptors in the choroid plexus. *J Pharmacol Exp Ther*, 247, 169-73.
- [209] CONN, P. J. & SANDERS-BUSH, E. 1984. Selective 5HT-2 antagonists inhibit serotonin stimulated phosphatidylinositol metabolism in cerebral cortex. *Neuropharmacology*, 23, 993-6.
- [210] BARNES, N. M., HALES, T. G., LUMMIS, S. C. & PETERS, J. A. 2009. The 5-HT<sub>3</sub> receptor--the relationship between structure and function. *Neuropharmacology*, 56, 273-84.
- [211] MIQUEL, M. C., EMERIT, M. B., NOSJEAN, A., SIMON, A., RUMAJOGEE, P., BRISORGUEIL, M. J., DOUCET, E., HAMON, M. & VERGÉ, D. 2002. Differential subcellular localization of the 5-HT<sub>3A</sub>s receptor subunit in the rat central nervous system. *Eur J Neurosci*, 15, 449-57.
- [212] YAKEL, J. L. & JACKSON, M. B. 1988. 5-HT<sub>3</sub> receptors mediate rapid responses in cultured hippocampus and a clonal cell line. *Neuron*, 1, 615-21.
- [213] KING, B. N., STONER, M. C., HAQUE, S. M. & KELLUM, J. M. 2004. A nitroergic secretomotor neurotransmitter in the chloride secretory response to serotonin. *Dig Dis Sci*, 49, 196-201.
- [214] MOHL, M. C. & GRAHAM, R. M. 2012. Chapter 9 -  $\alpha$ 1-Adrenergic Receptors. In: ROBERTSON, D., BIAGGIONI, I., BURNSTOCK, G., LOW, P. A. & PATON, J. F. R. (eds.) *Primer on the Autonomic Nervous System (Third Edition)*. San Diego: Academic Press.
- [215] BYLUND, D. B. 2004. Adrenergic Receptors. In: LENNARZ, W. J. & LANE, M. D. (eds.) *Encyclopedia of Biological Chemistry*. New York: Elsevier.
- [216] BYLUND, D. B. 2013. Adrenergic Receptors. In: LENNARZ, W. J. & LANE, M. D. (eds.) *Encyclopedia of Biological Chemistry (Second Edition)*. Waltham: Academic Press.
- [217] WANG, Q. 2012. Chapter 10 -  $\alpha$ 2-Adrenergic Receptors. In: ROBERTSON, D., BIAGGIONI, I., BURNSTOCK, G., LOW, P. A. & PATON, J. F. R. (eds.) *Primer on the Autonomic Nervous System (Third Edition)*. San Diego: Academic Press.
- [218] STEIN, C. M. 2012. Chapter 11 -  $\beta$ -Adrenergic Receptors. In: ROBERTSON, D., BIAGGIONI, I., BURNSTOCK, G., LOW, P. A. & PATON, J. F. R. (eds.) *Primer on the Autonomic Nervous System (Third Edition)*. San Diego: Academic Press.
- [219] DROUIN, C., BOBADILLA, A.-C. & TASSIN, J.-P. 2017. Norepinephrine. *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier.
- [220] HAYWARD, L. F., MUELLER, P. J. & HASSER, E. M. 2004. Adrenergic Receptors. In: MARTINI, L. (ed.) *Encyclopedia of Endocrine Diseases*. New York: Elsevier.
- [221] MITRANO, D. A., SCHROEDER, J. P., SMITH, Y., CORTRIGHT, J. J., BUBULA, N., VEZINA, P. & WEINSHENKER, D. 2012.  $\alpha$ -1 Adrenergic receptors are localized on presynaptic elements in the nucleus accumbens and regulate mesolimbic dopamine transmission. *Neuropsychopharmacology*, 37, 2161-72.



- [222] RICHARD, J. E., LÓPEZ-FERRERAS, L., CHANCLÓN, B., EEROLA, K., MICALLEF, P., SKIBICKA, K. P. & WERNSTEDT ASTERHOLM, I. 2017. CNS  $\beta(3)$ -adrenergic receptor activation regulates feeding behavior, white fat browning, and body weight. *Am J Physiol Endocrinol Metab*, 313, E344-e358.
- [223] AGGARWAL, S. & MORTENSEN, O. V. 2017. Overview of Monoamine Transporters. *Curr Protoc Pharmacol*, 79, 12.16.1-12.16.17.
- [224] LIN, Z., CANALES, J. J., BJÖRGVINSSON, T., THOMSEN, M., QU, H., LIU, Q. R., TORRES, G. E. & CAINE, S. B. 2011. Monoamine transporters: vulnerable and vital doorkeepers. *Prog Mol Biol Transl Sci*, 98, 1-46.
- [225] KRISTENSEN, A. S., ANDERSEN, J., JØRGENSEN, T. N., SØRENSEN, L., ERIKSEN, J., LOLAND, C. J., STRØMGAARD, K. & GETHER, U. 2011. SLC6 neurotransmitter transporters: structure, function, and regulation. *Pharmacol Rev*, 63, 585-640.
- [226] ALEXANDER, S. P., KELLY, E., MARRION, N. V., PETERS, J. A., FACCENDA, E., HARDING, S. D., PAWSON, A. J., SHARMAN, J. L., SOUTHAN, C. & DAVIES, J. A. 2017. The concise guide to Pharmacology 2017/18: Transporters. *Br J Pharmacol*, 174 Suppl 1, S360-s446.
- [227] YANG, D. & GOUAUX, E. 2021. Illumination of serotonin transporter mechanism and role of the allosteric site. *Sci Adv*, 7, eabl3857.
- [228] HOWELL, L. L. & NEGUS, S. S. 2014. Monoamine transporter inhibitors and substrates as treatments for stimulant abuse. *Adv Pharmacol*, 69, 129-76.
- [229] SONNERS, M. S., ZHU, S. J., ZAHNISER, N. R., KAVANAUGH, M. P. & AMARA, S. G. 1997. Multiple ionic conductances of the human dopamine transporter: the actions of dopamine and psychostimulants. *J Neurosci*, 17, 960-74.
- [230] SITTE, H. H., HUCK, S., REITHER, H., BOEHM, S., SINGER, E. A. & PIFL, C. 1998. Carrier-mediated release, transport rates, and charge transfer induced by amphetamine, tyramine, and dopamine in mammalian cells transfected with the human dopamine transporter. *J Neurochem*, 71, 1289-97.
- [231] HILBER, B., SCHOLZE, P., DOROSTKAR, M. M., SANDTNER, W., HOLY, M., BOEHM, S., SINGER, E. A. & SITTE, H. H. 2005. Serotonin-transporter mediated efflux: a pharmacological analysis of amphetamines and non-amphetamines. *Neuropharmacology*, 49, 811-9.
- [232] ROBERTSON, S. D., MATTHIES, H. J. & GALLI, A. 2009. A closer look at amphetamine-induced reverse transport and trafficking of the dopamine and norepinephrine transporters. *Mol Neurobiol*, 39, 73-80.
- [233] FLECKENSTEIN, A. E., VOLZ, T. J., RIDDLE, E. L., GIBB, J. W. & HANSON, G. R. 2007. New insights into the mechanism of action of amphetamines. *Annu Rev Pharmacol Toxicol*, 47, 681-98.
- [234] BAUMANN, M. H., BULLING, S., BENADERET, T. S., SAHA, K., AYESTAS, M. A., PARTILLA, J. S., ALI, S. F., STOCKNER, T., ROTHMAN, R. B., SANDTNER, W. & SITTE, H. H. 2014b. Evidence for a role of transporter-mediated currents in the depletion of brain serotonin induced by serotonin transporter substrates. *Neuropsychopharmacology*, 39, 1355-1365.
- [235] KALIX, P. 1980b. A constituent of khat leaves with amphetamine-like releasing properties. *European Journal of Pharmacology*, 68, 213-215.
- [236] KALIX, P. 1980a. Hyperthermic response to (—)-cathinone, an alkaloid of *Catha edulis* (khat). *Journal of Pharmacy and Pharmacology*, 32, 662-663.
- [237] PEHEK, E. A., SCHECHTER, M. D. & YAMAMOTO, B. K. 1990. Effects of cathinone and amphetamine on the neurochemistry of dopamine in vivo. *Neuropharmacology*, 29, 1171-1176.
- [238] ZELGER, J. L. & CARLINI, E. A. 1981. Influence of cathinone ( $\alpha$ -aminopropiophenone) and cathine (phenylpropanolamine) on circling behavior and on the uptake and release of [<sup>3</sup>H]dopamine in striatal slices of rats. *Neuropharmacology*, 20, 839-843.

- [239]WAGNER, G. C., PRESTON, K., RICAURTE, G. A., SCHUSTER, C. R. & SEIDEN, L. S. 1982. Neurochemical similarities between d,l-cathinone and d-amphetamine. *Drug and Alcohol Dependence*, 9, 279-284.
- [240]KALIX, P. 1981. Cathinone, an alkaloid from khat leaves with an amphetamine-like releasing effect. *Psychopharmacology (Berl)*, 74, 269-70.
- [241]KALIX, P. 1982. The amphetamine-like releasing effect of the alkaloid (-)cathinone on rat nucleus accumbens and rabbit caudate nucleus. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 6, 43-49.
- [242]MEREU, G. P., PACITTI, C. & ARGIOLOS, A. 1983. Effect of (-)-cathinone, a khat leaf constituent, on dopaminergic firing and dopamine metabolism in the rat brain. *Life Sci*, 32, 1383-9.
- [243]NIELSEN, J. A. & SCHECHTER, M. D. 1985. Behavioral and neurochemical effects of (-)- and (+/-)-cathinone: dose-response and time-course. *Prog Neuropsychopharmacol Biol Psychiatry*, 9, 739-43.
- [244]KALIX, P. 1984. Effect of the Alkaloid (-)-Cathinone on the Release of Radioactivity from Rat Striatal Tissue Prelabelled with 3H-Serotonin. *Neuropsychobiology*, 12, 127-129.
- [245]SHORTALL, S. E., GREEN, A. R., SWIFT, K. M., FONE, K. C. & KING, M. V. 2013. Differential effects of cathinone compounds and MDMA on body temperature in the rat, and pharmacological characterization of mephedrone-induced hypothermia. *Br J Pharmacol*, 168, 966-77.
- [246]FLECKENSTEIN, A. E., HAUGHEY, H. M., METZGER, R. R., KOKOSKA, J. M., RIDDLE, E. L., HANSON, J. E., GIBB, J. W. & HANSON, G. R. 1999. Differential effects of psychostimulants and related agents on dopaminergic and serotonergic transporter function. *Eur J Pharmacol*, 382, 45-9.
- [247]KALIX, P. 1983. Effect of the alkaloid (-) cathinone on the release of radioactivity from rabbit atria prelabelled with 3H-norepinephrine. *Life Sciences*, 32, 801-807.
- [248]ROTHMAN, R. B., VU, N., PARTILLA, J. S., ROTH, B. L., HUFEISEN, S. J., COMPTON-TOTH, B. A., BIRKES, J., YOUNG, R. & GLENNON, R. A. 2003. In vitro characterization of ephedrine-related stereoisomers at biogenic amine transporters and the receptorome reveals selective actions as norepinephrine transporter substrates. *J Pharmacol Exp Ther*, 307, 138-45.
- [249]CLEARY, L. & DOCHERTY, J. R. 2003. Actions of amphetamine derivatives and cathinone at the noradrenaline transporter. *Eur J Pharmacol*, 476, 31-4.
- [250]GLENNON, R. A. & LIEBOWITZ, S. M. 1982. Serotonin receptor affinity of cathinone and related analogues. *J Med Chem*, 25, 393-7.
- [251]GOŁEMBIOWSKA, K. & KAMIŃSKA, K. 2018. Effects of synthetic cathinones on brain neurotransmitters. *Synthetic Cathinones: Novel Addictive and Stimulatory Psychoactive Substances*, 117-124.
- [252]LISEK, R., XU, W., YUVASHEVA, E., CHIU, Y. T., REITZ, A. B., LIU-CHEN, L. Y. & RAWLS, S. M. 2012. Mephedrone ('bath salt') elicits conditioned place preference and dopamine-sensitive motor activation. *Drug and Alcohol Dependence*, 126, 257-262.
- [253]CAMERON, K. N., KOLANOS, R., SOLIS JR, E., GLENNON, R. A. & DE FELICE, L. J. 2013. Bath salts components mephedrone and methylenedioxypropylvalerone (MDPV) act synergistically at the human dopamine transporter. *British Journal of Pharmacology*, 168, 1750-1757.
- [254]LUETHI, D., KOLACZYNSKA, K. E., DOCCI, L., KRÄHENBÜHL, S., HOENER, M. C. & LIECHTI, M. E. 2018. Pharmacological profile of mephedrone analogs and related new psychoactive substances. *Neuropharmacology*, 134, 4-12.
- [255]ESHLEMAN, A. J., WOLFRUM, K. M., REED, J. F., KIM, S. O., SWANSON, T., JOHNSON, R. A. & JANOWSKY, A. 2017. Structure-activity relationships of substituted cathinones, with transporter binding, uptake, and release. *Journal of Pharmacology and Experimental Therapeutics*, 360, 33-47.

- [256] SAHA, K., PARTILLA, J. S., LEHNER, K. R., SEDDIK, A., STOCKNER, T., HOLY, M., SANDTNER, W., ECKER, G. F., SITTE, H. H. & BAUMANN, M. H. 2015. 'Second-Generation' Mephedrone Analogs, 4-MEC and 4-MePPP, Differentially Affect Monoamine Transporter Function. *Neuropsychopharmacology*, 40, 1321-1331.
- [257] ESHLEMAN, A. J., WOLFRUM, K. M., HATFIELD, M. G., JOHNSON, R. A., MURPHY, K. V. & JANOWSKY, A. 2013. Substituted methcathinones differ in transporter and receptor interactions. *Biochemical Pharmacology*, 85, 1803-1815.
- [258] COZZI, N. V. & FOLEY, K. F. 2003. Methcathinone is a Substrate for the Serotonin Uptake Transporter. *Pharmacology and Toxicology*, 93, 219-225.
- [259] BAUMANN, M. H., PARTILLA, J. S., LEHNER, K. R., THORNDIKE, E. B., HOFFMAN, A. F., HOLY, M., ROTHMAN, R. B., GOLDBERG, S. R., LUPICA, C. R., SITTE, H. H., BRANDT, S. D., TELLA, S. R., COZZI, N. V. & SCHINDLER, C. W. 2013. Powerful cocaine-like actions of 3,4-methylenedioxypropylvalerone (MDPV), a principal constituent of psychoactive 'bath salts' products. *Neuropsychopharmacology*, 38, 552-562.
- [260] MELTZER, P. C., BUTLER, D., DESCHAMPS, J. R. & MADRAS, B. K. 2006. 1-(4-Methylphenyl)-2-pyrrolidin-1-yl-pentan-1-one (pyrovalerone) analogues: A promising class of monoamine uptake inhibitors. *Journal of Medicinal Chemistry*, 49, 1420-1432.
- [261] SHALABI, A. R., WALTHER, D., BAUMANN, M. H. & GLENNON, R. A. 2017. Deconstructed Analogues of Bupropion Reveal Structural Requirements for Transporter Inhibition versus Substrate-Induced Neurotransmitter Release. *ACS Chemical Neuroscience*, 8, 1397-1403.
- [262] ZAWILSKA, J. B. & WOJCIESZAK, J. 2017.  $\alpha$ -Pyrrolidinophenones: a new wave of designer cathinones. *Forensic Toxicology*, 35, 201-216.
- [263] DUMANIAN, G. A., DASCOMBE, W., HONG, C., LABADIE, K., GARRETT, K., SAWHNEY, A. S., PATHAK, C. P., HUBBELL, J. A. & JOHNSON, P. C. 1995. A new photopolymerizable blood vessel glue that seals human vessel anastomoses without augmenting thrombogenicity. *Plast Reconstr Surg*, 95, 901-7.
- [264] CRUISE, G. M., HEGRE, O. D., LAMBERTI, F. V., HAGER, S. R., HILL, R., SHARP, D. S. & HUBBELL, J. A. 1999. In vitro and in vivo performance of porcine islets encapsulated in interfacially photopolymerized poly(ethylene glycol) diacrylate membranes. *Cell Transplant*, 8, 293-306.
- [265] HILL, R. S., CRUISE, G. M., HAGER, S. R., LAMBERTI, F. V., YU, X., GARUFIS, C. L., YU, Y., MUNDWILER, K. E., COLE, J. F., HUBBELL, J. A., HEGRE, O. D. & SHARP, D. W. 1997. Immunoisolation of adult porcine islets for the treatment of diabetes mellitus. The use of photopolymerizable polyethylene glycol in the conformal coating of mass-isolated porcine islets. *Ann N Y Acad Sci*, 831, 332-43.
- [266] KOSTORYZ, E. L., TONG, P. Y., CHAPPELOW, C. C., EICK, J. D., GLAROS, A. G. & YOURTEE, D. M. 1999. In vitro cytotoxicity of solid epoxy-based dental resins and their components. *Dent Mater*, 15, 363-73.
- [267] SHEN, D.-X., LIAN, H.-Z., DING, T., XU, J.-Z. & SHEN, C.-Y. 2009. Determination of low-level ink photoinitiator residues in packaged milk by solid-phase extraction and LC-ESI/MS/MS using triple-quadrupole mass analyzer. *Analytical and Bioanalytical Chemistry*, 395, 2359-2370.
- [268] ROTHENBACHER, T., BAUMANN, M. & FÜGEL, D. 2007. 2-Isopropylthioxanthone (2-ITX) in food and food packaging materials on the German market. *Food Addit Contam*, 24, 438-44.
- [269] KAWASAKI, Y., YAMAJI, K., MATSUNAGA, H. & SENDO, T. 2012. Cytotoxicity of the Polymerization Agent, 2-Methyl-4-(methylthio)-2-morpholinopropiophenone on Human Monocytes. *Biological and Pharmaceutical Bulletin*, 35, 256-259.

- [270]SANCHES-SILVA, A., PASTORELLI, S., CRUZ, J. M., SIMONEAU, C., CASTANHEIRA, I. & PASEIRO-LOSADA, P. 2008. Development of an Analytical Method for the Determination of Photoinitiators Used for Food Packaging Materials with Potential to Migrate into Milk. *Journal of Dairy Science*, 91, 900-909.
- [271]ALLEN, N. S. 1996. Photoinitiators for UV and visible curing of coatings: Mechanisms and properties. *Journal of Photochemistry and Photobiology A: Chemistry*, 100, 101-107.
- [272]SEGUROLA, J., ALLEN, N. S., EDGE, M., MCMAHON, A. & WILSON, S. 1999. Photoyellowing and discolouration of UV cured acrylated clear coatings systems: influence of photoinitiator type. *Polymer Degradation and Stability*, 64, 39-48.
- [273]LAGO, M. A., RODRÍGUEZ-BERNALDO DE QUIRÓS, A., SENDÓN, R., BUSTOS, J., NIETO, M. T. & PASEIRO, P. 2015. Photoinitiators: a food safety review. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*, 32, 779-98.
- [274]COOK, N. & FREEMAN, S. 2001. Report of 19 cases of photoallergic contact dermatitis to sunscreens seen at the Skin and Cancer Foundation. *Australas J Dermatol*, 42, 257-9.
- [275]ALANKO, K., JOLANKI, R., ESTLANDER, T. & KANERVA, L. 2001. Occupational allergic contact dermatitis from benzophenone-4 in hair-care products. *Contact Dermatitis*, 44, 188.
- [276]NEDOROST, S. T. 2003. Facial erythema as a result of benzophenone allergy. *Journal of the American Academy of Dermatology*, 49, 259-261.
- [277](RASFF), R. A. S. F. F. A. F. 2005. Isopropyl thioxanthone in milk for babies from Spain, Alert notification 2005.631.
- [278]SUN, C., CHAN, S. H., LU, D., LEE, H. M. W. & BLOODWORTH, B. C. 2007. Determination of isopropyl-9H-thioxanthone-9-one in packaged beverages by solid-phase extraction clean-up and liquid chromatography with tandem mass spectrometry detection. *Journal of Chromatography A*, 1143, 162-167.
- [279]MORLOCK, G. & SCHWACK, W. 2006. Determination of isopropylthioxanthone (ITX) in milk, yoghurt and fat by HPTLC-FLD, HPTLC-ESI/MS and HPTLC-DART/MS. *Anal Bioanal Chem*, 385, 586-95.
- [280]ELAMIN, A. 2005. Nestlé baby milk recall begins due to concerns over packaging ink. Dairy Reporter.com.
- [281]KAWASAKI, Y., YAGI, K., TSUBOI, C., MORIZANE, M., KITAMURA, Y. & SENDO, T. 2013. The Polymerization Agent, 2-Methyl-4'-(methylthio)-2-morpholinopropiophenone Induces Caspases-3/7 in Human Blood Mononuclear Cells *in Vitro*. *Biological and Pharmaceutical Bulletin*, 36, 1640-1645.
- [282]GARNER, M., TURNER, M. C., GHADIRIAN, P., KREWSKI, D. & WADE, M. 2008. Testicular cancer and hormonally active agents. *J Toxicol Environ Health B Crit Rev*, 11, 260-75.
- [283]HWANG, K. A., PARK, S. H., YI, B. R. & CHOI, K. C. 2011. Gene alterations of ovarian cancer cells expressing estrogen receptors by estrogen and bisphenol a using microarray analysis. *Lab Anim Res*, 27, 99-107.
- [284]TAKAI, M., KAWASAKI, Y., ARIMOTO, S., TANIMOTO, Y., KITAMURA, Y. & SENDO, T. 2018. UV-irradiated 2-methyl-4'-(methylthio)-2-morpholinopropiophenone-containing injection solution produced frameshift mutations in the Ames mutagenicity assay. *Environmental Science and Pollution Research*, 25, 10135-10140.
- [285]KOLANOS, R., SOLIS, E., JR., SAKLOTH, F., DE FELICE, L. J. & GLENNON, R. A. 2013. "Deconstruction" of the Abused Synthetic Cathinone Methylenedioxypropylvalerone (MDPV) and an Examination of Effects at the Human Dopamine Transporter. *ACS Chemical Neuroscience*, 4, 1524-1529.
- [286]COULTER, A. A., REBELLO, C. J. & GREENWAY, F. L. 2018. Centrally Acting Agents for Obesity: Past, Present, and Future. *Drugs*, 78, 1113-1132.

- [287] HENDRICKS, E. J., SRISURAPANONT, M., SCHMIDT, S. L., HAGGARD, M., SOUTER, S., MITCHELL, C. L., DE MARCO, D. G., HENDRICKS, M. J., ISTRATIY, Y. & GREENWAY, F. L. 2014. Addiction potential of phentermine prescribed during long-term treatment of obesity. *Int J Obes (Lond)*, 38, 292-8.
- [288] TALELE, T. T., KHEDKAR, S. A. & RIGBY, A. C. 2010. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr Top Med Chem*, 10, 127-41.
- [289] MARK ANDREW, P., MARISA, A. S., DARBY, L. W. & ZHONG-RU, X. 2018. Has Molecular Docking Ever Brought us a Medicine? In: DIMITRIOS, P. V. (ed.) *Molecular Docking*. Rijeka: IntechOpen.
- [290] Aaftaab, S., KHUSBHOO, J., SASIKALA, K. & MALLIKA, A. 2019. Molecular Docking in Modern Drug Discovery: Principles and Recent Applications. In: VISHWANATH, G., PARTHA, K. & ASHIT, T. (eds.) *Drug Discovery and Development*. Rijeka: IntechOpen.
- [291] BECKER, O. M., DHANOA, D. S., MARANTZ, Y., CHEN, D., SHACHAM, S., CHERUKU, S., HEIFETZ, A., MOHANTY, P., FICHMAN, M., SHARADENDU, A., NUDELMAN, R., KAUFFMAN, M. & NOIMAN, S. 2006. An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT<sub>1A</sub> agonist (PRX-00023) for the treatment of anxiety and depression. *J Med Chem*, 49, 3116-35.
- [292] RICCI, C. G. & NETZ, P. A. 2009. Docking Studies on DNA-Ligand Interactions: Building and Application of a Protocol To Identify the Binding Mode. *Journal of Chemical Information and Modeling*, 49, 1925-1935.
- [293] GILAD, Y. & SENDEROWITZ, H. 2014. Docking Studies on DNA Intercalators. *Journal of Chemical Information and Modeling*, 54, 96-107.
- [294] HALGREN, T. A., MURPHY, R. B., FRIESNER, R. A., BEARD, H. S., FRYE, L. L., POLLARD, W. T. & BANKS, J. L. 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, 47, 1750-1759.
- [295] JONES, G., WILLETT, P., GLEN, R. C., LEACH, A. R. & TAYLOR, R. 1997. Development and validation of a genetic algorithm for flexible docking. Edited by F. E. Cohen. *Journal of Molecular Biology*, 267, 727-748.
- [296] HANDOKO, S. D., OUYANG, X., SU, C. T., KWOH, C. K. & ONG, Y. S. 2012. QuickVina: accelerating AutoDock Vina using gradient-based heuristics for global optimization. *IEEE/ACM Trans Comput Biol Bioinform*, 9, 1266-72.
- [297] FORLI, S., HUEY, R., PIQUE, M. E., SANNER, M. F., GOODSSELL, D. S. & OLSON, A. J. 2016. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols*, 11, 905-919.
- [298] QUIROGA, R. & VILLARREAL, M. A. 2016. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS One*, 11, e0155183.
- [299] RAMÍREZ, D. & CABALLERO, J. 2018. Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules*, 23, 1038.
- [300] ELDRIDGE, M. D., MURRAY, C. W., AUTON, T. R., PAOLINI, G. V. & MEE, R. P. 1997. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*, 11, 425-45.
- [301] WARREN, G. L., ANDREWS, C. W., CAPELLI, A.-M., CLARKE, B., LALONDE, J., LAMBERT, M. H., LINDVALL, M., NEVINS, N., SEMUS, S. F., SENGER, S., TEDESCO, G., WALL, I. D., WOOLVEN, J. M., PEISHOFF, C. E. & HEAD, M. S. 2006. A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry*, 49, 5912-5931.

- [302]LIU, J. & WANG, R. 2015. Classification of Current Scoring Functions. *Journal of Chemical Information and Modeling*, 55, 475-482.
- [303]GOODSELL, D. S., MORRIS, G. M. & OLSON, A. J. 1996. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit*, 9, 1-5.
- [304]WEINER, S. J., KOLLMAN, P. A., CASE, D. A., SINGH, U. C., GHIO, C., ALAGONA, G., PROFETA, S. & WEINER, P. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106, 765-784.
- [305]CORNEILL, W. D., CIEPLAK, P., BAYLY, C. I., GOULD, I. R., MERZ, K. M., FERGUSON, D. M., SPELLMEYER, D. C., FOX, T., CALDWELL, J. W. & KOLLMAN, P. A. 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117, 5179-5197.
- [306]GOHLKE, H., HENDLICH, M. & KLEBE, G. 2000. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*, 295, 337-56.
- [307]VELEC, H., GOHLKE, H. & KLEBE, G. 2005. DrugScore CSD Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *Journal of medicinal chemistry*, 48, 6296-303.
- [308]ZHENG, Z. & MERZ, K. M., JR. 2013. Development of the Knowledge-Based and Empirical Combined Scoring Algorithm (KECSA) To Score Protein-Ligand Interactions. *Journal of Chemical Information and Modeling*, 53, 1073-1083.
- [309]HUANG, S. Y. & ZOU, X. 2006. An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *J Comput Chem*, 27, 1876-82.
- [310]HUANG, S. Y. & ZOU, X. 2006. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J Comput Chem*, 27, 1866-75.
- [311]CHARIFSON, P. S., CORKERY, J. J., MURCKO, M. A. & WALTERS, W. P. 1999. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *Journal of Medicinal Chemistry*, 42, 5100-5109.
- [312]ODA, A., TSUCHIDA, K., TAKAKURA, T., YAMAOTSU, N. & HIRONO, S. 2006. Comparison of Consensus Scoring Strategies for Evaluating Computational Models of Protein-Ligand Complexes. *Journal of Chemical Information and Modeling*, 46, 380-391.
- [313]BÖHM, H. J. 1994. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*, 8, 243-56.
- [314]TROTT, O. & OLSON, A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, 31, 455-61.
- [315]WANG, R., LAI, L. & WANG, S. 2002. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des*, 16, 11-26.
- [316]HUANG, S.-Y. & ZOU, X. 2010. Inclusion of Solvation and Entropy in the Knowledge-Based Scoring Function for Protein-Ligand Interactions. *Journal of Chemical Information and Modeling*, 50, 262-273.
- [317]ZHENG, M., XIONG, B., LUO, C., LI, S., LIU, X., SHEN, Q., LI, J., ZHU, W., LUO, X. & JIANG, H. 2011. Knowledge-Based Scoring Functions in Drug Design: 3. A Two-Dimensional Knowledge-Based Hydrogen-Bonding Potential for the Prediction of Protein-Ligand Interactions. *Journal of Chemical Information and Modeling*, 51, 2994-3004.

- [318]RAPPE, A. K., CASEWIT, C. J., COLWELL, K. S., GODDARD, W. A., III & SKIFF, W. M. 1992. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114, 10024-10035.
- [319]CASEWIT, C. J., COLWELL, K. S. & RAPPE, A. K. 1992. Application of a universal force field to organic molecules. *Journal of the American Chemical Society*, 114, 10035-10046.
- [320]HALGREN, T. A. 1996. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17, 490-519.
- [321]HALGREN, T. A. 1999. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J Comput Chem*, 20, 730-748.
- [322]JORGENSEN, W. L. & TIRADO-RIVES, J. 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc*, 110, 1657-66.
- [323]JORGENSEN, W. L., MAXWELL, D. S. & TIRADO-RIVES, J. 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 118, 11225-11236.
- [324]BANKS, J. L., BEARD, H. S., CAO, Y., CHO, A. E., DAMM, W., FARID, R., FELTS, A. K., HALGREN, T. A., MAINZ, D. T., MAPLE, J. R., MURPHY, R., PHILIPP, D. M., REPASKY, M. P., ZHANG, L. Y., BERNE, B. J., FRIESNER, R. A., GALLICCHIO, E. & LEVY, R. M. 2005. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem*, 26, 1752-80.
- [325]HARDER, E., DAMM, W., MAPLE, J., WU, C., REBOUL, M., XIANG, J. Y., WANG, L., LUPYAN, D., DAHLGREN, M. K., KNIGHT, J. L., KAUS, J. W., CERUTTI, D. S., KRILOV, G., JORGENSEN, W. L., ABEL, R. & FRIESNER, R. A. 2016. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation*, 12, 281-296.
- [326]ROOS, K., WU, C., DAMM, W., REBOUL, M., STEVENSON, J. M., LU, C., DAHLGREN, M. K., MONDAL, S., CHEN, W., WANG, L., ABEL, R., FRIESNER, R. A. & HARDER, E. D. 2019. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J Chem Theory Comput*, 15, 1863-1874.
- [327]SHERMAN, W., DAY, T., JACOBSON, M. P., FRIESNER, R. A. & FARID, R. 2006. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *Journal of Medicinal Chemistry*, 49, 534-553.
- [328]TRAMONTANO, A. 1998. Homology modeling with low sequence identity. *Methods*, 14, 293-300.
- [329]2023. Schrödinger 2023-3: LigPrep. New York, NY: Schrödinger, LLC.
- [330]O'BOYLE, N. M., BANCK, M., JAMES, C. A., MORLEY, C., VANDERMEERSCH, T. & HUTCHISON, G. R. 2011. Open Babel: An open chemical toolbox. *J Cheminform*, 3, 33.
- [331]VAINIO, M. J. & JOHNSON, M. S. 2007. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *Journal of Chemical Information and Modeling*, 47, 2462-2474.
- [332]PURANEN, J. S., VAINIO, M. J. & JOHNSON, M. S. 2010. Accurate conformation-dependent molecular electrostatic potentials for high-throughput in silico drug discovery. *J Comput Chem*, 31, 1722-32.
- [333]TOSCO, P., STIEFL, N. & LANDRUM, G. 2014. Bringing the MMFF force field to the RDKit: Implementation and validation. *Journal of Cheminformatics*, 6, 37.
- [334]SNYDER, P. W., MECINOVIĆ, J., MOUSTAKAS, D. T., THOMAS, S. W., HARDER, M., MACK, E. T., LOCKETT, M. R., HÉROUX, A., SHERMAN, W. & WHITESIDES, G. M. 2011. Mechanism of the hydrophobic effect in the biomolecular

- recognition of arylsulfonamides by carbonic anhydrase. *Proceedings of the National Academy of Sciences*, 108, 17889-17894.
- [335] BALIUS, T. E., FISCHER, M., STEIN, R. M., ADLER, T. B., NGUYEN, C. N., CRUZ, A., GILSON, M. K., KURTZMAN, T. & SHOICHET, B. K. 2017. Testing inhomogeneous solvation theory in structure-based ligand discovery. *Proceedings of the National Academy of Sciences*, 114, E6839-E6846.
- [336] CAPPEL, D., SHERMAN, W. & BEUMING, T. 2017. Calculating Water Thermodynamics in the Binding Site of Proteins - Applications of WaterMap to Drug Discovery. *Curr Top Med Chem*, 17, 2586-2598.
- [337] MADHAVI SASTRY, G., ADZHIGIREY, M., DAY, T., ANNABHIMOJU, R. & SHERMAN, W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 27, 221-234.
- [338] 2023. Schrödinger Release 2023-3: Protein Preparation Wizard; Epik. New York, NY: Schrödinger, LLC.
- [339] MCCAMMON, J. A., GELIN, B. R. & KARPLUS, M. 1977. Dynamics of folded proteins. *Nature*, 267, 585-90.
- [340] WARSHEL, A. & LEVITT, M. 1976. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol*, 103, 227-49.
- [341] OROZCO, M., ORELLANA, L., HOSPITAL, A., NAGANATHAN, A. N., EMPERADOR, A., CARRILLO, O. & GELPÍ, J. L. 2011. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. *Adv Protein Chem Struct Biol*, 85, 183-215.
- [342] DONIACH, S. & EASTMAN, P. 1999. Protein dynamics simulations from nanoseconds to microseconds. *Curr Opin Struct Biol*, 9, 157-63.
- [343] KOMBO, D. C., YOUNG, M. A. & BEVERIDGE, D. L. 2000. One nanosecond molecular dynamics simulation of the N-terminal domain of the lambda repressor protein. *Biopolymers*, 53, 596-605.
- [344] CHOWDHURY, S. & BANSAL, M. 2001. A nanosecond molecular dynamics study of antiparallel d(G)<sub>7</sub> quadruplex structures: effect of the coordinated cations. *J Biomol Struct Dyn*, 18, 647-69.
- [345] NURY, H., POITEVIN, F., VAN RENTERGHEM, C., CHANGEUX, J. P., CORRINGER, P. J., DELARUE, M. & BAADEN, M. 2010. One-microsecond molecular dynamics simulation of channel gating in a nicotinic receptor homologue. *Proc Natl Acad Sci U S A*, 107, 6275-80.
- [346] DON, C. G. & SMIEŠKO, M. 2018. Microsecond MD simulations of human CYP2D6 wild-type and five allelic variants reveal mechanistic insights on the function. *PLoS One*, 13, e0202534.
- [347] SHAW, D., DROR, R., SALMON, J., GROSSMAN, J. P., MACKENZIE, K., BANK, J., YOUNG, C., DENEROFF, M., BATSON, B., BOWERS, K., CHOW, E., EASTWOOD, M., IERARDI, D., KLEPEIS, J., KUSKIN, J., LARSON, R., LINDORFF-LARSEN, K., MARAGAKIS, P., MORAES, M. & TOWLES, B. 2009. *Millisecond-scale molecular dynamics simulations on Anton*.
- [348] LANE, T. J., SHUKLA, D., BEAUCHAMP, K. A. & PANDE, V. S. 2013. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr Opin Struct Biol*, 23, 58-65.
- [349] NOÉ, F. 2015. Beating the millisecond barrier in molecular dynamics simulations. *Biophys J*, 108, 228-9.
- [350] LIU, W., SCHMIDT, B., VOSS, G. & MÜLLER-WITTIG, W. 2008. Accelerating molecular dynamics simulations using Graphics Processing Units with CUDA. *Computer Physics Communications*, 179, 634-641.



- [351]MERMELSTEIN, D. J., LIN, C., NELSON, G., KRETSCH, R., MCCAMMON, J. A. & WALKER, R. C. 2018. Fast and flexible gpu accelerated binding free energy calculations within the amber molecular dynamics package. *J Comput Chem*, 39, 1354-1358.
- [352]KUTZNER, C., PÁLL, S., FECHNER, M., ESZTERMANN, A., DE GROOT, B. L. & GRUBMÜLLER, H. 2015. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J Comput Chem*, 36, 1990-2008.
- [353]ABRAHAM, M., MURTOLA, T., SCHULZ, R., PÁLL, S., SMITH, J., HESS, B. & LINDAHL, E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1.
- [354]MAZUR, J. & JERNIGAN, R. L. 1991. Distance-dependent dielectric constants and their application to double-helical DNA. *Biopolymers*, 31, 1615-29.
- [355]JORGENSEN, W., CHANDRASEKHAR, J., MADURA, J., IMPEY, R. & KLEIN, M. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, 79, 926-935.
- [356]HORN, H. W., SWOPE, W. C., PITERA, J. W., MADURA, J. D., DICK, T. J., HURA, G. L. & HEAD-GORDON, T. 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys*, 120, 9665-78.
- [357]MARK, P. & NILSSON, L. 2001. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A - J PHYS CHEM A*, 105.
- [358]MAHONEY, M. & JORGENSEN, W. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid. *The Journal of Chemical Physics*, 112, 8910-8922.
- [359]CHEN, F. & SMITH, P. E. 2007. Simulated surface tensions of common water models. *J Chem Phys*, 126, 221101.
- [360]SHIVAKUMAR, D., WILLIAMS, J., WU, Y., DAMM, W., SHELLEY, J. & SHERMAN, W. 2010. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *Journal of Chemical Theory and Computation*, 6.
- [361]CHEN, Y. C. 2015. Beware of docking! *Trends Pharmacol Sci*, 36, 78-95.
- [362]SHELLEY, J. C., CHOLLETI, A., FRYE, L. L., GREENWOOD, J. R., TIMLIN, M. R. & UCHIMAYA, M. 2007. Epik: a software program for pK( a ) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des*, 21, 681-91.
- [363]GREENWOOD, J. R., CALKINS, D., SULLIVAN, A. P. & SHELLEY, J. C. 2010. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J Comput Aided Mol Des*, 24, 591-604.
- [364]2023. Schrödinger 2023-3: Epik. New York, NY: Schrödinger, LLC.
- [365]2023. Schrödinger 2023-3: Prime. New York, NY: Schrödinger, LLC.
- [366]JACOBSON, M. P., FRIESNER, R. A., XIANG, Z. & HONIG, B. 2002. On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol*, 320, 597-608.
- [367]JACOBSON, M., PINCUS, D., RAPP, C., DAY, T., HONIG, B., SHAW, D. & FRIESNER, R. 2004. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins*, 55, 351-67.
- [368]KRAMER, B., RAREY, M. & LENGAUER, T. 1999. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins*, 37, 228-41.
- [369]CASTRO-ALVAREZ, A., COSTA, A. M. & VILARRASA, J. 2017. The Performance of Several Docking Programs at Reproducing Protein-Macrolide-Like Crystal Structures. *Molecules*, 22, 136.
- [370]SUN, B., FENG, D., CHU, M. L.-H., FISH, I., LOVERA, S., SANDS, Z. A., KELM, S., VALADE, A., WOOD, M., CESKA, T., KOBILKA, T. S., LEBON, F. & KOBILKA,

- B. K. 2021. Crystal structure of dopamine D1 receptor in complex with G protein and a non-catechol agonist. *Nature Communications*, 12, 3305.
- [371] WANG, S., CHE, T., LEVIT, A., SHOICHET, B. K., WACKER, D. & ROTH, B. L. 2018. Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature*, 555, 269-273.
- [372] XU, P., HUANG, S., MAO, C., KRUMM, B. E., ZHOU, X. E., TAN, Y., HUANG, X. P., LIU, Y., SHEN, D. D., JIANG, Y., YU, X., JIANG, H., MELCHER, K., ROTH, B. L., CHENG, X., ZHANG, Y. & XU, H. E. 2021. Structures of the human dopamine D3 receptor-G(i) complexes. *Mol Cell*, 81, 1147-1159.e4.
- [373] WANG, S., WACKER, D., LEVIT, A., CHE, T., BETZ, R. M., MCCORVY, J. D., VENKATAKRISHNAN, A. J., HUANG, X. P., DROR, R. O., SHOICHET, B. K. & ROTH, B. L. 2017. D4 dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science*, 358, 381-386.
- [374] XU, P., HUANG, S., KRUMM, B. E., ZHUANG, Y., MAO, C., ZHANG, Y., WANG, Y., HUANG, X. P., LIU, Y. F., HE, X., LI, H., YIN, W., JIANG, Y., ZHANG, Y., ROTH, B. L. & XU, H. E. 2023. Structural genomics of the human dopamine receptor system. *Cell Res*, 33, 604-616.
- [375] XU, P., HUANG, S., ZHANG, H., MAO, C., ZHOU, X. E., CHENG, X., SIMON, I. A., SHEN, D. D., YEN, H. Y., ROBINSON, C. V., HARPSØE, K., SVENSSON, B., GUO, J., JIANG, H., GLORIAM, D. E., MELCHER, K., JIANG, Y., ZHANG, Y. & XU, H. E. 2021. Structural insights into the lipid and ligand regulation of serotonin receptors. *Nature*, 592, 469-473.
- [376] WANG, C., JIANG, Y., MA, J., WU, H., WACKER, D., KATRITCH, V., HAN, G. W., LIU, W., HUANG, X. P., VARDY, E., MCCORVY, J. D., GAO, X., ZHOU, X. E., MELCHER, K., ZHANG, C., BAI, F., YANG, H., YANG, L., JIANG, H., ROTH, B. L., CHEREZOV, V., STEVENS, R. C. & XU, H. E. 2013. Structural basis for molecular recognition at serotonin receptors. *Science*, 340, 610-4.
- [377] HUANG, S., XU, P., TAN, Y., YOU, C., ZHANG, Y., JIANG, Y. & XU, H. E. 2021. Structural basis for recognition of anti-migraine drug lasmiditan by the serotonin receptor 5-HT(1F)-G protein complex. *Cell Res*, 31, 1036-1038.
- [378] KIMURA, K. T., ASADA, H., INOUE, A., KADJI, F. M. N., IM, D., MORI, C., ARAKAWA, T., HIRATA, K., NOMURA, Y., NOMURA, N., AOKI, J., IWATA, S. & SHIMAMURA, T. 2019. Structures of the 5-HT(2A) receptor in complex with the antipsychotics risperidone and zotepine. *Nat Struct Mol Biol*, 26, 121-128.
- [379] WACKER, D., WANG, C., KATRITCH, V., HAN, G. W., HUANG, X. P., VARDY, E., MCCORVY, J. D., JIANG, Y., CHU, M., SIU, F. Y., LIU, W., XU, H. E., CHEREZOV, V., ROTH, B. L. & STEVENS, R. C. 2013. Structural features for functional selectivity at serotonin receptors. *Science*, 340, 615-9.
- [380] PENG, Y., MCCORVY, J. D., HARPSØE, K., LANSU, K., YUAN, S., POPOV, P., QU, L., PU, M., CHE, T., NIKOLAJSSEN, L. F., HUANG, X. P., WU, Y., SHEN, L., BJØRN-YOSHIMOTO, W. E., DING, K., WACKER, D., HAN, G. W., CHENG, J., KATRITCH, V., JENSEN, A. A., HANSON, M. A., ZHAO, S., GLORIAM, D. E., ROTH, B. L., STEVENS, R. C. & LIU, Z. J. 2018. 5-HT(2C) Receptor Structures Reveal the Structural Basis of GPCR Polypharmacology. *Cell*, 172, 719-730.e14.
- [381] HUANG, S., XU, P., SHEN, D. D., SIMON, I. A., MAO, C., TAN, Y., ZHANG, H., HARPSØE, K., LI, H., ZHANG, Y., YOU, C., YU, X., JIANG, Y., ZHANG, Y., GLORIAM, D. E. & XU, H. E. 2022. GPCRs steer G(i) and G(s) selectivity via TM5-TM6 switches as revealed by structures of serotonin receptors. *Mol Cell*, 82, 2681-2695.e6.
- [382] ZHANG, S., CHEN, H., ZHANG, C., YANG, Y., POPOV, P., LIU, J., KRUMM, B. E., CAO, C., KIM, K., XIONG, Y., KATRITCH, V., SHOICHET, B. K., JIN, J., FAY, J. F. & ROTH, B. L. 2022. Inactive and active state structures template selective tools for the human 5-HT(5A) receptor. *Nat Struct Mol Biol*, 29, 677-687.

- [383]SU, M., WANG, J., XIANG, G., DO, H. N., LEVITZ, J., MIAO, Y. & HUANG, X.-Y. 2023. Structural basis of agonist specificity of  $\alpha$ 1A-adrenergic receptor. *Nature Communications*, 14, 4819.
- [384]XU, J., CAO, S., HÜBNER, H., WEIKERT, D., CHEN, G., LU, Q., YUAN, D., GMEINER, P., LIU, Z. & DU, Y. 2022. Structural insights into ligand recognition, activation, and signaling of the  $\alpha$ (2A) adrenergic receptor. *Sci Adv*, 8, eabj5347.
- [385]YUAN, D., LIU, Z., KAINDL, J., MAEDA, S., ZHAO, J., SUN, X., XU, J., GMEINER, P., WANG, H. W. & KOBILKA, B. K. 2020. Activation of the  $\alpha$ (2B) adrenoceptor by the sedative sympatholytic dexmedetomidine. *Nat Chem Biol*, 16, 507-512.
- [386]XU, X., KAINDL, J., CLARK, M. J., HÜBNER, H., HIRATA, K., SUNAHARA, R. K., GMEINER, P., KOBILKA, B. K. & LIU, X. 2021. Binding pathway determines norepinephrine selectivity for the human  $\beta$ (1)AR over  $\beta$ (2)AR. *Cell Res*, 31, 569-579.
- [387]RASMUSSEN, S. G., DEVREE, B. T., ZOU, Y., KRUSE, A. C., CHUNG, K. Y., KOBILKA, T. S., THIAN, F. S., CHAE, P. S., PARDON, E., CALINSKI, D., MATHIESEN, J. M., SHAH, S. T., LYONS, J. A., CAFFREY, M., GELLMAN, S. H., STEYAERT, J., SKINIOTIS, G., WEIS, W. I., SUNAHARA, R. K. & KOBILKA, B. K. 2011. Crystal structure of the  $\beta$ 2 adrenergic receptor-Gs protein complex. *Nature*, 477, 549-55.
- [388]NAGIRI, C., KOBAYASHI, K., TOMITA, A., KATO, M., KOBAYASHI, K., YAMASHITA, K., NISHIZAWA, T., INOUE, A., SHIHOYA, W. & NUREKI, O. 2021. Cryo-EM structure of the  $\beta$ 3-adrenergic receptor reveals the molecular basis of subtype selectivity. *Mol Cell*, 81, 3205-3215.e5.
- [389]WOOD, P., ALLEN, F. & PIDCOCK, E. 2009. Hydrogen-bond directionality at the donor H atom - Analysis of interaction energies and database statistics. *CrystEngComm*, 11, 1563-1571.
- [390]DANNENBERG, J. J. 1998. An Introduction to Hydrogen Bonding By George A. Jeffrey (University of Pittsburgh). Oxford University Press: New York and Oxford. 1997. ix + 303 pp. \$60.00. ISBN 0-19-509549-9. *Journal of the American Chemical Society*, 120, 5604-5604.
- [391]ANJANA, R., VAISHNAVI, M. K., SHERLIN, D., KUMAR, S. P., NAVEEN, K., KANTH, P. S. & SEKAR, K. 2012. Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. *Bioinformation*, 8, 1220-4.
- [392]MEYER, E. A., CASTELLANO, R. K. & DIEDERICH, F. 2003. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angewandte Chemie International Edition*, 42, 1210-1250.
- [393]KUMAR, S. & NUSSINOV, R. 2002. Close-range electrostatic interactions in proteins. *ChemBiochem*, 3, 604-17.
- [394]SIPPEL, K. H. & QUIOCHO, F. A. 2015. Ion-dipole interactions and their functions in proteins. *Protein Sci*, 24, 1040-6.
- [395]HASEGAWA, K. 2012. Introduction to single crystal X-ray analysis. *The Rigaku Journal*, 28, 14-18.
- [396]RÖNTGEN, W. C. 1896. ON A NEW KIND OF RAYS. *Science*, 3, 227-31.
- [397]TUBIANA, M. 1996. [Wilhelm Conrad Röntgen and the discovery of X-rays]. *Bull Acad Natl Med*, 180, 97-108.
- [398]ECKERT, M. 2012. Max von Laue and the discovery of X-ray diffraction in 1912. *Annalen der Physik*, 524, 83-A85.
- [399]AUTHIER, A. 2013. The Discovery Of X-Ray Diffraction and the Birth of X-Ray Analysis. In: AUTHIER, A. (ed.) *Early Days of X-ray Crystallography*. Oxford University Press.
- [400]MARTIRADONNA, L. 2014. A photograph of crystal order. *Nature*, 511, 7-7.
- [401]BRAGG, W. L. 1913. The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society*, 17, 43-57.
- [402]HODGKIN, D. C. 1949. The X-ray analysis of the structure of penicillin. *Adv Sci*, 6, 85-9.

- [403]BRINK, C., HODGKIN, D. C., LINDSEY, J., PICKWORTH, J., ROBERTSON, J. R. & WHITE, J. G. 1954. X-ray crystallographic evidence on the structure of vitamin B12. *Nature*, 174, 1169-71.
- [404]ADAMS, M. J., BLUNDELL, T. L., DODSON, E. J., DODSON, G. G., VIJAYAN, M., BAKER, E. N., HARDING, M. M., HODGKIN, D. C., RIMMER, B. & SHEAT, S. 1969. Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature*, 224, 491-495.
- [405]KENDREW, J. C., BODO, G., DINTZIS, H. M., PARRISH, R. G., WYCKOFF, H. & PHILLIPS, D. C. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181, 662-6.
- [406]PERUTZ, M. F., ROSSMANN, M. G., CULLIS, A. F., MUIRHEAD, H., WILL, G. & NORTH, A. C. T. 1960. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature*, 185, 416-422.
- [407]WATSON, J. D. & CRICK, F. H. C. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171, 737-738.
- [408]HODGKIN, D. C. 1971. X Rays And The Structure Of Insulin. *The British Medical Journal*, 4, 447-451.
- [409]BRAGG, W. H., BRAGG, W. L., JAMES, R. W. & LIPSON, H. 1934. *The crystalline state*, New York, Macmillan New York.
- [410]TAYLOR, G. L. 2010. Introduction to phasing. *Acta Crystallogr D Biol Crystallogr*, 66, 325-38.
- [411]SCHENK, H. An Introduction to Direct Methods . The Most Important Phase Relationships and their Application in Solving the Phase Problem. 2001.
- [412]LE PEVELEN, D. D. 2010. Small Molecule X-Ray Crystallography, Theory and Workflow. In: LINDON, J. C. (ed.) *Encyclopedia of Spectroscopy and Spectrometry (Second Edition)*. Oxford: Academic Press.
- [413]SPAGNA, R. 2005. Crystal Structure Determination. In: BASSANI, F., LIEDL, G. L. & WYDER, P. (eds.) *Encyclopedia of Condensed Matter Physics*. Oxford: Elsevier.
- [414]BRÜNGER, A. T. 1997. [32] Patterson correlation searches and refinement. *Methods in Enzymology*. Academic Press.
- [415]WEEKS, C. M., SHELDRIK, G., MILLER, R., US'N, I. & HAUPTMA, H. A. 2001. Ab initio phasing by dual-space direct methods.
- [416]PALATINUS, L. 2013. The charge-flipping algorithm in crystallography. *Acta Crystallogr B*, 69, 1-16.
- [417]SHELDRIK, G. M. 2015. SHELXT - integrated space-group and crystal-structure determination. *Acta Crystallogr A Found Adv*, 71, 3-8.
- [418]GIACOVAZZO, C. 2013. Isomorphous replacement techniques. *Phasing in Crystallography: A Modern Perspective*. Oxford University Press.
- [419]YAMASHITA, K., PAN, D., OKUDA, T., SUGAHARA, M., KODAN, A., YAMAGUCHI, T., MURAI, T., GOMI, K., KAJIYAMA, N., MIZOHATA, E., SUZUKI, M., NANGO, E., TONO, K., JOTI, Y., KAMESHIMA, T., PARK, J., SONG, C., HATSUI, T., YABASHI, M., IWATA, S., KATO, H., AGO, H., YAMAMOTO, M. & NAKATSU, T. 2015. An isomorphous replacement method for efficient de novo phasing for serial femtosecond crystallography. *Scientific Reports*, 5, 14017.
- [420]YAMANO, A. 2012. Introduction to single crystal X-ray analysis II. Mounting crystals. *The Rigaku Journal*, 28, 1-4.
- [421]FARRUGIA, L. J. 1999. WinGX suite for small-molecule single-crystal crystallography. *Journal of Applied Crystallography*, 32, 837-838.
- [422]DOLOMANOV, O., BOURHIS, L., GILDEA, R., HOWARD, J. & PUSCHMANN, H. 2009. OLEX2: A complete structure solution, refinement and analysis program. *J. Appl. Cryst. J. Appl. Cryst*, 42, 339-341.
- [423]SHELDRIK, G. M. 2008. A short history of SHELX. *Acta Crystallogr A*, 64, 112-22.

- [424]MACRAE, C. F., SOVAGO, I., COTTRELL, S. J., GALEK, P. T. A., MCCABE, P., PIDCOCK, E., PLATINGS, M., SHIELDS, G. P., STEVENS, J. S., TOWLER, M. & WOOD, P. A. 2020. Mercury 4.0: from visualization to analysis, design and prediction. *J Appl Crystallogr*, 53, 226-235.
- [425]HOFFMANN, F. 2020. *Introduction to Crystallography*, Gewerbestrasse 11, 6330 Cham, Switzerland, Springer Nature Switzerland
- [426]RIGAKU 2014. CrysAlisPRO. 1.171.41.97a ed. Yarnton, Oxfordshire, England: Agilent Technologies UK Ltd.
- [427]2005. SCALE3 ABSPACK - An Oxford Diffraction program Oxford Diffraction Ltd.
- [428]HERBST-IRMER, R., MÜLLER, P., HERBST-IRMER, R., SPEK, A. L., SCHNEIDER, T. R. & SAWAYA, M. R. 2006. Twinning. *Crystal Structure Refinement: A Crystallographer's Guide to SHELXL*. Oxford University Press.
- [429]BÖHM, H.-J., BRODE, S., HESSE, U. & KLEBE, G. 1996. Oxygen and Nitrogen in Competitive Situations: Which is the Hydrogen-Bond Acceptor? *Chemistry – A European Journal*, 2, 1509-1513.
- [430]LEVITT, M. & PERUTZ, M. F. 1988. Aromatic rings act as hydrogen bond acceptors. *Journal of Molecular Biology*, 201, 751-754.
- [431]WILLIAMS, D. E. & STARR, T. L. 1977. Calculation of the crystal structures of hydrocarbons by molecular packing analysis. *Computers & Chemistry*, 1, 173-177.
- [432]COCKROFT, S. L., HUNTER, C. A., LAWSON, K. R., PERKINS, J. & URCH, C. J. 2005. Electrostatic Control of Aromatic Stacking Interactions. *Journal of the American Chemical Society*, 127, 8594-8595.

# **Annexures**

# Simulation Interactions Diagram Report

## Simulation Details

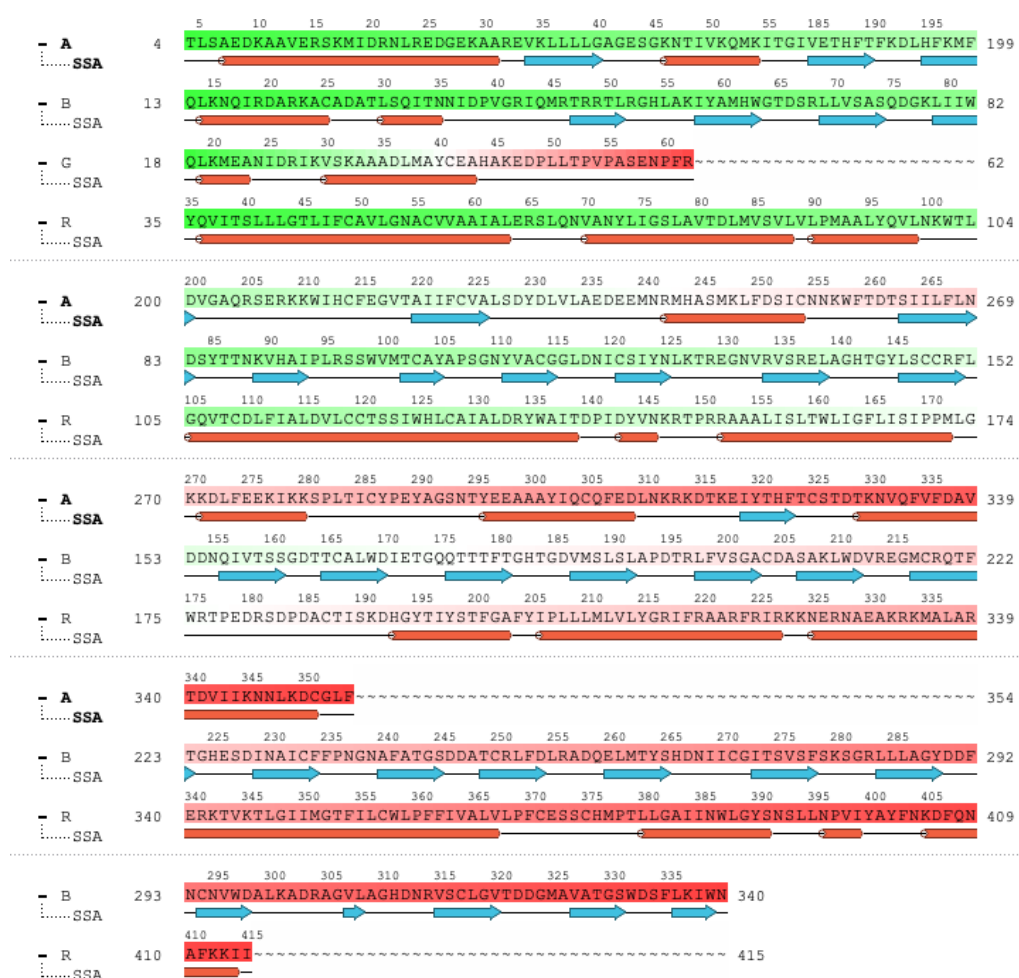
Jobname: merged  
Entry title: Ligand

| CPU #    | Job Type | Ensemble | Temp. [K] | Sim. Time [ns] | # Atoms | # Waters | Charge |
|----------|----------|----------|-----------|----------------|---------|----------|--------|
| Unknown* | Unknown* | Unknown* | 300.0     | 200.404        | 129212  | 38357    | 0      |

\* The configuration file (-out.cfg) was not found. Keep it in same directory as .aef file.

## Protein Information

| Tot. Residues | Prot. Chain(s)     | Res. in Chain(s)             | # Atoms | # Heavy Atoms | Charge |
|---------------|--------------------|------------------------------|---------|---------------|--------|
| 884           | 'A', 'B', 'G', 'R' | Residues (225, 328, 45, 286) | 3897    | 6941          | +3     |

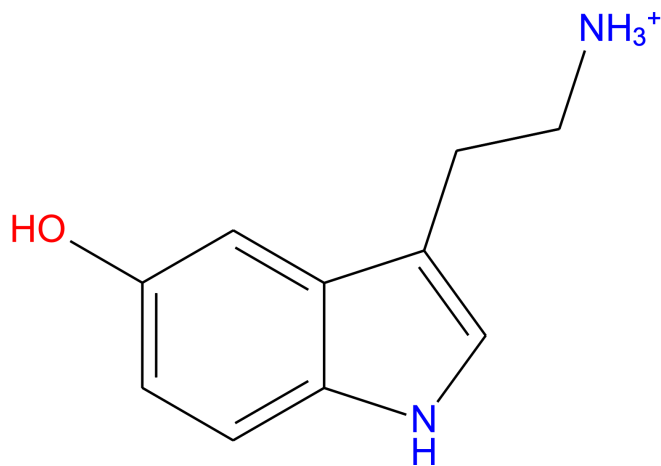


## Ligand Information

SMILES

[NH3+]CCc1c[nH]c(c12)ccc(c2)O

PDB Name 'SRO', 'UNK'  
Num. of Atoms 26 (total) 13 (heavy)  
Atomic Mass 177.228 au  
Charge +1  
Mol. Formula C<sub>10</sub>H<sub>13</sub>N<sub>2</sub>O  
Num. of Fragments 2  
Num. of Rot. Bonds 3

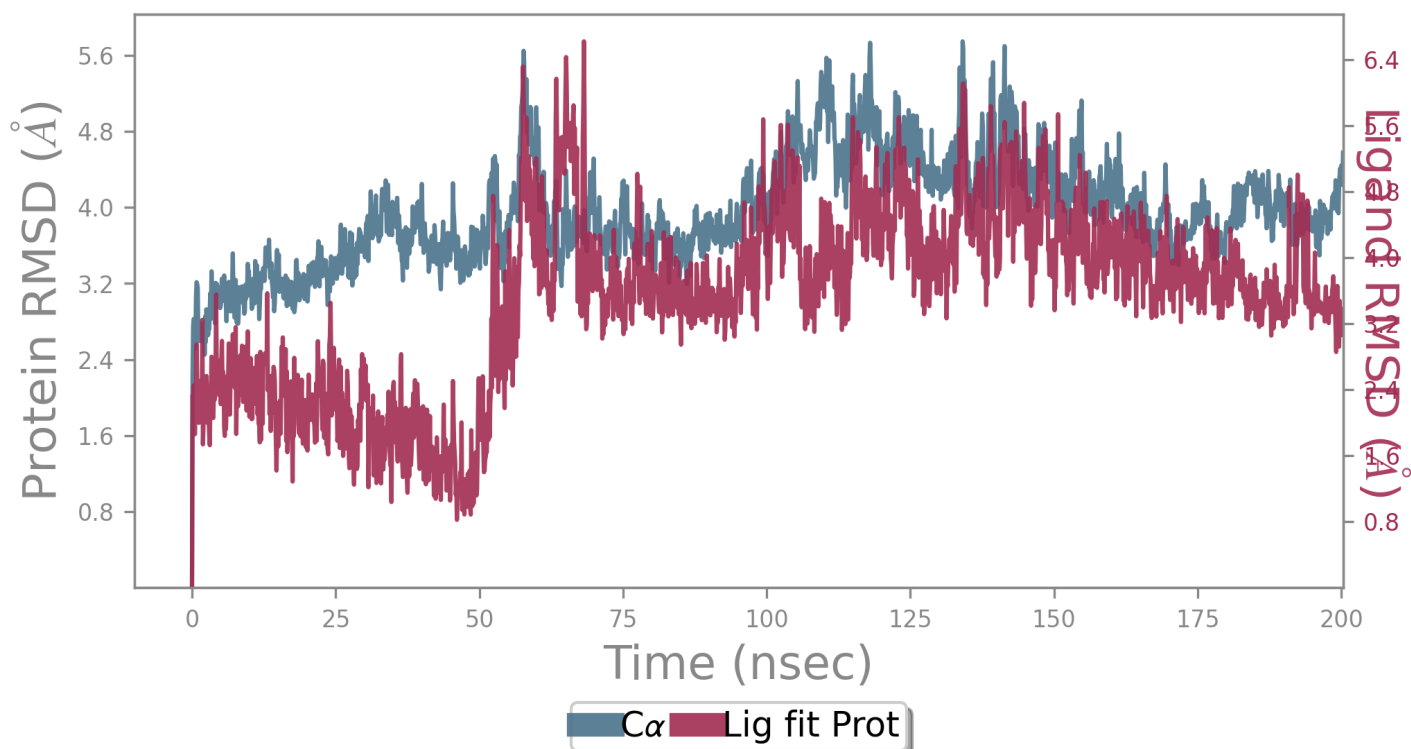


### Counter Ion/Salt Information

| Type | Num. | Concentration [mM] | Total Charge |
|------|------|--------------------|--------------|
| Cl   | 111  | 52.616             | -111         |
| Na   | 107  | 50.720             | +107         |



## Protein-Ligand RMSD



The Root Mean Square Deviation (RMSD) is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The RMSD for frame  $x$  is:

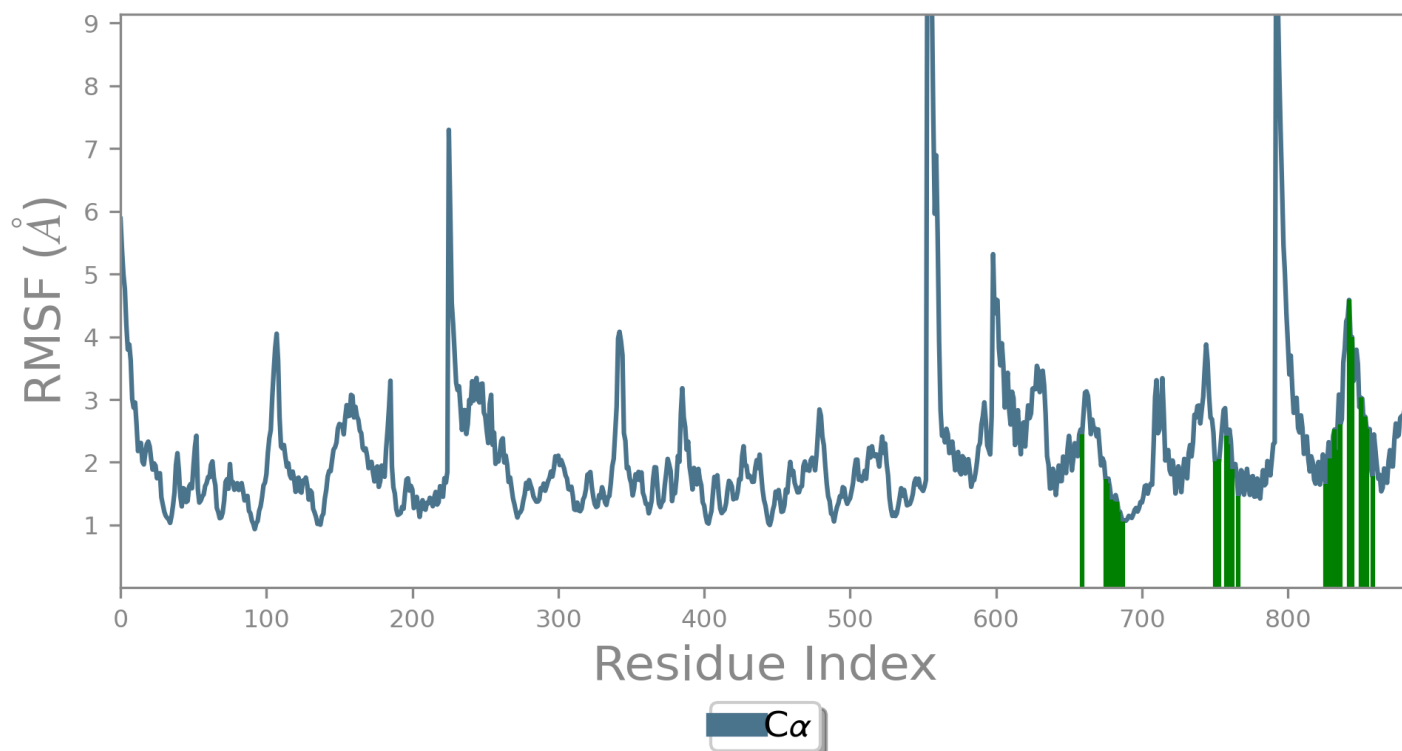
$$RMSD_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'_i(t_x) - r_i(t_{ref}))^2}$$

where  $N$  is the number of atoms in the atom selection;  $t_{ref}$  is the reference time, (typically the first frame is used as the reference and it is regarded as time  $t=0$ ); and  $r'_i$  is the position of the selected atoms in frame  $x$  after superimposing on the reference frame, where frame  $x$  is recorded at time  $t_x$ . The procedure is repeated for every frame in the simulation trajectory.

**Protein RMSD:** The above plot shows the RMSD evolution of a protein (left Y-axis). All protein frames are first aligned on the reference frame backbone, and then the RMSD is calculated based on the atom selection. Monitoring the RMSD of the protein can give insights into its structural conformation throughout the simulation. RMSD analysis can indicate if the simulation has equilibrated — its fluctuations towards the end of the simulation are around some thermal average structure. **Changes of the order of 1-3 Å are perfectly acceptable for small, globular proteins.** Changes much larger than that, however, indicate that the protein is undergoing a large conformational change during the simulation. **It is also important that your simulation converges — the RMSD values stabilize around a fixed value. If the RMSD of the protein is still increasing or decreasing on average at the end of the simulation, then your system has not equilibrated, and your simulation may not be long enough for rigorous analysis.**

**Ligand RMSD:** Ligand RMSD (right Y-axis) indicates how stable the ligand is with respect to the protein and its binding pocket. In the above plot, 'Lig fit Prot' shows the RMSD of a ligand when the protein-ligand complex is first aligned on the protein backbone of the reference and then the RMSD of the ligand heavy atoms is measured. If the values observed are significantly larger than the RMSD of the protein, then it is likely that the ligand has diffused away from its initial binding site.

## Protein RMSF



The Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain. The RMSF for residue  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \langle (r'_i(t) - r_i(t_{ref}))^2 \rangle}$$

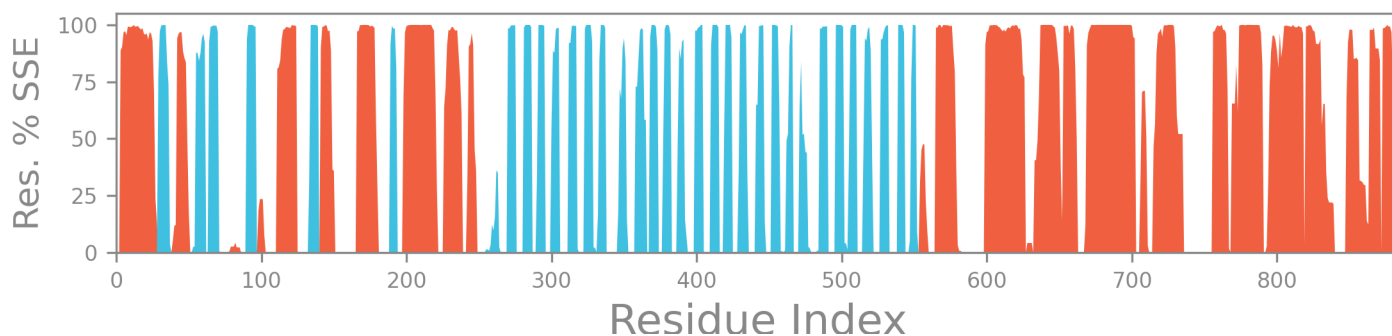
where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time,  $r_i$  is the position of residue  $i$ ;  $r'_i$  is the position of atoms in residue  $i$  after superposition on the reference, and the  $i$  angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

On this plot, peaks indicate areas of the protein that fluctuate the most during the simulation. Typically you will observe that the tails ( $N$ - and  $C$ -terminal) fluctuate more than any other part of the protein. Secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein, and thus fluctuate less than the loop regions.

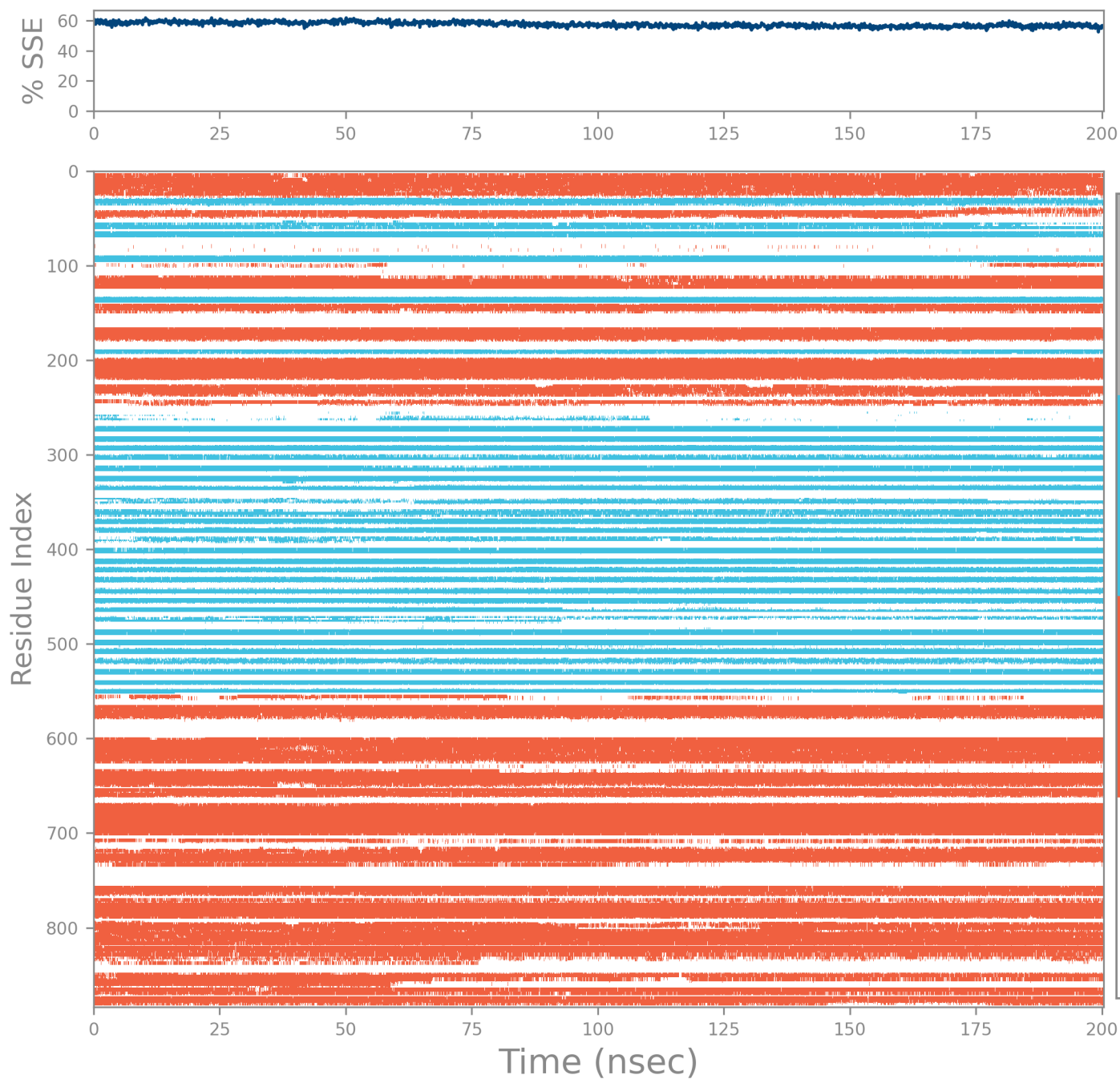
Ligand Contacts: Protein residues that interact with the ligand are marked with green-colored vertical bars.

# Protein Secondary Structure

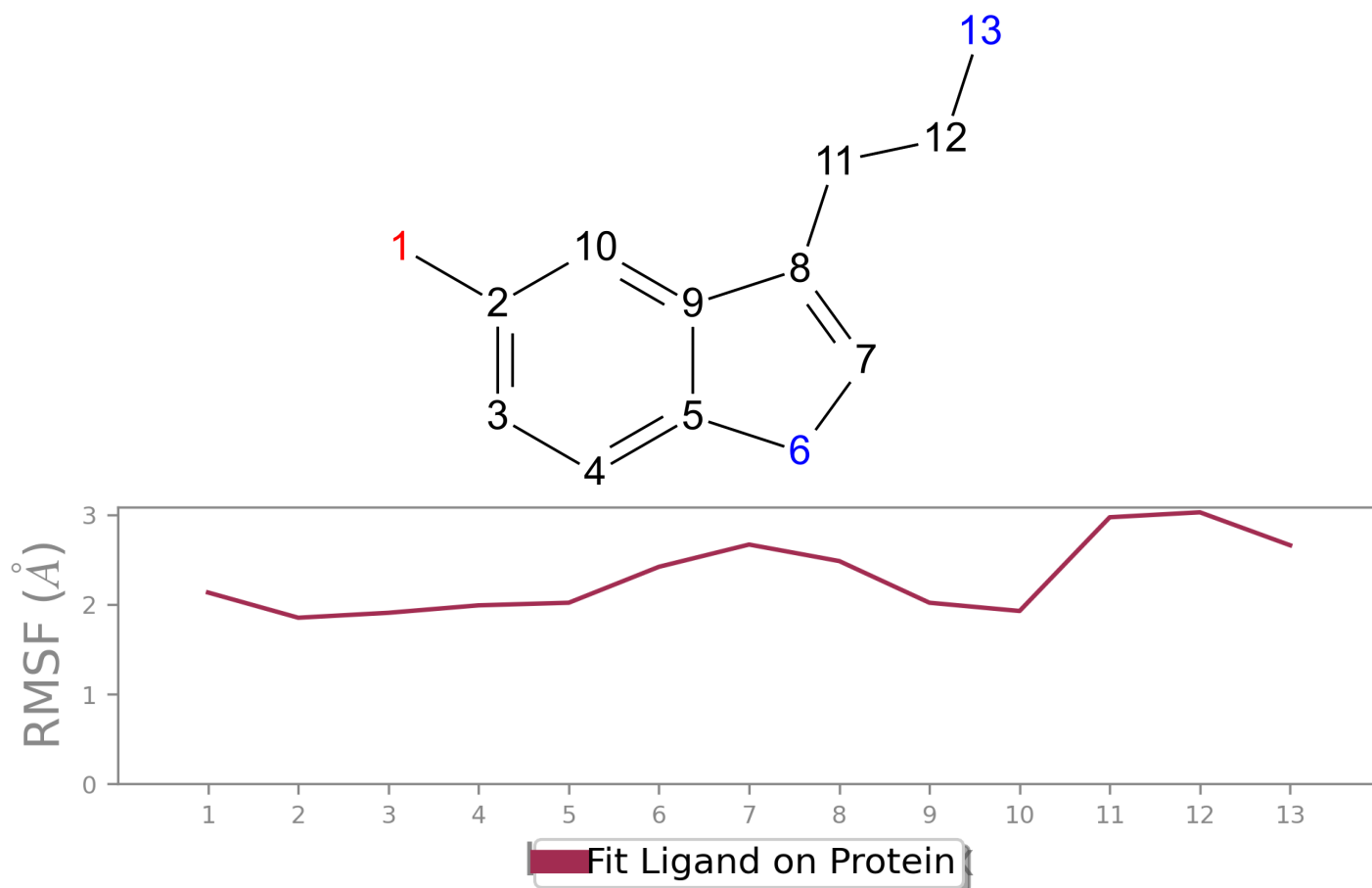
|         |          |             |
|---------|----------|-------------|
| % Helix | % Strand | % Total SSE |
| 36.53   | 20.87    | 57.40       |



Protein secondary structure elements (SSE) like **alpha-helices** and **beta-strands** are monitored throughout the simulation. The plot above reports SSE distribution by residue index throughout the protein structure. The plot below summarizes the SSE composition for each trajectory frame over the course of the simulation, and the plot at the bottom monitors each residue and its SSE assignment over time.



## Ligand RMSF



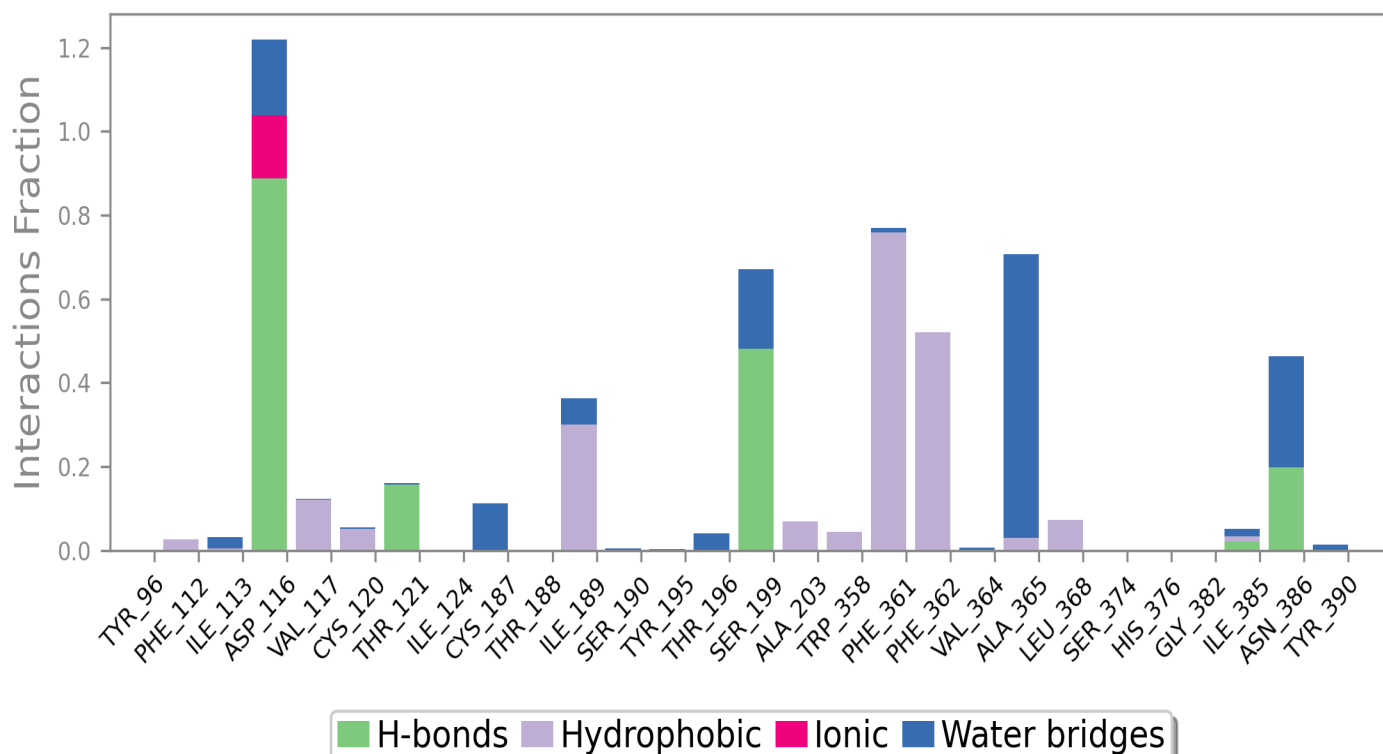
The Ligand Root Mean Square Fluctuation (L-RMSF) is useful for characterizing changes in the ligand atom positions. The RMSF for atom  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (r'_i(t) - r_i(t_{ref}))^2}$$

where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time (usually for the first frame, and is regarded as the *zero* of time);  $r$  is the position of atom  $i$  in the reference at time  $t_{ref}$  and  $r'$  is the position of atom  $i$  at time  $t$  after superposition on the reference frame.

Ligand RMSF shows the ligand's fluctuations broken down by atom, corresponding to the 2D structure in the top panel. The ligand RMSF may give you insights on how ligand fragments interact with the protein and their entropic role in the binding event. In the bottom panel, the 'Fit Ligand on Protein' line shows the ligand fluctuations, with respect to the protein. The protein-ligand complex is first aligned on the protein backbone and then the ligand RMSF is measured on the ligand heavy atoms.

## Protein-Ligand Contacts



Protein interactions with the ligand can be monitored throughout the simulation. These interactions can be categorized by type and summarized, as shown in the plot above. Protein-ligand interactions (or 'contacts') are categorized into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges. Each interaction type contains more specific subtypes, which can be explored through the 'Simulation Interactions Diagram' panel. The stacked bar charts are normalized over the course of the trajectory: for example, a value of 0.7 suggests that 70% of the simulation time the specific interaction is maintained. Values over 1.0 are possible as some protein residue may make multiple contacts of same subtype with the ligand.

**Hydrogen Bonds:** (H-bonds) play a significant role in ligand binding. Consideration of hydrogen-bonding properties in drug design is important because of their strong influence on drug specificity, metabolism and adsorption. Hydrogen bonds between a protein and a ligand can be further broken down into four subtypes: backbone acceptor; backbone donor; side-chain acceptor; side-chain donor.

The current geometric criteria for protein-ligand H-bond is: distance of 2.5 Å between the donor and acceptor atoms (D—H···A); a donor angle of  $\geq 120^\circ$  between the donor-hydrogen-acceptor atoms (D—H···A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H···A—X).

**Hydrophobic contacts:** fall into three subtypes:  $\pi$ -Cation;  $\pi$ - $\pi$ ; and Other, non-specific interactions. Generally these type of interactions involve a hydrophobic amino acid and an aromatic or aliphatic group on the ligand, but we have extended this category to also include  $\pi$ -Cation interactions.

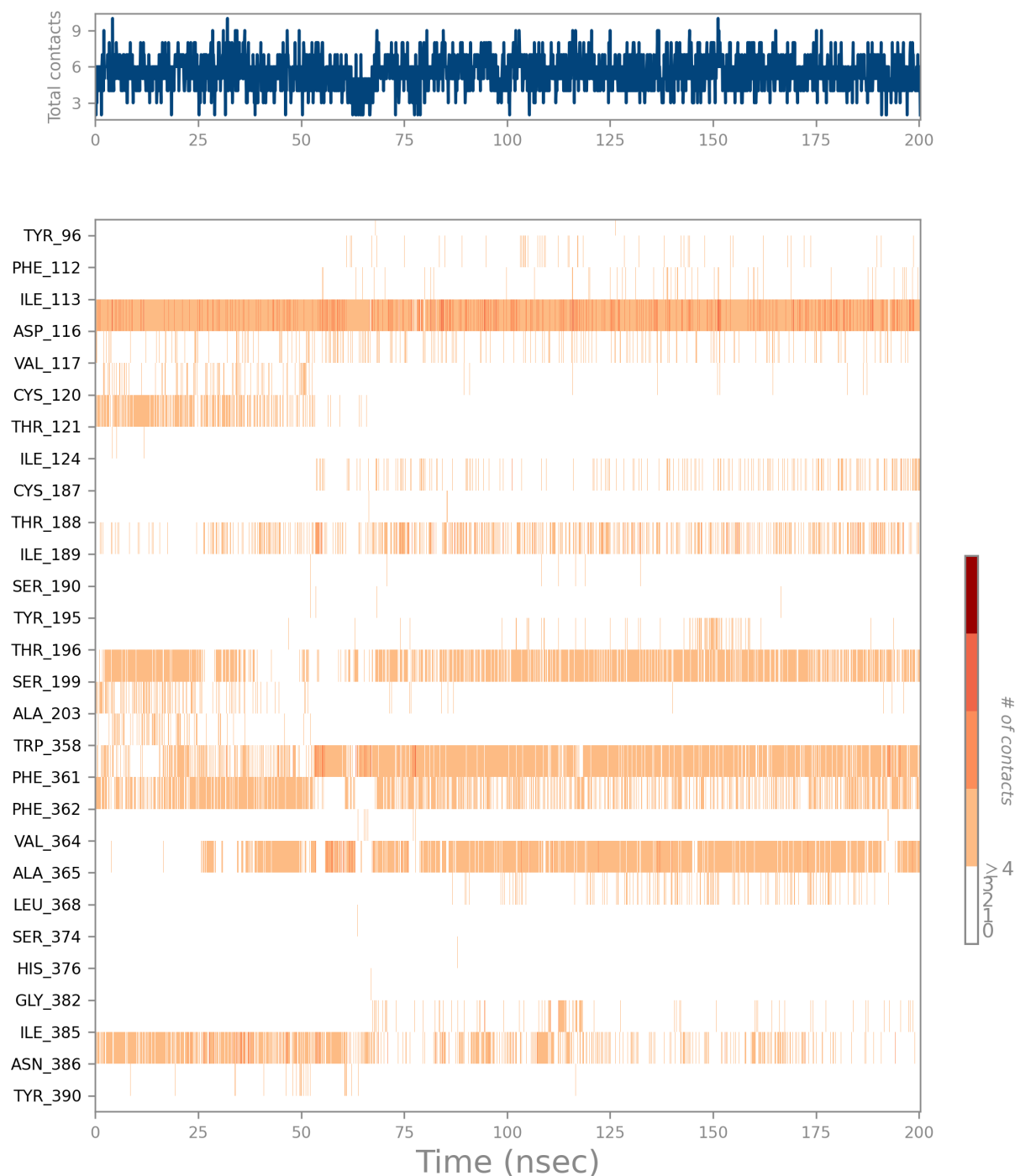
The current geometric criteria for hydrophobic interactions is as follows:  $\pi$ -Cation — Aromatic and charged groups within 4.5 Å;  $\pi$ - $\pi$  — Two aromatic groups stacked face-to-face or face-to-edge; Other — A non-specific hydrophobic sidechain within 3.6 Å of a ligand's aromatic or aliphatic carbons.

**Ionic interactions:** or polar interactions, are between two oppositely charged atoms that are within 3.7 Å of each other and do not involve a hydrogen bond. We also monitor Protein-Metal-Ligand interactions, which are defined by a metal ion coordinated within 3.4 Å of protein's and ligand's heavy atoms (except carbon). All ionic interactions are broken down into two subtypes: those mediated by a protein backbone or side chains.

**Water Bridges:** are hydrogen-bonded protein-ligand interactions mediated by a water molecule. The hydrogen-bond geometry is slightly relaxed from the standard H-bond definition.

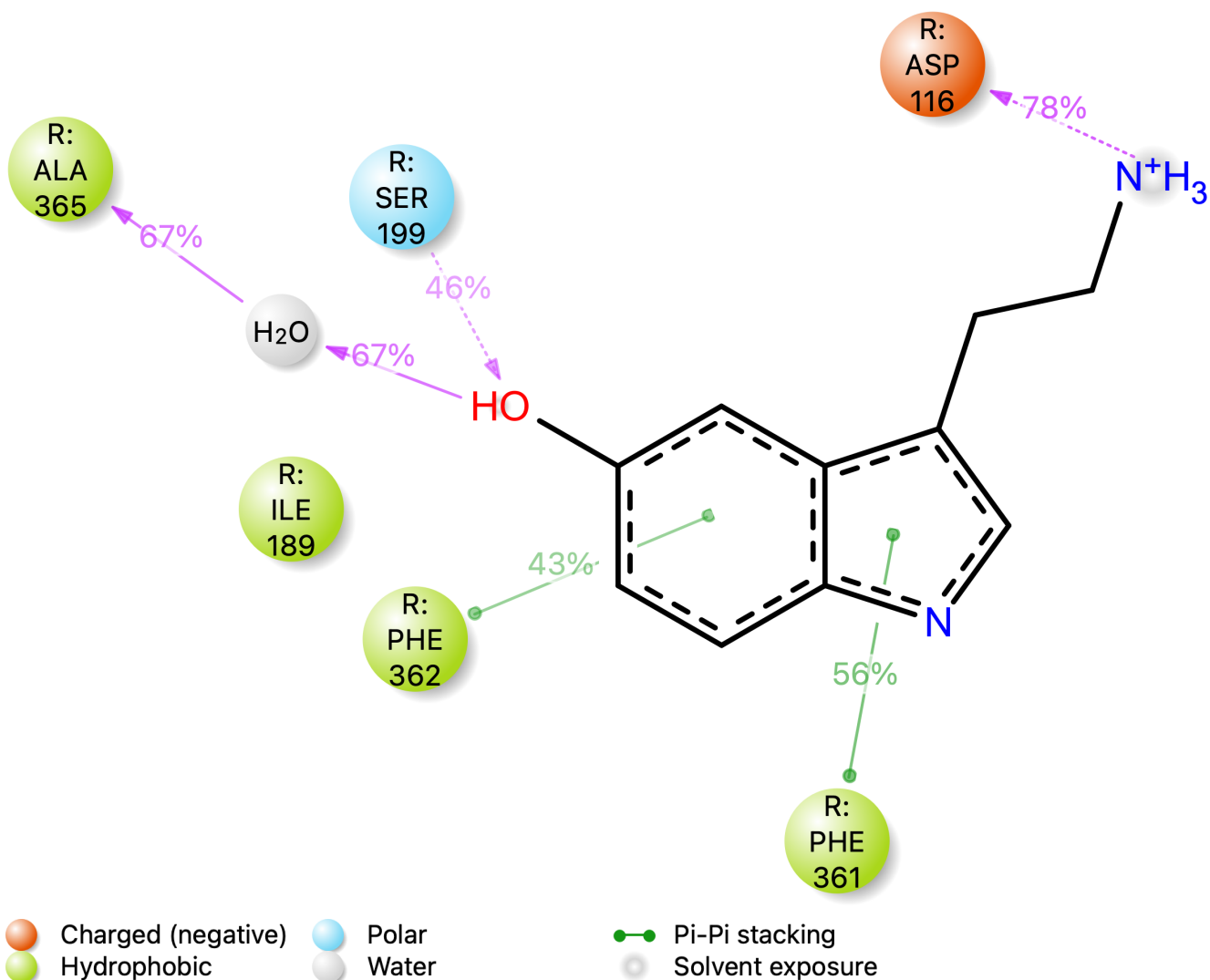
The current geometric criteria for a protein-water or water-ligand H-bond are: a distance of 2.8 Å between the donor and acceptor atoms (D—H···A); a donor angle of  $\geq 110^\circ$  between the donor-hydrogen-acceptor atoms (D—H···A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H···A—X).

## Protein-Ligand Contacts (cont.)



A timeline representation of the interactions and contacts (**H-bonds, Hydrophobic, Ionic, Water bridges**) summarized in the previous page. The top panel shows the total number of specific contacts the protein makes with the ligand over the course of the trajectory. The bottom panel shows which residues interact with the ligand in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot.

## Ligand-Protein Contacts

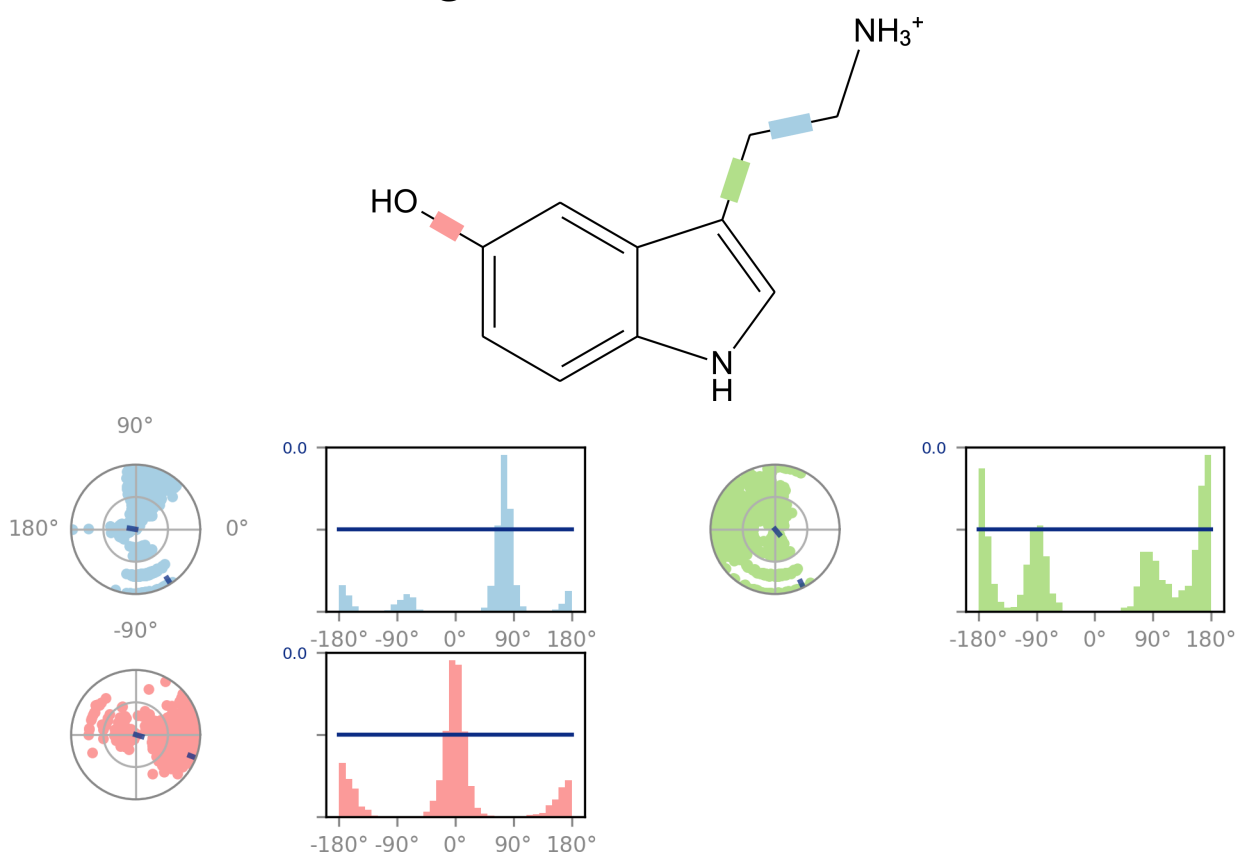


A schematic of detailed ligand atom interactions with the protein residues. Interactions that occur more than **30.0%** of the simulation time in the selected trajectory ( 0.00 through 200.30 nsec), are shown.

Note: it is possible to have interactions with >100% as some residues may have multiple interactions of a single type with the same ligand atom. For example, the ARG side chain has four H-bond donors that can all hydrogen-bond to a single H-bond acceptor.



## Ligand Torsion Profile

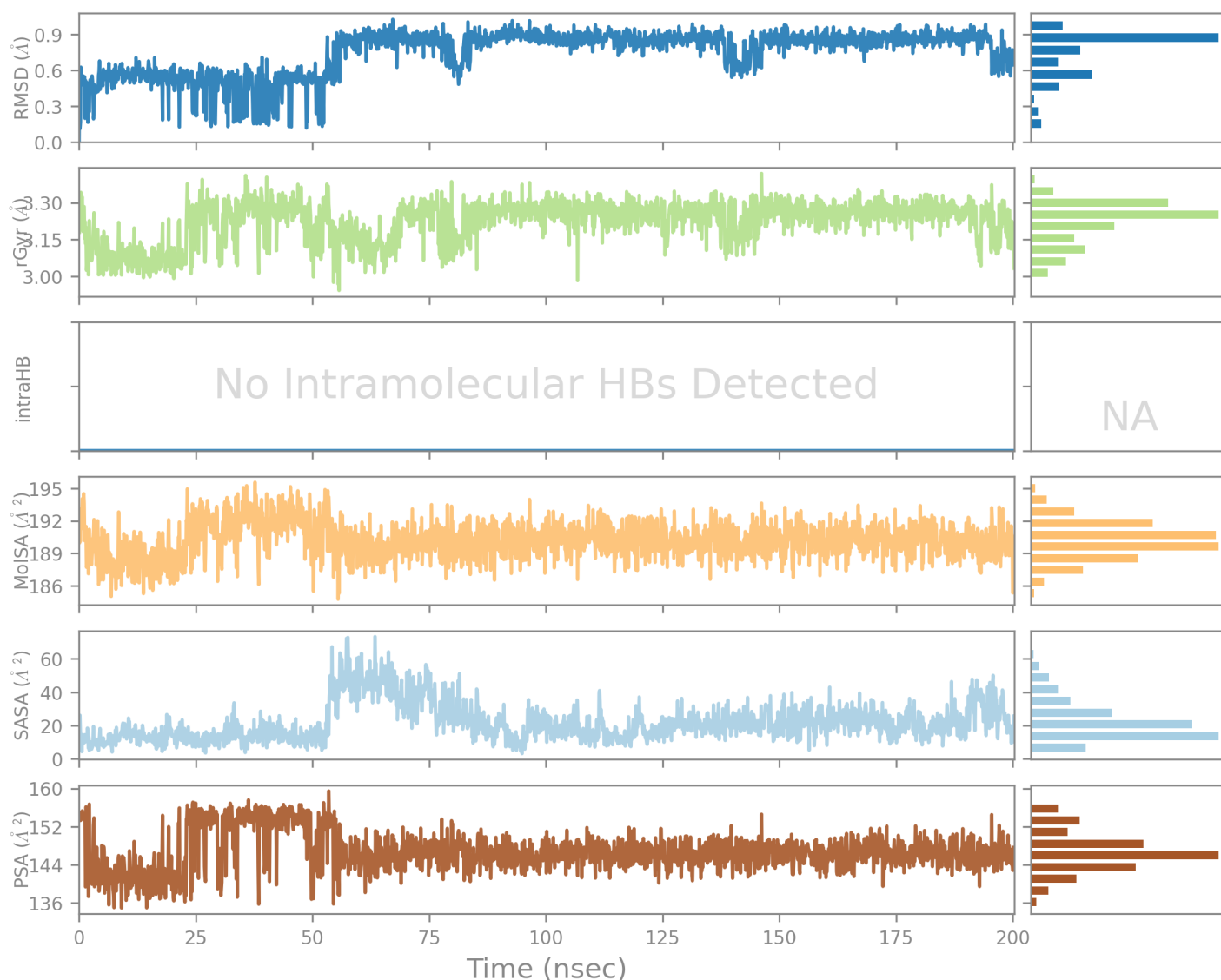


The ligand torsions plot summarizes the conformational evolution of every rotatable bond (RB) in the ligand throughout the simulation trajectory ( 0.00 through 200.30 nsec). The top panel shows the 2d schematic of a ligand with color-coded rotatable bonds. Each rotatable bond torsion is accompanied by a dial plot and bar plots of the same color.

Dial (or radial) plots describe the conformation of the torsion throughout the course of the simulation. The beginning of the simulation is in the center of the radial plot and the time evolution is plotted radially outwards.

The bar plots summarize the data on the dial plots, by showing the probability density of the torsion. If torsional potential information is available, the plot also shows the potential of the rotatable bond (by summing the potential of the related torsions). The values of the potential are on the left Y-axis of the chart, and are expressed in *kcal/mol*. Looking at the histogram and torsion potential relationships may give insights into the conformational strain the ligand undergoes to maintain a protein-bound conformation.

## Ligand Properties



**Ligand RMSD:** Root mean square deviation of a ligand with respect to the reference conformation (typically the first frame is used as the reference and it is regarded as time  $t=0$ ).

**Radius of Gyration (rGyr):** Measures the 'extendedness' of a ligand, and is equivalent to its principal moment of inertia.

**Intramolecular Hydrogen Bonds (intraHB):** Number of internal hydrogen bonds (HB) within a ligand molecule.

**Molecular Surface Area (MolISA):** Molecular surface calculation with 1.4 Å probe radius. This value is equivalent to a van der Waals surface area.

**Solvent Accessible Surface Area (SASA):** Surface area of a molecule accessible by a water molecule.

**Polar Surface Area (PSA):** Solvent accessible surface area in a molecule contributed only by oxygen and nitrogen atoms.

# Simulation Interactions Diagram Report

## Simulation Details

Jobname: merged  
Entry title: MMMP

| CPU #    | Job Type | Ensemble | Temp. [K] | Sim. Time [ns] | # Atoms | # Waters | Charge |
|----------|----------|----------|-----------|----------------|---------|----------|--------|
| Unknown* | Unknown* | Unknown* | 300.0     | 200.404        | 129225  | 38357    | 0      |

\* The configuration file (-out.cfg) was not found. Keep it in same directory as .aef file.

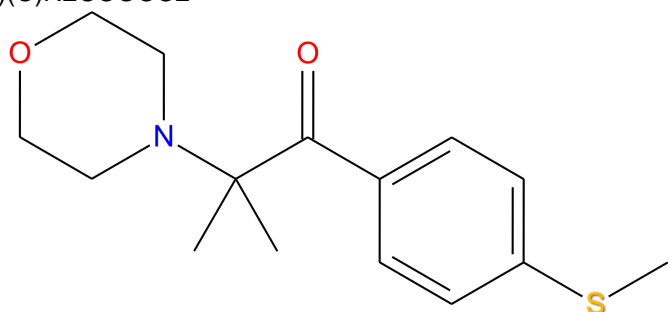
## Protein Information

| Tot. Residues | Prot. Chain(s)     | Res. in Chain(s)             | # Atoms | # Heavy Atoms | Charge |
|---------------|--------------------|------------------------------|---------|---------------|--------|
| 884           | 'A', 'B', 'G', 'R' | Residues (225, 328, 45, 286) | 3897    | 6941          | +3     |



## Ligand Information

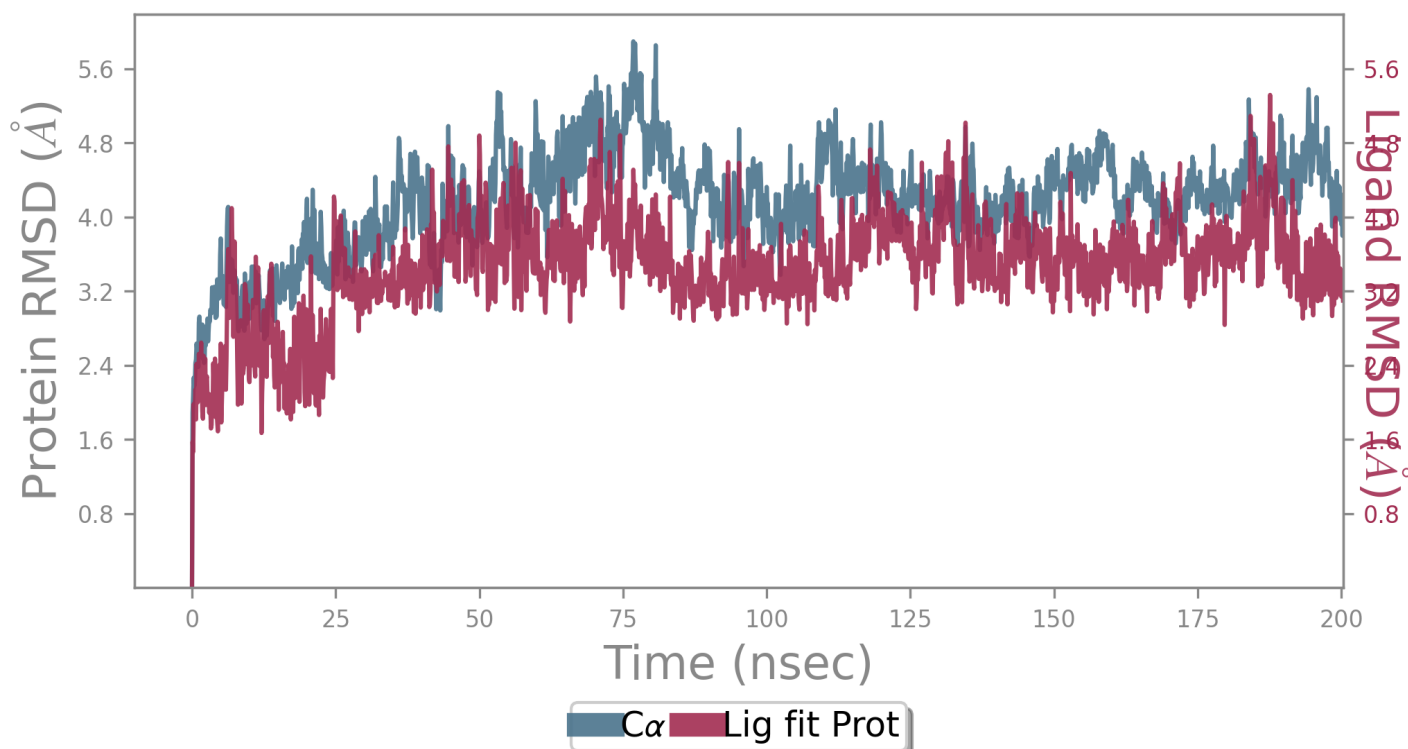
|                    |                                 |
|--------------------|---------------------------------|
| SMILES             | CSc(cc1ccc1C(=O)C(C)(C)N2CCOCC2 |
| PDB Name           | 'UNK'                           |
| Num. of Atoms      | 40 (total) 19 (heavy)           |
| Atomic Mass        | 279.404 au                      |
| Charge             | 0                               |
| Mol. Formula       | C15H21NO2S                      |
| Num. of Fragments  | 2                               |
| Num. of Rot. Bonds | 4                               |



## Counter Ion/Salt Information

| Type | Num. | Concentration [mM] | Total Charge |
|------|------|--------------------|--------------|
| Cl   | 110  | 52.142             | -110         |
| Na   | 107  | 50.720             | +107         |

## Protein-Ligand RMSD



The Root Mean Square Deviation (RMSD) is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The RMSD for frame  $x$  is:

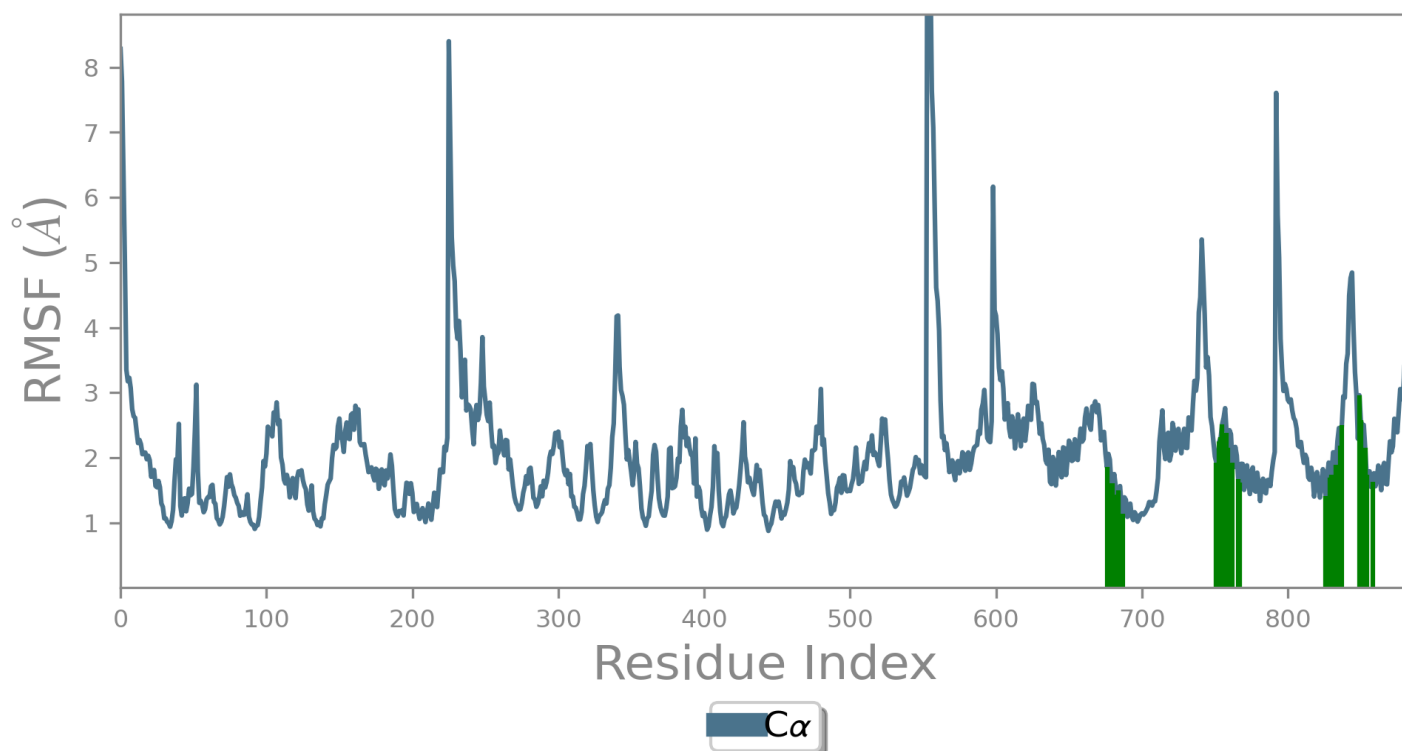
$$RMSD_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'_i(t_x) - r_i(t_{ref}))^2}$$

where  $N$  is the number of atoms in the atom selection;  $t_{ref}$  is the reference time, (typically the first frame is used as the reference and it is regarded as time  $t=0$ ); and  $r'_i$  is the position of the selected atoms in frame  $x$  after superimposing on the reference frame, where frame  $x$  is recorded at time  $t_x$ . The procedure is repeated for every frame in the simulation trajectory.

**Protein RMSD:** The above plot shows the RMSD evolution of a protein (left Y-axis). All protein frames are first aligned on the reference frame backbone, and then the RMSD is calculated based on the atom selection. Monitoring the RMSD of the protein can give insights into its structural conformation throughout the simulation. RMSD analysis can indicate if the simulation has equilibrated — its fluctuations towards the end of the simulation are around some thermal average structure. Changes of the order of 1-3 Å are perfectly acceptable for small, globular proteins. Changes much larger than that, however, indicate that the protein is undergoing a large conformational change during the simulation. It is also important that your simulation converges — the RMSD values stabilize around a fixed value. If the RMSD of the protein is still increasing or decreasing on average at the end of the simulation, then your system has not equilibrated, and your simulation may not be long enough for rigorous analysis.

**Ligand RMSD:** Ligand RMSD (right Y-axis) indicates how stable the ligand is with respect to the protein and its binding pocket. In the above plot, 'Lig fit Prot' shows the RMSD of a ligand when the protein-ligand complex is first aligned on the protein backbone of the reference and then the RMSD of the ligand heavy atoms is measured. If the values observed are significantly larger than the RMSD of the protein, then it is likely that the ligand has diffused away from its initial binding site.

## Protein RMSF



The Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain. The RMSF for residue  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \langle (r'_i(t) - r_i(t_{ref}))^2 \rangle}$$

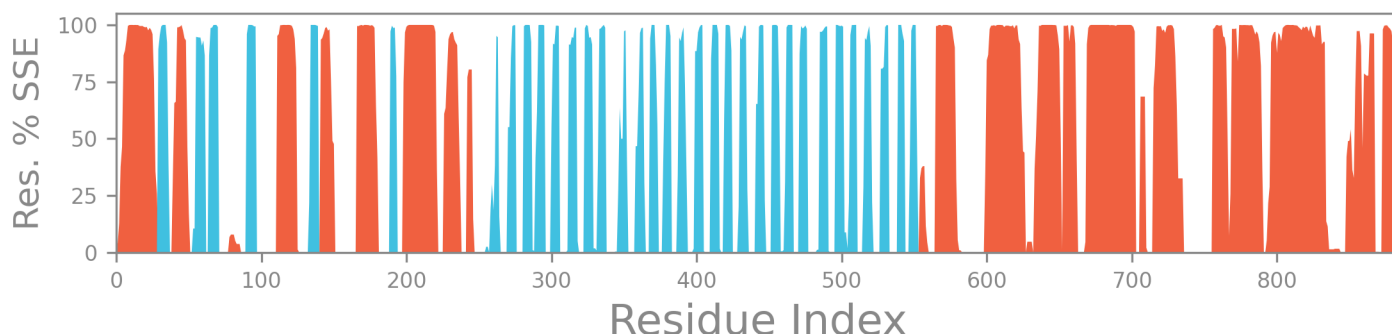
where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time,  $r_i$  is the position of residue  $i$ ;  $r'$  is the position of atoms in residue  $i$  after superposition on  $t_{ref}$  the reference, and the  $i$  angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

On this plot, peaks indicate areas of the protein that fluctuate the most during the simulation. Typically you will observe that the tails ( $N$ - and  $C$ -terminal) fluctuate more than any other part of the protein. Secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein, and thus fluctuate less than the loop regions.

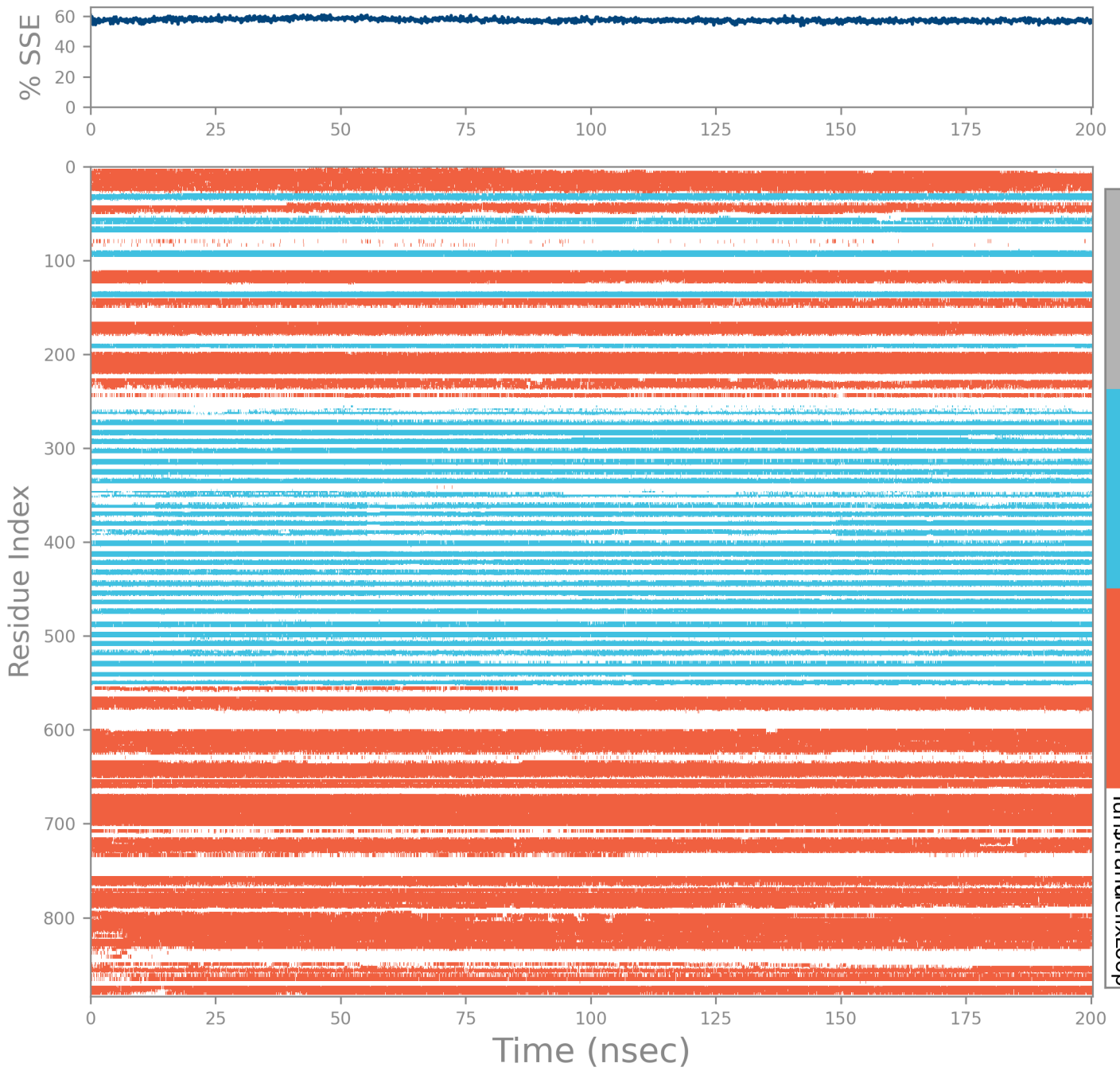
Ligand Contacts: Protein residues that interact with the ligand are marked with green-colored vertical bars.

# Protein Secondary Structure

|         |          |             |
|---------|----------|-------------|
| % Helix | % Strand | % Total SSE |
| 36.13   | 21.28    | 57.41       |

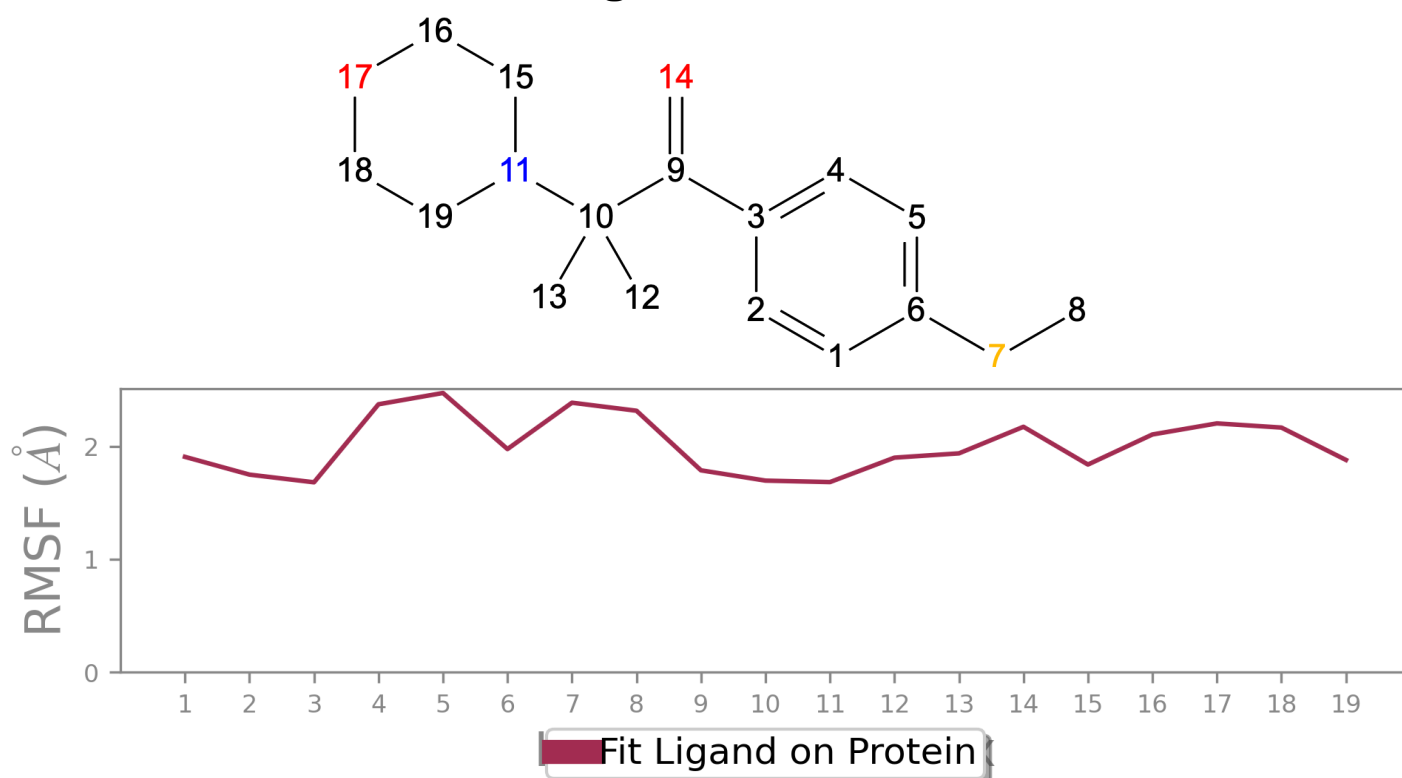


Protein secondary structure elements (SSE) like **alpha-helices** and **beta-strands** are monitored throughout the simulation. The plot above reports SSE distribution by residue index throughout the protein structure. The plot below summarizes the SSE composition for each trajectory frame over the course of the simulation, and the plot at the bottom monitors each residue and its SSE assignment over time.





## Ligand RMSF



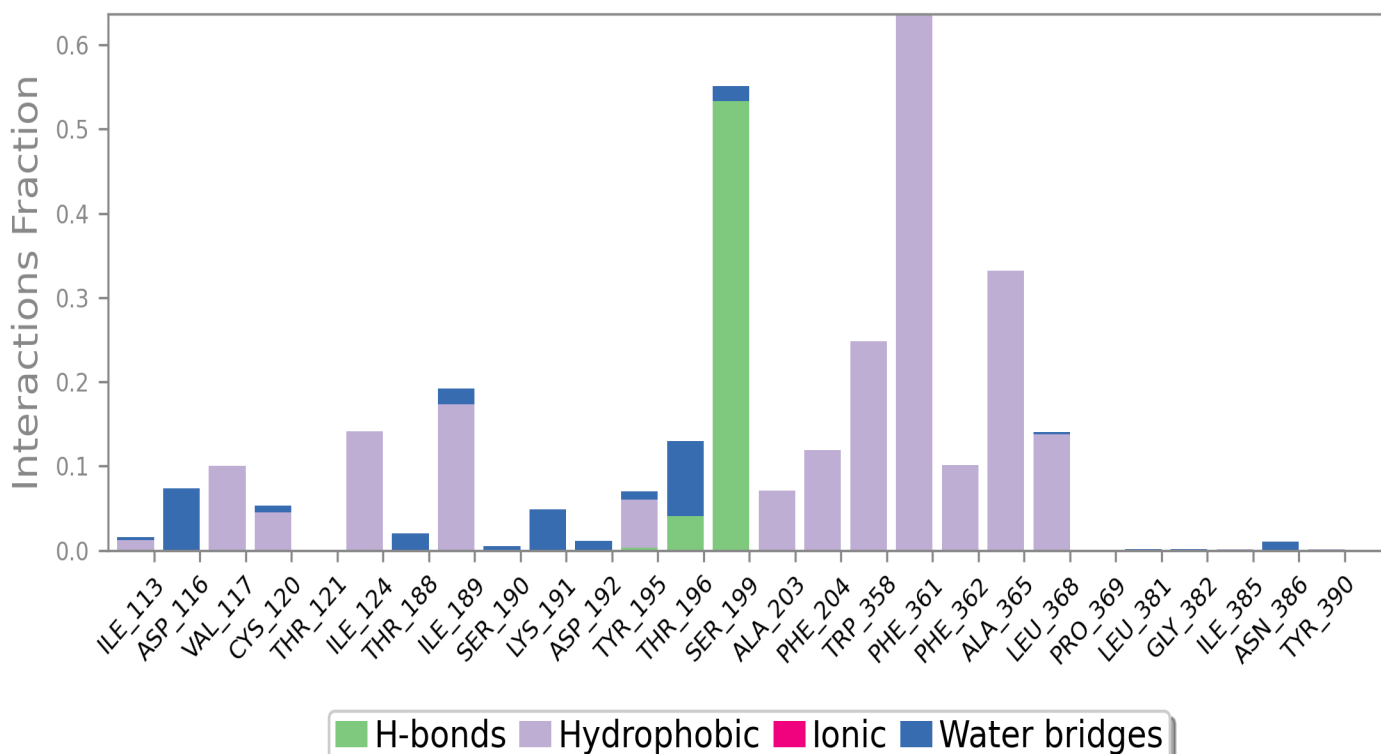
The Ligand Root Mean Square Fluctuation (L-RMSF) is useful for characterizing changes in the ligand atom positions. The RMSF for atom  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (r'_i(t) - r_i(t_{ref}))^2}$$

where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time (usually for the first frame, and is regarded as the zero of time);  $r$  is the position of atom  $i$  in the reference at time  $t_{ref}$  and  $r'$  is the position of atom  $i$  at time  $t$  after superposition on the reference frame.

Ligand RMSF shows the ligand's fluctuations broken down by atom, corresponding to the 2D structure in the top panel. The ligand RMSF may give you insights on how ligand fragments interact with the protein and their entropic role in the binding event. In the bottom panel, the 'Fit Ligand on Protein' line shows the ligand fluctuations, with respect to the protein. The protein-ligand complex is first aligned on the protein backbone and then the ligand RMSF is measured on the ligand heavy atoms.

## Protein-Ligand Contacts



Protein interactions with the ligand can be monitored throughout the simulation. These interactions can be categorized by type and summarized, as shown in the plot above. Protein-ligand interactions (or 'contacts') are categorized into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges. Each interaction type contains more specific subtypes, which can be explored through the 'Simulation Interactions Diagram' panel. The stacked bar charts are normalized over the course of the trajectory: for example, a value of 0.7 suggests that 70% of the simulation time the specific interaction is maintained. Values over 1.0 are possible as some protein residue may make multiple contacts of same subtype with the ligand.

**Hydrogen Bonds:** (H-bonds) play a significant role in ligand binding. Consideration of hydrogen-bonding properties in drug design is important because of their strong influence on drug specificity, metabolism and adsorption. Hydrogen bonds between a protein and a ligand can be further broken down into four subtypes: backbone acceptor; backbone donor; side-chain acceptor; side-chain donor.

The current geometric criteria for protein-ligand H-bond is: distance of 2.5 Å between the donor and acceptor atoms (D—H...A); a donor angle of  $\geq 120^\circ$  between the donor-hydrogen-acceptor atoms (D—H...A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H...A—X).

**Hydrophobic contacts:** fall into three subtypes:  $\pi$ -Cation;  $\pi$ - $\pi$ ; and Other, non-specific interactions. Generally these type of interactions involve a hydrophobic amino acid and an aromatic or aliphatic group on the ligand, but we have extended this category to also include  $\pi$ -Cation interactions.

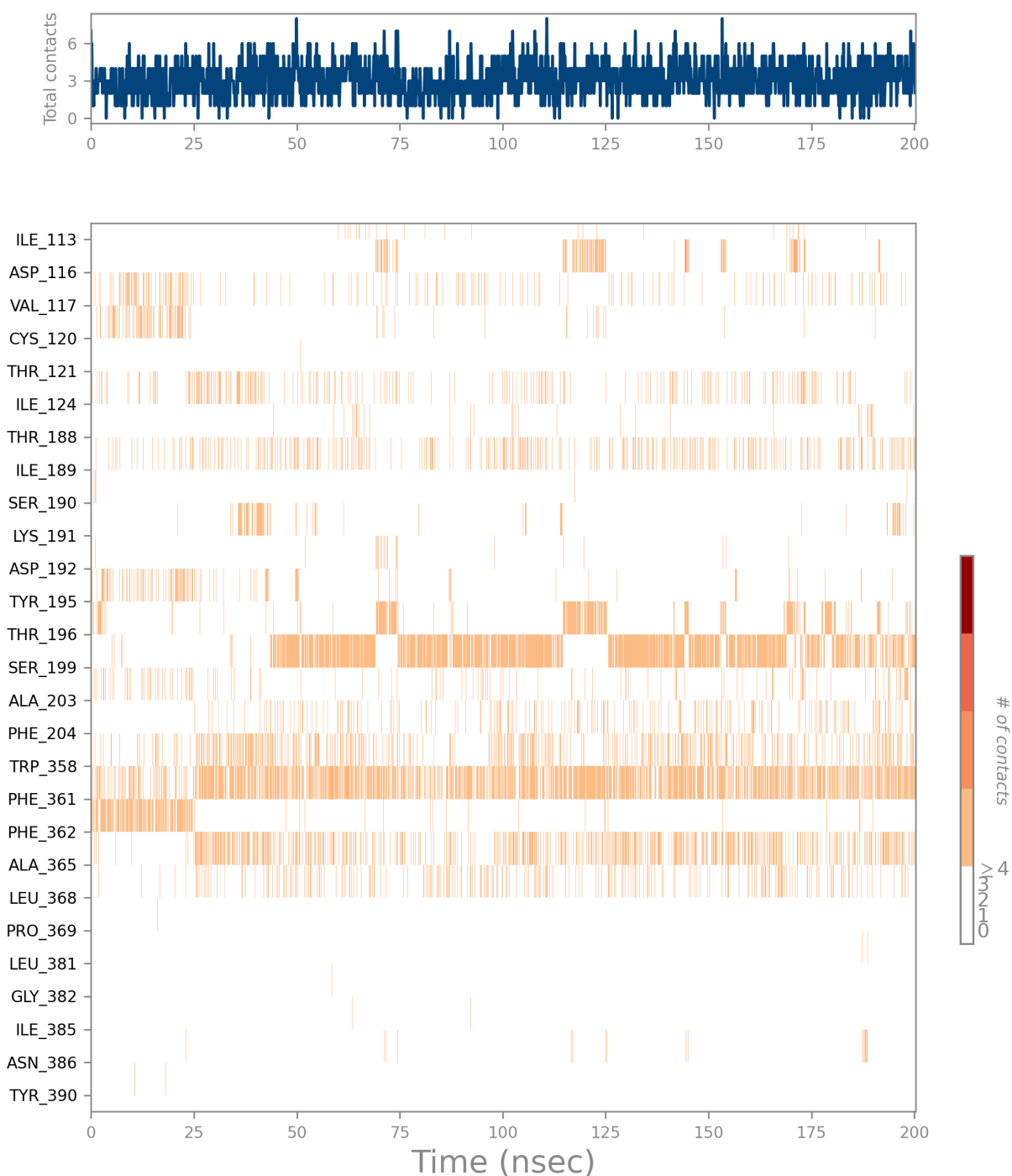
The current geometric criteria for hydrophobic interactions is as follows:  $\pi$ -Cation — Aromatic and charged groups within 4.5 Å;  $\pi$ - $\pi$  — Two aromatic groups stacked face-to-face or face-to-edge; Other — A non-specific hydrophobic sidechain within 3.6 Å of a ligand's aromatic or aliphatic carbons.

**Ionic interactions:** or polar interactions, are between two oppositely charged atoms that are within 3.7 Å of each other and do not involve a hydrogen bond. We also monitor Protein-Metal-Ligand interactions, which are defined by a metal ion coordinated within 3.4 Å of protein's and ligand's heavy atoms (except carbon). All ionic interactions are broken down into two subtypes: those mediated by a protein backbone or side chains.

**Water Bridges:** are hydrogen-bonded protein-ligand interactions mediated by a water molecule. The hydrogen-bond geometry is slightly relaxed from the standard H-bond definition.

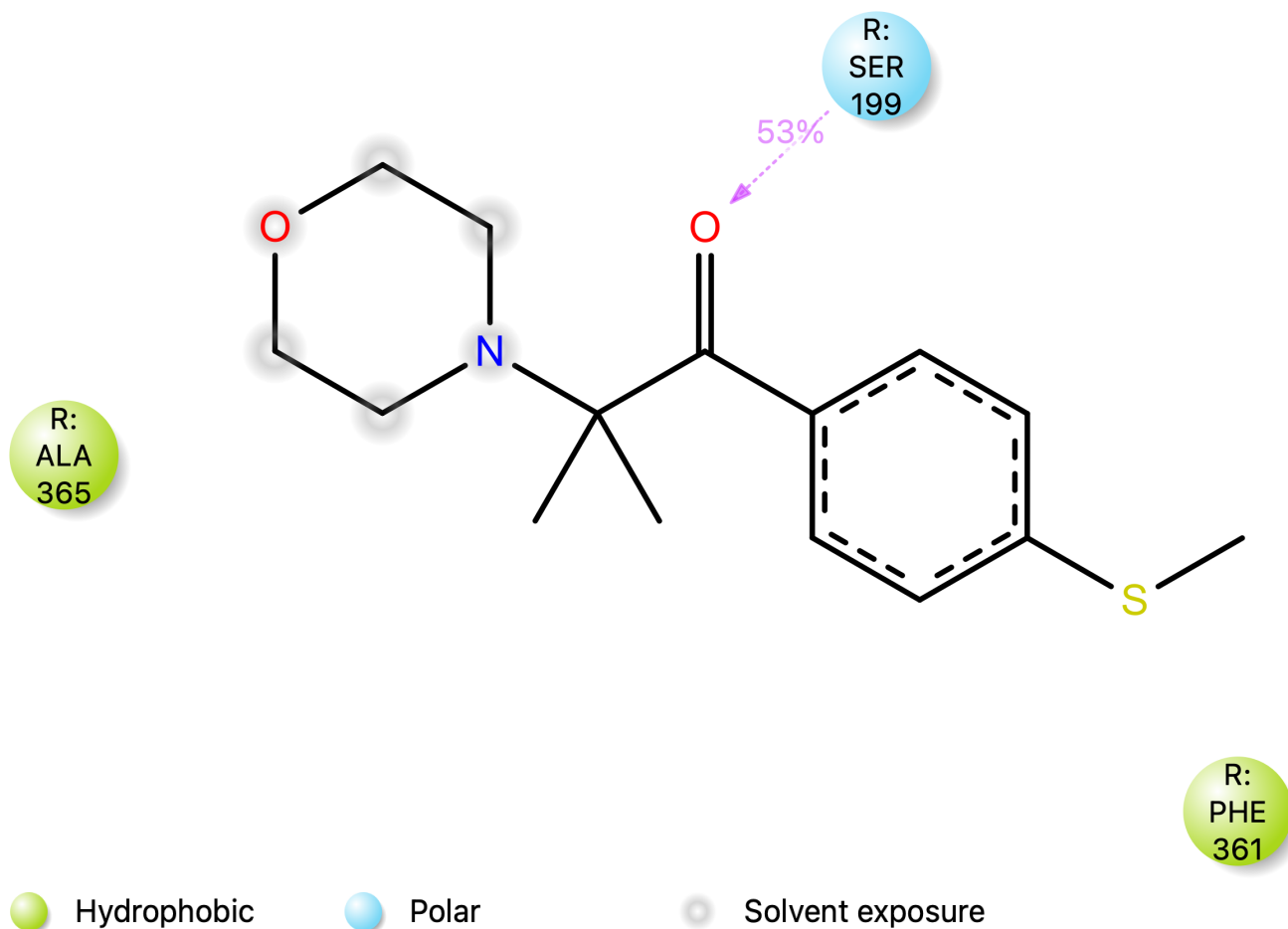
The current geometric criteria for a protein-water or water-ligand H-bond are: a distance of 2.8 Å between the donor and acceptor atoms (D—H...A); a donor angle of  $\geq 110^\circ$  between the donor-hydrogen-acceptor atoms (D—H...A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H...A—X).

## Protein-Ligand Contacts (cont.)



A timeline representation of the interactions and contacts (**H-bonds, Hydrophobic, Ionic, Water bridges**) summarized in the previous page. The top panel shows the total number of specific contacts the protein makes with the ligand over the course of the trajectory. The bottom panel shows which residues interact with the ligand in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot.

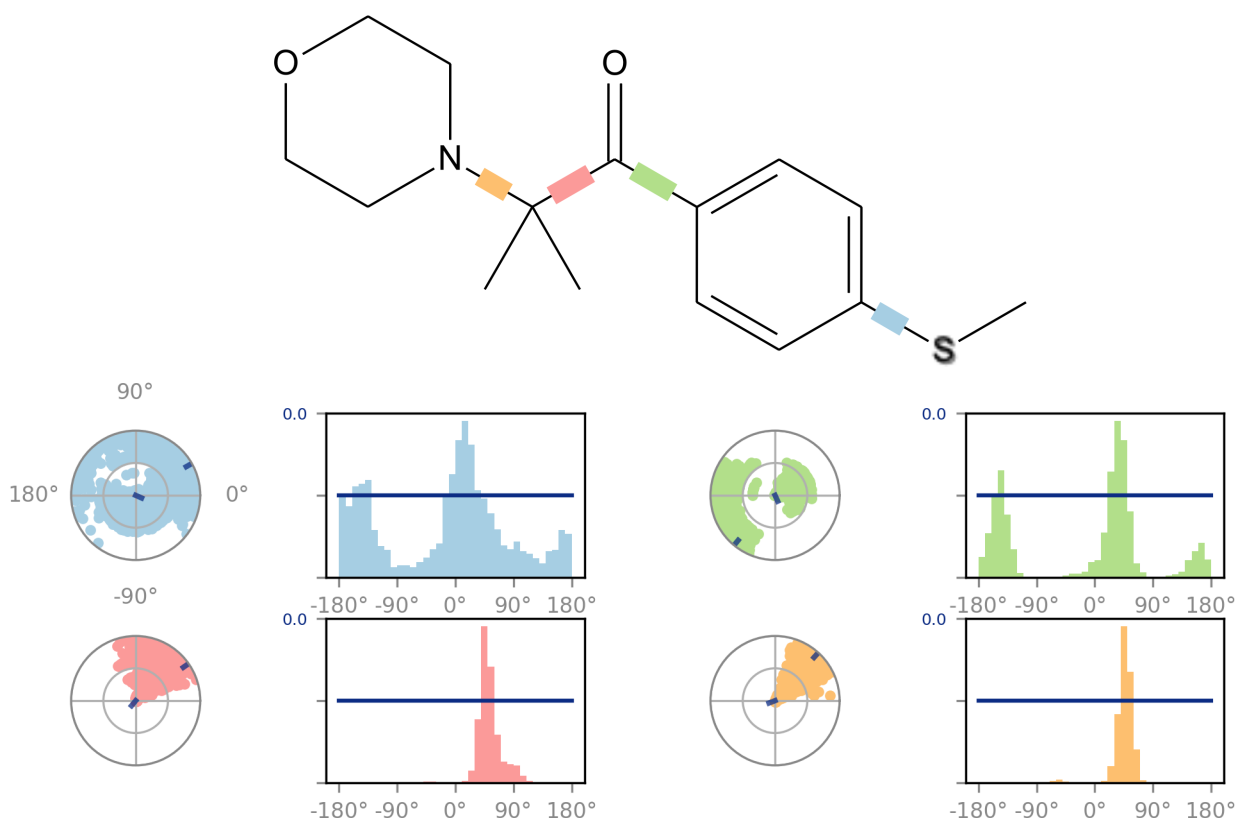
## Ligand-Protein Contacts



A schematic of detailed ligand atom interactions with the protein residues. Interactions that occur more than **30.0%** of the simulation time in the selected trajectory ( 0.00 through 200.30 nsec), are shown.

Note: it is possible to have interactions with >100% as some residues may have multiple interactions of a single type with the same ligand atom. For example, the ARG side chain has four H-bond donors that can all hydrogen-bond to a single H-bond acceptor.

## Ligand Torsion Profile

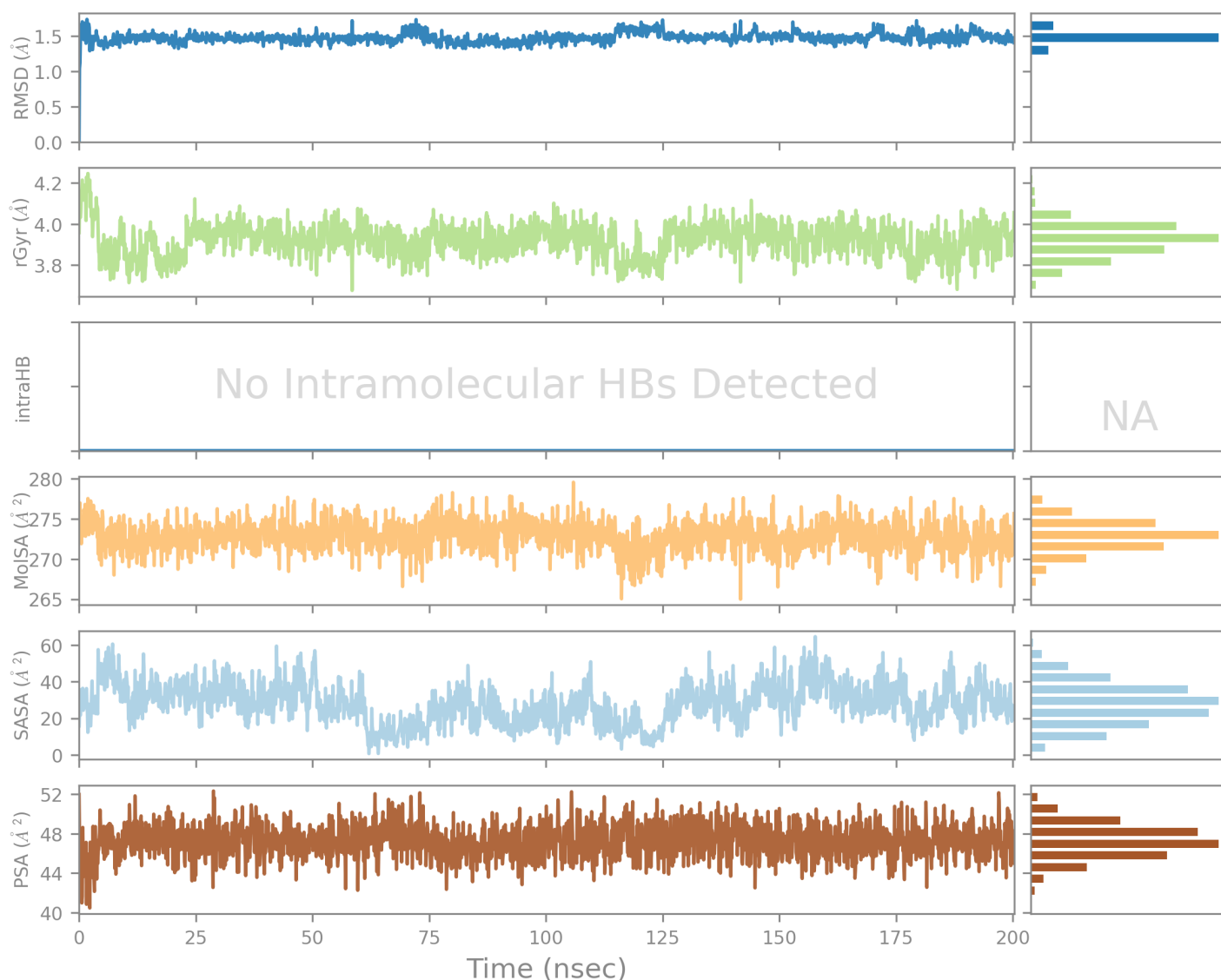


The ligand torsions plot summarizes the conformational evolution of every rotatable bond (RB) in the ligand throughout the simulation trajectory ( 0.00 through 200.30 nsec). The top panel shows the 2d schematic of a ligand with color-coded rotatable bonds. Each rotatable bond torsion is accompanied by a dial plot and bar plots of the same color.

Dial (or radial) plots describe the conformation of the torsion throughout the course of the simulation. The beginning of the simulation is in the center of the radial plot and the time evolution is plotted radially outwards.

The bar plots summarize the data on the dial plots, by showing the probability density of the torsion. If torsional potential information is available, the plot also shows the potential of the rotatable bond (by summing the potential of the related torsions). The values of the potential are on the left Y-axis of the chart, and are expressed in *kcal/mol*. Looking at the histogram and torsion potential relationships may give insights into the conformational strain the ligand undergoes to maintain a protein-bound conformation.

## Ligand Properties



**Ligand RMSD:** Root mean square deviation of a ligand with respect to the reference conformation (typically the first frame is used as the reference and it is regarded as time  $t=0$ ).

**Radius of Gyration (rGyr):** Measures the 'extendedness' of a ligand, and is equivalent to its principal moment of inertia.

**Intramolecular Hydrogen Bonds (intraHB):** Number of internal hydrogen bonds (HB) within a ligand molecule.

**Molecular Surface Area (MolISA):** Molecular surface calculation with 1.4 Å probe radius. This value is equivalent to a van der Waals surface area.

**Solvent Accessible Surface Area (SASA):** Surface area of a molecule accessible by a water molecule.

**Polar Surface Area (PSA):** Solvent accessible surface area in a molecule contributed only by oxygen and nitrogen atoms.

# Simulation Interactions Diagram Report

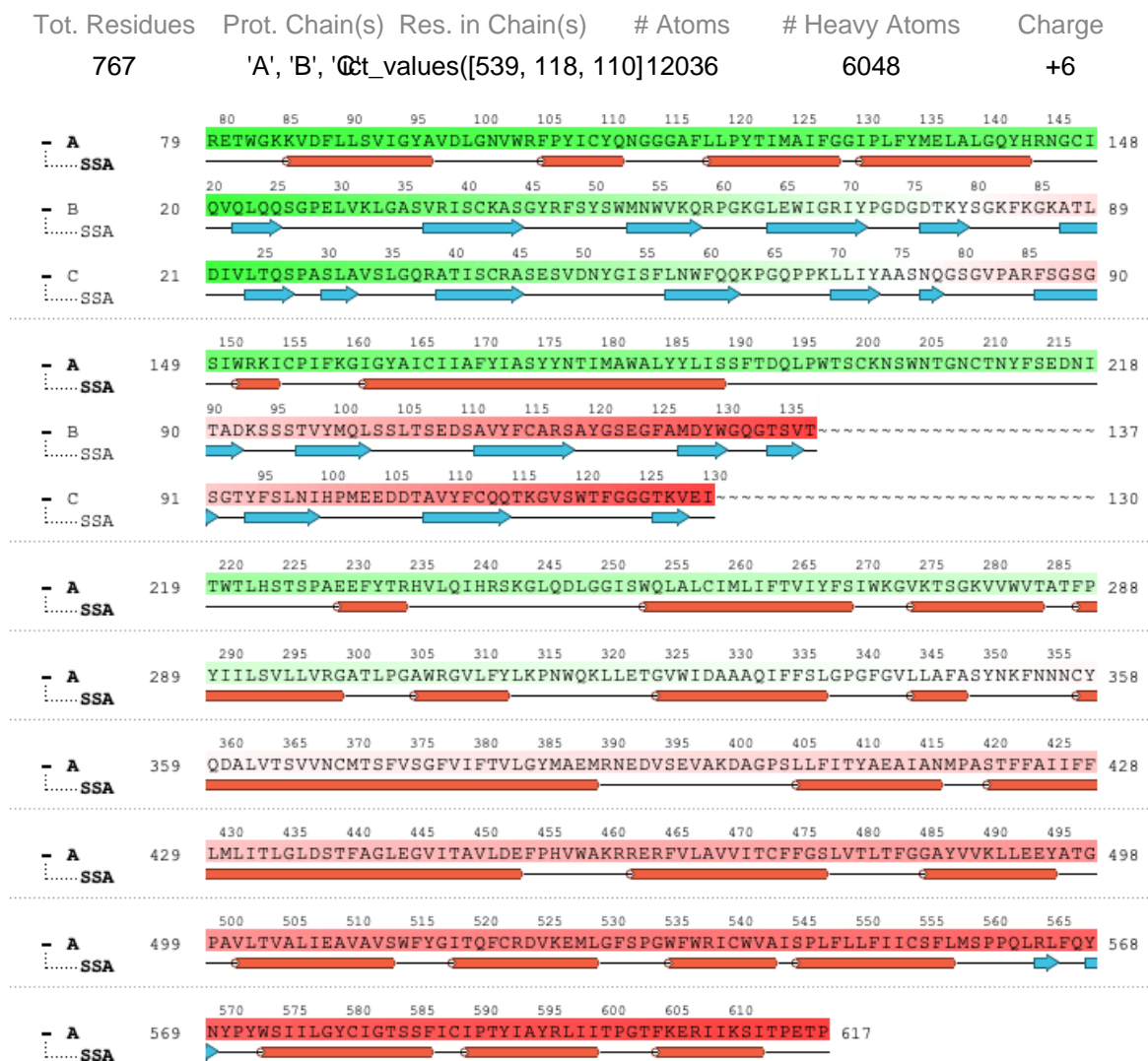
## Simulation Details

Jobname: merged  
Entry title: Ligand 1

| CPU #    | Job Type | Ensemble | Temp. [K] | Sim. Time [ns] | # Atoms | # Waters | Charge |
|----------|----------|----------|-----------|----------------|---------|----------|--------|
| Unknown* | Unknown* | Unknown* | 300.0     | 200.404        | 100650  | 29471    | 0      |

\* The configuration file (-out.cfg) was not found. Keep it in same directory as .aef file.

## Protein Information

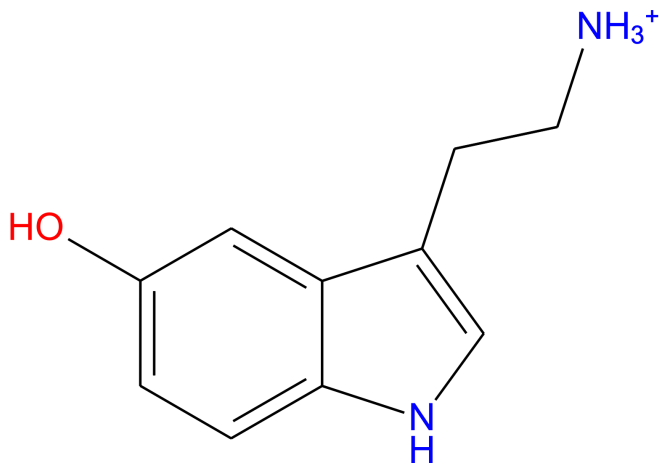


## Ligand Information

SMILES

[NH3+]CCc1c[nH]c(c12)ccc(c2)O

PDB Name 'SRO', 'UNK'  
Num. of Atoms 26 (total) 13 (heavy)  
Atomic Mass 177.228 au  
Charge +1  
Mol. Formula C<sub>10</sub>H<sub>13</sub>N<sub>2</sub>O  
Num. of Fragments 2  
Num. of Rot. Bonds 3

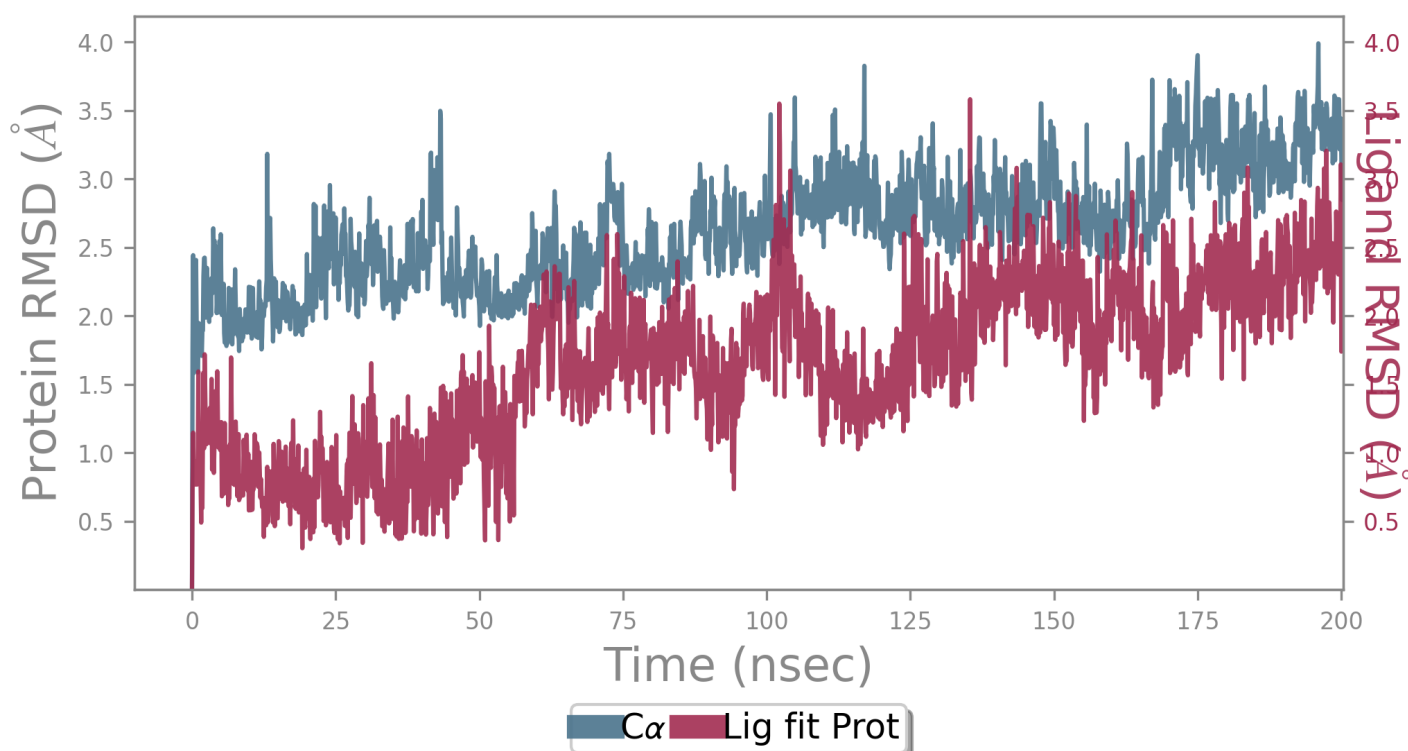


### Counter Ion/Salt Information

| Type | Num. | Concentration [mM] | Total Charge |
|------|------|--------------------|--------------|
| Cl   | 91   | 56.141             | -91          |
| Na   | 84   | 51.823             | +84          |



## Protein-Ligand RMSD



The Root Mean Square Deviation (RMSD) is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The RMSD for frame  $x$  is:

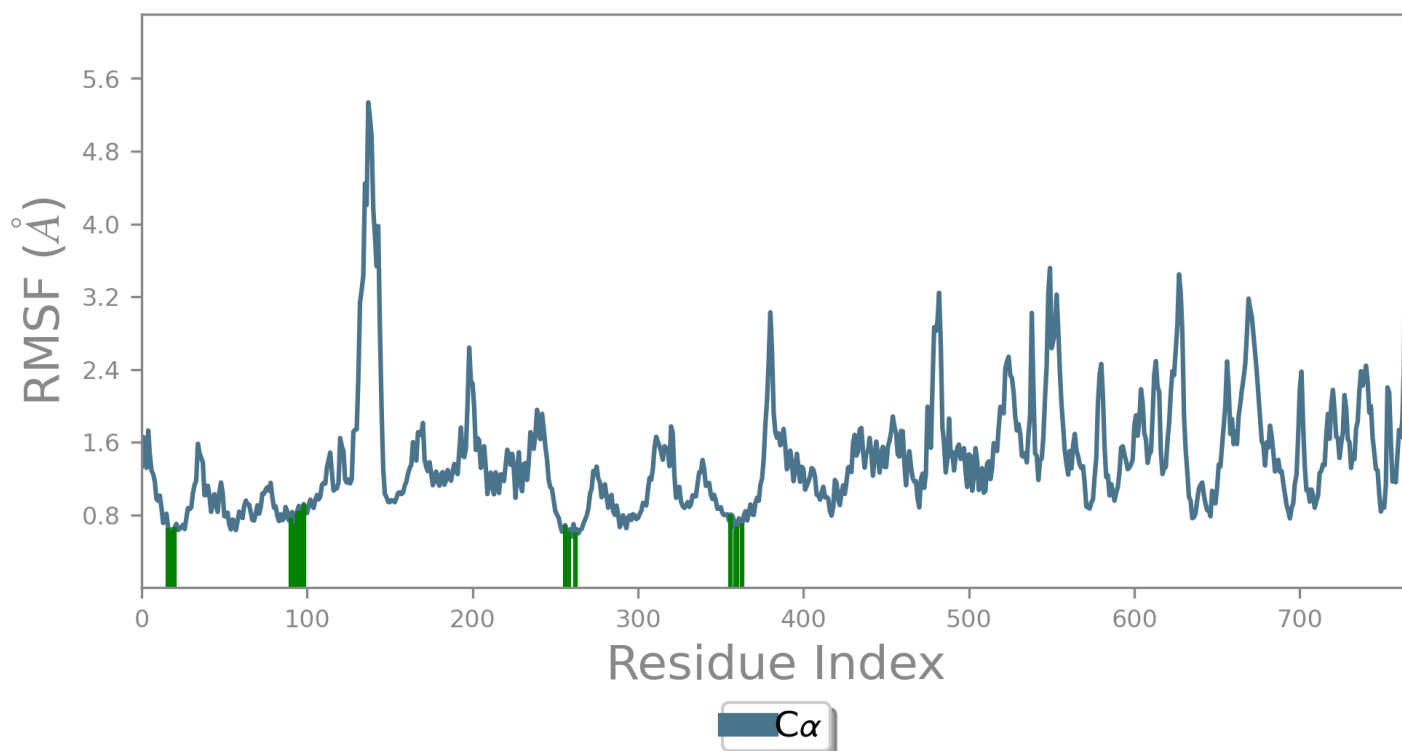
$$RMSD_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'_i(t_x) - r_i(t_{ref}))^2}$$

where  $N$  is the number of atoms in the atom selection;  $t_{ref}$  is the reference time, (typically the first frame is used as the reference and it is regarded as time  $t=0$ ); and  $r'_i$  is the position of the selected atoms in frame  $x$  after superimposing on the reference frame, where frame  $x$  is recorded at time  $t_x$ . The procedure is repeated for every frame in the simulation trajectory.

**Protein RMSD:** The above plot shows the RMSD evolution of a protein (left Y-axis). All protein frames are first aligned on the reference frame backbone, and then the RMSD is calculated based on the atom selection. Monitoring the RMSD of the protein can give insights into its structural conformation throughout the simulation. RMSD analysis can indicate if the simulation has equilibrated — its fluctuations towards the end of the simulation are around some thermal average structure. Changes of the order of 1-3 Å are perfectly acceptable for small, globular proteins. Changes much larger than that, however, indicate that the protein is undergoing a large conformational change during the simulation. It is also important that your simulation converges — the RMSD values stabilize around a fixed value. If the RMSD of the protein is still increasing or decreasing on average at the end of the simulation, then your system has not equilibrated, and your simulation may not be long enough for rigorous analysis.

**Ligand RMSD:** Ligand RMSD (right Y-axis) indicates how stable the ligand is with respect to the protein and its binding pocket. In the above plot, 'Lig fit Prot' shows the RMSD of a ligand when the protein-ligand complex is first aligned on the protein backbone of the reference and then the RMSD of the ligand heavy atoms is measured. If the values observed are significantly larger than the RMSD of the protein, then it is likely that the ligand has diffused away from its initial binding site.

## Protein RMSF



The Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain. The RMSF for residue  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \langle (r'_i(t)) - r_i(t_{ref}) \rangle^2}$$

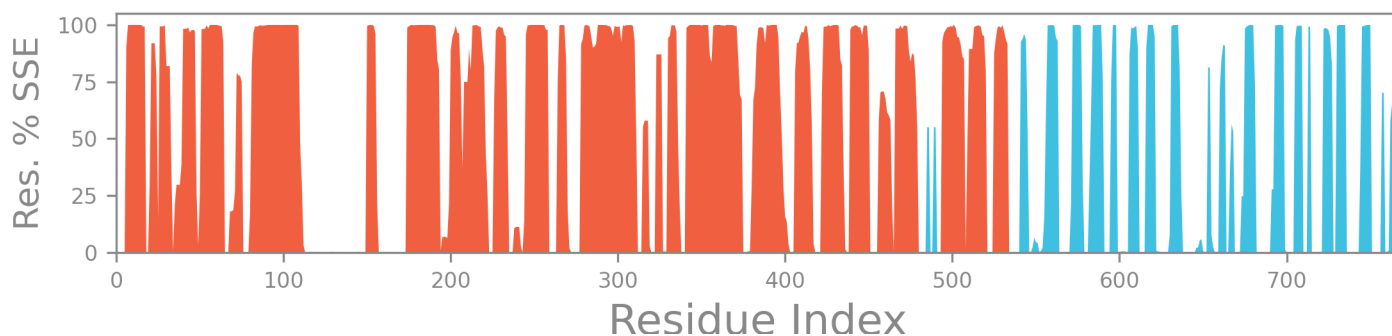
where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time,  $r_i$  is the position of residue  $i$ ;  $r'$  is the position of atoms in residue  $i$  after superposition on  $t_{ref}$ , and the  $\langle \rangle$  angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

On this plot, peaks indicate areas of the protein that fluctuate the most during the simulation. Typically you will observe that the tails ( $N$ - and  $C$ -terminal) fluctuate more than any other part of the protein. Secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein, and thus fluctuate less than the loop regions.

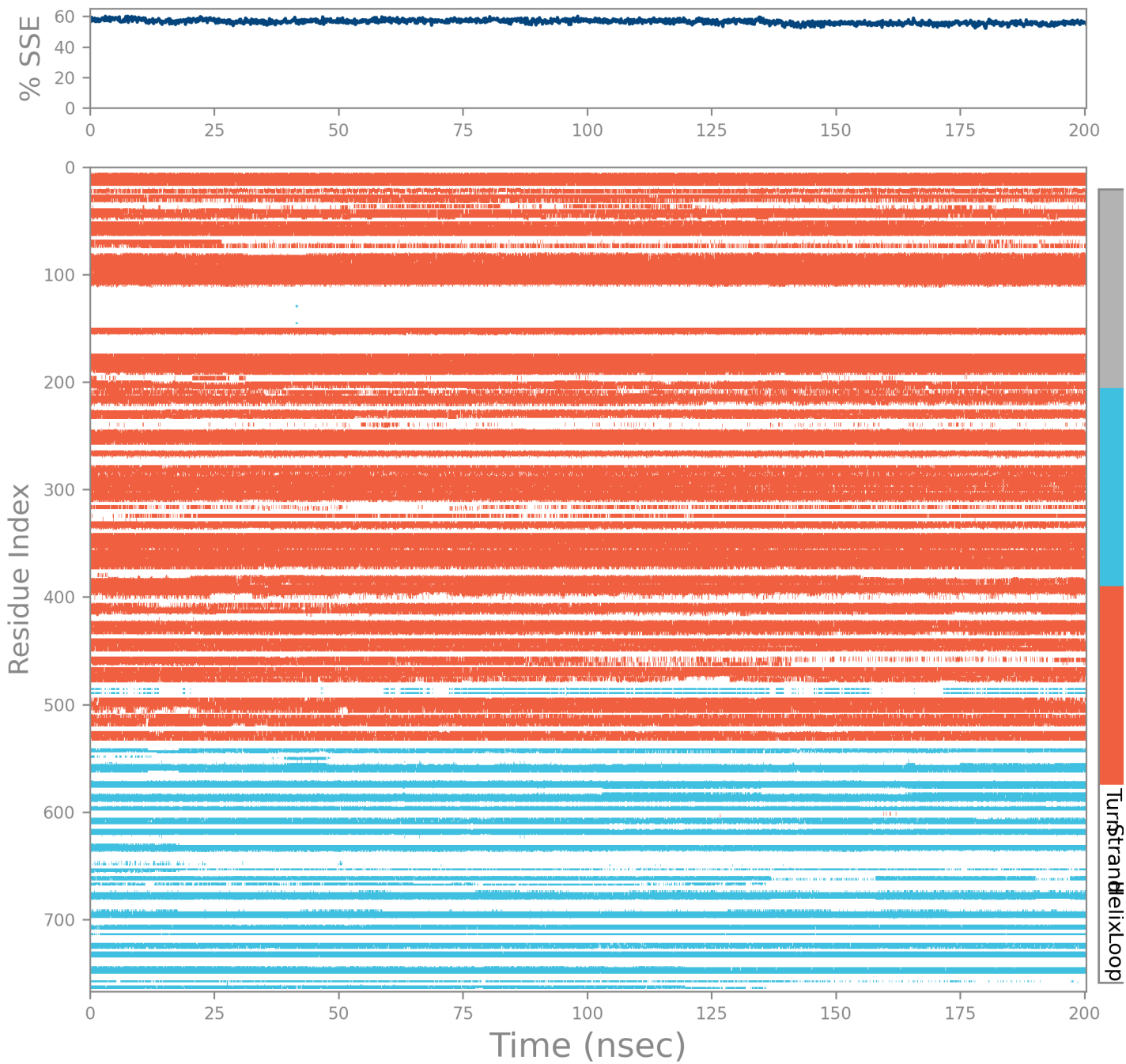
**Ligand Contacts:** Protein residues that interact with the ligand are marked with green-colored vertical bars.

# Protein Secondary Structure

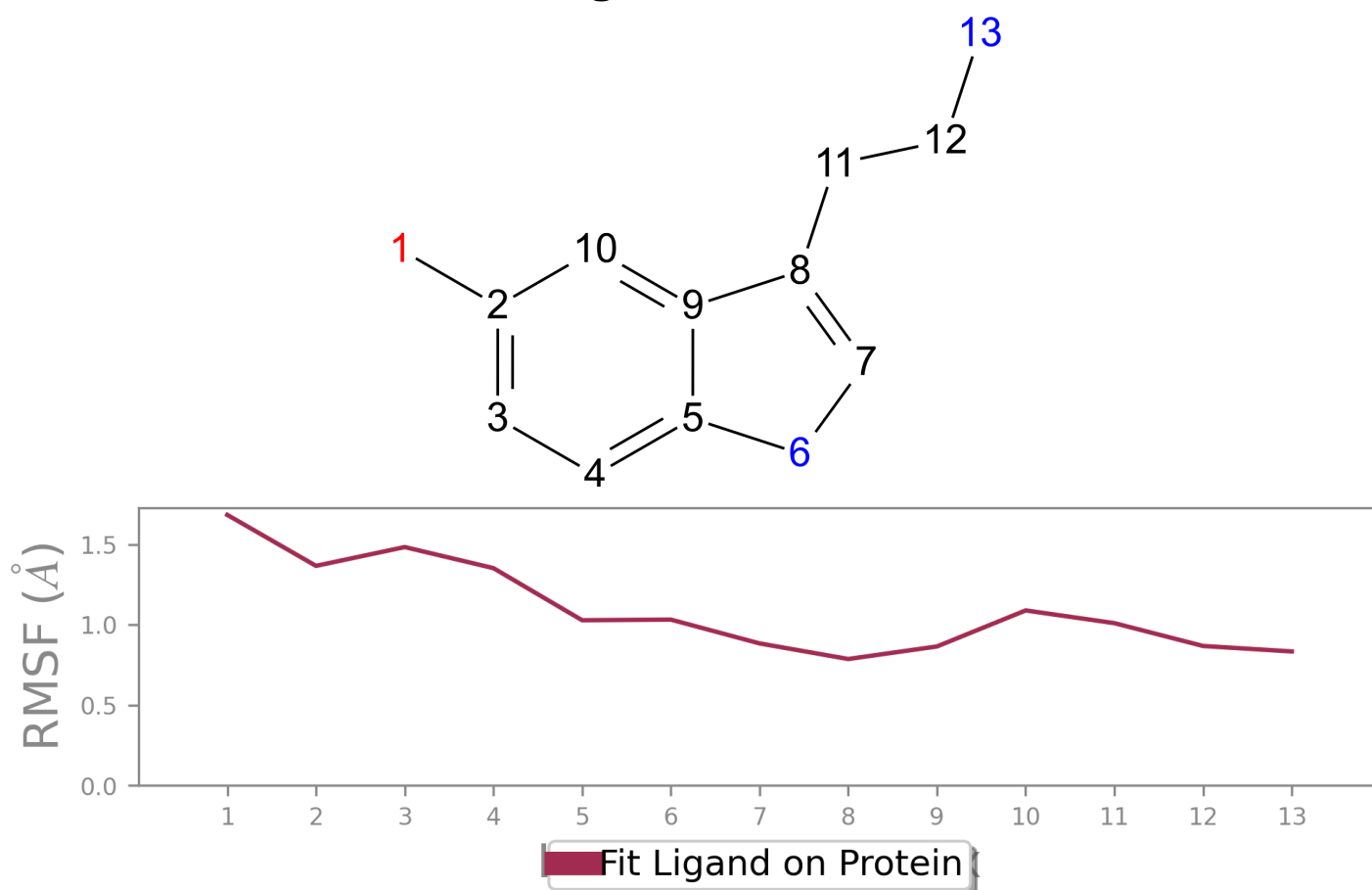
|         |          |             |
|---------|----------|-------------|
| % Helix | % Strand | % Total SSE |
| 43.58   | 12.87    | 56.45       |



Protein secondary structure elements (SSE) like **alpha-helices** and **beta-strands** are monitored throughout the simulation. The plot above reports SSE distribution by residue index throughout the protein structure. The plot below summarizes the SSE composition for each trajectory frame over the course of the simulation, and the plot at the bottom monitors each residue and its SSE assignment over time.



## Ligand RMSF



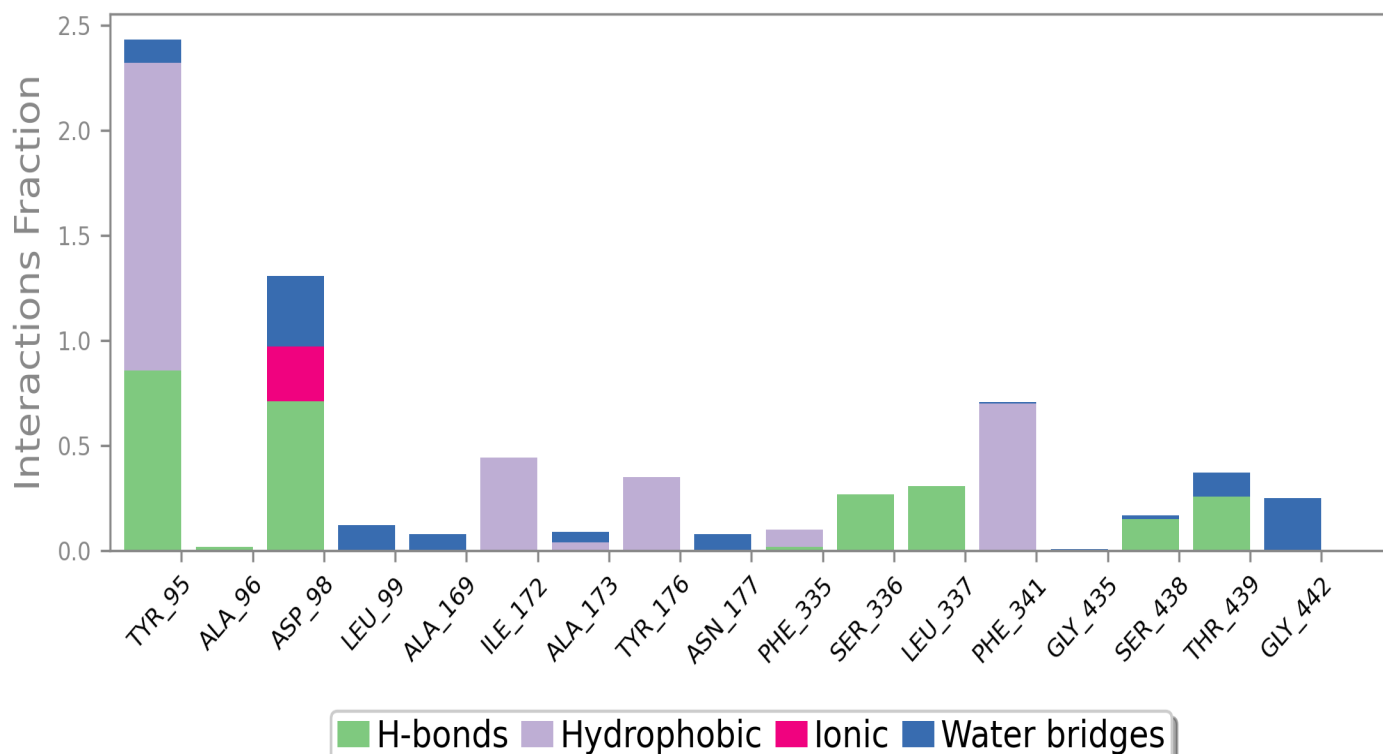
The Ligand Root Mean Square Fluctuation (L-RMSF) is useful for characterizing changes in the ligand atom positions. The RMSF for atom  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (r'_i(t) - r_i(t_{ref}))^2}$$

where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time (usually for the first frame, and is regarded as the zero of time);  $r$  is the position of atom  $i$  in the reference at time  $t_{ref}$  and  $r'$  is the position of atom  $i$  at time  $t$  after superposition on the reference frame.

Ligand RMSF shows the ligand's fluctuations broken down by atom, corresponding to the 2D structure in the top panel. The ligand RMSF may give you insights on how ligand fragments interact with the protein and their entropic role in the binding event. In the bottom panel, the 'Fit Ligand on Protein' line shows the ligand fluctuations, with respect to the protein. The protein-ligand complex is first aligned on the protein backbone and then the ligand RMSF is measured on the ligand heavy atoms.

## Protein-Ligand Contacts



Protein interactions with the ligand can be monitored throughout the simulation. These interactions can be categorized by type and summarized, as shown in the plot above. Protein-ligand interactions (or 'contacts') are categorized into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges. Each interaction type contains more specific subtypes, which can be explored through the 'Simulation Interactions Diagram' panel. The stacked bar charts are normalized over the course of the trajectory: for example, a value of 0.7 suggests that 70% of the simulation time the specific interaction is maintained. Values over 1.0 are possible as some protein residue may make multiple contacts of same subtype with the ligand.

**Hydrogen Bonds:** (H-bonds) play a significant role in ligand binding. Consideration of hydrogen-bonding properties in drug design is important because of their strong influence on drug specificity, metabolism and adsorption. Hydrogen bonds between a protein and a ligand can be further broken down into four subtypes: backbone acceptor; backbone donor; side-chain acceptor; side-chain donor.

The current geometric criteria for protein-ligand H-bond is: distance of 2.5 Å between the donor and acceptor atoms (D—H...A); a donor angle of  $\geq 120^\circ$  between the donor-hydrogen-acceptor atoms (D—H...A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H...A—X).

**Hydrophobic contacts:** fall into three subtypes:  $\pi$ -Cation;  $\pi$ - $\pi$ ; and Other, non-specific interactions. Generally these type of interactions involve a hydrophobic amino acid and an aromatic or aliphatic group on the ligand, but we have extended this category to also include  $\pi$ -Cation interactions.

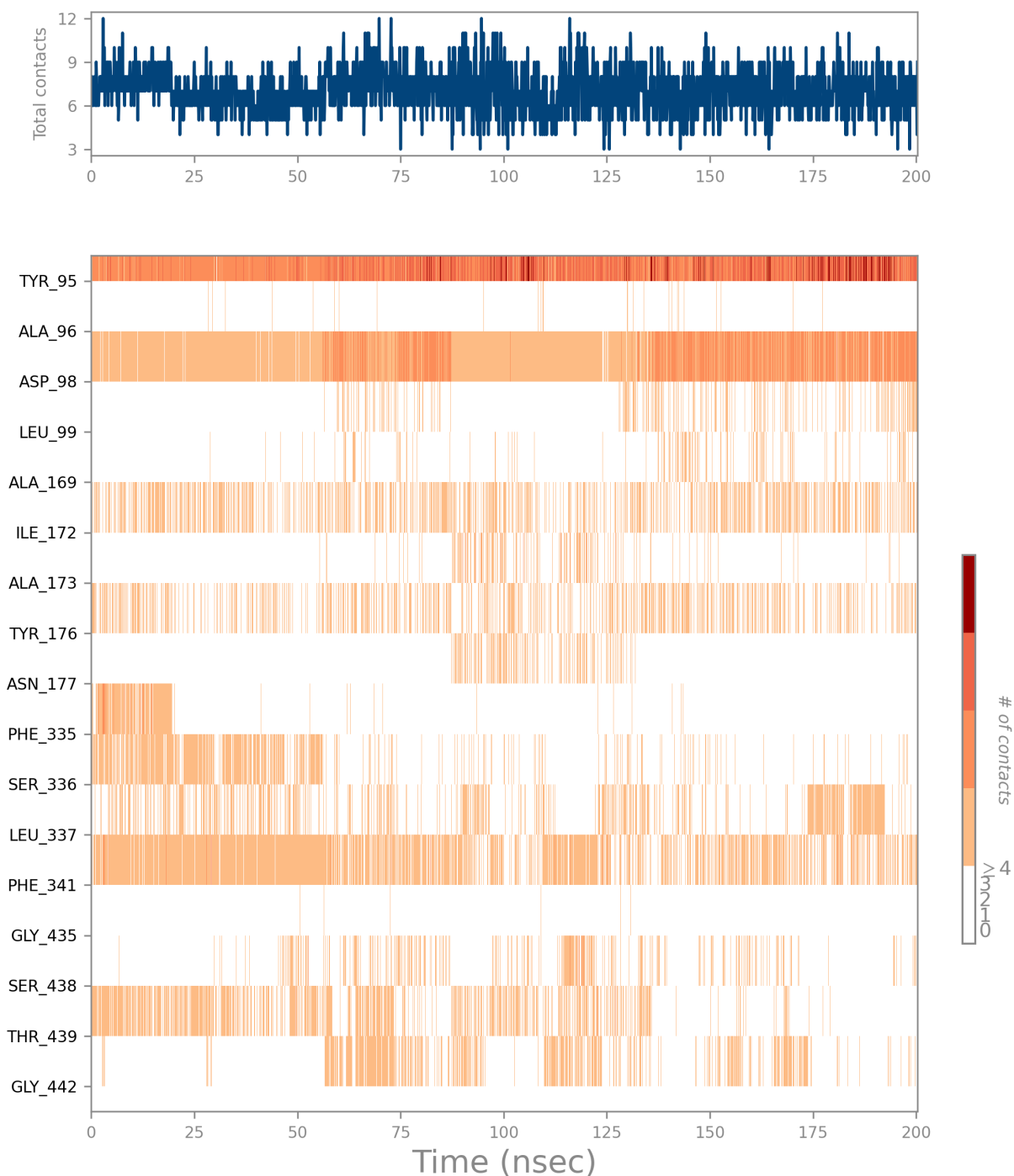
The current geometric criteria for hydrophobic interactions is as follows:  $\pi$ -Cation — Aromatic and charged groups within 4.5 Å;  $\pi$ - $\pi$  — Two aromatic groups stacked face-to-face or face-to-edge; Other — A non-specific hydrophobic sidechain within 3.6 Å of a ligand's aromatic or aliphatic carbons.

**Ionic interactions:** or polar interactions, are between two oppositely charged atoms that are within 3.7 Å of each other and do not involve a hydrogen bond. We also monitor Protein-Metal-Ligand interactions, which are defined by a metal ion coordinated within 3.4 Å of protein's and ligand's heavy atoms (except carbon). All ionic interactions are broken down into two subtypes: those mediated by a protein backbone or side chains.

**Water Bridges:** are hydrogen-bonded protein-ligand interactions mediated by a water molecule. The hydrogen-bond geometry is slightly relaxed from the standard H-bond definition.

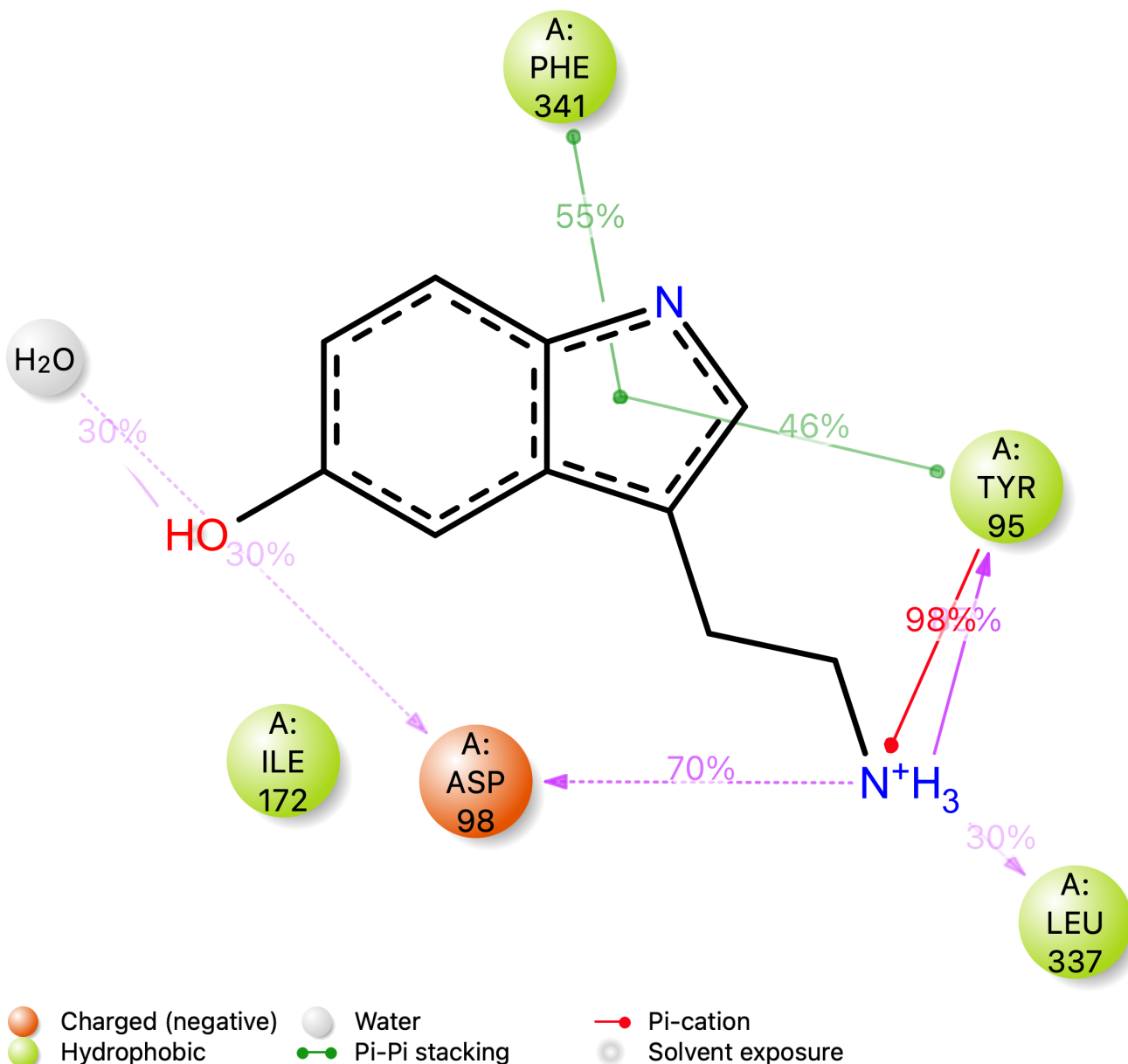
The current geometric criteria for a protein-water or water-ligand H-bond are: a distance of 2.8 Å between the donor and acceptor atoms (D—H...A); a donor angle of  $\geq 110^\circ$  between the donor-hydrogen-acceptor atoms (D—H...A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H...A—X).

## Protein-Ligand Contacts (cont.)



A timeline representation of the interactions and contacts (**H-bonds, Hydrophobic, Ionic, Water bridges**) summarized in the previous page. The top panel shows the total number of specific contacts the protein makes with the ligand over the course of the trajectory. The bottom panel shows which residues interact with the ligand in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot.

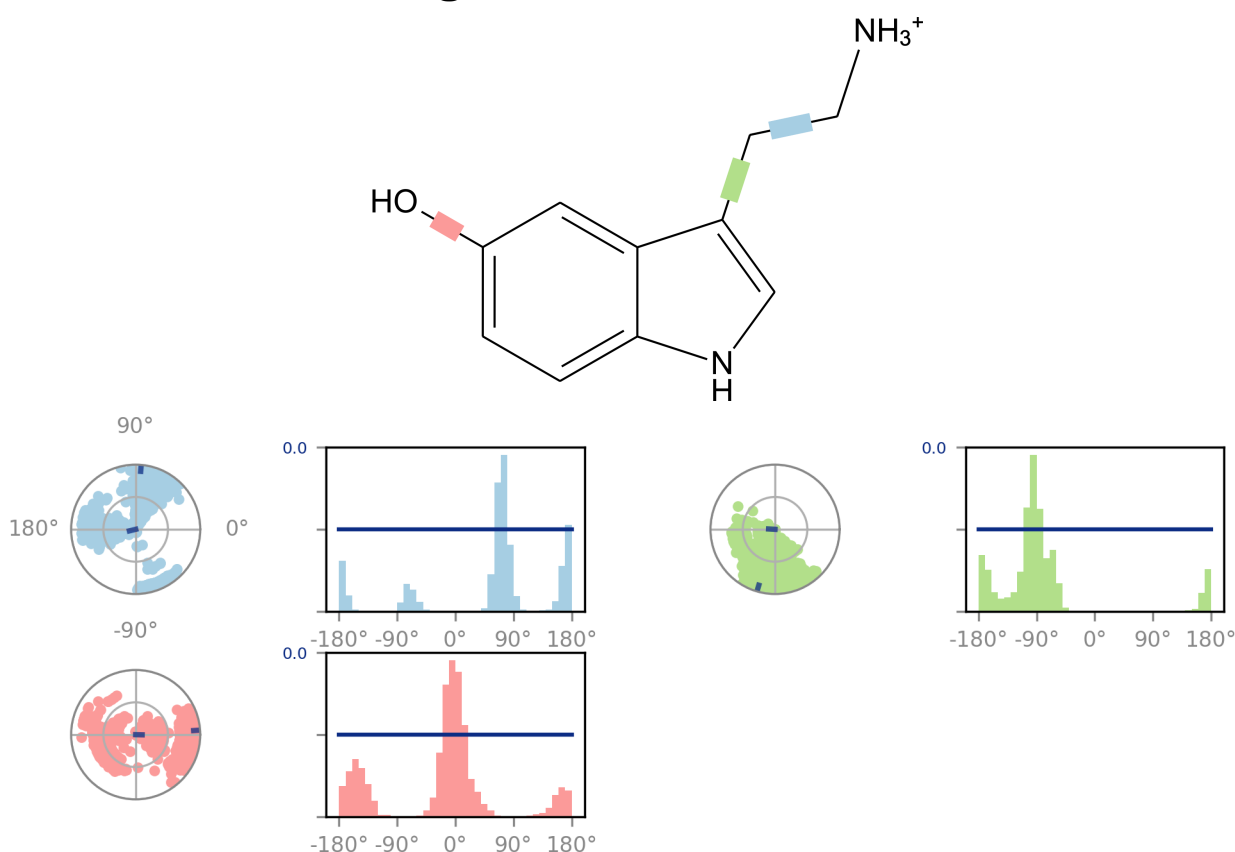
## Ligand-Protein Contacts



A schematic of detailed ligand atom interactions with the protein residues. Interactions that occur more than **30.0%** of the simulation time in the selected trajectory ( 0.00 through 200.30 nsec), are shown. Note: it is possible to have interactions with >100% as some residues may have multiple interactions of a single type with the same ligand atom. For example, the ARG side chain has four H-bond donors that can all hydrogen-bond to a single H-bond acceptor.



## Ligand Torsion Profile

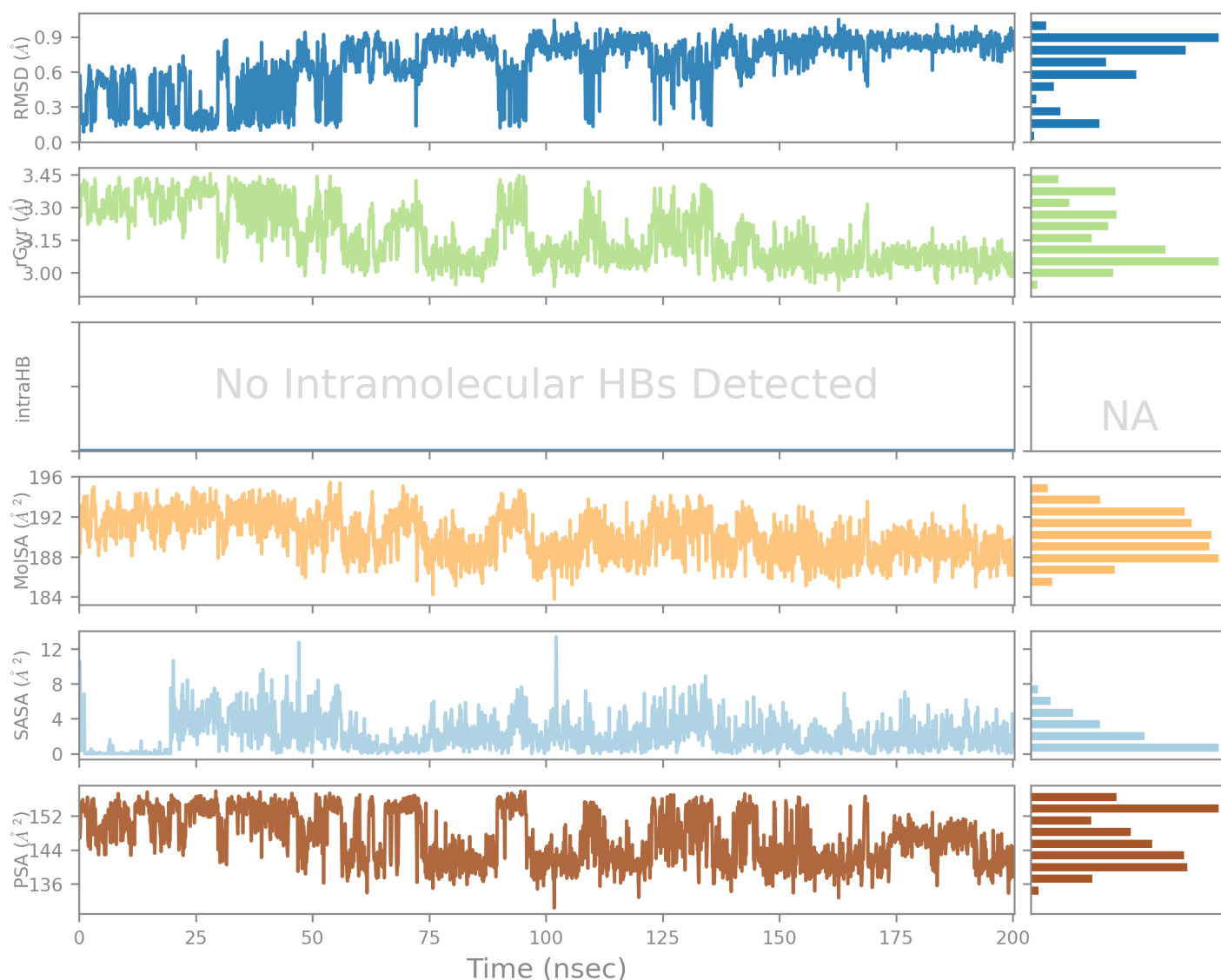


The ligand torsions plot summarizes the conformational evolution of every rotatable bond (RB) in the ligand throughout the simulation trajectory ( 0.00 through 200.30 nsec). The top panel shows the 2d schematic of a ligand with color-coded rotatable bonds. Each rotatable bond torsion is accompanied by a dial plot and bar plots of the same color.

Dial (or radial) plots describe the conformation of the torsion throughout the course of the simulation. The beginning of the simulation is in the center of the radial plot and the time evolution is plotted radially outwards.

The bar plots summarize the data on the dial plots, by showing the probability density of the torsion. If torsional potential information is available, the plot also shows the potential of the rotatable bond (by summing the potential of the related torsions). The values of the potential are on the left Y-axis of the chart, and are expressed in *kcal/mol*. Looking at the histogram and torsion potential relationships may give insights into the conformational strain the ligand undergoes to maintain a protein-bound conformation.

## Ligand Properties



**Ligand RMSD:** Root mean square deviation of a ligand with respect to the reference conformation (typically the first frame is used as the reference and it is regarded as time  $t=0$ ).

**Radius of Gyration (rGyr):** Measures the 'extendedness' of a ligand, and is equivalent to its principal moment of inertia.

**Intramolecular Hydrogen Bonds (intraHB):** Number of internal hydrogen bonds (HB) within a ligand molecule.

**Molecular Surface Area (MolISA):** Molecular surface calculation with 1.4 Å probe radius. This value is equivalent to a van der Waals surface area.

**Solvent Accessible Surface Area (SASA):** Surface area of a molecule accessible by a water molecule.

**Polar Surface Area (PSA):** Solvent accessible surface area in a molecule contributed only by oxygen and nitrogen atoms.

# Simulation Interactions Diagram Report

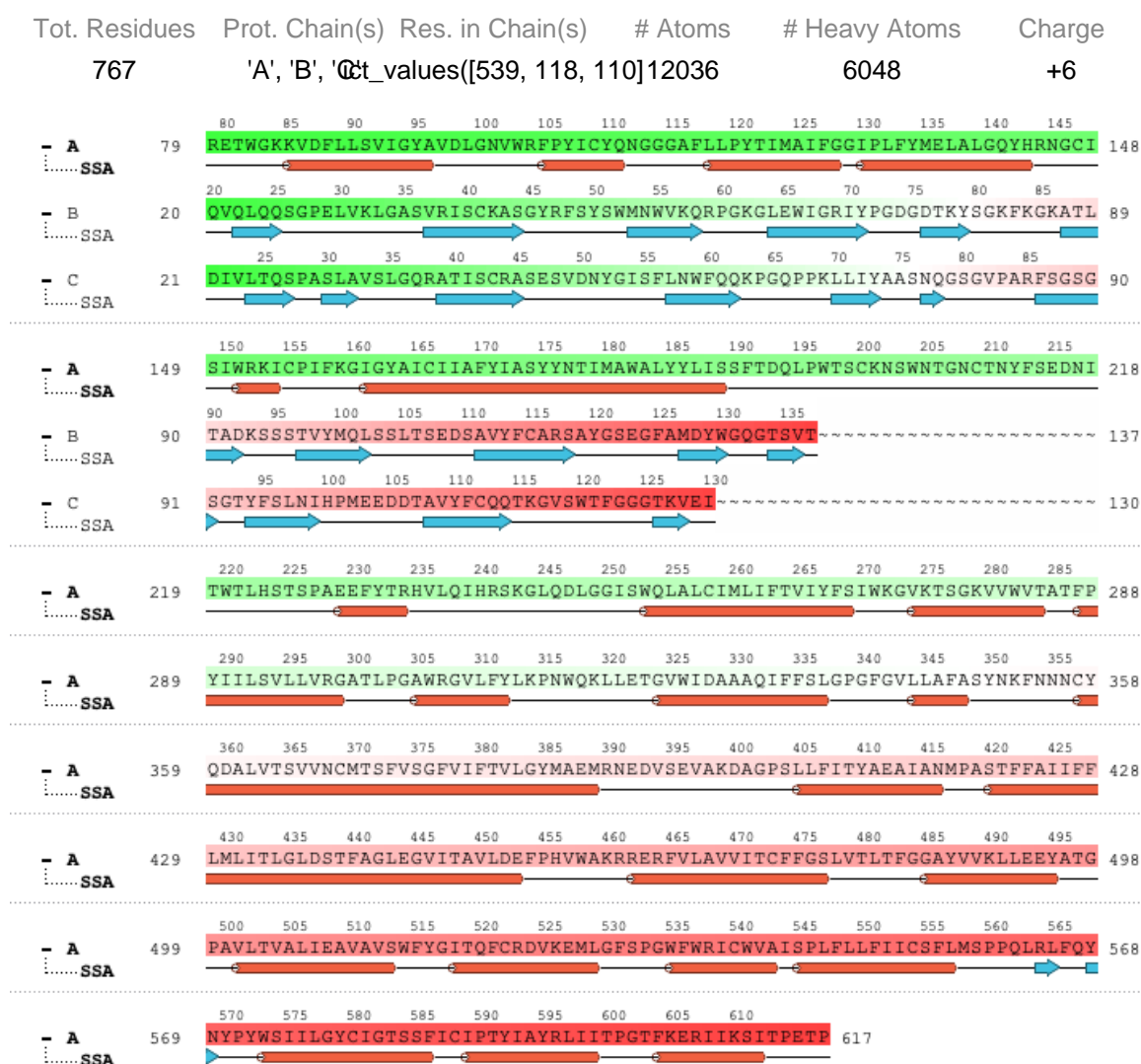
## Simulation Details

Jobname: merged  
Entry title: MMMP

| CPU #    | Job Type | Ensemble | Temp. [K] | Sim. Time [ns] | # Atoms | # Waters | Charge |
|----------|----------|----------|-----------|----------------|---------|----------|--------|
| Unknown* | Unknown* | Unknown* | 300.0     | 200.404        | 100665  | 29471    | 0      |

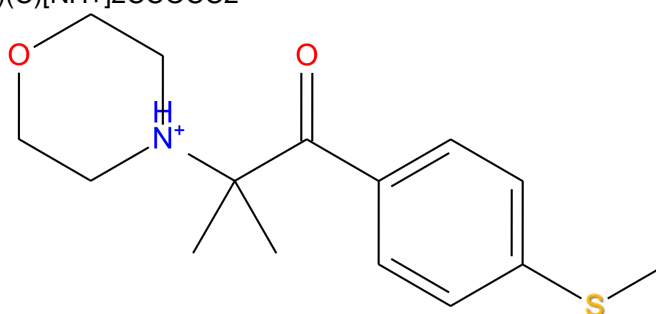
\* The configuration file (-out.cfg) was not found. Keep it in same directory as .aef file.

## Protein Information



## Ligand Information

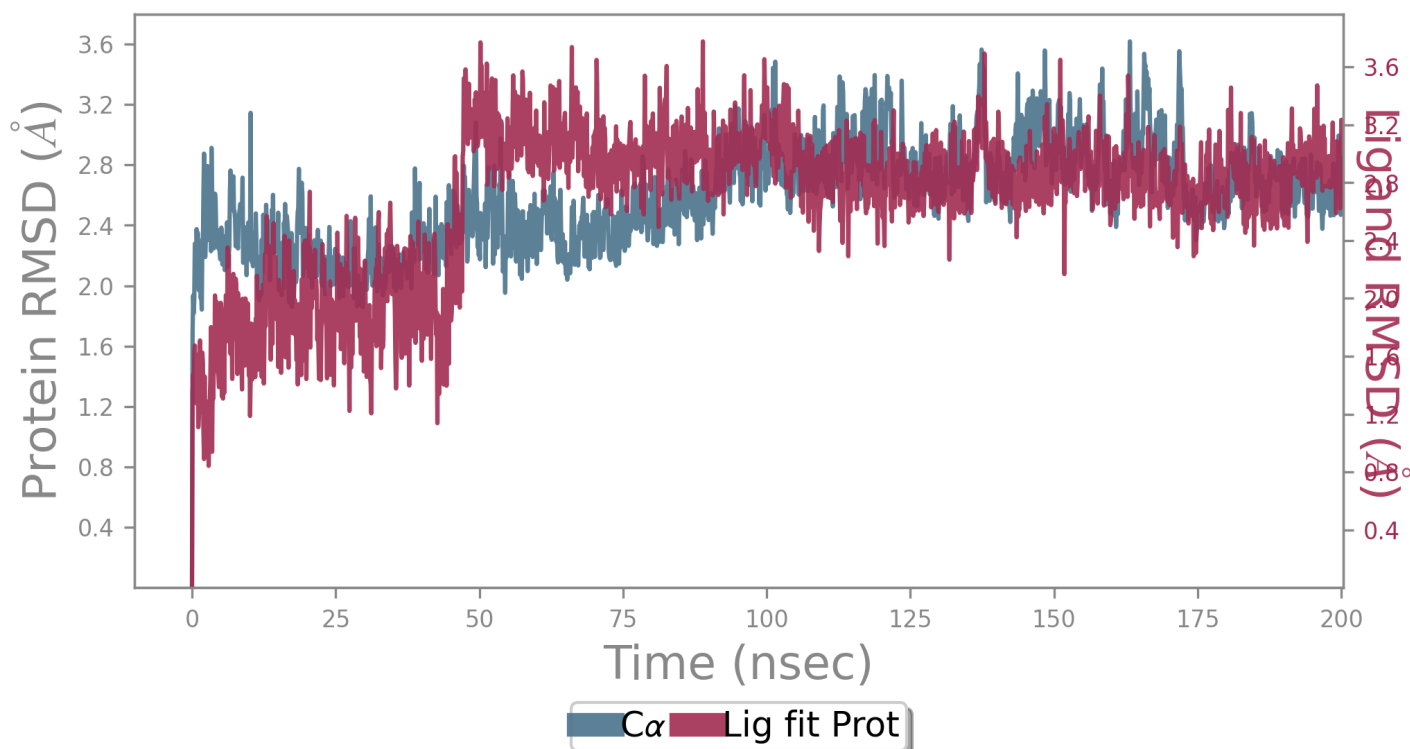
|                   |                                     |
|-------------------|-------------------------------------|
| SMILES            | CSc(cc1ccc1C(=O)C(C)(C)[NH+]2CCOCC2 |
| PDB Name          | 'UNK'                               |
| Num. of Atoms     | 41 (total) 19 (heavy)               |
| Atomic Mass       | 280.412 au                          |
| Charge            | +1                                  |
| Mol. Formula      | C15H22NO2S                          |
| Num. of Fragments | 2                                   |



## Counter Ion/Salt Information

| Type | Num. | Concentration [mM] | Total Charge |
|------|------|--------------------|--------------|
| Cl   | 91   | 56.141             | -91          |
| Na   | 84   | 51.823             | +84          |

## Protein-Ligand RMSD



The Root Mean Square Deviation (RMSD) is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The RMSD for frame  $x$  is:

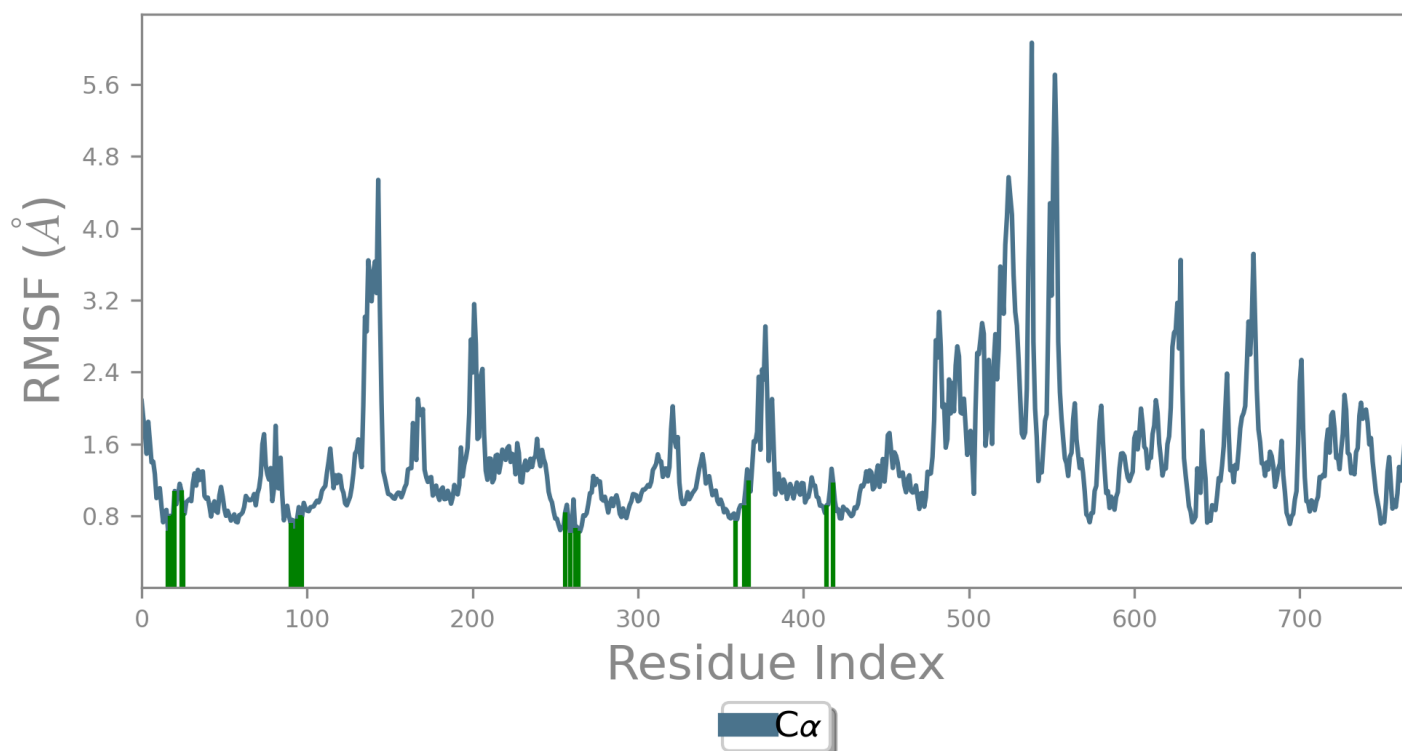
$$RMSD_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'_i(t_x) - r_i(t_{ref}))^2}$$

where  $N$  is the number of atoms in the atom selection;  $t_{ref}$  is the reference time, (typically the first frame is used as the reference and it is regarded as time  $t=0$ ); and  $r'_i$  is the position of the selected atoms in frame  $x$  after superimposing on the reference frame, where frame  $x$  is recorded at time  $t_x$ . The procedure is repeated for every frame in the simulation trajectory.

**Protein RMSD:** The above plot shows the RMSD evolution of a protein (left Y-axis). All protein frames are first aligned on the reference frame backbone, and then the RMSD is calculated based on the atom selection. Monitoring the RMSD of the protein can give insights into its structural conformation throughout the simulation. RMSD analysis can indicate if the simulation has equilibrated — its fluctuations towards the end of the simulation are around some thermal average structure. Changes of the order of 1-3 Å are perfectly acceptable for small, globular proteins. Changes much larger than that, however, indicate that the protein is undergoing a large conformational change during the simulation. It is also important that your simulation converges — the RMSD values stabilize around a fixed value. If the RMSD of the protein is still increasing or decreasing on average at the end of the simulation, then your system has not equilibrated, and your simulation may not be long enough for rigorous analysis.

**Ligand RMSD:** Ligand RMSD (right Y-axis) indicates how stable the ligand is with respect to the protein and its binding pocket. In the above plot, 'Lig fit Prot' shows the RMSD of a ligand when the protein-ligand complex is first aligned on the protein backbone of the reference and then the RMSD of the ligand heavy atoms is measured. If the values observed are significantly larger than the RMSD of the protein, then it is likely that the ligand has diffused away from its initial binding site.

## Protein RMSF



The Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain. The RMSF for residue  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \langle (r'_i(t)) - r_i(t_{ref})^2 \rangle}$$

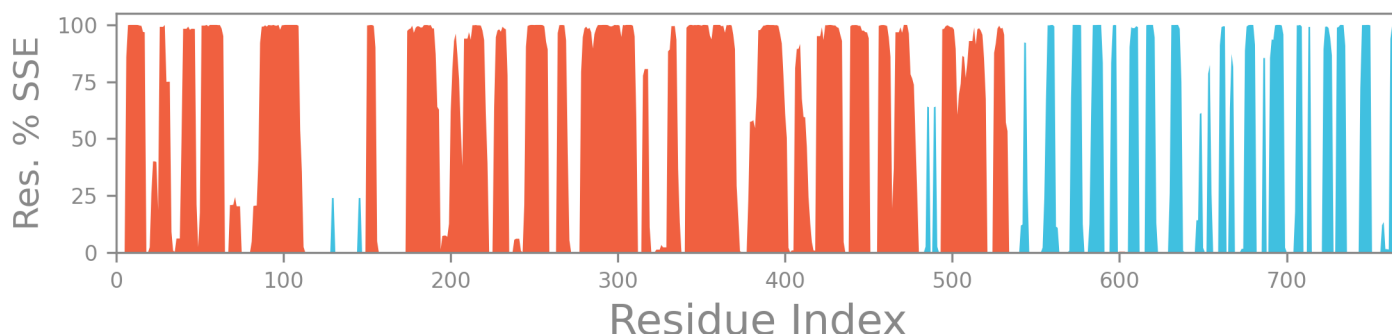
where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time,  $r_i$  is the position of residue  $i$ ;  $r'$  is the position of atoms in residue  $i$  after superposition on  $r_{ref}$ , and the  $\langle \rangle$  angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

On this plot, peaks indicate areas of the protein that fluctuate the most during the simulation. Typically you will observe that the tails ( $N$ - and  $C$ -terminal) fluctuate more than any other part of the protein. Secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein, and thus fluctuate less than the loop regions.

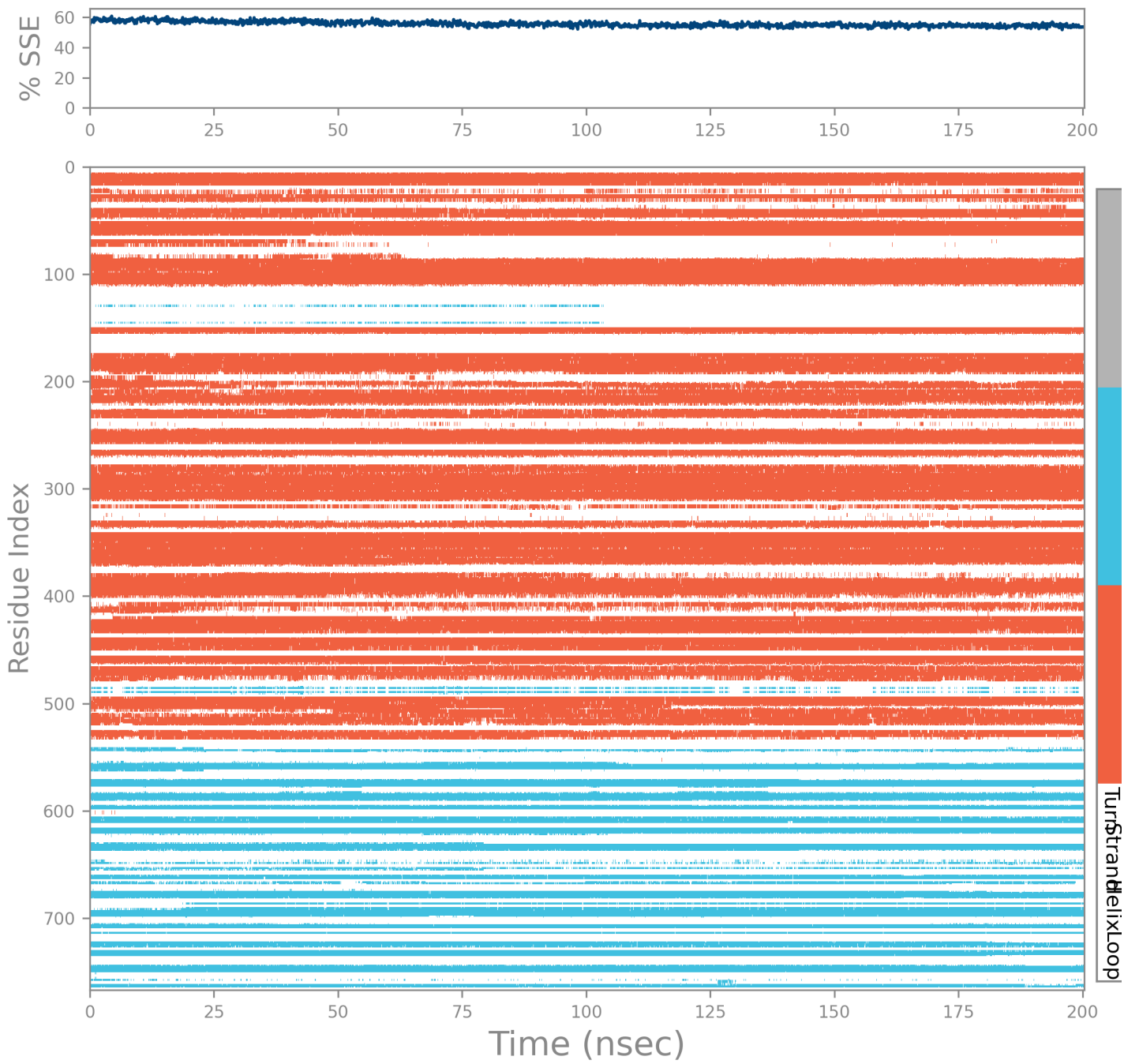
**Ligand Contacts:** Protein residues that interact with the ligand are marked with green-colored vertical bars.

# Protein Secondary Structure

|         |          |             |
|---------|----------|-------------|
| % Helix | % Strand | % Total SSE |
| 42.01   | 13.51    | 55.51       |

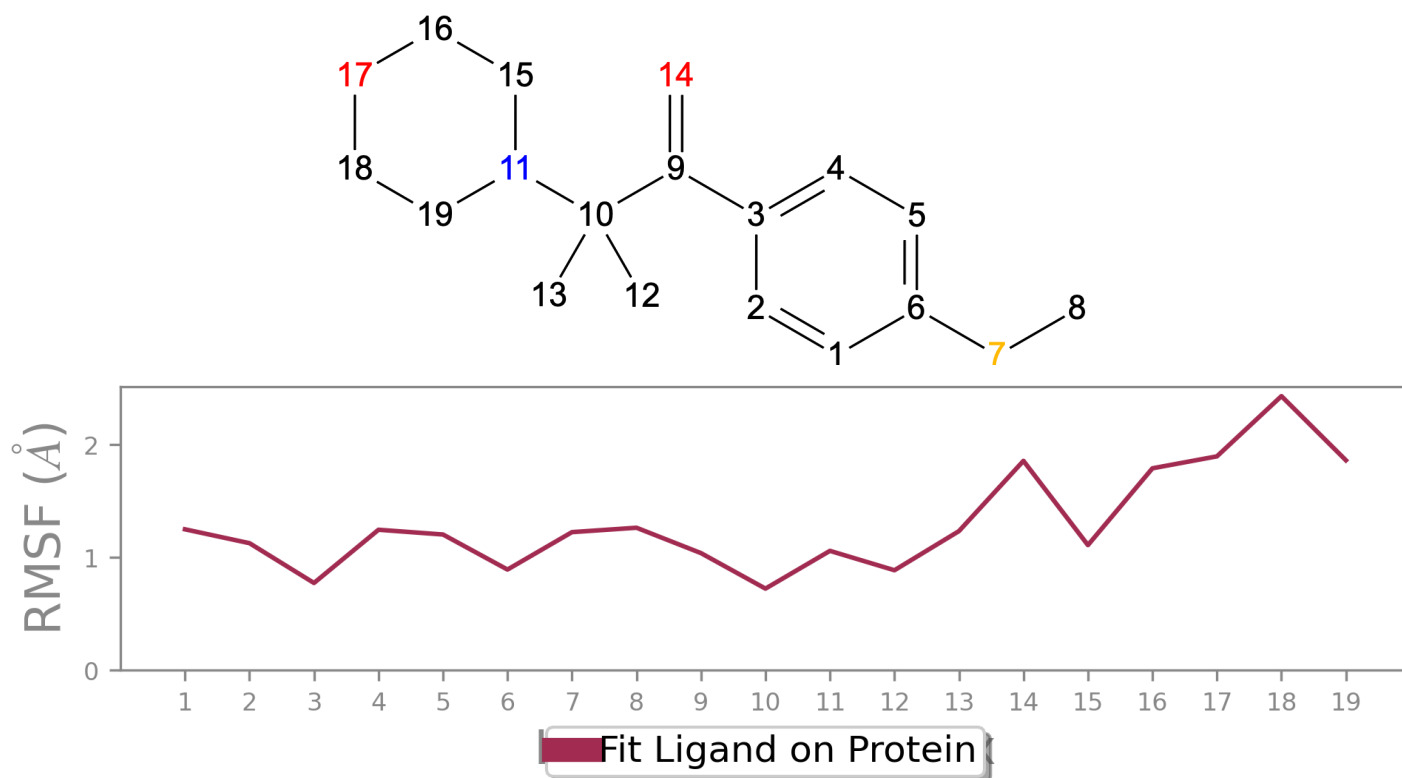


Protein secondary structure elements (SSE) like **alpha-helices** and **beta-strands** are monitored throughout the simulation. The plot above reports SSE distribution by residue index throughout the protein structure. The plot below summarizes the SSE composition for each trajectory frame over the course of the simulation, and the plot at the bottom monitors each residue and its SSE assignment over time.





## Ligand RMSF



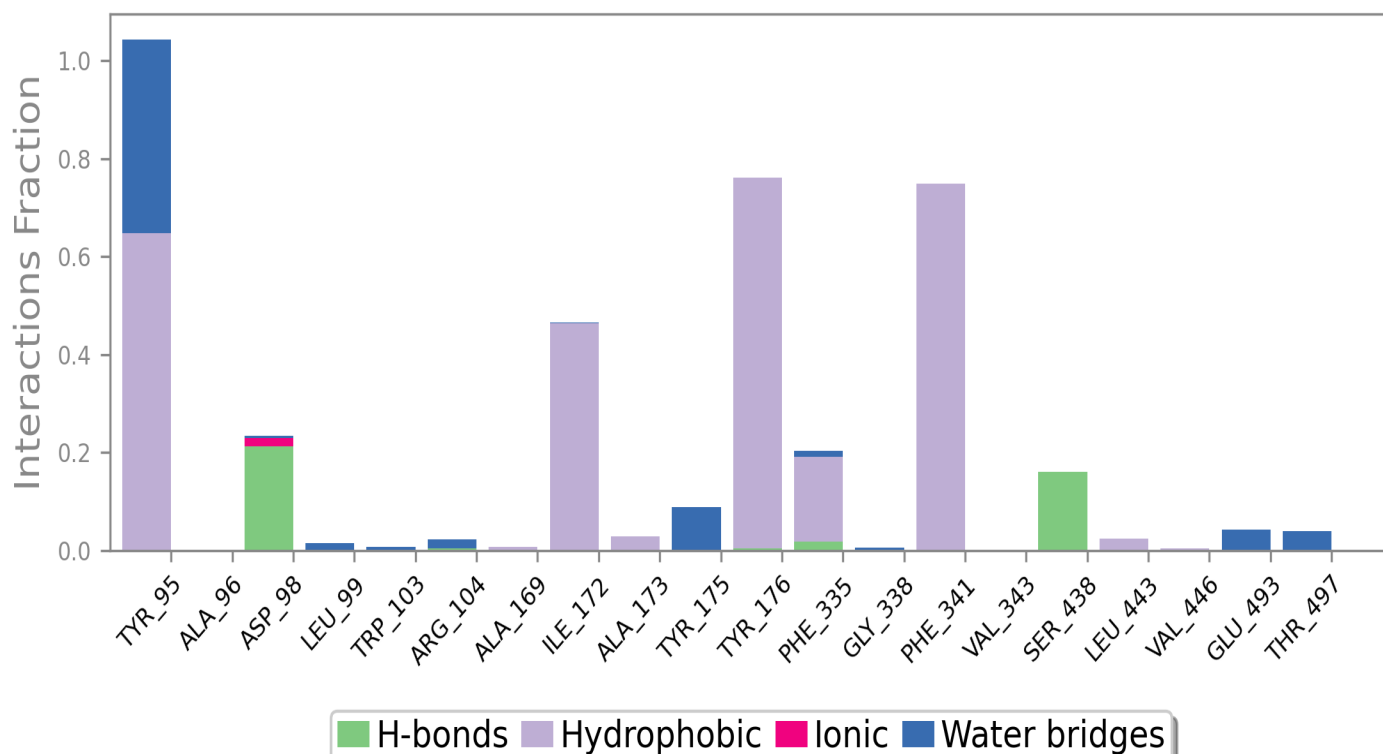
The Ligand Root Mean Square Fluctuation (L-RMSF) is useful for characterizing changes in the ligand atom positions. The RMSF for atom  $i$  is:

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (r'_i(t) - r_i(t_{ref}))^2}$$

where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{ref}$  is the reference time (usually for the first frame, and is regarded as the zero of time);  $r$  is the position of atom  $i$  in the reference at time  $t_{ref}$  and  $r'$  is the position of atom  $i$  at time  $t$  after superposition on the reference frame.

Ligand RMSF shows the ligand's fluctuations broken down by atom, corresponding to the 2D structure in the top panel. The ligand RMSF may give you insights on how ligand fragments interact with the protein and their entropic role in the binding event. In the bottom panel, the 'Fit Ligand on Protein' line shows the ligand fluctuations, with respect to the protein. The protein-ligand complex is first aligned on the protein backbone and then the ligand RMSF is measured on the ligand heavy atoms.

## Protein-Ligand Contacts



Protein interactions with the ligand can be monitored throughout the simulation. These interactions can be categorized by type and summarized, as shown in the plot above. Protein-ligand interactions (or 'contacts') are categorized into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges. Each interaction type contains more specific subtypes, which can be explored through the 'Simulation Interactions Diagram' panel. The stacked bar charts are normalized over the course of the trajectory: for example, a value of 0.7 suggests that 70% of the simulation time the specific interaction is maintained. Values over 1.0 are possible as some protein residue may make multiple contacts of same subtype with the ligand.

**Hydrogen Bonds:** (H-bonds) play a significant role in ligand binding. Consideration of hydrogen-bonding properties in drug design is important because of their strong influence on drug specificity, metabolism and adsorption. Hydrogen bonds between a protein and a ligand can be further broken down into four subtypes: backbone acceptor; backbone donor; side-chain acceptor; side-chain donor.

The current geometric criteria for protein-ligand H-bond is: distance of 2.5 Å between the donor and acceptor atoms (D—H...A); a donor angle of  $\geq 120^\circ$  between the donor-hydrogen-acceptor atoms (D—H...A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H...A—X).

**Hydrophobic contacts:** fall into three subtypes:  $\pi$ -Cation;  $\pi$ - $\pi$ ; and Other, non-specific interactions. Generally these type of interactions involve a hydrophobic amino acid and an aromatic or aliphatic group on the ligand, but we have extended this category to also include  $\pi$ -Cation interactions.

The current geometric criteria for hydrophobic interactions is as follows:  $\pi$ -Cation — Aromatic and charged groups within 4.5 Å;  $\pi$ - $\pi$  — Two aromatic groups stacked face-to-face or face-to-edge; Other — A non-specific hydrophobic sidechain within 3.6 Å of a ligand's aromatic or aliphatic carbons.

**Ionic interactions:** or polar interactions, are between two oppositely charged atoms that are within 3.7 Å of each other and do not involve a hydrogen bond. We also monitor Protein-Metal-Ligand interactions, which are defined by a metal ion coordinated within 3.4 Å of protein's and ligand's heavy atoms (except carbon). All ionic interactions are broken down into two subtypes: those mediated by a protein backbone or side chains.

**Water Bridges:** are hydrogen-bonded protein-ligand interactions mediated by a water molecule. The hydrogen-bond geometry is slightly relaxed from the standard H-bond definition.

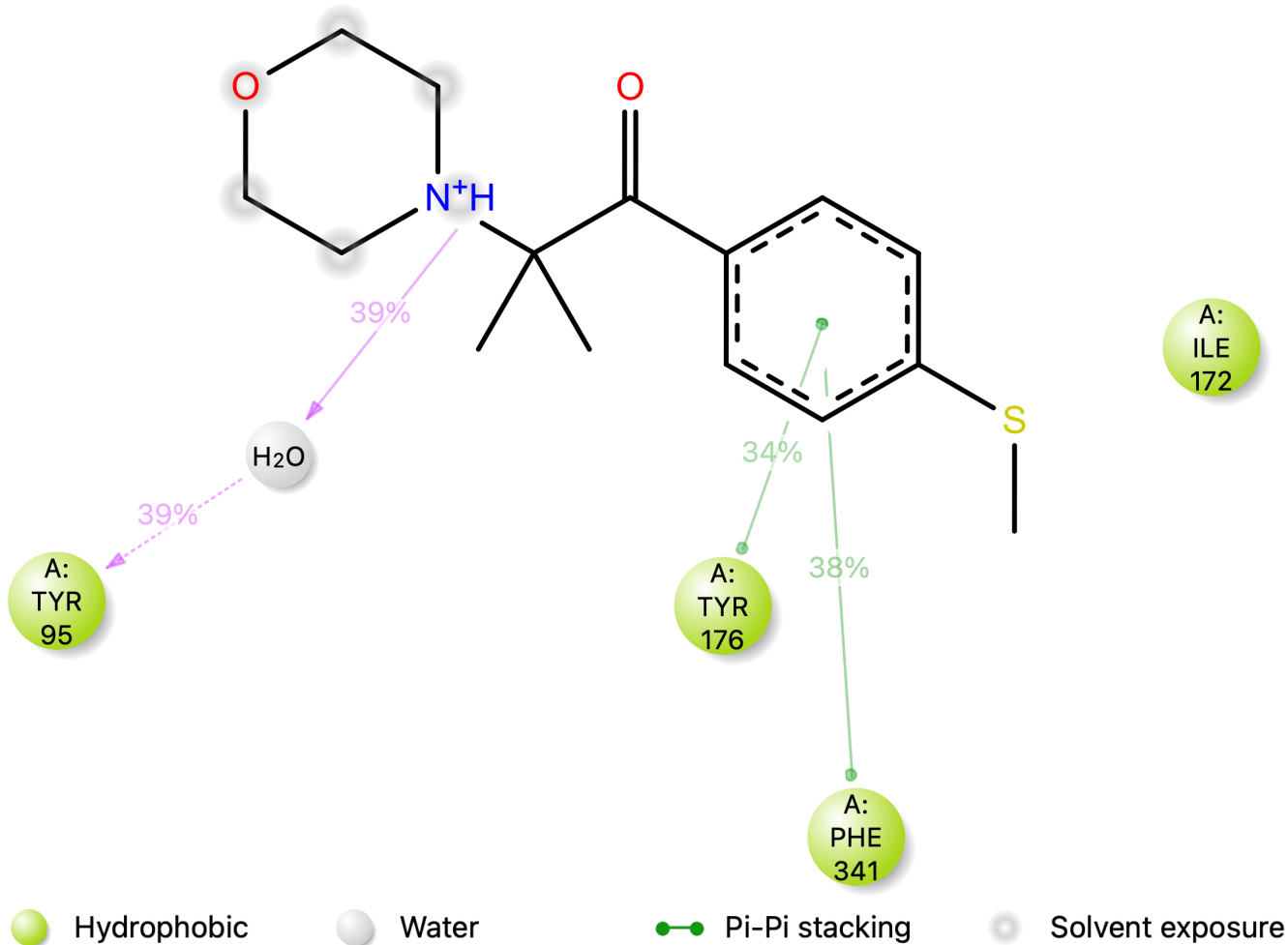
The current geometric criteria for a protein-water or water-ligand H-bond are: a distance of 2.8 Å between the donor and acceptor atoms (D—H...A); a donor angle of  $\geq 110^\circ$  between the donor-hydrogen-acceptor atoms (D—H...A); and an acceptor angle of  $\geq 90^\circ$  between the hydrogen-acceptor-bonded\_atom atoms (H...A—X).

## Protein-Ligand Contacts (cont.)



A timeline representation of the interactions and contacts (**H-bonds, Hydrophobic, Ionic, Water bridges**) summarized in the previous page. The top panel shows the total number of specific contacts the protein makes with the ligand over the course of the trajectory. The bottom panel shows which residues interact with the ligand in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot.

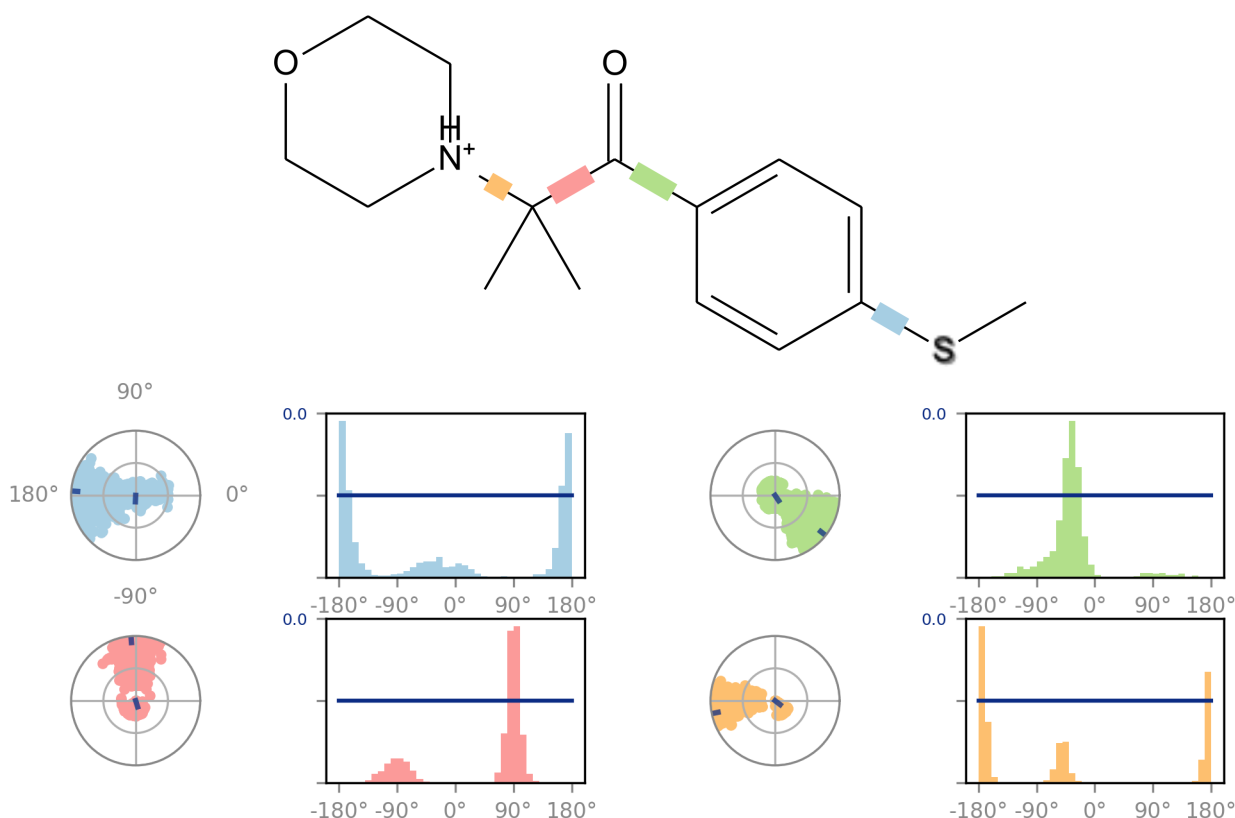
## Ligand-Protein Contacts



A schematic of detailed ligand atom interactions with the protein residues. Interactions that occur more than **30.0%** of the simulation time in the selected trajectory ( 0.00 through 200.30 nsec), are shown.

Note: it is possible to have interactions with >100% as some residues may have multiple interactions of a single type with the same ligand atom. For example, the ARG side chain has four H-bond donors that can all hydrogen-bond to a single H-bond acceptor.

## Ligand Torsion Profile

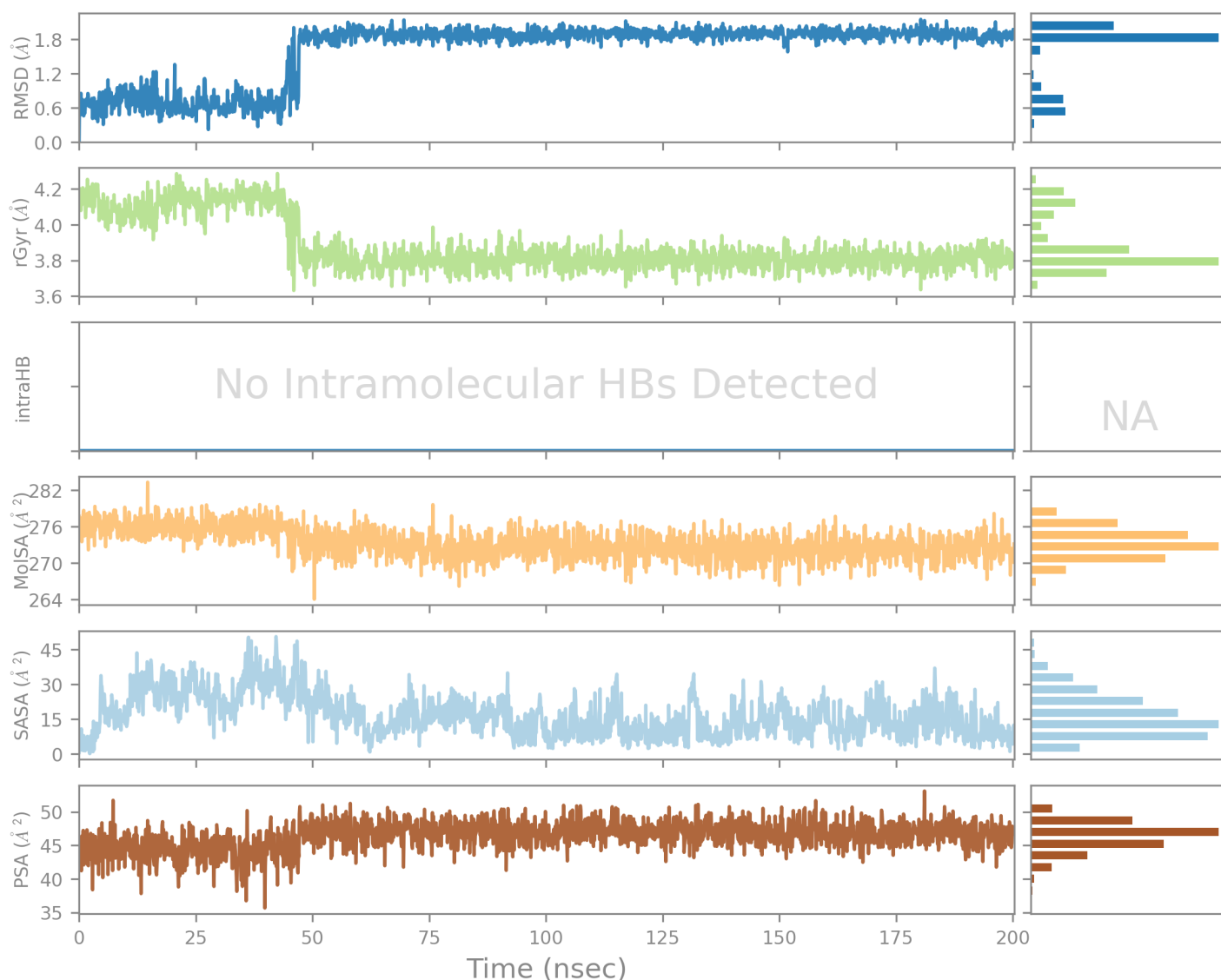


The ligand torsions plot summarizes the conformational evolution of every rotatable bond (RB) in the ligand throughout the simulation trajectory ( 0.00 through 200.30 nsec). The top panel shows the 2d schematic of a ligand with color-coded rotatable bonds. Each rotatable bond torsion is accompanied by a dial plot and bar plots of the same color.

Dial (or radial) plots describe the conformation of the torsion throughout the course of the simulation. The beginning of the simulation is in the center of the radial plot and the time evolution is plotted radially outwards.

The bar plots summarize the data on the dial plots, by showing the probability density of the torsion. If torsional potential information is available, the plot also shows the potential of the rotatable bond (by summing the potential of the related torsions). The values of the potential are on the left Y-axis of the chart, and are expressed in *kcal/mol*. Looking at the histogram and torsion potential relationships may give insights into the conformational strain the ligand undergoes to maintain a protein-bound conformation.

## Ligand Properties



**Ligand RMSD:** Root mean square deviation of a ligand with respect to the reference conformation (typically the first frame is used as the reference and it is regarded as time  $t=0$ ).

**Radius of Gyration (rGyr):** Measures the 'extendedness' of a ligand, and is equivalent to its principal moment of inertia.

**Intramolecular Hydrogen Bonds (intraHB):** Number of internal hydrogen bonds (HB) within a ligand molecule.

**Molecular Surface Area (MolISA):** Molecular surface calculation with 1.4 Å probe radius. This value is equivalent to a van der Waals surface area.

**Solvent Accessible Surface Area (SASA):** Surface area of a molecule accessible by a water molecule.

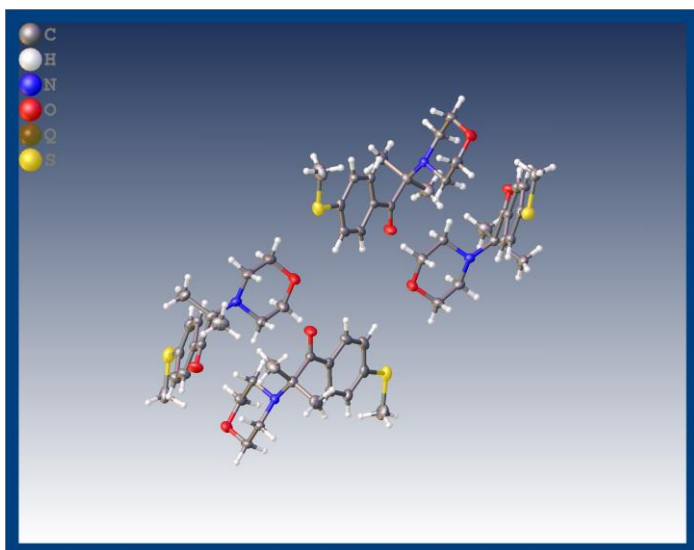
**Polar Surface Area (PSA):** Solvent accessible surface area in a molecule contributed only by oxygen and nitrogen atoms.

## Id021\_etoh\_lt\_auto

Sample ID: Id021\_etoh\_lt\_auto

 **$R_1=4.66\%$** 

## Crystal Data and Experimental



**Experimental.** Single None None None ? -shaped crystals of **Id021\_etoh\_lt\_auto** were obtained by recrystallisation from .... A suitable crystal ? ×? ×? mm<sup>3</sup> was selected and mounted on a suitable support on an XtaLAB Synergy R, DW system, HyPix diffractometer. The crystal was kept at a steady  $T = 152.0(2)$  K during data collection. The structure was solved with the **ShelXT** (Sheldrick, 2015) structure solution program using the dual solution method and by using **Olex2** (Dolomanov et al., 2009) as the graphical interface. The model was refined with version of olex2.refine 1.5 (Bourhis et al., 2015) using Gauss-Newton minimisation.

**Crystal Data.** C<sub>60</sub>H<sub>84</sub>N<sub>4</sub>O<sub>8</sub>S<sub>4</sub>,  $M_r = 1117.622$ , orthorhombic,  $Pca2_1$  (No. 29),  $a = 37.0231(3)$  Å,  $b = 6.5766(1)$  Å,  $c = 24.1761(3)$  Å,  $\alpha = \beta = \gamma = 90^\circ$ ,  $V = 5886.54(12)$  Å<sup>3</sup>,  $T = 152.0(2)$  K,  $Z = 4$ ,  $Z' = 1$ ,  $\mu(\text{Cu } K\alpha) = 1.934$ , 32547 reflections measured, 9767 unique ( $R_{int} = 0.0405$ ) which were used in all calculations. The final  $wR_2$  was 0.1261 (all data) and  $R_1$  was 0.0466 ( $I \geq \sigma(I)$ ).

| Compound                            | Id021_etoh_lt_auto   |
|-------------------------------------|--|
| Formula                             | C <sub>60</sub> H <sub>84</sub> N <sub>4</sub> O <sub>8</sub> S <sub>4</sub> |
| $D_{calc.}/\text{g cm}^{-3}$        | 1.261  |
| $\mu/\text{mm}^{-1}$                | 1.934  |
| Formula Weight                      | 1117.622   |
| Colour                              | None None None   |
| Shape                               | ?  |
| Size/mm <sup>3</sup>                | ? ×? ×?  |
| $T/\text{K}$                        | 152.0(2)   |
| Crystal System                      | orthorhombic   |
| Flack Parameter                     | 0.302(6)   |
| Hooft Parameter                     | 0.302(6)   |
| Space Group                         | $Pca2_1$   |
| $a/\text{\AA}$                      | 37.0231(3)   |
| $b/\text{\AA}$                      | 6.5766(1)  |
| $c/\text{\AA}$                      | 24.1761(3)   |
| $\alpha^\circ$                      | 90   |
| $\beta^\circ$                       | 90   |
| $\gamma^\circ$                      | 90   |
| $V/\text{\AA}^3$                    | 5886.54(12)  |
| $Z$                                 | 4  |
| $Z'$                                | 1  |
| Wavelength/Å                        | 1.54184  |
| Radiation type                      | Cu $K\alpha$   |
| $\theta_{min}/^\circ$               | 3.01   |
| $\theta_{max}/^\circ$               | 72.11  |
| Measured Refl.                      | 32547  |
| Independent Refl.                   | 9767   |
| Reflections with $I \geq \sigma(I)$ | 9312   |
| $R_{int}$                           | 0.0405   |
| Parameters                          | 697  |
| Restraints                          | 1  |
| Largest Peak                        | 0.7245   |
| Deepest Hole                        | -0.2797  |
| GooF                                | 1.0475   |
| $wR_2$ (all data)                   | 0.1261   |
| $wR_2$                              | 0.1233   |
| $R_1$ (all data)                    | 0.0490   |
| $R_1$                               | 0.0466   |

## Structure Quality Indicators

|                     |   |        |                 |      |                |       |            |         |
|---------------------|---|--------|-----------------|------|----------------|-------|------------|---------|
| <b>Reflections:</b> | d min (CuK $\alpha$ )<br>2 $\theta$ =144.2° | 0.81   | I/ $\sigma$ (I) | 29.1 | Rint<br>m=3.50 | 4.04% | CAP 144.0° | 99.5    |
| <b>Refinement:</b>  | Shift                                       | -0.001 | Max Peak        | 0.7  | Min Peak       | -0.3  | Goof       | 1.048   |
|                     |   |        |                 |      |                |       | Hooft      | .302(6) |

**Experimental Extended.** A None None None ? -shaped crystal with dimensions ?  $\times$ ?  $\times$ ? mm<sup>3</sup> was mounted on a suitable support. Data were collected using an XtaLAB Synergy R, DW system, HyPix diffractometer operating at  $T = 152.0(2)$  K.

Data were measured using  $\omega$  scans of 0.5° per frame for 0.1 s using Cu K $\alpha$  radiation. The diffraction pattern was indexed and the total number of runs and images was based on the strategy calculation from the program CrysAlisPro (Rigaku, V1.171.41.97a, 2021) The maximum resolution that was achieved was  $\theta = 72.11^\circ$  (0.81 Å).

The diffraction pattern was indexed The diffraction pattern was indexed and the total number of runs and images was based on the strategy calculation from the program CrysAlisPro (Rigaku, V1.171.41.97a, 2021) and the unit cell was refined using CrysAlisPro (Rigaku, V1.171.41.97a, 2021) on 23055 reflections, 71% of the observed reflections.

Data reduction, scaling and absorption corrections were performed using CrysAlisPro (Rigaku, V1.171.41.97a, 2021). The final completeness is 99.89 % out to 72.11° in  $\theta$ . A multi-scan absorption correction was performed using CrysAlisPro 1.171.41.97a (Rigaku Oxford Diffraction, 2021) using spherical harmonics,implemented in SCALE3 ABSPACK scaling algorithm. The absorption coefficient  $\mu$  of this material is 1.934 mm<sup>-1</sup> at this wavelength ( $\lambda = 1.542\text{Å}$ ) and the minimum and maximum transmissions are 0.839 and 1.000.

The structure was solved and the space group  $Pca2_1$  (# 29) determined by the **ShelXT** (Sheldrick, 2015) structure solution program using dual and refined by Gauss-Newton using version of olex2.refine 1.5 (Bourhis et al., 2015). All non-hydrogen atoms were refined anisotropically. Hydrogen atom positions were calculated geometrically and refined using the riding model. Hydrogen atom positions were calculated geometrically and refined using the riding model.

*\_exptl\_absorpt\_process\_details:* CrysAlisPro 1.171.41.97a (Rigaku Oxford Diffraction, 2021) using spherical harmonics,implemented in SCALE3 ABSPACK scaling algorithm.

*\_exptl\_absorpt\_process\_details:* CrysAlisPro 1.171.41.97a (Rigaku Oxford Diffraction, 2021) using spherical harmonics,implemented in SCALE3 ABSPACK scaling algorithm.

**Table 1:** Fractional Atomic Coordinates ( $\times 10^4$ ) and Equivalent Isotropic Displacement Parameters ( $\text{Å}^2 \times 10^3$ ) for **ld021\_etoht\_ito**.  $U_{eq}$  is defined as 1/3 of the trace of the orthogonalised  $U_{ij}$ .

| Atom | x         | y           | z          | $U_{eq}$  |
|------|-----------|-------------|------------|-----------|
| S41  | 8354.7(2) | 7102.9(13)  | 5818.3(4)  | 33.43(18) |
| S31  | 5852.3(2) | 7423.7(14)  | 7624.5(4)  | 34.61(19) |
| S11  | 6665.5(2) | -3644.2(15) | 2265.5(4)  | 37.2(2)   |
| S21  | 4075.1(2) | -2398.6(15) | 4246.0(4)  | 40.3(2)   |
| O11  | 5762.5(6) | -800(4)     | 4525.5(12) | 40.1(6)   |
| O32  | 7675.0(6) | 12338(3)    | 6302.1(12) | 38.4(6)   |
| O41  | 8137.3(7) | 8684(3)     | 3107.2(10) | 35.0(5)   |
| N11  | 5335.8(6) | -5361(4)    | 4236.6(11) | 21.6(5)   |
| O12  | 4762.0(6) | -7225(3)    | 3641.0(11) | 35.5(5)   |
| O21  | 4391.6(7) | -3232(3)    | 6959.5(10) | 37.1(6)   |
| O31  | 6713.0(6) | 6003(3)     | 5245.2(11) | 36.1(6)   |



| Atom | x          | y        | z          | $U_{eq}$ |
|------|------------|----------|------------|----------|
| O22  | 5570.8(6)  | 2707(4)  | 6367.9(13) | 45.6(7)  |
| O42  | 6975.3(6)  | 2248(3)  | 3490.6(11) | 36.3(6)  |
| N41  | 7659.1(6)  | 4095(4)  | 3277.7(10) | 22.3(5)  |
| N31  | 7116.2(6)  | 10540(4) | 5667.7(10) | 21.7(5)  |
| N21  | 4875.2(7)  | 1173(4)  | 6654.3(11) | 25.1(5)  |
| C35  | 6516.2(8)  | 7742(4)  | 6035.6(14) | 23.7(6)  |
| C415 | 7342.2(8)  | 5303(4)  | 3426.3(13) | 23.9(6)  |
| C34  | 6473.1(8)  | 9493(5)  | 6355.0(14) | 27.1(6)  |
| C36  | 6355.8(8)  | 5929(5)  | 6230.0(14) | 28.0(7)  |
| C47  | 8146.4(8)  | 5360(5)  | 4857.4(13) | 25.9(6)  |
| C18  | 5761.2(8)  | -2515(4) | 4335.3(14) | 27.2(7)  |
| C26  | 4415.9(8)  | -188(5)  | 5701.1(13) | 25.4(6)  |
| C27  | 4323.0(8)  | -425(5)  | 5150.7(14) | 29.8(7)  |
| C28  | 4455.0(8)  | -1737(5) | 6683.7(13) | 26.5(7)  |
| C112 | 5077.9(8)  | -4150(5) | 3920.0(14) | 27.5(7)  |
| C29  | 4583.0(9)  | 209(5)   | 6974.4(13) | 28.1(7)  |
| C13  | 6205.1(8)  | -5095(5) | 3089.0(14) | 29.9(7)  |
| C17  | 6395.3(8)  | -1616(5) | 3144.3(14) | 32.4(7)  |
| C15  | 5983.4(8)  | -2954(5) | 3831.2(14) | 26.4(6)  |
| C14  | 5998.2(8)  | -4840(5) | 3561.0(14) | 28.3(7)  |
| C45  | 8117.7(8)  | 7012(5)  | 3961.8(13) | 23.7(6)  |
| C44  | 8230.0(8)  | 8811(5)  | 4219.5(15) | 30.3(7)  |
| C19  | 5546.9(8)  | -4191(5) | 4643.1(13) | 27.2(7)  |
| C39  | 6910.8(8)  | 9426(5)  | 5236.9(13) | 25.7(6)  |
| C48  | 8063.7(8)  | 7096(4)  | 3346.8(14) | 24.2(6)  |
| C33  | 6277.2(8)  | 9467(5)  | 6840.4(14) | 28.0(7)  |
| C25  | 4379.1(8)  | -1823(5) | 6072.2(13) | 25.5(6)  |
| C42  | 8261.2(8)  | 7195(5)  | 5105.2(13) | 26.2(6)  |
| C38  | 6716.8(8)  | 7607(4)  | 5499.0(14) | 24.8(6)  |
| C37  | 6163.7(8)  | 5890(5)  | 6715.3(14) | 29.6(7)  |
| C12  | 6404.3(8)  | -3496(5) | 2872.5(14) | 30.1(7)  |
| C314 | 7505.6(9)  | 10498(5) | 6469.1(15) | 34.2(8)  |
| C46  | 8075.1(8)  | 5275(5)  | 4295.5(13) | 25.0(6)  |
| C115 | 5147.1(8)  | -7145(4) | 4459.0(14) | 25.7(6)  |
| C315 | 7371.5(8)  | 9295(5)  | 5976.8(14) | 27.9(7)  |
| C32  | 6119.0(8)  | 7658(5)  | 7026.9(14) | 27.4(7)  |
| C414 | 7088.7(8)  | 4043(5)  | 3777.7(15) | 30.9(7)  |
| C313 | 7436.1(9)  | 13546(5) | 5973.4(15) | 30.5(7)  |
| C312 | 7301.2(8)  | 12376(5) | 5473.0(14) | 26.8(6)  |
| C412 | 7534.6(9)  | 2323(4)  | 2966.8(14) | 27.7(7)  |
| C113 | 4932.5(9)  | -5400(5) | 3447.7(15) | 34.2(8)  |
| C49  | 7947.0(8)  | 5249(5)  | 2996.9(13) | 24.1(6)  |
| C43  | 8301.6(9)  | 8906(5)  | 4777.7(15) | 32.1(7)  |
| C413 | 7282.6(9)  | 1089(5)  | 3329.0(16) | 33.4(7)  |
| C114 | 5004.9(9)  | -8372(5) | 3972.5(14) | 28.7(7)  |
| C41  | 8443.0(9)  | 9715(5)  | 5991.9(15) | 33.6(7)  |
| C212 | 5179.4(8)  | -178(5)  | 6529.2(15) | 28.5(7)  |
| C21  | 4026.1(9)  | -5061(6) | 4115.7(16) | 42.6(9)  |
| C22  | 4194.6(8)  | -2282(5) | 4950.0(14) | 29.1(7)  |
| C11  | 6617.5(10) | -6235(6) | 2066.1(18) | 46.0(10) |
| C311 | 6611.3(9)  | 10795(5) | 4998.8(15) | 35.4(8)  |
| C16  | 6188.9(9)  | -1357(5) | 3613.4(14) | 30.6(7)  |
| C211 | 4248.1(10) | 1604(5)  | 6985.9(17) | 39.2(8)  |
| C215 | 5019.7(9)  | 3032(5)  | 6904.5(15) | 33.9(8)  |
| C24  | 4251.7(9)  | -3668(5) | 5865.9(15) | 33.8(7)  |
| C23  | 4159.7(9)  | -3911(5) | 5313.8(16) | 36.0(8)  |
| C310 | 7149.9(10) | 8680(6)  | 4761.2(15) | 39.7(8)  |
| C210 | 4687.6(12) | -295(6)  | 7571.3(15) | 44.3(9)  |
| C411 | 7841.3(10) | 6012(6)  | 2417.1(14) | 37.0(8)  |
| C213 | 5436.0(9)  | 868(5)   | 6137.8(17) | 38.3(8)  |
| C110 | 5829.4(10) | -5569(6) | 4920.9(16) | 39.0(8)  |
| C410 | 8288.6(9)  | 3911(5)  | 2948.9(16) | 35.2(8)  |
| C214 | 5277.1(9)  | 4012(5)  | 6500.6(18) | 41.3(9)  |

| Atom | x          | y        | z          | $U_{eq}$ |
|------|------------|----------|------------|----------|
| C111 | 5312.3(11) | -3170(6) | 5088.7(16) | 41.6(9)  |
| C31  | 5875.4(11) | 9895(7)  | 7934.8(17) | 50.9(11) |

**Table 2:** Anisotropic Displacement Parameters ( $\times 10^4$ ) **ld021\_etoH\_lt\_auto**. The anisotropic displacement factor exponent takes the form:  $-2\pi^2[h^2a^{*2} \times U_{11} + \dots + 2hka^* \times b^* \times U_{12}]$

| Atom | $U_{11}$ | $U_{22}$ | $U_{33}$ | $U_{23}$ | $U_{13}$  | $U_{12}$  |
|------|----------|----------|----------|----------|-----------|-----------|
| S41  | 36.2(4)  | 35.6(4)  | 28.5(4)  | 3.3(3)   | -5.2(3)   | -2.8(3)   |
| S31  | 29.5(4)  | 44.3(5)  | 30.1(4)  | -5.0(3)  | 3.2(3)    | 6.6(4)    |
| S11  | 28.0(4)  | 51.4(5)  | 32.2(4)  | 1.1(4)   | 3.6(3)    | 6.0(4)    |
| S21  | 37.3(4)  | 55.1(6)  | 28.6(4)  | 7.8(4)   | -5.0(4)   | -3.4(4)   |
| O11  | 36.0(12) | 24.4(12) | 59.8(17) | -4.9(10) | 6.6(11)   | -14.9(11) |
| O32  | 30.5(11) | 28.0(12) | 56.7(17) | -5.8(9)  | -16.2(11) | -3.5(11)  |
| O41  | 47.1(13) | 22.8(12) | 35.0(14) | -5.0(10) | -1.4(11)  | 6.4(10)   |
| N11  | 19.5(10) | 21.9(11) | 23.3(13) | 0.5(9)   | -1.5(9)   | -3.1(10)  |
| O12  | 29.7(11) | 27.5(11) | 49.2(16) | -4.2(9)  | -10.1(11) | -3.9(10)  |
| O21  | 51.3(14) | 25.4(11) | 34.7(14) | -4.1(11) | -0.6(11)  | 10.4(10)  |
| O31  | 37.9(12) | 23.7(11) | 46.7(15) | -4.8(9)  | 6.6(11)   | -11.6(10) |
| O22  | 30.0(12) | 27.7(13) | 79(2)    | -7.0(10) | 0.7(12)   | -14.7(13) |
| O42  | 27.0(11) | 27.9(12) | 54.1(17) | -6.1(9)  | 3.3(10)   | -6.7(11)  |
| N41  | 25.0(11) | 19.8(12) | 22.2(13) | 0.2(10)  | 0.1(10)   | -0.7(10)  |
| N31  | 19.6(11) | 19.6(12) | 25.9(13) | -2.1(9)  | 0.7(9)    | -1.7(10)  |
| N21  | 27.5(12) | 18.8(12) | 29.0(15) | 2.5(10)  | -3.5(10)  | -1.9(10)  |
| C35  | 20.2(13) | 19.1(14) | 31.9(17) | -1.0(11) | -2.6(12)  | 1.0(12)   |
| C415 | 24.2(13) | 22.6(14) | 24.9(16) | -0.4(12) | -1.5(12)  | 0.1(12)   |
| C34  | 23.2(13) | 23.5(15) | 34.6(18) | -2.3(12) | 1.4(12)   | -0.0(14)  |
| C36  | 26.0(14) | 21.7(15) | 36.2(18) | -5.6(12) | -7.8(13)  | 0.8(13)   |
| C47  | 23.8(13) | 25.7(15) | 28.4(17) | -1.6(12) | -1.1(12)  | 3.5(13)   |
| C18  | 21.8(13) | 21.1(15) | 39(2)    | -0.5(11) | -4.8(13)  | -3.8(13)  |
| C26  | 22.9(13) | 24.8(15) | 28.6(17) | -0.8(12) | 0.8(12)   | 6.6(12)   |
| C27  | 26.0(14) | 35.1(17) | 28.3(17) | 2.2(13)  | 2.7(12)   | 7.1(14)   |
| C28  | 25.1(14) | 22.3(15) | 32.1(18) | 3.6(12)  | 1.1(12)   | 5.5(13)   |
| C112 | 23.1(14) | 21.4(15) | 37.9(19) | 0.8(12)  | -2.5(12)  | -3.3(13)  |
| C29  | 36.3(16) | 24.2(15) | 23.9(16) | 4.7(13)  | 1.9(13)   | 0.0(12)   |
| C13  | 22.3(14) | 32.2(17) | 35.3(19) | 0.2(13)  | 3.3(13)   | -6.1(14)  |
| C17  | 27.5(15) | 35.2(18) | 34.5(19) | -9.5(13) | -5.4(13)  | 7.1(14)   |
| C15  | 21.6(13) | 26.1(15) | 31.6(17) | -1.1(12) | -5.8(12)  | 1.2(13)   |
| C14  | 21.6(13) | 25.4(16) | 37.8(19) | -3.3(12) | 4.2(13)   | -3.5(13)  |
| C45  | 20.5(13) | 22.6(14) | 27.9(17) | 0.8(11)  | -0.7(11)  | 1.7(12)   |
| C44  | 32.5(15) | 24.2(16) | 34.1(18) | -5.0(13) | -1.4(14)  | 4.0(13)   |
| C19  | 26.0(14) | 26.8(15) | 28.8(17) | -2.0(13) | -1.2(12)  | -3.8(13)  |
| C39  | 26.1(14) | 27.8(15) | 23.1(16) | -4.0(12) | -1.3(12)  | -2.3(12)  |
| C48  | 21.9(13) | 19.9(15) | 30.7(17) | 0.2(11)  | 0.3(12)   | 3.0(12)   |
| C33  | 25.5(14) | 22.9(15) | 35.5(19) | -0.5(12) | 2.8(13)   | -1.5(13)  |
| C25  | 22.4(14) | 23.8(15) | 30.5(17) | 1.0(12)  | -0.0(12)  | 4.6(13)   |
| C42  | 20.1(13) | 33.9(16) | 24.8(16) | 2.6(12)  | 1.1(11)   | -2.5(13)  |
| C38  | 18.6(12) | 22.1(15) | 33.5(17) | -1.4(11) | -4.1(12)  | -4.8(13)  |
| C37  | 26.9(14) | 25.3(16) | 36.7(19) | -6.0(12) | -6.6(13)  | 5.3(14)   |
| C12  | 19.9(13) | 36.7(18) | 33.7(18) | 0.8(13)  | -6.2(12)  | 4.9(14)   |
| C314 | 33.7(16) | 31.8(17) | 37(2)    | 0.9(14)  | -12.9(15) | 0.2(14)   |
| C46  | 22.2(13) | 24.5(15) | 28.2(16) | -1.8(11) | -0.9(12)  | -0.1(12)  |
| C115 | 28.9(14) | 18.4(14) | 29.8(17) | -1.2(12) | 4.4(12)   | -1.1(12)  |
| C315 | 24.1(14) | 20.9(14) | 38.7(19) | 0.3(11)  | -3.4(13)  | -1.7(13)  |
| C32  | 18.4(13) | 35.4(17) | 28.3(17) | -1.2(12) | -3.6(12)  | 4.6(13)   |
| C414 | 29.1(15) | 26.7(16) | 36.9(19) | -4.3(13) | 4.8(14)   | -5.6(14)  |
| C313 | 30.7(16) | 21.6(15) | 39(2)    | -2.1(12) | -2.8(14)  | -3.8(13)  |
| C312 | 26.4(14) | 21.7(15) | 32.3(17) | -3.8(12) | 4.3(13)   | -1.2(13)  |
| C412 | 30.3(15) | 22.4(15) | 30.5(17) | 1.8(12)  | -4.6(13)  | -5.1(13)  |
| C113 | 34.7(17) | 28.4(17) | 39(2)    | -3.7(13) | -12.4(15) | 0.3(15)   |
| C49  | 26.6(14) | 22.2(15) | 23.5(16) | 1.2(12)  | 3.3(12)   | 1.4(12)   |
| C43  | 36.5(17) | 22.7(16) | 37.2(19) | -4.5(13) | -5.6(14)  | -1.8(14)  |
| C413 | 33.7(16) | 22.2(16) | 44(2)    | -2.9(13) | 1.0(15)   | -1.9(14)  |

| Atom | $U_{11}$ | $U_{22}$ | $U_{33}$ | $U_{23}$  | $U_{13}$  | $U_{12}$  |
|------|----------|----------|----------|-----------|-----------|-----------|
| C114 | 30.8(15) | 17.8(15) | 37.5(19) | -3.1(12)  | 2.1(13)   | -2.0(13)  |
| C41  | 27.8(15) | 39.5(18) | 33.6(18) | 1.6(14)   | -5.0(13)  | -8.1(15)  |
| C212 | 25.4(14) | 24.1(16) | 35.9(18) | 4.8(12)   | -7.2(13)  | -5.0(13)  |
| C21  | 26.8(16) | 64(2)    | 37(2)    | 4.4(16)   | -0.9(14)  | -14.1(18) |
| C22  | 23.8(14) | 36.7(18) | 26.8(17) | 2.5(13)   | -1.2(12)  | 0.6(14)   |
| C11  | 35.4(18) | 51(2)    | 51(2)    | 7.5(17)   | 17.2(17)  | -3.4(19)  |
| C311 | 35.7(16) | 35.7(18) | 34.9(19) | -4.6(14)  | -13.0(14) | 3.2(15)   |
| C16  | 33.1(16) | 23.6(15) | 35.1(18) | -6.5(13)  | -10.1(13) | 0.6(13)   |
| C211 | 39.1(18) | 30.1(17) | 48(2)    | 7.2(15)   | 15.1(16)  | 0.5(16)   |
| C215 | 35.8(16) | 21.7(15) | 44(2)    | 3.7(13)   | -10.1(15) | -9.3(14)  |
| C24  | 40.1(17) | 24.6(16) | 37(2)    | -3.7(13)  | -0.5(15)  | 4.2(14)   |
| C23  | 39.9(18) | 28.6(17) | 39(2)    | -5.4(14)  | -4.4(15)  | -1.7(15)  |
| C310 | 47.8(19) | 41(2)    | 30.0(19) | -11.1(16) | 11.1(16)  | -10.7(16) |
| C210 | 67(3)    | 41(2)    | 24.8(18) | -3.2(18)  | -2.9(17)  | 3.0(16)   |
| C411 | 47.1(19) | 38.3(18) | 25.7(18) | -6.6(16)  | 1.3(14)   | 4.7(14)   |
| C213 | 31.7(16) | 31.8(18) | 51(2)    | -6.0(14)  | 1.0(15)   | -13.1(16) |
| C110 | 39.3(18) | 41(2)    | 37(2)    | -2.8(16)  | -15.7(15) | 2.0(16)   |
| C410 | 30.6(16) | 26.8(17) | 48(2)    | 2.2(13)   | 11.0(15)  | -3.3(15)  |
| C214 | 36.5(18) | 21.8(16) | 66(3)    | -1.3(14)  | -8.9(17)  | -5.1(16)  |
| C111 | 48(2)    | 42(2)    | 34(2)    | -9.1(17)  | 9.6(16)   | -15.0(17) |
| C31  | 47(2)    | 67(3)    | 39(2)    | -18(2)    | 15.5(17)  | -11(2)    |

**Table 3:** Bond Lengths in Å for `ld021_etoh_lt_auto`.

| Atom | Atom | Length/Å | Atom | Atom | Length/Å |
|------|------|----------|------|------|----------|
| S41  | C42  | 1.759(3) | C47  | C42  | 1.413(4) |
| S41  | C41  | 1.798(3) | C47  | C46  | 1.385(4) |
| S31  | C32  | 1.757(3) | C18  | C15  | 1.498(5) |
| S31  | C31  | 1.792(4) | C18  | C19  | 1.549(4) |
| S11  | C12  | 1.760(3) | C26  | C27  | 1.383(5) |
| S11  | C11  | 1.779(4) | C26  | C25  | 1.407(4) |
| S21  | C21  | 1.788(4) | C27  | C22  | 1.398(5) |
| S21  | C22  | 1.760(3) | C28  | C29  | 1.535(4) |
| O11  | C18  | 1.218(4) | C28  | C25  | 1.506(5) |
| O32  | C314 | 1.421(4) | C112 | C113 | 1.507(4) |
| O32  | C313 | 1.430(4) | C29  | C211 | 1.543(4) |
| O41  | C48  | 1.225(4) | C29  | C210 | 1.530(5) |
| N11  | C112 | 1.460(4) | C13  | C14  | 1.385(5) |
| N11  | C19  | 1.473(4) | C13  | C12  | 1.387(5) |
| N11  | C115 | 1.468(4) | C17  | C12  | 1.401(5) |
| O12  | C113 | 1.434(4) | C17  | C16  | 1.378(5) |
| O12  | C114 | 1.421(4) | C15  | C14  | 1.403(4) |
| O21  | C28  | 1.210(4) | C15  | C16  | 1.400(4) |
| O31  | C38  | 1.220(4) | C45  | C44  | 1.401(4) |
| O22  | C213 | 1.422(4) | C45  | C48  | 1.501(4) |
| O22  | C214 | 1.422(4) | C45  | C46  | 1.407(4) |
| O42  | C414 | 1.433(4) | C44  | C43  | 1.377(5) |
| O42  | C413 | 1.424(4) | C19  | C110 | 1.538(5) |
| N41  | C415 | 1.462(4) | C19  | C111 | 1.538(5) |
| N41  | C412 | 1.461(4) | C39  | C38  | 1.532(4) |
| N41  | C49  | 1.474(4) | C39  | C311 | 1.540(4) |
| N31  | C39  | 1.483(4) | C39  | C310 | 1.532(5) |
| N31  | C315 | 1.457(4) | C48  | C49  | 1.542(4) |
| N31  | C312 | 1.466(4) | C33  | C32  | 1.401(4) |
| N21  | C29  | 1.473(4) | C25  | C24  | 1.394(5) |
| N21  | C212 | 1.466(4) | C42  | C43  | 1.384(5) |
| N21  | C215 | 1.465(4) | C37  | C32  | 1.395(5) |
| C35  | C34  | 1.396(4) | C314 | C315 | 1.513(5) |
| C35  | C36  | 1.413(4) | C115 | C114 | 1.521(4) |
| C35  | C38  | 1.497(5) | C313 | C312 | 1.519(5) |
| C415 | C414 | 1.513(4) | C412 | C413 | 1.515(5) |
| C34  | C33  | 1.379(5) | C49  | C411 | 1.539(5) |
| C36  | C37  | 1.372(5) | C49  | C410 | 1.545(4) |

| Atom | Atom | Length/Å |
|------|------|----------|
| C212 | C213 | 1.507(5) |
| C22  | C23  | 1.392(5) |

| Atom | Atom | Length/Å |
|------|------|----------|
| C215 | C214 | 1.509(5) |
| C24  | C23  | 1.387(5) |

**Table 4:** Bond Angles in ° for **Id021\_etoH\_lt\_auto**.

| Atom | Atom | Atom | Angle/°    |
|------|------|------|------------|
| C41  | S41  | C42  | 103.39(17) |
| C31  | S31  | C32  | 103.77(17) |
| C11  | S11  | C12  | 102.95(17) |
| C22  | S21  | C21  | 103.81(18) |
| C313 | O32  | C314 | 111.0(2)   |
| C19  | N11  | C112 | 114.3(2)   |
| C115 | N11  | C112 | 108.5(2)   |
| C115 | N11  | C19  | 115.2(3)   |
| C114 | O12  | C113 | 110.4(2)   |
| C214 | O22  | C213 | 109.5(3)   |
| C413 | O42  | C414 | 109.9(2)   |
| C412 | N41  | C415 | 107.9(2)   |
| C49  | N41  | C415 | 114.4(2)   |
| C49  | N41  | C412 | 113.7(2)   |
| C315 | N31  | C39  | 114.5(2)   |
| C312 | N31  | C39  | 114.9(2)   |
| C312 | N31  | C315 | 108.9(2)   |
| C212 | N21  | C29  | 114.3(2)   |
| C215 | N21  | C29  | 114.2(3)   |
| C215 | N21  | C212 | 108.1(2)   |
| C36  | C35  | C34  | 117.7(3)   |
| C38  | C35  | C34  | 125.8(3)   |
| C38  | C35  | C36  | 116.5(3)   |
| C414 | C415 | N41  | 109.7(2)   |
| C33  | C34  | C35  | 121.4(3)   |
| C37  | C36  | C35  | 121.2(3)   |
| C46  | C47  | C42  | 120.5(3)   |
| C15  | C18  | O11  | 118.9(3)   |
| C19  | C18  | O11  | 118.7(3)   |
| C19  | C18  | C15  | 122.3(3)   |
| C25  | C26  | C27  | 120.2(3)   |
| C22  | C27  | C26  | 121.1(3)   |
| C29  | C28  | O21  | 119.0(3)   |
| C25  | C28  | O21  | 118.3(3)   |
| C25  | C28  | C29  | 122.6(3)   |
| C113 | C112 | N11  | 109.5(3)   |
| C28  | C29  | N21  | 110.2(2)   |
| C211 | C29  | N21  | 110.1(3)   |
| C211 | C29  | C28  | 104.8(3)   |
| C210 | C29  | N21  | 113.7(3)   |
| C210 | C29  | C28  | 109.2(3)   |
| C210 | C29  | C211 | 108.4(3)   |
| C12  | C13  | C14  | 120.9(3)   |
| C16  | C17  | C12  | 120.6(3)   |
| C14  | C15  | C18  | 124.8(3)   |
| C16  | C15  | C18  | 117.3(3)   |
| C16  | C15  | C14  | 117.8(3)   |
| C15  | C14  | C13  | 120.8(3)   |
| C48  | C45  | C44  | 116.7(3)   |
| C46  | C45  | C44  | 117.7(3)   |
| C46  | C45  | C48  | 125.6(3)   |
| C43  | C44  | C45  | 122.1(3)   |
| C18  | C19  | N11  | 108.9(3)   |
| C110 | C19  | N11  | 110.1(3)   |
| C110 | C19  | C18  | 106.3(2)   |
| C111 | C19  | N11  | 113.3(3)   |

| Atom | Atom | Atom | Angle/°  |
|------|------|------|----------|
| C111 | C19  | C18  | 108.4(3) |
| C111 | C19  | C110 | 109.6(3) |
| C38  | C39  | N31  | 109.6(2) |
| C311 | C39  | N31  | 110.0(3) |
| C311 | C39  | C38  | 105.9(2) |
| C310 | C39  | N31  | 112.9(2) |
| C310 | C39  | C38  | 109.3(3) |
| C310 | C39  | C311 | 108.8(3) |
| C45  | C48  | O41  | 118.1(3) |
| C49  | C48  | O41  | 118.3(3) |
| C49  | C48  | C45  | 123.5(3) |
| C32  | C33  | C34  | 120.3(3) |
| C28  | C25  | C26  | 125.4(3) |
| C24  | C25  | C26  | 118.0(3) |
| C24  | C25  | C28  | 116.6(3) |
| C47  | C42  | S41  | 116.4(2) |
| C43  | C42  | S41  | 124.6(3) |
| C43  | C42  | C47  | 119.0(3) |
| C35  | C38  | O31  | 118.8(3) |
| C39  | C38  | O31  | 118.2(3) |
| C39  | C38  | C35  | 123.0(3) |
| C32  | C37  | C36  | 120.5(3) |
| C13  | C12  | S11  | 124.4(3) |
| C17  | C12  | S11  | 116.9(2) |
| C17  | C12  | C13  | 118.7(3) |
| C315 | C314 | O32  | 111.5(3) |
| C45  | C46  | C47  | 120.5(3) |
| C114 | C115 | N11  | 107.8(3) |
| C314 | C315 | N31  | 108.8(2) |
| C33  | C32  | S31  | 125.0(3) |
| C37  | C32  | S31  | 116.0(2) |
| C37  | C32  | C33  | 119.0(3) |
| C415 | C414 | O42  | 111.2(3) |
| C312 | C313 | O32  | 111.4(3) |
| C313 | C312 | N31  | 108.4(3) |
| C413 | C412 | N41  | 108.9(3) |
| C112 | C113 | O12  | 111.5(3) |
| C48  | C49  | N41  | 110.8(2) |
| C411 | C49  | N41  | 113.8(3) |
| C411 | C49  | C48  | 108.3(3) |
| C410 | C49  | N41  | 109.4(2) |
| C410 | C49  | C48  | 105.1(2) |
| C410 | C49  | C411 | 109.0(3) |
| C42  | C43  | C44  | 120.2(3) |
| C412 | C413 | O42  | 111.3(3) |
| C115 | C114 | O12  | 111.9(2) |
| C213 | C212 | N21  | 109.7(3) |
| C27  | C22  | S21  | 117.3(3) |
| C23  | C22  | S21  | 123.6(3) |
| C23  | C22  | C27  | 119.0(3) |
| C15  | C16  | C17  | 121.2(3) |
| C214 | C215 | N21  | 108.6(3) |
| C23  | C24  | C25  | 121.9(3) |
| C24  | C23  | C22  | 119.8(3) |
| C212 | C213 | O22  | 111.3(3) |
| C215 | C214 | O22  | 111.8(3) |

**Table 5:** Torsion Angles in ° for **ld021\_etoh\_lt\_auto**.

| Atom | Atom | Atom | Atom | Angle/°   |
|------|------|------|------|-----------|
| S41  | C42  | C47  | C46  | -178.4(2) |
| S41  | C42  | C43  | C44  | 178.3(3)  |
| S31  | C32  | C33  | C34  | -178.0(3) |
| S31  | C32  | C37  | C36  | 177.5(2)  |
| S11  | C12  | C13  | C14  | -178.2(3) |
| S11  | C12  | C17  | C16  | 178.1(2)  |
| S21  | C22  | C27  | C26  | 178.7(2)  |
| S21  | C22  | C23  | C24  | -178.3(3) |
| O11  | C18  | C15  | C14  | -178.0(3) |
| O11  | C18  | C15  | C16  | 1.5(4)    |
| O11  | C18  | C19  | N11  | 134.6(3)  |
| O11  | C18  | C19  | C110 | -106.8(3) |
| O11  | C18  | C19  | C111 | 10.9(3)   |
| O32  | C314 | C315 | N31  | -58.8(3)  |
| O32  | C313 | C312 | N31  | 58.5(3)   |
| O41  | C48  | C45  | C44  | 3.5(3)    |
| O41  | C48  | C45  | C46  | -174.6(3) |
| O41  | C48  | C49  | N41  | -142.1(3) |
| O41  | C48  | C49  | C411 | -16.6(3)  |
| O41  | C48  | C49  | C410 | 99.8(3)   |
| N11  | C112 | C113 | O12  | 58.2(3)   |
| N11  | C19  | C18  | C15  | -49.0(3)  |
| N11  | C115 | C114 | O12  | -59.8(3)  |
| O21  | C28  | C29  | N21  | 138.4(3)  |
| O21  | C28  | C29  | C211 | -103.2(3) |
| O21  | C28  | C29  | C210 | 12.8(3)   |
| O21  | C28  | C25  | C26  | 175.8(3)  |
| O21  | C28  | C25  | C24  | -2.2(3)   |
| O31  | C38  | C35  | C34  | -175.3(3) |
| O31  | C38  | C35  | C36  | 4.4(3)    |
| O31  | C38  | C39  | N31  | -139.1(3) |
| O31  | C38  | C39  | C311 | 102.2(3)  |
| O31  | C38  | C39  | C310 | -14.8(3)  |
| O22  | C213 | C212 | N21  | 59.2(3)   |
| O22  | C214 | C215 | N21  | -60.3(3)  |
| O42  | C414 | C415 | N41  | -58.9(3)  |
| O42  | C413 | C412 | N41  | 60.5(3)   |
| N41  | C49  | C48  | C45  | 42.4(3)   |
| N31  | C39  | C38  | C35  | 44.3(3)   |
| N21  | C29  | C28  | C25  | -46.1(3)  |
| C35  | C34  | C33  | C32  | 0.6(4)    |
| C35  | C36  | C37  | C32  | 0.7(4)    |
| C35  | C38  | C39  | C311 | -74.4(3)  |
| C35  | C38  | C39  | C310 | 168.5(3)  |
| C34  | C33  | C32  | C37  | 0.2(4)    |
| C36  | C37  | C32  | C33  | -0.8(4)   |
| C47  | C42  | C43  | C44  | 0.1(3)    |
| C47  | C46  | C45  | C44  | -0.7(3)   |
| C47  | C46  | C45  | C48  | 177.4(3)  |
| C18  | C15  | C14  | C13  | 178.5(3)  |
| C18  | C15  | C16  | C17  | -178.7(3) |
| C26  | C27  | C22  | C23  | 0.2(4)    |
| C26  | C25  | C28  | C29  | 0.3(4)    |
| C26  | C25  | C24  | C23  | -0.5(3)   |
| C27  | C22  | C23  | C24  | 0.1(4)    |
| C28  | C25  | C24  | C23  | 177.6(3)  |
| C13  | C14  | C15  | C16  | -1.1(4)   |
| C13  | C12  | C17  | C16  | -0.8(3)   |
| C17  | C16  | C15  | C14  | 0.9(4)    |

| Atom | Atom | Atom | Atom | Angle/°  |
|------|------|------|------|----------|
| C45  | C44  | C43  | C42  | -0.5(4)  |
| C45  | C48  | C49  | C411 | 168.0(3) |
| C45  | C48  | C49  | C410 | -75.7(3) |
| C25  | C24  | C23  | C22  | 0.1(4)   |

**Table 6:** Hydrogen Fractional Atomic Coordinates ( $\times 10^4$ ) and Equivalent Isotropic Displacement Parameters ( $\text{\AA}^2 \times 10^3$ ) for **ld021\_etoht\_automato**.  $U_{eq}$  is defined as 1/3 of the trace of the orthogonalised  $U_{ij}$ .

| Atom | x          | y         | z          | $U_{eq}$ |
|------|------------|-----------|------------|----------|
| H41a | 7418.6(8)  | 6527(4)   | 3634.1(13) | 28.7(7)  |
| H41b | 7216.1(8)  | 5754(4)   | 3086.6(13) | 28.7(7)  |
| H34  | 6580.9(8)  | 10728(5)  | 6235.8(14) | 32.5(8)  |
| H36  | 6381.3(8)  | 4714(5)   | 6020.9(14) | 33.6(8)  |
| H47  | 8117.8(8)  | 4175(5)   | 5077.9(13) | 31.1(8)  |
| H26  | 4504.9(8)  | 1082(5)   | 5829.1(13) | 30.5(8)  |
| H27  | 4346.9(8)  | 694(5)    | 4904.9(14) | 35.8(8)  |
| H11a | 5198.4(8)  | -2917(5)  | 3774.3(14) | 33.0(8)  |
| H11b | 4876.9(8)  | -3715(5)  | 4162.7(14) | 33.0(8)  |
| H13  | 6210.7(8)  | -6382(5)  | 2911.0(14) | 35.9(8)  |
| H17  | 6532.6(8)  | -510(5)   | 3004.0(14) | 38.9(9)  |
| H14  | 5864.5(8)  | -5956(5)  | 3703.6(14) | 33.9(8)  |
| H44  | 8257.7(8)  | 10005(5)  | 4002.1(15) | 36.3(8)  |
| H33  | 6249.5(8)  | 10681(5)  | 7049.0(14) | 33.5(8)  |
| H37  | 6060.4(8)  | 4650(5)   | 6840.1(14) | 35.6(8)  |
| H31a | 7299.5(9)  | 10816(5)  | 6715.4(15) | 41.0(9)  |
| H31b | 7679.9(9)  | 9664(5)   | 6680.9(15) | 41.0(9)  |
| H46  | 7996.8(8)  | 4034(5)   | 4134.0(13) | 30.0(7)  |
| H11c | 4944.6(8)  | -6710(4)  | 4698.9(14) | 30.8(8)  |
| H11d | 5315.0(8)  | -7980(4)  | 4682.3(14) | 30.8(8)  |
| H31c | 7577.3(8)  | 8915(5)   | 5736.8(14) | 33.5(8)  |
| H31d | 7251.9(8)  | 8033(5)   | 6104.2(14) | 33.5(8)  |
| H41c | 6874.3(8)  | 4867(5)   | 3875.9(15) | 37.1(8)  |
| H41d | 7212.2(8)  | 3652(5)   | 4125.0(15) | 37.1(8)  |
| H31e | 7227.5(9)  | 13975(5)  | 6201.5(15) | 36.6(8)  |
| H31f | 7564.1(9)  | 14786(5)  | 5848.3(15) | 36.6(8)  |
| H31g | 7132.6(8)  | 13229(5)  | 5256.0(14) | 32.2(8)  |
| H31h | 7506.9(8)  | 11999(5)  | 5232.2(14) | 32.2(8)  |
| H41e | 7406.2(9)  | 2769(4)   | 2628.6(14) | 33.3(8)  |
| H41f | 7743.7(9)  | 1481(4)   | 2854.5(14) | 33.3(8)  |
| H11e | 4755.4(9)  | -4583(5)  | 3236.6(15) | 41.0(9)  |
| H11f | 5133.0(9)  | -5761(5)  | 3195.0(15) | 41.0(9)  |
| H43  | 8379.0(9)  | 10150(5)  | 4938.6(15) | 38.6(9)  |
| H41g | 7414.1(9)  | 632(5)    | 3663.2(16) | 40.1(9)  |
| H41h | 7202.2(9)  | -134(5)   | 3124.5(16) | 40.1(9)  |
| H11g | 5210.7(9)  | -8826(5)  | 3741.9(14) | 34.4(8)  |
| H11h | 4879.7(9)  | -9598(5)  | 4112.8(14) | 34.4(8)  |
| H41i | 8650(4)    | 10207(13) | 5777(8)    | 50.4(11) |
| H41j | 8230(3)    | 10540(8)  | 5906(10)   | 50.4(11) |
| H41k | 8498(7)    | 9821(8)   | 6388(3)    | 50.4(11) |
| H21a | 5089.0(8)  | -1449(5)  | 6358.9(15) | 34.2(8)  |
| H21b | 5307.4(8)  | -536(5)   | 6875.3(15) | 34.2(8)  |
| H21c | 3975(8)    | -5278(7)  | 3722(3)    | 63.9(14) |
| H21d | 4250(3)    | -5765(9)  | 4215(11)   | 63.9(14) |
| H21e | 3826(5)    | -5599(11) | 4338(9)    | 63.9(14) |
| H11i | 6765(7)    | -6499(13) | 1738(8)    | 69.0(14) |
| H11j | 6697(8)    | -7116(6)  | 2369(5)    | 69.0(14) |
| H11k | 6363.6(17) | -6516(14) | 1981(13)   | 69.0(14) |
| H31i | 6436(4)    | 9957(7)   | 4799(10)   | 53.1(12) |
| H31j | 6490(5)    | 11510(30) | 5301.3(18) | 53.1(12) |
| H31k | 6718.2(13) | 11790(30) | 4745(9)    | 53.1(12) |
| H16  | 6186.1(9)  | -70(5)    | 3792.3(14) | 36.7(8)  |
| H21f | 4185(5)    | 2000(30)  | 6607.5(19) | 58.8(13) |

| Atom | x          | y         | z          | $U_{eq}$ |
|------|------------|-----------|------------|----------|
| H21g | 4045(2)    | 873(16)   | 7152(11)   | 58.8(13) |
| H21h | 4301(3)    | 2820(20)  | 7205(10)   | 58.8(13) |
| H21i | 5147.8(9)  | 2699(5)   | 7252.4(15) | 40.7(9)  |
| H21j | 4820.0(9)  | 3980(5)   | 6992.2(15) | 40.7(9)  |
| H24  | 4227.1(9)  | -4790(5)  | 6110.5(15) | 40.6(9)  |
| H23  | 4073.2(9)  | -5185(5)  | 5184.2(16) | 43.2(9)  |
| H31l | 7309(5)    | 7590(30)  | 4894(3)    | 59.6(13) |
| H31m | 6997.3(11) | 8160(40)  | 4462(5)    | 59.6(13) |
| H31n | 7297(5)    | 9811(11)  | 4623(8)    | 59.6(13) |
| H21k | 4487(3)    | -1000(40) | 7753(4)    | 66.4(14) |
| H21l | 4901(5)    | -1180(40) | 7571.8(16) | 66.4(14) |
| H21m | 4742(8)    | 965(7)    | 7771(4)    | 66.4(14) |
| H41l | 8044(3)    | 6770(30)  | 2256(5)    | 55.6(12) |
| H41m | 7782(7)    | 4848(6)   | 2181(3)    | 55.6(12) |
| H41n | 7631(4)    | 6910(30)  | 2445.9(19) | 55.6(12) |
| H21n | 5640.4(9)  | -52(5)    | 6054.0(17) | 46.0(10) |
| H21o | 5309.4(9)  | 1170(5)   | 5786.6(17) | 46.0(10) |
| H11l | 5713.7(19) | -6360(30) | 5215(8)    | 58.4(12) |
| H11m | 5933(5)    | -6490(30) | 4645(3)    | 58.4(12) |
| H11n | 6022(4)    | -4730(6)  | 5080(10)   | 58.4(12) |
| H41o | 8244(3)    | 2810(20)  | 2685(8)    | 52.8(12) |
| H41p | 8491.3(19) | 4742(10)  | 2819(11)   | 52.8(12) |
| H41q | 8347(4)    | 3340(30)  | 3312(3)    | 52.8(12) |
| H21p | 5371.8(9)  | 5287(5)   | 6662.5(18) | 49.5(11) |
| H21q | 5145.1(9)  | 4363(5)   | 6157.6(18) | 49.5(11) |
| H11o | 5157(6)    | -2140(30) | 4916(2)    | 62.4(13) |
| H11p | 5162(6)    | -4199(9)  | 5271(8)    | 62.4(13) |
| H11q | 5468.4(11) | -2520(40) | 5364(7)    | 62.4(13) |
| H31o | 5799(8)    | 10922(8)  | 7665(5)    | 76.3(16) |
| H31p | 5715(7)    | 9943(17)  | 8257(8)    | 76.3(16) |
| H31q | 6124(2)    | 10170(20) | 8052(12)   | 76.3(16) |

### Citations

O.V. Dolomanov and L.J. Bourhis and R.J. Gildea and J.A.K. Howard and H. Puschmann, Olex2: A complete structure solution, refinement and analysis program, *J. Appl. Cryst.*, (2009), **42**, 339-341.

Sheldrick, G.M., ShelXT-Integrated space-group and crystal-structure determination, *Acta Cryst.*, (2015), **A71**, 3-8.

## Annexure 6

### checkCIF/PLATON report

Structure factors have been supplied for datablock(s) LD021\_EtOH\_LT\_auto

THIS REPORT IS FOR GUIDANCE ONLY. IF USED AS PART OF A REVIEW PROCEDURE FOR PUBLICATION, IT SHOULD NOT REPLACE THE EXPERTISE OF AN EXPERIENCED CRYSTALLOGRAPHIC REFEREE.

No syntax errors found.      CIF dictionary      Interpreting this report

### Datablock: LD021\_EtOH\_LT\_auto

---

Bond precision:    C-C = 0.0048 Å                      Wavelength=1.54184

Cell:                      a=37.0231(3)              b=6.5766(1)              c=24.1761(3)  
                                alpha=90                      beta=90                      gamma=90

Temperature:              150 K

|                        | Calculated     | Reported       |
|------------------------|----------------|----------------|
| Volume                 | 5886.55(13)    | 5886.54(12)    |
| Space group            | P c a 21       | P c a 21       |
| Hall group             | P 2c -2ac      | P 2c -2ac      |
| Moiety formula         | C15 H21 N O2 S | C15 H21 N O2 S |
| Sum formula            | C15 H21 N O2 S | C15 H21 N O2 S |
| Mr                     | 279.39         | 279.39         |
| Dx, g cm <sup>-3</sup> | 1.261          | 1.261          |
| Z                      | 16             | 16             |
| Mu (mm <sup>-1</sup> ) | 1.933          | 1.934          |
| F000                   | 2400.0         | 2411.9         |
| F000'                  | 2411.46        |                |
| h, k, lmax             |                | 45, 8, 29      |
| Nref                   |                | 9767           |
| Tmin, Tmax             | 0.683, 0.858   | 0.839, 1.000   |
| Tmin'                  | 0.618          |                |

Correction method= # Reported T Limits: Tmin=0.839 Tmax=1.000  
AbsCorr = MULTI-SCAN

Data completeness=                      Theta(max)= 72.110

R(reflections)= 0.0465( 9312)

wR2(reflections)=  
0.1258( 9767)

S = 1.047

Npar= 697



---

The following ALERTS were generated. Each ALERT has the format  
**test-name\_ALERT\_alert-type\_alert-level.**

Click on the hyperlinks for more details of the test.

---

 **Alert level A**

PLAT029\_ALERT\_3\_A \_diffn\_measured\_fraction\_theta\_full value Low . 0.845 Why?

---

**Author Response: Data was collected on a relatively small and weakly diffracting crysta**

---

 **Alert level C**

PLAT094\_ALERT\_2\_C Ratio of Maximum / Minimum Residual Density .... 2.52 Report  
PLAT340\_ALERT\_3\_C Low Bond Precision on C-C Bonds ..... 0.00481 Ang.  
PLAT790\_ALERT\_4\_C Centre of Gravity not Within Unit Cell: Resd. # 1 Note  
C15 H21 N O2 S  
PLAT911\_ALERT\_3\_C Missing FCF Refl Between Thmin & STh/L= 0.600 5 Report  
0 4 0, 1 4 0, 1 4 1, 0 4 2, 1 7 12,  
PLAT926\_ALERT\_1\_C Reported and Calculated R1 Differ by ..... 0.0011 Check  
PLAT927\_ALERT\_1\_C Reported and Calculated wR2 Differ by ..... 0.0024 Check  
PLAT987\_ALERT\_1\_C The Flack x is >> 0 - Do a BASF/TWIN Refinement Please Check

---

 **Alert level G**

PLAT068\_ALERT\_1\_G Reported F000 Differs from Calcd (or Missing)... Please Check  
PLAT073\_ALERT\_1\_G H-atoms ref, but \_hydrogen\_treatment Reported as constr Check  
PLAT398\_ALERT\_2\_G Deviating C-O-C Angle From 120 for O22 . 109.5 Degree  
PLAT398\_ALERT\_2\_G Deviating C-O-C Angle From 120 for O42 . 109.9 Degree  
PLAT802\_ALERT\_4\_G CIF Input Record(s) with more than 80 Characters 1 Info  
PLAT910\_ALERT\_3\_G Missing # of FCF Reflection(s) Below Theta(Min). 1 Note  
2 0 0,  
PLAT912\_ALERT\_4\_G Missing # of FCF Reflections Above STh/L= 0.600 23 Note  
PLAT915\_ALERT\_3\_G No Flack x Check Done: Low Friedel Pair Coverage 69 %  
PLAT978\_ALERT\_2\_G Number C-C Bonds with Positive Residual Density. 0 Info  
PLAT980\_ALERT\_1\_G No Anomalous Scattering Factors Found in CIF ... Please Check

---

- 1 **ALERT level A** = Most likely a serious problem - resolve or explain  
0 **ALERT level B** = A potentially serious problem, consider carefully  
7 **ALERT level C** = Check. Ensure it is not caused by an omission or oversight  
10 **ALERT level G** = General information/check it is not something unexpected

- 6 ALERT type 1 CIF construction/syntax error, inconsistent or missing data  
4 ALERT type 2 Indicator that the structure model may be wrong or deficient  
5 ALERT type 3 Indicator that the structure quality may be low  
3 ALERT type 4 Improvement, methodology, query or suggestion  
0 ALERT type 5 Informative message, check
- 
-

It is advisable to attempt to resolve as many as possible of the alerts in all categories. Often the minor alerts point to easily fixed oversights, errors and omissions in your CIF or refinement strategy, so attention to these fine details can be worthwhile. In order to resolve some of the more serious problems it may be necessary to carry out additional measurements or structure refinements. However, the purpose of your study may justify the reported deviations and the more serious of these should normally be commented upon in the discussion or experimental section of a paper or in the "special\_details" fields of the CIF. checkCIF was carefully designed to identify outliers and unusual parameters, but every test has its limitations and alerts that are not important in a particular case may appear. Conversely, the absence of alerts does not guarantee there are no aspects of the results needing attention. It is up to the individual to critically assess their own results and, if necessary, seek expert advice.

### **Publication of your CIF in IUCr journals**

A basic structural check has been run on your CIF. These basic checks will be run on all CIFs submitted for publication in IUCr journals (*Acta Crystallographica*, *Journal of Applied Crystallography*, *Journal of Synchrotron Radiation*); however, if you intend to submit to *Acta Crystallographica Section C* or *E* or *IUCrData*, you should make sure that full publication checks are run on the final version of your CIF prior to submission.

### **Publication of your CIF in other journals**

Please refer to the *Notes for Authors* of the relevant journal for any special instructions relating to CIF submission.

