

# **Simulation of protein dynamics for mechanistic insight and drug design**

A thesis submitted to The University of Manchester for the degree of

Master of Philosophy

in the Faculty of Biology, Medicine and Health

**2019**

**YANG, Zhang**

**School of Health Sciences/Division of Pharmacy**

# List of contents

List of figures.....	5
List of tables .....	14
List of abbreviations.....	16
Abstract.....	18
Declaration.....	20
Copyright Statement.....	21
Acknowledgements.....	22
Introduction to protein dynamics in drug design .....	23
In silico design of anti-cancer compounds targeting R-spondin using molecular dynamics and free energy calculations.....	24
<b>1. Background.....</b>	<b>24</b>
1.1 Wnt signalling pathway.....	24
1.2 R-spondin signalling and disease .....	26
1.3 Structure basis of R-spondin signalling .....	29
1.4 Protein dynamics.....	31
1.5 Aim of this work .....	33
<b>2. Computational Theory .....</b>	<b>35</b>
2.1 Protein-ligand binding site prediction .....	35
2.1.1 Shape-based methodology .....	36
2.1.2 Energy-based methodology .....	37
2.2 Molecular docking.....	38
2.2.1 Molecular docking algorithms .....	39
2.2.2 Scoring functions.....	41
2.3 Molecular dynamics .....	43
2.3.1 Molecular mechanics .....	43
2.3.2 Water solvent models .....	45
2.3.3 Periodical boundary conditions .....	46
2.3.4 Temperature and pressure regulation.....	47
2.4 Free energy calculation.....	48
2.4.1 Poisson Boltzmann potential .....	48
2.4.2 Generalized Born model .....	49
2.4.3 MM-GBSA method .....	50
<b>3. High throughput virtual screening.....</b>	<b>51</b>

3.1 Compound database refinement.....	53
3.2 Binding pocket identification .....	55
3.2.1 Closed and Open2 R-spondin binding pockets .....	55
3.2.2 RNF43 binding pockets .....	61
3.2.3 LGR5 binding pockets.....	63
3.3 Molecular docking and protein-ligand interactions.....	66
3.3.1 Methods .....	66
3.3.2 Closed R-spondin interactions .....	67
3.3.3 Open2 R-spondin interactions .....	70
3.3.4 RNF43 interactions.....	72
3.3.5 LGR5 interactions.....	76
<b>4 Molecular dynamics simulation.....</b>	<b>79</b>
4.1 Methods .....	80
4.1.1 Model construction.....	80
4.1.2 Molecular dynamics protocol .....	81
4.1.3 Compounds selection and optimization .....	81
4.1.4 Analysis tools.....	82
4.2 Results: <i>Closed</i> R-spondin ligands .....	83
4.2.1 Compound selection .....	83
4.2.2 Ligand Binding stabilities.....	90
4.2.3 Ligand structure refinement .....	94
4.2.4 Special cases.....	100
4.3 Results: <i>Open2</i> R-spondin ligands .....	105
4.3.1 Compound selection .....	105
4.3.2 Ligand Binding stabilities.....	110
4.3.3 Ligand structure refinement .....	113
4.3.4 Special cases.....	118
4.4 Results: RNF43 ligands .....	123
4.4.1 Compound selection .....	123
4.4.2 Ligand Binding stabilities.....	128
4.4.3 Ligand structure refinement .....	137
4.5 Results: LGR5 ligands .....	141
4.5.1 Compound selection .....	141
4.5.2 Ligand Binding stabilities.....	143
4.5.3 Ligand structure refinement .....	149
4.5.4 Special cases.....	155

<b>5. Free energy calculation .....</b>	157
5.1 Methods .....	157
5.2 Binding free energy and energy decomposition.....	159
5.2.1 Closed R-spondin .....	159
5.2.2 Open2 R-spondin .....	162
5.2.3 RNF43 .....	169
5.2.4 LGR5 .....	173
<b>6. Discussion .....</b>	178
References .....	183

**Word count: 25464**

## List of figures

<b>Fig 1.1</b> Schematic model of Wnt/β-catenin pathway. When the Wnt pathway is inactive, β-catenin is retained in its self-destruction complex. However, when Wnt pathway is on, Wnt proteins bind with a heterodimeric receptor formed by Frizzled and LRP improving the recruitment of Dishevelled (Dvl), Axin and CK1 ternary complex which is key for β-catenin self-destruction complex. Hence, the Wnt pathway makes β-catenin escape from degradation and hence, promote abnormal cell proliferation. The figure is adapted from Clevers et al <sup>5</sup> .....	25
<b>Fig 1.2</b> Simplified schematic representation of R-spondin involved the Wnt pathway. <b>a)</b> RNF43 binds with Frizzled and promotes the clearance of Wnt receptor. β-catenin is confined in its self-degradation complex. <b>b)</b> R-spondin promote the binding between LGR5 and RNF43. Without the inhibition by RNF43, the strength of the Wnt pathway increases which helps β-catenin escape from self-degradation, and hence, promotes cell abnormal proliferation.....	28
<b>Fig 1.3</b> The contact interface between R-spondin and LGR5, contact interaction marked black dotted line, LGR5 coloured cyan and R-spondin 1 coloured yellow ..	29
<b>Fig 1.4</b> The contact interface between R-spondin 1 and RNF43, molecular contacts are marked black dotted line, R-spondin 1 coloured yellow and RNF43 coloured pink.	30
<b>Fig 2.1</b> 2D schematic representation of different types of grid points in DoGSiteScorer. Blue points form the hull that directly contact the solvent (Exposed) or protein surface (Surface). Orange points represent the grid points that are inside the receptor. Image is adapted from Volkamer et al <sup>38</sup> .....	37
<b>Fig 2.2</b> Standard workflow of the FRED molecular docking program <sup>43</sup> .....	41
<b>Fig 3.1.1</b> Schematic representation of the high-throughput virtual screening process.	51
<b>Fig 3.1.2</b> Tranches in the <b>ZINC 15</b> compound database chopped by molecular weight and LogP. The red rectangle marks the <b>Drug-like</b> subset in the <b>ZINC 15</b> database. .	53
<b>Fig 3.2.1</b> Superposition of <b>Closed</b> , <b>Open</b> and <b>Open2</b> R-spondin structures. The hinge area is inset. The yellow structure represents the <b>Open</b> conformation of R-spondin-1 obtained from PDB code 4KNG. The green structure represents <b>Closed</b> R-spondin-1 obtained from PDB code 4BSO. The blue structure represents the <b>Open2</b> conformation obtained from the long-range molecular dynamics simulation of R-spondin 1.....	56
<b>Fig 3.2.2</b> Top three binding pockets on the structure of R-spondin <b>Closed</b> conformation predicted by FTsite. The protein structure was coloured green. The orange coloured pocket ranked the first place and the red one and blue pocket ranked the second and the third place.....	58

<b>Fig 3.2.3</b> Top three binding pockets on the structure of R-spondin <b><i>Open2</i></b> conformation predicted by FTSite. The protein structure was coloured blue. The pink colour pocket ranked the first place and the green one and cyan pocket ranked second and third place.....	58
<b>Fig 3.2.4</b> Two major binding pockets of <b>a) Closed</b> R-spondin and <b>b) Open2</b> R-spondin predicted by DoGSiteScorer.....	59
<b>Fig 3.2.5</b> Top three binding pockets on the structure of R-spondin <b><i>Open</i></b> conformation predicted by FTSite. The protein structure was coloured blue. The pink colour pocket ranked the first place and the green one and purple pocket ranked the second and the third place.....	60
<b>Fig 3.2.6</b> Eight binding pockets of <b><i>Open</i></b> R-spondin predicted by DoGSiteScorer. Binding pockets coloured light blue.....	60
<b>Fig 3.2.7</b> Top three binding pockets on the structure of RNF43 predicted by FTSite. The protein structure coloured magenta. The first-ranked binding site coloured pink and the second-ranked pocket coloured green and the third-ranked pocket coloured blue. ....	61
<b>Fig 3.2.8</b> RNF43 and three binding pockets predicted by DoGSiteScorer. Pocket 3 is coloured brown, Pocket 4 is coloured blue and Pocket 6 is coloured green. ....	63
<b>Fig 3.2.9</b> Top three binding pockets on the structure of LGR5 predicted by FTSite. The protein structure coloured green. The first binding site coloured orange, the second pocket coloured magenta and the third pocket coloured cyan.....	64
<b>Fig 3.2.10</b> LGR5 and all binding pockets predicted by DoGSiteScorer. The binding pocket near the binding interface between LGR5 and R-spondin coloured orange.	64
<b>Fig3.3.1</b> Highest populated protein-ligand interactions for top 1000 compounds docked into <b><i>Closed</i></b> R-spondin conformation (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction).....	68
<b>Fig 3.3.2</b> Complex of <b><i>Closed</i></b> R-spondin and RSP-40, representing a major binding pattern. The R-spondin structure is coloured green, RSP-40 is coloured orange and the molecular surface is coloured by electrostatics. ....	69
<b>Fig 3.3.3</b> Intermolecular clashes between ligand RSP-40 and the R-spondin binding site on RNF43. Molecular clashes between ligand atoms and R-spondin atoms are shown as black dotted lines. ....	69
<b>Fig 3.3.4</b> Highest populated protein-ligand interactions for top 1000 ranked compounds docked into R-spondin <b><i>Open2</i></b> conformation (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction) .....	70
<b>Fig 3.3.5</b> Superposition of <b><i>Open2</i></b> R-spondin to the binding site of <b><i>Open</i></b> R-spondin binding to RNF43. R-spondin were coloured blue, RNF43 were coloured pink and	

RSP-60 was coloured yellow. Molecular clashes were represented by black dotted lines .....	71
<b>Fig 3.3.6</b> Binding interface between R-spondin (yellow) and RNF43 (pink). The interface coloured by electrostatics. ....	72
<b>Fig 3.3.7</b> Highest populated protein-ligand interactions for top 1000 compounds docked into RNF43 protein. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction) .....	73
<b>Fig 3.3.8 a)</b> Binding position of RSP-18 on RNF43 in the presence of R-spondin. RSP-18 coloured light blue, R-spondin coloured yellow and RNF43 coloured pink. <b>b)</b> Intermolecular clashes between ligand RSP-18 and second $\beta$ -hairpin on R-spondin-1 represented by red lines; 239 clashes identified between ligand atoms and R-spondin atoms.....	74
<b>Fig 3.3.9 a)</b> Binding position of RSP-22 on RNF43 with presence of R-spondin. RSP-22 is coloured light blue, R-spondin is coloured yellow and RNF43 is coloured pink. <b>b)</b> Intermolecular clashes between ligand RSP-22 and two $\beta$ -hairpin on R-spondin-1 represented by red line; 219 clashes identified between ligand atoms and R-spondin atoms.....	75
<b>Fig 3.3.10 a)</b> Binding position of RSP-32 on RNF43 with presence of R-spondin. RSP-32 coloured light blue, R-spondin coloured yellow and RNF43 coloured pink. <b>b)</b> Intermolecular clashes between ligand RSP-32 and the first $\beta$ -hairpin on R-spondin-1 represented by red line; 87 clashes identified between ligand atoms and R-spondin atoms.....	75
<b>Fig 3.3.11</b> Surface contacts between R-spondin and LGR5. Atoms which formed intermolecular contacts coloured purple in R-spondin and orange in LGR5. The protein structure of LGR5 coloured cyan and R-spondin coloured yellow. Atomic contacts represented by black dotted lines.....	76
<b>Fig 3.3.12</b> Highest populated protein-ligand interactions for top 1000 compounds docked into LGR5 protein. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction) .....	78
<b>Fig 3.3.13</b> Binding position of RSP-90 on LGR5 with presence of R-spondin. RSP-90 coloured light blue, R-spondin coloured yellow and LGR5 coloured green. Intermolecular clashes between RSP-90 and R-spondin-1 marked by black dotted lines; 174 clashes identified between RSP-90 and R-spondin. Binding interface between RSP-90 and LGR5 coloured by electrostatics. ....	78
<b>Fig 4.2.1</b> Protein-ligand interaction fingerprint of top 1000 compounds targeting <b>Closed</b> R-spondin. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction).....	84
<b>Fig 4.2.2 a)</b> Molecular interface between RSP-05 binding with <b>Closed</b> R-spondin coloured by electrostatics. <b>b)</b> Graphical representation of RSP-05 binding site on <b>Closed</b> R-spondin, proteins structure coloured tan and RSP-05 coloured orange....	84

<b>Fig 4.2.3</b> Schematic 2D representation of binding mode 1: RSP-05 with <b>Closed</b> R-spondin.....	85
<b>Fig 4.2.4</b> Schematic 2D representation of the binding mode 2: RSP-09 with <b>Closed</b> R-spondin.....	86
<b>Fig 4.2.5</b> Core structures of binding mode 1 and 2 of <b>Closed</b> R-spondin compounds. M <sub>1</sub> , M <sub>2</sub> and M <sub>3</sub> were used to represent moieties that linked to core structures. Hydrogen bonds were represented by dotted arrow.....	87
<b>Fig 4.2.6</b> Schematic 2D representation of the binding mode 3: RSP-06 with <b>Closed</b> R-spondin.....	88
<b>Fig 4.2.7</b> The structures of RSP-03, RSP-05 and RSP-06 which are predicted from MD to form a stable protein-ligand complex with <b>Closed</b> R-spondin. ....	91
<b>Fig 4.2.8</b> Comparison between RSP-05 (good binding stability in MD) and RSP-02, RSP-10, RSP-12 and RSP-13 (poor binding stability in MD). The anchoring five-membered rings coloured red. Moiety linkers coloured magenta. Benzene ring branches of the righthand moieties coloured green. Structures of the lefthand moieties coloured blue. ....	94
<b>Fig 4.2.9</b> Compounds selected in the first round of structural optimization targeting <b>Closed</b> R-spondin.....	95
<b>Fig 4.2.10</b> <b>a)</b> Initial and <b>b)</b> final binding pose of RSP-40. Hydrogen bond represented by doted arrow.....	97
<b>Fig 4.2.11</b> Structure of RSP-52, RSP-53, RSP-54 and RSP-55 which performed well in the second round of structural optimization.....	97
<b>Fig 4.2.12</b> Time series of distance between backbone oxygen of Cys94 and its corresponding hydrogen bond partners in RSP-52 (pink) and RSP-53 (cyan). .....	99
<b>Fig 4.2.13</b> <b>a)</b> Initial binding posed and <b>b)</b> final binding pose of RSP-04 binding to <b>Closed</b> R-spondin. RSP-04 colored by orange red, <b>Closed</b> R-spondin coloured tan and surface colored electrostatics.....	100
<b>Fig 4.2.14</b> 2D schematic representation of protein-ligand interactions between RSP-04 and <b>Closed</b> R-spondin in the <b>a)</b> initial binding pose and <b>b)</b> final binding pose. Hydrogen bonds represented by dotted arrow from hydrogen bond donor to acceptor. ....	101
<b>Fig 4.2.15</b> <b>a)</b> The distances between RSP <sup>N3</sup> and Cys53 <sup>O</sup> (turquoise) and between RSP-06 <sup>N4</sup> and Cys94 <sup>O</sup> (coral). <b>b)</b> Binding pose of the 200 ns structure of RSP-06, hydrogen bonds are marked by dotted line. ....	102
<b>Fig 4.2.16</b> <b>a)</b> Four sets of distances between the ring system of RSP-06 and the ring system on His108. <b>b)</b> Final binding pose of RSP-06 binding with <b>Closed</b> R-spondin. hydrogen bonds are represented by black dotted line. Specific atoms that were used for distance measurement were labeled in <b>c)</b> . The distance between RSP-06 ring and His108 ring was measured by the distance between COMs of these two rings.....	103

<b>Fig 4.2.17 a)</b> Initial and <b>b)</b> final binding pattern of RSP-54. Hydrogen bonds represented by dotted arrow from hydrogen bond donor to acceptor.....	104
<b>Fig 4.2.18</b> 3D representation of initial and final binding pose of RSP-54. Protein structure coloured tan, RSP-54 coloured orange and interface colored electrostatics. ....	104
<b>Fig 4.3.1 a)</b> The binding interface between ligand and <b>Open2</b> R-spondin binding cavity. <b>b)</b> The location of Leu64 (orange, Furin-1 domain), Cys94 (navy blue, hinge area) and Lys115 (green, Furin-2 domain) in <b>Open2</b> R-spondin.....	105
<b>Fig 4.3.2</b> Protein-ligand interaction fingerprint of top 1000 compounds targeting <b>Open2</b> R-spondin. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction).....	106
<b>Fig 4.3.3</b> Two 4-point binding models which have been widely observed across the top 1000 <b>Open2</b> R-spondin compounds. Key points are coloured red. Hydrogen bonds represented by the dotted arrow from hydrogen bonds donor to hydrogen bond acceptor. ....	107
<b>Fig 4.3.4</b> Schematic 2D representation of (top) NH-π interactions formed by <b>Open2</b> R-spondin compound aromatic rings and Lys96 backbone nitrogen; and (bottom) hydrogen bonds with sidechain of Glu103. Hydrogen bonds represented by dotted arrow from hydrogen bond donor to acceptor. ....	108
<b>Fig 4.3.5</b> The structures of compounds that binding well with <b>Open2</b> R-spondin conformation.....	111
<b>Fig 4.3.6</b> Structures of five compounds including RSP-70, RSP-71, RSP-72, RSP-73 and RSP-74 which demonstrated good binding stability with <b>Open2</b> R-spondin....	114
<b>Fig 4.3.7</b> Schematic 2D representation of <b>a)</b> initial binding pattern and <b>b)</b> final binding pattern of RSP-73 bound to <b>Open2</b> R-spondin conformation.....	115
<b>Fig 4.3.8</b> Graphical representation of RSP-73 <b>a)</b> initial binding pattern predicted by molecular docking and <b>b)</b> final binding pattern after 100 ns MD simulation of the protein-ligand complex. RSP-73 coloured yellow and protein coloured pink. The protein-ligand interface coloured by electrostatics. ....	116
<b>Fig 4.3.9 a)</b> Time series of distance between COM of Leu64 and Lys115 from MD simulation of RSP-58/ <b>Open2</b> R-spondin complex. <b>b)</b> Time series of distance between COM of RSP-58 and COM of residues form Phe61 to Arg66 from MD simulation of RSP-58/ <b>Open2</b> R-spondin complex. ....	119
<b>Fig 4.3.10 a)</b> Initial pose of RSP-58 binding to <b>Open2</b> R-spondin. <b>b)</b> final pose of RSP-58 in the 100 ns simulation. The R-spondin protein structure coloured by pink and RSP-58 coloured blue. The protein surface within 4.5 Å of compound is shown and coloured by electrostatics.....	120
<b>Fig 4.3.11 a)</b> Time series of distance between Leu64 and Lys115 from MD simulation of RSP-62/ <b>Open2</b> R-spondin complex. <b>b)</b> The distance between COM of	

RSP-62 and COM of Glu103 (pink) and the distance between COM of RSP-62 and COM of Leu64 as a function of time (cyan). ....	121
<b>Fig 4.3.12</b> <b>a)</b> RSP-62 initial binding pose and <b>b)</b> final binding pose in the MD simulation. The R-spondin protein was colour by pink and RSP-62 was yellow. Protein-ligand interface was coloured by electrostatics. ....	121
<b>Fig 4.3.13</b> Graphical representation of <b>a)</b> initial and <b>b)</b> final binding pose of RSP-76. RSP-76 coloured yellow and <b>Open2</b> R-spondin coloured pink. Arg66, Cys94 and Glu103 coloured lime gree, navy blue and purple respectively. ....	122
<b>Fig 4.3.14</b> Distances between COM of RSP-76 and COM of Arg66 (pink), Glu103 (cyan) and Cys94 (green) respectively as a function of time.....	122
<b>Fig 4.4.1</b> <b>a)</b> Two major binding spaces on RNF43 predicted by molecular docking - two compounds are used to represent the location of these spaces. The Zinc ID of the blue compound is ZINC000843195955. The violet compound is RSP-20. The molecular interface between the two compounds and RNF43 was coloured by electrostatics. <b>b)</b> Highly populated hydrogen bond interaction partners including Gln84, His86, ASP97, Lys108, Glu110 and ASP140 were coloured red in the protein structure of RNF43.....	124
<b>Fig 4.4.2</b> Protein-ligand interaction fingerprint of top 1000 compounds targeting RNF43. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction).....	124
<b>Fig 4.4.3</b> Two compounds representing binding mode-1 of RNF43.....	125
<b>Fig 4.4.4</b> Typical binding mode-1 represented by RSP-20. Hydrogen bonds represented by dotted line from hydrogen bond donor to acceptor. ....	125
<b>Fig 4.4.5</b> Binding pose of RSP-22 representing the binding mode-2 of RNF43 ligand. ....	126
<b>Fig 4.4.6</b> 2D schematic representation of binding mode-2 compounds represented by RSP-23. Hydrogen bonds represented by dotted line from hydrogen bond donor to acceptor. ....	126
<b>Fig 4.4.7</b> 2D structures of RSP-18, RSP-20, RSP-22, RSP-27 and RSP-29 which could bind stably with RNF43. RSP-18, RSP-20, RSP-27 and RSP-29 belongs to binding mode-1 and the part of their structures buried in the upper cavity of RNF43 are coloured red.....	131
<b>Fig 4.4.8</b> Structure of failed RNF43 binding mode-1 compounds. The structure that near outside of the upper cavity was coloured by red and structure that projected into the upper cavity were coloured by green. ....	133
<b>Fig 4.4.9</b> <b>a)</b> Asp97 form three hydrogen bonds with RSP-18. <b>b)</b> Asp97 forms three hydrogen bonds with RSP-20. ....	133
<b>Fig 4.4.10</b> Upper cavity hydrophobic residues and <b>a)</b> RSP-16 and <b>b)</b> RSP-27. ....	134

<b>Fig 4.4.11 a)</b> Time series of two pairs of distances between RSP-26 and RNF43. <b>b)</b> The first distance is the distance between RSP-26 chloride atom and Ala129 CB atom in upper cavity and the second distances is between RSP-26 C5 atom and Ala182 C atom in lower cavity. These two distances were represented as dotted lines.....	136
<b>Fig 4.4.12 a)</b> Binding pattern between RSP-22 and RNF43. <b>b)</b> Top view of the initial binding pose of RSP-22 binding pose. <b>c)</b> Top view of the 100ns' simulation of RSP-22 binding pose. The new cavity formed during the MD simulation marked by red circle.....	136
<b>Fig 4.4.13</b> 2D structure of RSP-43, RSP-45, RSP-46, RSP-47 and RSP-49. Structures that were buried by upper cavities were coloured by red.....	138
<b>Fig 4.4.14</b> Binding pattern of RSP-46 and RSP-47 binding to RNF43. Core structures were coloured red which forms a hydrogen bond interaction between Asp97 and Gln84.....	139
<b>Fig 4.4.15</b> 3D binding pattern of RSP-49 with the new cavity zoomed up.....	140
<b>Fig 4.5.1</b> Protein-ligand interaction fingerprint of top 1000 compounds targeting LGR5. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction).....	141
<b>Fig 4.5.2</b> Binding patterns of RSP-83 (yellow) and RSP-84 (light blue) binding with LGR5. Two pairs of major hydrogen bonds were zoomed up respectively. Hydrogen bond with Thr192 and Glu261 were represented by black line.....	142
<b>Fig 4.5.3</b> Ligand positional shift of RSP-78, RSP-79, RSP-80 and RSP-81 from their binding site.....	145
<b>Fig 4.5.4 a)</b> Time series of distances between RSP-80 and two major cavities. <b>b)</b> Superposition of initial binding pose (yellow), 60 ns binding pose (blue), 100 ns binding pose (orange).....	145
<b>Fig 4.5.5</b> Ligand positional shift of RSP-82, RSP-83, RSP-84 and RSP-85 from their binding site.....	146
<b>Fig 4.5.6 a)</b> Binding pattern of RSP-83 binding to LGR5 at 80 ns of simulation; intermolecular surface colored by electrostatics. <b>b)</b> Schematic molecular interactions between LGR5 and RSP-83.....	148
<b>Fig 4.5.7</b> Structure of RSP-83 and four failed compounds: RSP-78, RSP-79, RSP-81 and RSP-84.....	148
<b>Fig 4.5.8</b> Time series of three sets of distances between two major structures and their corresponding binding pockets. The distance between nitrogen from benzonitrile group and Leu167 $\alpha$ -carbon was to measure the distance to the Thr192 cavity. The distance between ligand chloride atom and Asp240 CG atom on sidechain was to measure the distance between the benzyl chloride moiety and Asp240 cavity and the distance between the chloride atom was to measure the distance between benzyl chloride moiety and His166 cavity.....	149

<b>Fig 4.5.9</b> Ligand positional shift of RSP-90, RSP-91 and RSP-92 from their binding site.	.....	151
<b>Fig 4.5.10</b> The distances between RSP-90, RSP-91 and RSP-92 atoms and their corresponding binding cavities. The CA atom in Leu167 was to measure the distance between cyano group/halogen group to Thr192 cavity. The CG atom in Asp240 was to measure the distance between the opposite moiety and Glu261 cavity distance. Specifically, RSP-92 binds with the cavity near His166, it was also measured by the distance between O2 atom of RSP-92 and NE2 atom on His166. ....	.....	152
<b>Fig 4.5.11</b> 2D structures of RSP-90, RSP-91 and RSP-92. The moieties that binds with Thr192 cavity coloured red. ....	.....	153
<b>Fig 4.5.12 a)</b> Time series of distance from RSP-80 Thr192-cavity-binding moiety to its pocket. Distance between chloride atom and the $\alpha$ -carbon of Leu167. Initial (b) and final (c) binding pattern of RSP-88. R-spondin coloured yellow, LGR5 coloured green and RSP-88 coloured blue.....	.....	154
<b>Fig 4.5.13</b> Superposition of RSP-82, RSP-86 and RSP-89 on the ligand binding site of LGR5. Structures that formed hydrogen bonds with Glu261 are inset; hydrogen bonds represented by black dotted lines. The orange ligand is RSP-82, the light blue structure is RSP-86 and the purple ligand is RSP-89. ....	.....	156
<b>Fig 4.5.14</b> 2D structures of RSP-82, RSP-86 and RSP-89. Key structures that formed hydrogen bonds with Glu261 coloured red. ....	.....	156
<b>Fig 5.2.1</b> Free energy contribution per residues in <i>Closed</i> R-spondin ligand MD simulations. Major residues marked on figure. ....	.....	160
<b>Fig 5.2.2</b> Energy decomposition of RSP-05 and RSP-40, RSP-52 and RSP-53 showing the free energy contribution as a function of residue number. ....	.....	161
<b>Fig 5.2.3</b> Schematic representation of RSP-40 binding with Ile62, Leu63, Leu64 and Cys94. Hydrogen bond between RSP-40 and Cys94 marked by black line.....	.....	161
<b>Fig 5.2.4</b> Energy decomposition of RSP-05 and RSP-40, RSP-52 and RSP-53 showing the energy contribution of residues as a function of residue number. ....	.....	162
<b>Fig 5.2.5</b> Free energy decomposition per residues in <i>Open2</i> R-spondin ligand MD simulation. Major residues had been marked on the figure. ....	.....	164
<b>Fig 5.2.6</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-57, RSP-60, RSP-64, RSP-68, RSP-70, RSP-74 and RSP-76.....	.....	164
<b>Fig 5.2.7</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-62. ....	.....	165
<b>Fig 5.2.8</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-72. ....	.....	167
<b>Fig 5.2.9</b> Key residues that contribute to the binding energy of RSP-72.....	.....	167
<b>Fig 5.2.10</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-58, RSP-71 and RSP-73.....	.....	168

<b>Fig 5.2.11</b> Free energy decomposition per residues in RNF43 ligand MD simulations. Major residues marked on figure.....	170
<b>Fig 5.2.12</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-18, RSP-20, RSP-27, RSP-29, RSP-43, RSP-45, RSP-46 and RSP-47. ....	171
<b>Fig 5.2.13</b> RSP-20 (orange) binding to RNF43 (pink). Important residues labeled..	171
<b>Fig 5.2.14</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-22 and RSP-49.....	172
<b>Fig 5.2.15</b> Free energy decomposition per residues in LGR5 ligand MD simulations. Major residues marked on figure.....	175
<b>Fig 5.2.16</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-83, RSP-85, RSP-91 and RSP-92. ....	175
<b>Fig 5.2.17</b> Total binding free energy of RSP-83 binding to LGR5 protein as a function of time.....	175
<b>Fig 5.2.18</b> Decomposition of protein-ligand binding free energies in the simulations of RSP-82, RSP-86, RSP-88 and RSP-89. ....	176
<b>Fig 5.2.19</b> LGR5 binding patterns of RSP-85 and RSP-89. Major binding free energy contributors near RSP-85 coloured by purple and contributors near RSP-90 colored yellow. LGR5 protein structure coloured green, RSP-85 coloured light blue and RSP-89 coloured orange. ....	176
<b>Fig 5.2.20</b> Energy decomposition of RSP-90/LGR5 interaction by residue.....	177
<b>Fig 5.2.21</b> Binding pattern of RSP-90 binding to LGR5. Protein structure coloured green and RSP-90 were coloured light blue. Key residues were labeled. ....	177

## List of tables

<b>Table 3.1</b> Major filtration parameters in the first-round compounds filtration. ....	54
<b>Table 3.2</b> Major filtration parameters in the second-round compounds filtration. .	55
<b>Table 3.3</b> Major binding pocket descriptors calculated by DoGSiteScorer targeting <b>Closed</b> and <b>Open2</b> R-spondin protein structure. ....	58
<b>Table 3.4</b> Major descriptors of three binding pockets calculated by DoGSiteScorer targeting RNF43 protein.....	62
<b>Table 4.1.1</b> Protein models and number of simulated compounds for each protein structure.....	79
<b>Table 4.1.2</b> Basic information of four simulation systems .....	81
<b>Table 4.2.1</b> Name and ZINC ID of the thirteen selected <b>Closed</b> R-spondin ligands and their corresponding molecular docking scores and ranking among the top 1000 compounds.....	89
<b>Table 4.2.2</b> Ligand binding stability descriptors of thirteen <b>Closed</b> R-spondin ligands in the first-round selection. ....	90
<b>Table 4.2.3</b> Highly populated hydrogen bond interaction in RSP-03, RSP-05 and RSP-06 MD simulation.....	92
<b>Table 4.2.4</b> Binding stability descriptors of <b>Closed</b> R-spondin ligands in the third-round selection. ....	96
<b>Table 4.2.5</b> Binding stability descriptors of eight <b>Closed</b> R-spondin ligands in the third-round selection. ....	98
<b>Table 4.2.6</b> Highly populated hydrogen bond interactions in RSP-52, RSP-53, RSP-54 and RSP-55 MD simulation.....	99
<b>Table 4.3.1</b> Name and ZINC ID of thirteen <b>Open2</b> R-spondin ligands and their corresponding molecular docking scores and rankings among the top 1000 compounds.....	109
<b>Table 4.3.2</b> Binding stability descriptors of thirteen <b>Open2</b> R-spondin ligands in the first-round selection.....	110
<b>Table 4.3.3</b> Highly populated hydrogen bonds formed by four stably bound compounds and <b>Open2</b> R-spondin during the course of 100ns MD simulation. ....	112
<b>Table 4.3.4</b> Binding stability descriptors of <b>Open2</b> R-spondin ligands in the second-round selection. ....	114
<b>Table 4.3.5</b> Highly populated hydrogen bond interaction in RSP-70, RSP-71, RSP-72, RSP-73 and RSP-74 MD simulation. ....	117
<b>Table 4.4.1</b> Ligand name and Zinc ID of twenty-three selected RNF43 ligands and their corresponding molecular docking scores and rankings. ....	127

<b>Table 4.4.2</b> Major binding stability descriptors of twenty-three RNF43 ligands in the first-round selection.....	129
<b>Table 4.4.3</b> Highly populated hydrogen bond interaction in RSP-18, RSP-20, RSP-22, RSP-27 and RSP-29 MD simulation. ....	130
<b>Table 4.4.4</b> Binding stability descriptors of eight RNF43 ligands in the second-round selection. ....	137
<b>Table 4.4.5</b> Highly populated hydrogen bond interactions in RSP-45, RSP-46, RSP-47 and RSP-49 MD simulation with RNF43. ....	140
<b>Table 4.5.1</b> Name, ZINC ID of eight compounds of LGR5 compounds and their corresponding molecular docking scores and rankings among the top 1000 compounds.....	143
<b>Table 4.5.2</b> Binding stability properties of eight LGR5 ligands in the first-round selection. ....	144
<b>Table 4.5.3</b> Highly populated hydrogen bond interactions in RSP-82, RSP-82 and RSP-85 MD simulation.....	147
<b>Table 4.5.4</b> Binding stability properties of nine LGR5 ligands in the second-round selection. ....	150
<b>Table 4.5.5</b> Highly populated hydrogen bonds between RSP-86, RSP-89, RSP-90, RSP-91 and RSP-92 and LGR5.....	155
<b>Table 5.1</b> Number of compounds selected for free energy calculation by protein target/conformation. ....	157
<b>Table 5.2</b> Binding free energy and standard deviation of nine <b>Closed</b> R-spondin ligands. ....	159
<b>Table 5.3</b> Binding free energy and standard deviation of thirteen <b>Open2</b> R-spondin ligands. ....	163
<b>Table 5.4</b> Energy of residues with highest contribution in RSP-60/ <b>Open2</b> R-spondin simulation. Energies in kcal/mol. ....	165
<b>Table 5.5</b> Energy of residues with highest contribution in RSP-62/ <b>Open2</b> R-spondin simulation. Energies in kcal/mol. ....	166
<b>Table 5.6</b> Energy of residues with highest contribution to in RSP-58/ <b>Open2</b> R-spondin simulation. Energy in kcal/mol. ....	168
<b>Table 5.7</b> Binding free energy and standard deviation of eleven RNF43 ligands....	169
<b>Table 5.8</b> Binding free energy and standard deviation of ten LGR5 ligands. ....	173

## List of abbreviations

APC	adenomatous polyposis coli
ACE	analytic continuum electrostatic
AMBER	Assisted Model Building with Energy Refinement
BR	basic region
CK1	casein kinase 1
COM	center of mass
CRD	cysteine-rich domains
CS	consensus sites
CUDA	Compute Unified Device Architecture
DKK1	dickkopf-related protein 1
Dvl/Dsh	dishevelled
EGF	epidermal growth factor
Fz	Frizzled
GA	genetic algorithm
GAFF	Generalized Amber Force Field
GB	Generalized Born
GBSA	generalized Born surface area
Gln	Glutamine
GPCR	G protein-coupled receptor
GSK3 $\beta$	Glycogen synthase kinase 3 $\beta$
HSPG	heparan sulfate proteoglycans
LGR	leucine-rich repeat-containing G-protein coupled receptor
LRP	low-density lipoprotein receptor related protein
MD	molecular dynamics
MM	Molecular Mechanics
MM-PBSA	Molecular Mechanics/Poisson-Boltzmann Surface Area
MMTV	mouse mammary tumour virus
MOE	Molecular Operating Environment
NPT	constant number of atoms, pressure and temperature

NVE	constant number of atoms, volume and energy
PARP	poly ADP-ribose polymerase
PBE	Poisson-Boltzmann equation
PBSA	Poisson Boltzmann surface area
PCP	planar cell polarity
PDB	Protein Data Bank
PMEMD	Particle mesh Ewald molecular dynamics
PORCN	Porcupine
RMSD	root mean square deviation
RMSF	root-mean-square fluctuation
RNF43	ring finger 43
R-spondin/RSPO	roof plate-specific spondin
SD	standard deviation
TCF/LEF	T-Cell Factor/lymphoid enhancer factors
TIP3P	transferable intermolecular potential with 3 points
TSP	thrombospondin
TSR	thrombospondin type 1 repeat
VMD	visual molecular dynamics
WT	wild type
ZNRF3	zinc and ring finger 3
$\beta$ -TrCP	$\beta$ -transducin repeat containing protein

## Abstract

Wnt pathway is a key cellular signal responsible for carcinogenesis and embryonic development. Excessive Wnt signal induces intracellular accumulation of  $\beta$ -catenin which improves DNA replication and cell abnormal proliferation. R-spondin is an essential fine-tuning signal for Wnt signal; studies have suggested that it could perform as a hinge to connect E3 Ubiquitin ligase RNF43 and LGR5 protein with high affinity and then promote the elimination of RNF43 in the cell surface. RNF43 is responsible for the degradation of the Wnt receptor, and with too much R-spondin signal, the degradation of the Wnt receptor is inhibited and Wnt strength is improved which would cause significant cell abnormal proliferation.

With the aim of modulating R-spondin interactions, we firstly examined the druggability of R-spondin, LGR5 and RNF43 through energy-based and shape-based binding pocket prediction methodologies; we find that RNF43 and R-spondin in an open conformation (which we call "**Open2**") are predicted as highly druggable. After refining 1.3 billion compounds to 20.7 million, molecular docking of this subset was performed against these proteins, based on their drug binding site information. Ninety-three top-ranked compounds were then simulated by molecular dynamics: 43 of these compounds were successfully identified that could bind well with their corresponding protein structures. We then performed free energy calculations and energy decomposition to quantify the binding affinity and energy contribution of each residue in these complexes.

This research aims to identify compounds that could potentially inhibit the binding between R-spondin, RNF43 and LGR5 in order to probe its anti-cancer activity. From binding pocket identification and molecular dynamics (MD) simulation, RNF43 is predicted to be the most druggable among the protein structures. In our ligand optimization process, increasing the hydrophobicity of a compound to promote binding with the upper cavity on RNF43 could dramatically increase the binding stability of the ligand. An observation also occurred in the simulation of compounds bound to protein LGR5, as benzonitrile and chlorobenzene group tend to form highly stable hydrophobic interactions with the Trp168/Thr192 cavity on LGR5.

The binding stability of compounds to the **Closed** conformation of R-spondin is aided by Cys94 hydrogen bond interactions and by hydrophobic residues including Ile63, Leu63 and Leu64. However, **Closed** R-spondin is sensitive to the length of

ligand, as many compounds with different length were tested and only limited length molecules succeeded.

The cavity on ***Open2*** R-spondin is deep but unstable: in some simulations, the cavity was not stable enough to sustain interactions. Three groups of residues including residues near Leu64 (Furin-like 1 domain), Cys94 (hinge area) and Glu103 (Furin-like 2 domain) are important for the maintenance of ***Open2*** R-spondin and ligand binding stability.

In conclusion, this study provides insight into the structure, energy and dynamics of ligands targeting four R-spondin signalling related protein structures; and can provide valuable input into the choice and design of inhibitors of Wnt signalling, as chemical tools and potential drug molecules.

## **Declaration**

The author declares that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

## Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions", which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf>), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see <http://www.manchester.ac.uk/library/aboutus/regulations>) and in The University's policy on presentation of Theses.

## Acknowledgements

The author would like to express his gratitude to a number of people, without whom the completion of this work is impossible.

I would like to express my deepest thanks for my supervisor Dr Richard Bryce. Thanks for your careful guidance, immense patience and professional support along my study at the University of Manchester. I would also like to thank Dr Rupesh Chikhale for helping me out with great patience when I first started my work in Manchester and teaching me so much on molecular modelling.

Thanks must be given to all current and former members in computational-aided drug design group for caring for me and teaching me so much. I would also like to thank Patricia Pacios Centeno and Dr Donald Ward for cooperating with me on the calcium sensing receptor work. I would like to thank the Computational Shared Facility at the University of Manchester for providing computational resource and IT support.

Finally, I would like to thank my family and friends; their deepest love and understanding enable me to complete my study and research successfully.

## Introduction to protein dynamics in drug design

X-ray crystallography has largely facilitated structural research on biomolecules and their interactions. These structures, however, are usually static snapshots of molecular structure. Molecular dynamics (MD) simulation<sup>1</sup> is a powerful computational technique that applies Newtonian physics to computationally mimic the behaviour of flexible macro-biomolecules and predict their motion in a given set of conditions. The MD method is gaining increasing recognition in drug discovery research.

A protein under normal conditions is in a dynamical state; one fixed snapshot of the protein is not fully reliable for drug research. When a drug compound is approaching its target, the protein is not a frozen structure but rather a dynamical biological system. The major benefit of the molecular dynamics method is that it can simulate protein flexibility and help the researcher to gain insight into protein motion. Molecular dynamics can also provide a reference for protein binding pocket prediction. For example, a new HIV binding trench was identified through using molecular dynamics simulation<sup>2</sup>. MD can also be used to provide a flexible protein structure for docking, rather than traditional methodology using only one single still structure<sup>3</sup>. MD can also be applied as a validator of novel ligand binding once the ligand has been docked.

Molecular dynamics simulations are useful in revealing details that cannot be obtained from experiments. With increasing computational power becoming available, MD simulation will be a very promising tool in pharmaceutical research.

# In silico design of anti-cancer compounds targeting R-spondin using molecular dynamics and free energy calculations

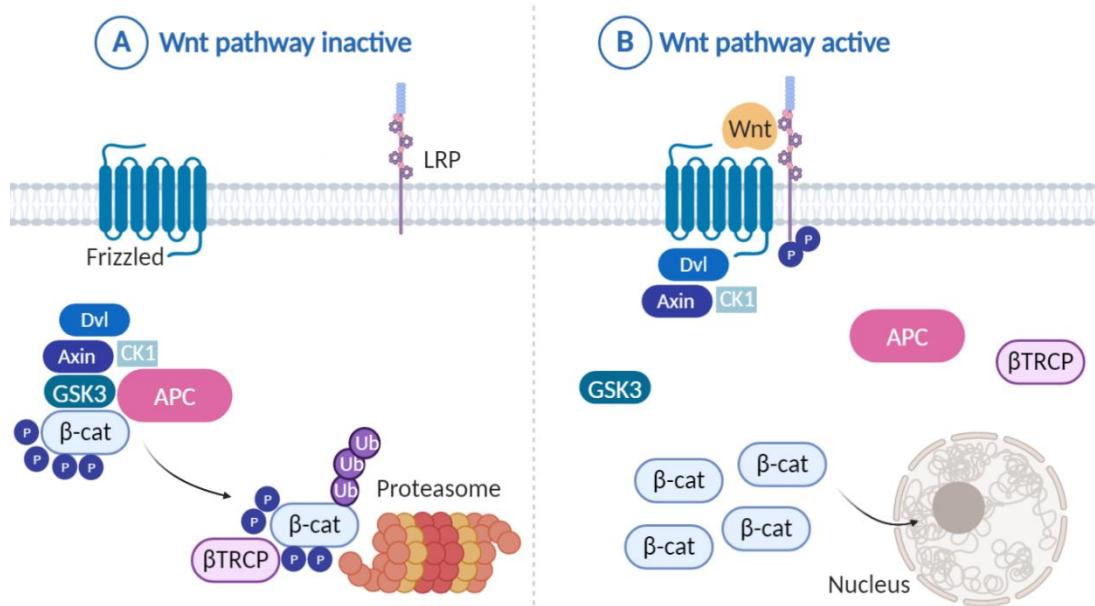
## 1. Background

### 1.1 Wnt signalling pathway

Wnt1 gene is the initial member of the Wnt family and was firstly discovered in 1982. It was named int1 as a product in the oncogenesis induced by mouse mammary tumour virus(MMTV)<sup>4,5</sup>. Later in 1987, int1 was found to be identical to Drosophila wingless gene<sup>6</sup>. To avoid being verbose, the Wingless/int protein family gradually became the Wnt family. Wnt is a highly conserved secreted protein family which is a crucial mediator in embryogenesis and adult tissue homeostasis. The Wnt family can also participate in many other biological processes including mitogenic stimulation, cell fate determination and differentiation<sup>7</sup>.

The most elucidated Wnt pathway is the Wnt/β-catenin pathway which is also called the canonical Wnt pathway (**Fig 1.1**)<sup>5</sup>. The Wnt/β-catenin pathway has demonstrated significant effects in the body's axis development, cell self-renewal and cancer genesis. Abnormal expression of β-catenin has been reported in many cancer types ranging from colorectal cancers to breast cancers<sup>8,9</sup>. In the absence of Wnt signal, β-catenin is kept in the self-destruction complex formed by APC, Axin, GSK3β, CK1 and Dvl which could activate the phosphorylation, ubiquitination and degradation of β-catenin. Secreted Wnt protein transfers its signal to downstream proteins via the combination of Frizzled and its co-receptor LRP5/6. Frizzled's

cytoplasmic tail interacts with Dishevelled (Dvl) and phosphorylated LRP cytoplasmic tail recruits Axin making the self-destruction complex fall apart.



**Fig 1.1** Schematic model of Wnt/β-catenin pathway. When the Wnt pathway is inactive, β-catenin is retained in its self-destruction complex. However, when Wnt pathway is on, Wnt proteins bind with a heterodimeric receptor formed by Frizzled and LRP improving the recruitment of Dishevelled (Dvl), Axin and CK1 ternary complex which is key for β-catenin self-destruction complex. Hence, the Wnt pathway makes β-catenin escape from degradation and hence, promote abnormal cell proliferation. The figure is adapted from Clevers et al<sup>5</sup>.

In the canonical Wnt/β-catenin signalling pathway, Wnt proteins are secreted to the extracellular space and bind to the seven-transmembrane helix receptor, Frizzled (Fz), to activate the Wnt cascade in the human body<sup>7</sup>. In normal tissues, the concentration of β-catenin remains at a stable level through self-degradation; however, Wnt could interrupt β-catenin maintenance and hence dramatically increase cytoplasmic β-catenin concentration. β-catenin is considered as an important biomarker of carcinoma<sup>10</sup> and abnormal Wnt signal promotes the translocation of β-catenin into the nucleus. In the nucleus, β-catenin will interact

with transcription factor TCF and prohibit the combination between TCF and Groucho which is considered to be a suppressor of gene transcription. It is also noticeable that the Wnt/β-catenin pathway could enhance the ovarian resistance to platinum chemotherapeutical agents<sup>11</sup>. As a consequence, overexpression of the Wnt/β-catenin pathway activates cell abnormal proliferation<sup>12</sup>.

## **1.2 R-spondin signalling and disease**

R-spondin (roof plate-specific spondin, also RSPO) belongs to the thrombospondin type 1 repeat (TSR-1)-containing protein superfamily. It was first discovered in 2004 as a family of secreted proteins in Xenopus. The prefix R in R-spondin subfamily derives from the persistent appearance of this protein family in the neural tube after conception, in the region between the roof plate and neuroepithelium<sup>13</sup>. The Wnt signalling pathway is highly sensitive to regulation by R-spondin. The discovery of R-spondin receptor leucine-rich repeat-containing G-protein coupled receptor (LGR), E3 ubiquitin ligase zinc and ring finger 3 (ZNRF3) and ring finger 43 (RNF43), has been a breakthrough in elucidating Wnt signalling and regulating Wnt signal strength<sup>14</sup>.

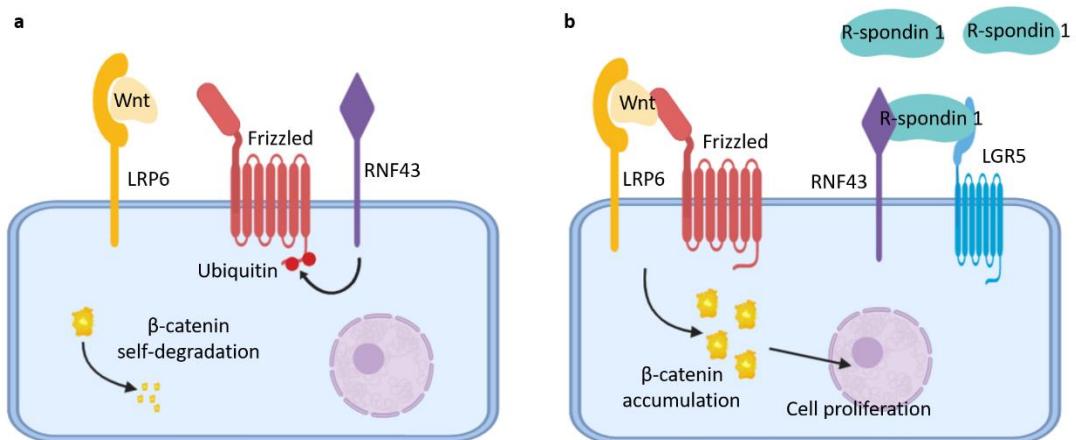
There are four distinct domains in R-spondin protein including two Furin-like cysteine-rich domains (CRD), a thrombospondin (TSP) domain and a basic region (BR). The Furin 1/2 domain is essential for the promotion of the canonical Wnt signalling pathway. Loss of either Furin domain would induce severe change, whereas loss of TSP and BR could not induce the same effect as the Furin domains<sup>15,16</sup>. The TSP domain and BR domain directly bind to glycosaminoglycans or

proteoglycans<sup>15</sup>. An interesting study suggested the existence of an LGR-independent way of antagonizing E3 ubiquitin ligase by binding with heparan sulfate proteoglycans (HSPG) using the TSP domain and BR domain<sup>17</sup>.

R-spondin is an upstream protein of the Wnt/β-catenin pathway which is closely related to cell proliferation, differentiation and embryonic development. R-spondin can also perform a wide range of fate decision bioactivities. The expression of R-spondin is closely related to the phenotypic sex determination of an embryo which is determined by gonadal development. It can disrupt gonadal development and induce XX sex reversal<sup>18</sup>. R-spondin 2 protein could enhance Wnt3A and R-spondin 1 and Wnt3A could jointly amplify the strength of Wnt pathway<sup>16,19</sup>. R-spondin 2 can also positively regulate bone metabolism and osteogenesis through the canonical Wnt pathway. Exogenous R-spondin 2 can rescue bone loss and improve bone microarchitecture<sup>20</sup>. R-spondin 2 is also required for the respiratory tract and limb development and morphogenesis<sup>21</sup>.

The leucine-rich repeat-containing G-protein coupled receptor (LGR) was first identified to be associated with the self-renewal of small intestine stem cells and became a biomarker for small intestinal and colon stem cells<sup>22</sup>. LGR protein acts as R-spondin receptor with high affinity and performs as an important mediator in the Wnt/β-catenin pathway. Dysfunction of LGR is related to body developmental processes and normal tissue maintenance. Overexpression of LGR protein has been reported to accompany several types of cancers<sup>23,24</sup>. Knockout of the LGR gene can induce neonatal lethality<sup>25,26</sup>.

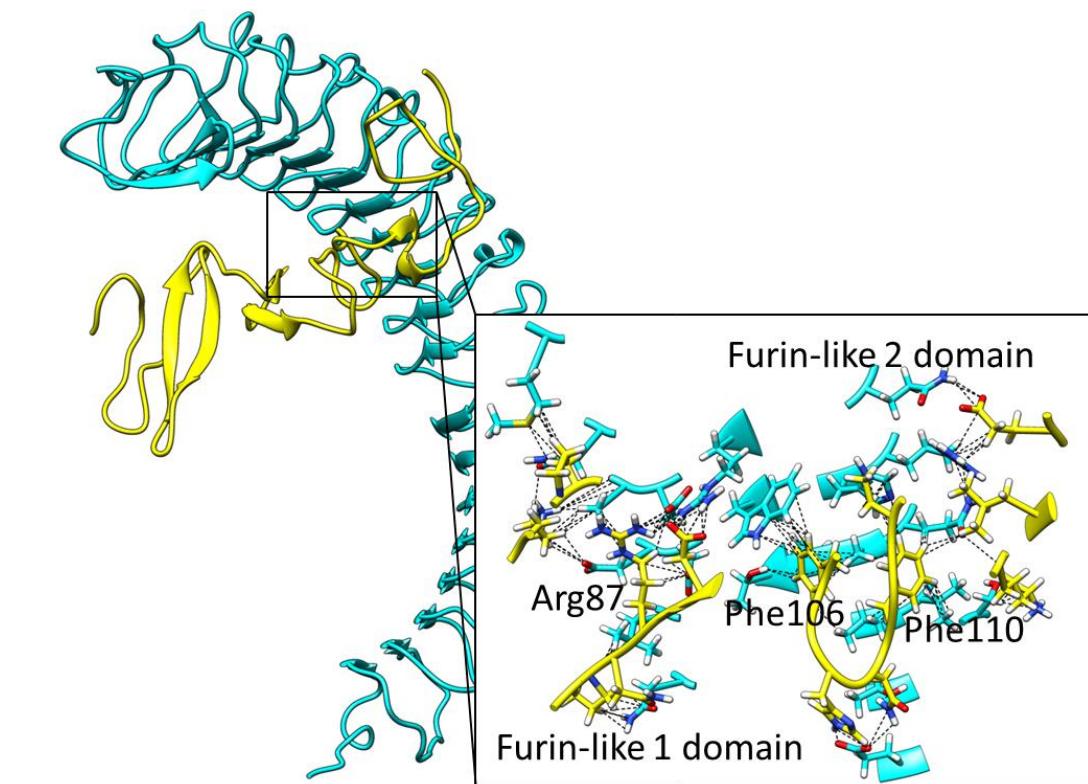
Early studies on the distribution of Wnt product showed a distinct feature of Wnt receptor: internalization of the receptors plays an important role in the regulation of  $\beta$ -catenin<sup>23,24</sup>. Later research has shown that Clathrin-mediated endocytosis coincides with R-spondin 3 and LGR4 co-internalization and is required for the R-spondin 3 induced  $\beta$ -catenin accumulation<sup>27</sup>. ZNRF3 and its homologue RNF43 are cell-surface transmembrane E3 ubiquitin ligases. By promoting the ubiquitylation and degradation of Wnt signalling receptors, RNF43 perform as a potent negative regulator of the canonical Wnt pathway (Fig 1.2)<sup>28</sup>. Cytosolic scaffold protein dishevelled (DVL) is required for the function of RNF43. DVL bind with the cytoplasmic tail of RNF43 and act as a link between RNF43 and Wnt receptors<sup>29</sup>. The ability of R-spondin to recruit RNF43 determines the effect of this activity, rather than the engagement of LGR protein<sup>14</sup>.



**Fig 1.2** Simplified schematic representation of R-spondin involved the Wnt pathway. **a)** RNF43 binds with Frizzled and promotes the clearance of Wnt receptor.  $\beta$ -catenin is confined in its self-degradation complex. **b)** R-spondin promote the binding between LGR5 and RNF43. Without the inhibition by RNF43, the strength of the Wnt pathway increases which helps  $\beta$ -catenin escape from self-degradation, and hence, promotes cell abnormal proliferation.

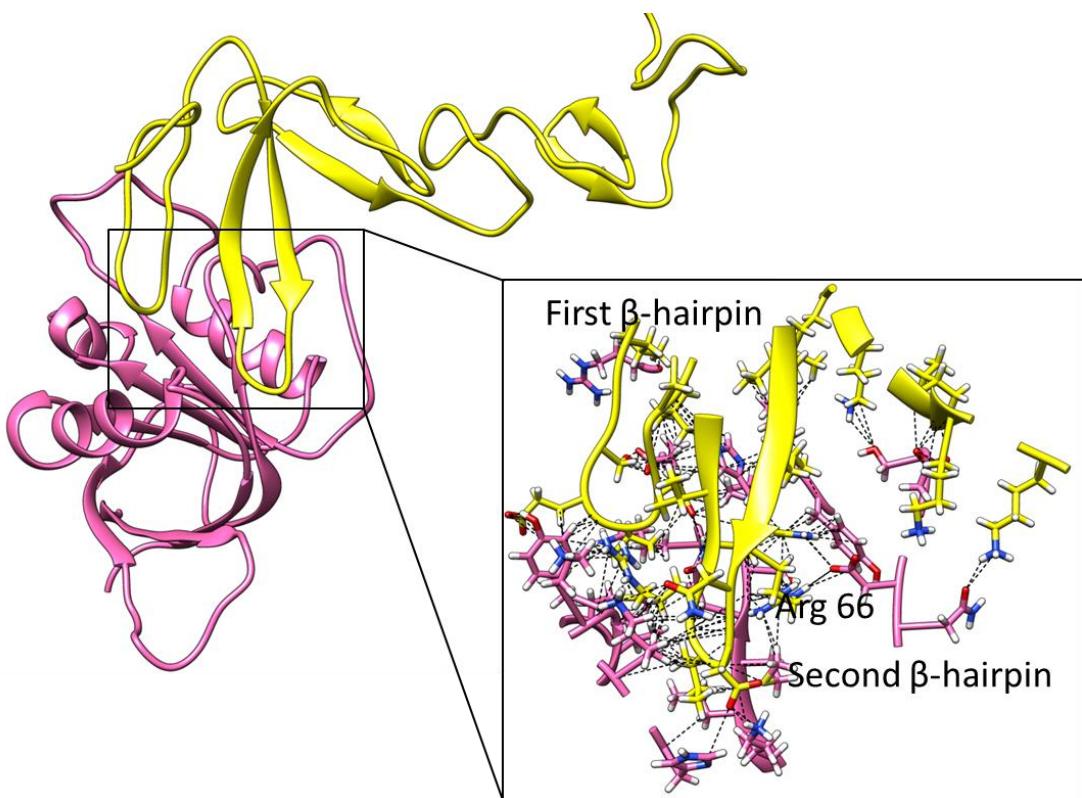
### 1.3 Structure basis of R-spondin signalling

LGR5 and RNF43 do not directly contact with each other but use R-spondin as a mediator to link them together. Hence, interrupting the contact between these three components could be a promising solution for Wnt inhibition. R-spondin has been reported to combine with LGR5 at a nanomolar level<sup>25</sup>. LGR5 is a large scaffold protein forming an interfacial area of up to 2000 Å<sup>2</sup> with R-spondin, the majority of which is hydrophobic. The two aromatic rings of Phe106 and Phe110 on R-spondin protrude into the concavities on LGR5. The interface between LGR5 and R-spondin is shown in **Fig 1.3**.



**Fig 1.3** The contact interface between R-spondin and LGR5, contact interaction marked black dotted line, LGR5 coloured cyan and R-spondin 1 coloured yellow.

The interface between R-spondin and RNF43 is mainly mediated by two  $\beta$ -hairpin structures of the Furin-1 domain. The conserved residues Leu64, Arg66 and Gln71 of R-spondin are essential for the recognition between these two proteins<sup>30</sup>. A hydrophilic interaction is observed in Arg66 and Gln71 forming hydrogen bond and salt bridge with Gln84, His86 and Asp97 on RNF43. A small hydrophilic area is formed between Leu64 of R-spondin and His86 and Leu88 of RNF43. The loop of R-spondin  $\beta$ -hairpin Arg70 to Asn76 inserts into a hydrophobic groove on RNF43 (**Fig 1.4**).



**Fig 1.4** The contact interface between R-spondin 1 and RNF43, molecular contacts are marked black dotted line, R-spondin 1 coloured yellow and RNF43 coloured pink.

The interface between R-spondin and LGR5 could be chosen to be a potential drug target to inhibit the binding between these two proteins. Similarly, studies has

suggested that RNF43 serves as the primary target of R-spondin<sup>29</sup> and activates the internalization of Wnt receptors. Hence, inhibiting the binding between RNF43 and R-spondin Furin-1 domain could also help to prevent RNF43 from exhaustion and down-regulate Wnt signal.

A recent study has suggested that R-spondin could potentiate Wnt signalling in the absence of LGR. This research proposes the possibility of the existence of an unknown mechanism that could also induce the R-spondin-mediated RNF43/ZNRF3 combination<sup>31</sup>. Another study proposes a novel explanation of an LGR-independent mechanism, suggesting heparan sulfate proteoglycans (HSPG) could bind to the thrombospondin type 1 (TSP) domain and basic region (BR) domain<sup>17</sup>. This potential mechanism could cast doubt to the choice of binding sites between LGR5 and R-spondin. Although there might exist an alternative functional mechanism, the role of R-spondin in the ternary complex will not alter. It will block the LGR-dependent promotion of the Wnt signal.

#### **1.4 Protein dynamics**

Due to the unique amino acid sequences, proteins generally have different structures. However, these structures are not static, and they are often flexible enough to transit from one status to another. Protein dynamics is concerned about the behaviour that protein conformations keep changing between different states and the alteration from their equilibrium state. Protein dynamics is important for the biological activity and functionality of proteins. Allosteric regulation is one of the most representative behaviour in this protein dynamics, for example, one

subunit of haemoglobin tetramer binds with oxygen could improve the allosterically activate the affinity of oxygen affinity of the whole protein.

Protein dynamics could be described in different scales. The first scale is in atomic and residue level which could be concluded into harmonic and anharmonic excursion from the equilibrium state. The harmonic excursion includes stochastic fluctuation of bonds or bond angle and anharmonic contains different energy minima or rotamers of residues.

The second level of protein dynamics contains residues coupling in one or multiple subdomains of protein. Since the sidechains of residues have different special direction, contiguous residues or distal residues brought together by protein folding are approximal enough to form multiple types of interactions including hydrogen bonds, van der Waals interaction, ionic bond, metal bonds.

The last level is protein domain dynamics which is caused by the motion and flexibility of multiple domains in one protein. The inter-domain flexibility is important for a lot of protein functions including ion channel, intercellular signalling transduction, biological catalysis, the formation of protein complex and metabolites transportation. There are usually hinge areas that connect different secondary structures together and the movement of the hinge areas would cause significant changes in protein conformation. This phenomenon happens in the conformation of R-spondin (**Chapter 3.2**). The protein plasticity is also caused by domain motion

in a functional site, helical extension in the linkage of  $\alpha$ -helix, sheer motion in a flat architecture.

### **1.5 Aim of this work**

It has been over 30 years since the first discovery of Wnt signalling pathway; a large variety of signal transduction mechanisms have been elucidated and some Wnt regulators are under clinical trials like Porcupine inhibitor, WNT974<sup>32</sup>. However, none of the drugs targeting this pathway has passed into the market. One of the major reasons is the complex regulatory network of Wnt/ $\beta$ -catenin pathway on carcinogenesis and normal tissue maintenance<sup>33,34</sup>. It is not clear whether it is possible to safely target the  $\beta$ -catenin pathway. Hence, although there are various proteins that could significantly affect the regulation of Wnt signalling, the drug target should be chosen carefully.

The model of LGR5-dependent R-spondin-mediated Wnt potentiation has been extensively studied and it has been recognized as a standard regulatory process of the Wnt pathway<sup>35</sup>. R-spondin can promote canonical Wnt signalling by clearing out the cell surface E3 ubiquitin ligase; this makes it a potential drug target for further drug research. It has been reported that the affinity between R-spondin and LGR-5 is high - the IC<sub>50</sub> of R-spondin 1 binding with LGR5 is at nanomolar level<sup>25</sup>. The Furin 1 domain of R-spondin could also bind with RNF43 and induce depletion of cell membrane RNF43.

The purpose of this work is to identify effective small molecule candidates that aim to modulate R-spondin activity and then regulate downstream  $\beta$ -catenin. Due to the pivotal role of the LGR5, R-spondin and RNF43 collaboration in regulating Wnt receptors, these three proteins were targeting respectively. The following are key steps to determine the top hits.

1. Identify key pockets for targeting of virtual screening of open/closed conformations of R-spondin 1, RNF43 and LGR5;
2. Perform virtual screening against these targets using a suitable ligand library;
3. Apply MD simulation of top-ranked VS hits to assess stability and further filter;
4. Perform free energy analysis to rationalise predicted potency of top hits and to identify key binding residues in the proteins.

## **2. Computational Theory**

The main objective of this research was to obtain the information about potential inhibitive compounds of R-spondin proteins. To achieve this objective, multiple methodologies were used collectively. Firstly, the protein-ligand binding sites of proteins were predicted by different algorithms and software in order to determine the best binding site. After determining the protein-ligand binding site and refining the initial compound database, molecular docking was applied to rank the binding affinities of compounds. As top-scored compounds being selected, molecular dynamics simulation and free energy calculation of protein-ligand complexes were performed to test their binding affinities. This chapter introduced the theoretical backgrounds of methodologies used in this research.

### **2.1 Protein-ligand binding site prediction**

Identifying the correct protein-ligand binding site is one of the most essential and challenging tasks in drug design. There are two widely used protein binding pocket prediction approaches: the first method is a shape-based methodology that builds grid points and calculates the geometrical properties of each grid point to characterize the druggability of the protein and predict potential binding pockets. The second method is the energy-based prediction, which uses a library of small fragments to dock into the potential binding pocket and then search for those with the lowest energy. The underlying theories and representative software of these two methodologies will be discussed below.

### 2.1.1 Shape-based methodology

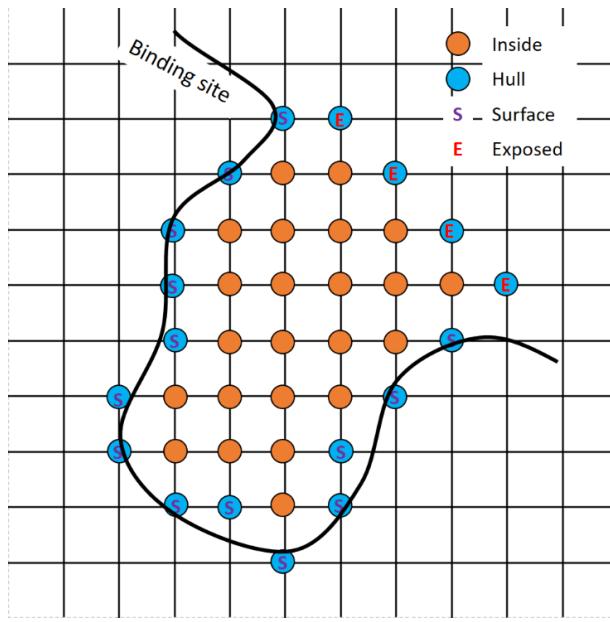
Most successful drug binding pockets share several common features like high hydrophobicity, a deep binding cavity, a high extent of enclosure and so forth. One of the simplest methods to predict ligand binding pocket is to measure 3D space features of the protein surface. This method relies heavily on the protein atom coordinates of protein.

DoGSiteScorer is one of the most representative programs to predict protein binding pockets by the shape of the protein. DoGSite maps the protein onto a grid system with 2.0 Å as the bounding box and 0.4 Å as grid spacing. Grid points are labelled with two states including free and occupied-within the van der Waals radius (**Fig 2.1**). After merging continuous free grid space, cavities whose volume are less than 100 Å<sup>3</sup> are discarded and the remaining pockets are sorted by their volume.<sup>36–38</sup>

Multiple spatial properties are automatically calculated for each pocket, including volume, depth of pocket, quotient of fitted ellipsoid to pocket volume, quotient of two ellipsoid main axis etc. According to these special descriptors, pocket scores are calculated by the following function:

$$SimpleScore = -0.62 + 0.035\sqrt{volume} - 0.016gps\_se\_h + 0.4lipo\_siac$$

where volume is pocket volume and *gps\_se\_h* is the cavity enclosure and *lipo\_siac* is the hydrophobicity of pocket.



**Fig 2.1** 2D schematic representation of different types of grid points in DoGSiteScorer. Blue points form the hull that directly contact the solvent (Exposed) or protein surface (Surface). Orange points represent the grid points that are inside the receptor. Image is adapted from Volkamer et al<sup>38</sup>.

$$\text{SimpleScore} = -0.62 + 0.035\sqrt{\text{volume}} - 0.016\text{gps\_se\_h} + 0.4\text{lipo\_siac}$$

### 2.1.2 Energy-based methodology

Fragment-based methodology is used to identify binding sites by probing multiple low-energy molecules into the surface of the protein and calculate the energy expression by the cluster of small probes<sup>39</sup>. FTMap is one of the representative energy-based binding pocket searching algorithms and the fragment docking energy is calculated as follow:

$$E = E_{vdw} + w_2 E_{elec} + w_3 E_{cavity} + w_4 E_{pair}$$

where  $E_{vdw}$  is the van der Waals interaction energy,  $E_{elec}$  is the electrostatic interaction energy,  $E_{cavity}$  characterizes the hydrophobic enclosure of the cavity and  $E_{pair}$  represents the solvent effect.

Since a hydrophobic cavity is favoured for ligand binding, to characterize the hydrophobicity of a cavity, a gaussian set is put onto grid points and calculate the

correlation with near hydrophobic residues. The following function is used to calculate the fraction of the gaussian set occupied by the non-polar region of receptor and ligand:

$$R_p(l, m, n) = \sum_{i=1}^{N_r} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-r_{i,(lmn)}^2}{\sigma^2}\right)$$
$$L_p(l, m, n) = \begin{cases} 1 & \text{if } (l, m, n) \ni A_L \\ 0 & \text{otherwise} \end{cases}$$

where  $N_r$  is the number of receptor atoms,  $r_{i,(lmn)}$  is the distance between the grid point and atom i and  $A_L$  are atoms in ligand.

## 2.2 Molecular docking

Protein-ligand docking is usually considered the first step of drug high-throughput screening and lead optimization<sup>40</sup>. Two major molecular docking methods were used in this part including the stochastic optimization method and the multi-conformer docking method respectively. Each type of molecular docking methods and important scoring functions will be discussed in the following sections. The objective of applying molecular docking in this research work is as follow:

1. Explore the 3D space of target and predict possible binding poses with a reasonable cost of computing capacity.
2. Explicitly evaluate the binding affinity of each compound through their numerical scores.
3. Identify best compound structures which demonstrate the greatest shape complementarity and hydrogen bond contribution.
4. Obtain key information on protein-ligand interaction and pharmacophores.

### 2.2.1 Molecular docking algorithms

GOLD<sup>41</sup> (Genetic Optimization for Ligand Docking) is a widely used molecular docking software program that applies a genetic algorithm with an evolutionary strategy to automatically explore the conformational space of ligands and dock them to a partially flexible receptor. Genetic algorithm (GA) is an optimization algorithm that mimics the process of evolution of genes by manipulating the data as chromosomes. Each chromosome assigns of the fitness function and helps the program find the most merited solution. GOLD docking software utilizes a steady-state-with-no-duplication GA with a fixed population of 500 chromosomes. Each chromosome records the internal conformation, protein active site and mapping of hydrogen bonds. Based on the information on chromosome, the fitness function of GOLD is as follow:

$$GOLD \text{ fitness function} = S_{hb-ext} + S_{vdw-ext} + S_{hb-int} + S_{vdw-int}$$

where **hb** represents the hydrogen bond interaction, and **vdw** represents the van der Waals interaction; **ext** represents the protein-ligand interaction and **int** represents the ligand intra-molecular interaction.

Each chromosome contains four string including two binary strings recording the angle of rotation of ligand and protein, one integer to map the lone pair in the ligand to receptor hydrogen atoms and one integer to map the hydrogen in ligand to receptor long pairs. And two genetic operators are used in the optimization: mutation and crossover. Mutation requires one parent and produces one child which performs radon mutation in the binary and integer mutation from its parents.

Crossover requires two parents and produces two children which ensure that parent chromosomes copied to children. Using multiple iterations, the docking program could explore the best fitting conformation of compounds.

FRED is a molecular docking that uses an extensive search algorithm to systematically rotate within the receptor and translate into scores by ChemGauss 3 scoring function. Protein structure and ligand information were used to do FRED docking. Firstly, different conformers are generated by the OEOmega program in the OpenEye software by rotating bonds and flipping chiral centres. Then, FRED extensively searches for binding poses of difference conformers in the receptor with a resolution of 1 Å. The ligand flexibility of ligand is implicitly represented by multiple conformers and protein structure is treated as rigid body in the docking process<sup>42,43</sup>. After exhaustive searching, the top 100 scored posed are preserved and further solid body optimizations are conducted, with the best scored posed output to compare with other compounds (**Fig 2.2**).

1. A set of 1D structure is used as input to exhaustively search for 3D conformers.
2. Rotamers of each conformer are enumerated.
3. Translations of compounds near the protein binding site are enumerate.
4. Poses that clash with protein or are far from the protein are eliminated.
5. Poses are scored and low scored poses are eliminated.  
Remaining binding poses are optimized.
6. Output top scored poses

**Fig 2.2** Standard workflow of the FRED molecular docking program<sup>43</sup>.

### 2.2.2 Scoring functions

Good contacts between protein and ligand are the prerequisite of their binding. The ChemScore scoring function is an empirical scoring function designed to calculate the binding free energy of protein-ligand complex. The function is based on the following function<sup>44,45</sup>:

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} \sum_{il} g_1(\Delta r) g_2(\Delta \alpha) + \Delta G_{metal} \sum_{aM} f(r_{am}) + \Delta G_{lipo} \sum_{il} f(r_{il}) + \Delta G_{rot} H_{rot}$$

where  $\sum_{il} g_1 g_2$  is calculated for the possibility of complementary hydrogen bonds; the metal term was calculated by  $\sum_{aM} f(r_{am})$  when there is metal atom in the receptor; lipophilic term  $\sum_{il} f(r_{il})$  is calculated for lipophilic long range interaction between ligand and receptor atoms;  $\Delta G_{rot} H_{rot}$  term identifies frozen rotatable

bonds to avoid too much the entropy loss when the rotatable bond do not form contact with protein.

ChemPLP scoring function<sup>46</sup> was designed for the Protein-Ligand ANT System (PLANTS) molecular docking algorithm which utilizes ant colony optimization. The scoring function of ChemPLP is calculated based on the following function:

$$f_{PLANTS_{CHEMPLP}} = f_{plp} + f_{hb} + f_{hb-ch} + f_{hb-CHO} + f_{met} + f_{met-coord} + f_{met-ch} \\ + f_{met-coord-ch} + f_{clash} + f_{tors} + c_{site}$$

where the  $f_{plp}$  (Piecewise Linear Potential) term is used to describe the complementarity of ligand to protein;  $f_{clash} = \omega_{clash} \sum_{c \in C_{clash}} f_{dist} \left( \left| \left| \overrightarrow{c_{pa}c_{pb}} \right| \right|, c_w, c_{r_{clash}} \right)$  is calculated to avoid internal ligand clash;  $f_{tors}$  is the torsional potential of all rotatable bonds.  $f_{hb}$ ,  $f_{hb-ch}$  and  $f_{hb-CHO}$  are distance dependent and angle dependent hydrogen bond terms based on ChemScore potentials.  $f_{met}$ ,  $f_{met-coord}$ ,  $f_{met-ch}$  and  $f_{met-coord-ch}$  are metal interaction terms for magnesium and calcium.

ChemGauss is the main scoring function used in FRED molecular docking software; the score is smoothed by Gaussian functions to delineate the shape and chemistry of molecules. The following are the main components of ChemGauss 3 scoring function<sup>47</sup>:

$$S_{chemgauss3} = S_{steric} + S_{acceptor} + S_{donor} + S_{metal} + S_{desolvation}$$

where steric term contains two types of effects including an approximation of VdW interaction and protein desolvation energy ignoring any favourable hydrogen bonds with the binding site. An acceptor term and donor term measure the interaction

between ligand acceptor/protein donor and donor/protein acceptor respectively.

Metal term measures the interaction between ligand and metal atom in the active site.  $S_{desolvation}$  is a penalty term when donors and acceptors of the ligand are blocked by solvent hydrogen bond in the active site.

## 2.3 Molecular dynamics

Molecular dynamics is defined as a computational method to study the movements of atoms and molecules. By solving Newtonian's equation of motion, molecular dynamics simulation can mimic and predict the behaviour of molecules or molecular systems.

$$F = ma$$

where F is the force applied on the atom, and a is the acceleration of atom motion.

There are multiple methods to calculate the force and potential energy between atoms. In this research, the molecular mechanics force field was applied in potential energy calculation. In the following sections, molecular mechanics and molecular dynamics algorithms and models will be discussed.

### 2.3.1 Molecular mechanics

Molecular dynamics simulation needs a way of calculating the energy and forces acting on the molecules in the system of interest. A very widely used theory for computing the potential energy of molecules is molecular mechanics. The Born-Oppenheimer approximation<sup>48</sup> (BOA) assumes that atomic nuclei and electronic motion can be treated separately. Based on this assumption, the potential energy

of a molecule could be presented as a form of summation of energy terms. The intramolecular potential of a molecule can be divided into two parts, covalent contribution and noncovalent contribution.

$$E_{total} = E_{covalent} + E_{non-covalent}$$

where  $E_{covalent} = E_{bond} + E_{angle} + E_{dihedral}$  ;

$$E_{non-covalent} = E_{electrostatics} + E_{vdw}.$$

The bond and angle terms are modelled by harmonic oscillators and atoms tend to maintain their bonds and angles in a fixed value. Hence, a set of parameters including equilibrated values obtained from experiments or higher-level ab initio simulations is used to calculate their potential energy. Since bonds rarely deviate from its equilibrated position, to achieve higher calculation efficiency, Hooke's law is used:

$$v(l) = \frac{k}{2}(l - l_0)^2$$

where  $k$  is force constant;  $l$  is bond length and  $l_0$  is the equilibrated status bond length.

The calculation of covalent terms is often more complicated than covalent terms. The Lennard-Jones potential, also called a 6-12 potential, is an approximation to the interaction between a pair of atoms or molecules and it is often used to model van der Waals interaction between atoms. The attractive force decreases with distance as  $r^{-6}$  and repulsive force decreases as  $r^{-12}$ . Since the van der Waals force falls off

rapidly as the distance between atoms increases, hence, to simplify the calculation, a cutoff value is often set to reduce van der Waals force to zero.

Coulomb's Law is one of the simplest ways to approximate electrostatic interactions. Unlike van der Waals forces, long-range electrostatic interactions do not dramatically decrease as the distance increases, and it played an important role in biomolecular system simulation. To directly set a cutoff as for the VdW term would cause severe discontinuity to energy. Therefore, other methods like the Particle Mesh Ewald (PME) method or multipole algorithms are applied.

Following is a typical potential energy function for use in molecular dynamics simulation:

$$V(r^N) = \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \\ + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) + \sum_{i=1}^N \sum_{j=i+1}^N \left( 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

The potential energy function together with parameters like bond/angle/dihedral parameters, force constant etc. could be referred to be force field which is used for potential energy calculation.

### 2.3.2 Water solvent models

Solvent effects play an essential role in the interaction between protein and ligand<sup>49</sup>. There are a lot of models designed to computationally simulate the behaviour of the condensed phase and most of them could be classified into explicit solvent models and implicit solvent models.

Implicit solvent also called a continuum model, is based on the assumption that the solvent could be treated as a homogenous polarizable medium that can give comparable properties as modelling individual solvent molecules.<sup>50</sup> The four interaction operators are cavity creation, electrostatic energy, quantum mechanical dispersion energy and quantum mechanical exchange repulsion. Applying an implicit solvent model can save a lot of calculation time. The polarizable continuum model (PCM) is commonly used as an implicit solvent model and has several variations like the Poisson-Boltzmann equation-based and Generalized Born equation-based which will be used in the following chapter.

Explicit solvent has a simple representation of solvent molecules which could directly interact with simulated molecule. Transferable intermolecular potential functions (TIPs) and the SPC water solvent model<sup>51</sup> are widely used in molecular dynamics simulation, molecular mechanics and Monte Carlo simulation. These models are usually constrained by multiple parameters, for example, in TIP3P water model, the length of OH is 0.9572 Å and the H-O-H bond angle is 104.52 degrees.

### 2.3.3 Periodical boundary conditions

Due to the limitation of computational capacity, it is impossible to simulate the explicit water solvent in an infinite space condition. As a consequence, periodic boundary conditions (PBC) are applied to solve this problem: a chamber is surrounded by multiple images of the chamber itself. As one particle goes through

the boundary of the chamber, the particle reappears from the opposite surface which forms a pseudo-infinite space.

#### 2.3.4 Temperature and pressure regulation

MD simulation is often performed in a micro-canonical ensemble (NVE), which describes a closed system with a constant number of particles, energy and volume. To connect the system to the isothermal-isobaric ensemble (NPT) or canonical ensemble (NVT), several algorithms can be used. To regulate the temperature in MD simulation, correct velocity and random collision of particles at each step can help the system stay at the desired temperature. Langevin dynamics<sup>52</sup> was used to regulate the temperature in this work. Two additional terms were introduced into the equation of motion for Langevin dynamics. The first term is the damping coefficient to mimic the viscous solvent. The second term is the time-dependent random variable to mimic high-velocity collision of solvent molecules. By adjusting the particles' kinetic energy, the system can be maintained at the set temperature.

The Langevin dynamics equation is written as follow:

$$M\ddot{X} = -\Delta U(X) - \gamma\dot{X} + \sqrt{2\gamma k_B T} R(t)$$

where  $U(X)$  is the particle interaction potential,  $\gamma$  is the viscosity,  $T$  is temperature,  $k_B$  is the Boltzmann's constant,  $\dot{X}$  is the velocity and  $\ddot{X}$  is the acceleration.

To maintain the system at a fixed pressure, the simplest way is to change the volume of the system. The Berendsen barostat was used in this work to control the pressure, which rescales the volume of the simulation box by each time step and lets the system oscillate around the desired pressure.

## 2.4 Free energy calculation

Free energy calculation is an important aid to address biological and chemical problems e.g. in determining preferred conformations of biomolecules, evaluate external influences etc. To calculate the solution free energy, the generalized Born solvent model is magnitudes faster than a free energy perturbation (FEP) simulation in the presence of the explicit model.<sup>53</sup> In this work, free energy calculations using an implicit solvent model were used to evaluate the stability of ligands and determine the energy contribution to the protein-ligand binding process.

### 2.4.1 Poisson Boltzmann potential

Poisson-Boltzmann (PB) equation is useful when studying physiological interfaces and the major goal of this equation is to calculate the distribution of electric potential in solvent. When describing the electrostatic contribution of the solvent molecule, it could be transformed into the distribution of charge in a solvent inaccessible surface. For this model, electrostatic potential  $\phi(r)$  is calculated rigorously by the following Poisson equation:

$$\nabla[\epsilon(r)\nabla\phi(r)] = -4\pi\rho(r)$$

where  $\rho(r)$  is the solute charge distribution and  $\epsilon(r)$  is the dielectric constant. There are many forms of Poisson-Boltzmann equations, and as the Poisson equation extend to salt condition, the equation is called Poisson-Boltzmann equation.

## 2.4.2 Generalized Born model

Although the Poisson-Boltzmann equation has good performance in characterizing solvent energy contribution, it is relatively computationally expensive. Generalized Born model is a linear approximation to the PB equation to simplify the calculation and avoid solving the numerical solution of the PB equation to describe continuum solvent.

Generalized Born solvation model is the most widely used implicit solvent model that is divided into two major components: the electrostatic part and non-electrostatic part<sup>54,55</sup>.

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel}$$

The Born solvation energy could be solved by the following function<sup>56</sup>:

$$\Delta G_{solvation}^{el} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i,j}^N \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp\left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right)}}$$

where  $\epsilon$  is dielectric constant;  $q_i$  and  $q_j$  are electrostatic charges of particles;  $r_{ij}$  is the distance between two particles;  $\alpha_i$  and  $\alpha_j$  are called *effective Born radii* which reflects the burial of the molecule.

Successful definition and calculation of Born radius in some cases are keys to access the success of the GB model. Coulombic field approximation of born radii is expressed in the following function:

$$\frac{1}{\alpha_i} = \frac{1}{R_i} - \frac{1}{4\pi} \int \frac{1}{r^4} dV$$

where  $R_i$  is the atomic radii like VdW radius; the volume of the cavity is integrated.

### **2.4.3 MM-GBSA method**

The molecular mechanics-Generalized Born surface area (MMGBSA) method applies a continuum solvent model to the trajectory produced by molecular dynamics simulation to calculate the binding free energy of a protein-ligand complex. The free energy calculation can be written as follows<sup>53</sup>:

$$\Delta G_{total} = G_{complex} - G_{ligand} - G_{protein}$$

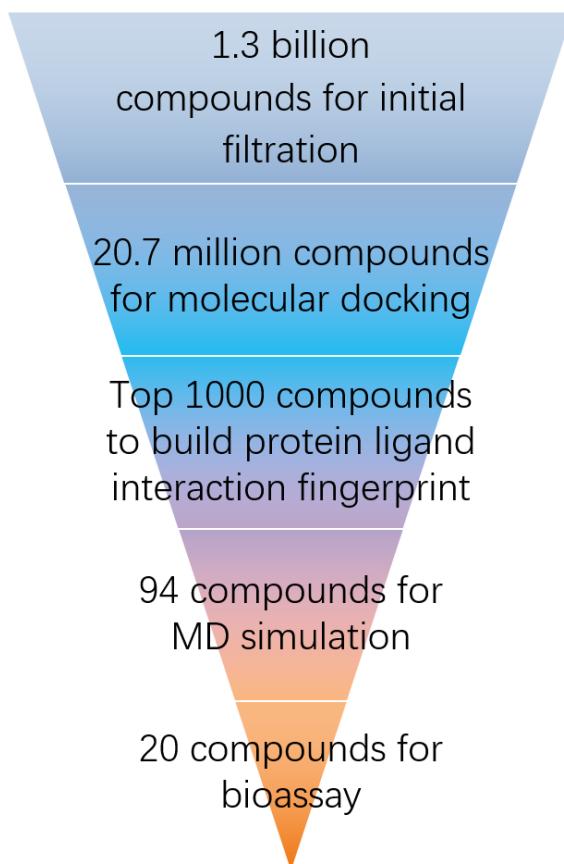
The free energy terms here can be calculated by the following function:

$$G = E_{MM} + G_{psolv} + G_{npsolv} - TS$$

where  $E_{MM}$  is the molecular mechanics energy term calculated from the internal energy, vdw energy and electrostatics;  $G_{psolv}$  and  $G_{npsolv}$  are polar and non-polar solvent energy contributions respectively;  $T$  is the temperature of the system and  $S$  stands for the entropy of molecules.

### 3. High throughput virtual screening

Our overall aim of the work in this section is to identify small-molecular inhibitors of the Wnt pathway by inhibiting the interaction between R-spondin 1, LGR5 and RNF43 proteins. The first step is to initially refine a large compound database to manageable scale by eliminating undesired compounds for further virtual screening; then determine the best binding sites on protein structures; consequently, we use molecular docking to rank by the shape complementarity and hydrogen bond interaction between protein and compounds; by calculating the protein-ligand interaction fingerprint, we summarized some pharmacophores of the potential modulator that might increase the binding affinity to proteins. An overview of the high-throughput virtual screening process is listed in the **Fig 3.1.1.**



**Fig 3.1.1** Schematic representation of the high-throughput virtual screening process.

Firstly, over 1.3 billion compounds from the **ZINC15 database** were used for initial screening. We used the FILTER module in the Openeye software to cut down the database to 20.7 million entities based on their molecular properties including molecular weight, number of ring system, number of chiral centre, number of H-bond donors/acceptors etc.

To obtain the information on the optimal ligand binding site, we used energy-based and shape-based methodologies to determine the most druggable binding sites on Closed and Open2 R-spondin, LGR5 and RNF43. 3. 20.7 million compounds were then docked to each protein site and ranked by molecular docking. Protein-ligand interaction fingerprints (PLIFs) were used to summarise pharmacophores of potential modulator molecules. PLIF is a built-in function in MOE software that directly uses the 3D structure of protein-ligand complex to identify their hydrogen bonds and aromatic interaction. There are five commonly used protein-ligand interactions including: donor to backbone; donor to sidechain; acceptor from backbone; acceptor from sidechain and arene interaction.

The **ZINC 15** database is a publicly accessible compound database designed for ligand virtual screening. By the time we carried the research, the **ZINC 15** database contained 1.38 billion compounds and we used this compound database as input for the initial compound refinement. The database is divided into tranches based on compounds' molecular weight and LogP. We used the **Drug-Like** compound subset marked by the red rectangle in the **Fig 3.1.2**, as input for our high-throughput

virtual screening. This subset contains the compounds whose molecular weight is between 200 and 500 as well as LogP less than or equal to 5.

		Molecular Weight (up to, Daltons)												Totals, by LogP
		200	250	300	325	350	375	400	425	450	500	>500		
-1	36,279	279,807	1,319,183	1,913,073	3,702,316	1,056,946	304,100	70,354	44,461	24,999	5,352	8,756,870		
0	177,244	1,482,832	6,369,098	8,808,603	17,259,185	4,968,583	2,044,226	602,185	402,899	224,910	3,964	42,343,729		
1	504,344	4,642,181	20,366,563	26,946,096	51,168,026	17,552,849	9,214,730	3,522,721	2,402,060	1,354,959	8,543	137,683,072		
2	681,925	7,723,993	38,776,111	51,081,402	101,275,495	40,563,850	25,980,277	12,228,506	8,841,984	5,320,820	22,604	292,496,967		
2.5	263,609	3,772,343	22,243,430	30,062,326	60,668,180	29,200,419	21,065,468	11,712,792	8,809,555	5,530,120	23,822	193,352,064		
3	152,782	2,915,963	19,642,735	27,767,845	54,928,486	32,051,865	25,218,735	15,844,541	12,423,198	7,960,128	39,180	198,945,458		
3.5	67,835	1,849,633	14,766,258	21,907,012	42,592,349	30,725,881	26,784,197	19,167,628	15,474,237	10,310,662	64,003	183,709,695		
4	20,447	830,014	8,715,619	11,870,098	18,186,812	22,714,906	24,444,222	20,149,755	16,923,281	11,790,226	95,185	135,740,565		
4.5	2,545	233,892	4,090,232	6,842,387	11,726,987	16,330,198	18,996,402	18,213,467	16,094,606	11,775,365	131,741	104,437,822		
5	93	35,132	1,271,800	2,922,160	6,070,628	9,754,590	12,748,384	13,716,822	12,887,055	10,036,684	160,390	69,603,738		
>5	28	891	45,626	179,182	556,667	1,238,210	2,075,670	2,666,169	2,967,925	2,481,135	818,316	13,029,819		
Totals, by Weight		1,907,131	23,766,681	137,606,655	190,300,184	368,135,131	206,158,297	168,876,411	117,894,940	97,271,261	66,810,008	1,373,100	1380M Substances 1.7K Tranches	

**Fig 3.1.2** Tranches in the **ZINC 15** compound database chopped by molecular weight and LogP. The red rectangle marks the **Drug-like** subset in the **ZINC 15** database.

Multiple modules including filter, omega, FRED, vida and make\_receptor from OpenEye software were used to process the compound database, prepare protein binding sites, explore the conformational space of compounds and dock them to proteins. PyMOL<sup>57</sup> and UCSF Chimera<sup>58</sup> were used for graphical representation of receptors. MOE was used for protein-ligand interaction fingerprint clustering and analysis.

### 3.1 Compound database refinement

Compounds in the **ZINC 15 drug-like** database<sup>59</sup> have a wide variety of properties: the molecular weight ranges from 250 to 500 and log P ranged from -1 to 5. However, to mimic the protein-protein interaction, desired compounds are

supposed to have larger molecular weight than normal drug molecules. Hence, the initial refinement of compounds focused on filtering out low molecular weight compounds. Also, several other patterns like ring system number, heavy atom number, hydrogen bond donor/acceptor and so forth were defined to eliminate simple compounds structures.

**Table 3.1** Major filtration parameters in the first-round compounds filtration.

Structure pattern	Minimal	Maximal
<b>Molecular weight</b>	330 Da	550 Da
<b>Number of heavy atoms</b>	22	45
<b>Number of ring systems</b>	2	5
<b>Number of functional groups</b>	0	12
<b>Number of heteroatoms</b>	4	12
<b>Number of rotatable bonds</b>	3	12
<b>Number of chiral cents</b>	0	2
<b>Number of H-bond acceptor</b>	0	10
<b>Number of H-bond donor</b>	0	5
<b>XLogP</b>	0	7

Approximately 162 million compounds passed the first compounds filtration. To further refine the volume of the compound database, a second-round compounds filtration was performed to increase the structural diversity of compounds. And after the second database refinement, the volume of the compound databased decreased to 20.7 million.

**Table 3.2** Major filtration parameters in the second-round compounds filtration.

Structure pattern	Minimal value	Maximal value
<b>Number of functional group</b>	2	10
<b>Number of H-bond acceptor</b>	2	5
<b>Number of H-bond donor</b>	3	10
<b>XLogP</b>	3.0	7.5

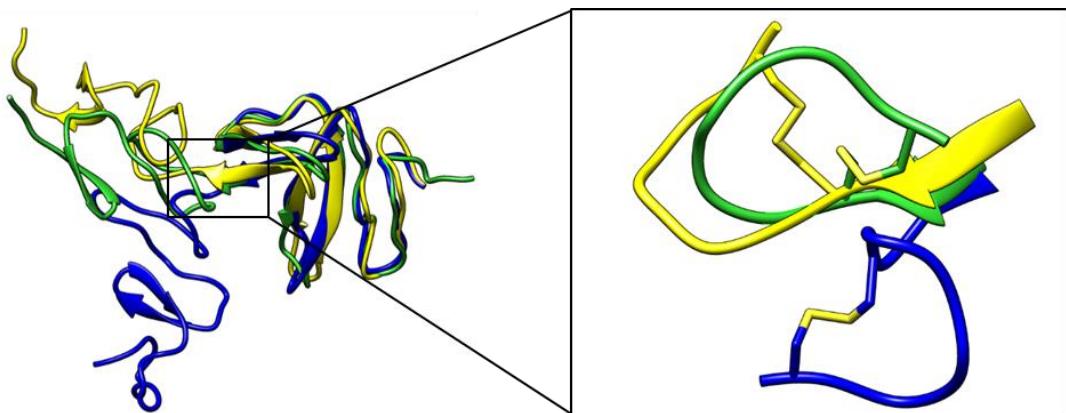
## 3.2 Binding pocket identification

### 3.2.1 *Closed* and *Open2* R-spondin binding pockets

In the crystal structure of R-spondin, there are two major R-spondin conformations: the *Open* conformation and *Closed* conformation. However, according to our molecular dynamics simulation of R-spondin (R. Chikhale, unpublished data), a new possible conformation might exist which we give the name *Open2*. In R-spondin, Furin-like 1 and 2 domains, and 8 disulfide bonds help stabilize the 3D structure of R-spondin; however, the hinge area of R-spondin demonstrates high flexibility. From **Fig 3.2.1**, the hinge area that links the two Furin-like domains requires to twist as R-spondin's conformation changes from *Open* to *Closed*. During a multi-microsecond molecular dynamics simulation of *Open*, the hinge area twists dramatically making the two Furin-like domains locate in a totally different region. This conformation is named *Open2* R-spondin.

To predict the binding cavities of R-spondin protein, we used FTSite<sup>60</sup> which utilizes a small molecule probe/interaction energy methodology for binding pocket prediction. The orange pocket was ranked first among all the cavities predicted (**Fig 3.2.2**) for *Closed* R-spondin. For the orange binding pocket, 41 probe molecules

were identified; however, only 16 and 14 probe molecules were observed in the second (red) and third (blue) binding pockets respectively. Furthermore, these two pockets were far from the binding interface between R-spondin and RNF43. Hence the orange cavity was selected as the most interesting binding pocket for **Closed** R-spondin.

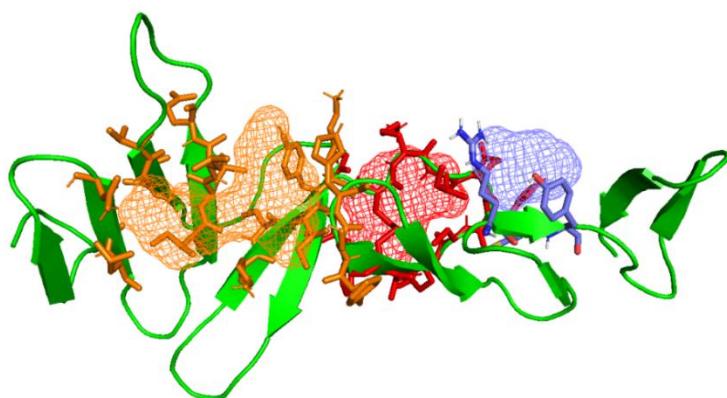


**Fig 3.2.1** Superposition of **Closed**, **Open** and **Open2** R-spondin structures. The hinge area is inset. The yellow structure represents the **Open** conformation of R-spondin-1 obtained from PDB code 4KNG. The green structure represents **Closed** R-spondin-1 obtained from PDB code 4BSO. The blue structure represents the **Open2** conformation obtained from the long-range molecular dynamics simulation of R-spondin 1.

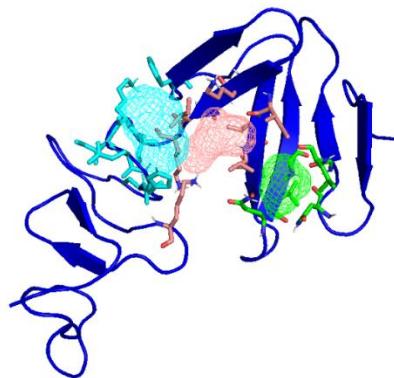
The three top-ranked predicted binding pockets of **Open2** R-spondin (**Fig 3.2.3**) were small compared to **Closed**: 18 probe molecules were found in the first pocket and 16 and 19 probes were found in the second and the third binding pocket respectively; however, the these pockets were close to each other and they could be linked to become a bigger cavity, surrounded by the Furin-like 1 and 2 domains and hinge.

DoGSiteScorer was also used to predict binding pockets of three R-spondin structures mentioned above. DoGSiteScorer is a shape-based binding site prediction software, and it used a quite different algorithm compared to FTSite (see **Chapter 2.1**). DoGSiteScorer found the same results as FTSite. Ten descriptors including pocket volume, enclosure (the ratio of surface grid points to total hull grid points), pocket surface area, pocket depth, the surface/volume ratio, the number of atoms in the pocket, the number of hydrophobic contacts, hydrophobicity, simple score and druggability score of the pocket were used to characterize the druggability of possible binding sites. Druggability score is predicted by machine learning and its value ranges from 0 (undruggable) to 1 (highly druggable).

The two top binding sites of **Closed** and **Open2** R-spondin (**Fig 3.2.4**) demonstrated favourable topological features. The volumes of these pockets are over 715 Å<sup>3</sup> and the depths are around 20 Å (**Table 3.3**). The druggability score of these two pockets were 0.83 and 0.78, meaning that the pocket is highly druggable.



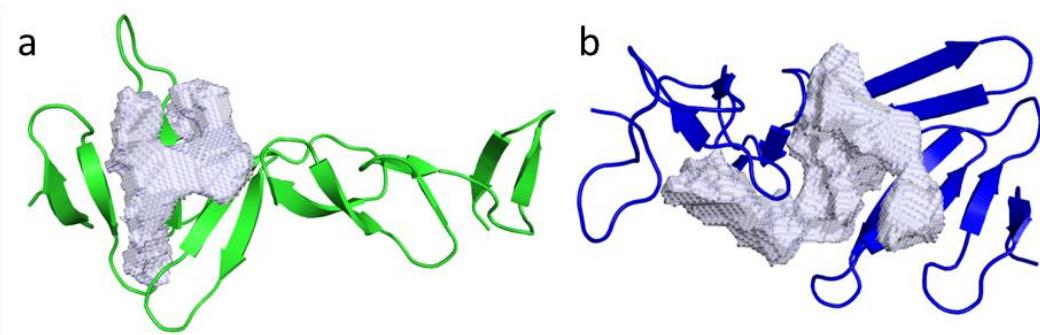
**Fig 3.2.2** Top three binding pockets on the structure of R-spondin **Closed** conformation predicted by FTSite. The protein structure was coloured green. The orange coloured pocket ranked the first place and the red one and blue pocket ranked the second and the third place.



**Fig 3.2.3** Top three binding pockets on the structure of R-spondin **Open2** conformation predicted by FTSite. The protein structure was coloured blue. The pink colour pocket ranked the first place and the green one and cyan pocket ranked second and third place.

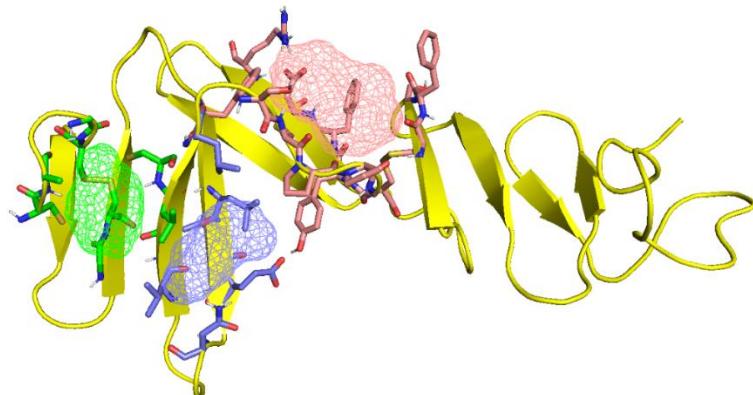
**Table 3.3** Major binding pocket descriptors calculated by DoGSiteScorer targeting **Closed** and **Open2** R-spondin protein structure.

Pocket property	<b>Closed</b> R-spondin	<b>Open2</b> R-spondin
<b>Volume (Å³)</b>	715.1	1015.2
<b>Enclosure</b>	0.19	0.21
<b>Surface (Å²)</b>	1157.8	1578.6
<b>Depth (Å)</b>	20.1	17.1
<b>Surface/volume (1/Å)</b>	1.6	1.6
<b>Site atoms</b>	163	191
<b>Hydrophobic interactions</b>	60	88
<b>Hydrophobicity</b>	0.48	0.50
<b>Simple score</b>	0.51	0.68
<b>Drug score</b>	0.83	0.78

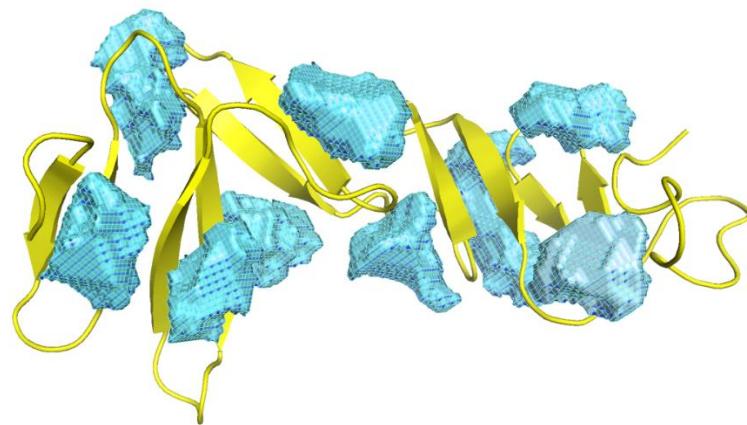


**Fig 3.2.4** Two major binding pockets of **a)** **Closed** R-spondin and **b)** **Open2** R-spondin predicted by DoGSiteScorer.

Finally, we considered **Open** R-spondin. Using FTSite, although there are 32 probe molecules predicted to locate in the first binding pocket (pink, **Fig 3.2.5**), the volume of the pocket was not big enough to be a drug binding pocket. Also, 15 and 13 probe molecules were located in the second (green) and the third (purple) binding pocket and their volumes were also not large. These two pockets are also separated by the residues that are near the interface between two  $\beta$ -hairpin structures, hence these two pockets could not be linked together. This arises because the conformation of **Open** R-spondin is flat; therefore, **Open** R-spondin was not used to screen for R-spondin modulators. DoGSiteScorer was also used to predict binding pockets for the **Open** structure; however, the same problem occurs. There is no major binding pocket and all sub-pockets were widely distributed over the whole protein structure (**Fig 3.2.6**).



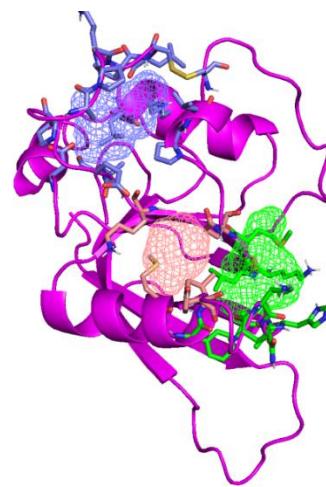
**Fig 3.2.5** Top three binding pockets on the structure of R-spondin ***Open*** conformation predicted by FTSite. The protein structure was coloured blue. The pink colour pocket ranked the first place and the green one and purple pocket ranked the second and the third place.



**Fig 3.2.6** Eight binding pockets of ***Open*** R-spondin predicted by DoGSiteScorer. Binding pockets coloured light blue.

### **3.2.2 RNF43 binding pockets**

FTSite was then used to predict binding sites on RNF43 protein. Although the pink binding site in **Fig 3.2.7** only has four probe molecules, it was ranked first place among all binding sites. It was formed by an  $\alpha$ -helix and the main body of RNF43 protein. The second-ranked binding site is the green one which contains 16 probe molecules. These two binding pockets are near each other and could be linked to form a bigger binding pocket. The third-ranked binding pocket was the purple one which contains 19 probes; however, this pocket is far from the first two binding pockets and this research focuses on the binding pocket near the first and the second binding pocket.



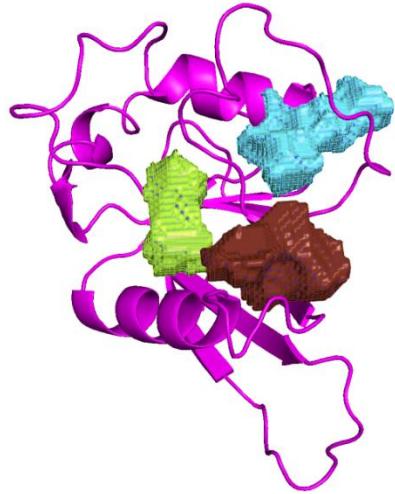
**Fig 3.2.7** Top three binding pockets on the structure of RNF43 predicted by FTSite. The protein structure coloured magenta. The first-ranked binding site coloured pink and the second-ranked pocket coloured green and the third-ranked pocket coloured blue.

DogSiteScorer was also used to test for binding pockets in RNF43. Six pockets were predicted to be druggable; and three sub-pockets located near the interface between R-spondin and RNF43 were of great interest (**Fig 3.2.8**). The binding pocket 3 is equivalent to the green binding site in **Fig 3.2.7**, and the binding pocket

6 is equivalent to the pink pocket in **Fig 3.2.7**. An additional binding pocket, pocket 4, is located in the interface between RNF43 and second β-hairpin of R-spondin 1; furthermore, this binding pocket is closed to binding pocket 3 (green pocket in FTSite result) and 4 (pink pocket in FTSite result) suggesting pocket 6 along with pocket 3 and 4 could be treated as one pocket when screening molecules. The total volume of these three sites is to 697 Å<sup>3</sup>; their site atom count is 192 and total number of hydrophobic interactions is 91. In the following steps, a large pocket composed by these three pockets will be used.

**Table 3.4** Major descriptors of three binding pockets calculated by DoGSiteScorer targeting RNF43 protein.

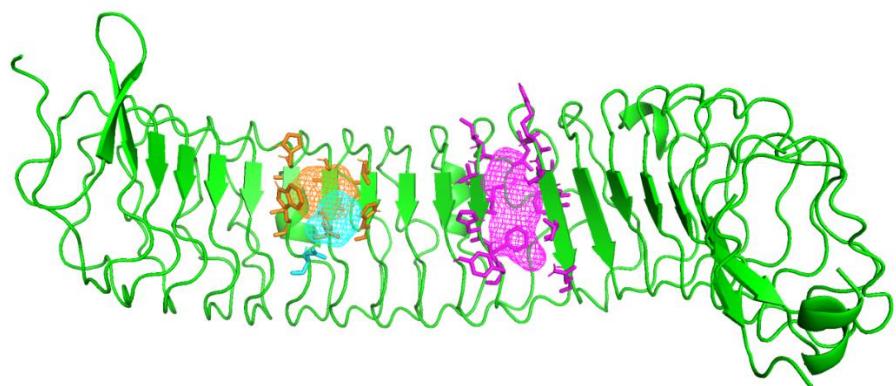
	Pocket3	Pocket4	RSP-06
<b>Volume (Å<sup>3</sup>)</b>	286.4	265.5	145.9
<b>Enclosure</b>	0.23	0.09	0.31
<b>Surface (Å<sup>2</sup>)</b>	610.6	570.4	363.6
<b>Depth (Å)</b>	10.9	11.3	9.8
<b>Surface/volume (1/Å)</b>	2.1	2.1	2.5
<b>Site atoms</b>	67	82	43
<b>Hydrophobic interactions</b>	48	23	20
<b>Hydrophobicity</b>	0.62	0.45	0.53
<b>Simple score</b>	0.21	0.13	0.01
<b>Drug score</b>	0.48	0.48	0.32



**Fig 3.2.8** RNF43 and three binding pockets predicted by DoGSiteScorer. Pocket 3 is coloured brown, Pocket 4 is coloured blue and Pocket 6 is coloured green.

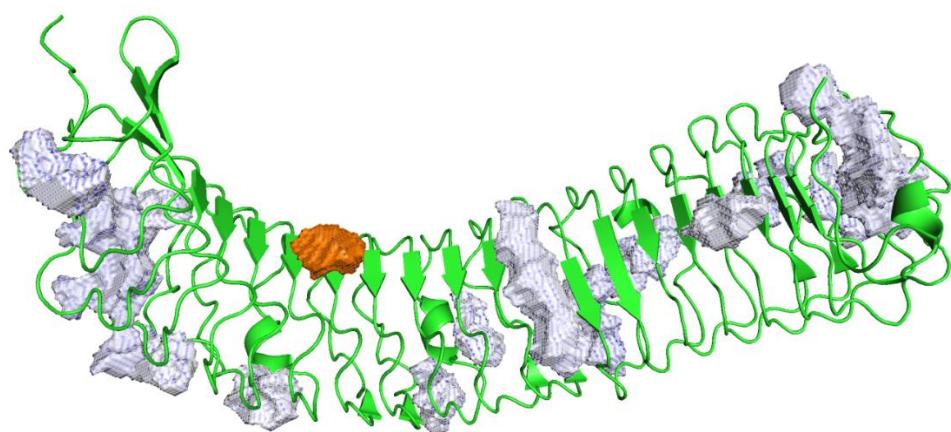
### **3.2.3 LGR5 binding pockets**

FTSite was then used to find pockets on LGR5. The first-ranked (orange) and the third-ranked (cyan) binding pocket were located near the binding interface between R-spondin and LGR5 (**Fig 3.2.9**). These two pockets have some overlap. There were 17 probe molecules in the first binding site and eight residues help to form the binding cavity. Two probe molecules were found in the third binding pocket and four residues helped to form the cavity. These two cavities share some common surrounding residues: Trp168, Thr192 and His216. These two cavities will be considered as one binding pocket for screening further compounds. The second cavity is far from the binding interface between LGR5 and R-spondin, hence, this pocket is not considered.



**Fig 3.2.9** Top three binding pockets on the structure of LGR5 predicted by FTSite. The protein structure coloured green. The first binding site coloured orange, the second pocket coloured magenta and the third pocket coloured cyan.

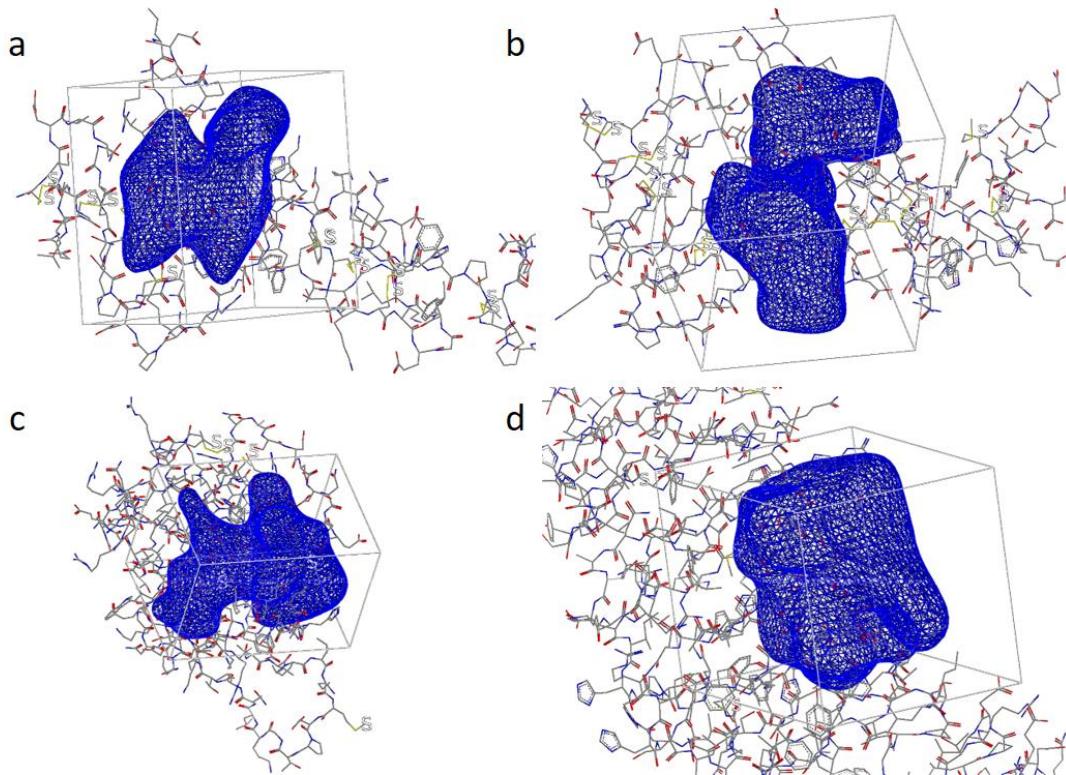
DoGSiteScorer was also used to predict pockets on the structure of LGR5 (**Fig 3.2.10**). Fifteen binding pockets were predicted by DoGSiteScorer; only one pocket was located near the binding interface between R-spondin and LGR5. The pocket ranks fourteenth among the pockets. The volume of the binding site is only  $111 \text{ \AA}^3$  and the surface area is  $224 \text{ \AA}^2$ ; thus the ratio of surface to volume is 2.2, indicating that the size of the binding pocket is small. The simple score of the pocket is 0.05 and the druggability score (drug score) of the pocket is 0.21; both are low scores.



**Fig 3.2.10** LGR5 and all binding pockets predicted by DoGSiteScorer. The binding pocket near the binding interface between LGR5 and R-spondin coloured orange.

This pocket is in the same binding position as the first and third ranked sites from the FTSite software, although here it ranks 14<sup>th</sup> place using DoGSiteScorer. This may be explained by the mechanism of binding site prediction: DogSiteScorer relies heavily on the geometric descriptors of the binding pocket including pocket size and enclosure. A small binding pocket is not likely to get a good score. However, FTSite relies on energy calculations which mean that even small pockets could score well.

Based on above findings, make\_receptor module in OpenEye software was used to prepare protein binding pockets for molecular docking (**Fig 3.2.11**). The volume of the **Closed** R-spondin binding site is 1808 Å<sup>3</sup> and the volume of **Open2** R-spondin is 2213 Å<sup>3</sup> (**Table 4.1.2**). The volume of RNF43 binding site is 2309 Å<sup>3</sup> and the binding site on LGR5 counts to 2268 Å<sup>3</sup>. ChemGauss 4 scoring function was used to rank compounds and only the two highest rated conformations of the top 1000 ranked compounds were recorded during molecular docking.



**Fig 3.2.11** Binding pockets for molecular docking: **a)** binding pocket of Closed R-spondin, **b)** OpenR-spondin, **c)** RNF43, **d)** LGR5. Binding pocket contour represented by blue grid lines.

### 3.3 Molecular docking and protein-ligand interactions

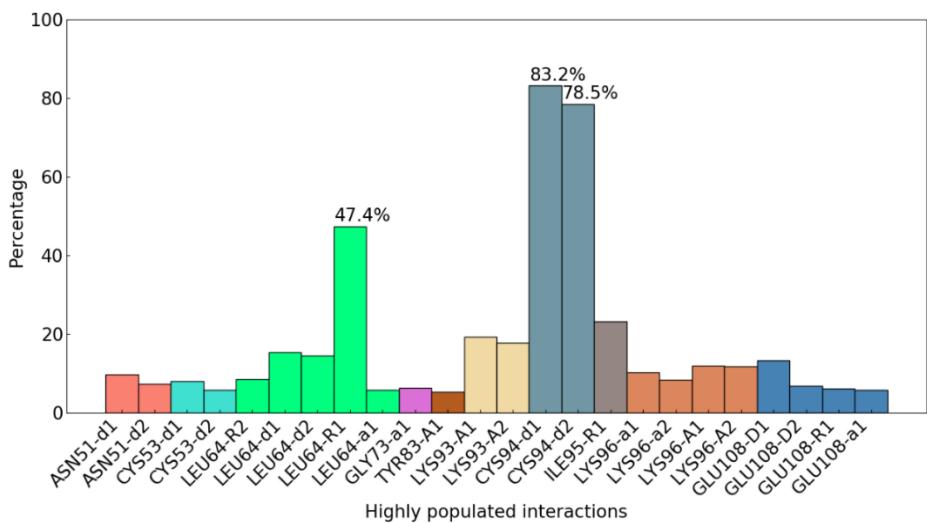
#### 3.3.1 Methods

FRED is a rigid body molecular docking algorithm and does not generate new ligand conformations during the course of molecular docking. Hence, before molecular docking, the Omega module in the OpenEye software package was used to explore the conformational space of compounds. A maximum of six hundred conformations per compound was generated using the modified MMFF94S force field without electrostatics. The energy window was set to 10 kcal/mol and the minimum RMSD of conformers was set to 0.5 Å. Up to four stereocentres per compound could be flipped exhaustively during the conformation enumeration.

ZINC ID of compounds is too long and verbose; hence, selected compounds were renamed according to their time sequence of MD simulation with RSP- as prefix; for example, the RSP-01 is the first selected and simulated compound and it is not related to compound molecular docking score, ranking or target protein etc. Since compound selection were conducted parallelly, compounds belonging to one target might be discontinuous; for example, compounds: RSP-01-RSP-13 targeting **Closed** R-spondin were selected, and RSP-14-RSP-36 were selected targeting RNG43; after finishing molecular dynamics simulation of RSP-01-RSP-13, second round **Closed** R-spondin compound were selected from RSP-37 to RSP-42.

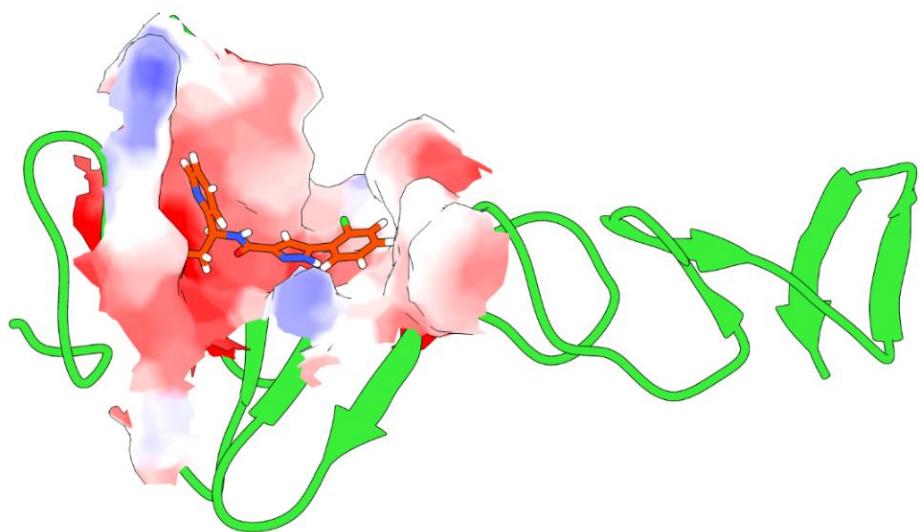
### 3.3.2 Closed R-spondin interactions

20.7 million compounds filtered form **ZINC15 drug-like** databased were docked using FRED to each protein site and ranked by ChemGauss 4 scoring function. From this, the top 1000 compounds were saved. Protein-ligand interaction fingerprints (PLIFs) were then used to summarize good binding features. For **Closed** R-spondin, the highest populated protein-ligand interactions were the hydrogen bond interactions between ligand and Cys94 (**Fig 3.3.1**). Over 78.5% of compounds form two hydrogen bonds with Cys94. Apart from this interaction, the aromatic interaction between ligand and Leu64 also occupies 47.4% of the top 1000 compounds. These findings highlight that there is only one major binding space (**Fig 3.2.2**) between ligand and **Closed** R-spondin.

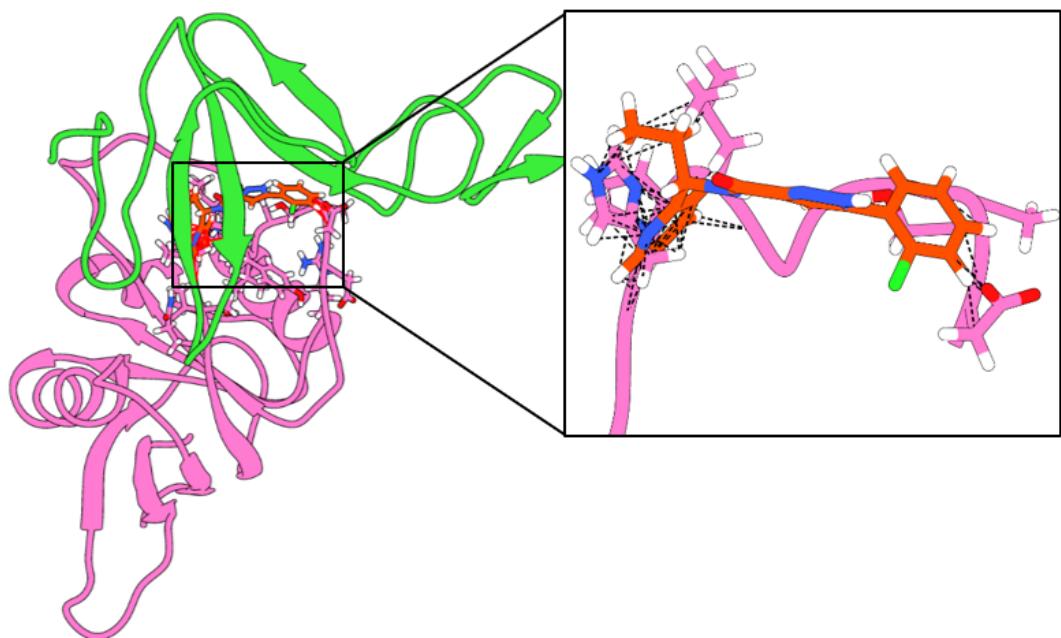


**Fig3.3.1** Highest populated protein-ligand interactions for top 1000 compounds docked into **Closed** R-spondin conformation (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)

As an example of interaction with this major site, docked compound RSP-40 lies in the binding space near the  $\beta$ -hairpin in the Furin-like 1 (**Fig 3.3.2**), which means RSP-40 could potentially disrupt the binding of R-spondin and RNF43. The binding space is formed by the Furin-like 1 and 2 domains and the hinge area between domains. The structure of the RSP-40-**Closed** R-spondin complex was superposed onto the position of **Open** R-spondin in the 4KNG X-ray structure: RSP-40 forms 101 intermolecular clashes with a loop area on RNF43 from His86 to Asp95 (**Fig 3.3.3**). This suggests this compound when bound to R-spondin would prevent it from contacting with RNF43.



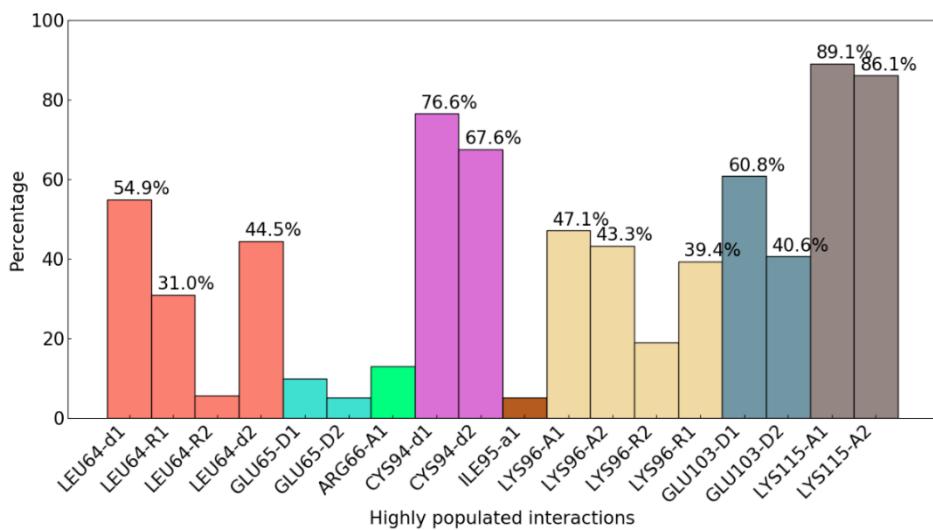
**Fig 3.3.2** Complex of **Closed** R-spondin and RSP-40, representing a major binding pattern. The R-spondin structure is coloured green, RSP-40 is coloured orange and the molecular surface is coloured by electrostatics.



**Fig 3.3.3** Intermolecular clashes between ligand RSP-40 and the R-spondin binding site on RNF43. Molecular clashes between ligand atoms and R-spondin atoms are shown as black dotted lines.

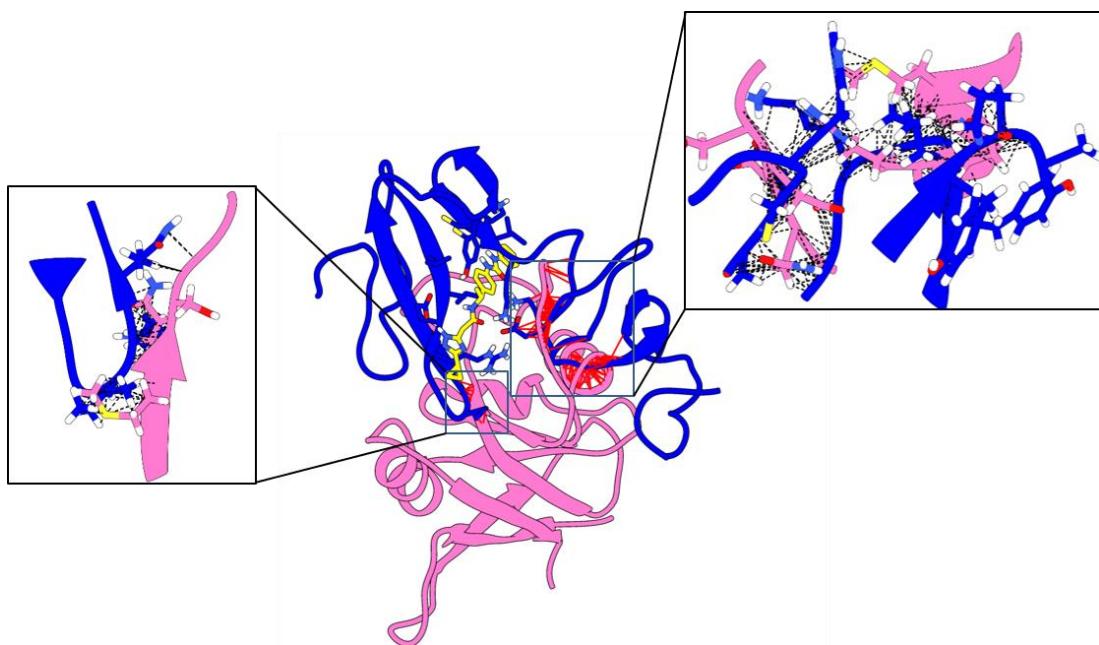
### 3.3.3 *Open2* R-spondin interactions

There are more highly populated protein ligand interactions for compounds docked into ***Open2*** R-spondin. The interaction between Lys115 and compounds was the most populated, with over 89% of compounds forming at least one hydrogen bond with Lys115 (**Fig 3.3.4**). The second most highly populated protein-ligand interaction partner is Cys94 which was also highly populated in ***Closed*** R-spondin PLIFs; this residue could form at least one hydrogen bonds with over 67% of compounds. Another three highly populated interaction partners are Leu64, Lys96 and Glu103. The volume of the ***Open2*** binding site is larger than that of ***Closed*** R-spondin and the diversity and total number of protein-ligand interaction as a consequence is higher than for ***Closed*** R-spondin.



**Fig 3.3.4** Highest populated protein-ligand interactions for top 1000 ranked compounds docked into R-spondin ***Open2*** conformation (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)

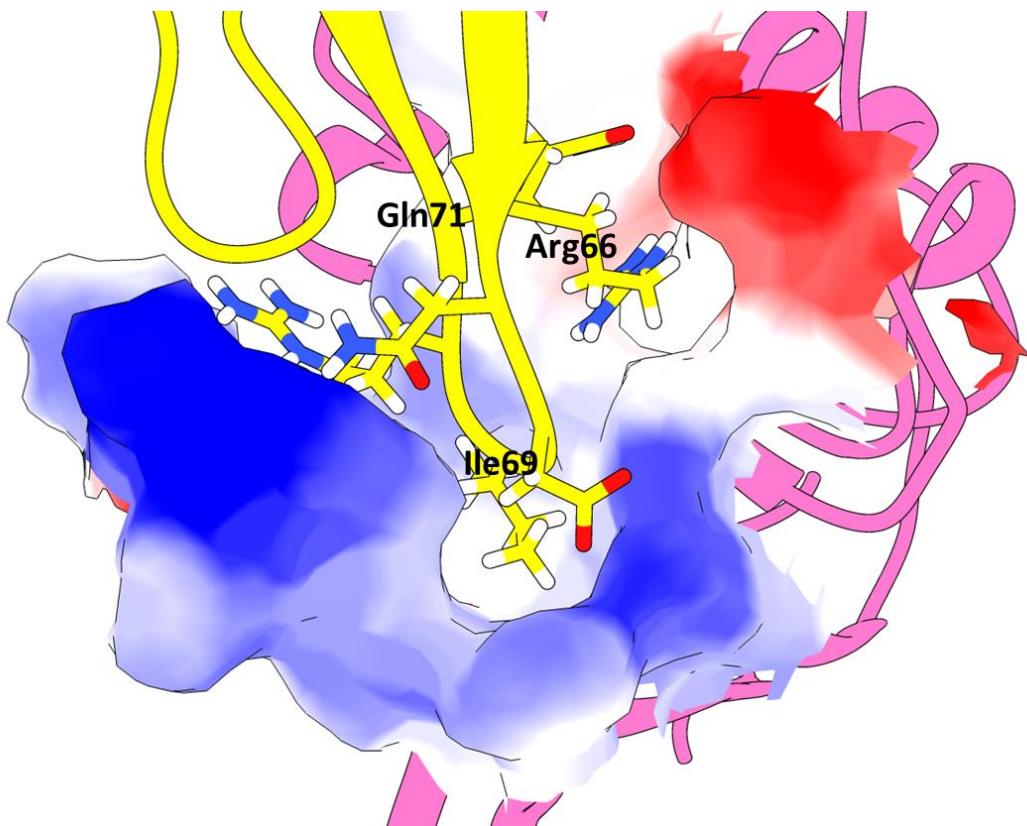
The ***Open2*** conformation of R-spondin was superposed onto the R-spondin binding site of RNF43 (**Fig 3.3.5**). However, two regions demonstrated significant molecular clashes between R-spondin and RNF43 and over 800 molecular clashes were identified in these two regions. The first group of molecular clashes was located in the Furin-like 1 domain of R-spondin, between residues Leu82 - His86 (RNF43) and Asp68 - Gln71 (***Open2*** R-spondin). The second group of molecular clashes was located in the Furin-like 2 domain: residues Asp95 - Asn98 and Met128 - Arg132 in RNF43 overlap the position of residues Ala104 - Glu100, Lys113 - Tyr119 and Tyr126 - Ala128 in ***Open2*** R-spondin. This means that ***Open2*** R-spondin is very unlikely to bind to LGR5. Hence, unlike ***Closed*** R-spondin, a modulator of ***Open2*** R-spondin should bind to ***Open2*** R-spondin such that it prevents recovery to its ***Open*** conformation.



**Fig 3.3.5** Superposition of ***Open2*** R-spondin to the binding site of ***Open*** R-spondin binding to RNF43. R-spondin were coloured blue, RNF43 were coloured pink and RSP-60 was coloured yellow. Molecular clashes were represented by black dotted lines.

### 3.3.4 RNF43 interactions

The surface between the second  $\beta$ -hairpin of R-spondin and RNF43 is shown in **Fig 3.3.6**; there are two cavities on the surface of RNF43. The cavity in an upper location near Arg66 and Gln71 of R-spondin will be called the “upper cavity” in comparison with the “lower cavity” near Ile69 (**Fig 3.3.6**). Subsequent binding sites will be discussed based on this interface between R-spondin and RNF43.

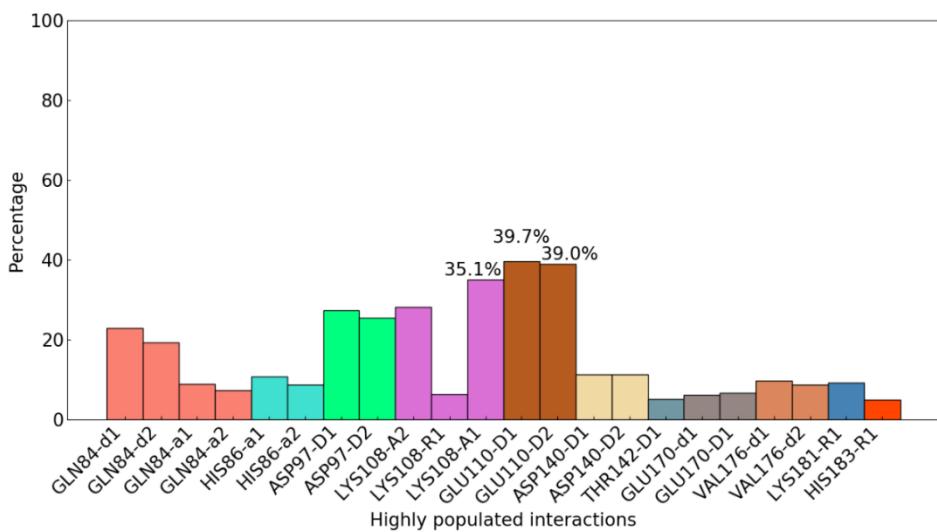


**Fig 3.3.6** Binding interface between R-spondin (yellow) and RNF43 (pink). The interface coloured by electrostatics.

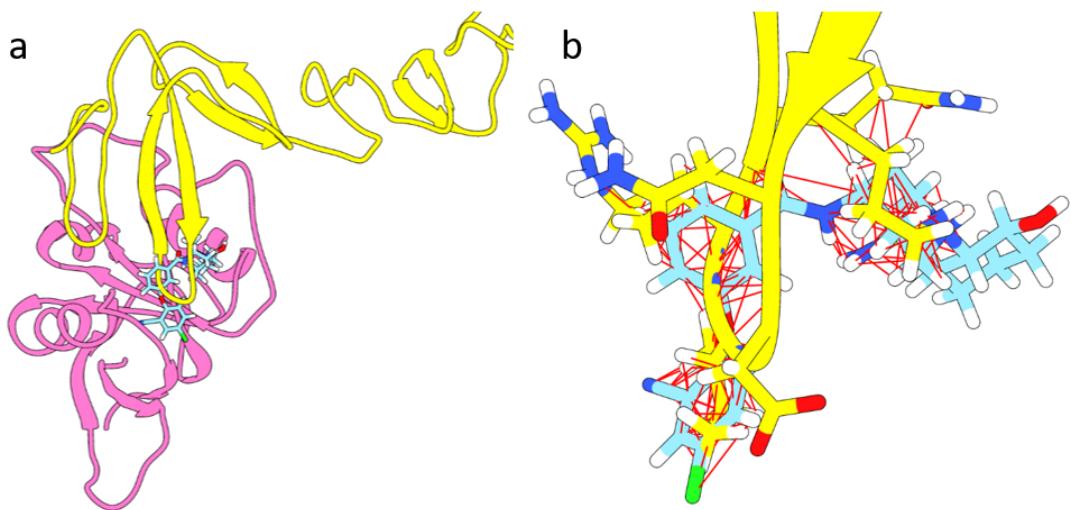
According to the highest populated hydrogen bonds (**Fig 3.3.7**), the populations of hydrogen bonds from docked compounds to RNF43 are relatively low. The maximum occupancy of a hydrogen bond only reached 39.7%. From observation of

the top-ranked molecules, two major binding locations and three major binding patterns were identified.

The first binding location is near the second  $\beta$ -hairpin of R-spondin (**Fig 3.3.6**); two cavities are located in this binding space. These two cavities form the first binding pattern: RSP-22 is a representative compound of this binding pattern and it links the two cavities together. As RSP-20 was superposed onto the binding interface between R-spondin and RNF43 (**Fig 3.3.8 (b)**), there is a large overlap between RSP-22 and R-spondin: 239 molecular clashes were identified which suggests that RSP-22 could disrupt the interaction between RNF43 and R-spondin.



**Fig 3.3.7** Highest populated protein-ligand interactions for top 1000 compounds docked into RNF43 protein. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)

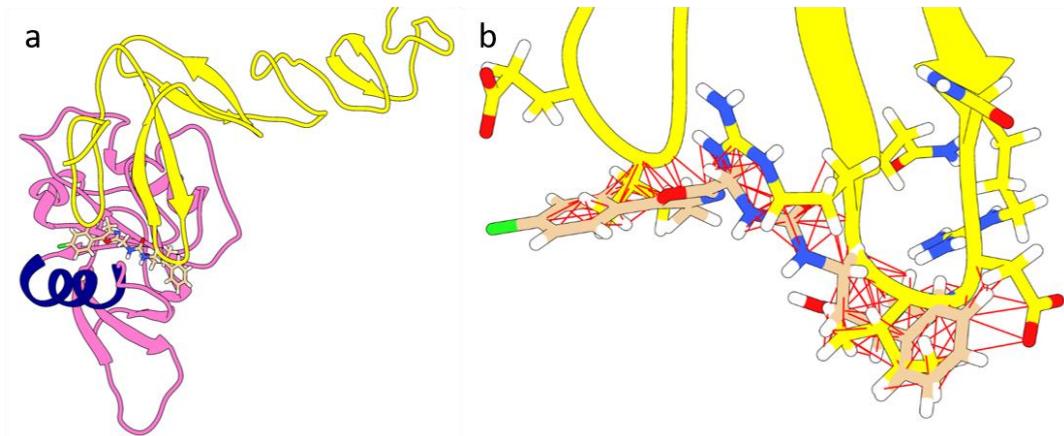


**Fig 3.3.8** **a)** Binding position of RSP-18 on RNF43 in the presence of R-spondin. RSP-18 coloured light blue, R-spondin coloured yellow and RNF43 coloured pink. **b)** Intermolecular clashes between ligand RSP-18 and second  $\beta$ -hairpin on R-spondin-1 represented by red lines; 239 clashes identified between ligand atoms and R-spondin atoms.

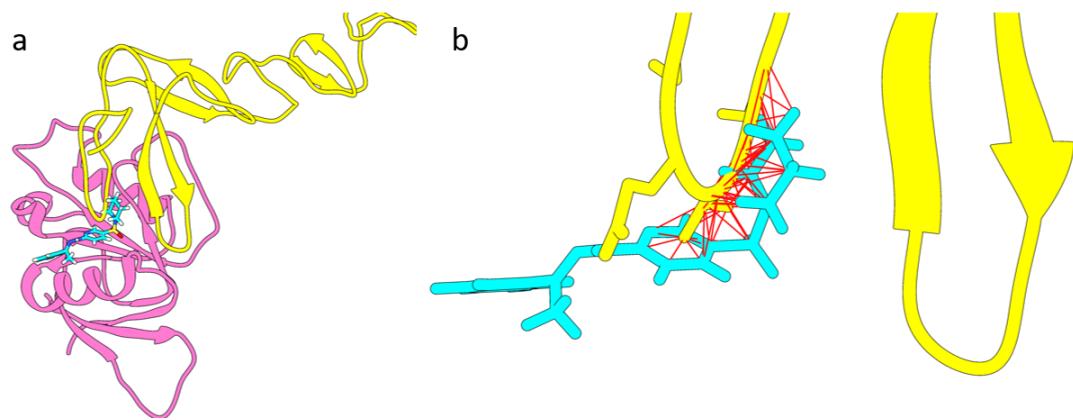
The second binding pattern is also near the lower cavity: an  $\alpha$ -helix from Gly166 to Val176 (dark blue, **Fig 3.3.9 (a)**) could form a groove with the main body of RNF43 allowing compounds to lie in this groove. RSP-22 forms 219 inter-molecular clashes (**Fig 3.3.9 (b)**) with R-spondin after it was superposed onto the binding interface between **Open** R-spondin and RNF43.

Another binding location is shown in **Fig 3.3.10 (a)**, and it is near the binding interface between the first  $\beta$ -hairpin of R-spondin and RNF43. Although nearly half of the top 1000 compounds were located in this binding space, the performance of this binding pattern is not as good as the two binding patterns mentioned before. RSP-32 is a representative compound located in this binding space and it forms only 87 intermolecular clashes (**Fig 3.3.10 (b)**) with R-spondin, even less than half of the

first two binding patterns. Also, without a deep cavity, the third binding pocket is relatively shallow compared with the first and the second binding patterns.



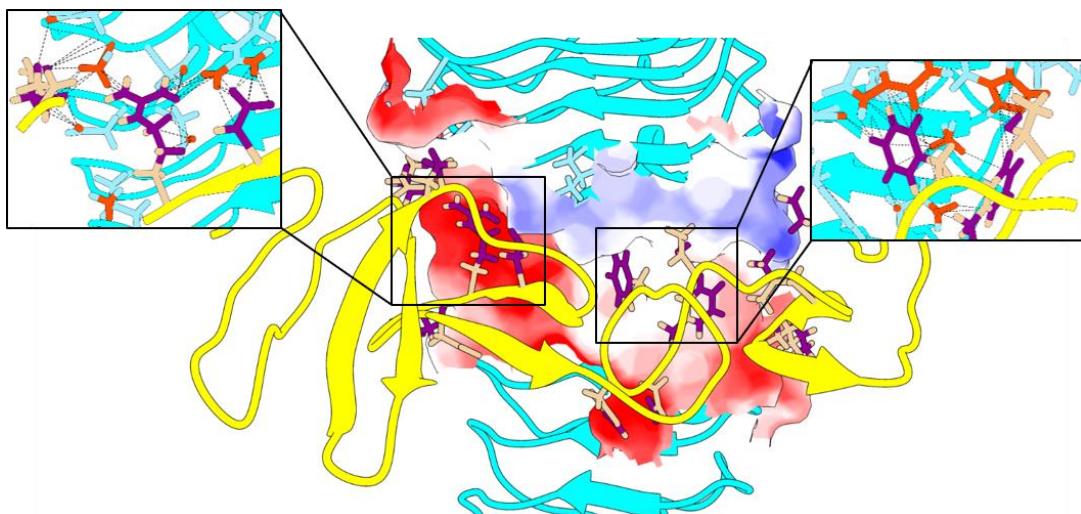
**Fig 3.3.9** a) Binding position of RSP-22 on RNF43 with presence of R-spondin. RSP-22 is coloured light blue, R-spondin is coloured yellow and RNF43 is coloured pink. b) Intermolecular clashes between ligand RSP-22 and two  $\beta$ -hairpin on R-spondin-1 represented by red line; 219 clashes identified between ligand atoms and R-spondin atoms



**Fig 3.3.10** a) Binding position of RSP-32 on RNF43 with presence of R-spondin. RSP-32 coloured light blue, R-spondin coloured yellow and RNF43 coloured pink. b) Intermolecular clashes between ligand RSP-32 and the first  $\beta$ -hairpin on R-spondin-1 represented by red line; 87 clashes identified between ligand atoms and R-spondin atoms.

### 3.3.5 LGR5 interactions

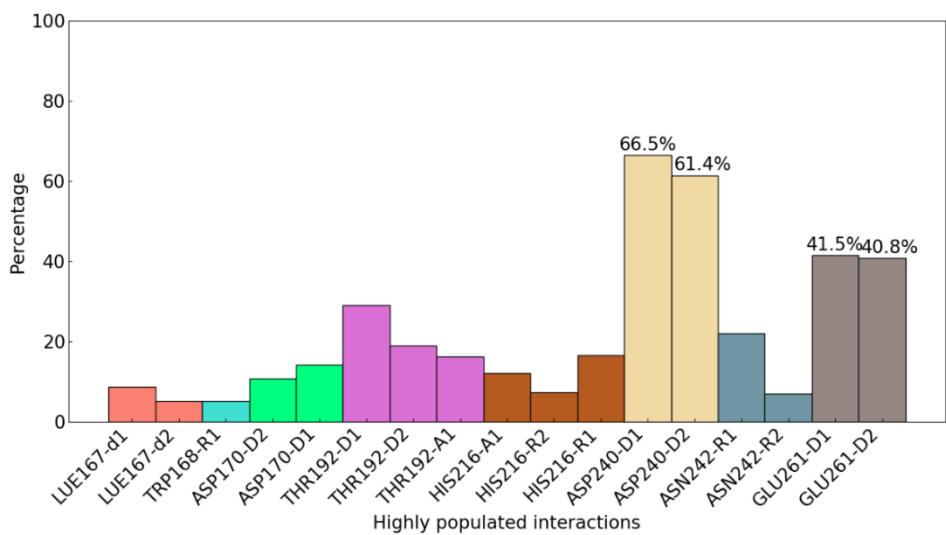
There is a large binding interface between R-spondin and LGR5 (**Fig 3.3.11**), with a solvent accessible surface area of  $1622.8 \text{ \AA}^2$ . Approximately 135 molecular contacts were detected between R-spondin and LGR5. There are two clusters of intermolecular contacts: the first group contains the contacts between Lys59, Pro77, Asp85, Arg87, Asn88 and Pro89 (Furin-like-1 domain of R-spondin) and Met75, Asn123, Arg144, Asp146, Ala147, Asp170, Asp171, Leu195 and Asn219 (LGR5). Lys59, Arg87 (R-spondin) and Arg144, Ala147 (LGR5) formed intensive molecular contact (**Fig 3.3.11**, right sub-pocket) suggesting that these residues are important for ligand binding. The second intensive contact group contains the contacts between Phe106, His108, Asn109, Phe110, Thr112, Leu120, Lys122 and Glu141 (Furin-like-2 domain of **Open** R-spondin) and Gln141, Arg165, His166, Trp168, Gln189, Ala190, Met191, Thr192, Val213, Val214, Ser235, Thr238 and Glu261 (LGR5). Among these contacts, the most significant residue contacts were between Phe106 and Phe110 (R-spondin) and His166, Trp168, Ala190 and Val213 (LGR5).



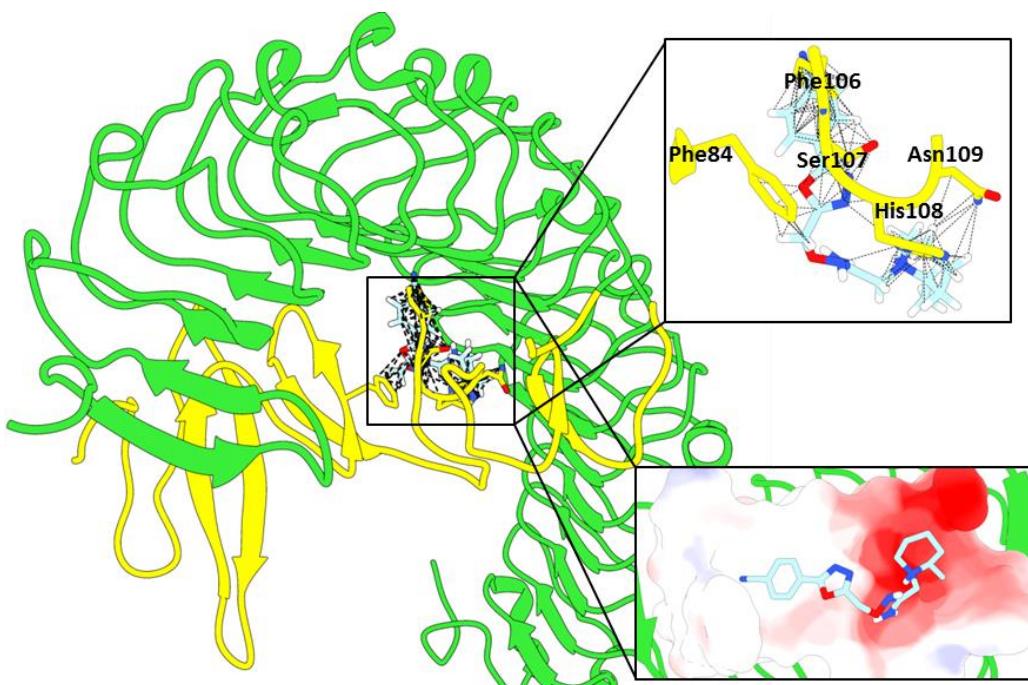
**Fig 3.3.11** Surface contacts between R-spondin and LGR5. Atoms which formed intermolecular contacts coloured purple in R-spondin and orange in LGR5. The protein structure of LGR5 coloured cyan and R-spondin coloured yellow. Atomic contacts represented by black dotted lines.

Based on the binding pocket identification results (**Chapter 3.2.3: LGR5 binding pockets**), the cavity near the second binding contact cluster was considered. After docking molecules to LGR5, the highest populated interaction between LGR5 and compounds is the hydrogen bond with the sidechain of Asp240 (**Fig 3.3.12**); over 61% of compounds formed at least one hydrogen bond with Asp240. The second and the third highest populated interactions were between compounds and Glu261 and Thr192 respectively. Specifically, Thr192 contributed to the surface contact between R-spondin and LGR5; it also formed the cavity which buries Phe106 on R-spondin. Near this cavity, Asp240 helps to form a cavity with Thr238 and Glu261.

RSP-90 is a representative compound that binds to LGR5. When it is superposed onto the binding interface between R-spondin and LGR5, it forms 174 molecular clashes, mainly with Phe84, Phe106, Ser107, His108 and Asn109 of R-spondin (**Fig 3.3.13**). It binds with the two cavities mentioned in the last paragraph and interferes with the interaction between R-spondin and LGR5.



**Fig 3.3.12** Highest populated protein-ligand interactions for top 1000 compounds docked into LGR5 protein. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)



**Fig 3.3.13** Binding position of RSP-90 on LGR5 with presence of R-spondin. RSP-90 coloured light blue, R-spondin coloured yellow and LGR5 coloured green. Intermolecular clashes between RSP-90 and R-spondin-1 marked by black dotted lines; 174 clashes identified between RSP-90 and R-spondin. Binding interface between RSP-90 and LGR5 coloured by electrostatics.

## 4 Molecular dynamics simulation

In this chapter, the main goal is to select compounds that could form stable protein-ligand complex with R-spondin signalling proteins including **Closed** R-spondin, **Open2** R-spondin, RNF43 and LGR5. To achieve this purpose, protein-ligand interaction fingerprints were analysed to build highly populated binding models and select representative compounds. Then, perform molecular dynamics simulation to model the behaviour of different solvated protein-ligand complexes and evaluate their binding stabilities. Following this, structural optimization based on hydrogen bond interaction, key interactions etc.

In the last section, molecular docking had been used to predict the binding poses of 20.7 million compounds. Compounds tested in this section are all selected from the top 1000 ranked compounds among which 93 compounds were tested by molecular dynamics simulation.

**Table 4.1.1** Protein models and number of simulated compounds for each protein structure.

Protein models	Number of ligands
<b>Closed</b> R-spondin	25
<b>Open2</b> R-spondin	21
RNF43	31
LGR-5	16

## 4.1 Methods

### 4.1.1 Model construction

This work is based on molecular dynamics simulations; this was performed by *PMEMD.cuda* module in AMBER 16<sup>61</sup> with the ff14SB force field<sup>62</sup> and General AMBER force field (GAFF)<sup>63</sup> to simulate protein and ligand respectively. Calculations were accelerated by a NVIDIA V100 graphics processing unit.

The X-ray structure of LGR5 and RNF43 was obtained from Protein Data Bank (PDB) using PDB code 4KNG (resolution 2.5 Å)<sup>30</sup>. Residues between Phe482 and His537 were missing in the crystal structure of LGR5 which is located in the C-terminus of LGR5. As the C-terminus is far from the R-spondin binding site, residues after His537 were removed. Hydrogen atoms and missing residues in LGR5 and RNF43 were added and corrected by MOE<sup>64</sup>. The structure of **Closed** and **Open2** conformations of R-spondin 1 were generated from microsecond scale molecular dynamics simulations of R-spondin 1<sup>65</sup>.

Molecular charges of compounds were calculated by antechamber module with the AM1-BCC charge model<sup>66</sup>. Systems were solvated in TIP3P water<sup>67</sup> and further details of systems simulated are given in **Table 4.1.2**. Chloride ions and sodium ions were used to neutralize the systems depending on the charge of compounds. The TLEaP module in AmberTools was used to generate AMBER topology and coordinate files.

**Table 4.1.2** Basic information of four simulation systems

Protein	Average volume of box ( $\text{\AA}^3$ )	Number of residues	Number of Added water molecules	Binding pocket volume ( $\text{\AA}^3$ )
<b>Closed</b> R-spondin	466189	104	13463	1808
<b>Open2</b> R-spondin	359855	104	10071	2213
RNF43	382729	151	10490	2309
LGR5	1419934	453	41143	2268

#### **4.1.2 Molecular dynamics protocol**

Two minimization processes with and then without 500 kcal/(mol  $\text{\AA}^2$ ) positional restraints on the protein structure were performed to relax the protein system and water molecules. Non-bounded cutoff was set to 8.0  $\text{\AA}$ . After two minimization processes, the system was heated from 0 K to 300 K with the heating speed of 150 K/20 ps. The SHAKE algorithm was applied to constrain bond lengths. Temperature scaling methodology applies Langevin dynamics with collision frequency set to 2.0  $\text{ps}^{-1}$ . Protein and ligands were restrained by 25 kcal/(mol  $\text{\AA}^2$ ) harmonic potential during the course of heating process. Six equilibration steps were performed with the 25 kcal/(mol  $\text{\AA}^2$ ) harmonic potential gradually reduced to zero. Each equilibration process lasted 50 ps. Production of 100 ns for each the protein-ligand system was performed at 300 K and at 1 bar.

#### **4.1.3 Compounds selection and optimization**

To obtain the best predicted compounds for binding with R-spondin signalling proteins, multiple rounds of simulations and ligand optimizations were performed.

Compounds selection and ligand optimization followed the following steps. Firstly, highly populated protein-ligand interactions were identified, and major binding modes were identified based on their interactions. Then, representative compounds from different binding patterns were selected, while also seeking to keep structural diversity. Subsequently, the protein-ligand complexes were simulated, and their binding stabilities were evaluated. After analysing the main features of computationally successful compounds, other compounds were selected that may improve their binding stabilities further; these were then simulated and their binding affinities computed. Multiple cycles of ligand optimization were followed by to seek to improve the performance of compounds.

#### ***4.1.4 Analysis tools***

The trajectory generated from the simulation is quantitatively analysed by CPPTRAJ<sup>68</sup> module of Amber 16. MOE<sup>64</sup> and UCSF Chimera<sup>58</sup> were used to generate graphical representation of molecular structures and analysis protein-ligand complex. The visual inspection of molecular trajectory is processed by VMD<sup>69</sup>.

Four numeral descriptors were calculated to quantify the R-spondin 1 binding performance of compounds. First of all, ligand positional shift was measured by the distance between centre of mass (COM) of ligand and COM of residues within 4.5 Å; this roughly characterizes the distance from the supposed binding site. Large positional shifts could directly reflect whether a ligand binds with protein or dissociates. The standard deviation of ligand positional shift was used to determine whether a compound could bind stably with protein in a single stable location. This

descriptor is useful when encountering ligands that explored an unexpected binding site, whose ligand positional shift is large but the binding mode is stable. In some cases, ligands stay in their binding sites; however, their poses keep changing during the simulation. The RMSF of ligand atoms is a useful descriptor to eliminate this false positive condition. The RMSD of ligand atoms was also calculated to reflect whether ligand has experienced a significant conformational change compared with its original binding pose.

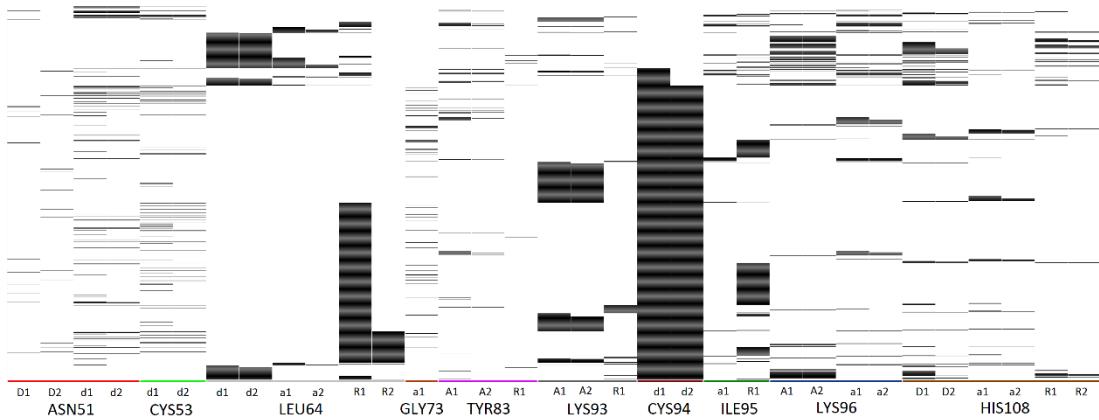
## 4.2 Results: *Closed* R-spondin ligands

### 4.2.1 Compound selection

According to highly populated protein-ligand interactions shown in **Fig 3.3.1** (**Chapter 3.3.2: *Closed* R-spondin interactions**), the most significant hydrogen bond interaction partner was Cys94. Over 83.2% of compounds formed at least one hydrogen bond with Cys94. The second most important interaction partner was Leu64, which formed aromatic interactions with 47.4% of compounds. Leu64 could also function as hydrogen bond acceptor to form hydrogen bonds with over 15% of compounds.

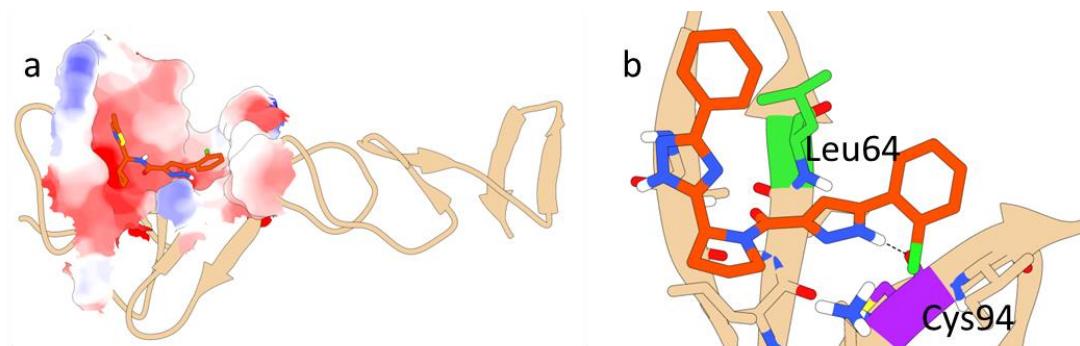
The protein-ligand interaction fingerprint of *Closed* R-spondin is presented in **Fig 4.2.1**. As these two residues are the most important partner of compounds, an interesting correlation between Cys94 and Leu64 was observed. Except for a few compounds, a majority of Leu64-based hydrogen bonds does not form hydrogen bonds with Cys94; however, there is a large overlap between Leu64 based aromatic interactions and a Cys94 hydrogen bond interaction. Based on the correlation

between Leu64 interactions and Cys94 interactions, three key binding modes were identified to facilitate compound selection.



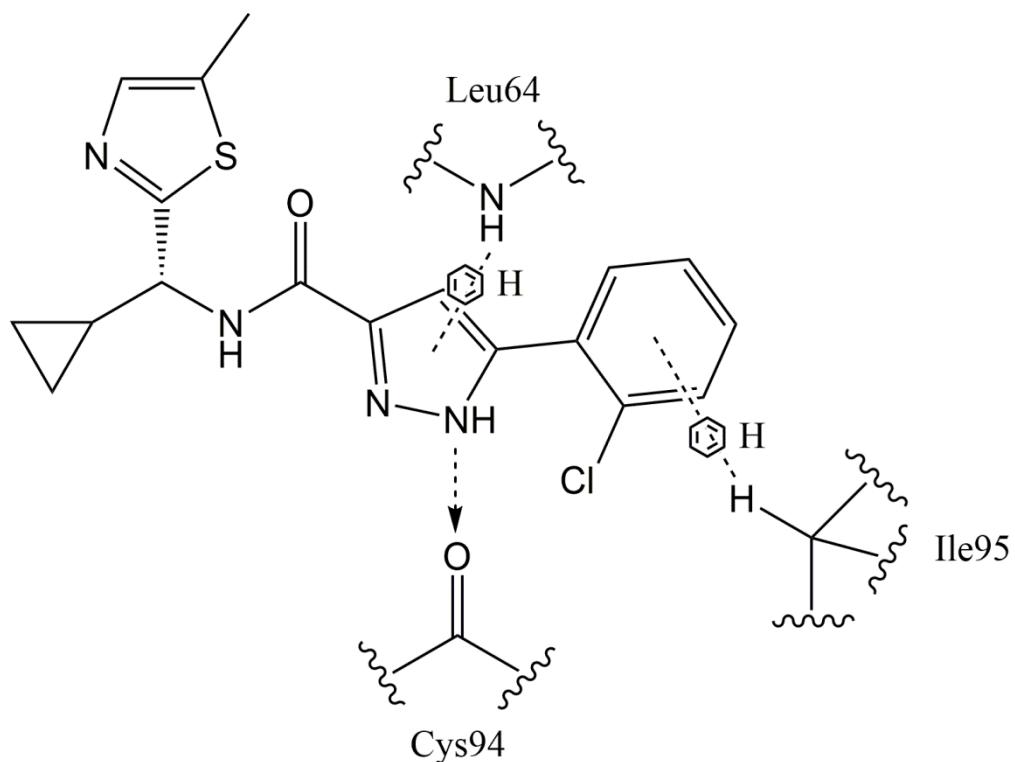
**Fig 4.2.1** Protein-ligand interaction fingerprint of top 1000 compounds targeting **Closed** R-spondin. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)

As is shown in **Fig 4.2.2**, the binding site is surrounded by two  $\beta$ -hairpin structures in the Furin-like-1 domain and a corner of Furin-like-2 domain and the hinge area of R-spondin. Compared with typical drug binding sites, the binding site on **Closed** R-spondin is  $1808 \text{ \AA}^3$ , relatively shallower than other proteins (**Table 4.1.2**).



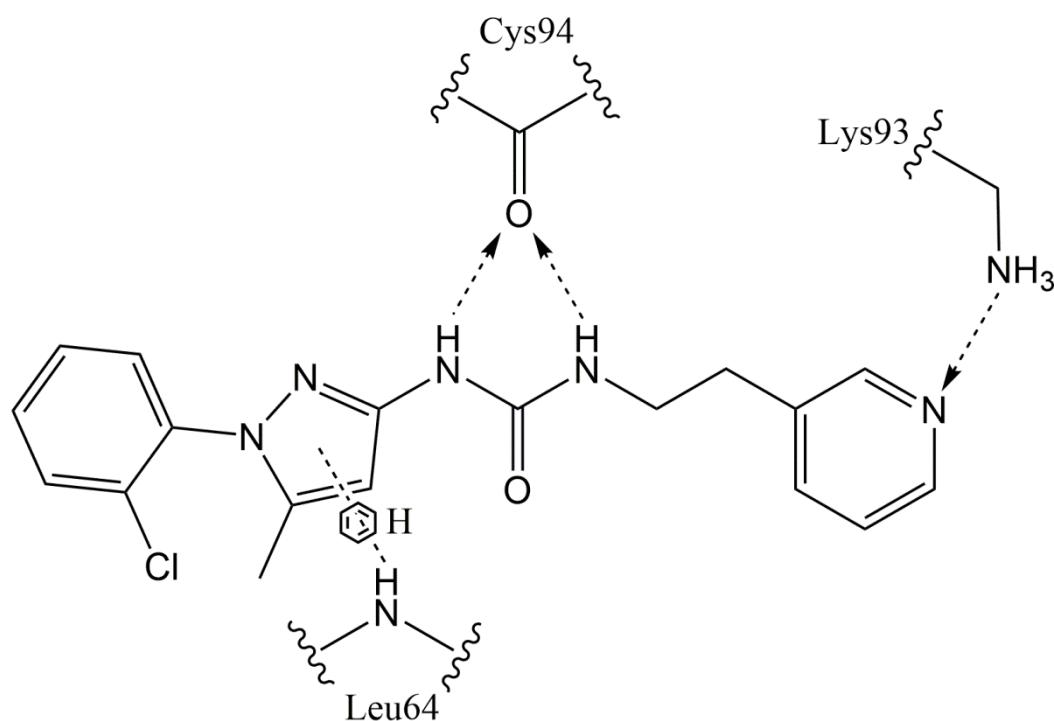
**Fig 4.2.2** a) Molecular interface between RSP-05 binding with **Closed** R-spondin coloured by electrostatics. b) Graphical representation of RSP-05 binding site on **Closed** R-spondin, proteins structure coloured tan and RSP-05 coloured orange.

Binding mode-1 of compounds was shown in **Fig 4.2.3**. The structures of binding mode-1 compounds often contain a two- or three-nitrogen five-member ring structure to form a hydrogen bond interaction with the carbonyl group on the backbone of Cys94. Often the ring structures were perpendicular to the backbone nitrogen atom of Leu64 which helped protein and ligand form a  $\pi$ -H bond at the same time (**Fig 4.2.2 (b)**). This phenomenon may explain why Cys94 demonstrated a significant correlation with Leu64.



**Fig 4.2.3** Schematic 2D representation of binding mode 1: RSP-05 with **Closed** R-spondin.

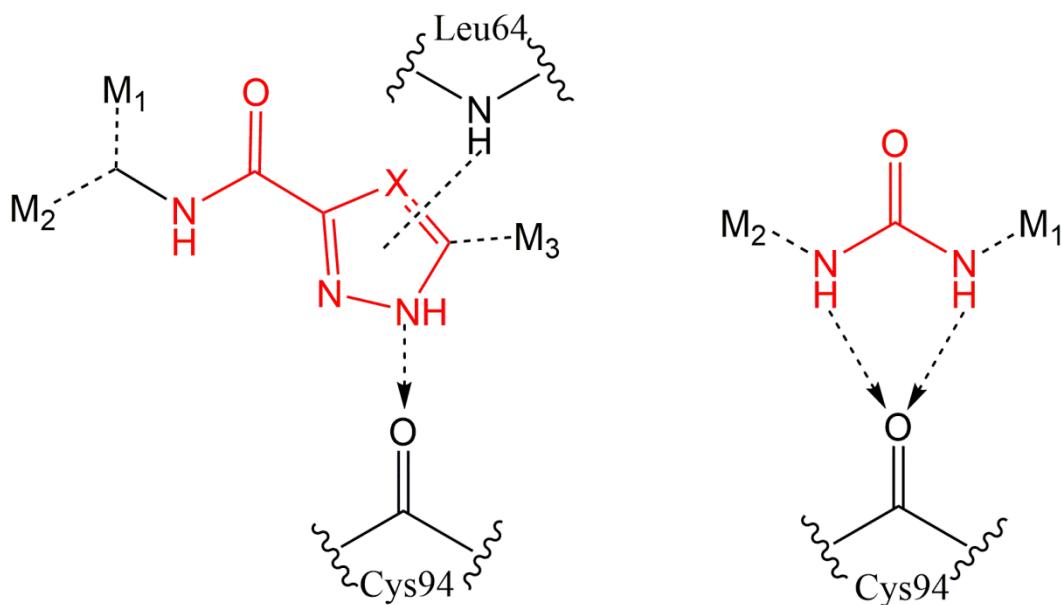
Like binding mode-1, binding mode-2 (**Fig 4.2.4**) is also based on the hydrogen bond interactions between compounds and Cys94, however, in another form. Unlike a heterocyclic ligand structure making hydrogen bond interactions, binding mode 2 compounds usually form two hydrogen bonds with the carbonyl group of Cys94 with a urea group. As a consequence, binding mode 2 compounds usually have a long backbone.



**Fig 4.2.4** Schematic 2D representation of the binding mode 2: RSP-09 with **Closed** R-spondin.

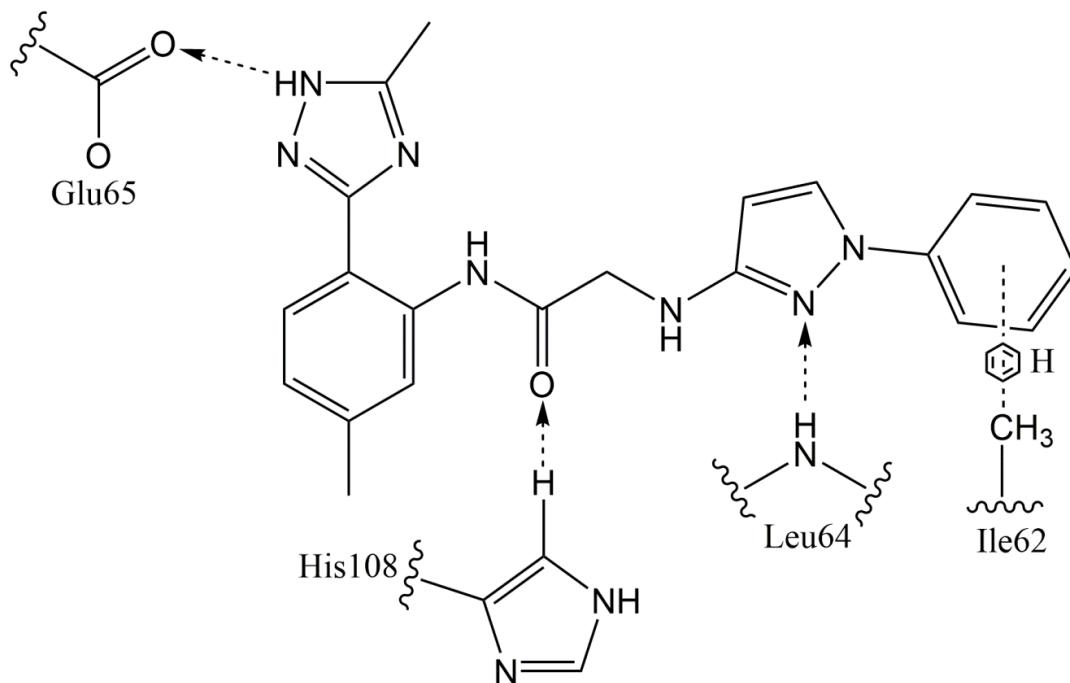
The core structures of binding mode 1 and binding mode 2 are compared in **Fig 4.2.5**.

**4.2.5.** The core structure of the binding mode 1 compounds is a five-membered ring structure and an amide group. Three moieties were also linked to the core structure of binding mode 1 including two hydrophobic moieties linked to the amide group and one ring moiety linked to the five-member ring. Also, most of the selected compounds form aromatic interaction with Leu64 at the same time. For binding mode-2, a urea group is often observed to form two hydrogen bond interactions with the backbone oxygen of Cys94.



**Fig 4.2.5** Core structures of binding mode 1 and 2 of *Closed* R-spondin compounds. M<sub>1</sub>, M<sub>2</sub> and M<sub>3</sub> were used to represent moieties that linked to core structures. Hydrogen bonds were represented by dotted arrow.

The third binding mode (**Fig 4.2.6**) is based on the hydrogen bond interactions between compounds and Leu64. As is seen on the protein-ligand fingerprint, if compounds form hydrogen bonds with Leu64, they are not likely to bind with Cys94. It prefers to form other less highly populated interactions e.g. with His108, Ile62 or Glu65.



**Fig 4.2.6** Schematic 2D representation of the binding mode 3: RSP-06 with **Closed** R-spondin.

Compared with binding mode-1, the two other binding modes are of less frequency; hence, more of the binding mode-1 compounds were selected for further study. In the first round of simulations, 13 compounds were selected (**Table 4.2.1**). Of these, seven compounds (RSP-02, RSP-04, RSP-05, RSP-07, RSP-10, RSP-12 and RSP-13) belong to binding mode 1. Three compounds (RSP-08, RSP-09 and RSP-11) belong to binding mode 2. Another three compounds (RSP-01, RSP-03, RSP-06) belong to binding mode 3.

**Table 4.2.1** Name and ZINC ID of the thirteen selected *Closed* R-spondin ligands and their corresponding molecular docking scores and ranking among the top 1000 compounds.

Ligand name	Compound Zinc ID	Molecular docking score	Ranking	Binding mode
RSP-01	ZINC001459863801	-11.8	22	3
RSP-02	ZINC000795474930	-13.4	1	1
RSP-03	ZINC000486902815	-12.8	2	3
RSP-04	ZINC000089851301	-11.8	25	1
RSP-05	ZINC000461235917	-11.8	30	1
RSP-06	ZINC000486427111	-11.7	37	3
RSP-07	ZINC000436854920	-11.5	65	1
RSP-08	ZINC001214704385	-11.5	67	2
RSP-09	ZINC000180133595	-11.1	213	2
RSP-10	ZINC001156941895	-11.1	185	1
RSP-11	ZINC000756325488	-11.4	79	2
RSP-12	ZINC000564207284	-11.2	129	1
RSP-13	ZINC001643536612	-11.6	46	1

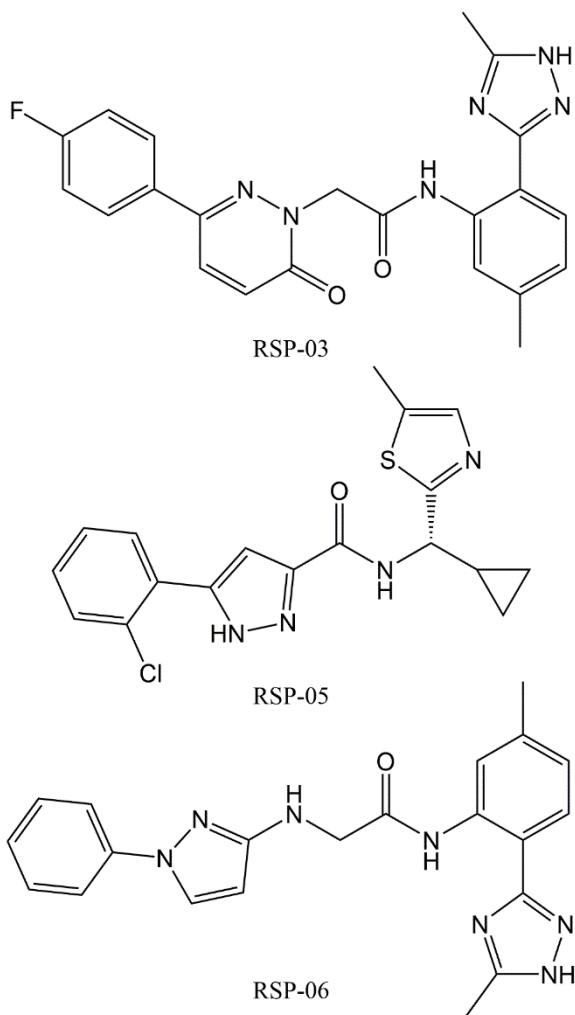
#### **4.2.2 Ligand Binding stabilities**

MD simulations of the 13 selected compounds were then performed for 100 ns; from the trajectories, protein-ligand binding descriptors were computed, including the RMSF and RMSD of ligand atoms, ligand positional shift as well as its standard deviation (**Table 4.2.2**).

**Table 4.2.2** Ligand binding stability descriptors of thirteen **Closed** R-spondin ligands in the first-round selection.

Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
<b>RSP-01</b>	2.5	4.4	5.7	1.5
<b>RSP-02</b>	2.1	7.5	14.3	4.6
<b>RSP-03</b>	1.7	2.7	4.3	0.5
<b>RSP-04</b>	1.8	1.5	3.3	0.6
<b>RSP-05</b>	1.0	2.3	3.4	0.4
<b>RSP-06</b>	2.0	2.1	5.1	1.2
<b>RSP-07</b>	2.3	3.4	3.6	1.8
<b>RSP-08</b>	2.9	6.4	8.8	3.1
<b>RSP-09</b>	2.4	3.7	5.7	0.9
<b>RSP-10</b>	1.8	5.9	6.8	3.0
<b>RSP-11</b>	3.0	26.4	20.8	13.8
<b>RSP-12</b>	2.2	32.8	29.3	14.4
<b>RSP-13</b>	2.3	13.0	22.8	8.9

Based on performance of binding stabilities, the 13 simulated compounds were divided into four groups. The first group includes RSP-03, RSP-05 and RSP-06, which demonstrated good binding stability. The structures of these three compounds are shown **Fig 4.2.7**. Their RMSF values ranged from 2.1 Å to 2.7 Å and their ligand positional shift ranged from 3.4 Å to 5.1 Å showing that these compounds could bind with a location during the simulations.



**Fig 4.2.7** The structures of RSP-03, RSP-05 and RSP-06 which are predicted from MD to form a stable protein-ligand complex with **Closed** R-spondin.

RSP-05 is one of the most typical binding mode-1 compounds (**Fig 4.2.7**). With the five-membered ring and amide group as core structure, a chlorobenzene group is

linked to the pyrazoline group and methyl-thiazole ring and a cyclopropane ring are connected to the amide group joined by a methylene group.

The highest populated hydrogen bonds over 10% time of simulation are listed in

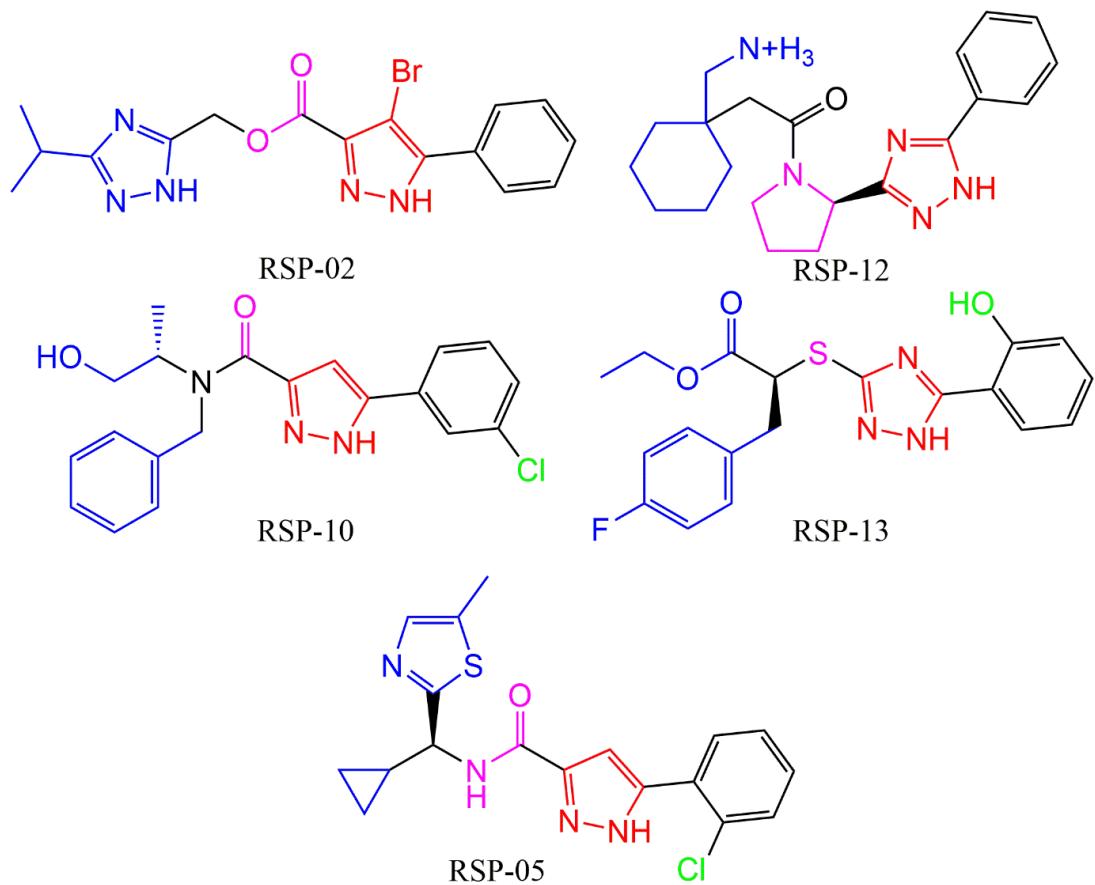
**Table 4.2.3.** RSP-05 formed hydrogen bonds with Cys94 for 83% of the 100 ns simulation, which is in accordance with binding mode-1. RSP-03 and RSP-06 belong to the binding mode 3, however; molecular docking failed to predict the most important hydrogen bond partner for these compounds, Glu65. RSP-07 forms hydrogen bonds with Glu65 for 76% of time (**Table 4.2.3**). However, the occupancy of the RSP-03 was only 20% which means that hydrophobic interactions may play a crucial role in the protein-ligand binding.

**Table 4.2.3** Highly populated hydrogen bond interaction in RSP-03, RSP-05 and RSP-06 MD simulation.

Ligand name	Acceptor	Donor	Fraction
RSP-03	E65 <sup>OE</sup>	RSP-03 <sup>N3</sup>	20%
	C94 <sup>O</sup>	RSP-05 <sup>N2</sup>	83%
	RSP-05 <sup>O</sup>	K93 <sup>NZ</sup>	12%
RSP-05	E65 <sup>OE</sup>	RSP-06 <sup>N3</sup>	76%
	RSP-06 <sup>O</sup>	R66 <sup>NE</sup>	36%
	C94 <sup>O</sup>	RSP-06 <sup>N6</sup>	14%
RSP-06	RSP-06 <sup>O</sup>	Q71 <sup>NE2</sup>	12%
	Q71 <sup>OE1</sup>	RSP-06 <sup>N6</sup>	11%

The next group demonstrated intermediate binding stability. RSP-01, RSP-07 and RSP-09 did not form very stable complexes with **Closed** R-spondin; however, they fluctuated within a tolerable distance from the initial binding site. Their RMSF values are higher than the three compounds mentioned previously, ranging from 3.4 Å to 4.4 Å. They also shifted from the original binding site by 3.6 Å to 5.7 Å.

The third group contains compounds that failed to pass the molecular dynamics assay and fell off the cavity on **Closed** R-spondin. These six compounds are RSP-02, RSP-08, RSP-10, RSP-11, RSP-12 and RSP-13 whose ligand positional shifts were higher than 5.6 Å showing that they could not bind with the **Closed** R-spondin binding site. Among these compounds, four compounds (RSP-02, RSP-10, RSP-12, RSP-13) are binding mode-1 compounds. However, it is worth noticing that RSP-05 (binding mode-1) was the best compound among the three stable binding compounds. This means that the range of performance of binding mode 1 compounds is broad and their performance could be improved through structure optimization. Hence, in the following, ligand optimization will focus on binding mode-1 optimization. Following are some major differences that may affect the performance of compounds (**Fig 4.2.8**): the number of nitrogen atoms in the anchoring five-membered ring structure; if there are any halogen atoms in the anchor ring; the structure of linker between lefthand and righthand moieties; if there are any branches on the benzene ring in the righthand moiety; the number of branches in the lefthand moiety. In the following ligand optimization process, a few tests of these features were explored to optimize the structure of binding mode-1 compounds.



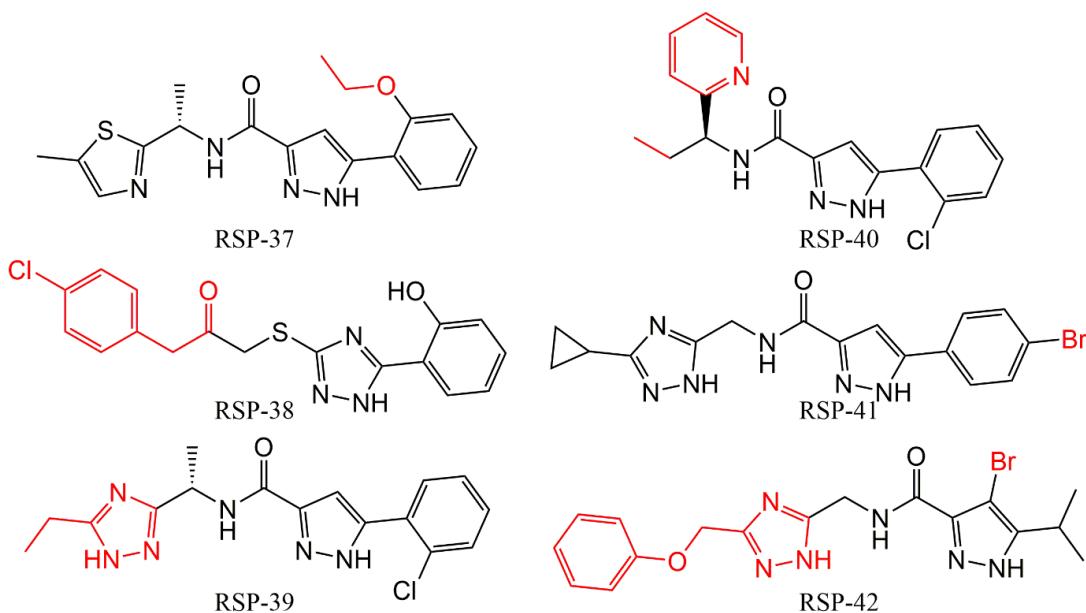
**Fig 4.2.8** Comparison between RSP-05 (good binding stability in MD) and RSP-02, RSP-10, RSP-12 and RSP-13 (poor binding stability in MD). The anchoring five-membered rings coloured red. Moiety linkers coloured magenta. Benzene ring branches of the righthand moieties coloured green. Structures of the lefthand moieties coloured blue.

The last group contains special cases observed from the MD simulations. Two compounds including RSP-04 and RSP-06 will be discussed in the following **Special Cases** section.

#### 4.2.3 Ligand structure refinement

In the first MD simulation round of 13 compounds, binding mode 1 was found to have greater potential for structural optimization. To test some features of different moieties of binding mode 1 compounds, the following compounds shown

in **Fig 4.2.9** were selected based on their structural features. These six compounds were selected based on the structure of RSP-05 to determine which kind of change is favoured for binding stability. For example: in RSP-37, an ethoxy group took the place of the chloride atoms in RSP-05 to test the effect of branches on the benzene ring in the righthand moiety. A prolonged lefthand moiety in RSP-38 and RSP-41 were used to test if a longer lefthand moiety could facilitate protein-ligand binding. RSP-40 has a changed ring structure in the lefthand moiety. RSP-41 has a changed position of the halogen atom.



**Fig 4.2.9** Compounds selected in the first round of structural optimization targeting **Closed** R-spondin.

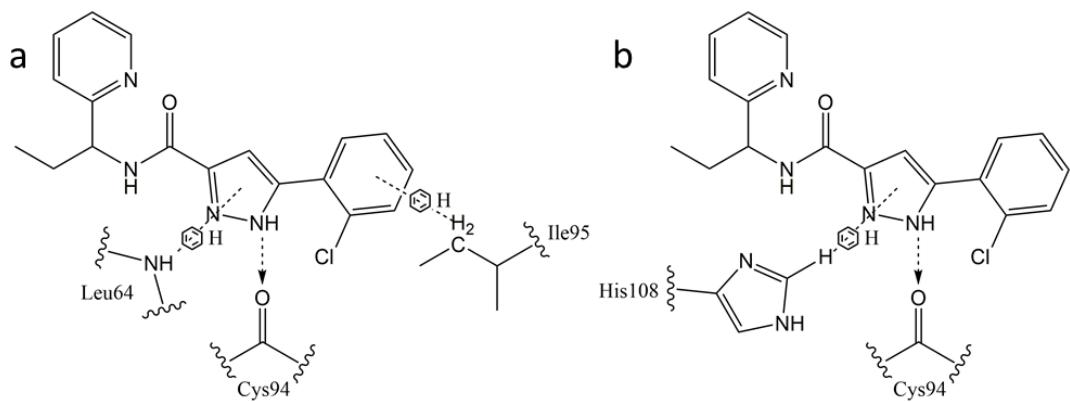
According to **Table 4.2.4**, apart from RSP-40, five of these compounds did not bind stably with **Closed** R-spondin in 100 ns MD simulation of the complex, showing high ligand positional shifts. The binding pattern of RSP-40 is shown in **Fig 4.2.10**; RSP-40 shares a very similar structure with RSP-05 and both compounds form high

occupancy hydrogen bonds with Cys94. RSP-40 forms hydrogen bonds with Cys94 for 86% of the simulation time.

**Table 4.2.4** Binding stability descriptors of *Closed* R-spondin ligands in the third-round selection.

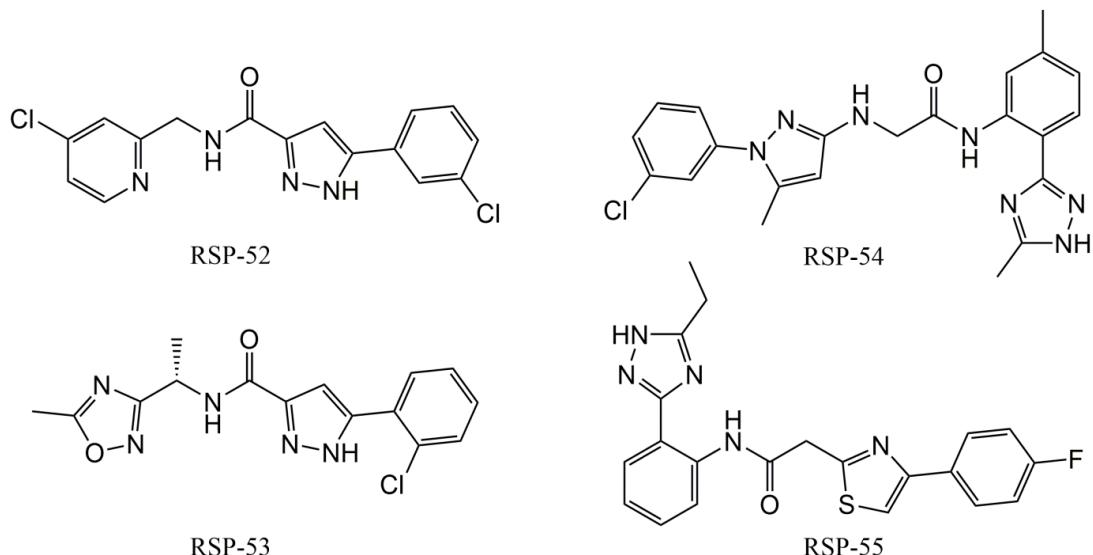
Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
<b>RSP-37</b>	1.7	4.6	5.6	1.6
<b>RSP-38</b>	2.3	3.4	8.8	1.8
<b>RSP-39</b>	1.7	6.4	8.5	4.1
<b>RSP-40</b>	2.2	1.1	2.8	0.4
<b>RSP-41</b>	2.3	19.0	16.1	13.4
<b>RSP-42</b>	2.4	7.0	6.9	2.9

Six compounds were selected in the second selection of compound, however, only RSP-40 demonstrated good binding stability. Among all binding mode 1 compounds, only RSP-05 and RSP-04 demonstrated good binding properties; also, they share several similarities: they have benzyl chloride group as the righthand moiety; 1*H*-pyrazole-3-carboxylamide group the as anchor system. Specifically, two short hydrophobic groups are observed on the lefthand moiety of RSP-05 and RSP-40. In contrast to the RSP-39 which has a long ethyl-triazole group, short hydrophobic groups in the lefthand moiety were favoured. For the anchor ring system, all triazole-based and halogenic diazole binding mode 1 compounds failed in the MD simulations e.g. RSP-13, RSP-02 and RSP-42; which suggested that triazole and halogenic diazole ring systems were not favoured.



**Fig 4.2.10** a) Initial and b) final binding pose of RSP-40. Hydrogen bond represented by doted arrow.

With the features summarized from the first structural optimization, three binding mode-1 compounds (RSP-51, RSP-52 and RSP-53) were selected. Also, two binding mode-3 compounds (RSP-55, 56) and one special compound, RSP-54, were selected. The binding descriptors of the six compounds are listed in **Table 4.2.5**. Four compounds: RSP-52, RSP-53, RSP-54 and RSP-55 were considered to bind well with **Closed** R-spondin, and the structures of these four compounds are given in **Fig 4.2.11**.



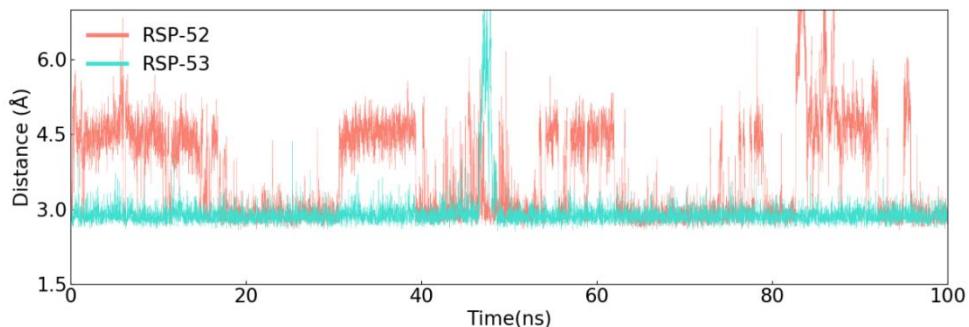
**Fig 4.2.11** Structure of RSP-52, RSP-53, RSP-54 and RSP-55 which performed well in the second round of structural optimization.

**Table 4.2.5** Binding stability descriptors of eight *Closed* R-spondin ligands in the third-round selection.

Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
RSP-51	1.9	5.1	6.6	1.9
RSP-52	2.0	3.2	4.3	1.1
RSP-53	1.0	1.9	3.7	0.7
RSP-54	1.5	1.6	3.7	0.7
RSP-55	2.2	1.6	3.8	0.5
RSP-56	2.5	22.7	18.8	13.5

Two successful compounds, RSP-52 and RSP-53, share the same features as RSP-05 and RSP-40; this suggests that these features are important for the good interaction of binding mode-1 compounds. Highly populated hydrogen bonds formed by these compounds and *Closed* R-spondin are given in **Table 4.2.6**. From the table, RSP-53 forms hydrogen bonds with Cys94 for 82% of time; however, RSP-52 only forms this interaction for 41% of the simulation. Also, the time series of the distance between the Cys94 backbone oxygen and its corresponding ligand hydrogen bond partner (**Fig 4.2.12**) also confirms that this interaction for RSP-53 is stronger than for RSP-52. These two compounds share significant structural similarity with each other. For example, the length of the lefthand moiety, the type of linker and anchoring ring, and the structure of righthand moiety are the same. However, two differences may contribute to the difference in their binding performance: the first is the position of

the chloride atom on the benzene ring (**Figure 4.2.11**). The second difference is the presence of a second branch on the lefthand moiety.



**Fig 4.2.12** Time series of distance between backbone oxygen of Cys94 and its corresponding hydrogen bond partners in RSP-52 (pink) and RSP-53 (cyan).

**Table 4.2.6** Highly populated hydrogen bond interactions in RSP-52, RSP-53, RSP-54 and RSP-55 MD simulation.

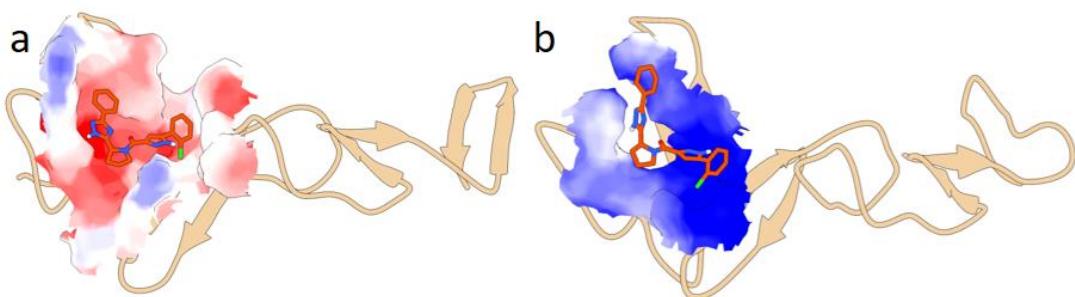
Ligand name	Acceptor	Donor	Fraction
<b>RSP-52</b>	C94 <sup>O</sup>	RSP-52 <sup>N2</sup>	41%
	C94 <sup>O</sup>	RSP-53 <sup>N3</sup>	82%
<b>RSP-53</b>	RSP-53 <sup>O</sup>	K93 <sup>NZ</sup>	30%
	C94 <sup>O</sup>	RSP-54 <sup>N6</sup>	86%
<b>RSP-54</b>	E65 <sup>OE</sup>	RSP-54 <sup>N3</sup>	77%
	L64 <sup>O</sup>	RSP-54 <sup>N5</sup>	50%
	RSP-54 <sup>N1</sup>	K96 <sup>NZ</sup>	41%
	RSP-54 <sup>N</sup>	L64 <sup>N</sup>	21%
<b>RSP-55</b>	L64 <sup>O</sup>	RSP-55 <sup>N4</sup>	37%

The binding mode of RSP-54 is different from other compounds, which will now be discussed in the **Special cases** section.

#### 4.2.4 Special cases

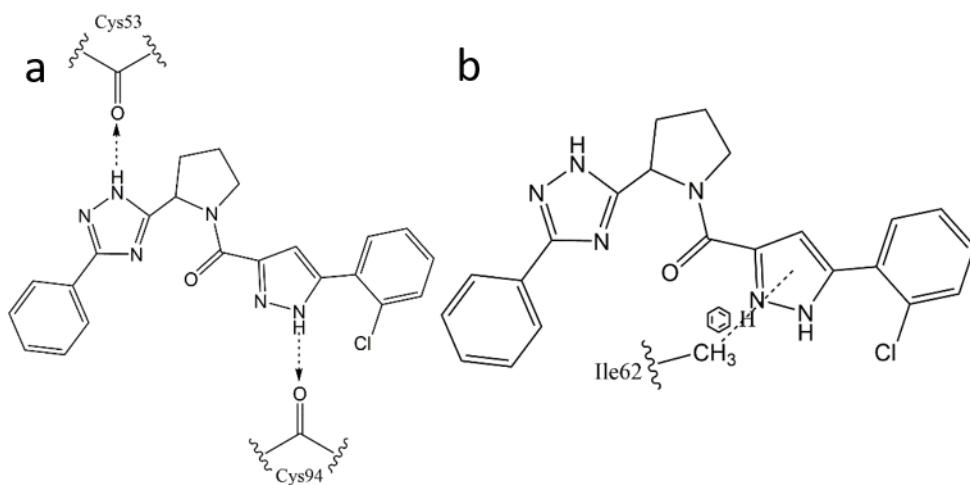
##### Special case: RSP-04

First special compound is RSP-04. Compared with other compounds, the RMSF of this compound is only 1.5 Å (**Table 4.2.2**) which is the lowest one among ligands RSP-01 to RSP-13. Its ligand positional shift is 3.3 Å, which is also the smallest among the initially selected compounds. This means that RSP-04 binds with **Closed** R-spondin stably but the binding site of RSP-04 changed during the course of MD simulation. The initial and final snapshots of RSP-04 and its binding site are shown in **Fig 4.2.13**. In the initial structure of **Closed** R-spondin, two β-hairpins were stacked with each other and there is no gap between these two β-hairpins. However, after 100 ns of MD simulation, RSP-04 worked open a gap between two β-hairpins and slipped into this cavity rather than staying in the initial binding position.

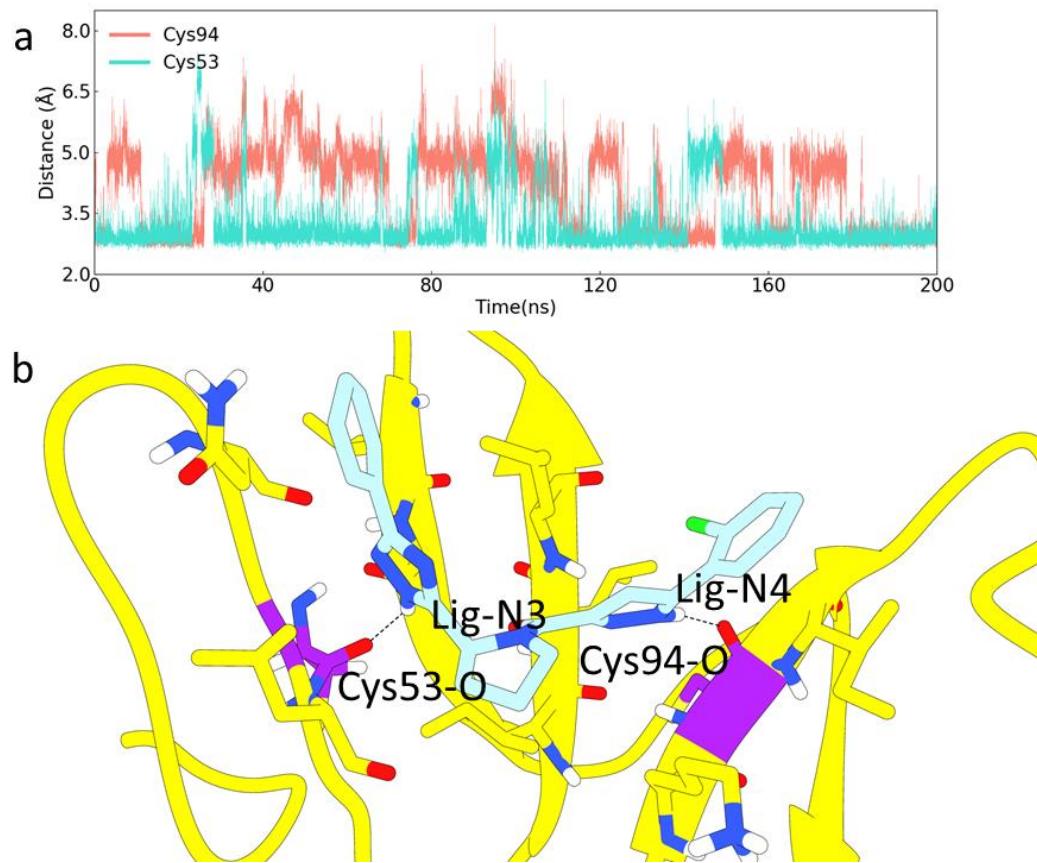


**Fig 4.2.13** a) Initial binding posed and b) final binding pose of RSP-04 binding to **Closed** R-spondin. RSP-04 colored by orange red, **Closed** R-spondin coloured tan and surface colored electrostatics.

According to **Fig 4.2.14 (a)**, initially RSP-04 forms two hydrogen bonds with Cys-94 and Cys53; however, latterly, it only forms one CH- $\pi$  interaction with Ile62. To examine the stability of the 100ns binding pose, an additional 100 ns of MD simulation was prolonged. The formation of two hydrogen bonds: RSP-04-Cys53 and RSP-04-Cys94 were examined by the distances between these two hydrogen bonds partners (**Fig 4.2.15 (a)**). During the first 90 ns, the distances maintained at 3.2 Å; which confirmed the formation of these two hydrogen bonds. At around 100ns, the distances increased to 4.8 Å suggesting that RSP-04 broke the hydrogen bonds with Cys53 and Cys94. However, after 100 ns, the distances decreased to around 2.9 Å suggesting that these two hydrogen bonds formed again (**Fig 4.2.15 (b)**). Based on the 200 ns structure, the binding pose of RSP-04 recovered to typical binding mode 1 binding pose rather than the 100 ns special binding pattern.



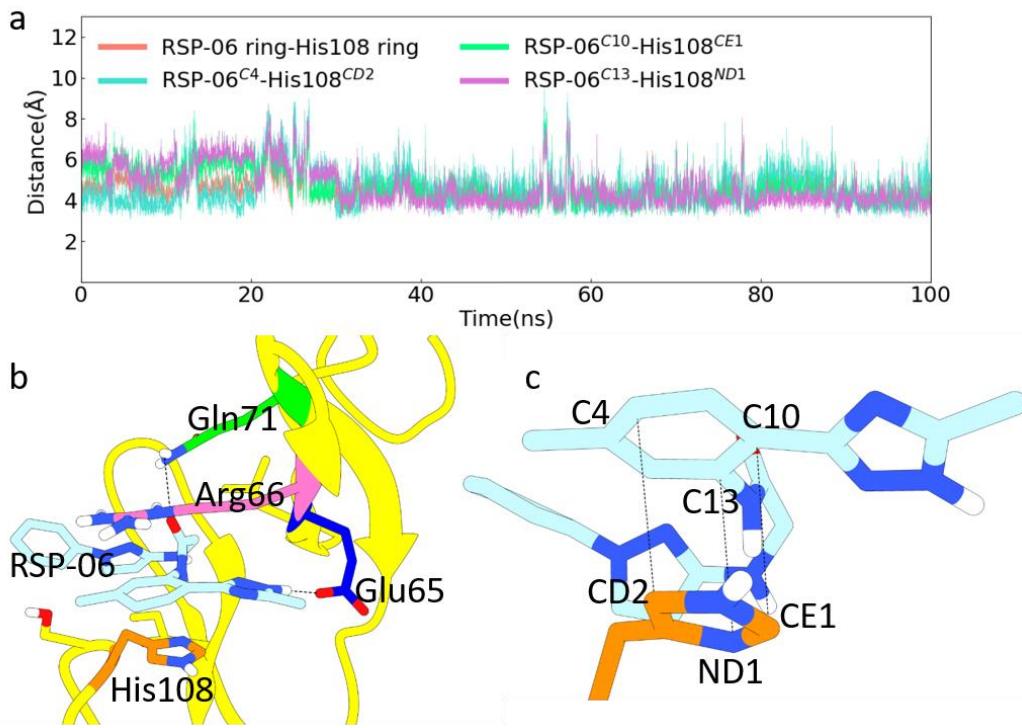
**Fig 4.2.14** 2D schematic representation of protein-ligand interactions between RSP-04 and **Closed R-spondin** in the **a)** initial binding pose and **b)** final binding pose. Hydrogen bonds represented by dotted arrow from hydrogen bond donor to acceptor.



**Fig 4.2.15** a) The distances between RSP<sup>N3</sup> and Cys53<sup>O</sup> (turquoise) and between RSP-06<sup>N4</sup> and Cys94<sup>O</sup> (coral). b) Binding pose of the 200 ns structure of RSP-06, hydrogen bonds are marked by dotted line.

### Special case: RSP-06

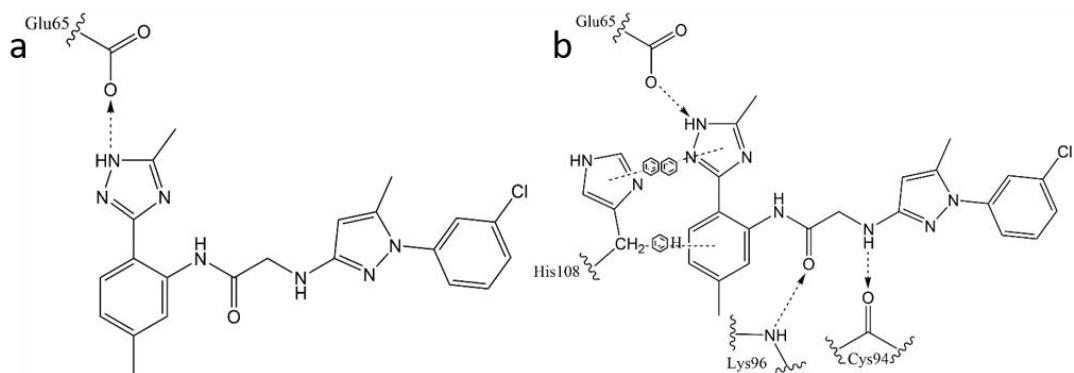
During the course of the simulation of RSP-06, one of the ring structure on RSP-06 moved into a position that was parallel to the ring structure of His108 forming a stable π-π stack (**Fig 4.2.16 (b)**). According to **Fig 4.2.16 (a)**, the distance between two ring systems reached a stable level after 30 ns' MD simulation; which suggests that the π-π stack maintained for a long period of time and may play an important role in stabilizing the binding of RSP-06. In the meantime, RSP-06 also forms a highly populated hydrogen bond with Glu65 and a minor hydrogen bond with Gln71 (**Table 4.2.3**). The binding pattern of RSP-06 shared a totally different feature compared with other compounds.



**Fig 4.2.16 a)** Four sets of distances between the ring system of RSP-06 and the ring system on His108.  
**b)** Final binding pose of RSP-06 binding with **Closed** R-spondin. hydrogen bonds are represented by black dotted line. Specific atoms that were used for distance measurement were labeled in **c**). The distance between RSP-06 ring and His108 ring was measured by the distance between COMs of these two rings.

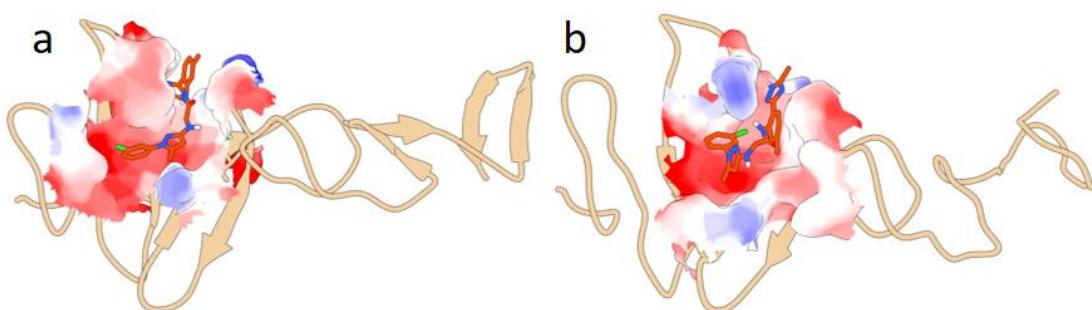
### Special case: RSP-54

The interactions between **Closed** R-spondin and RSP-54 changed a lot during the course of 100 ns of MD simulation (**Fig 4.2.17 (a)**). Only one hydrogen bond between RSP-54 and **Closed** R-spondin was predicted by molecular docking; however, in the MD simulation, multiple hydrogen bonds emerged. Compared with binding mode-1 compounds, RSP-54 possesses a more complex hydrogen bond network: it not only forms sustained hydrogen bonds with Cys94, but also with Glu65, Leu64 and Lys96.

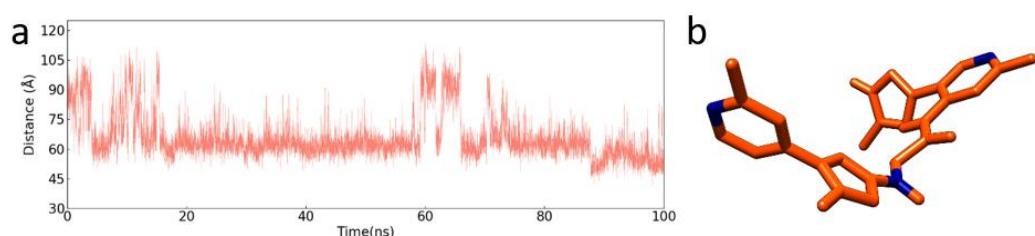


**Fig 4.2.17 a)** Initial and **b)** final binding pattern of RSP-54. Hydrogen bonds represented by dotted arrow from hydrogen bond donor to acceptor.

From inspection, the initial conformation is extended (**Fig 4.2.18 (a)**); however, after the simulation, the conformation of the compound twisted to a more compact conformation (**Fig 4.2.18 (b)**). This can be shown by the angle measured by C4, N6 and C2 atoms of the RSP-54 compound, which decrease from  $90^\circ$  to  $50^\circ$  (**Fig 4.2.19**).



**Fig 4.2.18** 3D representation of initial and final binding pose of RSP-54. Protein structure coloured tan, RSP-54 coloured orange and interface colored electrostatics.

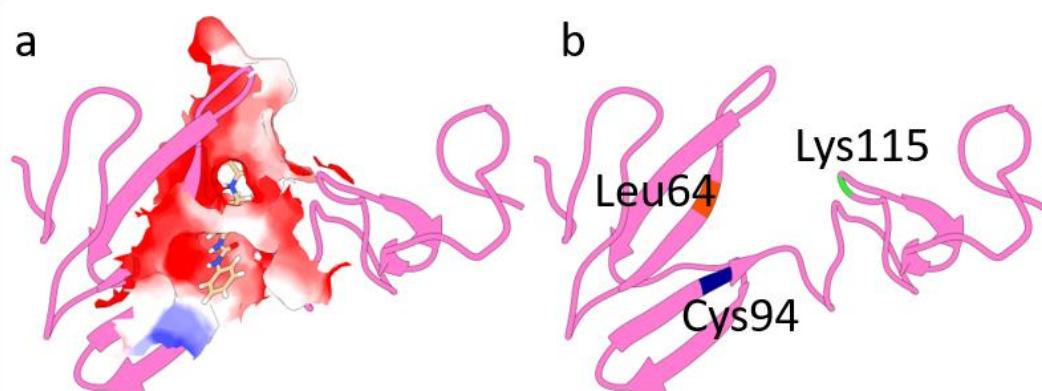


**Fig 4.2.19 a)** The angle measured by three atoms in RSP-54 as a function of time. **b)** RSP-54 coloured orange, and the atoms used in a) coloured blue: they are C4, N6 and C2 atom from left to right.

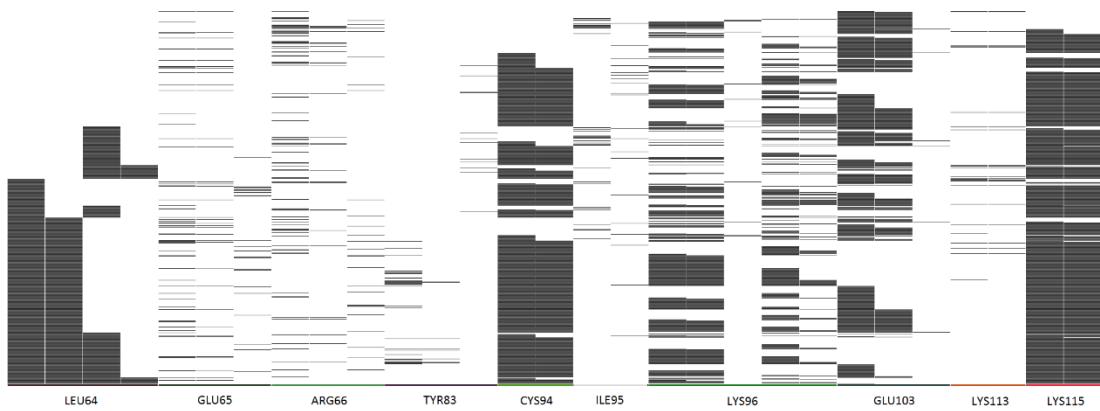
## 4.3 Results: *Open2* R-spondin ligands

### 4.3.1 Compound selection

As stated in **Chapter 3.1.1 (Closed and *Open2* R-spondins)**, ***Open2*** R-spondin is different from two other R-spondin conformations: in the microsecond MD simulation, R-spondin's hinge twisted to an extent that the Furin-like-2 domain was no longer parallel to Furin-like-1 domain; its flat conformation gradually turned into a twisted conformation which forms a deep cavity in ***Open2*** R-spondin. With this deep cavity, ***Open2*** R-spondin demonstrated greater predicted druggability. The surface of the cavity is shown in **Fig 4.3.1 (a)**. The protein-ligand interaction fingerprint indicates a significant correlation between Leu64, Cys94 and Lys115 (**Fig 4.3.2**); this shows a large number of compounds form hydrogen bonds with these three residues at the same time. As shown in **Fig 4.3.1 (b)**, Leu64 is located in the Furin-like 1 domain, Lys115 in the Furin-like 2 domain and Cys94 in the hinge area; this suggests that these compounds might possibly bind these three regions together and function as a modulator of R-spondin.

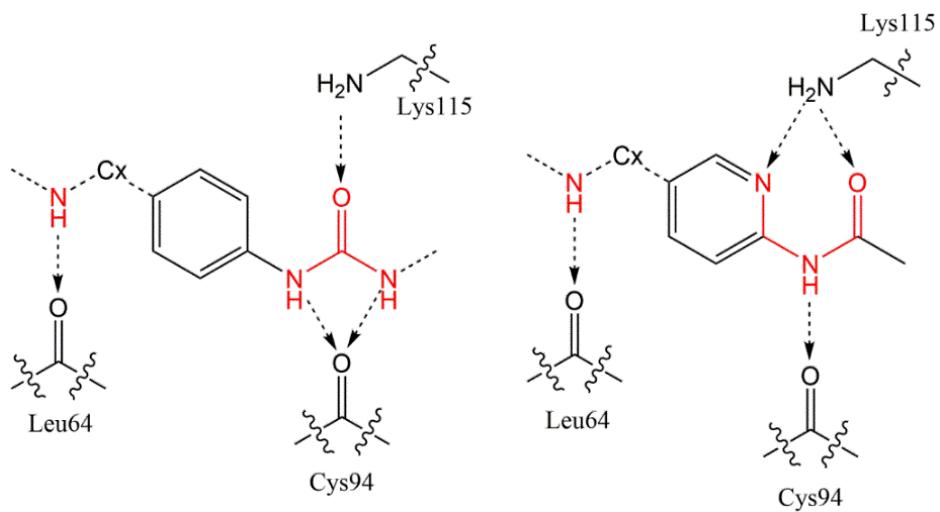


**Fig 4.3.1 a)** The binding interface between ligand and ***Open2*** R-spondin binding cavity. **b)** The location of Leu64 (orange, Furin-1 domain), Cys94 (navy blue, hinge area) and Lys115 (green, Furin-2 domain) in ***Open2*** R-spondin.



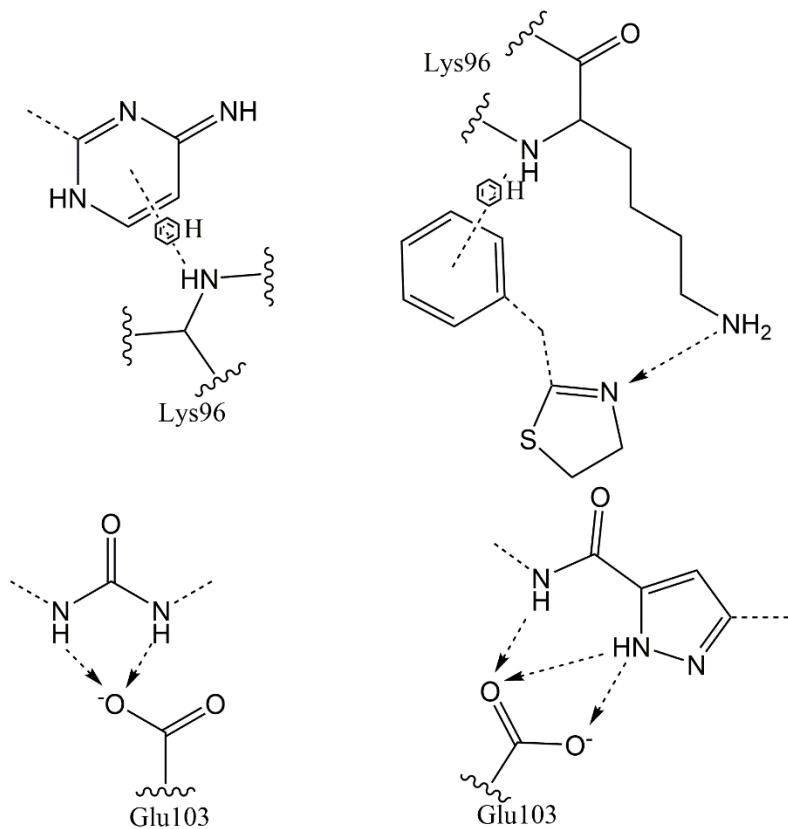
**Fig 4.3.2** Protein-ligand interaction fingerprint of top 1000 compounds targeting *Open2* R-spondin.  
 (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)

Based on compounds that shared high similarity in fingerprints, two 4-point binding modes were established (**Fig 4.3.3**). The first binding mode will be called binding mode 1 in the following. The core structure of binding mode 1 is a triangular structure which is part of a urea group or N-(pyridine-2-yl)amide group (red structure in **Fig 4.3.3**). The triangular structure is placed beside a ring system or forms a part of the ring system. The three nitrogen/oxygen heteroatoms in the triangular structure could bind with Lys115 and Cys94. The fourth point of the binding mode 1 is a nitrogen atom that was linked by a flexible hinge, consisting of one to two carbon atoms which bind with Leu64.



**Fig 4.3.3** Two 4-point binding models which have been widely observed across the top 1000 **Open2** R-spondin compounds. Key points are coloured red. Hydrogen bonds represented by the dotted arrow from hydrogen bonds donor to hydrogen bond acceptor.

Apart from the three residues mentioned in the last paragraph, most other highly populated hydrogen bond partners are sidechain atoms (**Fig 4.3.2**) e.g. Lys96 and Glu103. Over 60% of compounds form at least one hydrogen bond with Glu103. Over 40% of compounds form more than two hydrogen bonds or aromatic interactions with Lys96. Binding mode 2 is established based on protein-ligand interactions between compounds and Lys96 or Glu103. Core structures of binding mode-2 are shown in **Fig 4.3.4**. Compared with binding mode 1, there is no significant interaction correlation between compounds and these two residues.



**Fig 4.3.4** Schematic 2D representation of (top) NH- $\pi$  interactions formed by **Open2** R-spondin compound aromatic rings and Lys96 backbone nitrogen; and (bottom) hydrogen bonds with sidechain of Glu103. Hydrogen bonds represented by dotted arrow from hydrogen bond donor to acceptor.

In the selection process, 13 compounds were screened; among which six compounds belong to binding mode 1 i.e. RSP-57 RSP-58, RSP-59, RSP-60, RSP-64 and RSP-66. Five compounds (RSP-61, RSP-62, RSP-67, RSP-68 and RSP-69) belong to binding mode 2; and two compounds, RSP-63 and RSP-65, were deemed special as they did not belong to the two major binding modes. Most compounds selected from the molecular docking had a long backbone. Further information of those selected compounds is given in **Table 4.3.1** below.

**Table 4.3.1** Name and ZINC ID of thirteen *Open2* R-spondin ligands and their corresponding molecular docking scores and rankings among the top 1000 compounds.

Ligand name	Compound ZINC ID	Molecular docking score	Ranking (top 1000)	Binding pose
RSP-57	ZINC000756105835	-12.8	4	1
RSP-58	ZINC000756115073	-13.1	2	1
RSP-59	ZINC000284078002	-12.3	10	1
RSP-60	ZINC001290920185	-12.0	49	1
RSP-61	ZINC001216223386	-12.2	24	2
RSP-62	ZINC001678032632	-12.0	44	2
RSP-63	ZINC001168642430	-12.0	54	Other
RSP-64	ZINC000756112565	-11.2	345	1
RSP-65	ZINC000813354566	-11.1	445	Other
RSP-66	ZINC001661991519	-11.5	162	1
RSP-67	ZINC000761649431	-11.4	342	2
RSP-68	ZINC000615448906	-10.9	696	2
RSP-69	ZINC001118733425	-12.5	242	2

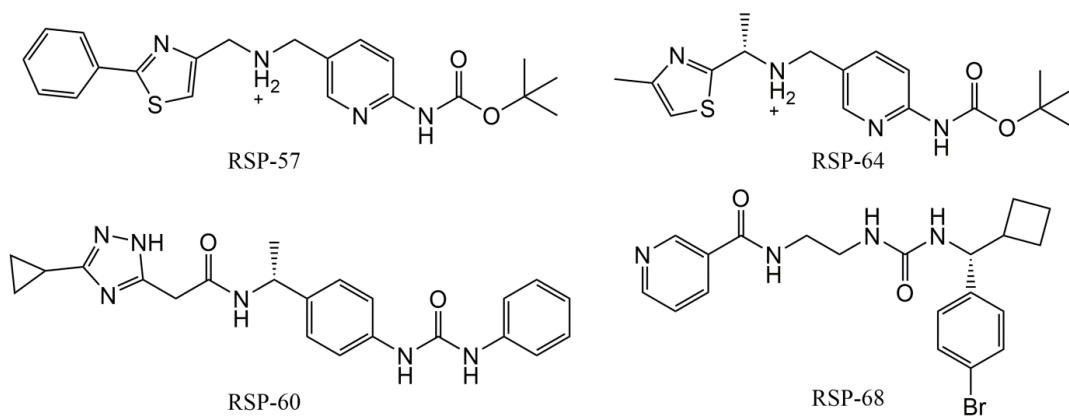
### **4.3.2 Ligand Binding stabilities**

After compounds had been selected, molecular dynamics simulations of the complexes with **Open-2** R-spondin were performed for 100 ns to predict the binding stability of the compounds. The main descriptors of protein-ligand stability are presented in **Table 4.3.2**.

**Table 4.3.2** Binding stability descriptors of thirteen **Open2** R-spondin ligands in the first-round selection.

Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
<b>RSP-57</b>	2.4	3.3	4.7	1.1
<b>RSP-58</b>	1.7	1.3	2.1	0.4
<b>RSP-59</b>	2.3	14.9	9.0	11.3
<b>RSP-60</b>	2.1	2.0	2.0	0.9
<b>RSP-61</b>	1.8	5.6	7.8	3.5
<b>RSP-62</b>	1.8	2.1	2.8	0.6
<b>RSP-63</b>	1.4	1.8	4.7	0.9
<b>RSP-64</b>	2.1	2.8	3.9	1.4
<b>RSP-65</b>	1.2	4.1	4.5	2.1
<b>RSP-66</b>	2.5	18.3	17.3	7.8
<b>RSP-67</b>	2.6	3.3	4.3	1.4
<b>RSP-68</b>	1.9	1.9	3.9	1.2
<b>RSP-69</b>	2.2	3.2	6.9	1.5

Based on the analysis of the 13 MD trajectories, four compounds appeared to bind well with **Open2** R-spondin: they are RSP-57, RSP-60, RSP-64 and RSP-68. The 2D structures of these compounds are shown in **Fig 4.3.5**. The RMSF and ligand positional shift of these four compounds were in a reasonable range: RMSF ranged from 2.0 Å to 3.3 Å and ligand positional shift ranged from 2.0 Å to 4.7 Å. Also, the binding mode of these compounds did not significantly change during the simulation. RSP-57, RSP-60 and RSP-64 belong to binding mode-1 and RSP-68 belongs to the Glu103-based binding mode-2. To further explore the potential of each binding pattern, hydrogen bonds between protein and **Open2** R-spondin were analysed (**Table 4.3.3**). RSP-57 and RSP-60 formed hydrogen bonds with the sidechain oxygen atoms of Glu103 for 72% and 84% of the time respectively rather than interacting with Leu64, Cys94 or Lys115. RSP-64 was the Glu103-based binding mode-2 compound and clearly interacts with Leu64 most. This suggests that Glu103 plays an important role in protein ligand binding.



**Fig 4.3.5** The structures of compounds that bind well with **Open2** R-spondin conformation.

**Table 4.3.3** Highly populated hydrogen bonds formed by four stably bound compounds and **Open2** R-spondin during the course of 100ns MD simulation.

Ligand name	Acceptor	Donor	Fraction
<b>RSP-57</b>	E103 <sup>OE</sup>	RSP-57 <sup>N3</sup>	72%
	Q71 <sup>OE</sup>	RSP-57 <sup>N2</sup>	36%
<b>RSP-60</b>	E103 <sup>OE</sup>	RSP-60 <sup>N2</sup>	84%
	RSP-60 <sup>N</sup>	R66 <sup>N</sup>	18%
<b>RSP-64</b>	L64 <sup>O</sup>	RSP-64 <sup>N5</sup>	18%
	L64 <sup>O</sup>	RSP-64 <sup>N3</sup>	46%
<b>RSP-68</b>	Q71 <sup>OE</sup>	RSP-64 <sup>N2</sup>	21%
	RSP-64 <sup>O</sup>	L64 <sup>N</sup>	11%
<b>RSP-59</b>	E103 <sup>OE</sup>	RSP-68 <sup>N2</sup>	52%
	E103 <sup>OE</sup>	RSP-68 <sup>N3</sup>	45%
<b>RSP-66</b>	RSP-68 <sup>O</sup>	Y83 <sup>OH</sup>	29%
	L64 <sup>O</sup>	RSP-68 <sup>N1</sup>	19%
<b>RSP-63</b>	RSP-68 <sup>O</sup>	R66 <sup>N</sup>	10%

The RMSF of ligands RSP-59 and RSP-66 were over 10 Å which suggests these two compounds dissociate from the protein and fluctuate freely in water solvent. The ligand positional shifts of these compounds are 9.0 Å and 17.3 Å respectively which also confirms that they detached from the binding site and could not stably form protein-ligand complexes with **Open2** R-spondin. Although the RMSF values for RSP-63, RSP-67 and RSP-69 are within a reasonable range (from 1.8 Å to 3.3 Å), the positional shifts of these compounds are high (from 4.3 Å to 6.9 Å); this means their

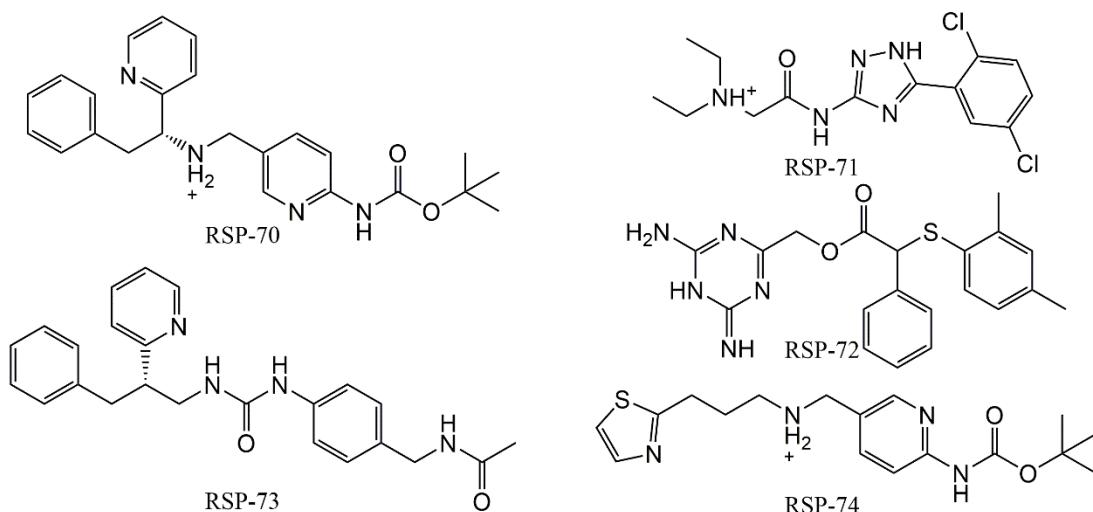
position changed a lot and they did not bind stably with the original binding site. Separate from the five compounds mentioned above, RSP-61 and RSP-65 were found to have induced a change in the conformation of ***Open2*** R-spondin. During the simulation of RSP-61 and protein complex, the cavity formed by the hinge area and two Furin-like domains were opened, making the ligand exposed to water and the protein-ligand complex less stable. For example, the RMSF of RSP-61 reached 5.5 Å and ligand positional shift reached 7.8 Å which means the ligand is highly flexible and unable to bind well with protein.

Two compounds, RSP-58 and RSP-62, showed distinct binding properties to the other simulated compounds and cannot simply be labelled stable or unstable. As a consequence, these two compounds will be discussed in the following ***Special Cases*** section.

#### ***4.3.3 Ligand structure refinement***

In the first-round simulation, most binding mode-1 compounds were not able to bind with Leu64, Cys94 and Lys115 at the same time. Furthermore, the interactions between compounds and Glu103 were important to stabilize the protein-ligand complex. Hence, in this selection, more compounds that form interaction with Glu103 were selected. Two compounds, RSP-75 and RSP-77, failed to bind with ***Open2*** R-spondin. The binding site of these two compounds opens, making ligand fully exposed to solvent. In addition, their RMSF values were over 4 Å, indicating that they have high flexibility and might not able to form stable interactions with the protein. Five compounds including RSP-70, RSP-71, RSP-72, RSP-73 and RSP-74

from the eight selected compounds showed good binding affinity (**Fig 4.3.6**). The RMSF values of these four compounds were all under 2 Å showing that their structures were stable when binding to **Open2** R-spondin (**Table 4.3.4**). We note that the minor tautomer of RSP-72 shown in **Fig 4.3.6** was simulated by MD, and it could form a good hydrogen bond network with the protein's residues.



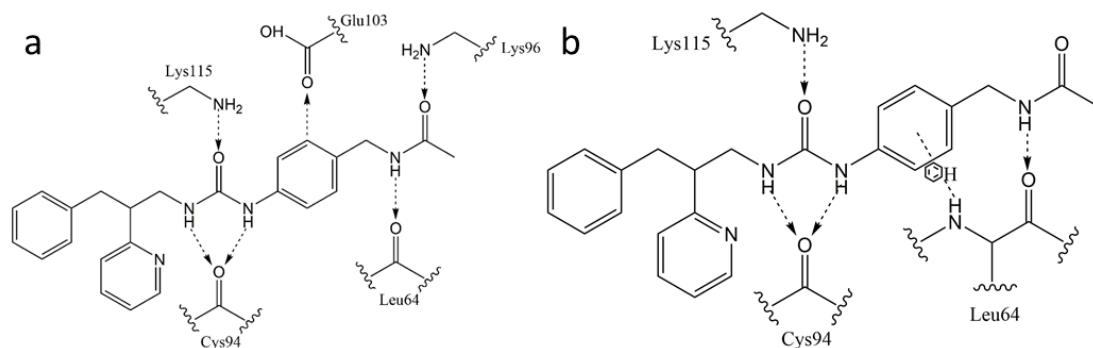
**Fig 4.3.6** Structures of five compounds including RSP-70, RSP-71, RSP-72, RSP-73 and RSP-74 which demonstrated good binding stability with **Open2** R-spondin.

**Table 4.3.4** Binding stability descriptors of **Open2** R-spondin ligands in the second-round selection.

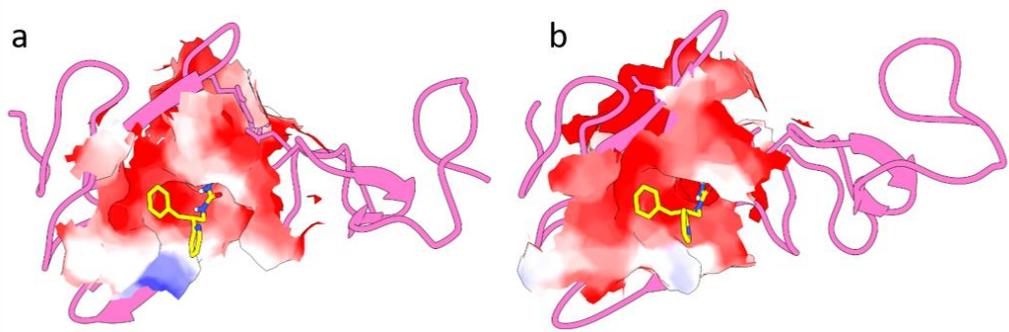
Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
<b>RSP-70</b>	2.3	1.7	2.8	0.7
<b>RSP-71</b>	1.3	1.8	1.5	0.6
<b>RSP-72</b>	2.3	1.5	3.8	0.6
<b>RSP-73</b>	1.7	1.3	1.3	0.4
<b>RSP-74</b>	1.9	1.8	2.3	0.6

<b>RSP-75</b>	2.2	6.4	8.5	2.5
<b>RSP-76</b>	2.2	2.2	3.5	0.8
<b>RSP-77</b>	2.6	4.5	3.6	2.2

Noticeably, RSP-71 and RSP-73 formed hydrogen bonds with Cys94 with high occupancies: RSP-71 hydrogen bonded with the residue over 84% of the time; and RSP-73 formed two hydrogen bonds with Cys94 for 80% and 56% of the time respectively (**Table 4.3.5**). As schematically shown in **Fig 4.3.7**, three major interactions, between RSP-73 and Leu64, Cys94 and Lys115, are maintained during the course of MD simulation. The occupancy of hydrogen bonds interactions between RSP-73 and Leu64, Cys94 and Lys115 are among highest three interactions; this is in accord with the proposed binding mode 1, suggesting its likelihood. The initial binding pose and the final binding pose of RSP-73 also confirm that the binding pose did not experience a significant change (**Fig 4.3.8**).



**Fig 4.3.7** Schematic 2D representation of **a)** initial binding pattern and **b)** final binding pattern of RSP-73 bound to **Open2** R-spondin conformation.



**Fig 4.3.8** Graphical representation of RSP-73 **a)** initial binding pattern predicted by molecular docking and **b)** final binding pattern after 100 ns MD simulation of the protein-ligand complex. RSP-73 coloured yellow and protein coloured pink. The protein-ligand interface coloured by electrostatics.

**Table 4.3.5** Highly populated hydrogen bond interaction in RSP-70, RSP-71, RSP-72, RSP-73 and RSP-74 MD simulation.

Ligand name	Acceptor	Donor	Fraction
<b>RSP-70</b>	E103 <sup>O</sup> E	RSP-70 <sup>N3</sup>	88%
	E65 <sup>O</sup> E	RSP-70 <sup>N3</sup>	42%
	RSP-70 <sup>N1</sup>	K115 <sup>NZ</sup>	29%
	L64 <sup>O</sup>	RSP-70 <sup>N2</sup>	18%
<b>RSP-71</b>	C94 <sup>O</sup>	RSP-71 <sup>N2</sup>	84%
	RSP-71 <sup>O</sup>	Y83 <sup>OH</sup>	75%
	L64 <sup>O</sup>	RSP-71 <sup>N4</sup>	48%
	E65 <sup>O</sup> E	RSP-71 <sup>N4</sup>	45%
<b>RSP-72</b>	L64 <sup>O</sup>	RSP-72 <sup>N3</sup>	26%
	L64 <sup>O</sup>	RSP-72 <sup>N2</sup>	14%
	RSP-72 <sup>N2</sup>	K96 <sup>N</sup>	11%
	C94 <sup>O</sup>	RSP-73 <sup>N1</sup>	80%
<b>RSP-73</b>	C94 <sup>O</sup>	RSP-73 <sup>N3</sup>	56%
	L64 <sup>O</sup>	RSP-73 <sup>N2</sup>	34%
	RSP-73 <sup>O1</sup>	K115 <sup>NZ</sup>	25%
	E65 <sup>O</sup> E	RSP-74 <sup>N3</sup>	94%
<b>RSP-74</b>	E103 <sup>O</sup> E	RSP-74 <sup>N3</sup>	86%
	L64 <sup>O</sup>	RSP-74 <sup>N2</sup>	31%

The new hydrogen bond partner, Glu65, is of interest. RSP-74 forms this interaction with Glu65 for over 94% of the simulation time. Also, Glu65 also forms hydrogen bonds with RSP-70 and RSP-71 for over 42% and 45% of the time respectively. However, this is quite different from compounds selected initially. The first-round selected compounds did not form hydrogen bonds with Glu65. Also, molecular docking failed to identify the interactions between RSP-70, RSP-71 and RSP-74 and Glu65. The main reason for this is the high flexibility of ***Open2*** R-spondin. As R-spondin fluctuated and the compounds adjusted their initial positions, such that a ligand might break its interaction with one residue and form a new interaction with another. For example, the N4 atom in RSP-71 forms hydrogen with Leu64 for 48% of time and with Glu65 for 45% of time (**Table 4.3.5**).

RSP-76 demonstrated a different binding pose compared with these compounds, and it will be discussed in the following **Special cases** section.

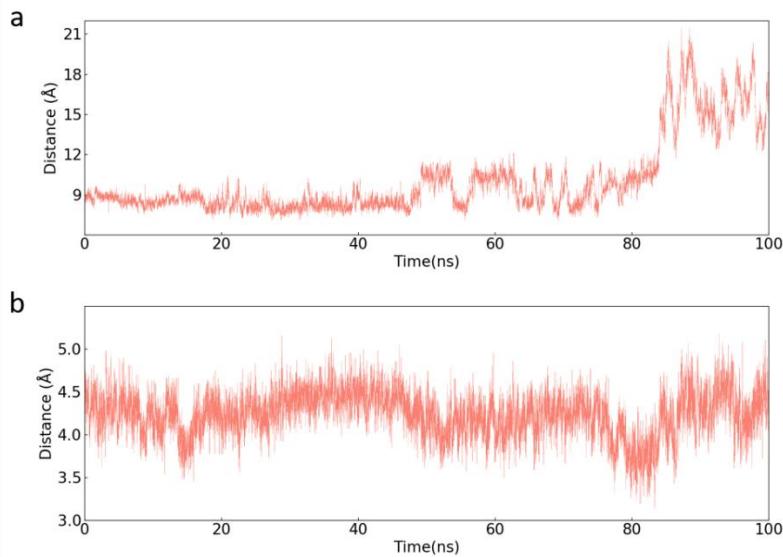
#### ***4.3.4 Special cases***

In molecular dynamics simulations, some compounds may bind with protein in a way that was not expected, however, they showed acceptable binding stability to the protein. As a consequence, these examples will be discussed in this section.

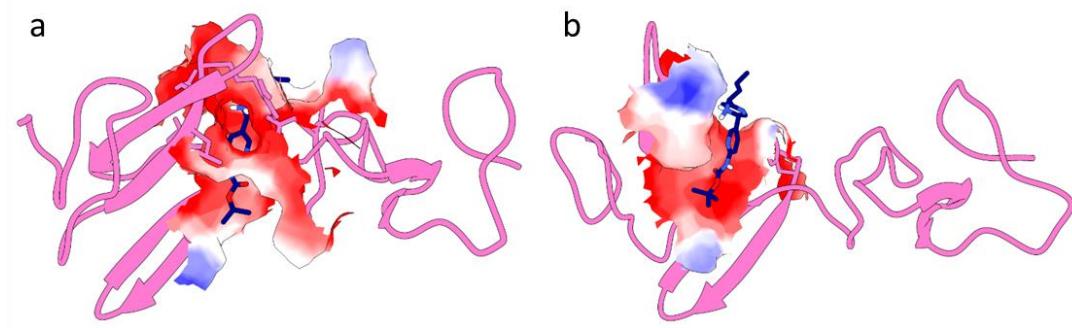
##### **Special case: RSP-58**

The first compound of interest is RSP-58, which demonstrated highly stability when considering its binding descriptors. The RMSF of the compounds is only 1.3 Å and its positional shift from binding site is only 2.0 Å, which suggests that RSP-58 binds to

**Open2** R-spondin with high stability. The reason why it was considered special is that the conformation of **Open2** R-spondin was stable in the first 80 ns; however, the furin-like-2 domain began to rotate away from the Furin-like-1 domain and broke the binding cavity for RSP-58. According to **Fig 4.3.9 (a)**, the distance between the COM of Leu64 (located in Furin-like-1 domain) and the COM of Lys115 (located in Furin-like-2 domain) was maintained at around 8.8 Å in the first 80ns; this value then increased to 15.5 Å in the last 20 ns. Although the cavity for ligand binding was opened, it did not affect the binding stability of RSP-58. The initial and final internal conformations of RSP-58 did not change significantly (**Fig 4.3.10**). The distance between the COM of RSP-58 and COM of the Phe61 to Arg66 region (i.e. binding site residues in the Furin-like-1 domain) are maintained at 4.2 Å (**Fig 4.3.9 (b)**). This means that the position of RSP-58 did not greatly alter with respect to that domain, such that the conformational change of **Open2** R-spondin did not affect the binding stability of RSP-58.



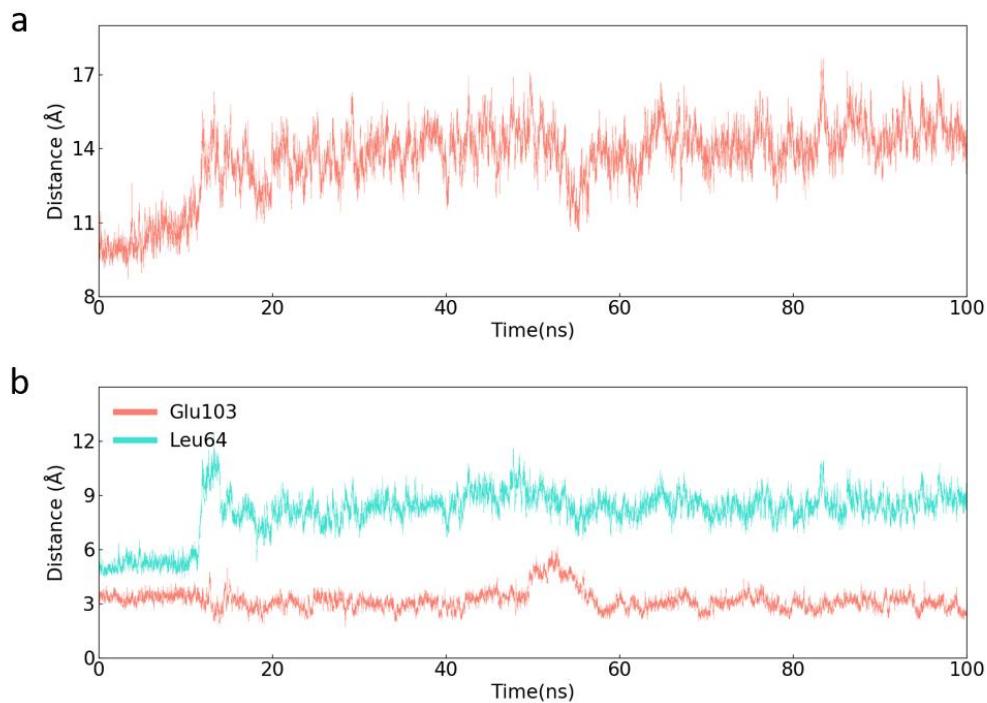
**Fig 4.3.9 a)** Time series of distance between COM of Leu64 and Lys115 from MD simulation of RSP-58/**Open2** R-spondin complex. **b)** Time series of distance between COM of RSP-58 and COM of residues form Phe61 to Arg66 from MD simulation of RSP-58/**Open2** R-spondin complex.



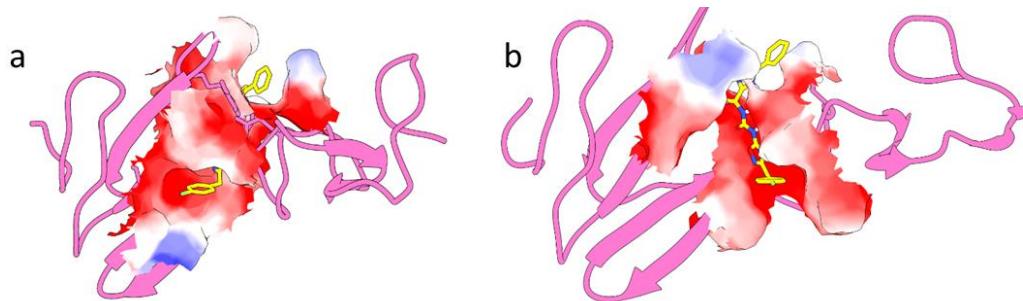
**Fig 4.3.10** **a)** Initial pose of RSP-58 binding to ***Open2*** R-spondin. **b)** final pose of RSP-58 in the 100 ns simulation. The R-spondin protein structure coloured by pink and RSP-58 coloured blue. The protein surface within 4.5 Å of compound is shown and coloured by electrostatics.

#### Special case RSP-62

RSP-62 shares some similar features with RSP-58. Firstly, its RMSF value and ligand positional shift were low (**Table 4.3.2**) indicating that RSP-62 formed a stable protein-ligand complex with ***Open2*** R-spondin. Secondly, the cavity formed by Furin-like 1 and 2 domains and the hinge area opened during the course of simulation. After 10 ns of simulation, the distance between Leu64 and Lys115 increased from 10.3 Å to 14.1 Å (**Fig 4.3.11**), indicating that the conformation opened. However, RSP-62 shared a completely different binding pattern to RSP-58: it binds closer to the Furin 2 domain rather than Furin-like 1 domain; the initial and final binding poses are shown in **Fig 4.3.12**. The distance between the centres of mass of RSP-62 and Glu103, located in the Furin 2 domain, is maintained at ~3.2 Å over the course of simulation (**Fig 4.3.11 (b)**). However, the distance between RSP-62 and Leu64, located in Furin-like 1 domain, increased from 5.2 Å at the beginning, to 8.5 Å after 10 ns of MD simulation (**Fig 4.3.11 (b)**).



**Fig 4.3.11** a) Time series of distance between Leu64 and Lys115 from MD simulation of RSP-62/*Open2* R-spondin complex. b) The distance between COM of RSP-62 and COM of Glu103 (pink) and the distance between COM of RSP-62 and COM of Leu64 as a function of time (cyan).

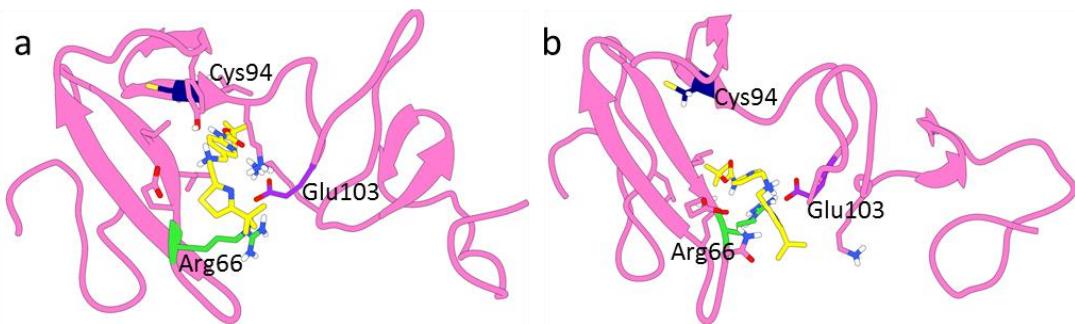


**Fig 4.3.12** a) RSP-62 initial binding pose and b) final binding pose in the MD simulation. The R-spondin protein was colour by pink and RSP-62 was yellow. Protein-ligand interface was coloured by electrostatics.

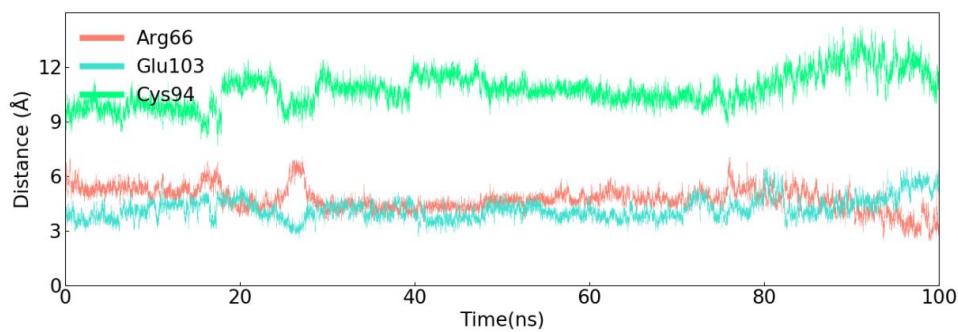
#### Special case RSP-76

For RSP-76, there is a change of binding pose. In the initial protein-ligand complex (**Fig 4.3.13 (a)**), RSP-76 was closely surrounded by Furin-1 and -2 domains and the hinge area; however, during the course of the simulation, RSP-76 gradually moved further from the hinge area to closely bind with residues near the interface of Furin-

like 1 and 2 domains (**Fig 4.3.13 (b)**). Arg66 and Glu103 are two residues located near the interface between the Furin 1 and Furin 2 domains. Cys94 is a residue located in the hinge area linking the two furin-like domains. Considering the time series of distance from these three residues to RSP-68 (**Fig 4.3.14**) shows that RSP-68 tends to stay closer to Arg66 and Glu103 than Cys94. The average distance between centres of mass of RSP-68 and Arg66 is 4.7 Å and to Glu103 is 4.2 Å. During the course of simulation, the average distance between RSP-76 and Cys94 increased from 9.5 Å to 12.0 Å (**Fig 4.3.14**).



**Fig 4.3.13** Graphical representation of **a)** initial and **b)** final binding pose of RSP-76. RSP-76 coloured yellow and **Open2** R-spondin coloured pink. Arg66, Cys94 and Glu103 coloured lime green, navy blue and purple respectively.



**Fig 4.3.14** Distances between COM of RSP-76 and COM of Arg66 (pink), Glu103 (cyan) and Cys94 (green) respectively as a function of time.

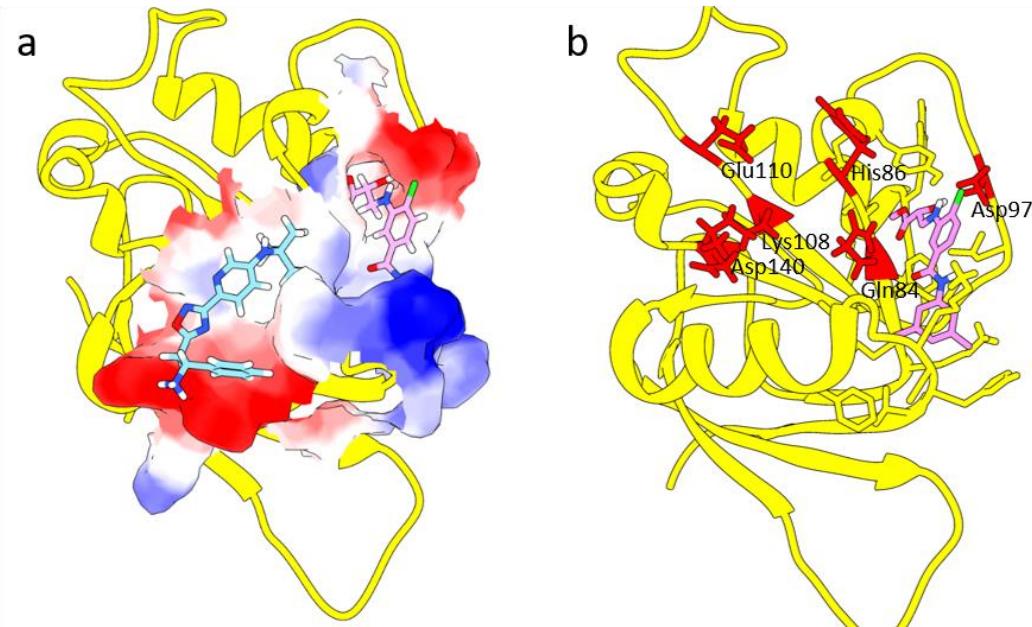
## 4.4 Results: RNF43 ligands

### 4.4.1 Compound selection

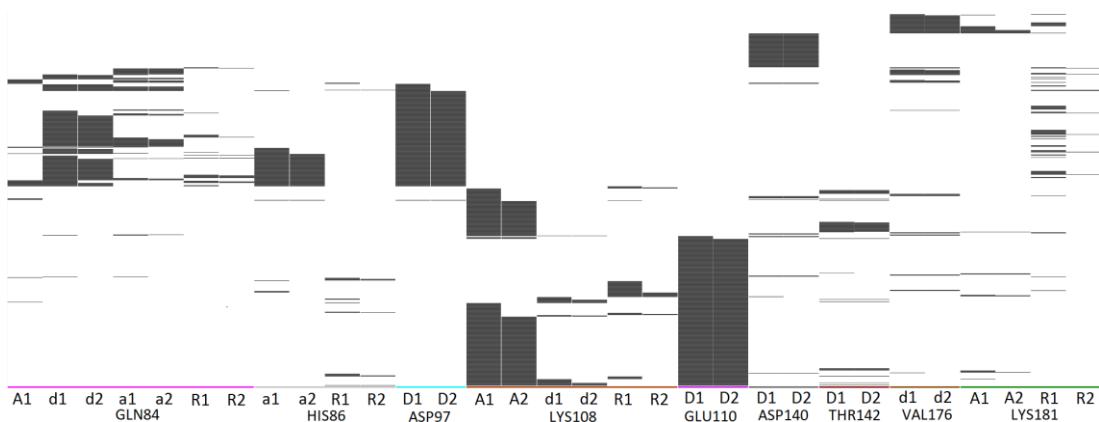
From the consideration of the population of protein-ligand interactions with RNF43 (**Fig 3.3.7**), there is no dominant protein-ligand interaction with RNF43. In comparison with **Closed** and **Open2** R-spondin, whose highest occupied one or two interactions reach 80% to 90%, the highest interaction occupancy of RNF43 only reached 39% and many interaction populations were around 30%. One of the main reasons for this phenomenon is that the binding pocket was sufficiently large that compounds could dock into two major binding spaces. **Fig 4.4.1 (a)** shows the two major binding spaces explored by molecular docking. As stated in **Chapter 3.3.3** (“RNF43 interactions”), the cavity near the binding site of R-spondin and RNF43 was favoured (violet ligand region, **Fig 4.4.1 (a)**); hence, only a few compounds for the blue compound binding space were pursued. The protein-ligand interaction fingerprint (**Fig 4.4.2**) indicates a significant correlation between Gln84, His86 and Asp97; these three residues are located around the violet compound (**Fig 4.4.1 (b)**). Also, Lys108 and Glu110 demonstrated correlation, although these two residues are located in the blue compound binding location which was discarded.

The violet compound binding space (**Fig 4.4.1 (a)**) could be divided into two regions including the upper cavity and the lower cavity (**Fig 4.4.3**). The upper cavity was surrounded by Gln84, His86, Asp97 and their nearby residues. Typical binding mode-1 compounds form major hydrogen bonds with Gln84 and Asp97 (**Fig 4.4.4**). The main structure of binding mode-1 ligands contains two moieties that project into the upper and lower cavity and a linker that connects these two moieties

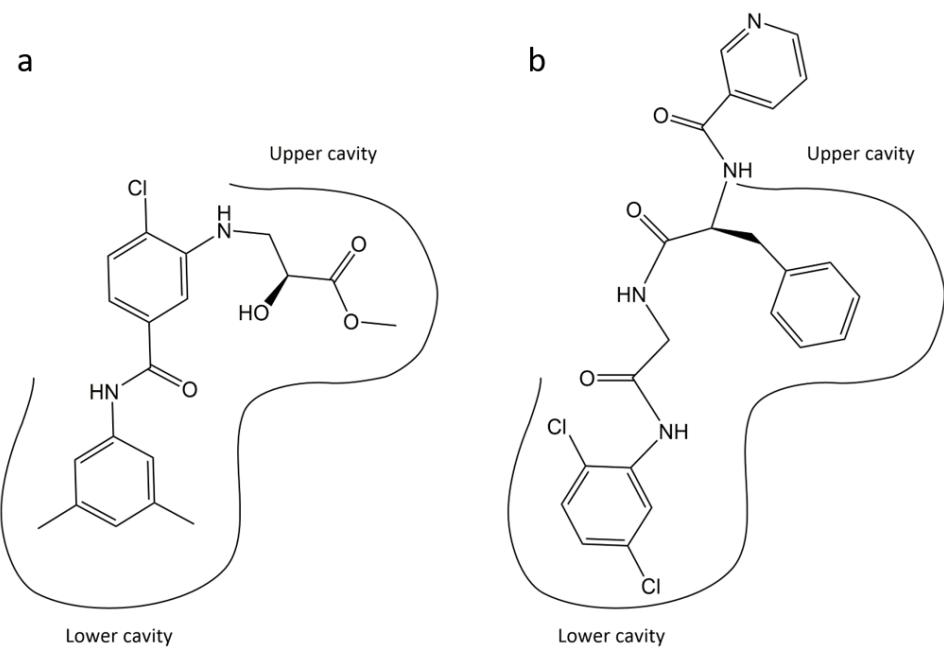
together. As shown in **Fig 4.4.3 (b)**, a third moiety located outside the upper cavity is also highly populated.



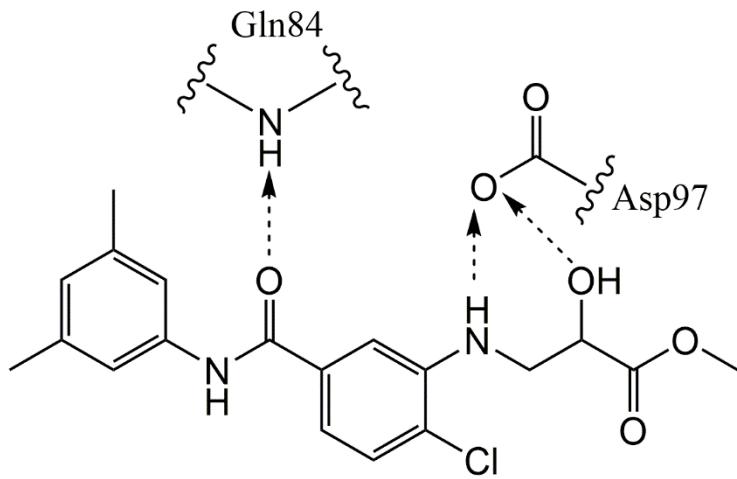
**Fig 4.4.1 a)** Two major binding spaces on RNF43 predicted by molecular docking - two compounds are used to represent the location of these spaces. The Zinc ID of the blue compound is ZINC000843195955. The violet compound is RSP-20. The molecular interface between the two compounds and RNF43 was coloured by electrostatics. **b)** Highly populated hydrogen bond interaction partners including Gln84, His86, ASP97, Lys108, Glu110 and ASP140 were coloured red in the protein structure of RNF43.



**Fig 4.4.2** Protein-ligand interaction fingerprint of top 1000 compounds targeting RNF43. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)



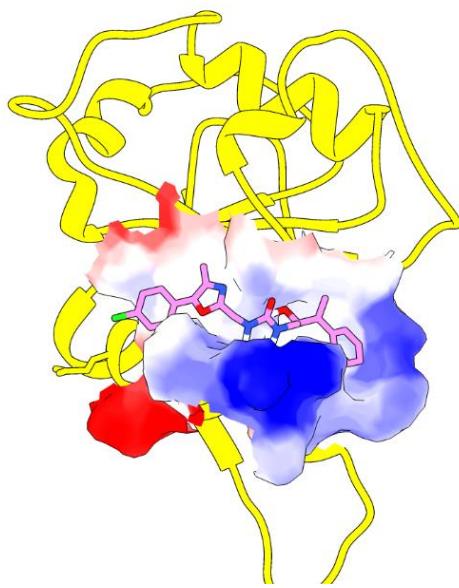
**Fig 4.4.3** Two compounds representing binding mode-1 of RNF43.



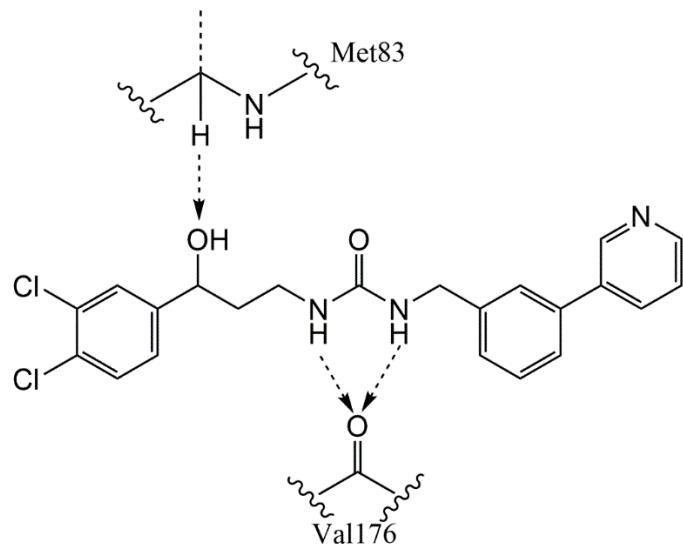
**Fig 4.4.4** Typical binding mode-1 represented by RSP-20. Hydrogen bonds represented by dotted line from hydrogen bond donor to acceptor.

The interaction pattern of binding mode-2 compounds is shown in **Fig 4.4.5**. Binding mode-2 compounds usually have a long backbone lying along the groove formed the  $\alpha$ -helix from Gly166 to Val176 and the main body of the RNF43 protein. There is also a ring system that binds with the lower cavity which was mentioned in the last

paragraph. Binding mode-2 compounds form two hydrogen bonds with Val176, e.g. as in the binding mode of RSP-23 (**Fig 4.4.6**).



**Fig 4.4.5** Binding pose of RSP-22 representing the binding mode-2 of RNF43 ligand.



**Fig 4.4.6** 2D schematic representation of binding mode-2 compounds represented by RSP-23. Hydrogen bonds represented by dotted line from hydrogen bond donor to acceptor.

Based on this analysis, 23 compounds were selected: 12 compounds belong to binding mode-1 and five compounds belong to binding mode-2. Although the blue compound binding space (**Fig 4.4.1**) was not favoured, six compounds bound with this region were examined and they were marked as ‘other’ binding mode.

**Table 4.4.1** Ligand name and Zinc ID of twenty-three selected RNF43 ligands and their corresponding molecular docking scores and rankings.

Ligand name	Compound ZINC ID	Molecular docking score	Ranking	Binding pose
RSP-14	ZINC000742340458	-11.6	151	1
RSP-15	ZINC001183403556	-11.0	778	1
RSP-16	ZINC000108118863	-11.7	89	1
RSP-17	ZINC000521420351	-11.2	432	1
RSP-18	ZINC001566400932	-11.2	476	1
RSP-19	ZINC001464841445	-11.2	400	1
RSP-20	ZINC001252434557	-10.9	943	1
RSP-21	ZINC000790510455	-11.2	418	2
RSP-22	ZINC001224034878	-11.0	740	2
RSP-23	ZINC000759118012	-10.9	865	2
RSP-24	ZINC000642648897	-11.7	112	2
RSP-25	ZINC001416570905	-11.1	523	2
RSP-26	ZINC000673466558	-10.9	979	1
RSP-27	ZINC000152842456	-11.8	81	1
RSP-28	ZINC001166093033	-11.2	438	1
RSP-29	ZINC000752281634	-11.2	479	1
RSP-30	ZINC000125485855	-11.0	678	1
RSP-31	ZINC000543525017	-11.3	318	Other
RSP-32	ZINC000916358643	-11.6	140	Other
RSP-33	ZINC000879188512	-11.6	129	Other
RSP-34	ZINC000523159928	-11.4	248	Other
RSP-35	ZINC001186878380	-11.0	748	Other
RSP-36	ZINC000175328345	-11.2	404	Other

#### **4.4.2 Ligand Binding stabilities**

MD simulations of the protein-ligand complex were performed, and ligand binding descriptors were computed (**Table 4.4.2**). The RMSF of five compounds including RSP-18, RSP-20, RSP-22, RSP-27 and RSP-29 were under 1.7 Å, suggesting could they could form a stable protein-ligand complex with RNF43. Among these compounds, RSP-20 showed the most promising binding properties. The RMSF of RSP-20 is only 1.0 Å and its positional shift from the binding site is only 3.0 Å, also the best among all 23 compounds. The 2D structures of some of these compounds are shown in **Fig 4.4.7** and highly populated hydrogen bonds are presented in **Table 4.4.3**.

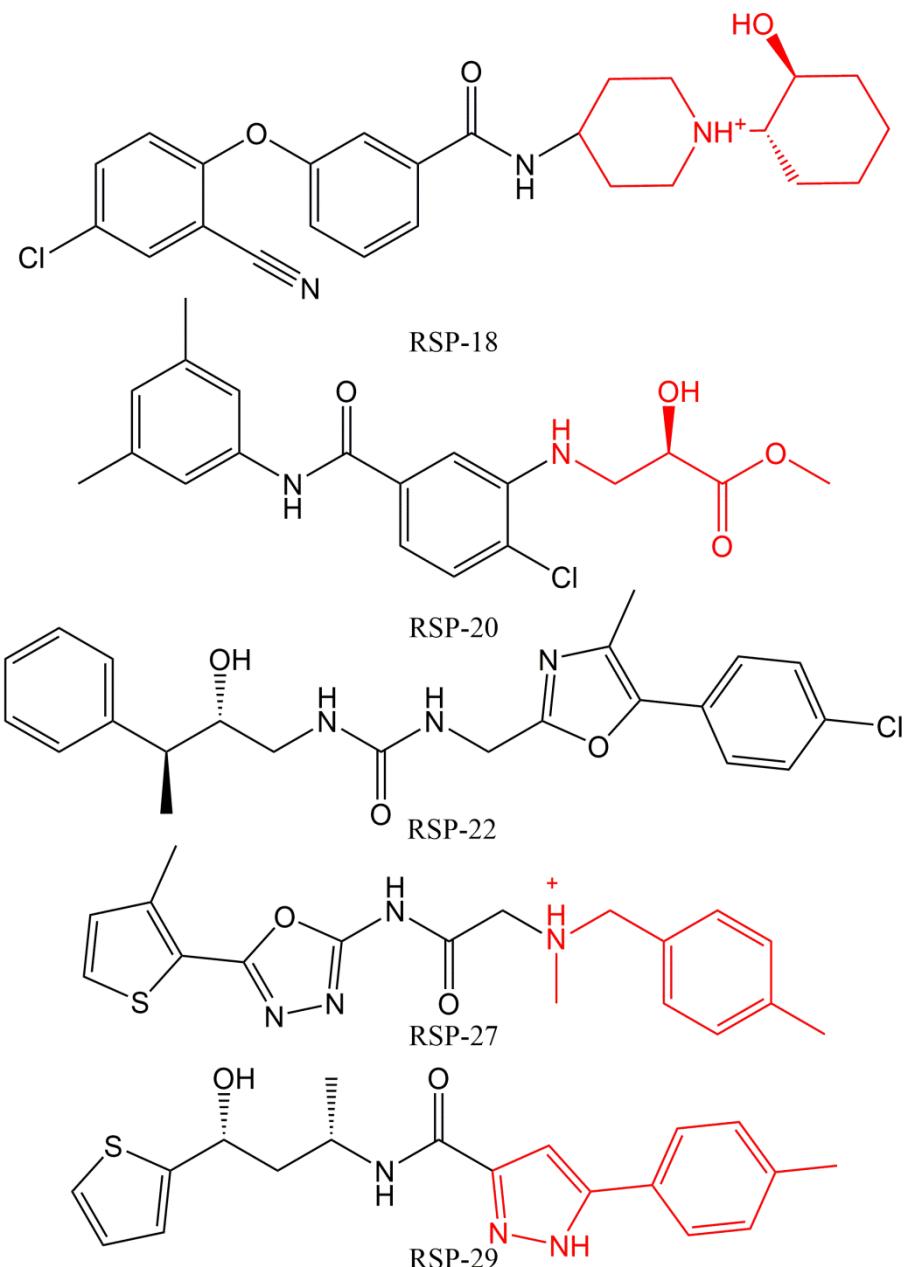
For the binding mode-1 compounds, of the 12 compounds selected, only four of them passed through MD tests. The comparison between successful and failed compounds will be discussed in the following paragraph. As shown in **Table 4.4.3**, RSP-18, RSP-20 and RSP-29 formed hydrogen bonds with Asp97 with high occupancy and the hydrogen bonds with Gln84 were also highly populated. This suggests that Asp97 and Gln84 are important residues that help stabilize the protein-ligand complex.

**Table 4.4.2** Major binding stability descriptors of twenty-three RNF43 ligands in the first-round selection.

Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
<b>RSP-14</b>	2.7	2.8	6.2	0.9
<b>RSP-15</b>	3.7	10.7	18.8	6.5
<b>RSP-16</b>	3.0	2.0	4.8	0.6
<b>RSP-17</b>	1.9	3.2	4.3	1.5
<b>RSP-18</b>	0.8	1.6	4.4	0.5
<b>RSP-19</b>	1.2	6.5	6.4	4.9
<b>RSP-20</b>	1.1	1.0	3.0	0.2
<b>RSP-21</b>	3.0	5.9	6.4	1.9
<b>RSP-22</b>	1.8	1.6	3.3	0.8
<b>RSP-23</b>	2.3	5.5	5.7	3.9
<b>RSP-24</b>	3.7	7.5	10.1	3.2
<b>RSP-25</b>	2.5	4.6	6.5	2.5
<b>RSP-26</b>	2.9	3.5	8.1	1.6
<b>RSP-27</b>	2.0	1.4	4.1	0.4
<b>RSP-28</b>	1.7	5.3	11.1	3.0
<b>RSP-29</b>	1.9	1.7	4.1	0.6
<b>RSP-30</b>	2.0	3.6	6.6	1.7
<b>RSP-31</b>	1.7	4.5	11.8	2.5
<b>RSP-32</b>	2.3	3.3	5.1	2.2
<b>RSP-33</b>	2.4	24.1	18.2	14.2
<b>RSP-34</b>	3.2	3.3	7.1	0.9
<b>RSP-35</b>	2.3	18.7	16.5	12.1
<b>RSP-36</b>	2.1	22.4	18.7	11.0

**Table 4.4.3** Highly populated hydrogen bond interaction in RSP-18, RSP-20, RSP-22, RSP-27 and RSP-29 MD simulation.

Ligand name	Acceptor	Donor	Fraction
<b>RSP-18</b>	D97 <sup>OD</sup>	RSP-18 <sup>O1</sup>	97%
	D97 <sup>OD</sup>	RSP-18 <sup>N2</sup>	41%
	RSP-18 <sup>O</sup>	H86 <sup>N</sup>	10%
<b>RSP-20</b>	RSP-20 <sup>O</sup>	Q84 <sup>N</sup>	69%
	D97 <sup>OD</sup>	RSP-20 <sup>O2</sup>	68%
	RSP-20 <sup>O1</sup>	S105 <sup>OG</sup>	67%
<b>RSP-22</b>	D97 <sup>OD</sup>	RSP-20 <sup>N1</sup>	41%
	D97 <sup>OD</sup>	RSP-20 <sup>O2</sup>	11%
	V176 <sup>O</sup>	RSP-22 <sup>N1</sup>	53%
<b>RSP-27</b>	V176 <sup>O</sup>	RSP-22 <sup>N2</sup>	47%
	Q84 <sup>O</sup>	RSP-27 <sup>N3</sup>	88%
<b>RSP-29</b>	RSP-27 <sup>O</sup>	Q84 <sup>N</sup>	33%
	D97 <sup>OD</sup>	RSP-29 <sup>N1</sup>	62%
	Q84 <sup>O</sup>	RSP-29 <sup>N2</sup>	24%
	Q84 <sup>O</sup>	RSP-29 <sup>O1</sup>	13%



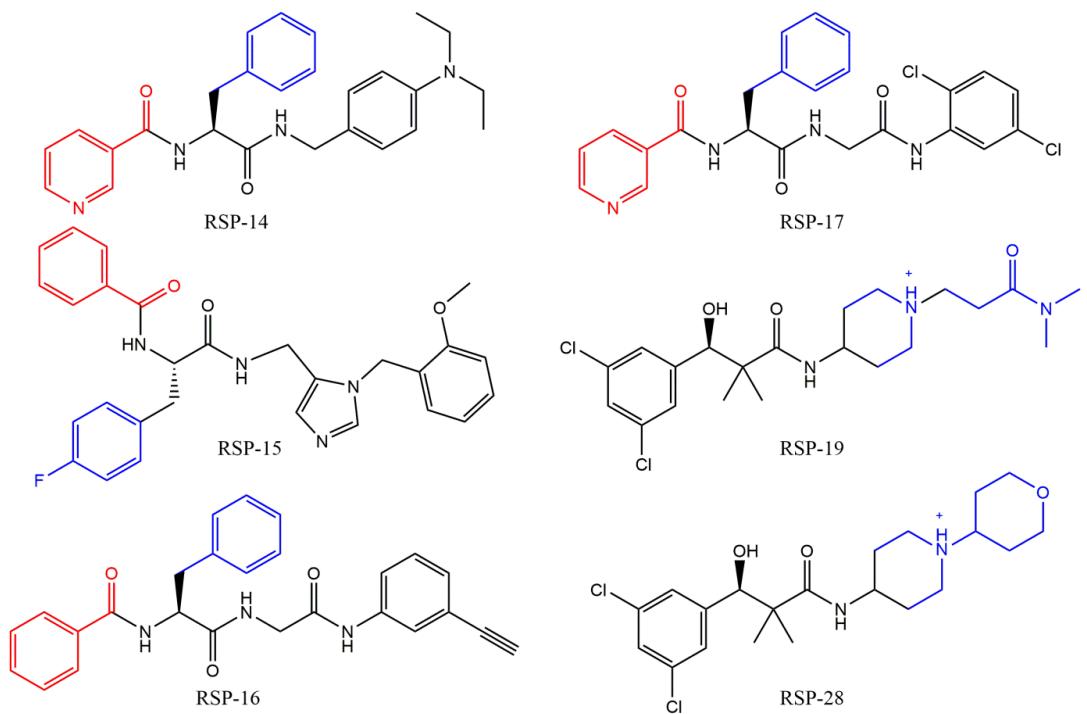
**Fig 4.4.7** 2D structures of RSP-18, RSP-20, RSP-22, RSP-27 and RSP-29 which could bind stably with RNF43. RSP-18, RSP-20, RSP-27 and RSP-29 belongs to binding mode-1 and the part of their structures buried in the upper cavity of RNF43 are coloured red.

Six binding mode-1 compounds: RSP-14, RSP-15, RSP-16, RSP-17, RSP-19 and RSP-28 (**Fig 4.4.8**), failed to bind well with RNF 43 in MD simulations. For the ligand structures in **Fig 4.4.7** and **Fig 4.4.8**, the moieties that bind in the upper cavity are marked in red. We noted that RSP-14, RSP-15, RSP-16 and RSP-17 have an

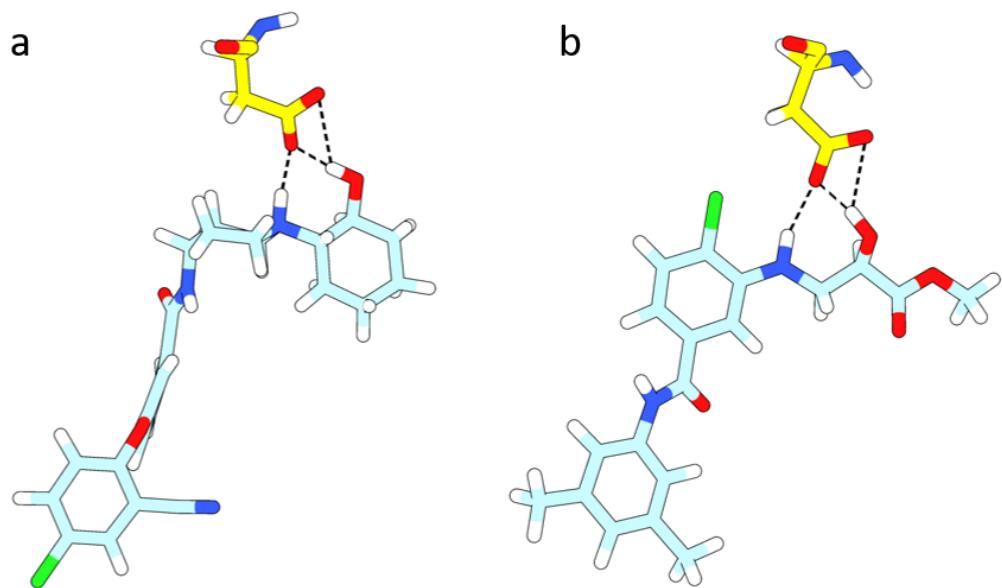
additional ring structure compared to the successful binding mode-1 compounds in

**Fig 4.4.7.** R-14, RSP-15 and RSP-17 do not bind to the upper cavity on RNF43, suggesting that the additional ring structure interferes with the interaction between compounds and the upper cavity. Although RSP-19 and RSP-28 do not have the additional ring structure, they do not form hydrogen bond interactions with Asp97. RSP-19 forms hydrogen bonds with Ser85 with 43% occupancy and Ser108 with 12%; RSP-28 forms hydrogen bonds with Gln84 for only 20% of the time. This is different from successful compounds RSP-18, RSP-20 and RSP-29, which form hydrogen bonds with Asp97 or Gln84 with high occupancy (**Table 4.4.3**).

Although RSP-18 (**Fig 4.4.7**) and RSP-28 (**Fig 4.4.8**) share similar structures where bind with the RNF43 upper cavity, RSP-18 could bind stably and RSP-28 failed. The reason is that RSP-18 possesses an additional hydroxyl group in the upper cavity moiety which might play an important role in stabilizing the binding interactions of compounds. The RSP-18 hydroxyl group forms a hydrogen bond for over 97% of the time (**Table 4.4.3, Figure 4.4.9 (a)**), which means that the hydroxyl group plays a major role in RSP-18/RNF43 binding. This point is supported by RSP-20, which also has a hydroxyl group that binds with Asp97 (**Fig 4.4.9 (b)**).

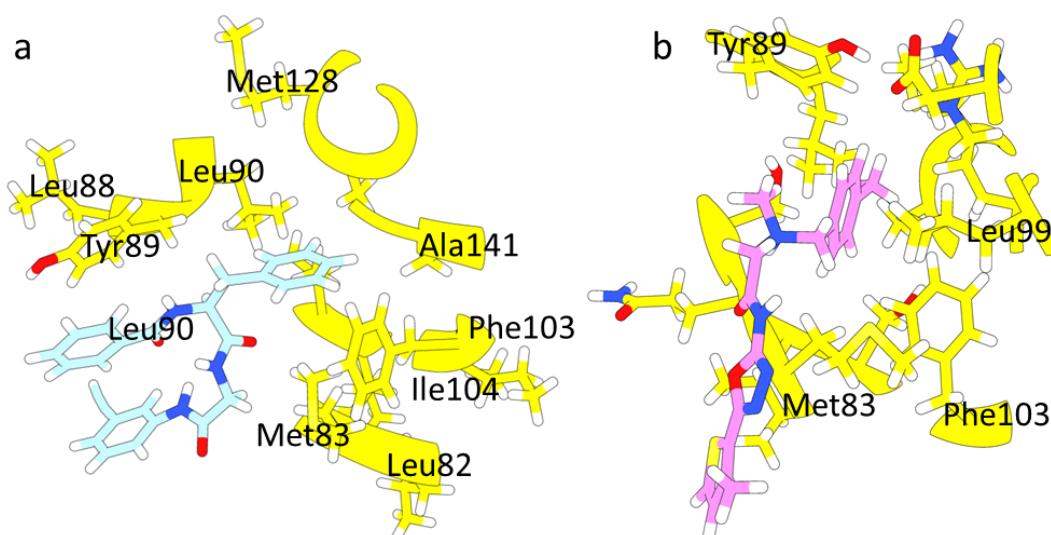


**Fig 4.4.8** Structure of failed RNF43 binding mode-1 compounds. The structure that near outside of the upper cavity was coloured by red and structure that projected into the upper cavity were coloured by green.



**Fig 4.4.9** **a)** Asp97 form three hydrogen bonds with RSP-18. **b)** Asp97 forms three hydrogen bonds with RSP-20.

RSP-16 has an additional ring structure near the upper cavity which is not good at forming a stable complex with protein. Furthermore, as shown in **Fig 4.4.10 (a)**, it cannot bind stably with the lower cavity: RSP-16 only forms 33% of time hydrogen bonds with His86 and 25% of time hydrogen bonds with Ser85 (**Table 4.4.3**). Although RSP-16 has many unfavourable aspects, it still binds firmly with the upper cavity, such that the compound adheres to the protein. In MD simulation, RSP-27 binds firmly with RNF43; however, it did not form hydrogen bonds with Asp97, unlike RSP-18, RSP-20 and RSP-29. Ligand RSP-27 forms two hydrogen bonds with Gln43 for 88% of the time and 33% of the time respectively. It shares a significant similarity with RSP-16; that is, they both have a highly hydrophobic group buried in the upper cavity. As shown in **Fig 4.4.10**, the upper cavity is surrounded by multiple hydrophobic residues, including Met83, Leu90, Leu99, Phe103, Ile106, Met128 and Ala129. These two facts suggest that hydrophobic is favoured in the upper cavity.



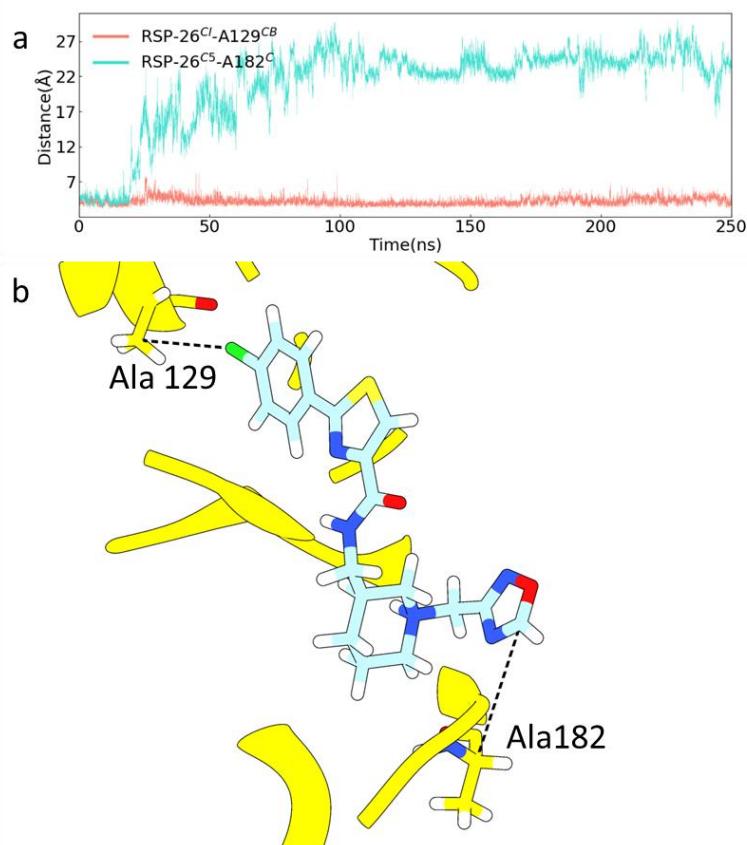
**Fig 4.4.10** Upper cavity hydrophobic residues and **a)** RSP-16 and **b)** RSP-27.

RSP-26 also suggests that the hydrophobicity of the upper cavity is important for protein-ligand interaction. As shown in **Fig 4.4.11 (a)**, the distance between chloride atom and Ala129 CB atom remains short, at around 4.2 Å, which means that the benzyl chloride group binds well with the upper cavity. However, the distance between ligand C5 atom and Ala182 in lower cavity fluctuated largely at 20.8 Å, which means RSP-26 could not bind with the lower cavity at all. Furthermore, RSP-26 did not form any hydrogen bonds with RNF43; thus, the hydrophobic interaction between the benzyl chloride group and the upper cavity is the dominant force that binds RSP-26 to protein. This suggests a hydrophobic moiety in the upper cavity is favoured.

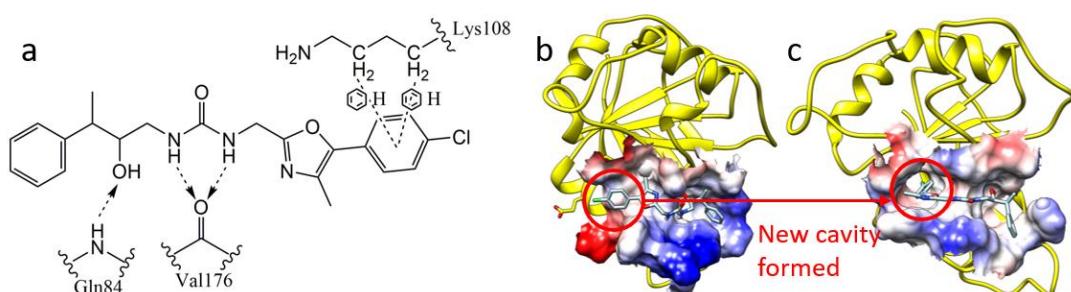
The binding stability of RSP-22 is as good as other successful binding mode-1 compounds. Its RMSF was 1.6 Å and its ligand positional shift was only 3.3 Å (**Table 4.4.2**). RSP-22 mainly binds with Gln84 and Val176 (**Fig 4.4.12 (a)**). The most highly populated hydrogen bonds were between two ligand nitrogen atoms and the Val176 backbone oxygen which occupies 53% and 47% of time respectively (**Table 4.4.3**). It is worth noticing that in the **Fig 4.4.12 (b)**, there is no cavity in the groove where RSP-22 lies; however, after the simulation, a new cavity forms and the benzyl chloride group projects into it. This suggests that using two ring structures as the termini of the compounds could help stabilize their binding to RNF43.

Other compounds like RSP-21, RSP-23, RSP-24, RSP-25 and RSP-30 demonstrated high flexibility during the course of the MD simulation: their RMSF reached 7.5 Å and positional shifts from the binding site were over 5.7 Å, suggesting they could

not bind stably to RNF43. All of the compounds which bind to the blue compound binding pocket (**Fig 4.4.1**) (RSP-31, RSP-32, RSP-33, RSP-34, RSP-35 and RSP-36) failed to pass the molecular dynamics simulation assay. As a consequence, in the next round of simulations, no compounds targeting this pocket were selected.



**Fig 4.4.11** a) Time series of two pairs of distances between RSP-26 and RNF43. b) The first distance is the distance between RSP-26 chloride atom and Ala129 CB atom in upper cavity and the second distances is between RSP-26 C5 atom and Ala182 C atom in lower cavity. These two distances were represented as dotted lines.



**Fig 4.4.12** a) Binding pattern between RSP-22 and RNF43. b) Top view of the initial binding pose of RSP-22 binding pose. c) Top view of the 100ns' simulation of RSP-22 binding pose. The new cavity formed during the MD simulation marked by red circle.

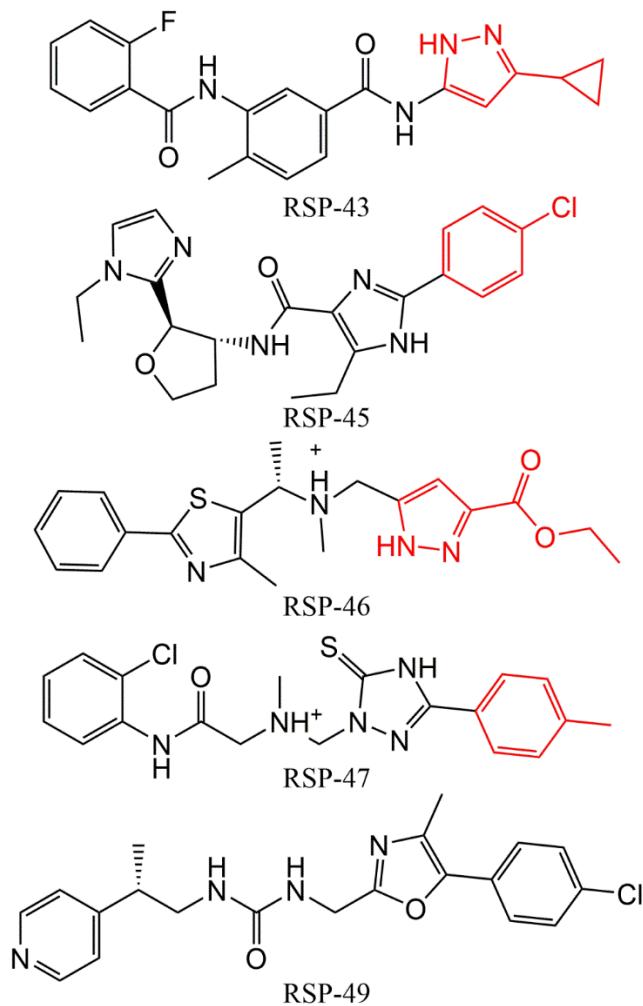
#### **4.4.3 Ligand structure refinement**

Based on the favourable and unfavourable features mentioned in the last section, eight virtually screened compounds with favourable features were selected and tested by MD simulation: RSP-43, RSP-44, RSP-45, RSP-46 and RSP-47 were chosen as binding mode-1 compounds which firstly eliminated the additional ring system near the upper cavity and secondly possess a highly hydrophobic structure for binding to the upper cavity. RSP-50 is a special binding mode which possesses a ring system to bind with the upper cavity and another moiety that lies in a shallow groove near the upper cavity. Two chosen compounds, RSP-48 and RSP-49, belong to binding mode-2. After MD simulation, major binding descriptors of these eight compounds are given in **Table 4.4.4**.

**Table 4.4.4** Binding stability descriptors of eight RNF43 ligands in the second-round selection.

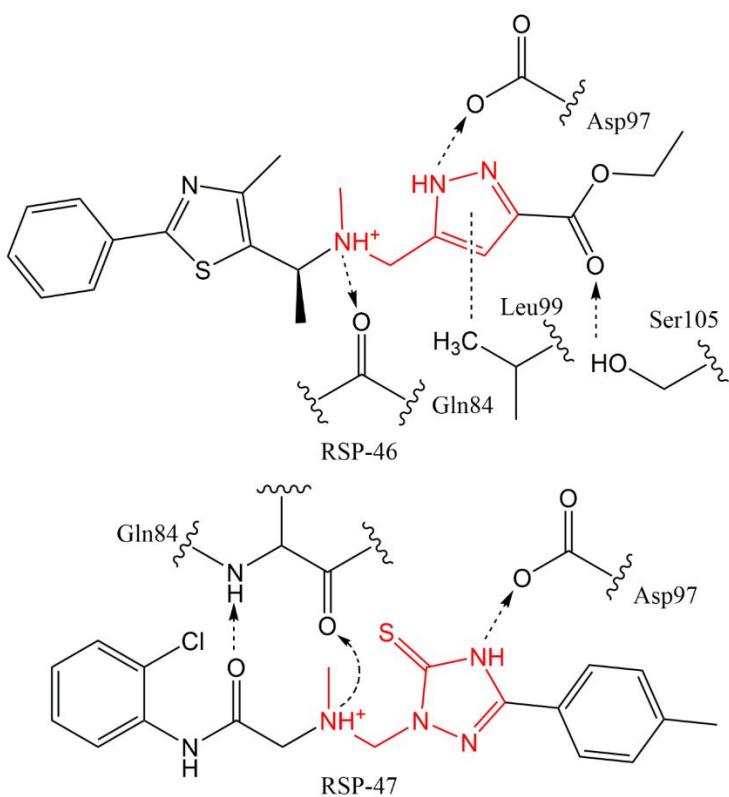
Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
RSP-43	1.5	2.1	4.6	0.5
RSP-44	2.0	4.4	6.3	1.4
RSP-45	1.1	1.0	3.5	0.4
RSP-46	1.5	1.3	3.9	0.7
RSP-47	1.3	1.2	3.4	0.3
RSP-48	2.8	20.9	15.9	11.9
RSP-49	1.8	2.4	3.1	0.6
RSP-50	2.5	3.1	6.7	1.0

Five compounds including: RSP-43, RSP-45, RSP-46, RSP-47 and RSP-49, demonstrated very good binding stability. RSP-43, RSP-45, RSP-46 and RSP-47 are binding mode-1 compounds and their RMSF ranged from 1.0 Å to 2.1 Å (**Table 4.4.4**). Their positional shifts were also under 4.6 Å. In addition, these four compounds all have hydrophobic groups (marked red in **Fig 4.4.13**). Although RSP-44 did not bind well with the lower cavity of RNF43, the 3-(methoxymethyl)-triazole group binds stably with the upper cavity of RNF43. This suggested that a hydrophobic group is important in stabilizing RNF43-ligand complex.

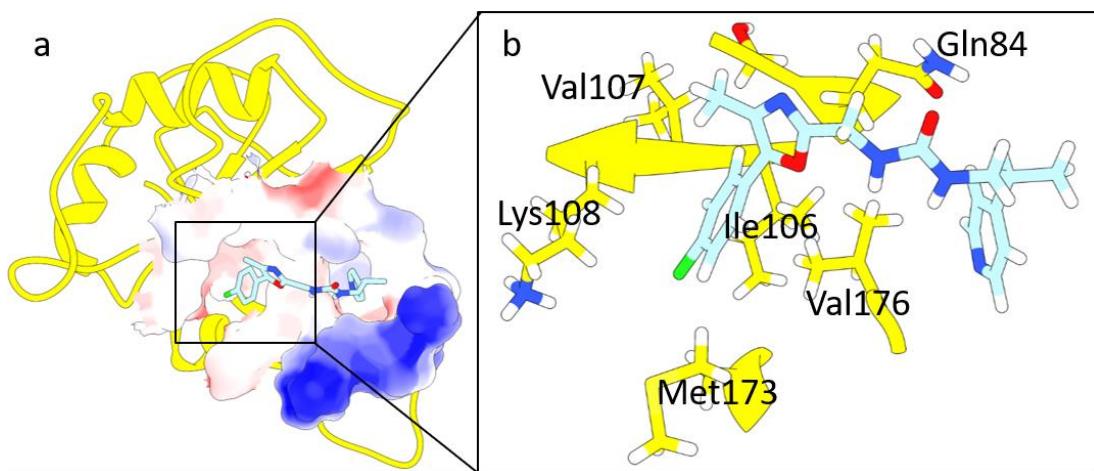


**Fig 4.4.13** 2D structure of RSP-43, RSP-45, RSP-46, RSP-47 and RSP-49. Structures that were buried by upper cavities were coloured by red.

According to **Table 4.4.5**, RSP-46 and RSP-47 shared similar hydrogen bond interactions and both formed good hydrogen bonds with Asp97 and Gln84. As a consequence, a new core structure of binding mode-1 compounds was proposed according to the structures of RSP-46 and RSP-47 (**Fig 4.4.14**). Similar to RSP-22, RSP-49 maintained two main hydrogen bond interactions with Val176 for 71% and 38% of the time (**Table 4.4.5**). These hydrogen bonds were formed by the urea groups and during the course of simulation, a new cavity opened on the surface of RNF43. Two ring structures were projected into the lower cavity and the new cavity as predicted. The binding pose of the 100 ns simulation is shown in **Fig 4.4.15**.



**Fig 4.4.14** Binding pattern of RSP-46 and RSP-47 binding to RNF43. Core structures were coloured red which forms a hydrogen bond interaction between Asp97 and Gln84.



**Fig 4.4.15** 3D binding pattern of RSP-49 with the new cavity zoomed up.

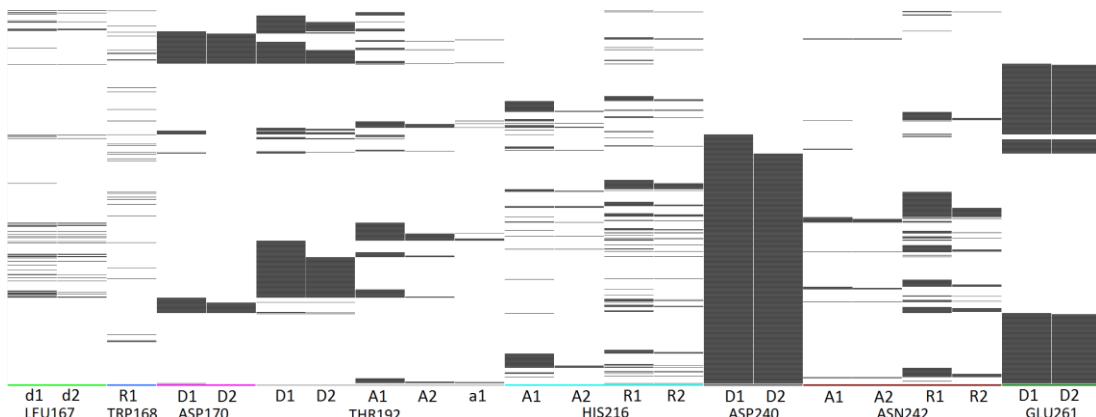
**Table 4.4.5** Highly populated hydrogen bond interactions in RSP-45, RSP-46, RSP-47 and RSP-49 MD simulation with RNF43.

Ligand name	Acceptor	Donor	Fraction
<b>RSP-43</b>	RSP-43 <sup>N</sup>	R132 <sup>NH</sup>	37%
	D97 <sup>OD</sup>	RSP-43 <sup>N1</sup>	16%
<b>RSP-45</b>	D97 <sup>OD</sup>	RSP-45 <sup>N2</sup>	45%
	D97 <sup>O</sup>	RSP-45 <sup>N2</sup>	15%
<b>RSP-46</b>	D97 <sup>OD</sup>	RSP-46 <sup>N2</sup>	94%
	RSP-46 <sup>O</sup>	S105 <sup>OG</sup>	79%
	Q84 <sup>O</sup>	RSP-46 <sup>N3</sup>	75%
<b>RSP-47</b>	D97 <sup>OD</sup>	RSP-47 <sup>N1</sup>	86%
	Q84 <sup>O</sup>	RSP-47 <sup>N4</sup>	84%
	RSP-47 <sup>O</sup>	Q84 <sup>N</sup>	67%
<b>RSP-49</b>	V176 <sup>O</sup>	RSP-49 <sup>N2</sup>	71%
	V176 <sup>O</sup>	RSP-49 <sup>N3</sup>	38%
	RSP-49 <sup>N</sup>	Q84 <sup>N</sup>	14%

## 4.5 Results: LGR5 ligands

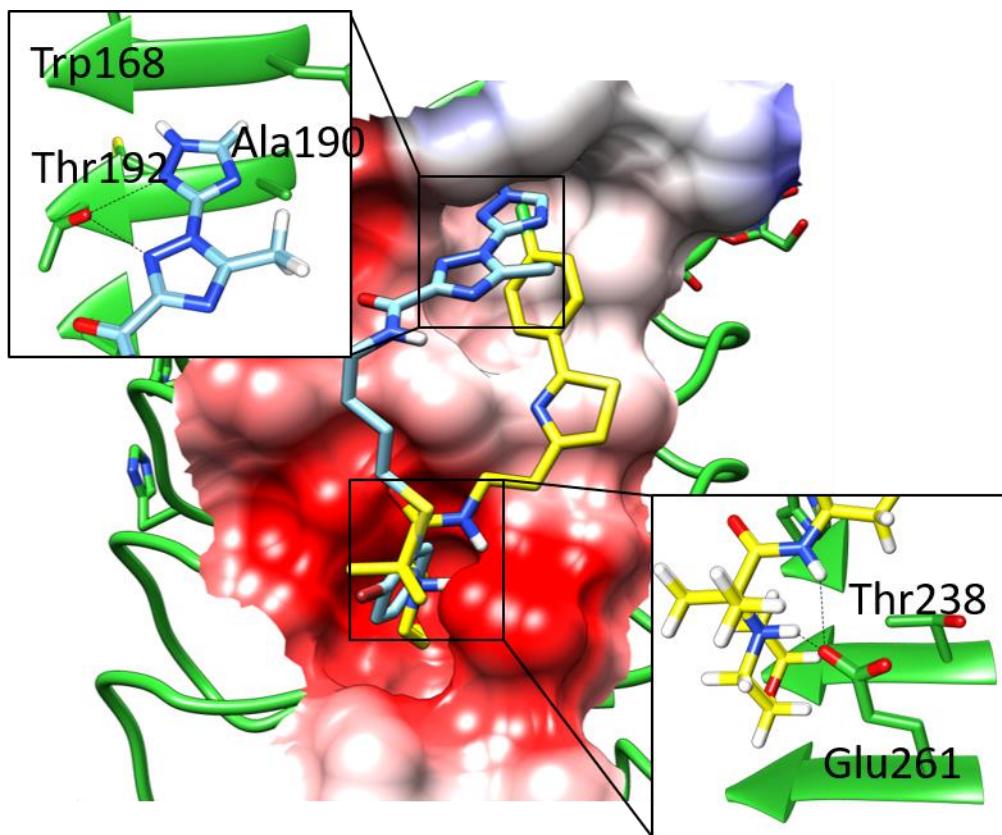
### 4.5.1 Compound selection

For the molecular docking with LGR5, ligands tend to bind well in two cavities: one is the cavity that binds with Phe106 of R-spondin; and the other is the cavity formed by Thr238, Asp240 and Glu261 (see **Chapter 3.3.5**). According to PLIF analysis of LGR5 ligands, Asp240 shows two significant correlations with Thr192 and Glu261; these two correlations did not overlap with each other. Based on these two residues, two binding patterns were built for LGR5 ligands to link the two cavities together (**Fig 4.5.1**). Since there is a small protrusion between the two cavities, and compounds should bridge the protrusion - either along the upper left groove (**Fig 4.5.2**, blue compound position) or the lower right groove (**Fig 4.5.2**, yellow compound position), to connect two cavities together. Thr192 is located in the upper left side of the binding site and Glu261 is located in the lower right corner of the binding site. Binding pattern 1 mainly goes through the left groove and forms hydrogen bonds with Thr192, and binding pattern 2 mainly goes through the right groove and forms hydrogen bonds with Glu261 (**Fig 4.5.2**).



**Fig 4.5.1** Protein-ligand interaction fingerprint of top 1000 compounds targeting LGR5. (d=donor to backbone; D=donor to sidechain; a=acceptor from backbone; A=acceptor from sidechain; R=arene interaction)

Apart from a few backbone hydrogen bonds with Leu167, almost all other protein-ligand interactions are sidechain hydrogen bonds or aromatic interactions (**Fig 4.5.1**, bond type). To mimic the protein-protein interaction between R-spondin Phe106 and LGR5, the terminus that (**Fig 4.5.2**, upper left cavity) binds with Trp168/Thr192 cavity is chosen to be highly hydrophobic.



**Fig 4.5.2** Binding patterns of RSP-83 (yellow) and RSP-84 (light blue) binding with LGR5. Two pairs of major hydrogen bonds were zoomed up respectively. Hydrogen bond with Thr192 and Glu261 were represented by black line.

Eight compounds (**Table 4.5.1**) were selected in the first round of compound selection - four compounds belong to the binding pattern 1 (**Fig 4.5.2**, yellow compound binding pose) and four compounds belong to the binding pattern 2 (**Fig 4.5.2**, blue compound binding pose).

**Table 4.5.1** Name, ZINC ID of eight compounds of LGR5 compounds and their corresponding molecular docking scores and rankings among the top 1000 compounds.

Ligand name	Compound ZINC ID	Molecular docking score	Ranking (top 1000)	Binding pose
RSP-78	ZINC000019317254	-13.2	4	1
RSP-79	ZINC001281562693	-13.3	3	1
RSP-80	ZINC001601137247	-13.9	1	2
RSP-81	ZINC001190629113	-13.0	6	1
RSP-82	ZINC000468484311	-12.1	47	2
RSP-83	ZINC000513636966	-11.9	90	2
RSP-84	ZINC001565817202	-11.6	199	1
RSP-85	ZINC000174100547	-12.6	12	2

#### **4.5.2 Ligand Binding stabilities**

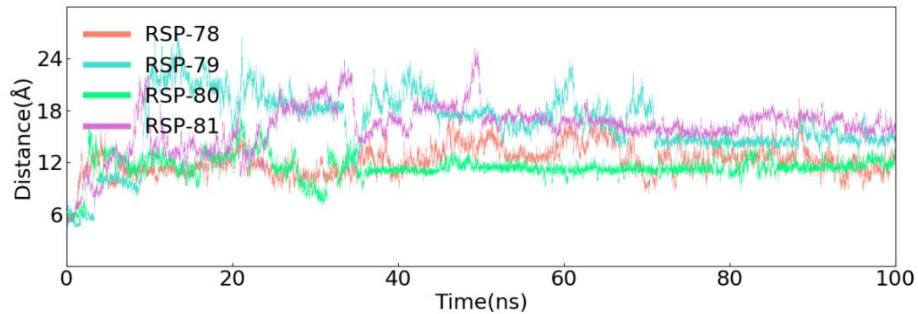
The overall performance of LGR5 ligands was not as good as for ligands of the other three protein/conformation targets. The minimum RMSF of compounds was 2.3 Å (**Table 4.5.2**). Because the cavity of LGR5 is shallow and ligands were not buried by the protein, ligand positional shifts were higher - half of ligand positional shifts were larger than 10 Å. In this case, the standard deviation (SD) is more helpful when determining the binding stability of compounds. From previous MD simulations, the SD of the most stable binding compounds are under 1 Å; for example, the SDs of RSP-05 and RSP-40 (**Closed** R-spondin) were 0.4 Å in both cases; the SD of RSP-60 (**Open2** R-spondin) was 0.9 Å and the SD of RSP-20 (RNF43) was 0.2 Å.

**Table 4.5.2** Binding stability properties of eight LGR5 ligands in the first-round selection.

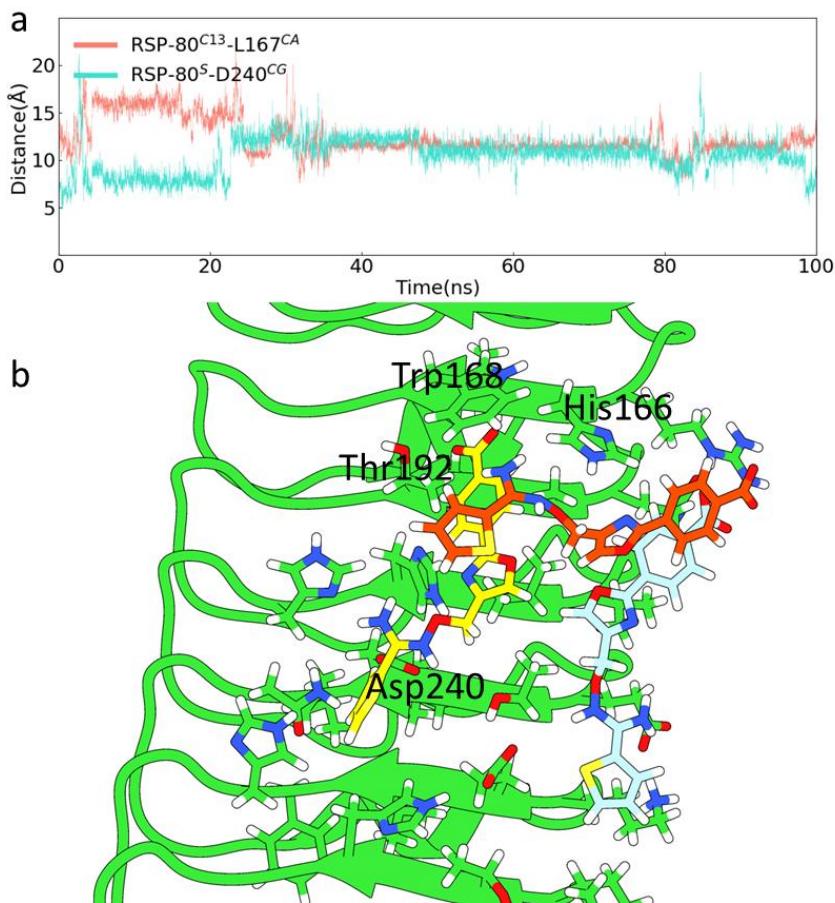
Ligand name	RMSD of ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
<b>RSP-78</b>	2.6	5.7	12.2	1.8
<b>RSP-79</b>	3.3	7.4	16.5	3.7
<b>RSP-80</b>	2.0	4.5	11.4	1.4
<b>RSP-81</b>	3.0	7.2	16.0	2.7
<b>RSP-82</b>	2.3	2.3	8.7	0.7
<b>RSP-83</b>	1.2	3.4	5.7	2.2
<b>RSP-84</b>	2.7	4.7	8.5	1.2
<b>RSP-85</b>	2.4	3.3	9.0	1.0

The ligand positional shifts of RSP-78, RSP-79, RSP-80 and RSP-81 were over 10 Å and their RMSFs were over 4.5 Å which means that these ligands demonstrated high flexibility. As the ligand positional shifts are shown in **Fig 4.5.3**, the ligand positional shift values of RSP-78, RSP-79 and RSP-81 keeps fluctuating during the simulation and only RSP-80 stayed in a stable distance suggesting it might bind with one position stably on LGR5. However, RSP-80 demonstrated poor but interesting binding property. In the following context, the moiety that binds with Trp168/Thr192 cavity is called the first cavity and the moiety that binds with Asp240/Glu261 cavity is called the second moiety. RSP-80 has carboxyl group on the first moiety and this compound was to test if hydrophilic group could bind with Trp169/Thr192 cavity. However, the benzoate group fell off the cavity early in the MD simulation (**Fig 4.5.4 (a)**); and also, the moiety that binds with Glu261 cavity

also fell off. After 40 ns of MD simulation, the ligand positional shift of RSP-80 maintained at 11.4 Å and its standard deviation was 0.7 Å; this suggests that RSP-80 could bind stably with LGR5 during this time (**Fig 4.5.4 (b)**). It is worth noticing that in the last 5 ns, the thiophene group started to move around again.



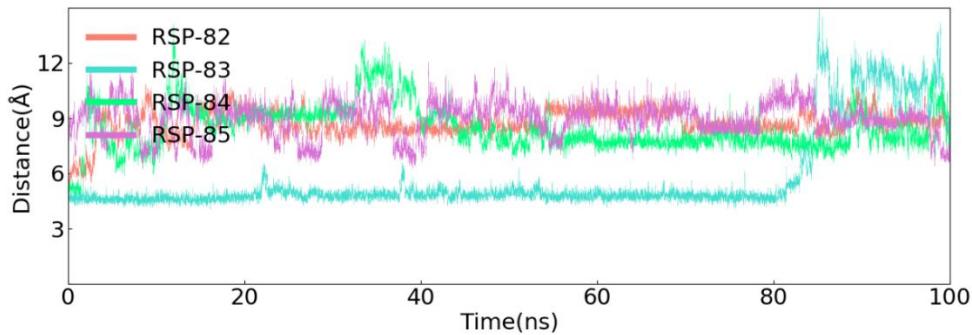
**Fig 4.5.3** Ligand positional shift of RSP-78, RSP-79, RSP-80 and RSP-81 from their binding site.



**Fig 4.5.4** a) Time series of distances between RSP-80 and two major cavities. b) Superposition of initial binding pose (yellow), 60 ns binding pose (blue), 100 ns binding pose (orange).

RSP-78, RSP-79 and RSP-81 failed to pass the MD-based assay. Their RMSFs were 5.7 Å, 7.4 Å and 7.2 Å respectively, indicating that they were highly flexible; their ligand positional shifts (**Fig 4.5.3**) confirm that they cannot bind stably with any position on LGR5.

The average ligand positional shifts of RSP-82, RSP-83, RSP-84 and RSP-85 were under 10 Å (**Table 4.5.2**); ligand positional shifts of these compounds as a function of time are plotted in **Fig 4.5.5** and these compounds will be introduced separately. Based on the computed descriptors, the performance of RSP-82 was the best among the eight compounds; its RMSF was only 2.3 Å which is the lowest. Although the positional shift is 8.7 Å, the standard deviation of ligand positional shift was 0.7 Å, also the lowest. Glu261 is an important hydrogen bond acceptor for RSP-82 which formed over 98% time of hydrogen bonds (**Table 4.5.3**). However, during the course of simulations, RSP-82 moved to a site beside the original cavity, a site which binds to Phe110 on R-spondin. RSP-82 will be discussed in the following **Special cases** section.

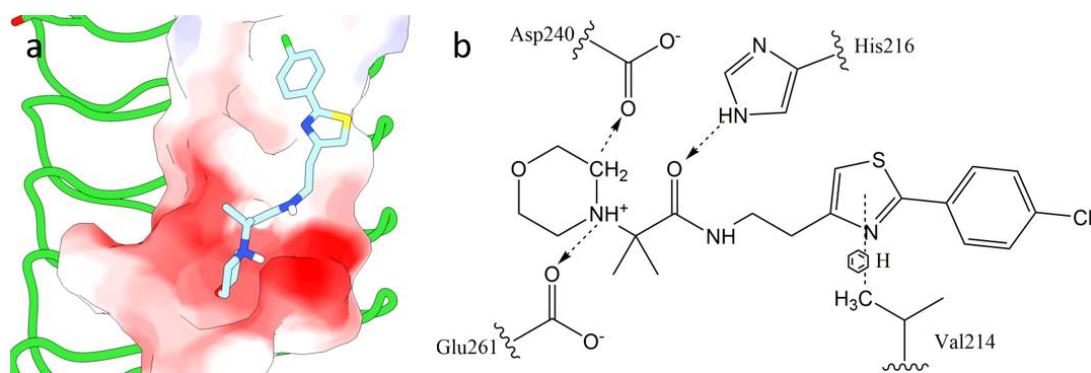


**Fig 4.5.5** Ligand positional shift of RSP-82, RSP-83, RSP-84 and RSP-85 from their binding site.

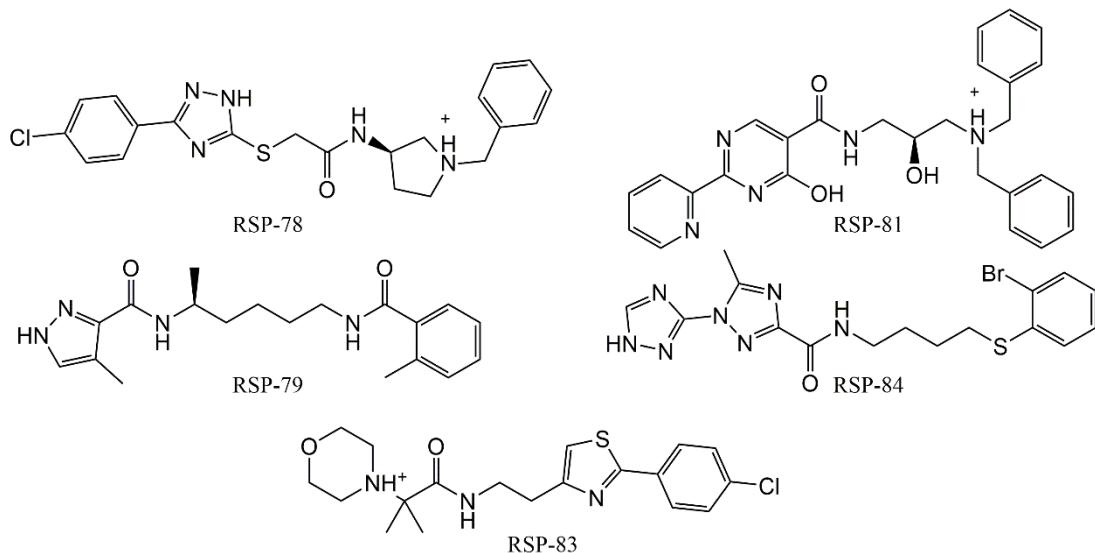
**Table 4.5.3** Highly populated hydrogen bond interactions in RSP-82, RSP-82 and RSP-85 MD simulation.

Ligand name	Acceptor	Donor	Fraction
<b>RSP-82</b>	E261 <sup>OE</sup>	RSP-82 <sup>N2</sup>	98%
	E261 <sup>OE</sup>	RSP-82 <sup>N3</sup>	83%
<b>RSP-83</b>	E261 <sup>OE</sup>	RSP-83 <sup>N2</sup>	73%
	RSP-83 <sup>O</sup>	H216 <sup>NE2</sup>	67%
<b>RSP-85</b>	RSP-83 <sup>O1</sup>	N242 <sup>ND2</sup>	15%
	-	-	-

Compared with RSP-82, RSP-84 and RSP-85, RSP-83 demonstrated an interesting binding feature (**Fig 4.5.5**). In the first 80 ns of simulation, the ligand positional shift was maintained at 4.8 Å and the standard deviation of ligand positional shift was 0.3 Å; this structure was the same as the initial binding pose (**Fig 4.5.6 (a)**). However, the ligand positional shift increased to 6.0 Å and its standard deviation increased to 2.4 Å in the final 20 ns of MD simulation. RSP-83 also formed 73% time of hydrogen bond with Glu261 suggesting that Glu261 plays an important role in protein-ligand maintenance. RSP-83 possesses a relatively short backbone compared with other failed compounds (**Fig 4.5.7**) suggesting that shorter backbones might be favoured for protein-ligand binding.



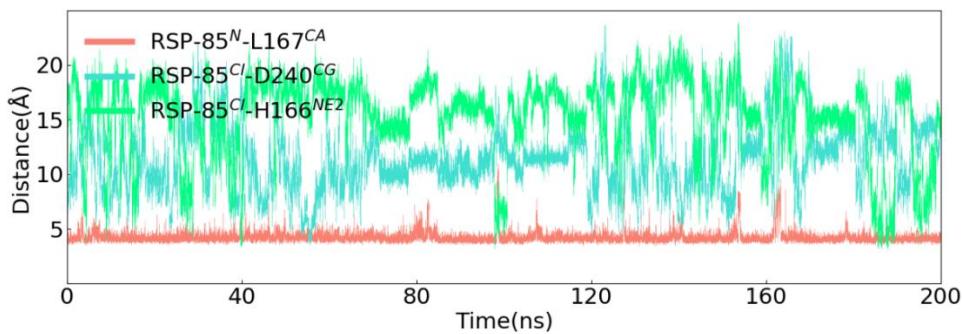
**Fig 4.5.6** **a)** Binding pattern of RSP-83 binding to LGR5 at 80 ns of simulation; intermolecular surface colored by electrostatics. **b)** Schematic molecular interactions between LGR5 and RSP-83.



**Fig 4.5.7** Structure of RSP-83 and four failed compounds: RSP-78, RSP-79, RSP-81 and RSP-84.

RSP-85 also demonstrated interesting binding behaviour. The benzonitrile group on RSP-85 binds well to the cavity near Thr192; however, the second moiety that binds with the cavity near Glu261 was not stable. Although the second moiety was unstable, it did not pull the benzonitrile group out of the cavity (**Fig 4.5.8**, pink line). Also, there is not significant hydrogen bonding that forms for more than 10% of simulation (**Table 4.5.3**). These facts suggest that the benzonitrile group contributed most to the protein-ligand binding of RSP-85. After 70 ns of simulation, the distance between RSP-85 chloride atom and Asp240/His166 reached relative

stability, in particular, near 100 ns, the benzyl chloride group binds to the cavity near His166. Hence, an extra 100 ns simulation was performed to further test the stability 100ns binding pose. From the 200 ns simulation, it is clear that the benzonitrile group could bind well with the Leu167 cavity and the second moiety could neither bind with the Asp240 cavity nor with the His166 cavity.



**Fig 4.5.8** Time series of three sets of distances between two major structures and their corresponding binding pockets. The distance between nitrogen from benzonitrile group and Leu167  $\alpha$ -carbon was to measure the distance to the Thr192 cavity. The distance between ligand chloride atom and Asp240 CG atom on sidechain was to measure the distance between the benzyl chloride moiety and Asp240 cavity and the distance between the chloride atom was to measure the distance between benzyl chloride moiety and His166 cavity.

RSP-84 failed to bind well to LGR5 in the MD simulation because it floated around the pocket (**Fig 4.5.4**); it did not stably bind with LGR5 in any pose.

### 4.5.3 Ligand structure refinement

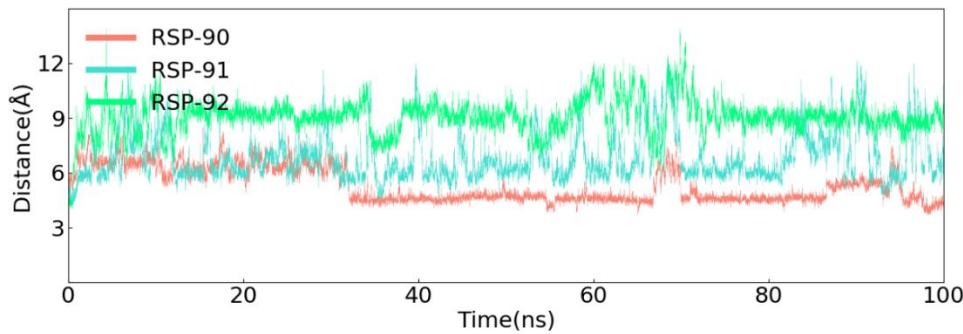
In the previous section, RSP-85 demonstrated an interesting property as the benzonitrile group attaches to the cavity near Thr192 firmly, regardless of other fluctuating moieties. Hence, in this section, more compounds with cyano groups were selected. Also, since RSP-83 binds stably with the LGR5 over the first 80 ns and

it has a halogen atom in the moiety that binds with the Thr192 cavity, this suggests that halogen atoms could also help stabilize protein-ligand interactions.

**Table 4.5.4** Binding stability properties of nine LGR5 ligands in the second-round selection.

Ligand name	RMSD of Ligand (Å)	RMSF of ligand (Å)	Ligand positional shift (Å)	SD of ligand positional shift (Å)
<b>RSP-86</b>	0.7	2.3	6.3	1.3
<b>RSP-87</b>	3.1	21.7	18.8	14.9
<b>RSP-88</b>	1.7	4.6	6.5	2.1
<b>RSP-89</b>	1.8	2.7	8.9	1.1
<b>RSP-90</b>	2.2	2.1	5.3	0.9
<b>RSP-91</b>	2.3	3.5	6.8	1.3
<b>RSP-92</b>	2.6	3.2	9.1	1.0
<b>RSP-93</b>	3.4	20.6	28.0	9.9

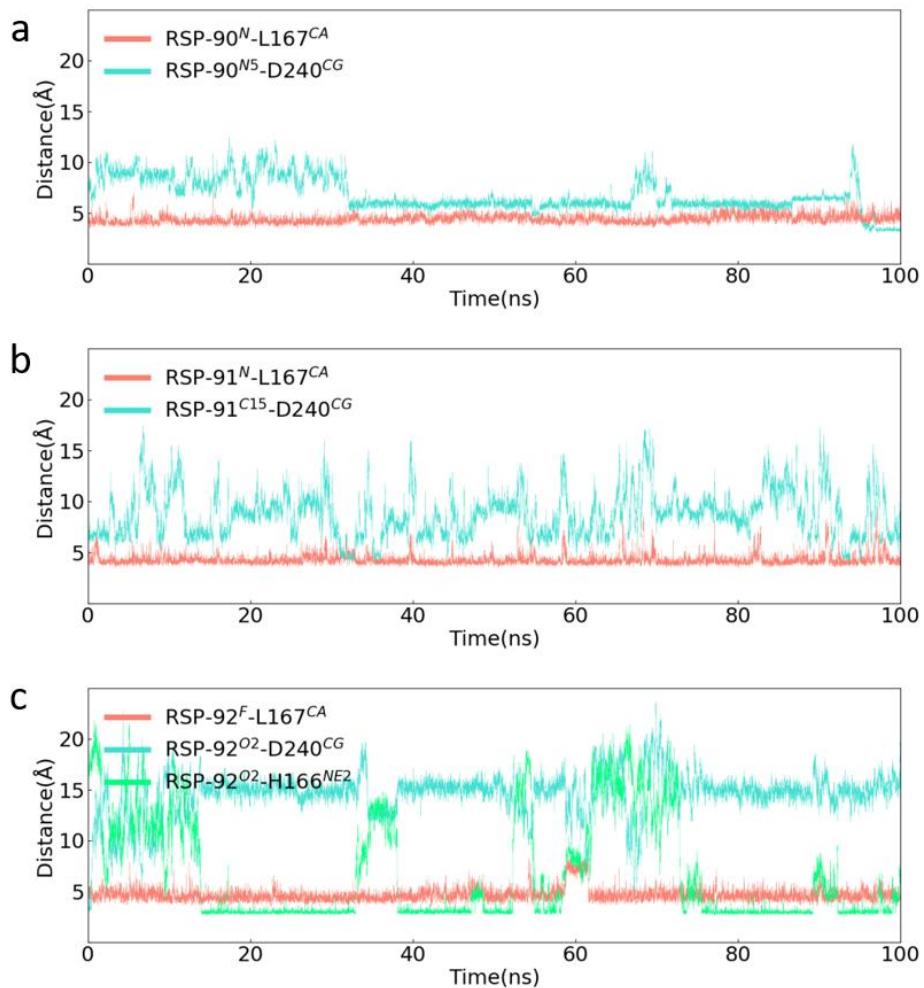
RSP-90, RSP-91 and RSP-92 were three representative compounds to test the contribution of cyano and halogen groups to binding the cavity near Thr192. The RMSF values of these compounds were 2.1 Å, 3.5 Å and 3.2 Å respectively. Also, the standard deviations of ligand positional shift of these compounds were at 0.9 Å, 1.3 Å and 1.0 Å, which means that they could bind with LGR5 in a fixed position (**Fig 4.5.9 and Table 4.5.4**).



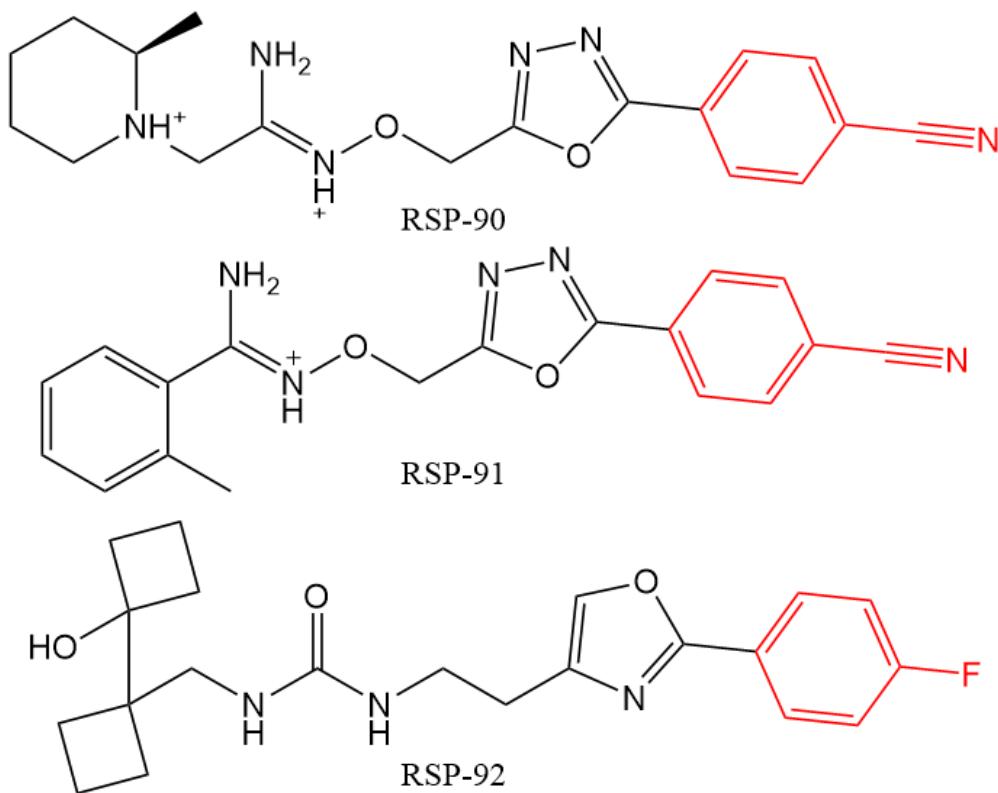
**Fig 4.5.9** Ligand positional shift of RSP-90, RSP-91 and RSP-92 from their binding site.

The distance between the cyano/halogen groups of RSP-90, RSP-91 and RSP-92 and Trp168 cavity remained relatively short (**Fig 4.5.10**, pink line in three figures). In RSP-91, the moiety that binds to the Glu261 cavity was unstable (**Fig 4.5.10 (b)**). Significantly, RSP-91 also did not form any hydrogen bond with an occupancy of more than 10% (**Table 4.5.5**). At the beginning of RSP-90 simulation, the second moiety in Asp240 cavity was unstable (**Fig 4.5.10 (b)**). However, after 30 ns, the distance decreased to 5.8 Å; and after 95 ns, the distance further decreased to 3.6 Å. Plus the distance between the first moiety and Leu167 cavity maintained at around 4.5 Å indicating that RSP-90 binds stably with LGR5. RSP-90 could form hydrogen bond with Asp240 for 73% of time (**Table 4.5.5**). The protein-ligand binding of RSP-92 is unusual because the second moiety binds with His166 for nearly half of the simulation time. His166 is located right beside the Thr192 cavity which binds with Phe110 of R-spondin. This suggests that compounds might be able to bind with the Thr192 cavity and His166 cavity at the same time (**Fig 4.5.10 (c)**). From these three compounds, we could confirm that cyano and halogenic group could help ligand bind with Trp168 cavity, the structure of RSP-90, RSP-91 and RSP-92 are listed in **Fig 4.5.11**. Based on the tautomer ranking of the RSP-90 and RSP-91,

the main tautomer is shown in the figure, especially, the ring in RSP-90 is a R-isomer. The tautomer ranking is performed by the **tautomers** module in the QUACPAC package of the OpenEye software<sup>70</sup>.

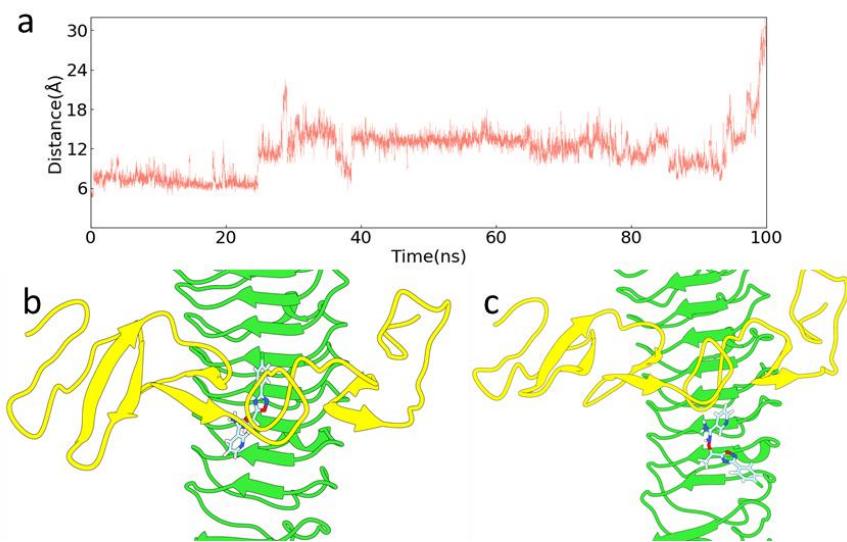


**Fig 4.5.10** The distances between RSP-90, RSP-91 and RSP-92 atoms and their corresponding binding cavities. The CA atom in Leu167 was to measure the distance between cyano group/halogen group to Thr192 cavity. The CG atom in Asp240 was to measure the distance between the opposite moiety and Glu261 cavity distance. Specifically, RSP-92 binds with the cavity near His166, it was also measured by the distance between O2 atom of RSP-92 and NE2 atom on His166.



**Fig 4.5.11** 2D structures of RSP-90, RSP-91 and RSP-92. The moieties that binds with Thr192 cavity coloured red.

RSP-87 and RSP-93 failed to bind well in the MD simulation: the ligand positional shifts of RSP-87 and RSP-93 were 18.8 Å and 28.0 Å and standard deviations were 14.9 Å and 9.9 Å respectively, confirming that they could not bind with LGR5. RSP-88 also failed the MD assay; its RMSF values was 4.7 Å and SD of ligand positional shift was 2.1 Å. More importantly, the moiety that binds with the cavity near Thr192 detached (**Fig 4.5.12 (a)**) and binds to another position far from the binding interface between R-spondin and LGR5 (**Fig 4.5.12 (c)**).



**Fig 4.5.12** a) Time series of distance from RSP-80 Thr192-cavity-binding moiety to its pocket. Distance between chloride atom and the  $\alpha$ -carbon of Leu167. Initial (b) and final (c) binding pattern of RSP-88. R-spondin coloured yellow, LGR5 coloured green and RSP-88 coloured blue.

Two compounds, RSP-86 and RSP-89, demonstrated different binding properties compared with other compounds, and they will be discussed in the following

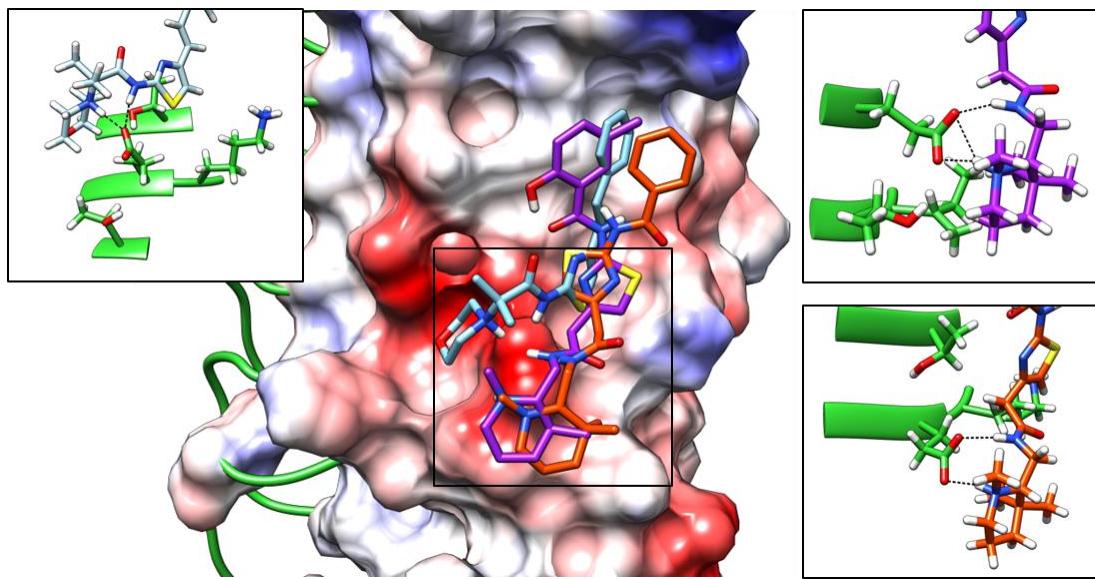
**Special cases** section.

**Table 4.5.5** Highly populated hydrogen bonds between RSP-86, RSP-89, RSP-90, RSP-91 and RSP-92 and LGR5.

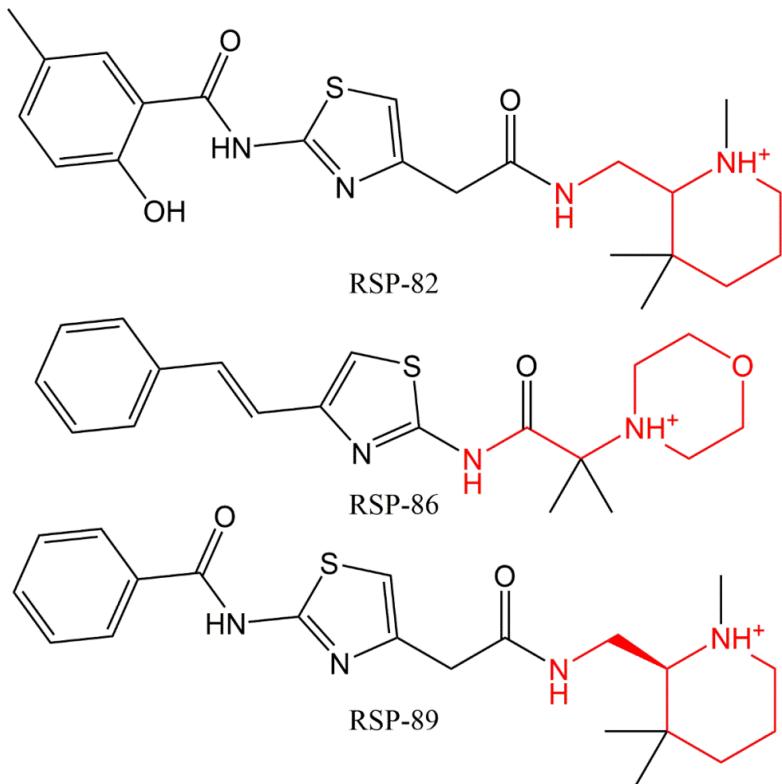
Ligand name	Acceptor	Donor	Fraction
RSP-86	E261 <sup>OE</sup>	RSP-86 <sup>N2</sup>	100%
	E261 <sup>OE</sup>	RSP-86 <sup>N1</sup>	83%
	RSP-86 <sup>O</sup>	H216 <sup>HE</sup>	18%
RSP-89	E261 <sup>OE</sup>	RSP-89 <sup>N3</sup>	98%
	E261 <sup>OE</sup>	RSP-89 <sup>N2</sup>	81%
RSP-90	E261 <sup>OE</sup>	RSP-90 <sup>N2</sup>	73%
	RSP-90 <sup>O</sup>	H216 <sup>NE2</sup>	67%
	RSP-90 <sup>O1</sup>	N242 <sup>ND2</sup>	15%
RSP-91	-	-	-
RSP-92	RSP-92 <sup>O2</sup>	H166 <sup>NE2</sup>	32%

#### 4.5.4 Special cases

RSP-82, RSP-86 and RSP-89 shared a great similarity in their behaviour during the MD simulations. Firstly, they all formed highly occupied hydrogen bonds with Glu261 (**Fig 4.5.13**). RSP-82 formed two hydrogen bonds for 98% and 83% of simulation time respectively (**Table 4.5.3**); RSP-86 formed two hydrogen bonds for 100% and 83% of the time and RSP-89 formed two hydrogen bonds for 98% and 81% of the time (**Table 4.5.5**). Secondly, these compounds share similar backbones and similar key structures to bind with Glu261 (**Fig 4.5.14**). Finally, all of them lost their binding to the Thr192 cavity and slipped into the pocket next to the original cavity, eg. the His166 cavity which binds with Phe110 of R-spondin.



**Fig 4.5.13** Superposition of RSP-82, RSP-86 and RSP-89 on the ligand binding site of LGR5. Structures that formed hydrogen bonds with Glu261 are inset; hydrogen bonds represented by black dotted lines. The orange ligand is RSP-82, the light blue structure is RSP-86 and the purple ligand is RSP-89.



**Fig 4.5.14** 2D structures of RSP-82, RSP-86 and RSP-89. Key structures that formed hydrogen bonds with Glu261 coloured red.

## 5. Free energy calculation

From the previous chapter, the protein binding stability of 93 compounds was tested by molecular dynamics simulations. The main goal for this chapter is to select compounds with the best calculated binding affinity based on these simulations. To achieve this, the MMGBSA method was used to calculate the binding free energy of successful compounds selected in Chapter 4. Energy decomposition was also performed to assess the contributions to binding by individual protein residues.

In the previous MD simulations, some compounds failed to bind with R-spondin proteins; hence their free energy was not calculated. Here, 43 compounds were selected to compute their binding affinities by MMGBSA (**Table 5.1**)

**Table 5.1** Number of compounds selected for free energy calculation by protein target/conformation.

Protein name	Number of compounds
<b><i>Closed</i> R-spondin</b>	9
<b><i>Open2</i> R-spondin</b>	13
<b>RNF43</b>	11
<b>LGR5</b>	10

### 5.1 Methods

The whole trajectory of 100 ns protein-ligand complex simulation was used for free energy calculation with Generalized Born model developed by Onufriev et al<sup>71</sup> and

Amber ff14SB force field. Cpptraj module in Amber 16 was used to prepare MMGBSA input files. MMPBSA.py.MPI module was used for multi-core MMGBSA calculation. The interval of calculation was set to 10 ps and the salt concentration of the system was set to 0.1 M.

To obtain further insight into the energy contribution of residues to protein-ligand interaction, energy decomposition was also performed during the free energy calculation. Four terms including electrostatics interaction energy ( $\Delta G_{elec}$ ), Van der Waals interaction energy ( $\Delta G_{vdw}$ ), polar solvent energy ( $\Delta G_{polar}$ ), and non-polar energy ( $\Delta G_{non-polar}$ ) were calculated to characterize residue contribution.

$$\Delta G_{residue} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{polar} + \Delta G_{non-polar}$$

## 5.2 Binding free energy and energy decomposition

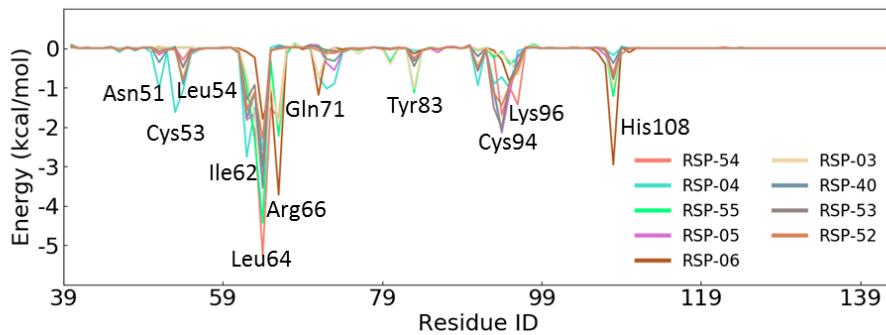
### 5.2.1 *Closed R-spondin*

Nine compounds that demonstrated reasonable binding stabilities with *Closed R-spondin* were selected for computation of binding free energy and energy decomposition by residue (**Table 5.2**). Among these nine compounds, RSP-54 had the best binding free energy reaching -36.2 kcal/mol. The binding free energies of the following compounds including RSP-04, RSP-55, RSP-05, RSP-06, RSP-03 and RSP-40, and ranged from -26.2 kcal/mol to -29.6 kcal/mol.

**Table 5.2** Binding free energy and standard deviation of nine *Closed R-spondin* ligands.

Compound	Free energy (kcal/mol)	Standard deviation (kcal/mol)
RSP-54	-36.2	4.3
RSP-04	-29.6	5.9
RSP-55	-29.5	3.7
RSP-05	-28.7	4.2
RSP-06	-27.8	6.2
RSP-03	-26.8	4.1
RSP-40	-26.2	2.7
RSP-53	-24.0	3.7
RSP-52	-23.0	4.6

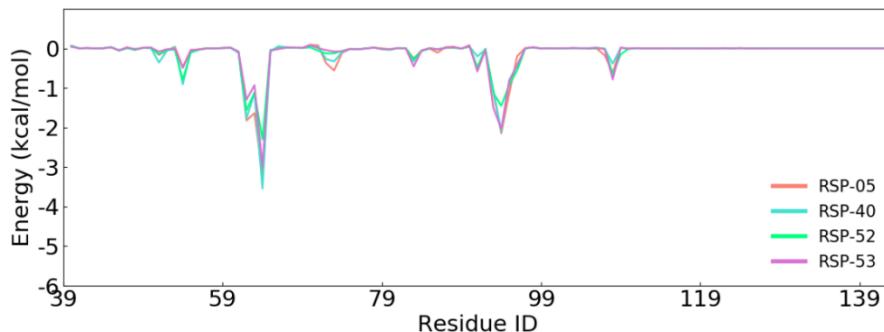
From energy decomposition analysis (**Fig 5.2.1**), several clusters of residues appeared to be important for protein-ligand binding free energy. The most important residue cluster is from Ile62 to Arg66 with the most important residue being Leu64. The second residue cluster is from Lys93 to Lys96 and the most important residue is Cys94. Residues near Leu54, Tyr83 and His108 also appeared in the energy decomposition of compounds; however, they are not as significant contributors to binding as the two residue clusters mentioned above (**Fig 5.2.1**).



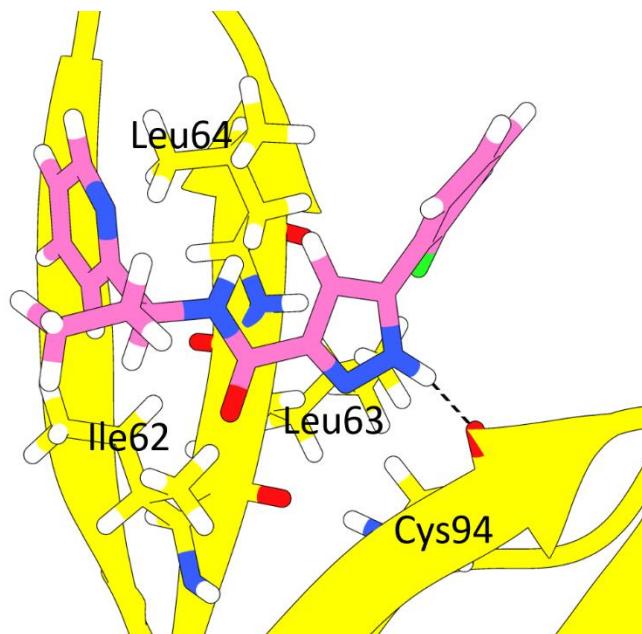
**Fig 5.2.1** Free energy contribution per residues in **Closed** R-spondin ligand MD simulations. Major residues marked on figure.

From MD simulation, RSP-05, RSP-40, RSP-52 and RSP-53 share the same binding mode (**Fig 4.2.8**, **Fig 4.2.9** and **Fig 4.2.10**); and here, from energy decomposition, as expected they also share the same key residues that contribute to their binding free energy (**Fig 5.2.2**). The most important residue is Leu64; the second important residue is Cys94, and the third and the fourth residue are Ile62 and Leu63 respectively. This is in accordance with the core structure of binding mode-1 compounds (**Fig 4.2.5**): Leu64, Ile62, and Leu63 form a hydrophobic binding interface between the compounds' hydrophobic ring structures and **Closed** R-

spondin (**Fig 5.2.3**); the backbone nitrogen of Leu64 forms a NH- $\pi$  bond; Cys94 forms the major hydrogen bonds with these ligands (**Table 4.2.3** and **Table 4.2.6**).



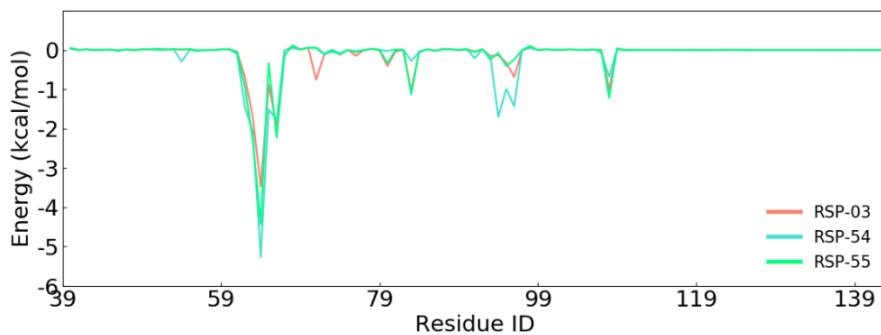
**Fig 5.2.2** Energy decomposition of RSP-05 and RSP-40, RSP-52 and RSP-53 showing the free energy contribution as a function of residue number.



**Fig 5.2.3** Schematic representation of RSP-40 binding with Ile62, Leu63, Leu64 and Cys94. Hydrogen bond between RSP-40 and Cys94 marked by black line.

Leu64 contributes the most to the protein-ligand binding stability of RSP-03, RSP-54 and RSP-55. The contribution of Leu63 and Arg66 ranks second or third place among all residues. This is in accordance with the hydrogen bond analysis (**Table 4.2.3** and **Table 4.2.6**): RSP-03 only forms 20% hydrogen bonds with **Closed R-**

spondin and RSP-55 only had 37% hydrogen bond occupancy, which suggests the hydrophobic interactions near Leu64 play an important role in stabilizing protein-ligand binding.



**Fig 5.2.4** Energy decomposition of RSP-05 and RSP-40, RSP-52 and RSP-53 showing the energy contribution of residues as a function of residue number.

RSP-06 and RSP-04 were two special cases - they bind with **Closed** R-spondin in a unique manner and their residue energy contribution fingerprint were also different from other compounds.

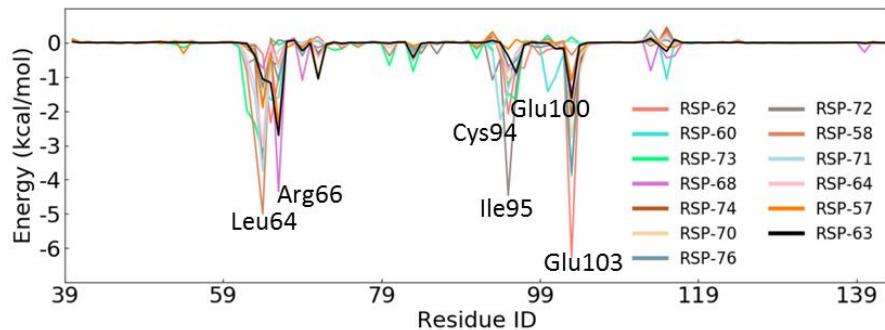
### 5.2.2 *Open2 R-spondin*

Binding free energy and energy decomposition were performed for thirteen compounds that showed acceptable binding stability with **Open2** R-spondin in MD simulation. Among those compounds, RSP-62 demonstrated the best binding affinity with a computed value of -41.2 kcal/mol (**Table 5.3**). The binding free energies of four compounds RSP-60, RSP-73, RSP-68 and RSP-74 were over -30 kcal/mol; and five compounds RSP-70, RSP-76, RSP-72, RSP-58 and RSP-71 were over -25 kcal/mol (**Table 5.3**).

**Table 5.3** Binding free energy and standard deviation of thirteen *Open2* R-spondin ligands.

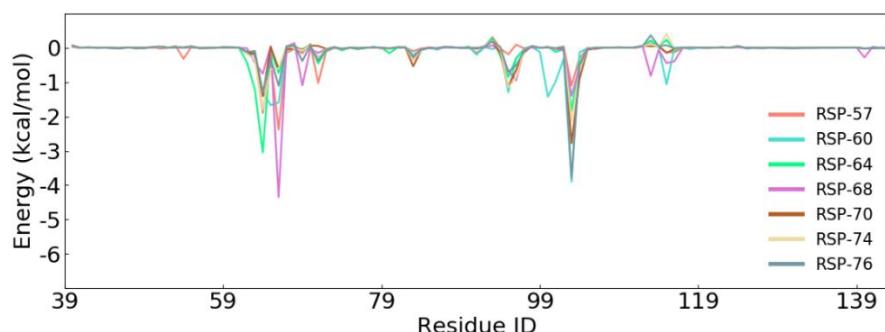
Compound	Free energy (kcal/mol)	Standard deviation (kcal/mol)
RSP-62	-41.2	5.3
RSP-60	-34.3	5.1
RSP-73	-33.8	3.1
RSP-68	-33.0	5.3
RSP-74	-32.3	4.9
RSP-70	-29.2	5.5
RSP-76	-28.6	7.2
RSP-72	-27.0	5.6
RSP-58	-26.6	3.4
RSP-71	-25.3	3.2
RSP-64	-24.9	4.6
RSP-57	-24.2	6.1
RSP-63	-20.8	4.1

There were some similarities between the energy decomposition of *Closed* R-spondin and *Open2* R-spondin: the residue cluster near Leu64 and Cys94 also plays an important role in the protein-ligand binding of *Open2* R-spondin. Apart from these two residue groups, Glu103 in the Furin-like 2 domain also plays an important role (Fig 5.2.5).



**Fig 5.2.5** Free energy decomposition per residues in **Open2** R-spondin ligand MD simulation. Major residues had been marked on the figure.

Seven compounds including: RSP-57, RSP-60, RSP-64, RSP-68, RSP-70, RSP-74 and RSP-76, share the same key residues. Three residue groups near Leu64 (located in Furin-like-1 domain), Ile95 (located in the hinge area linking two Furin-like domains) and Glu103 (located in Furin-like 2 domain) are important for the binding stability of these compounds. **Fig 5.2.6** shows that the contribution of residues near Leu64 and Glu103 are much higher than for Ile95, suggesting that the energy contribution of Furin-like-1 and -2 domains were higher than residues in the hinge area. RSP-60 is one of the most typical compounds that did not alter the R-spondin conformation, detailed energy decomposition of RSP-60 was listed in **Table 5.4**.

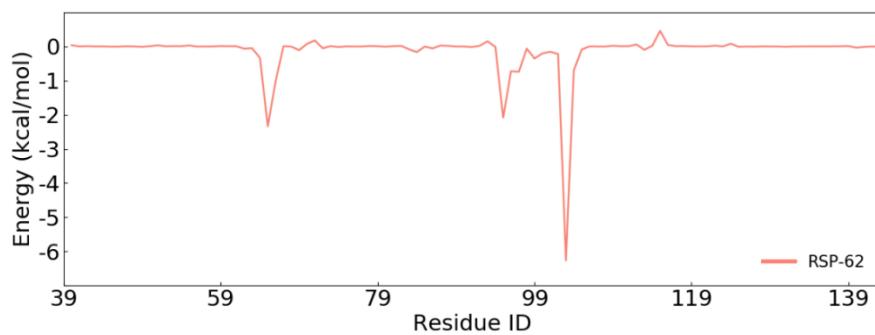


**Fig 5.2.6** Decomposition of protein-ligand binding free energies in the simulations of RSP-57, RSP-60, RSP-64, RSP-68, RSP-70, RSP-74 and RSP-76.

**Table 5.4** Energy of residues with highest contribution in RSP-60/*Open2* R-spondin simulation. Energies in kcal/mol.

Residue	$\Delta G_{vdw}$	$\Delta G_{elec}$	$\Delta G_{polar}$	$\Delta G_{non-polar}$	$\Delta G_{total}$
<b>Glu103</b>	-1.6	-7.7	5.7	-0.3	-3.9
<b>Glu65</b>	-1.6	-0.4	0.5	-0.2	-1.7
<b>Arg66</b>	-2.6	0.4	0.9	-0.3	-1.6
<b>Glu100</b>	-0.7	-6.7	5.6	-0.1	-1.4
<b>Ile95</b>	-1.4	-0.1	0.4	-0.2	-1.3
<b>Leu64</b>	-2.0	-0.5	1.5	-0.3	-1.3
<b>Lys115</b>	-1.3	-4.6	5.0	-0.2	-1.1
<b>His101</b>	-1.9	0.1	1.0	-0.2	-1.0

RSP-62 was a special case that binds well with Furin-like-2 domain but is far from Furin-like 1 domain. Hence, the contribution of Glu103 (located in the Furin-like-2 domain) was -6.3 kcal/mol (**Fig 5.2.7** and **Table 5.5**), which was far higher than the contribution of Glu103 for any other compound (**Fig 5.2.5**) and other residues in the RSP-62/*Open2* R-spondin simulation.

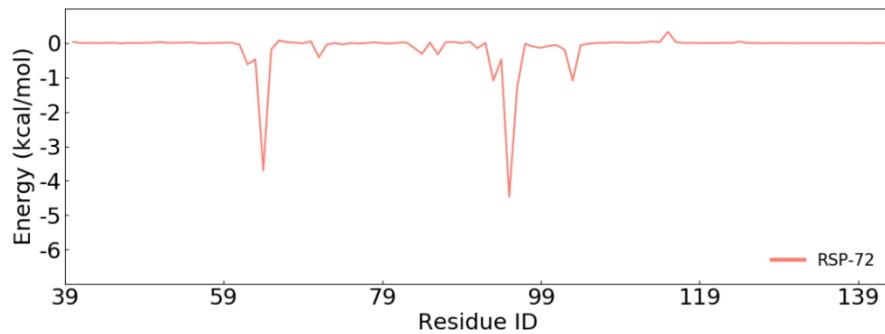


**Fig 5.2.7** Decomposition of protein-ligand binding free energies in the simulations of RSP-62.

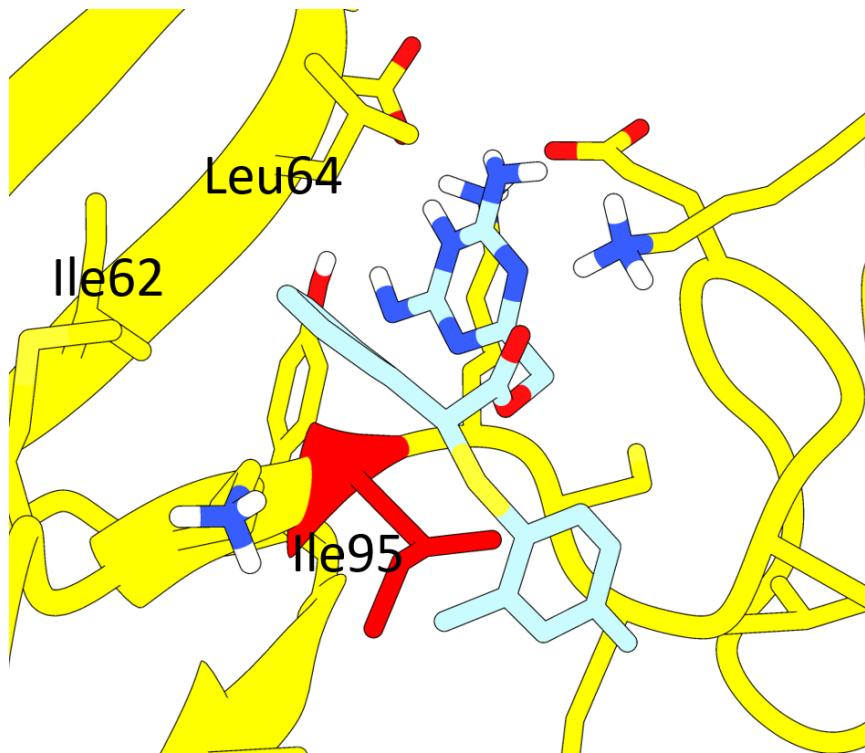
**Table 5.5** Energy of residues with highest contribution in RSP-62/**Open2** R-spondin simulation. Energies in kcal/mol.

Residue	$\Delta G_{vdw}$	$\Delta G_{elec}$	$\Delta G_{polar}$	$\Delta G_{non-polar}$	$\Delta G_{total}$
<b>Glu103</b>	-1.8	-56.0	51.9	-0.4	-6.3
<b>Glu65</b>	-0.5	-37.4	35.7	-0.2	-2.3
<b>Ile95</b>	-1.6	-0.4	0.2	-0.3	-2.1
<b>Aeg66</b>	-2.5	27.1	-25.2	-0.4	-1.0
<b>Cys97</b>	-1.3	-1.2	1.9	-0.1	-0.7
<b>Lys96</b>	-2.1	21.9	-20.3	-0.3	-0.7
<b>Ala104</b>	-0.6	-0.1	0.1	-0.1	-0.7
<b>Lys115</b>	-1.0	21.0	-19.3	-0.2	0.5

For most compounds, the contribution of Ile95 and its nearby residues was low; however, this was not the case in the RSP-72/**Open2** R-spondin complex simulation, where the highest affinity contributor was Ile95 (**Fig 5.2.8**). This may be because Ile95 provides a hydrophobic environment to bind the m-xylene group on RSP-72. Ile62 and Leu64 also provide a hydrophobic interface for the benzene ring in RSP-72 (**Fig 5.2.9**).



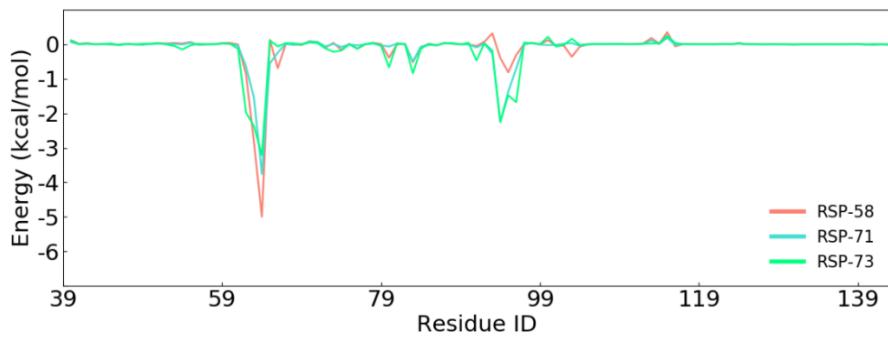
**Fig 5.2.8** Decomposition of protein-ligand binding free energies in the simulations of RSP-72.



**Fig 5.2.9** Key residues that contribute to the binding energy of RSP-72.

Two groups of residues contribute the most for the binding free energy of **Open2** R-spondin with RSP-58, RSP-71 and RSP-73: the first residue group is Ile62, Leu63 and Leu64; and the second residue group is Cys94, Ile95 and Lys96 (**Fig 5.2.10** and **Table 5.5**). The first group is near Leu63 and is located in the Furin-like 1 domain. The second group is located Ile95 in the hinge area. They do not form significant interactions with Glu103 (located in Furin-like 2 domain). RSP-58 is one of the most

representative compounds which binds with the furin-like-1 domain and went far from the Furin-like 2 domain. For these compounds, the protein-ligand affinities rely heavily on interactions between ligand and the Furin-like-1 domain (**Table 5.6**).



**Fig 5.2.10** Decomposition of protein-ligand binding free energies in the simulations of RSP-58, RSP-71 and RSP-73.

**Table 5.6** Energy of residues with highest contribution to in RSP-58/*Open2* R-spondin simulation. Energy in kcal/mol.

Residue	$\Delta G_{vdw}$	$\Delta G_{elec}$	$\Delta G_{polar}$	$\Delta G_{non-polar}$	$\Delta G_{total}$
<b>Leu64</b>	-2.8	-7.7	6.0	-0.5	-5.0
<b>Leu63</b>	-2.4	0.2	-0.5	-0.2	-2.8
<b>Ile62</b>	-0.9	-0.2	0.3	-0.1	-0.9
<b>Ile95</b>	-1.0	0.3	0.0	-0.1	-0.8
<b>Arg66</b>	-1.0	19.9	-19.3	-0.2	-0.7
<b>Tyr83</b>	-1.5	-0.2	1.5	-0.3	-0.5

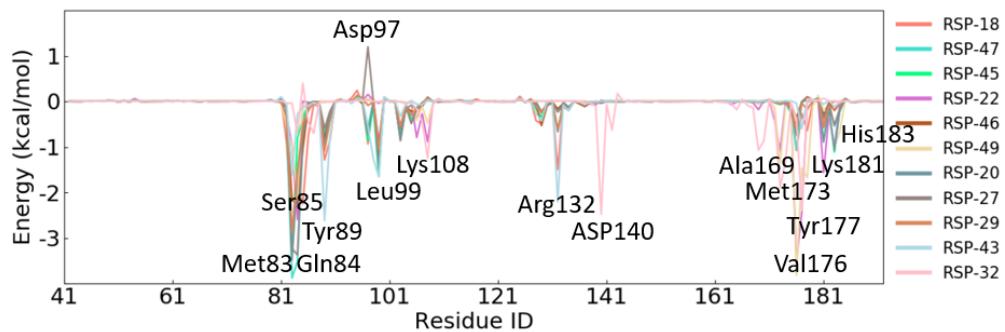
### **5.2.3 RNF43**

Compared with the binding free energies of R-spondin ligands, RNF43 ligands demonstrate higher computed binding affinities. The binding free energy and energy decomposition of eleven selected compounds were performed. Among these compounds, RSP-18 demonstrates the best binding affinity, with a binding free energy of -45.9 kcal/mol (**Table 5.7**). Significantly, the binding free energies of eight compounds, RSP-47, RSP-45, RSP-22, RSP-46, RSP-49, RSP-20, RSP-27, RSP-29, were under -30 kcal/mol. All compounds' binding free energies exceeded -25 kcal/mol (**Table 5.7**).

**Table 5.7** Binding free energy and standard deviation of eleven RNF43 ligands.

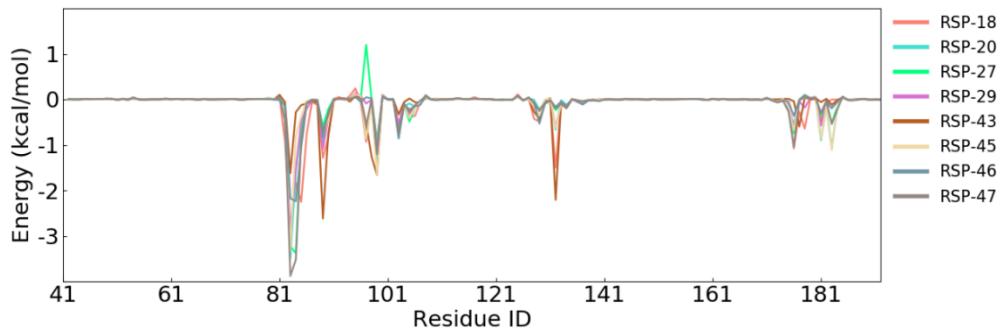
<b>Compound</b>	<b>Free energy (kcal/mol)</b>	<b>Standard deviation (kcal/mol)</b>
<b>RSP-18</b>	-45.9	7.1
<b>RSP-47</b>	-38.1	3.5
<b>RSP-45</b>	-36.9	3.5
<b>RSP-22</b>	-34.7	4.8
<b>RSP-46</b>	-33.8	4.5
<b>RSP-49</b>	-33.3	5.8
<b>RSP-20</b>	-32.8	3.5
<b>RSP-27</b>	-31.8	4.7
<b>RSP-29</b>	-30.6	4.6
<b>RSP-43</b>	-28.6	5.2
<b>RSP-32</b>	-26.1	8.5

Five major residue clusters near Met83, Tyr89, Asp97, Arg132 and Val176 contributed the most to the binding free energy of the selected compounds (**Fig 5.2.11**). The energy decomposition of these compounds is in accordance with their RNF43 binding modes proposed in **Chapter 4.3.1**. The first group includes binding mode-1 compounds; they share a great similarity in energy decomposition and likewise for binding mode-2 compounds.

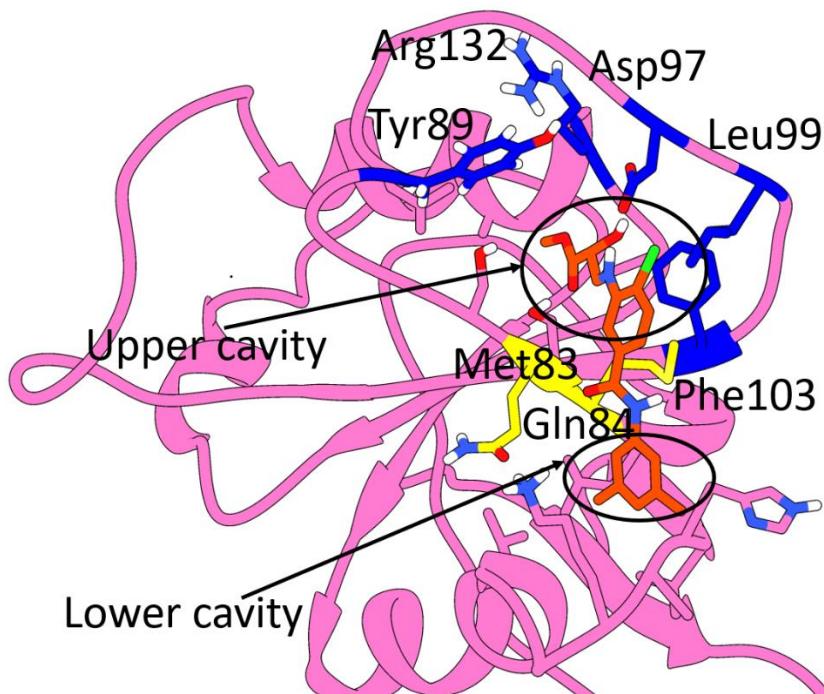


**Fig 5.2.11** Free energy decomposition per residues in RNF43 ligand MD simulations. Major residues marked on figure.

RSP-18, RSP-20, RSP-27, RSP-29, RSP-43, RSP-45, RSP-46 and RSP-47 belong to the binding mode 1 and they share a great similarity in energy decomposition (**Fig 5.2.12**). The most important residues are Met83 and Gln84 which are located in the middle of the upper and lower cavity and function as a groove for compounds to bind (**Fig 5.2.13**). Another three clusters of residues are near Tyr89, Leu99 and Arg132; these residues are all near the upper cavity of RNF43. Residues near His183 are near the lower cavity; however, their contributions to the compound binding free energy are relatively low. This suggests that the binding free energy contributions of the upper cavity and the hinge area are higher than that of the lower cavity.



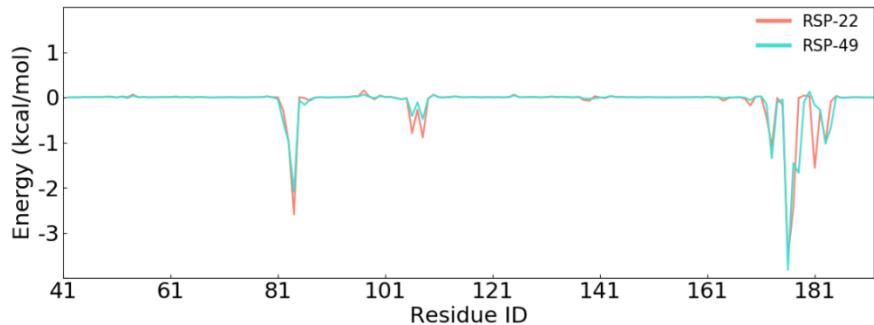
**Fig 5.2.12** Decomposition of protein-ligand binding free energies in the simulations of RSP-18, RSP-20, RSP-27, RSP-29, RSP-43, RSP-45, RSP-46 and RSP-47.



**Fig 5.2.13** RSP-20 (orange) binding to RNF43 (pink). Important residues labeled.

RSP-22 and RSP-49 successfully bound to protein in the MD simulation and belong to binding mode 2. The most important residue cluster that facilitates protein-ligand binding is Val167 and its proximal residues Met173, Tyr177, Lys178 and His183 (**Fig 5.2.14**). This may because these two compounds form two hydrogen bonds with Val167 (**Table 4.4.3** and **Table 4.4.5**). Also, they lie on the lower groove near Val176 (**Fig 4.4.12** and **Fig 4.4.15**) and hence can interact with residues near

Val176. The second key residue cluster is near Gln84, located in the lower cavity of RNF43; this binds with one end of RSP-22 and RSP-49.



**Fig 5.2.14** Decomposition of protein-ligand binding free energies in the simulations of RSP-22 and RSP-49.

The binding free energy of RSP-32 was poor and it was not stable enough to bind in one position with RNF43; hence its energy decomposition features were not informative.

### **5.2.4 LGR5**

Free energy calculation and energy decomposition of ten compounds which showed reasonable binding stabilities with LGR5 were performed. The binding free energies of LGR5 ligands were not as favourable as for the ligands mentioned above. Among the ten compounds, RSP-90 possessed highest binding affinity (**Table 5.8**); yet its binding free energy was only -33.0 kcal/mol; the second-highest compound binding free energy was only -30.2 kcal/mol. Apart from RSP-90 and RSP-83, the binding free energies of only five compounds including RSP-86, RSP-82, RSP-89, RSP-85 and RSP-92 were over -20 kcal/mol (**Table 5.8**).

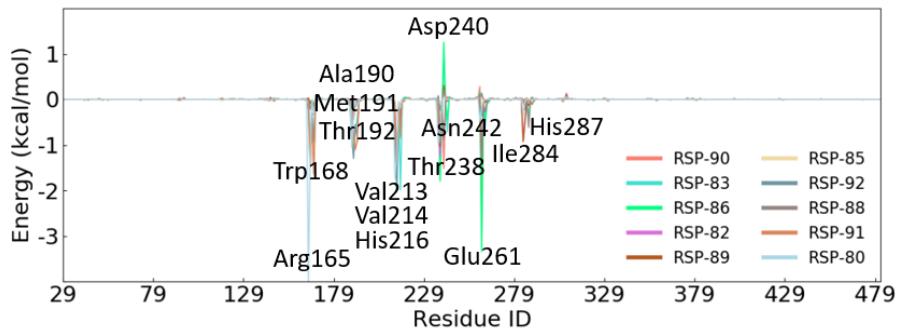
**Table 5.8** Binding free energy and standard deviation of ten LGR5 ligands.

Compound	Free energy (kcal/mol)	Standard Deviation (kcal/mol)
RSP-90	-33.0	5.8
RSP-83	-30.2	8.8
RSP-86	-28.0	5.6
RSP-82	-24.0	3.3
RSP-89	-23.2	3.2
RSP-85	-21.4	4.9
RSP-92	-20.5	4.8
RSP-88	-18.9	6.3
RSP-91	-18.6	5.0
RSP-80	-16.4	7.5

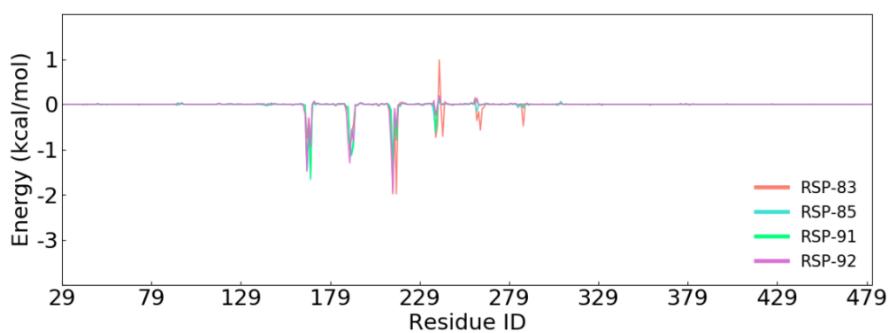
There are six clusters of key residues that contribute to the binding stabilities of LGR5 compounds; these residues are located near His166/Trp168, Met191/Thr192, Val214/His216, Thr238/Asp240, Glu261 and Gly283/Thr285 (**Fig 5.2.15**). Trp168 and Met191 are located near the binding cavity which binds with the Phe106 with R-spondin; and Asp240 and Glu261 are near the second binding cavity of binding mode 1 compounds (**Fig 4.5.2**).

RSP-83, RSP-85, RSP-91 and RSP-92 share the same energy decomposition features (**Fig 5.2.16**). There are four major groups of residues near His166/Trp168, Met191/Thr192, Val214/His216 and Thr238/Asp240. According to the hydrogen bond interaction analysis (**Table 4.5.3** and **Table 4.5.5**), RSP-85 and RSP-91 do not form any significant hydrogen bond that occupies more than 10% of the MD simulation; and RSP-92 only formed 32% of a hydrogen bond with His166, which was dependent on the fluctuation of RSP-92. This suggests that the hydrophobicity of the cavity near Trp168, Thr192 and Val214 is crucial for protein-ligand binding.

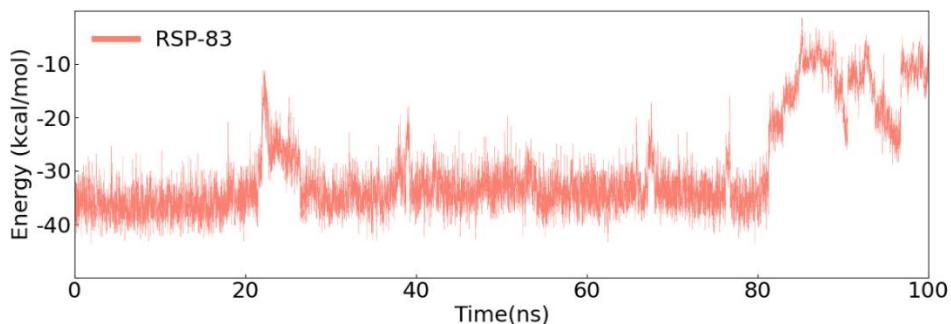
RSP-83 binds well with its binding site in the first 80 ns; however, it detached after this time. The time series of computed binding affinity (**Fig 5.2.17**) reflects this change. In the first 80 ns, the average binding free energy is around -33.9 kcal/mol which is similar to RSP-90, the highest-ranking compound. However, after 80 ns, RSP-83 moves out of the binding site and hence, the binding free energy reduces in magnitude dramatically (**Fig 5.2.17**).



**Fig 5.2.15** Free energy decomposition per residues in LGR5 ligand MD simulations. Major residues marked on figure.



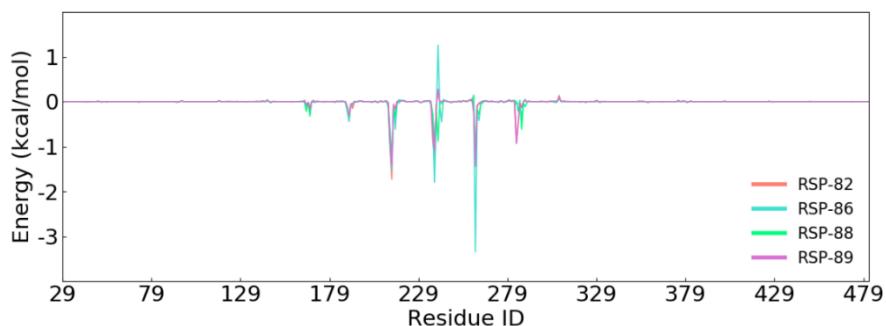
**Fig 5.2.16** Decomposition of protein-ligand binding free energies in the simulations of RSP-83, RSP-85, RSP-91 and RSP-92.



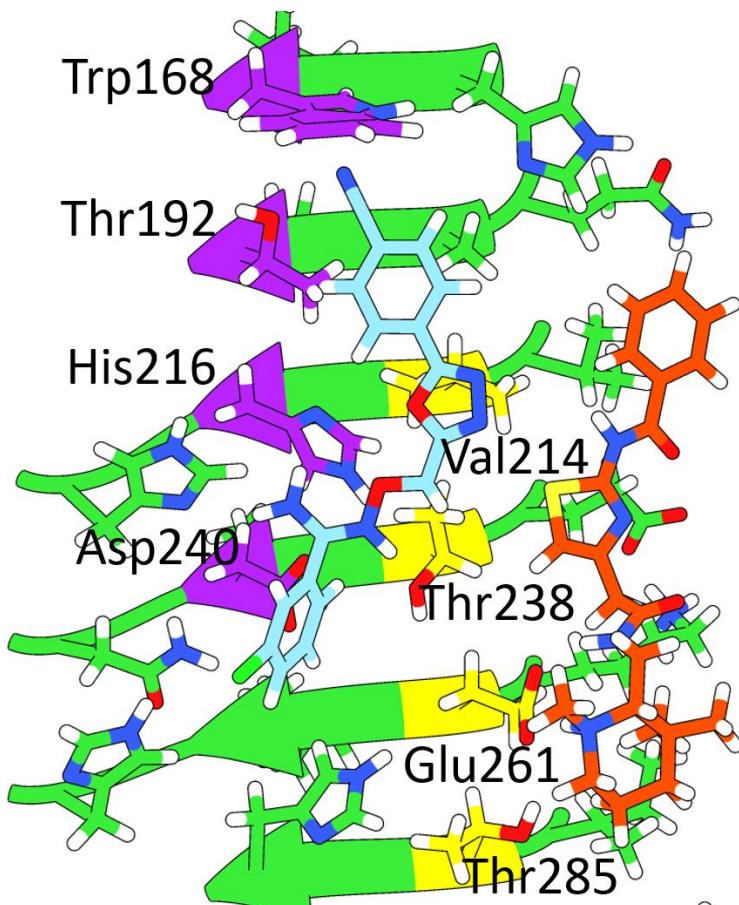
**Fig 5.2.17** Total binding free energy of RSP-83 binding to LGR5 protein as a function of time.

The energy decomposition of RSP-82, RSP-86, RSP-88 and RSP-89 share similar features (**Fig 5.2.18**), with four groups of residues that contribute most for protein-ligand binding stability, including residues near Val214, Thr238, Glu261 and Thr285. These residues located at the groove beside the original binding cavity (**Fig 5.2.19**).

This is in accordance with the observation that these compounds move out of the original binding site and bind to the groove nearby (**Fig 4.5.13**).



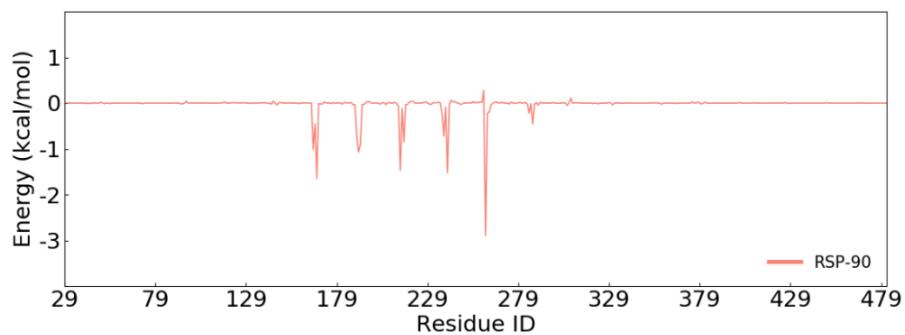
**Fig 5.2.18** Decomposition of protein-ligand binding free energies in the simulations of RSP-82, RSP-86, RSP-88 and RSP-89.



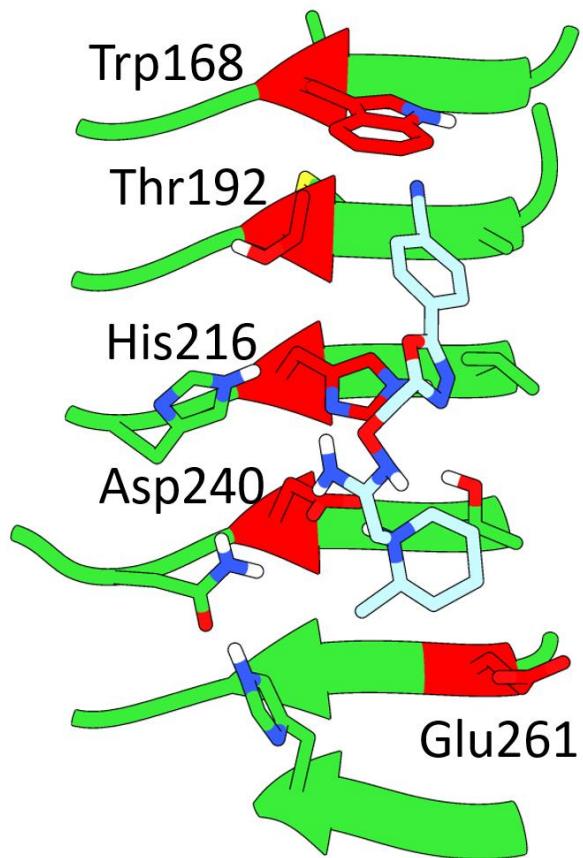
**Fig 5.2.19** LGR5 binding patterns of RSP-85 and RSP-89. Major binding free energy contributors near RSP-85 coloured by purple and contributors near RSP-90 colored yellow. LGR5 protein structure coloured green, RSP-85 coloured light blue and RSP-89 coloured orange.

RSP-90 was the compound that had the lowest binding free energy. Different from other compounds, it had five major binding residue clusters that contributed to its

protein-ligand binding affinity (**Fig 5.2.20**). These facts suggest that Glu261 is the main binding free energy contributor in the Glu261 cavity. The binding pattern of RSP-90 is shown in **Fig 5.2.21**.



**Fig 5.2.20** Energy decomposition of RSP-90/LGR5 interaction by residue.



**Fig 5.2.21** Binding pattern of RSP-90 binding to LGR5. Protein structure coloured green and RSP-90 were coloured light blue. Key residues were labeled.

## 6. Discussion

With interest in the functions of R-spondin protein in myogenic differentiation<sup>72</sup>, intestine repair<sup>73</sup>, placental development<sup>74</sup>, this work aims to identify compounds predicted to be capable of modulating R-spondin interactions through targeting R-spondin, LGR5 and RNF43 ternary complex. To achieve this, we have sought to provide insight into the relationship between ligand structure, protein structure and their interaction, using docking, MD, free energy calculations and energy decomposition.

RNF43 is of the greatest potential among all tested proteins: ligands that are able to bind with the upper and lower domain at the same are favoured (**Chapter 4.4.1**). The upper cavity is highly hydrophobic and as a consequence, the structure that binds with the upper cavity also has to be hydrophobic. In the meantime, two hydrogen bond donors near Asp97 were able to form two hydrogen bonds and help stabilize protein-ligand binding (**Fig 4.4.9**). RSP-18 is highly recommended to put into experimental testing because it has the lowest binding free energy, of -45.9 kcal/mol (**Table 5.7**). It has a highly hydrophobic moiety to bind with the RNF43 upper cavity; it also has an additional hydroxyl group to form a hydrogen bond with Asp97, occupied in MD for 97% of the time (**Fig 4.4.9** and **Table 4.4.3**). RSP-22 should also be considered for experimental testing because its binding free energy is -34.7 kcal/mol, the highest among all binding mode 2 compounds (**Fig 4.4.12** and **Table 5.7**). Its RMSF and standard deviation of ligand positional shift are 1.6 Å and 0.8 Å respectively, which is among the lowest level (**Table 4.4.2**); it also formed two

hydrogen bonds with Val176 for 53% and 47% of the time in MD respectively (**Table 4.4.3**).

**Open2** R-spondin is of greater potential for drug binding than **Closed** R-spondin. The hydrophobic residues include Leu63, Leu64, Ile62 and Leu54 and contributed the most to both **Closed** and **Open2** R-spondin protein-ligand binding. These residues provide a hydrophobic environment for binding hydrophobic moieties of compounds. **Closed** R-spondin has a relatively shallow cavity; hence, the hydrogen bond between ligand and Cys94 is important for protein-ligand binding. Compound RSP-05 is recommended for experimental assay; it possesses the most typical binding pattern of **Closed** R-spondin. It has two hydrophobic groups binding to Ile62, Leu63, Leu64; its binding free energy is -28.7 kcal/mol (**Table 5.2**) and it forms a hydrogen bond with Cys94 for 93% of time (**Table 4.2.3**). RSP-54 possesses a special binding pattern and it is also highly recommended for experimental assay: the RMSF of RSP-54 is 1.6 Å – one of the lowest values (**Table 4.2.5**) which means that it could bind to the cavity with high stability (**Fig 4.2.18**). **Open2** R-spondin possesses a deep cavity; Glu103 is located in the Furin-like 2 domain and plays a key role in the pocket formation. RSP-60 is recommended for experimental assay because it is one of the most representative compounds that does not alter the **Open2** cavity and demonstrates favourable binding free energy, of -34.3 kcal/mol (**Table 5.3**); it formed a hydrogen bond with Glu103 for 84% of the time (**Table 4.3.3**). From energy decomposition of RSP-60 (**Table 5.4**), Glu103 contributed the highest energy contribution -3.9 kcal/mol; other residues like Glu65, Arg66 and Ile95 contribute

equally to the binding free energy at 1.0 - 1.5 kcal/mol, suggesting that RSP-60 interacted with them equally well.

Two important binding patterns have been identified during the course of MD simulation of LGR5 compounds. Four residues, i.e. Trp168, Thr192, His216 and Asp240, are important for the first binding pattern (**Fig 5.2.21**). The cavity near Trp168 and Thr190 demonstrated significant affinity to a benzonitrile group and chlorobenzene group. RSP-90 belongs to this binding mode and has the best binding free energy, of -33.0 kcal/mol (**Table 5.8**). RSP-90 has a benzonitrile group at one end of the compound (**Fig 4.5.10** and **Fig 4.5.11**) which demonstrated significant binding stability with the Trp168/Thr192 cavity; it also formed a hydrogen bond with Glu261 for 73% and 67% of time in MD (**Table 4.5.5**). The second binding pattern was not predicted in the compound selection process. It is formed by Val214, Thr238, Glu261 and Thr285; this cluster is just beside the first binding pattern. RSP-82, RSP-86 and RSP-89 ligands form this binding pattern (**Fig 4.5.13**) and their binding free energies are -24.0 kcal/mol, -28.0 kcal/mol and -23.2 kcal/mol respectively (**Table 5.8**). RSP-86 is recommended for experimental testing, because it forms two hydrogen bonds with Glu261 for 100% and 86% of time respectively (**Table 4.5.5**). Two hydrogen bond interactions were formed between compounds and Glu261, making Glu261 the major energy contributor to LGR5 compounds (**Chapter 5.2.4**).

A number of limitations should be considered in this study. The first is the accuracy of FRED molecular docking, because FRED is a rigid body docking program and the

binding pockets are pre-defined. FRED could fail to predict good compounds for shallow pockets. In the FRED docking results for RNF43, the blue compound binding pocket (**Fig 4.4.1 (a)**) was open to solvent which makes the pocket unsuitable for ligand binding. However, a lot of high scoring compounds were predicted to bind with that position, like RSP-32 (**Table 4.4.1**). Also, we find there is no significant correlation between molecular docking scores and MD predicted binding stability. RSP-20 is ranked 943<sup>th</sup> place among RNF43's top 1000 compounds, and it demonstrated good binding affinity (**Table 4.4.1** and **Table 5.7**). The second limitation is the length of simulation time: for extremely stable binding compounds, 100 ns is enough to reach a stable status; however, for less highly stable binding compounds, the time should be extended to allow MD to adopt more stable binding poses of compounds. In the first 100 ns MD simulation of RSP-53 and **Closed** R-spondin, the binding pattern kept changing during the simulation. After extending the simulation time to 200 ns, the binding pose of RSP-53 finally reached a stable status (**Table 4.2.15**).

This research provides an insight into the structural basis and dynamics of protein-ligand binding between compounds and R-spondin signalling proteins. Since the safety to target Wnt pathway is unclear<sup>33</sup>, compounds selected in this research is helpful for the choice, design and structural optimization of chemical tool or drug that can modulate Wnt strength through changing the function of R-spondin and could provide a valuable reference for the study of Wnt drug safety assay. This research is a virtual screening and theoretical examination of potential compounds, further inter-molecular affinity assay and biological activity assay is needed to

evaluate the binding affinity and the toxicology of the compounds. Once acquired the biological activities, improvement in chemical synthesis and structural optimization is required to make the compounds accessible for all researchers who study the biology of the R-spondin-modulated Wnt regulation. Then validation of the R-spondin signal as a drug target is a key step for further application of the research in pharmaceutical research.

## References

- (1) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biology* **2011**, 9 (1). <https://doi.org/10.1186/1741-7007-9-71>.
- (2) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a Novel Binding Trench in HIV Integrase. *Journal of Medicinal Chemistry* **2004**, 47 (8), 1879–1881. <https://doi.org/10.1021/jm0341913>.
- (3) Amaro, R. E.; Baron, R.; McCammon, J. A. An Improved Relaxed Complex Scheme for Receptor Flexibility in Computer-Aided Drug Design. *Journal of Computer-Aided Molecular Design* **2008**, 22 (9), 693–705. <https://doi.org/10.1007/s10822-007-9159-2>.
- (4) Nusse, R.; Varmus, H. E. Many Tumors Induced by the Mouse Mammary Tumor Virus Contain a Proivirus Integrated in the Same Region of the Host Genome. *Cell* **1982**, 31 (1), 99–109. [https://doi.org/10.1016/0092-8674\(82\)90409-3](https://doi.org/10.1016/0092-8674(82)90409-3).
- (5) Clevers, H.; Nusse, R. Wnt/β-Catenin Signaling and Disease. *Cell* **2012**, 149 (6), 1192–1205. <https://doi.org/10.1016/j.cell.2012.05.012>.
- (6) Rijsewijk, F.; Wagenaar, E.; Parren, P.; Weigel, D. The Drosophila Homolog of the Mouse Mammary Oncogene Int-1 Is Identical to the Segment Polarity Gene Wingless. *Cell* **1987**, 50 (4), 649–657.
- (7) Logan, C. Y.; Nusse, R. THE WNT SIGNALING PATHWAY IN DEVELOPMENT AND DISEASE. *Annual Review of Cell and Developmental Biology* **2004**, 20 (1), 781–810. <https://doi.org/10.1146/annurev.cellbio.20.010403.113126>.
- (8) Björklund, P.; Svedlund, J.; Olsson, A.-K.; Åkerström, G.; Westin, G. The Internally Truncated LRP5 Receptor Presents a Therapeutic Target in Breast Cancer. *PLOS ONE* **2009**, 4 (1), e4243. <https://doi.org/10.1371/journal.pone.0004243>.
- (9) Polakis, P. The Many Ways of Wnt in Cancer. *Current Opinion in Genetics & Development* **2007**, 17 (1), 45–51. <https://doi.org/10.1016/j.gde.2006.12.007>.
- (10) The Role of Beta-Catenin in Cell Adhesion, Signal Transduction, and Cancer.
- (11) Nagaraj, A. B.; Joseph, P.; Kovalenko, O.; Singh, S.; Armstrong, A.; Redline, R.; Resnick, K.; Zanotti, K.; Waggoner, S.; DiFeo, A. Critical Role of Wnt/β-Catenin Signaling in Driving Epithelial Ovarian Cancer Platinum Resistance. *Oncotarget* **2015**, 6 (27). <https://doi.org/10.18632/oncotarget.4690>.
- (12) Takahashi-Yanaga, F.; Sasaguri, T. The Wnt/β-Catenin Signaling Pathway as a Target in Drug Discovery. *Journal of Pharmacological Sciences* **2007**, 104 (4), 293–302. <https://doi.org/10.1254/jphs.CR0070024>.
- (13) Kamata, T.; Katsume, K.; Michikawa, M.; Yamada, M.; Takada, S.; Mizusawa, H. R-Spondin, a Novel Gene with Thrombospondin Type 1 Domain, Was Expressed in the Dorsal Neural Tube and Affected in Wnts Mutants. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **2004**, 1676 (1), 51–62. <https://doi.org/10.1016/j.bbapara.2003.10.009>.
- (14) Zebisch, M.; Xu, Y.; Krastev, C.; MacDonald, B. T.; Chen, M.; Gilbert, R. J. C.; He, X.; Jones, E. Y. Structural and Molecular Basis of ZNRF3/RNF43 Transmembrane Ubiquitin

- Ligase Inhibition by the Wnt Agonist R-Spondin. *Nature Communications* **2013**, *4* (1). <https://doi.org/10.1038/ncomms3787>.
- (15) de Lau, W. B.; Snel, B.; Clevers, H. C. The R-Spondin Protein Family. *Genome Biology* **2012**, *13* (3), 242. <https://doi.org/10.1186/gb-2012-13-3-242>.
- (16) Kazanskaya, O.; Glinka, A. R-Spondin2 Is a Secreted Activator of Wnt/β-Catenin Signaling and Is Required for Xenopus Myogenesis. *Developmental Cell* **10**.
- (17) Lebensohn, A. M.; Rohatgi, R. R-Spondins Can Potentiate WNT Signaling without LGRs. *eLife* **2018**, *7*. <https://doi.org/10.7554/eLife.33126>.
- (18) Parma, P.; Radi, O.; Vidal, V.; Chaboissier, M. C.; Dellambra, E.; Valentini, S.; Guerra, L.; Schedl, A.; Camerino, G. R-Spondin1 Is Essential in Sex Determination, Skin Differentiation and Malignancy. *Nature Genetics* **2006**, *38* (11), 1304–1309. <https://doi.org/10.1038/ng1907>.
- (19) Krönke, G.; Uderhardt, S.; Kim, K.-A.; Stock, M.; Scholtysek, C.; Zaiss, M. M.; Surmann-Schmitt, C.; Luther, J.; Katzenbeisser, J.; David, J.-P.; et al. R-Spondin 1 Protects against Inflammatory Bone Damage during Murine Arthritis by Modulating the Wnt Pathway. *Arthritis & Rheumatism* **2010**, *62* (8), 2303–2312. <https://doi.org/10.1002/art.27496>.
- (20) Zhu, C.; Zheng, X.-F.; Yang, Y.-H.; Li, B.; Wang, Y.-R.; Jiang, S.-D.; Jiang, L.-S. LGR4 Acts as a Key Receptor for R-Spondin 2 to Promote Osteogenesis through Wnt Signaling Pathway. *Cellular Signalling* **2016**, *28* (8), 989–1000. <https://doi.org/10.1016/j.cellsig.2016.04.010>.
- (21) Bell, S. M.; Schreiner, C. M.; Wert, S. E.; Mucenski, M. L.; Scott, W. J.; Whitsett, J. A. R-Spondin 2 Is Required for Normal Laryngeal-Tracheal, Lung and Limb Morphogenesis. *Development* **2008**, *135* (6), 1049–1058. <https://doi.org/10.1242/dev.013359>.
- (22) Barker, N.; van Es, J. H.; Kuipers, J.; Kujala, P.; van den Born, M.; Cozijnsen, M.; Haegembirth, A.; Korving, J.; Begthel, H.; Peters, P. J.; et al. Identification of Stem Cells in Small Intestine and Colon by Marker Gene Lgr5. *Nature* **2007**, *449* (7165), 1003–1007. <https://doi.org/10.1038/nature06196>.
- (23) Kikuchi, A.; Yamamoto, H. Regulation of Wnt Signalling by Receptor-Mediated Endocytosis. *Journal of Biochemistry* **2007**, *141* (4), 443–451. <https://doi.org/10.1093/jb/mvm061>.
- (24) van den Heuvel, M.; Nusse, R.; Johnston, P.; Lawrence, P. A. Distribution of the Wingless Gene Product in Drosophila Embryos: A Protein Involved in Cell-Cell Communication. *Cell* **1989**, *59* (4), 739–749. [https://doi.org/10.1016/0092-8674\(89\)90020-2](https://doi.org/10.1016/0092-8674(89)90020-2).
- (25) Carmon, K. S.; Gong, X.; Lin, Q.; Thomas, A.; Liu, Q. R-Spondins Function as Ligands of the Orphan Receptors LGR4 and LGR5 to Regulate Wnt/β-Catenin Signaling. *CELL BIOLOGY* **2011**, *108* (28), 11452–11457.
- (26) Morita, H.; Mazerbourg, S.; Bouley, D. M.; Luo, C.-W.; Kawamura, K.; Kuwabara, Y.; Baribault, H.; Tian, H.; Hsueh, A. J. W. Neonatal Lethality of LGR5 Null Mice Is Associated

- with Ankyloglossia and Gastrointestinal Distension. *Molecular and Cellular Biology* **2004**, *24* (22), 9736–9743. <https://doi.org/10.1128/MCB.24.22.9736-9743.2004>.
- (27) Glinka, A.; Dolde, C.; Kirsch, N.; Huang, Y.-L.; Kazanskaya, O.; Ingelfinger, D.; Boutros, M.; Cruciat, C.-M.; Niehrs, C. LGR4 and LGR5 Are R-Spondin Receptors Mediating Wnt/β-Catenin and Wnt/PCP Signalling. *EMBO reports* **2011**, *12* (10), 1055–1061. <https://doi.org/10.1038/embor.2011.175>.
- (28) Hao, H.-X.; Xie, Y.; Zhang, Y.; Charlat, O.; Oster, E.; Avello, M.; Lei, H.; Mickanin, C.; Liu, D.; Ruffner, H.; et al. ZNRF3 Promotes Wnt Receptor Turnover in an R-Spondin-Sensitive Manner. *Nature* **2012**, *485* (7397), 195–200. <https://doi.org/10.1038/nature11019>.
- (29) Jiang, X.; Charlat, O.; Zamponi, R.; Yang, Y.; Cong, F. Dishevelled Promotes Wnt Receptor Degradation through Recruitment of ZNRF3/RNF43 E3 Ubiquitin Ligases. *Molecular Cell* **2015**, *58* (3), 522–533. <https://doi.org/10.1016/j.molcel.2015.03.015>.
- (30) Chen, P.-H.; Chen, X.; Lin, Z.; Fang, D.; He, X. The Structural Basis of R-Spondin Recognition by LGR5 and RNF43. *Genes & Development* **2013**, *27* (12), 1345–1350. <https://doi.org/10.1101/gad.219915.113>.
- (31) Szenker-Ravi, E.; Altunoglu, U.; Leushacke, M.; Bosso-Lefèvre, C.; Khatoo, M.; Thi Tran, H.; Naert, T.; Noelanders, R.; Hajamohideen, A.; Beneteau, C.; et al. RSPO2 Inhibition of RNF43 and ZNRF3 Governs Limb Development Independently of LGR4/5/6. *Nature* **2018**, *557* (7706), 564–569. <https://doi.org/10.1038/s41586-018-0118-y>.
- (32) Boone, J. D.; Arend, R. C.; Johnston, B. E.; Cooper, S. J.; Gilchrist, S. A.; Oelschlager, D. K.; Grizzle, W. E.; McGwin, G.; Gangrade, A.; Straughn, J. M.; et al. Targeting the Wnt/β-Catenin Pathway in Primary Ovarian Cancer with the Porcupine Inhibitor WNT974. *Laboratory Investigation* **2016**, *96* (2), 249–259. <https://doi.org/10.1038/labinvest.2015.150>.
- (33) Kahn, M. Can We Safely Target the WNT Pathway? *Nature Reviews Drug Discovery* **2014**, *13* (7), 513–532. <https://doi.org/10.1038/nrd4233>.
- (34) Niehrs, C. The Complex World of WNT Receptor Signalling. *Nature Reviews Molecular Cell Biology* **2012**, *13* (12), 767–779. <https://doi.org/10.1038/nrm3470>.
- (35) de Lau, W.; Peng, W. C.; Gros, P.; Clevers, H. The R-Spondin/Lgr5/Rnf43 Module: Regulator of Wnt Signal Strength. *Genes & Development* **2014**, *28* (4), 305–316. <https://doi.org/10.1101/gad.235473.113>.
- (36) Volkamer, A.; Griewel, A.; Grombacher, T.; Rarey, M. Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. *Journal of Chemical Information and Modeling* **2010**, *50* (11), 2041–2052. <https://doi.org/10.1021/ci100241y>.
- (37) Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: A Web Server for Automatic Binding Site Prediction, Analysis and Druggability Assessment. *Bioinformatics* **2012**, *28* (15), 2074–2075. <https://doi.org/10.1093/bioinformatics/bts310>.
- (38) Volkamer, A.; Kuhn, D.; Grombacher, T.; Rippmann, F.; Rarey, M. Combining Global and Local Measures for Structure-Based Druggability Predictions. *Journal of Chemical Information and Modeling* **2012**, *52* (2), 360–372. <https://doi.org/10.1021/ci200454v>.

- (39) Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-Based Identification of Druggable ‘Hot Spots’ of Proteins Using Fourier Domain Correlation Techniques. *Bioinformatics* **2009**, 25 (5), 621–627. <https://doi.org/10.1093/bioinformatics/btp036>.
- (40) Chen, Y.-C. Beware of Docking! *Trends in Pharmacological Sciences* **2015**, 36 (2), 78–95. <https://doi.org/10.1016/j.tips.2014.12.001>.
- (41) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *Journal of Molecular Biology* **1995**, 245 (1), 43–53. [https://doi.org/10.1016/S0022-2836\(95\)80037-9](https://doi.org/10.1016/S0022-2836(95)80037-9).
- (42) McGann, M. FRED and HYBRID Docking Performance on Standardized Datasets. *J Comput Aided Mol Des* **2012**, 26 (8), 897–906. <https://doi.org/10.1007/s10822-012-9584-8>.
- (43) McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, 51 (3), 578–596. <https://doi.org/10.1021/ci100436p>.
- (44) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. 21.
- (45) Murray, C. W.; Auton, T. R.; Eldridge, M. D. Empirical Scoring Functions. II. The Testing of an Empirical Scoring Function for the Prediction of Ligand-Receptor Binding Affinities and the Use of Bayesian Regression to Improve the Quality of the Model. 18.
- (46) Korb, O.; Stützle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling* **2009**, 49 (1), 84–96. <https://doi.org/10.1021/ci800298z>.
- (47) *FRED (Fast Rigid Exhaustive Docking)*; Openeye Scientific Software: Santa Fe, NM, 2006.
- (48) Born, M.; Oppenheimer, R. Zur Quantentheorie der Moleküle. *Ann. Phys.* **1927**, 389 (20), 457–484. <https://doi.org/10.1002/andp.19273892002>.
- (49) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. Solvent Models for Protein-Ligand Binding: Comparison of Implicit Solvent Poisson and Surface Generalized Born Models with Explicit Solvent Simulations. *Journal of Computational Chemistry* **2001**, 22 (6), 591–607. <https://doi.org/10.1002/jcc.1031>.
- (50) Skyner, R. E.; McDonagh, J. L.; Groom, C. R.; van Mourik, T.; Mitchell, J. B. O. A Review of Methods for the Calculation of Solution Free Energies and the Modelling of Systems in Solution. *Phys. Chem. Chem. Phys.* **2015**, 17 (9), 6174–6191. <https://doi.org/10.1039/C5CP00288E>.
- (51) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, 79 (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (52) Lemons, D. S.; Gythiel, A. Paul Langevin’s 1908 Paper “On the Theory of Brownian Motion” [“Sur La Théorie Du Mouvement Brownien,” C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]. *American Journal of Physics* **1997**, 65 (11), 1079–1081. <https://doi.org/10.1119/1.18725>.

- (53) Rastelli, G.; Rio, A. D.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry* **2009**, NA-NA. <https://doi.org/10.1002/jcc.21372>.
- (54) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–6129. <https://doi.org/10.1021/ja00172a038>.
- (55) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a Pairwise Generalized Born Model to Proteins and Nucleic Acids: Inclusion of Salt Effects. *9*.
- (56) Constanciel, R.; Contreras, R. Self Consistent Field Theory of Solvent Effects Representation by Continuum Models: Introduction of Desolvation Contribution. *Theoret. Chim. Acta* **1984**, *65* (1), 1–11. <https://doi.org/10.1007/BF02427575>.
- (57) *The PyMOL Molecular Graphics System*; Schrödinger, LLC.
- (58) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- (59) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling* **2015**, *55* (11), 2324–2337.  
<https://doi.org/10.1021/acs.jcim.5b00559>.
- (60) Ngan, C.-H.; Hall, D. R.; Zerbe, B.; Grove, L. E.; Kozakov, D.; Vajda, S. FTSite: High Accuracy Detection of Ligand Binding Sites on Unbound Protein Structures. *Bioinformatics* **2012**, *28* (2), 286–287. <https://doi.org/10.1093/bioinformatics/btr651>.
- (61) D.A. Case; I.Y. Ben-Shalom; S.R. Brozell; D.S. Cerutti; T.E. Cheatham; III; V.W.D. Cruzeiro; T.A. Darden; R.E. Duke; D. Ghoreishi; et al. AMBER 2018. University of California, San Francisco 2018.
- (62) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696–3713.  
<https://doi.org/10.1021/acs.jctc.5b00255>.
- (63) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25* (9), 1157–1174. <https://doi.org/10.1002/jcc.20035>.
- (64) *Molecular Operating Environment (MOE)*; Chemical Computing Group ULC: 1010 Sherbrooke St. West, Suit #910, Montreal, QC, Canada, H3A 2R7, 2018.
- (65) Rupesh, C.; Kepa, B.-G.; Alessio, A.; Richard, B. Molecular Dynamics of R-Spondin 1: Conformation and Druggability Analysis; AMER CHEMICAL SOC, 2018; Vol. 256.
- (66) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *Journal of Computational Chemistry* **2002**, *23* (16), 1623–1641. <https://doi.org/10.1002/jcc.10128>.
- (67) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. *10*. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers.

Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103* (2), 335–340.  
<https://doi.org/10.1021/ja00392a016>.

- (68) Roe, D. R.; Cheatham, T. E. PTraj and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. <https://doi.org/10.1021/ct400341p>.
- (69) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (70) QUACPAC 2.0.0.3; OpenEye Scientific Software: Santa Fe, NM.
- (71) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104* (15), 3712–3720.  
<https://doi.org/10.1021/jp994072s>.
- (72) Han, X. H.; Jin, Y.-R.; Seto, M.; Yoon, J. K. A WNT/β-Catenin Signaling Activator, R-Spondin, Plays Positive Regulatory Roles during Skeletal Myogenesis. *Journal of Biological Chemistry* **2011**, *286* (12), 10649–10659. <https://doi.org/10.1074/jbc.M110.169391>.
- (73) Storm, E. E.; Durinck, S.; de Sousa e Melo, F.; Tremayne, J.; Kljavin, N.; Tan, C.; Ye, X.; Chiu, C.; Pham, T.; Hongo, J.-A.; et al. Targeting PTPRK-RSPO3 Colon Tumours Promotes Differentiation and Loss of Stem-Cell Function. *Nature* **2016**, *529* (7584), 97–100.  
<https://doi.org/10.1038/nature16466>.
- (74) Aoki, M.; Mieda, M.; Ikeda, T.; Hamada, Y.; Nakamura, H.; Okamoto, H. R-Spondin3 Is Required for Mouse Placental Development. *Developmental Biology* **2007**, *301* (1), 218–226. <https://doi.org/10.1016/j.ydbio.2006.08.018>.