



João Carlos Moreno Ramos

Licenciado em Bioquímica

**Structure based drug design for the
discovery of promising inhibitors of
human Bcl-2 and *Streptococcus
dysgalactiae* LytR proteins**

Dissertação para obtenção do Grau de Mestre em
Bioquímica

Orientador: Doutora Teresa Sacadura Santos Silva, Investigadora
Auxiliar, FCT-NOVA
Co-orientador: Doutor Filipe Miguel dos Santos Freire, Investigador
Associado Sénior na Unidade de Biologia para a Descoberta de
Fármacos no IBET/Merck Healthcare Laboratory

Presidente: Prof. Doutor José Ricardo Ramos Franco Tavares
Arguente: Doutor Pedro Manuel Henriques Marques Matias
Vogal: Doutora Teresa Sacadura Santos-Silva



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2017



João Carlos Moreno Ramos

Licenciado em Bioquímica

**Structure based drug design for the
discovery of promising inhibitors of
human Bcl-2 and *Streptococcus
dysgalactiae* LytR proteins**

Dissertação para obtenção do Grau de Mestre em
Bioquímica

Orientador: Doutora Teresa Sacadura Santos Silva, Investigadora
Auxiliar, FCT-NOVA
Co-orientador: Doutor Filipe Miguel dos Santos Freire, Investigador
Associado Sénior na Unidade de Biologia para a Descoberta de
Fármacos no IBET/Merck Healthcare Laboratory

Presidente: Prof. Doutor José Ricardo Ramos Franco Tavares
Arguente: Doutor Pedro Manuel Henriques Marques Matias
Vogal: Doutora Teresa Sacadura Santos-Silva



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2017

“Structure based drug design for the discovery of promising inhibitors of human Bcl-2 and *Streptococcus dysgalactiae* LytR proteins”

“Copyright” em nome de João Carlos Moreno Ramos, da FCT/UNL e da UNL

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

Dedico esta dissertação aos meus pais, Carlos e Helena, aos meus irmãos, Pedro e Sara, e aos meus avós, Zé, Rosa e Neno. Obrigado por tudo o que fizeram e continuam a fazer por mim. Devo-vos muito do que tenho e principalmente do que sou.

Agradeço à Professora Maria João Romão por me ter dado a oportunidade de trabalhar, tanto durante a Licenciatura como no Mestrado, no grupo de Cristalografia de Macromoléculas (Xtal).

À Dr. Ana Luísa e ao Professor Ricardo Franco por me terem introduzido ao trabalho de investigação e ao método de Cristalografia de raios-X, o qual tenciono manter como um dos meus principais interesses científicos no futuro.

A todo o grupo Xtal: Teresa, Ana Luísa, Angelina, Catarina, Benedita, Marino, Márcia, Muthukumaran, Cristiano, Rita, Francisco, Viviana, Diana, Raquel, Filipa, Sónia e aos meus colegas de mestrado Frederico e Diogo por todo o apoio e pelo excelente ambiente que, certamente, ajudou-me a ultrapassar as dificuldades que surgiram ao longo deste ano.

À Dr. Teresa Santos Silva pela sua orientação durante este ano, que implicou muito trabalho, exigência e imensas oportunidades de enriquecimento académico e pessoal.

Ao Dr. Filipe Freire pela incessante paciência e disponibilidade durante a minha aprendizagem da parte experimental desta dissertação. O principal motivo de interesse que me direcionou para este trabalho foi a possibilidade de aprender sobre várias técnicas relacionadas com a cristalografia e a concepção de fármacos e, sem dúvida, não podia ter melhor professor. Continuarás a ser um exemplo de profissionalismo e organização que espero seguir no futuro.

To Dr. Jayaraman Muthukumaran for is remarkable patience and kindness throughout my learning process regarding the bioinformatic and computational approach in this work. You taught so much and I hope to keep learning more about all the techniques that we used and are common in drug design.

Ao Dr. Marino Santos pelo incansável apoio, pelas intermináveis discussões e pelo espírito aberto e humorístico, quase ao nível do Presidente, presente a cada dia com o qual recebeu até as mais desconcertantes dúvidas. Lamento as inúmeras falhas na entrega de protocolos e no cumprimento de preciosos conselhos, um dos quais, a escrita antecipada destes agradecimentos. És, paradoxalmente, um exemplo de esforço, dedicação, devoção e glória, ao ponto de te poder considerar um Adrien deste grupo. Devo-te muito do que será o meu pós-mestrado, mas talvez tenhamos de reflectir sobre a tua proposta de compensação.

Ao Francisco Leisico por todo o apoio durante o trabalho experimental e pelas bem-humoradas convivências perante a bancada do laboratório. Fico-te a dever incontáveis futeboladas e tu a dever-me umas quantas aulas de surf.

À Dr. Márcia Correia pelo extraordinário trabalho na gestão do laboratório e pela paciência com a qual enfrentou todas as minhas dúvidas sobre reagentes, equipamentos e *software*.

Aos alunos com os quais trabalhei, João Ferreira, Kivénia Ferreira e Sara Carvalho, e aos quais tentei ensinar parte do meu trabalho e das técnicas que utilizei. Espero que tenha conseguido passar algum conhecimento que vos ajude no futuro, independentemente de todas as minhas falhas.

Aos meus amigos, Rafael Rippel, Luis Silva, Diogo Gonçalves, Alexandra Francisco, António Lopéz, Mauro Duarte, Miguel Correia, Dário Valezim, João Veléz, Bruno Guerreiro, Miguel Palhas e Emanuel Moreira, pelos fantásticos tempos que passámos juntos e todas as recordações que levarei comigo. Sem vocês, a faculdade ter-me-ia dado muito pouco.

A todos os professores que tive durante a minha Licenciatura e Mestrado que, certamente, me influenciaram e inspiraram ao longo destes anos.

Resumo

A descoberta de fármacos tem evoluído nas últimas décadas no sentido da concepção racional de moléculas activas. A possibilidade do estudo de interacções moleculares ao nível atómico e a compreensão de que este conhecimento pode ser aplicado ao desenvolvimento de fármacos são as premissas da concepção estrutural de fármacos. Associando esta abordagem aos métodos computacionais disponíveis actualmente, é concretizada a oportunidade de acelerar o complexo processo de concepção de fármacos. Na presente tese, esta metodologia é empregue no estudo de promissores candidatos a inibidores das proteínas Bcl-2 humana e LytR de *Streptococcus dysgalactiae*.

Estima-se que metade dos cancros em humanos estão relacionados com a sobre-expressão de Bcl-2. Esta proteína é responsável pela inibição do mecanismo apoptótico, o qual é essencial para a irradicação de células disfuncionais. Quando ocorre sobre-expressão de Bcl-2, estas células não respondem a estímulos apoptóticos, quer endógenos ou exógenos, como agentes quimioterapêuticos. Derivados das famílias 4H-cromeno e indol foram estudados em relação à sua capacidade promissora de inibir a Bcl-2. Estudos de *docking* molecular revelaram afinidades sub-micromolares dos compostos 4H-cromeno *activemethine* e indol para o local de ligação fisiológico da Bcl-2. A caracterização biofísica das interacções não resultou em provas evidentes de ligação, provavelmente devido à escassa rede de interacções com os resíduos do local de ligação. Cristalografia de raios-X foi direccionada para a determinação da estrutura destes possíveis complexos proteína-ligando, sendo que condições preliminares de co-cristalização foram alcançadas.

Diversas patologias infecciosas estão relacionadas com o fenótipo de biofilme, o qual consiste na aglomeração de bactérias envoltas numa matriz. O biofilme confere às bactérias um aumento em relação à resistência ao sistema imunitário inato do hospedeiro e ao tratamento com antibióticos genéricos. A LytR pertence à família de proteínas LCP, as quais se pensa estarem envolvidas na adição de polímeros aniónicos ao peptidoglicano, protegendo as bactérias Gram-positivas de fagocitose e da lise celular. Estudos anteriores de procura bioinformática reportaram o ácido elágico e a fisetina como inibidores promissores da LytR, demonstrando actividade anti-biofilme. O *docking* molecular revelou a ligação destes compostos com afinidade micromolar ao hipotético sítio activo da LytR, interagindo com resíduos catalíticos fundamentais. As técnicas biofísicas empregues não comprovaram ligação destas moléculas à proteína, o que pode ser explicado pela co-purificação de um substrato lipídico, fenómeno que já foi reportado anteriormente. A aplicação de espectrometria de massa ou a determinação estrutural, através de cristalografia de raios-X ou RMN, pode ser conclusiva relativamente à ocupação do sítio activo da LytR por parte desta molécula.

Palavras-chave: concepção estrutural de fármacos, procura computacional, actividade anticancerígena, linfoma 2 de células-B, resistência a antibióticos, LytR, caracterização biofísica, cristalografia de raios-X.

Abstract

Drug research has evolved significantly in the last decades toward the concept of the rational design of drugs. The capability to study molecular interactions at the atomic level and to rationalize this knowledge to construct and improve drug candidates provided the premises of structure-based drug design (SBDD). This approach allied to the computational methods available nowadays yields the opportunity to expedite the intricate process of drug discovery. In the present thesis, the SBDD approach was implemented to study promising candidate inhibitors of the human Bcl-2 and the *Streptococcus dysgalactiae* LytR proteins.

Half of the cancers in humans are estimated to be related with overexpression of Bcl-2 protein. This macromolecule is responsible for the inhibition of the apoptotic process, which is pivotal for the elimination of abnormal cells. When Bcl-2 is overexpressed, these abnormal cells don't respond to death stimuli, either endogenous or exogenous, such as chemotherapeutic, and become immortal. Promising 4H-chromene and indole derivatives were studied regarding their potential to inhibit Bcl-2. Molecular docking studies revealed sub-micromolar binding of the 4H-chromene activemethine and the indole derivatives in the binding groove essential for Bcl-2 biological function. Biophysical characterization did not demonstrate significant evidence of binding between Bcl-2 and the compounds under study, probably due to their small network of interactions with the binding pocket residues. The structure determination process of the protein-ligand complexes achieved preliminary co-crystallization conditions that require further optimization.

Numerous infectious diseases are associated to the bacterial biofilm phenotype, which consists of agglomerates of cells enclosed in a self-produced matrix. Biofilms confer bacteria improved resistance to the host's innate immune system and to conventional antibiotics. LytR belongs to the LCP family of proteins, which are thought to be responsible for the attachment of anionic polymers to the peptidoglycan, protecting the Gram-positive bacteria from phagocytosis and lysis. Previous virtual screening studies yielded ellagic acid and fisetin as promising inhibitors of LytR, displaying anti-biofilm activity. Molecular docking revealed binding of these compounds in the hypothetical active site of LytR, with micromolar affinities, and specific interactions with crucial protein residues for catalysis. Biophysical techniques failed to provide evidence of protein-ligand interactions, although this may be related to the possible co-purification with a lipidic substrate, which has been reported before. Mass spectrometry or structural determination, through X-ray crystallography or NMR, should be pivotal to establish evidence of this molecule's accommodation in the binding pocket.

Keywords: structure-based drug design, virtual screening, anti-cancerous activity, B-cell lymphoma 2, antibiotic resistance, LytR, biophysical characterization, X-ray crystallography.

Table of contents

Agradecimientos.....	I
Resumo.....	III
Abstract.....	V
Table of contents	VII
List of figures	XI
1. Introduction.....	1
1.1. Drug Discovery	3
1.1.1. Historical Perspective	3
1.1.2. Overview.....	5
1.1.2.1. Target Identification	5
1.1.2.2. Target Validation	6
1.1.2.3. Hit Discovery Process	6
1.1.2.4. Hit series selection.....	8
1.1.2.5. Hit-to-lead phase	9
1.1.2.6. Lead optimization phase	10
1.2. Structure-based Virtual Screening (SBVS).....	11
1.2.1. Target structural determination	11
1.2.1.1. X-ray crystallography	11
1.2.1.1.1. Protein crystallization	13
1.2.1.1.2. X-ray diffraction.....	15
1.2.1.1.3. Structure determination.....	17
1.2.1.2. Homology Modeling	19
1.2.2. Virtual Screening	21
1.2.2.1. Ligand-based Virtual Screening	21
1.2.2.2. Structure-based Virtual Screening.....	23
1.2.3. Target-ligand complex characterization	24
1.2.3.1. Thermal Shift Assay (TSA)	24
1.2.3.2. Urea-Polyacrylamide Gel Electrophoresis	25
1.2.3.3. Isothermal Titration Calorimetry (ITC).....	26
1.2.3.4. Circular Dichroism (CD).....	27
1.2.4. Target-ligand structural determination	28
1.2.4.1. X-ray crystallography of protein-ligand complexes	28
1.2.4.2. Saturation Transfer Difference – NMR (STD-NMR)	29
1.2.4.3. Molecular Dynamics (MD)	31
1.3. State-of-the-Art	33

1.3.1.	Human Bcl-2 protein.....	33
1.3.2.	<i>Streptococcus dysgalactiae</i> LytR.....	39
2.	Materials and Methods.....	43
2.1.	Target structural determination.....	45
2.1.1.	Recombinant protein expression and purification	45
2.1.1.1.	Human Bcl-2 protein.....	45
2.1.1.2.	<i>Streptococcus dysgalactiae</i> LytR.....	46
2.1.2.	X-ray crystallography	47
2.1.2.1.	Protein crystallization	47
2.1.2.1.1.	Human Bcl-2 protein.....	47
2.1.2.1.2.	<i>Streptococcus dysgalactiae</i> LytR.....	48
2.1.2.2.	X-ray diffraction.....	49
2.1.3.	Homology Modeling	49
2.1.3.1.	Human Bcl-2 protein.....	49
2.1.3.2.	<i>Streptococcus dysgalactiae</i> LytR.....	49
2.2.	Virtual Screening and Molecular Docking.....	50
2.2.1.	Ligand-based.....	50
2.2.1.1.	Human Bcl-2 protein.....	50
2.2.2.	Structure-based	50
2.2.2.1.	Human Bcl-2 protein.....	50
2.2.3.	Molecular Docking.....	50
2.2.3.1.	Human Bcl-2 protein.....	50
2.2.3.2.	<i>Streptococcus dysgalactiae</i> LytR.....	51
2.3.	Target-ligand complex characterization	51
2.3.1.	Thermal Shift Assay (TSA)	51
2.3.2.	Urea-Polyacrylamide Gel Electrophoresis	52
2.3.2.1.	Human Bcl-2 protein.....	52
2.3.3.	Isothermal Titration Calorimetry (ITC).....	52
2.3.3.1.	Human Bcl-2 protein.....	52
2.3.3.2.	<i>Streptococcus dysgalactiae</i> LytR.....	53
2.3.4.	Circular Dichroism (CD).....	53
2.3.4.1.	<i>Streptococcus dysgalactiae</i> LytR.....	53
2.4.	Target-ligand structural determination	53
2.4.1.	X-ray crystallography	53
2.4.1.1.	Human Bcl-2 protein.....	53
2.4.2.	Saturation Transfer Difference NMR (STD-NMR)	54

2.4.3.	Molecular Dynamics (MD)	55
2.4.3.1.	Bcl-2 protein-ligand complexes	55
2.4.3.2.	Structural impact of non-synonymous single nucleotide polymorphisms (nsSNP) in Human Bcl-2 protein	56
3.	Results and Discussion	59
3.1.	Target structural determination	61
3.1.1.	Recombinant protein purification	61
3.1.1.1.	Human Bcl-2 protein	61
3.1.1.2.	<i>Streptococcus dysgalactiae</i> LytR	62
3.1.2.	X-ray crystallography.....	64
3.1.2.1.	Human Bcl-2 protein	64
3.1.2.2.	<i>Streptococcus dysgalactiae</i> LytR	64
3.1.3.	Homology Modeling.....	67
3.1.3.1.	Human Bcl-2 protein	67
3.1.3.2.	<i>Streptococcus dysgalactiae</i> LytR	68
3.2.	Virtual Screening and Molecular Docking.....	69
3.2.1.	Human Bcl-2 protein	69
3.2.3.	<i>Streptococcus dysgalactiae</i> LytR	80
3.3.	Target-ligand complex characterization	83
3.3.1.	Thermal Shift Assay (TSA)	83
3.3.1.1.	Human Bcl-2 protein	83
3.3.1.2.	<i>Streptococcus dysgalactiae</i> LytR	86
3.3.2.	Urea-Polyacrylamide Gel Electrophoresis.....	87
3.3.2.1.	Human Bcl-2 protein	87
3.3.3.	Isothermal Titration Calorimetry (ITC)	87
3.3.3.1.	Human Bcl-2 protein	88
3.3.3.2.	<i>Streptococcus dysgalactiae</i> LytR	88
3.3.4.	Circular Dichroism (CD)	89
3.3.4.1.	<i>Streptococcus dysgalactiae</i> LytR	89
3.4.	Target-ligand structural determination.....	90
3.4.1.	X-ray crystallography.....	90
3.4.1.1.	Human Bcl-2 protein	90
3.4.1.2.	<i>Streptococcus dysgalactiae</i> LytR	91
3.4.2.	Saturation Transfer Difference (STD-NMR).....	92
3.4.2.1.	Human Bcl-2 protein	92
3.4.2.2.	<i>Streptococcus dysgalactiae</i> LytR	93

3.4.3. Molecular Dynamics (MD).....	94
3.4.3.1. Protein-ligand complexes of Human Bcl-2	94
3.4.3.2. Structural and functional impact of non-synonymous single nucleotide polymorphisms (nsSNP) on Human Bcl-2 protein	96
4. Conclusions and future perspectives.....	99
5. Bibliography	105
6. Appendix	117
6.1. TSA buffer screening	119
6.2. TSA additive screening	120
6.3. TSA results of buffer and additive screenings with Bcl-2	121
6.4. ¹ H NMR spectra of the ligands under study	124

List of figures

Figure 1.1 - Chemical structures of the first two active compounds isolated from medicinal plants, morphine and papaverine.	3
Figure 1.2 - Schematic representation of a drug discovery process.	5
Figure 1.3 - Schematic representation of the properties assessed during a hit series selection.	9
Figure 1.4 - Schematic representation of the X-ray crystallography workflow.	12
Figure 1.5 - Representation of a (triclinic) unit cell with its defining parameters.	13
Figure 1.6 - Schematic representation of the vapor diffusion method.	14
Figure 1.7 - Representation of the Phase Diagram used to guide a protein crystallization endeavour.	14
Figure 1.8 - Schematic representation of the X-ray diffraction experiment.	15
Figure 1.9 – Schematic representation of a synchrotron.	16
Figure 1.10 - Schematic representation of the Bragg's law.	16
Figure 1.11 - Schematic representation of the normal mode protocol followed by Phyre2.	19
Figure 1.12 - Schematic representation of the intensive mode protocol used by Phyre2.	21
Figure 1.13 - Example of bit strings generated for two molecules using a small set of substructures.	22
Figure 1.14 - Representation of a protein unfolding curve from a TSA.	24
Figure 1.15 - Schematic representation of an ITC experiment.	27
Figure 1.16 - Far U.V. CD spectra characteristic for various secondary structures.	27
Figure 1.17 - Schematic representation of the STD-NMR technique.	30
Figure 1.18 - Energy equation applied to calculate the atom's motion in a Molecular Dynamics simulation.	31
Figure 1.19 - Schematic diagram of the six hallmarks of cancer.	33
Figure 1.20 - Schematic representation of the sequence homology present in Bcl-2 family members.	34
Figure 1.21 - Schematic representation of the embedded together model.	35
Figure 1.22 - Crystal structures of human Bcl-2 protein and Bcl-2 with Bax's BH3 domain.	36
Figure 1.23 - Chemical structures of the three generations of arylsulfonamides' most representative examples.	37
Figure 1.24 - Chemical structure of the reference 4H-chromene molecule, HA14-1.	38
Figure 1.25 - Chemical structures of the 4H-chromene derivatives under study.	38
Figure 1.26 - Chemical structure of the indole derivative under study.	38
Figure 1.27 - Model of the development of mature biofilm from planktonic cells.	39
Figure 1.28 - Schematic representation of the attachment of wall teichoic acids to the peptidoglycan cell wall by LCP proteins.	40
Figure 1.29 - Crystal structure of the extracellular domains of Cps2A from serotype 2 <i>S. pneumoniae</i> D39.	41
Figure 1.30 - Crystal structure of YwtF from <i>Bacillus subtilis</i>	41

Figure 1.31 - Schematic representation of the sequence homology between the proteins of the LCP family.	42
Figure 1.32 - Chemical structures of ellagic acid and fisetin.	42
Figure 3.1 - Bcl-2 IMAC chromatogram with corresponding fractions noted on the SDS-PAGE gel.	61
Figure 3.2 - Bcl-2 SEC chromatogram with corresponding peaks noted on SDS-PAGE gel. ...	62
Figure 3.3 - LytR IMAC chromatogram with corresponding peaks noted on SDS-PAGE gel. ...	62
Figure 3.4 - LytR SEC chromatogram for the buffer exchange.	63
Figure 3.5 - LytR SEC chromatogram with corresponding peaks noted on SDS-PAGE gel. ...	63
Figure 3.6 - Crystals of multiple nature obtained in a derived condition with ammonium sulphate and glycerol as precipitant agents.	65
Figure 3.7 – LytR crystals obtained from a condition with 50% MPD as precipitant agent.	65
Figure 3.8 – LytR crystals obtained from two different conditions.	66
Figure 3.9 - Bcl-2 full-length model predicted by the Phyre2 server.	67
Figure 3.10 - Three-dimensional structure of LytR LCP domain, predicted by homology modeling.	68
Figure 3.11 - Representation of each ligand best pose from the molecular docking results of Bcl-2 with the candidate inhibitors.	70
Figure 3.12 - Protein-ligand interaction mapping of the docking results for venetoclax, ethoxy and activemethine derivatives.	71
Figure 3.13 - Best pose docked for the indole 15 molecule to Bcl-2.	75
Figure 3.14 - Representation of the best pose from each ligand on the LytR modeled structure.	81
Figure 3.15 - Mapping of the LytR residue network interacting with the substrate, decaprenylphosphate.	82
Figure 3.16 - Mapping of protein-ligand interactions of LytR with ellagic acid and fisetin.	82
Figure 3.17 - TSA results of Bcl-2 incubated with the 4H- chromene derivatives.	84
Figure 3.18 - TSA of Bcl-2 in different percentages of DMSO.	84
Figure 3.19 - TSA of Bcl-2 incubated with venetoclax and the indole derivative in 2% DMSO. .	85
Figure 3.20 - TSA results of Bcl-2 with venetoclax and the indole derivative.	85
Figure 3.21 - TSA results of LytR incubated with ellagic acid and fisetin.	86
Figure 3.22 - TSA results of LytR incubated with ellagic acid and fisetin in 2% DMSO.	86
Figure 3.23 - Urea gel of unbounded Bcl-2 and upon incubation with the ethoxy, activemethine and indole derivatives.	87
Figure 3.24 - Urea gel of unbounded Bcl-2 and upon incubation with the indole derivative and venetoclax.	87
Figure 3.25 - Results of the ITC experiments of Bcl-2 with the indole derivative and venetoclax.	88
Figure 3.26 - ITC experiment of LytR with fisetin.	89
Figure 3.27 - CD spectra of LytR in the absence and presence of fisetin.	90

Figure 3.28 - Co-crystal of Bcl-2 with the indole derivative.	91
Figure 3.29 - Co-crystal of Bcl-2 with venetoclax.	91
Figure 3.30 - ¹ H and STD NMR spectra of Bcl-2 incubated with the indole derivative.	93
Figure 3.31 - ¹ H and STD NMR spectra of LytR incubated with ellagic acid.	93
Figure 3.32 - ¹ H and STD NMR spectra of LytR incubated with fisetin.	94
Figure 3.33 - Mapping of the Bcl-2 residue occupancy probability of venetoclax, ethoxy and indole derivatives, from MD simulations.	94
Figure 3.34 - Rmsd and essential dynamics analysis of the MD simulations of native form Bcl-2 and in the presence of the ligands under study.	95
Figure 6.1 - ¹ H NMR spectrum of the indole derivative.	124
Figure 6.2 - ¹ H NMR spectrum of ellagic acid.	124
Figure 6.3 - ¹ H NMR spectrum of the fisetin.	125

List of tables

Table 2.1 - Precipitant solutions derived from published structures for Bcl-2 crystallization.	48
Table 2.2 - Experiment details of the STD-NMR spectra of Bcl-2 and LytR.	55
Table 2.3 - Specific input parameters for MD simulations of each protein-ligand complex studied for Bcl-2.	56
Table 2.4 - Specific input parameters for MD simulations of each Bcl-2 nsSNP studied.	57
Table 3.1 - Docking results of Bcl-2 and the 4H-chromene derivatives, including venetoclax as a reference.	70
Table 3.2 - List of the best hits from the ligand-based virtual screening of the 4H-chromene moiety and the HA14-1 molecule.	71
Table 3.3 List of docking results of Bcl-2 with the azaindole derivatives' library, using AutoDock Vina.	75
Table 3.4 - List of docking results from structure-based virtual screening of Bcl-2.	78
Table 3.5 - Docking results of LytR with the reference lipidic substrates and the candidate inhibitors under study.	81
Table 3.6 - Potential energy and essential dynamics (ED) analysis of the MD simulations of the Bcl-2 complexes.	95
Table 3.7 - List of Bcl-2 nsSNP with deleterious phenotype prediction.	96
Table 3.8 - List of Bcl-2 nsSNP under study and correspondent potential energy and essential dynamics analysis from the MD simulations.	97
Table 3.9 - Docking results of Bcl-2 nsSNP with venetoclax and the indole derivative.	97
Table 6.1 - List of buffer solutions present in the buffer screening performed for Bcl-2.	119
Table 6.2 - List of additive solutions present in the additive screening performed for Bcl-2. ...	120
Table 6.3 - List of TSA results for Bcl-2 in the buffer and additive screenings.	121

List of abbreviations

α - Wave phase
 λ - Wavelength
 ΔG – Gibbs free energy
ADME – Absorption, Distribution, Metabolism, Excretion
AP – Anionic polymers
Bcl-2 – B-cell lymphoma 2
CCD – Cyclic coordinate descent
CD – Circular dichroism
CYP450 - Cytochrome P450
Da - Dalton
HTS – High throughput screening
HMM – Hidden Markov models
ITC – Isothermal titration calorimetry
 K_D – Dissociation constant
LCP – LytR-Cps2A-Psr
MAD – Multiple wavelength anomalous dispersion
MD – Molecular dynamics
MIR – Multiple isomorphous replacement
MOM – Mitochondrial outer membrane
MR – Molecular replacement
MW – Molecular weight
NMR – Nuclear magnetic resonance
nsSNP – Non-synonymous single nucleotide polymorphism
OD – Optical density
PCD – Programmed cell death
PDB – Protein data bank
PG - Peptidoglycan
RMSD – Root-mean-square deviation
SAD – Single wavelength anomalous dispersion
SAXS – Small angle X-ray scattering
SBDD – Structure-based drug design
SBVS - Structure-based virtual screening
SPR – Surface plasmon resonance
STD – Saturation transfer difference
 T_m – Melting temperature
TSA – Thermal shift assay
WTA – Wall theicoic acids
UV – Ultra-violet

1. Introduction

1.1. Drug Discovery

1.1.1. Historical Perspective

Since the 19th century, drug research has gained exponential interest and resourcefulness, allowing numerous triumphs in understanding and treating several diseases. At that time, chemistry had become more insightful and its principles started to be applied to different fields, specially pharmacology¹. By the 1890s, the foundations of chemistry, such as Avogadro's atomic hypothesis, the periodic table of elements, the theory of acids and bases by Arrhenius and, also, Kekulé's theory on the structure of aromatic compounds, had been settled^{2,3}.

In medicine, the studies on the selective affinity of dyes for biological tissues led by Paul Ehrlich were a fundamental step, since they allowed to postulate the existence of "chemoreceptors"⁴. Later, Ehrlich proposed that a therapeutic advantage could be gained from the hypothetical differences between analogous chemoreceptors of parasites, microorganisms and cancer cells in comparison with host tissues.

Analytical chemistry also played an important role in drug research, since, during the 19th century, several active compounds from medicinal plants were isolated and purified. These were the cases of for example morphine and papaverine, in 1815 and 1848, respectively^{2,5} (Figure 1.1).

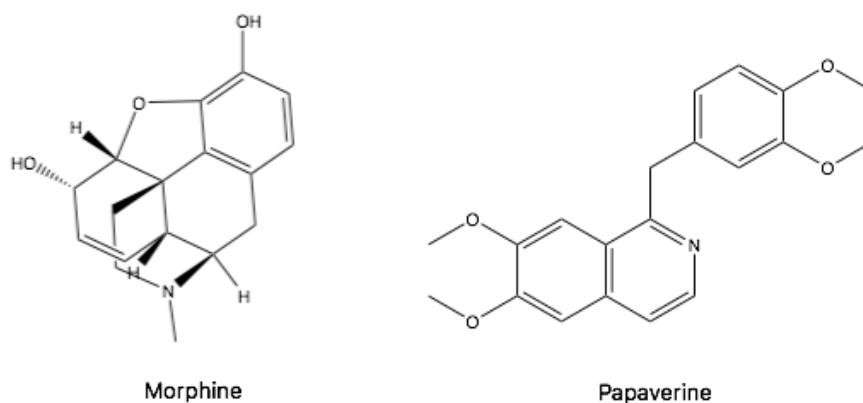


Figure 1.1 - Chemical structures of the first two active compounds isolated from medicinal plants, morphine and papaverine.

In the case of endogenous bioactive ligands, namely steroid hormones, they were identified even before their biomolecular target was established, isolated and structurally characterized⁶. However, their biological relevance could often be inferred, despite the lack of understanding of their underlying physiological mechanisms⁶. To overcome this trend, biochemistry was of pivotal importance, firstly, by proposing enzymes and receptors as good drug targets², and later, by enlightening and connecting the pathways and mechanisms of these biomolecular receptors⁶. In the beginning of the 20th century, discoveries concerning steroids and their impact in numerous physiological processes gave rise to several Nobel Prizes, through the 1920s and 1930s. In 1928, Adolf Windaus was rewarded for his research into the constitution of

sterols and their connection with vitamins. Paul Karrer's work on carotenoids, flavins and on vitamins A and B2 was acknowledged in 1937. In the year after, Richard Kuhn received the Noble Prize also for his discoveries regarding carotenoids and vitamins. Finally, in 1939, Adolf Butenandt was recognized for his work on sex hormones.

During the 20th century, X-ray crystallography emerged as a ground-breaking technique to study the three-dimensional structure of macromolecules, namely proteins. The realization that this knowledge could bring important light to drug research came in the 1980s and 1990s⁷. The rational design of potential drug candidates using the structure of their target proteins became a reachable goal. In 1990, the first examples where the use of this methodology was essential to yield successful inhibitors of HIV-1 protease were reported^{8,9}. Structure-based drug design (SBDD) became, therefore, a crucial step in the pharmaceutical industry drug discovery programs and one of the main research focuses in academic laboratories⁷.

In the past decades, two approaches emerged for the discovery of novel drug candidates, among the plethora of small molecules available, for a specific biomolecular receptor. As a predominantly empirical method, high throughput screening (HTS) promised to deliver new interacting compounds that demonstrate *in vitro* activity for the studied target. Through biophysical techniques, large libraries are screened giving rise to a few promising hits, that may require further optimization¹⁰. Meanwhile, advances in computer engineering and computational methods allowed the intervention of *in silico* techniques in drug discovery. Virtual screening comprises the application of computational methods, as quantum chemistry, molecular docking and molecular dynamics (MD), to screen extensive libraries, delivering novel drug candidates more quickly and with fewer costs⁶.

Nowadays, drug research is predominantly dependent on biological targets, genetic studies, transgenic animal models, molecular biology, gene technology and protein science⁶. Biological targets govern the drug discovery process, since their biological role, localization, physicochemical and structural properties define the potential drug molecules characteristics and their development. In the past decades, genetic studies have gained increasing importance in drug research, due to their insightful contributions to the genome constitution, helping to identify new targets and the phenomena that regulate their expression. Transgenic animal models are a fundamental part of understanding a target's biological impact and the clinical effects of potential drug candidates. The structural characterization of a target and its interactions with active compounds became a common approach for the rational design of potential drugs, giving to molecular biology a central role in drug research. Gene technology contributed with extensive knowledge regarding gene cloning, mutagenesis and expression, which are routine procedures, nowadays essential for molecular biology and protein science. The latter encompasses numerous biophysical and biochemical methods capable of characterizing a biological target and establish valuable information for rational drug design.

1.1.2. Overview

Drug discovery is driven by the need to find a therapeutic solution for a specific pathology. The initial step of this endeavour is often conducted by academia¹⁰, which attempts to develop a hypothesis that the inhibition or activation of a protein or pathway may result in an effective treatment of the disease state. The result of this effort is the identification and selection of a biological target which may require further validation prior to progression to the lead discovery phase¹⁰. The drug discovery process can be described by the combination of multiple phases: target identification, target validation, hit discovery process, hit series selection, hit-to-lead phase and lead optimization phase¹⁰ (Figure 1.2).

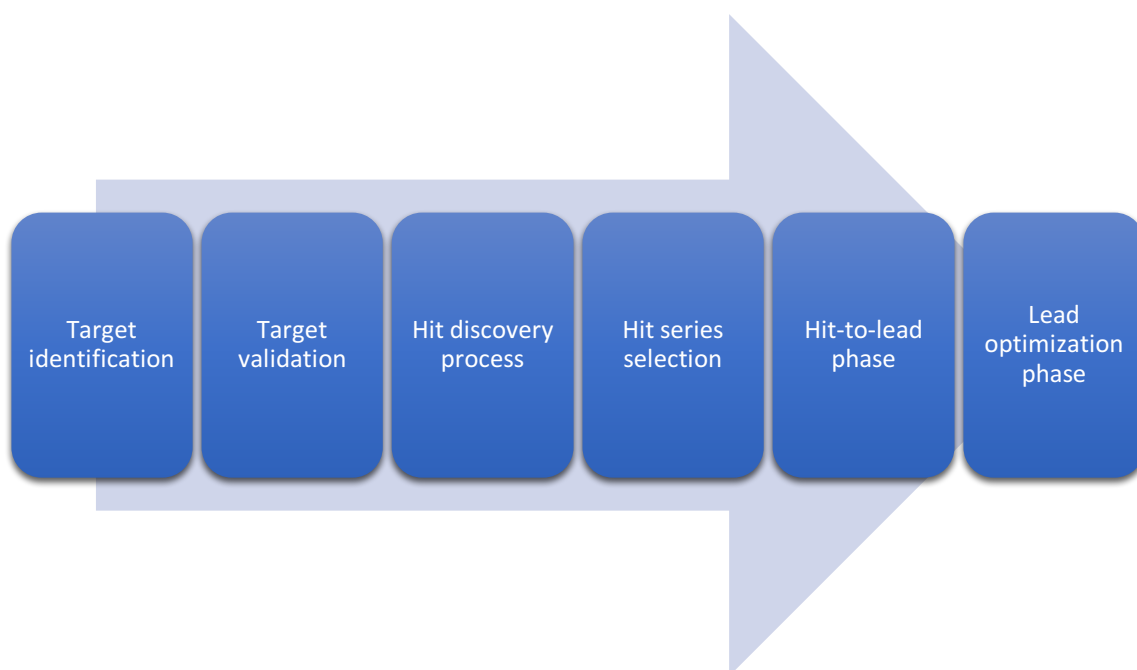


Figure 1.2 - Schematic representation of a drug discovery process.

1.1.2.1. Target Identification

The first step in drug discovery is target identification, which entails the selection and characterization of a macromolecule involved in the disease phenotype that either through inhibition or activation can generate the desired therapeutic response. Besides proteins, such as enzymes or receptors, genes and RNA may also be considered as targets¹⁰. A crucial parameter that defines a good therapeutic target is its “druggability”. A “druggable” target must be accessible to the drug molecule and provide a biological response that can be measured in both *in vitro* and *in vivo* assays¹⁰.

There are several approaches that may lead to successful target identification. A phenotypic screening can be performed to conclude, for example, which protein is associated to a specific disease state¹⁰. This can be the case of a protein which, upon overexpression, causes a pathology. Its identification is possible through techniques such as mass spectrometry¹¹. Another powerful approach is through genomics studies, which may indicate that a disease state

is caused by a genetic modification, as polymorphisms or translocations¹⁰. These alterations can jeopardise the protein's stability and function, and also promote or silence a protein expression. A concerted method is possible through computational efforts, namely data mining, which allows the integration of relevant biomedical data from publications, patents, genomics, proteomics, transgenic phenotyping and compound profiling to identify potential disease targets^{10,12}.

1.1.2.2. Target Validation

Once the target has been identified, its implications on the disease phenotype must be validated through *in vitro* techniques. One of the methodologies to validate a target is using antisense technology, in which RNA-like oligonucleotides are designed to be complementary to the mRNA precursor of the target¹⁰. Hence, the target's translation is suppressed by blocking the translation machinery. One major advantage of this technique is its reversibility when compared to the gene knockout approach¹⁰. The knockout approach is based on *in vivo* experiments, where transgenic animals that don't possess the gene of interest are used. A similar approach is called knock-in, where a non-enzymatically functioning protein replaces the endogenous target¹³. This approach allows the observation of the animal's response to an effective treatment, since the protein is expressed but functionally inhibited. The need to make tissue restricted and inducible knockouts led to the application of small-interference RNA (siRNA)^{10,14}. The process starts with the injection of double-stranded RNA (dsRNA) which is recognized by the cell as exogenous genetic material, activating the RNAi pathway. After dsRNA cleavage into small fragments, called siRNA, they are separated into single strands and a RNA-induced silencing complex is associated. This leads to recognition and cleavage of the endogenous target's mRNA, preventing its translation. A different approach to target validation is the use of monoclonal antibodies, as they can interact directly with the target¹⁰. Their specificity is a major advantage because they may discriminate between very closely related proteins, recognizing unique epitopes.

1.1.2.3. Hit Discovery Process

The hit discovery process relies on the detection of a small set of compounds that demonstrate effective interactions with the target under study. These "hits" are found through screening assays of large libraries of small molecules. Nowadays, numerous screening strategies are adopted¹⁰.

HTS is commonly used in the pharmaceutical industry. This technique involves the screening of entire libraries of compounds against the target molecule^{10,15}, usually through biophysical assays, such as absorbance-, fluorescent- and chemiluminescent-based methods¹⁶. An alternative is to screen the compounds in a more complex system, such as cell-based assays. However, this approach requires secondary assays to confirm the target specific interaction¹⁵. The HTS strategy entails a high-level of laboratory automation for the development of the screening assays, although no prior knowledge is required regarding the chemical affinity of the target's binding site^{10,15,17}.

Fragment screening focuses on building libraries of molecules with lower molecular weight, increased polarity and with enhanced solubility regarding drug-like compounds. These small molecules are screened at high concentrations against the target of interest and the successful hits allow the identification of building blocks capable of forming an inhibitor candidate¹⁸. The structure determination of the complexes between the target molecule and the hits, through X-ray crystallography or nuclear magnetic resonance (NMR), is valuable to enable compound progression¹⁰.

Physiological screening is an *in vitro* approach that screens compounds in tissue-based assays aiming to find hits that provide a response similar to the *in vivo* desired effect¹⁰. As in the cell-based assays in HTS, this methodology requires secondary experiments to determine target specificity.

Focused or knowledge-based screening comprises filtering large libraries to provide restricted small molecule subsets with chemical properties known to have activity toward the target molecule¹⁹. This knowledge is gathered from the literature and gave rise to the computational approach of virtual screening¹⁰. The search for common antagonists of the MDM2/4-p53 systems is a great example of the application of the knowledge-based screening approach²⁰. In this case, MDM2 and MDM4 are homologue proteins which inhibit the tumor-suppressing protein p53 by different mechanisms. Thus, combining inhibition of both homologues would activate p53 more significantly than only antagonize one of them. Through X-ray crystallography and NMR, structural information was obtained regarding the MDM2-p53 complex interactions and also between MDM2 and ligands with nanomolar potency. This knowledge led to the realization that fundamental interacting residues were conserved, at least in their hydrophobic nature, between MDM2 and MDM4. This was the premise for the identification of novel selective and dual MDM4 and MDM2 inhibitors, through knowledge-based screening, using the MDM2-p53 system as reference.

Virtual screening is an *in silico* method that allows the screening of vast small molecule databases more quickly and cost-efficiently, filtering the promising candidates that should be tested in the laboratory^{6,7,21}. Through molecular docking and scoring algorithms, the screened compounds are ranked according to the predicted affinity to the target molecule^{6,21}. To do so, the target's three-dimensional structure is required, which may be provided by experimental methods, as X-ray crystallography and NMR, or by computational methods, such as homology modeling and *ab initio* calculations^{7,10,21}.

All the above-mentioned methodologies require biophysical and biochemical techniques for hit detection. These experiments may be applied to the individual target, cell cultures, or even grown tissues.

HTS that encompass a wide variety of active compounds, by definition are unlikely to yield hits for the specific target compared to the plethora of biomolecules present in a cell²². This evidence suggests that cell-based assays may not be amenable to HTS screening strategies. The development of techniques to isolate and maintain primary cell cultures and the devising of

appropriate probes to detect compound activity toward the target of interest would greatly benefit cell-based techniques' implementation²².

Assays concerning only the individual target may provide potential chemotypes with simple structure-activity relationship and mechanism of action, resulting in the best approach to the hit discovery and hit-to-lead optimization phases²². Further cell-based studies would contribute with cell penetration, activity and stability evaluations²². However, in the case of an overwhelming number of hits emerging in the hit discovery phase screenings, the cell-based approach becomes a suitable alternative to preselect a subset of compounds with presumed specific cellular activity²².

1.1.2.4. Hit series selection

To avoid an exaggerated number of hits resulting from an active compound screening, which would diminish the chances of reaching an effectively promising hit, a triaging process must be performed. The first measure to be taken is the removal of compounds which emerge frequently as hits in various assays from the hit series library¹⁰. Another important approach is to use computational chemistry algorithms developed to group compounds according to their structural similarity¹⁰. Through this filtering, promiscuous compounds will be excluded from the hit series, which will possess a broad spectrum of chemical classes to be tested.

Several aspects impact the hit series selection in order to provide promising hits during the drug discovery process. These properties are associated with drug-likeness, toxicity and pharmacokinetics (ADME) (Figure 1.3).

The compounds present in a small molecule library are usually in accordance with some criteria that confers them drug-likeability properties, such as the Lipinski rule of five²³. This combination of characteristics used to select compounds arises from the statistical realization that most successful commercially available drugs can be described by a few parameters¹⁰. In the case of the Lipinski rule of five, these conditions are: molecular weight lower than 500 Da, cLogP (measure of lipophobicity which affects pharmacokinetics) less than 5, no more than 5 hydrogen bond donor atoms and less than 10 hydrogen bond acceptor atoms²³.

In many cases, reversibility is of great interest regarding the interactions between the hit and its target, ensuring that the drug is metabolized and excreted from the patient's body¹⁰. To assess this property it's pivotal to perform dose-responsive curves in the primary hit discovery assays¹⁰. Secondary assays may be focused on examining the surviving hits in a cell-based approach, which would provide a functional response to the compound¹⁰.

Chemical synthesis has also an important role in defining a hit series, since criteria like synthetic route, derivatization potential and amenability to parallel synthesis, are looked for, in a hit-to-lead optimization phase¹⁰.

Further filtering of a hit series is progressively carried out mainly by *in vitro* assays designed to provide information regarding the pharmacokinetics parameters of absorption, distribution, metabolism and excretion (ADME) and, also, the specificity profile of the hits under study¹⁰.

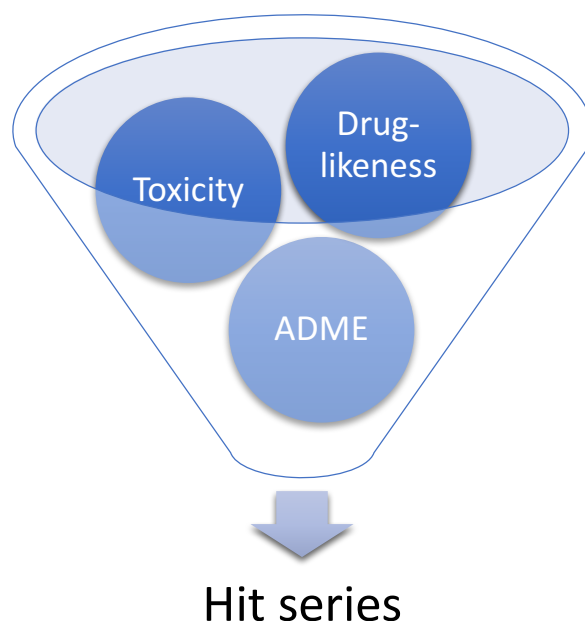


Figure 1.3 - Schematic representation of the properties assessed during a hit series selection.

1.1.2.5 Hit-to-lead phase

The purpose of the hit-to-lead phase is to refine the promising hit series detected previously. This refinement is meant to improve compound potency and selectivity besides its pharmacodynamic and pharmacokinetic properties¹⁰. The latter can be assessed through cell-based techniques, using *in vivo* models, while the target-hit relationship is studied by biophysical and biochemical assays.

To establish the structure-activity relationship between the hits and the target, a structure-based drug design approach is usually implemented¹⁰. This entails structural characterization through experimental techniques such as X-ray crystallography and NMR, namely saturation transfer difference NMR (STD-NMR), but can also imply *in silico* methods like molecular dynamics (MD)^{10,21,24}. This methodology provides details at the atomic level, which are pivotal to describe the mechanism of interaction between a hit and its target.

Compound potency and selectivity investigation can be achieved by several biophysical and biochemical techniques that focus on different aspects of the protein-ligand binding phenomenon. Some examples of such techniques are: thermal shift assay (TSA), urea-polyacrylamide gel electrophoresis, isothermal titration calorimetry (ITC), circular dichroism (CD), STD-NMR, small angle X-ray scattering (SAXS), surface plasmon resonance (SPR), UV/Visible spectroscopy, and fluorescence polarization/anisotropy²⁴⁻³¹.

In the case of TSA, the outcome of the protein-ligand interaction probed is the increase in protein stability upon ligand binding, which is reflected in the protein's augmented resistance to thermal denaturation²⁷. Regarding techniques such as urea-polyacrylamide gel electrophoresis, CD and SAXS the biophysical property observed is the structural difference between the native protein and the protein-ligand complex^{24,30,31}. SPR, UV/Visible spectroscopy,

and fluorescence polarization/anisotropy binding assays are based on the premise that a protein-ligand complex will change the electronic properties of the protein, promoting variations upon interaction with light²⁴.

Other considerations must be taken in the hit-to-lead phase regarding the pharmacodynamic and pharmacokinetic processes. Solubility and permeability of a compound is evaluated in order to devise the appropriate delivery strategy to its target, either by injection or oral uptake¹⁰. Cytochrome P450 (CYP450) inhibition is also investigated, since it is a pivotal element of the patient's metabolism, which may suffer from undesired drug interference¹⁰. Other metabolizing enzymes such as aldehyde oxidase are becoming more popular among medicinal chemists, due to their ability to react with scaffolds that were originally devised to circumvent CYP450 activity^{32,33}.

1.1.2.6. Lead optimization phase

The aim of this final stage of the drug discovery process is the improvement of the leading compound's undesirable characteristics while maintaining its promising properties¹⁰.

Optimization may be focused on several issues such as compound potency, selectivity, stability or solubility. After surpassing these hurdles the leading candidate may move to a preclinical stage, while other assessments continue being made. Genotoxicity models, such as the Ames test³⁴, are commonly used to examine the lead besides *in vivo* behaviour evaluation through, for example, the Irwin's test³⁵.

Pharmacodynamic and pharmacokinetic studies must be conducted in this final stage, including high-dose response, dose linearity and repeat dosing pharmacokinetic assessments¹⁰. These generate drug-induced metabolism and metabolic profiling evidence of the lead compound that may be extrapolated to the patient's organism.

1.2. Structure-based Virtual Screening (SBVS)

The pharmaceutical industry approach to drug discovery is primarily focused on HTS, which comprises several drawbacks such as high-cost, time-demand and mechanistic uncertainty^{6,21}. Academia's research is usually directed toward SBDD and computational methods to avoid those issues and to allow a more rational and iterative design of drug molecules.

Virtual screening, as an alternative to HTS, provides hits for a target from commercially available compound libraries, prior to experimental testing, in a more quick and cost-effective way²¹. These compounds are selected by their predicted affinity to the target of interest and progress to the hit-to-lead phase by means of SBDD. SBVS is based on the search for active compounds of a structural characterized target, through a computational approach. It comprises several stages that may be identified as: target structural determination, virtual screening, target-ligand complex characterization and target-ligand structural determination.

1.2.1. Target structural determination

The first step of a SBVS strategy entails the determination of the three-dimensional structure of the target of interest. This endeavour is predominantly achieved through X-ray crystallography, although NMR and computational techniques, such as homology modeling and *ab initio* methods, may also contribute to this end^{7,10,21}. In the work here described, X-ray crystallography and homology modeling, together with *ab initio* methods, were the main contributors to the structural knowledge required for the virtual screening approach.

1.2.1.1. X-ray crystallography

To determine the three-dimensional arrangement of macromolecules, namely proteins, X-ray crystallography has always been the main technique employed. It is based on the principle that to interact with matter, light must have a relatable wavelength³⁶. So, in order to obtain atomic resolution, minding that the atomic bond has around 1.5 Å of length, X-rays are the electromagnetic radiation of choice. The basic principle behind X-ray crystallography is that X-rays are scattered by the electron density of the protein's atoms, resulting in experimental evidence that lead to the determination of the atom's positions. These experimental results are not enough, since some of the information required for structure determination is lost during the X-ray diffraction data collection, giving rise to the "Phase Problem". Nowadays, there are several methods to overcome this hurdle, however, reaching this stage implies that another challenging part of X-ray crystallography has already been solved. This limiting step is protein crystallization, which delivers the experimental sample submitted to the X-rays. Therefore, this technique comprises several steps: protein crystallization, X-ray diffraction and structure determination (Figure 1.4).

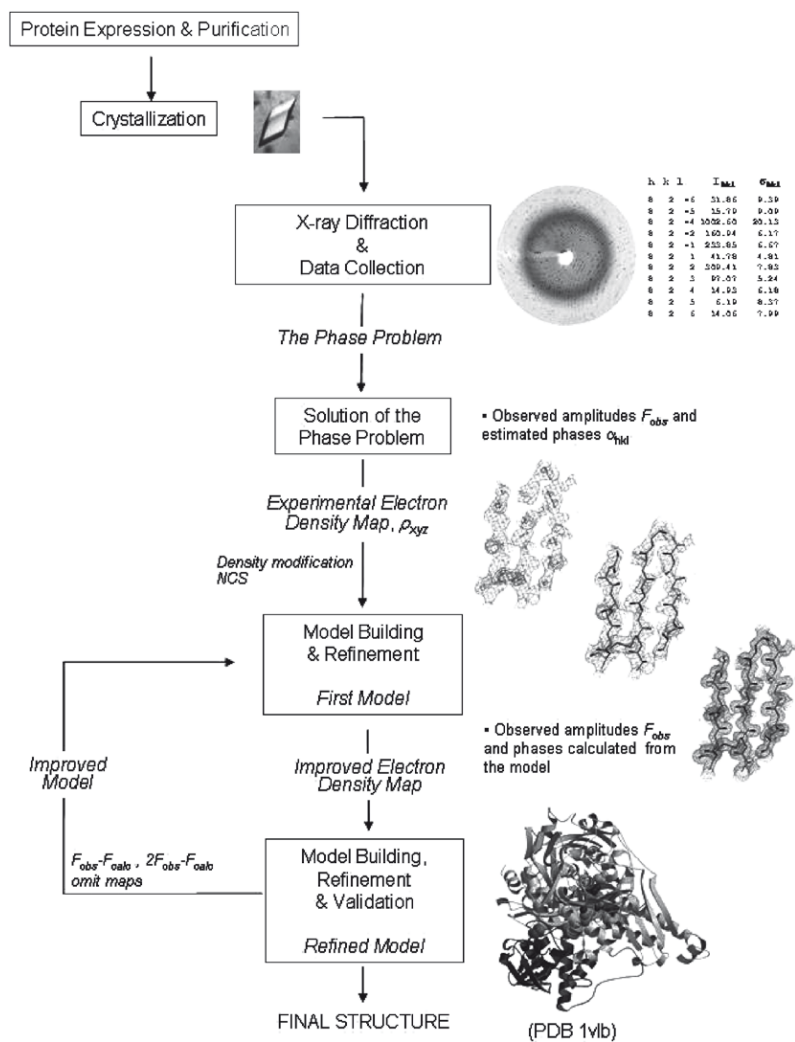


Figure 1.4 - Schematic representation of the X-ray crystallography workflow. The first step comprises the protein crystallization, which should be optimized to yield high-quality crystals. The scattering of the X-rays caused by the atom's electrons in the crystal are recorded and the "Phase Problem" is solved through well-established methods. Afterwards, the first protein model is built and successive stages of refinement and validation are performed to yield the final crystallographic model.⁴⁰

1.2.1.1.1. Protein crystallization

Protein crystals are three-dimensional ordered entities of repeated units, containing the protein atoms. They are formed through controlled precipitation of protein molecules from aqueous solution in specific physicochemical conditions. Crystals can be divided in their basic unit, named unit cell, which corresponds to its smallest and simplest volume element, described by its parameters a , b and c as edges and α , β and γ for the angles³⁶ (Figure 1.5). The unit cell contains the asymmetric unit which represents the smallest entity capable of forming the unit cell through crystallographic symmetry operations³⁶.

The systematic repetition of the unit cell that forms a protein crystal is the key to X-ray diffraction amenable to provide information for structure determination.

The reason why protein crystallization is such a challenging endeavour is related to the uncountable physicochemical variables involved in this process, that ranges from sample preparation, solution pH and ionic strength to temperature or precipitating agent used³⁷. Also, proteins are very complex macromolecules stabilized by intramolecular hydrogen bonds and solvent shells³⁶. The goal is to promote protein

stabilization at extremely high concentrations which allows their controlled precipitation, maintaining non-covalent interactions essential to the crystal's ordered nature.

Protein crystallization can be achieved through several methods: vapor diffusion, microbatch under oil, or microdialysis³⁶⁻³⁸. The most common method used is vapor diffusion, because it uses low amounts of protein and can be implemented through two different techniques, hanging or sitting drop (Figure 1.6). The vapor diffusion method consists in having a protein drop into which is added a precipitant solution, also present in the reservoir. The reservoir and the protein-precipitant drop are sealed, usually with a coverslip, and water in the form of vapor is transferred from the protein drop to the reservoir solution, since the precipitant concentration is lower in the drop and an equilibrium is forced in this closed system.³⁶

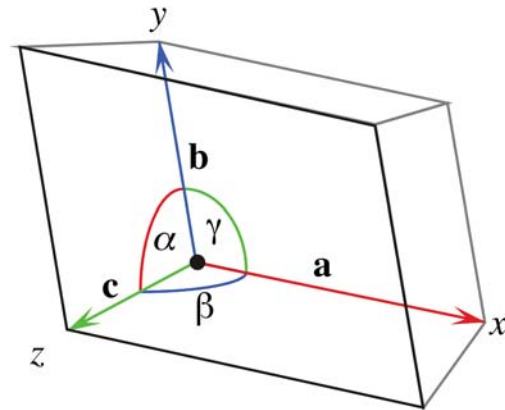


Figure 1.5 - Representation of a (triclinic) unit cell with its defining parameters: edges (a , b , c) and angles (α , β , γ).³⁶

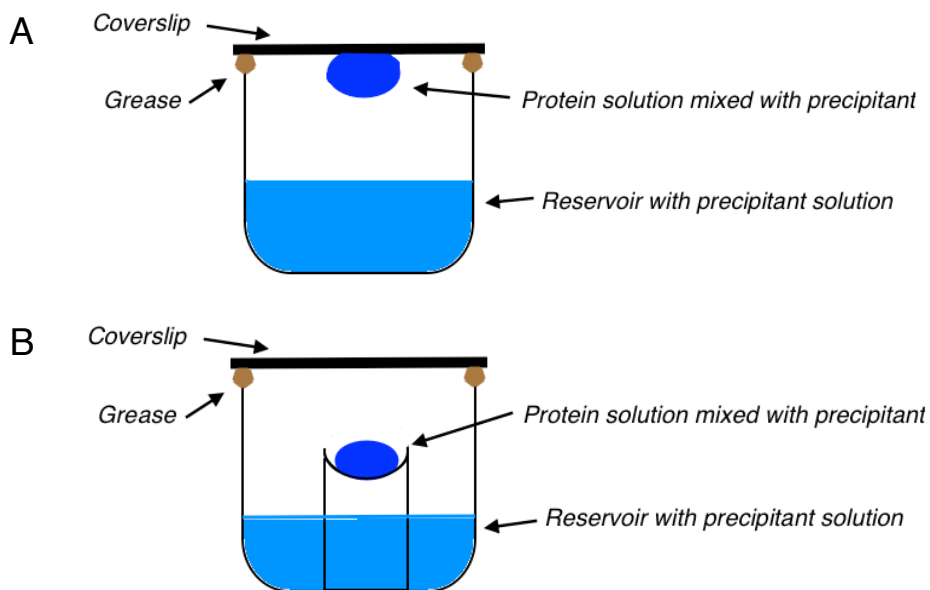


Figure 1.6 - Schematic representation of the vapor diffusion method associated with the hanging (A) and sitting (B) drop techniques.

The protein crystallization process entails many stages, which can be followed and understood through a phase diagram (Figure 1.7)³⁶⁻³⁸. Starting with a protein solution at high concentration and purity, the precipitant is added and the system is closed. At this initial stage, the drop is undersaturated. Then, as the drop's water content starts to decrease, the solution becomes supersaturated, hopefully reaching the labile region of the phase diagram. In the case of excessive protein and/or precipitant concentrations, the system finds itself in the precipitation region, where undesired amorphous aggregation of protein molecules occurs. If the conditions are favourable, nucleation begins, meaning that the first protein molecules start to interact non-covalently, creating ordered and stable nuclei. Afterwards, the nuclei turn into crystals, with fewer protein molecules in solution, which corresponds to the metastable zone.

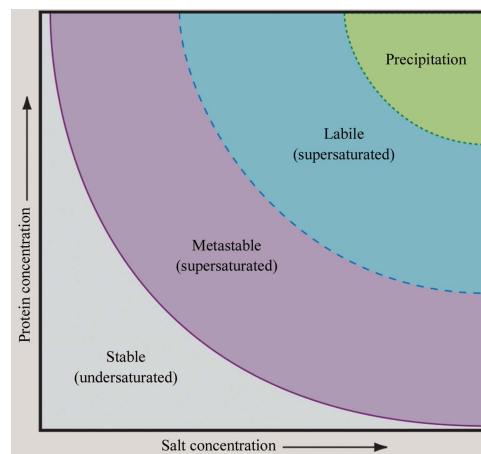


Figure 1.7 - Representation of the Phase Diagram used to guide a protein crystallization endeavour.³⁷

In the case of microbatch under oil, the protein solution is mixed with the precipitant and submerged under mineral oil. This prevents any vapor diffusion events and only the precipitant agents are responsible for the crystallization phenomenon.³⁸

The above-mentioned phase diagram is an effective guide for the crystallization process, however it concerns solely the relationship between protein and precipitant concentrations. One of the most time-consuming steps is finding the promising preliminary precipitant solutions that

through optimization can lead to suitable crystals for X-ray diffraction. Currently there is a plethora of crystallization screens commercially available, from different companies as Hampton Research, Jena Biosciences and others, covering a broad range of conditions and designed to provide the desired preliminary crystallization conditions. Further optimization of these conditions is usually required and can be achieved by variation of an enormous number of parameters: methodology, protein and precipitant concentrations, reservoir volume, protein-precipitant drop ratio, buffer pH, addition of salts/polymers/organic molecules/detergents, temperature and many others^{37,38}.

1.2.1.1.2. X-ray diffraction

After obtaining an appropriate protein crystal, it is mounted on a goniometer head and exposed to a collimated, monochromatic and intense X-ray beam. The result is a diffraction pattern captured in a detector plate where the diffracted X-rays collide (Figure 1.8). The three most common sources of X-rays are: X-ray tubes, rotating anode tubes and synchrotrons.³⁶

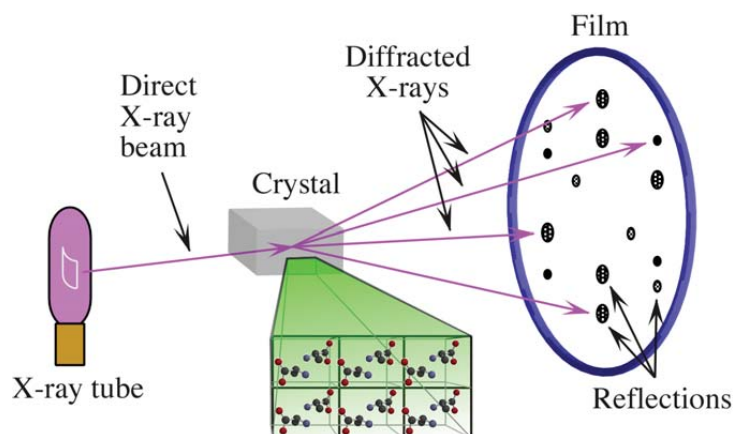


Figure 1.8 - Schematic representation of the X-ray diffraction experiment. An intense X-ray beam is directed at the protein crystal (a three-dimensional array of repeated and ordered units). The X-rays are diffracted by the atom's electrons and are recorded in a detector (film).³⁶

In house diffractometers commonly use X-ray tubes to produce X-rays, where a heated filament generates electrons that are accelerated by an electric field toward a metal target (usually copper, molybdenum or chromium). This high-energy electron collides with the metal atoms and displaces an electron from a low-lying atomic orbital. Then, an electron from a higher orbital occupies the vacant lower orbital position, emitting its excessive energy as an X-ray photon. The characteristic orbitals from each metallic element provide a wide range of wavelengths of the resulting X-rays.³⁶

Synchrotron sources (Figure 1.9) produce X-rays as a consequence of electron acceleration through increasing magnetic fields, that are synchronized with the electrons kinetic energy³⁹. These electrons travel through a closed-loop path and are conserved in particle storage rings. The bending phenomenon of the accelerated electrons through the application of magnets produces tangential X-ray photons that are captured by beamlines. Synchrotron radiation has

several advantages compared to *in house* sources: higher brilliance, tuneable wavelength and bandwidth. Indeed, collection of a complete data set from a crystal using synchrotron radiation can be achieved in several seconds, while the same experiment can take hours using *in house* sources, which are also limited to lower resolution limits.³⁶

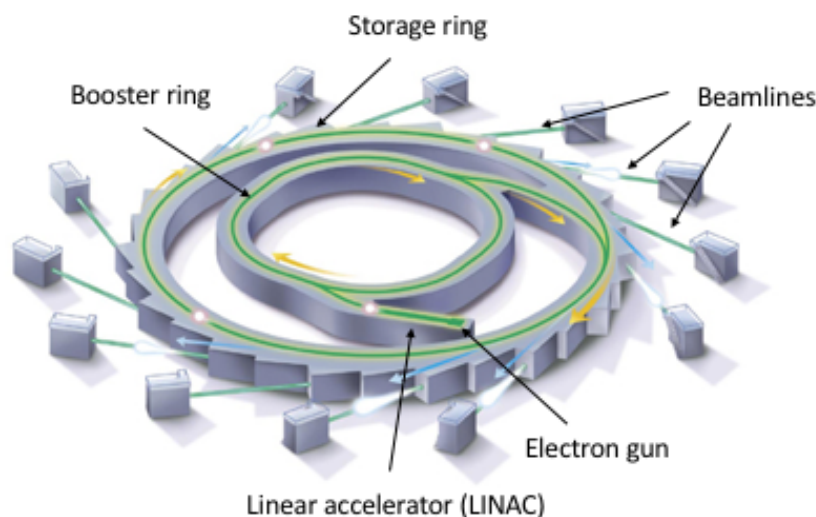


Figure 1.9 – Schematic representation of a synchrotron

(adapted from: <http://www.synchrotron.org.au/synchrotron-science/what-is-a-synchrotron>).

The X-ray diffraction experiment generates a diffraction pattern comprising reflections, which are the spots recorded in the detector plate where the diffracted X-rays collided. This phenomenon occurs according to the Bragg's law, through the constructive and destructive interference of electromagnetic waves. These reflections correspond to the reciprocal space and are indexed using the coordinates hkl (Miller indices). The reciprocal space is depicted in the diffraction pattern recorded and has an inverse relationship with the real space where the protein structure is represented. The Bragg's law postulates that the reflections in a diffraction pattern are the result of constructive interference between the electromagnetic waves diffracted by all the atoms of the crystal (Figure 1.10). The constructive combination of the diffracted waves gives rise to different intensities for each reflection. This is the other important parameter for structure determination, the reflection's intensity (I_{hkl}).³⁶

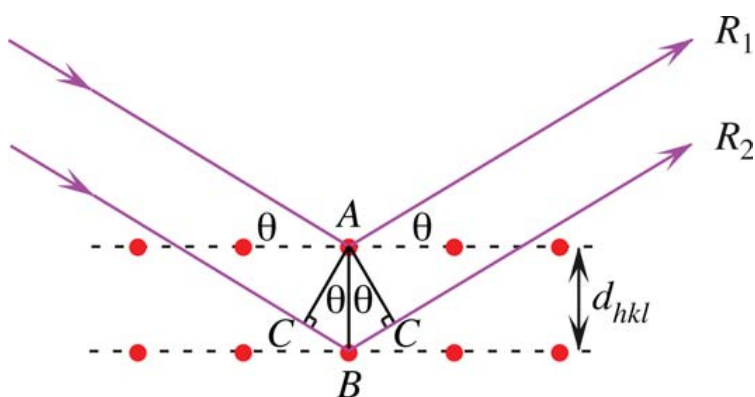


Figure 1.10 - Schematic representation of the Bragg's law, which is based on the constructive interference of waves. The reflection resulting of the sum of X-ray 1 and 2 (R_1 and R_2) has strong intensity if the distance between planes (d_{hkl}) is an integral multiple of the wavelength.³⁶

1.2.1.1.3. Structure determination

What crystallographers try to accomplish is to describe the electromagnetic waves diffracted by the atom's electron densities, which requires three physical parameters: wavelength (λ), amplitude (F), and phase (α). The first one is defined during the X-ray diffraction experiment, since it is tuneable by the user and intrinsic of the apparatus. The second one is measured experimentally, being directly related to the reflection intensities recorded on the detector plate. The last one is the missing information that originates the "Phase Problem" in crystallography.^{36,40}

Electromagnetic waves are periodic functions and even the most intricate ones can be described by the sum of sine and cosine functions. These sums are called Fourier series and are used to represent the X-rays diffracted by the unit cell's atoms as structure-factor equations, applied to each reflection recorded (F_{hkl}) (Equation 1). Experimentally, these structure-factors are determined by calculating the square-root of the relative intensities of each reflection and taking into account correction factors (F_{obs}).³⁶

$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i(hx+ky+lz)} dV$$

Equation 1 - Structure-factor equation for the reflection hkl , where $\rho(x, y, z)$ is the electron density and V is the unit cell volume.

Through a Fourier transform, it is possible to convert the structure-factors into the average electron density of a volume element, centred at x, y, z (Equation 2). This transform switches the integral into a triple sum, because the F_{hkl} represent discrete values, which are the reflections in the diffraction pattern. The equation expresses the desired electron densities as function of the known $|F_{hkl}|$ and the unknown α_{hkl} . To calculate the electron density map, the estimation of the phase angles for each reflection is required. This challenge can be surpassed by different methods: multiple isomorphous replacement (MIR), single or multiple anomalous dispersion (SAD/MAD) or molecular replacement (MR). MIR is used by the preparation of heavy atoms derivatives in crystals with identical unit cell parameters. These heavy atoms originate differences in the intensity of some reflections that when applied in Patterson maps enable the localization of these atoms in the unit cell, thus allowing the estimation of initial phases. SAD or MAD are based on the characteristic X-ray absorption of some atoms, such as selenium, which alters the intensity of symmetry-related reflections, the Friedel's pairs. After obtaining the atoms' positions in the unit cell, through Patterson methods, it is possible to estimate the phases of all reflections. However, if there are solved homologous structures of the protein of interest, MR is commonly used. This method comprises the application of a known structure as a phasing model. Since the phases of atomic structure-factors depend on the structure position and orientation in the unit cell, the superposition of the model to the unknown structure is searched. Using Patterson maps of both the known and the unknown structures, rotational and translational operations are performed to achieve the best superposition, which is judged by the agreement between the amplitudes of the

calculated structure-factors, $|F_{calc}|$, and the measured amplitudes, $|F_{obs}|$. The computed phases of the model structure-factors are used as estimates for structure determination of the unknown structure.^{36,40}

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i (hx + ky + lz - \alpha'_{hkl})}$$

Equation 2 - Equation of the atom's electron density as a function of the structure-factors, where α'_{hkl} is the unknown phase that must be estimated to solve the structure of interest.

After solving the phase problem and solving the structure through the experimental structure-factors and the estimated phases, the model is submitted to several cycles of refinement to yield a reliable crystallographic model. To achieve this, the crystallographer must interpret the model and adjust the atom's coordinates consonant to the suggested electron densities. Then, through Fourier transforms, these electron densities are reconverted into calculated structure-factors (F_{calc}) and better phases can be retrieved from these data. Statistical parameters are important to evaluate a model's quality, such as the R and R_{free} (Equation 3). Both terms reflect the amount of data that is not corroborated by the experimental results. In the case of the R_{free} , a percentage of the original data set, usually 5 to 10%, is used for comparison with the F_{calc} , since this data was removed from the refinement calculations and hence it is not affected by misinterpretation of the model.^{36,40} Other parameters as root-mean-square deviation (rmsd) of bond lengths and angles or the Ramachandran plot are also used to evaluate model geometry and, more generally, reliability.

$$R = \frac{\sum ||\mathbf{F}_{obs}| - |\mathbf{F}_{calc}||}{\sum |\mathbf{F}_{obs}|}$$

Equation 3 - Equation of the R-factor, which determines the amount of data that do not agree with the experimental results.

1.2.1.2. Homology Modeling

Currently, if the therapeutic target's structure is unknown and some homologue protein structures have been already solved, homology modeling poses as an effective approach for structure prediction in the context of SBVS. The basic principles of this technique are:

- protein structure is more conserved throughout evolution compared to protein sequence;
- there is evidence of finite number of unique protein folds in nature⁴². Hence, proteins with similar sequence, at least 30% homology, tend to adopt the same fold⁴³.

Programs like Phyre2⁴⁴, ModWeb⁴⁵ and SWISS-MODEL⁴⁶ can be used for structure prediction through homology modeling. In the present work, the Phyre2⁴⁴ server was chosen and its normal procedure encompasses the following steps (Figure 1.11): 1) sequence homology search against a non-redundant database with less than 20% identity between sequences, multiple sequence alignment, using PSI-BLAST⁴⁷, secondary structure prediction through PSIPRED⁴⁸ and conversion of both into a hidden Markov model (HMM) profile; 2) HMM profiles' search against a database and main-chain model building; 3) insertions and deletions in the sequence built through loop modeling; 4) addition of side-chains to the model.

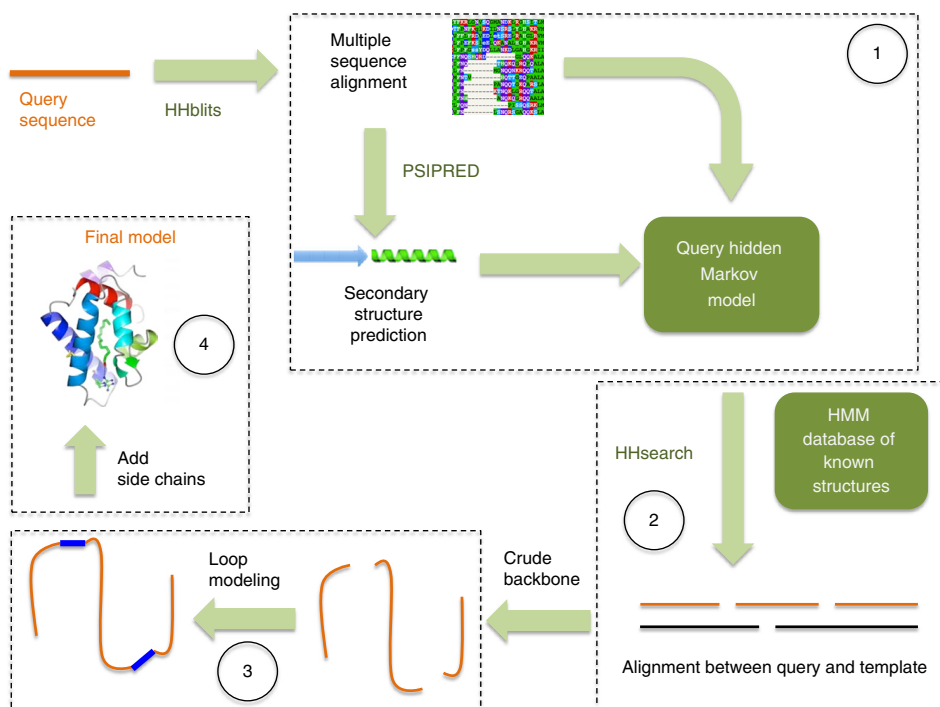


Figure 1.11 - Schematic representation of the normal mode protocol followed by the Phyre2⁴⁴ server to predict a protein structure through homology modeling.⁴⁴

The first stage of homology modeling performed by Phyre2⁴⁴ is finding the templates based on sequence homology between the protein of interest and an extensive database. The only input required for the Phyre2⁴⁴ server is the protein sequence of interest inserted manually or in a FASTA format file. HHblits⁴⁹ is the method used for templates' retrieval from the protein

sequence database. It comprises the conversion of the query sequence to a HMM profile, by producing evolutionary sequence alternatives through mutations to physicochemical similar residues and calculating the probabilities of each amino acid for each position. Then, the HMM profile is searched against the sequence-to-profile database and successive refinements of the HMM profile are performed⁴⁹. In addition, the protein's secondary structure is predicted based on the query sequence by PSIPRED⁴⁸. This process involves neural networks trained in protein sequence profiles and allows the prediction of structural elements, such as α -helices, β -sheets and coils.⁴⁴

After the HMM profile calculation, the secondary structure predicted is converted into a HMM profile, giving rise to a merged HMM profile that defines the protein of interest. Afterwards, an HMM-HMM search is performed, by HHsearch⁵⁰, resulting in a list of high-scored templates of known structure. This generates crude backbone models that do not contain side-chains.⁴⁴

To predict the target protein structure with insertions or deletions, or even with regions where the sequence differs substantially from the templates, loop modeling is applied⁵¹. In the Phyre2⁴⁴ protocol for insertions, a library of fragments with known structures is used based on sequence-profile matching, creating a list of potential useful fragments with similar sequence and endpoint distances. Similarly, for deletions a sequence window on either endpoints is used for fragment search in the mentioned library⁴⁴. To model these fragments the cyclic coordinate descent (CCD) algorithm is applied, where each degree of freedom is adjusted at a time to move the end effector toward the target endpoint⁵¹.

The final step of the process involves side-chain's placement and fitting to the backbone model derived in the previous stages. This is achieved using the residue-rotamer-reduction (R3)⁵² protocol which is a graph-based technique that allows elimination of residues and rotamers via global optimization procedures. In this way, the lower energy conformations are preferred and successive iterations eliminate alternatives and simplify the residue graph generated⁵².

In the case of the intensive mode of Phyre2⁴⁴ modeling (Figure 1.12), the standard protocol is followed by generating several structures concerning homology modeling to different regions of the protein sequence. Then, the C α -C α distance-constraints from each structure are extracted. These portions are maintained while the sequences without templates are modeled through *ab initio* methods, using the Poing algorithm⁵³. Afterwards, the protein model is built from the main-chain representation, through the Pulchra software⁵⁴. Finally, the side-chains are added and fitted by the R3 protocol⁵².

The main difference in the homology modeling procedure concerns the application of the Poing⁵³ algorithm, that includes *ab initio* modeling of the sequence regions without reliable templates. This method aims to replicate physiological dynamics by synthesizing the protein's backbone through a virtual ribosome. While the regions with identified templates are restrained and modeled as inelastic springs, the portions without homologue sequences are modeled by Poing's solvent bombardment model, including secondary structure springs prediction and penalization of steric clashes^{44,53}. This algorithm is based on Newtonian equations derived from Langevin's dynamics, allowing the protein's folding prediction through solvent-residue interaction

calculations and secondary structure prediction using PSIPRED^{48,53}. The solvent bombardment model accounts for solvation effects of each residue, ensuring that hydrophobic amino acids are buried in the protein folding⁵³.

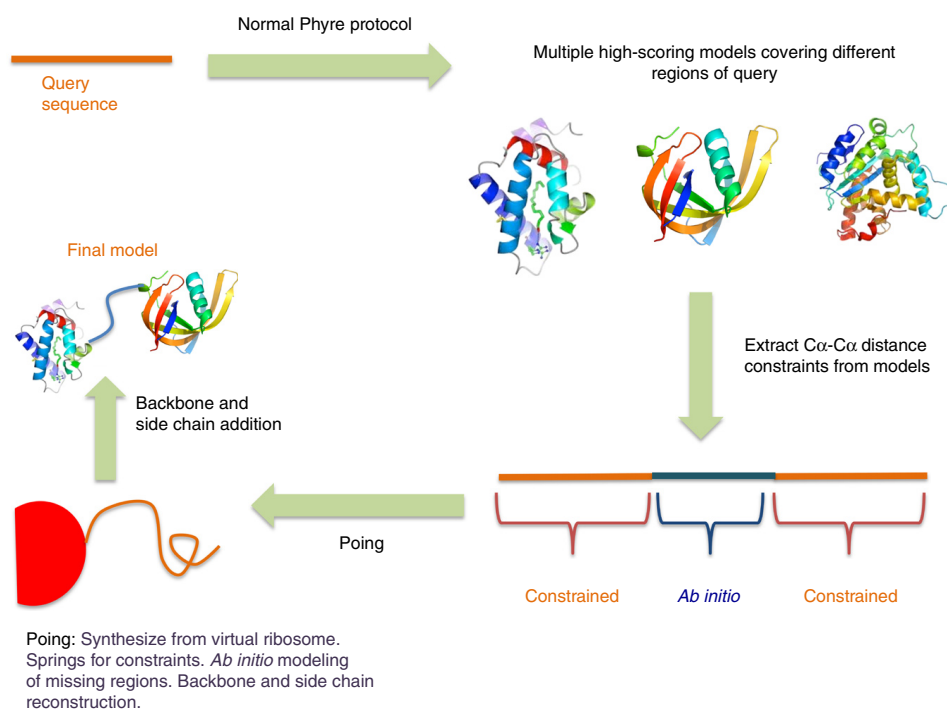


Figure 1.12 - Schematic representation of the intensive mode protocol used by the Phyre2⁴⁴ server, which differs from the normal mode by allowing the modeling of multiple regions from different templates and by including *ab initio* methods to build regions without available templates.⁴⁴

1.2.2. Virtual Screening

Virtual screening relies on docking methods to predict compound binding to a pharmaceutical target of interest. This approach attempts to yield hit molecules by analysing their posing and scoring results regarding its target^{55,56}. It can be an insightful and efficient technique for probing large compound libraries, before progressing to the experimental stages. There are two main variants of virtual screening consonant with previously obtained knowledge: ligand-based and structure-based⁵⁵. Docking and scoring is a fundamental part of either approach in the hit discovery process.

1.2.2.1. Ligand-based Virtual Screening

Ligand-based virtual screening is based on the similarity principle, which implies that similar compounds have identical biological effects. This means that libraries can be probed for novel promising compounds with similar structure in comparison to one or a few previously determined hits. The measure of similarity becomes the main difference between several ligand-

based virtual screening methodologies. This evaluation can be established by two-dimensional descriptors, in particular fingerprints and shape comparisons, or three-dimensional descriptors, such as pharmacophore-based.⁵⁵

Two-dimensional descriptors are commonly used in ligand-based virtual screening due to their simplicity and efficiency. These allow searching for substructures in large compound libraries, finding ligands that contain a privileged motif for a specific target. This methodology can also be important to exclude compounds with undesired characteristics present in the probed library. Molecular fingerprints are widely used to determine structure similarity between two compounds. They are stored as bit strings, wherein the absence ("0") or presence ("1") of a list of substructures is recorded for each molecule (Figure 1.13). Thus, the similarity comparison is computed regarding the individual bits of their bit strings and calculating similarity indices, such as the Tanimoto coefficient (Equation 4).⁵⁵

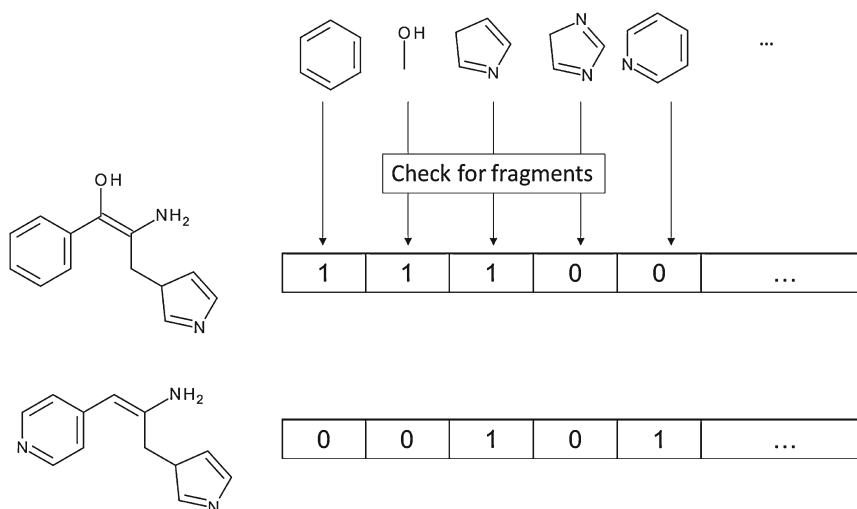


Figure 1.13 - Example of bit strings generated for two molecules using a small set of substructures.⁵⁵

$$T_{AB} = \frac{n_{AB}}{n_A + n_B - n_{AB}}$$

Equation 4 - Tanimoto coefficient, where n is the number of bits set to "1" for molecules A and B.

The strength of a protein-ligand interaction is intimately related to the shape of the binding pocket and the capability of the compound to fill that space. This steric complementary can be an important factor for ligand-based search, since two molecules with identical shape might possess similar biological activity⁵⁵. Using this premise, there is software able to retrieve similarly shaped compounds from libraries, as rapid overlay of chemical structures (ROCS)⁵⁷.

Pharmacophore-based screening relies on previously obtained structural evidences of an active compound with specific three-dimensional characteristics. These descriptors are then used to probe libraries for promising ligands that are capable of establishing similar interactions with the target's binding pocket.⁵⁵

1.2.2.2. Structure-based Virtual Screening

Structure-based virtual screening comprises the probing of entire compound libraries using docking methods. In this case, the target's structure is pivotal and may be determined as mentioned before by X-ray crystallography, NMR or homology and *ab initio* modeling^{7,10,21}. Intrinsically, docking establishes the target as a rigid entity, while promoting ligand flexibility. Several categories of search algorithms have been developed for docking applications, which can be classified as systematic, stochastic and deterministic simulation-based search approaches.^{55,56,58}

Systematic search algorithms consider all possible motions of the compounds, which generally becomes an impractically large number of ligand conformations to be computed. Besides translational and rotational motion, rotations around the torsional bonds of the molecule are typically allowed. As alternatives to this approach, fragment-based incremental construction and place-and-join algorithms can be applied. In incremental construction algorithms, the molecule's core is considered relatively rigid and apart from the "side-chains". Then, the core is docked in the target's binding pocket and the side-chains are subsequently "grown" from the core to reconstruct the original ligand. In the place-and-join approach, the molecule is split into several fragments which are individually docked. Afterwards, the fragments are linked and only the ligand poses with linkages that result in energetically low conformations are retrieved.^{55,56}

In stochastic search methods, such as Monte Carlo and genetic algorithms, new poses are generated from previous protein-ligand complexes by random changes to the dihedrals and translational and rotational degrees of motions^{55,56,58}. These solutions are ranked according to their energy function prediction and the energetically low poses are preferred. AutoDock⁵⁹ and GOLD⁶⁰ are examples of docking programs that use stochastic algorithms. In the specific case of AutoDock⁵⁹ the stochastic method used is a genetic algorithm.

Deterministic simulation-based methods, such as molecular dynamics simulations, are inefficient in generating diverse ligand poses and thus are not commonly used as search method in virtual screening. However, deterministic methods, in particular energy minimization, are frequently used to refine and evaluate protein-ligand complexes obtained from other search methods.^{55,56}

Scoring functions are fundamental in providing protein-ligand interaction's energies of all sampled poses and ligands from the different search techniques^{55,56,58}. In the case of AutoDock⁵⁹, the scoring function used is a force-field, which relies on molecular mechanics' energy terms to characterize the interactions between protein and ligand, besides accounting for the ligands' internal energy. Typically, they contain terms describing van der Waals and electrostatic interactions, but include also empirical parameters such as solvation and entropy^{55,56}.

1.2.3. Target-ligand complex characterization

After filtering the compound libraries for promising ligands of the desired pharmaceutical target, protein-ligand complex characterization must be performed to validate the hits and to gain quantitative insight regarding ligand's affinity and mode of binding. To achieve this goal, biophysical methods are commonly implemented, such as those mentioned in section 1.1.2.5. of the present work.

In the upcoming sections, the techniques employed in this thesis' experimental work will be briefly discussed. These methods, of biophysical and biochemical nature, can be divided by the effects caused on the protein upon ligand binding, regarding stability or conformation.

1.2.3.1. Thermal Shift Assay (TSA)

Thermal shift assay is a technique capable of providing relevant biophysical information regarding protein stability in a wide set of conditions. It is based on the principle that a compound present in solution, either salt or small molecule, can establish non-covalent interactions with several protein residues, thus increasing its structural stability and resistance to denaturation. In this case, protein denaturation is promoted through the increase in temperature, which eventually will cause internal hydrophobic residues to be exposed and activate the fluorescence of a dye, present in the mixture. The fluorescence intensity increases with protein denaturation and this process is followed as a dependency of temperature (Figure 1.14). This process is quantifiable through the melting temperature value (T_m) of the protein, which corresponds to the temperature at which half the protein's molecules are denatured.⁶¹⁻⁶³

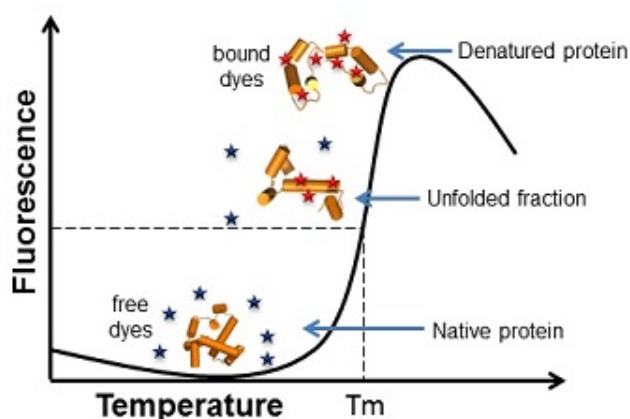


Figure 1.14 - Representation of a protein unfolding curve from a TSA (source: <https://www.jenabioscience.com/crystallography-cryo-em/screening/thermofluor-screens>).

TSA is commonly used for screening conditions in which a protein is more stable, with the purpose of tuning buffer either for purification or crystallization experiments^{62,63}. There are commercially available screens, for example from Jena Bioscience, that contain a significant variety of buffers, as well as additives, that can be tested. This technique can also be insightful for protein-ligand binding assays. Although its sensitivity may not be the best, compared with ITC or STD-NMR, it possesses great advantages, such as the requirement of less protein, the faster experiment times, including preparation and execution, and the possibility of testing many different conditions in a single assay.⁶¹

As any fluorescence based technique, TSA is very sensitive to contaminants, which may cause quenching of the dye used or enhance/decrease protein susceptibility to denaturation and

so, lead to false negative or positive results. Organic solvents required for small molecule solubilisation may entail a great challenge in result validation, because they may establish undesired interactions with the protein and/or the dye. Generally, the acceptable value for protein's destabilization promoted by organic solvents is 1° C, which corresponds to half of the 2° C cut-off established to report a positive hit⁶⁴.

Dissociation constant (K_D) values for protein-ligand interaction can be calculated by plotting the ΔT_m values for several ligand concentrations ($\Delta T_m = T_m(\text{protein} + \text{ligand/organic solvent}) - T_m(\text{control})$). If binding is observable, a hyperbolic curve is observed, which is associated with a Michaelis-Menten equation (referred to in GraphPad software as saturation binding, one-site specific binding) (Equation 5).⁶⁵

$$\Delta T_m = \frac{\Delta T_m(\text{max.}) \times [\text{ligand}]}{K_D + [\text{ligand}]}$$

Equation 5 – Equation derived from Michaelis-Menten kinetics for K_D determination through TSA, where $\Delta T_m(\text{max.})$ is the maximum ΔT_m value extrapolated at higher concentrations of ligand and K_D is the equilibrium dissociation constant, which corresponds to the concentration of ligand needed to achieve half-maximum binding at equilibrium.

1.2.3.2. Urea-Polyacrylamide Gel Electrophoresis

This method developed by Makey and Seal in 1976 consists of a polyacrylamide-gel electrophoresis in the presence of 6 M urea⁶⁶. In this way, proteins will have different mobility profiles in the gel according to their conformation, either in a compact or relaxed form. This allows the comparison between native and ligand-bound targets, in the cases where protein-ligand complexes assume a more closed conformation.

An example of the insights provided by this technique is the work reported by Makey and Seal, related to iron binding of human serum transferrin protein. Here, the existence of two different monoferric forms of transferrin were proven, by the different mobility shifts observed. As references, native transferrin and its diferric form were also distinguished by having opposite mobility shifts.⁶⁶ Recently the same approach was used to characterize binding of vanadium-based complexes to human serum transferrin and of Mo/W oxyanions to periplasmic components of ABC transporters^{67–69}.

1.2.3.3. Isothermal Titration Calorimetry (ITC)

Chemical reactions or physical changes are usually accompanied by alterations in heat or enthalpy. The amount of heat exchanged during these processes is directly proportional to the rate of the reaction and its enthalpy²⁹. Thus, protein-ligand interactions are amenable to be probed by this method, since they imply displacement of residues and water molecules for the accommodation of the compound in the binding pocket, alongside with possible conformation changes in the protein's tertiary structure upon binding.

As the technique's denomination suggests, it involves a titration experiment, where usually the ligand solution is added to the protein in an isothermal cell, which means that any heat exchanges will be compensated in order to maintain the cell's temperature (Figure 1.15). Therefore, heat alterations upon ligand binding are measured by the power applied to compensate them as a function of time, commonly in $\mu\text{cal}/\text{sec}$ or $\mu\text{J}/\text{sec}$. The effect of solution addition during the titration experiment is a concern in ITC, since small heat exchanges might happen due to system equilibration. To prevent or at least diminish this dilution effect, sample dialysis is recommended for the protein solution and the dialysis buffer is then used to solubilize the ligand. Another approach to solve this problem is to perform a blank experiment, where the conditions are replicated, but in the absence of ligand. If the heat signals remain, both solutions are not properly matched.²⁹ Another hurdle concerning the ITC experiment is the protein requirements⁶¹, which implicate highly pure protein in considerable volume (around 1 mL for each experiment) and concentration (commonly, 50 μM).

Determination of the protein-ligand complex constants, dissociation constant (K_D) and enthalpy (ΔH), is possible through several dedicated programs, such as TA Instruments NanoanalyzeTM and Origin-7TM (OriginLab, Northampton, MA). The most common model to fit the heat data as a function of the injection's volume is the one-site binding model and the fitting is achieved by a non-linear regression. The input variables and the output values are: number of binding ligands (N), dissociation constant (K_D), enthalpy (ΔH) and entropy (ΔS).²⁹ The ITC technique is best suited for medium-to-high affinity systems, since its detection limits of the complex's dissociation constant range from 100 μM to 10 nM⁷⁰.

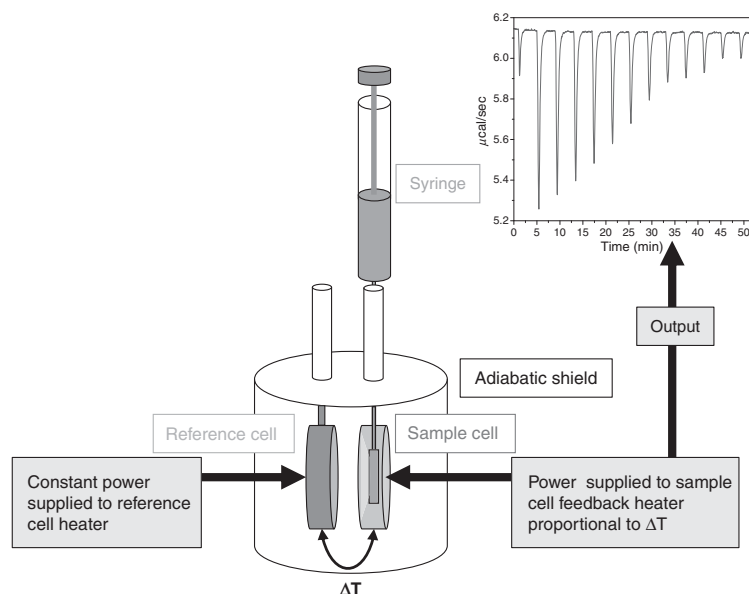


Figure 1.15 - Schematic representation of an ITC experiment in a power compensation instrument. The ligand solution in the syringe is injected consecutively in the sample cell, filled with the protein solution. The power needed to compensate the heat variations in the cell are plotted as a function of time.²⁹

1.2.3.4. Circular Dichroism (CD)

Plane polarized light can be viewed as two circularly polarized components of equal magnitude, one rotating counter-clockwise (left-handed, L) and the other rotating clockwise (right-handed, R). Circular dichroism is a technique capable of detecting differential absorption of these two components. This difference in absorbance of each component, either L or R, dictates a change in the circular polarized light to an elliptical shape. A CD signal is observed when a chromophore is chiral (optically active) by virtue of one of the following properties: it is intrinsically chiral, having an asymmetrical structure; it is covalently linked to a chiral center in the molecule; it is placed in an asymmetric environment, such as a three-dimensional structure adopted by the molecule.⁷¹

A protein three-dimensional structure provides an asymmetric environment that displays a CD signal in the far UV spectrum. It is well established that protein secondary structures, like α -helices and β -sheets, are observable through CD experiments at specific wavelengths (Figure 1.16)⁷¹. In the case of α -helices, they can be studied at 193, 208 and 222

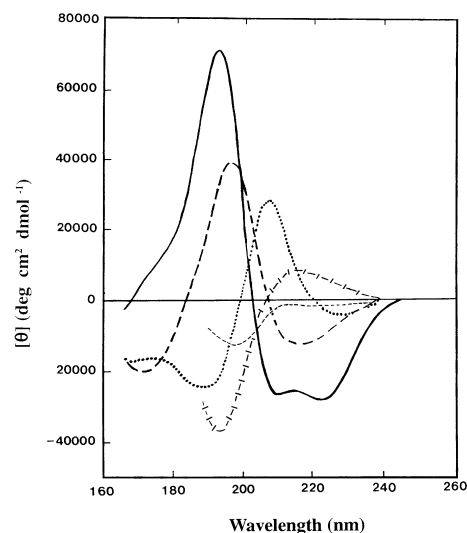


Figure 1.16 - Far U.V. CD spectra characteristic for various secondary structures. Solid line: α -helices; long dashed line: anti-parallel β -sheets; dotted line: type I β -turns; cross dashed line: extended 3_1 -helices or poly (Pro) II helices; short dashed line: irregular structures.⁷¹

nm and their content can also be estimated through bioinformatic algorithms such as the ones used by DICHROWEB^{72,73}. To enable this, experimental details are required to be known with appreciable precision, like protein concentration, molecular weight and number of residues.

CD might be an insightful technique regarding the characterization of protein-ligand interactions. If a ligand binding phenomenon is associated with protein structural alterations, those changes can be followed at the far UV spectrum. The range of ligand concentrations and the speed at which structural changes are induced can be examined through time-resolved CD studies.⁷¹

1.2.4. Target-ligand structural determination

After validating the hits retrieved from the virtual screenings through several biophysical and biochemical techniques, an in-depth understanding of protein-ligand interactions is sought. At this point, energetics and kinetics information have been collected regarding the ligand's affinity and specificity toward its target. The following stage is concerned with characterizing the underlying spatial arrangement of the protein-ligand complex. To achieve this purpose, there are two main techniques available: X-ray crystallography and NMR. An alternative approach, based solely on *in silico* methods is molecular dynamics.²⁴

1.2.4.1. X-ray crystallography of protein-ligand complexes

As described before, X-ray crystallography has the capability to study molecules at the atomic level. Thus, this technique poses as a great opportunity to structurally characterize protein interactions with ligands, such as substrates, activators or inhibitors.

Protein crystals have a high solvent content (between 30 and 80%⁷⁴), which is present in interstitial channels, contributing to their fragile nature^{24,75}. However, this characteristic provides reasonable flexibility to the crystal's protein molecules and often allow catalytic activity²⁴. Hence, protein crystals not only possess the means to incorporate small molecules, but, indeed, offer the chance to investigate, at high resolution, protein interactions toward those ligands. In crystallography, there are two main techniques to incorporate the ligand of interest in protein crystals: co-crystallization and soaking^{24,75,76}.

Co-crystallization involves incubation of the ligand with the protein prior to the crystallization experiment. Co-crystals can be achieved by adding ligand solution to the protein and precipitant's drop, or by incubating the protein and ligand for a certain period of time. This process diminishes the concerns regarding ligand solubility and prevents crystal degradation from soaking experiments⁷⁵. In fact, some proteins may present a tendency to form aggregates and only by incubation with a ligand that enhances their stability, they become amenable to crystallization⁷⁶. The main disadvantage of this approach is that a protein-ligand complex may possess different structural arrangements that avoid crystallization at the established apo form's crystallization conditions⁷⁵. Hence, the complex might require new screenings of crystallization conditions and posterior optimization to yield high-quality crystals.

In the soaking approach, the compound is incubated with preformed protein crystals. Often, these crystals have already been characterized and the structure of the protein apo form has already been determined. This technique can also be used to replace a co-crystallized ligand by a second molecule of interest in the protein structure. Soaking ensures some advantages in comparison with co-crystallization, such as convenience and reproducibility, since it allows soaking of a large number of high-quality crystals in drops with many different ligands. However, it may entail some challenges and limitations, since upon ligand binding, a protein cannot adopt very distinct conformations, and the soaking procedures can cause damage or even disruption of the crystal lattice-packing interactions. This can result in crystal cracking leading to poor diffraction or, in extreme cases, crystal dissolution.⁷⁵

1.2.4.2. Saturation Transfer Difference – NMR (STD-NMR)

Nuclear magnetic resonance is a technique able to investigate the structural details of macromolecules and their interaction with ligands. It is based on the nuclear magnetic moment of some elements such as ^1H and, in the case of biological macromolecules, isotopes such as ^{13}C and ^{15}N . When placed into a static magnetic field (B), the distinct nuclear spin states become quantized with energies proportional to their projection on B (Zeeman splitting). This energy difference depends on several variables: type of nucleus, field strength and chemical environment. The transition between these states can be induced by irradiation with the specific radio-frequency for each type of nucleus and its chemical environment. The frequency of the NMR signal is very sensitive to covalent bonds, for example between neighbouring groups, but also, and most importantly for protein-ligand interaction, to non-covalent bonding. Regarding the study of protein-ligand complexes, there are two approaches, using NMR techniques: STD-NMR and water-ligand optimized gradient spectroscopy (WaterLOGSY). In the present work, only STD-NMR was employed to assess possible protein-ligand interactions.²⁴

STD-NMR provides the structural characterization of protein-ligand complexes from the ligand perspective. It is based on the nuclear Overhauser effect and in the observation of the ligand resonance signals. This experiment allows not only to screen compounds for lead structures, but also to identify important ligand moieties for binding. It is possible to characterize a ligand binding by its binding epitope, which corresponds to the compound's hydrogens involved in protein-ligand interactions. The STD-NMR experiment comprises the saturation of the protein's hydrogens and the transfer of this saturation to the bound ligand hydrogens (Figure 1.17). In this way, a difference between the intensities of the protein's spectrum signals which were selectively saturated (on-resonance), I_{SAT} , and the intensities of the protein's spectrum signals without saturation (off-resonance), I_0 , can be calculated. The difference spectrum ($I_{\text{STD}} = I_{\text{SAT}} - I_0$) shows the signals of the ligand that received the saturation from the protein. Ligand binding can be quantified by the intensity of the signals present in the difference spectrum, which is related to the proximity of the ligand and protein hydrogens. This effect allows ligand mapping, yielding the moieties and atoms involved in these interactions and quantifying their participation.⁷⁷

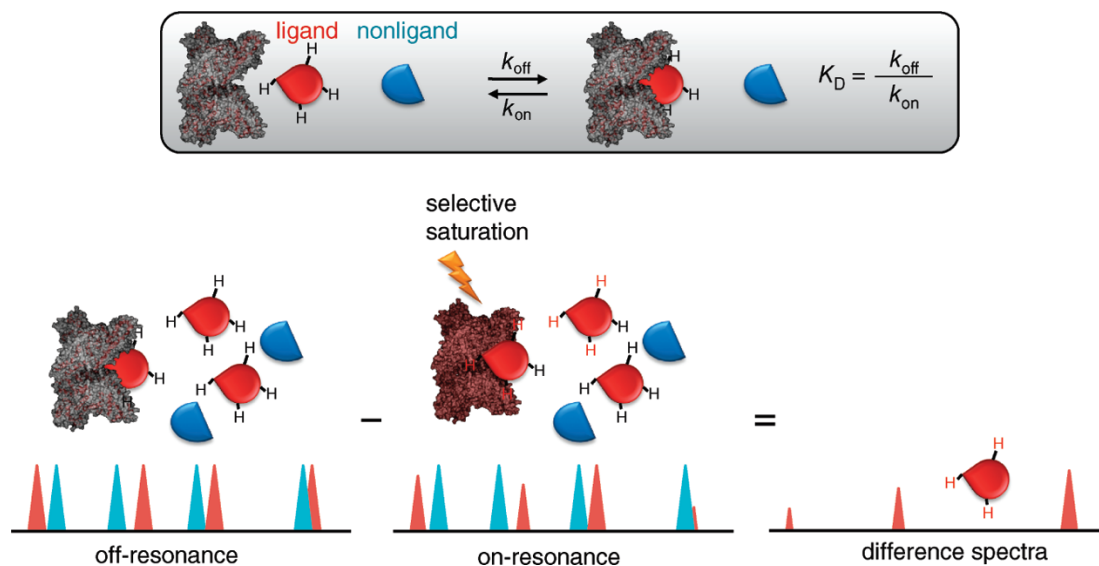


Figure 1.17 - Schematic representation of the STD-NMR technique, denoting the bounded and non-bounded states and their equilibrium. Also, the subtracting nature of the method is represented, where the on- and off-resonance spectra are subtracted to yield the STD spectrum.⁷⁷

Despite being a structural technique, STD-NMR can also provide kinetic parameters of protein-ligand binding, namely its dissociation constant (K_D). One approach is to obtain several STD-NMR spectra with increasing concentrations of ligand. This will generate a Michaelis-Menten's hyperbolic curve, when the STD amplification factor (A_{STD}), which is related to the signal's intensity, is plotted against the ligand concentration. The experimental data can be adjusted using the following equation, by changing the variables K_D and α_{STD} (Equation 6):⁷⁷

$$A_{STD} = \frac{\alpha_{STD} [L]}{K_D + [L]}$$

Equation 6 - Derived equation from the Michaelis-Menten kinetics that provides the estimation of the K_D , where α_{STD} is the maximum STD amplification factor, extrapolated at higher concentrations, and $[L]$ is the ligand concentration.

An important limitation of the STD-NMR experiment is its sensitivity, which ranges from $0.1 \text{ nM} < K_D < 100 \text{ mM}$. This means that above and below these limits the saturation transfer is not very efficient, yielding weak or no STD-NMR signals. In the lower-limit, the ligand's binding is too tight and consequentially the off-rates are very low, while in the upper-limit, the interaction is too weak and so the probability of the ligand being in the receptor site is very low.²⁸

1.2.4.3. Molecular Dynamics (MD)

X-ray crystallography and NMR provide valuable insight regarding macromolecular structures through static models. However, molecular recognition and drug binding are very dynamic processes, since a small molecule in solution encounters its target in constant motion and structural rearrangement⁷⁸. Molecular dynamics is an *in silico* method that aims to compute and estimate macromolecular motion in solution, and is able to provide relevant information about protein-ligand interactions.

The simulated motion of the system's atoms is based on Newtonian physics, which allows the reduction of the computational complexity. These motions are governed by the interaction forces from every atom which comprises the parameters of bonded and non-bonded interactions (Figure 1.18). Bonded interactions encompass chemical bond lengths, atomic angles and dihedral angles. While the non-bonded forces represent van de Waals and electrostatic interactions. The energy equation presented requires the input of several energy terms that fit quantum mechanical calculations and experimental data, provided by the so called "force fields". Once the forces acting on each of the system's atoms is calculated, the atom's positions are moved according to Newton's laws of motion for a period of few picoseconds. Then, the energies are computed again and this cycle is performed until the simulation time is finished.⁷⁸

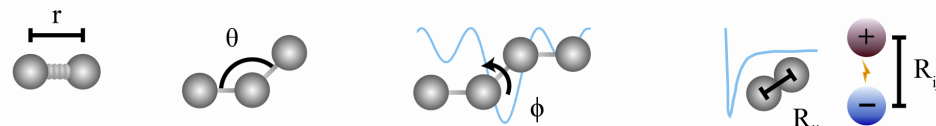
$$E_{total} = \underbrace{\sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]}_{\text{Bonded}} + \underbrace{\sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]}_{\text{Non-bonded}}$$


Figure 1.18 - Energy equation applied to calculate the atom's motion in a molecular dynamics simulation. The atomic forces are divided into bonded and non-bonded interactions. The first energy terms refer to bond lengths, atomic angles and dihedral angles, modeled by virtual springs and a sinusoidal function, in the case of the latter. The non-bonded energy terms correspond to van der Waals and electrostatic interactions.⁷⁸

One approach that takes advantage of MD simulations is energy minimization. Throughout a MD simulation, a protein experiences numerous conformations and rearrangements. Thus, after the simulation it is possible to extract the global minimum energy structure which can be employed in docking studies or enable different structural stability analysis^{21,79}. One example of the latter is the study of the impact of non-synonymous single nucleotide polymorphisms (nsSNP) which was reported for the TP53 gene by Chitralla Yeguvapalli⁷⁹. This type of work involves extensive analysis of MD simulations regarding parameters such as rmsd (root-mean-square deviation), potential energy, intramolecular hydrogen bonds, motion projection and solvent-accessible surface area.

Regarding the study of protein-ligand interactions, there are two main contributions offered by MD simulations: identification of cryptic and allosteric binding sites; prediction of compound binding affinity and relevant interacting protein residues⁷⁸. Taking advantage of the target and ligand flexibilities, non-traditional binding pockets of a protein may be found, revealing new potentially “druggable” sites. This surpasses the problem of the static models provided by X-ray crystallography and NMR, revealing rearranged structures for putative ligand accommodation. The binding affinity of a compound toward its target can be predicted through several techniques involving MD simulations. One example is the “alchemical transformation”, which aims to estimate the free energy difference that result from the protein’s conformational variations required for ligand accommodation and binding⁷⁸. To achieve this, it’s necessary to perform two different simulations, one where the solvated ligand is free and one where the ligand is bounded to the protein. These simulations entail the ligand’s subtraction after the protein’s conformation is stabilized. Therefore, the free energy of binding is the difference between the free energies of the non-bounded and bounded states of the protein.

1.3. State-of-the-Art

In the following sections, the state-of-the-art regarding the subjects of the present thesis, human Bcl-2 and *Streptococcus dysgalactiae* LytR proteins, is discussed. The most relevant topics to this work concern the rational design of small molecule inhibitors of these proteins and therefore the studied promising compounds are described.

1.3.1. Human Bcl-2 protein

Cancer continues to be one of the leading causes of death in humans, owing to the plethora of events responsible for its activation, the lateness of diagnostics and the ineffective treatments available. In 2000, the hallmarks of cancer were postulated: sustaining proliferative signalling; evading growth suppressors; resisting cell death; enabling replicative immortality; inducing angiogenesis; activating invasion and metastasis (Figure 1.19)⁸⁰. More recently, due to the growing understanding of cancer mechanisms and physiology, two emerging hallmarks were considered as potential additions to this list. The biology of tumors is no longer understood by the cancer cells traits but must encompass the contributions of the “tumor microenvironment” to tumorigenesis. Therefore, the two potential hallmarks to be included are: reprogramming of energy metabolism and evading immune destruction.⁸¹

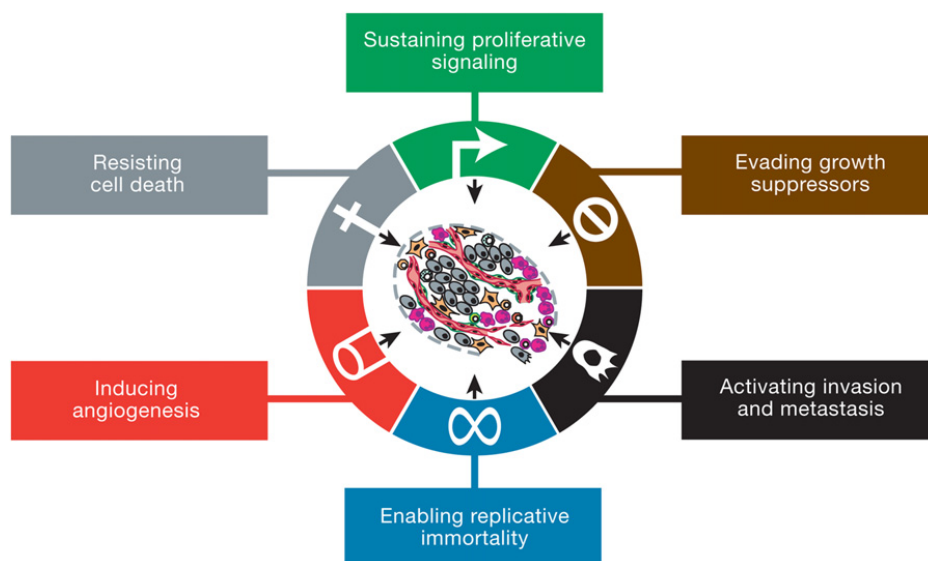


Figure 1.19 - Schematic diagram of the six hallmarks of cancer proposed in 2000, by Hanahan and Weinberg.⁸⁰

Tissue homeostasis comprises a strict balance between cell proliferation and cell death and changes in the rate of the latter may implicate tumor formation⁸². As established previously, evasion or resistance to cell death is a hallmark of cancer and this happens due to genetic mutations that alter either the expression or function of proteins⁸³. Cell death, referred to as apoptosis or programmed cell death (PCD), is a genetically defined mechanism that allows abnormal cells to commit suicide. It is a fundamental feature for multicellular organism survival,

since it promotes the eradication of damaged cells which may interfere with the organism's normal functions and even promote tumor formation. The apoptotic processes can be divided into two well-studied pathways: intrinsic and extrinsic.⁸⁴

One of the most well-known forms of the apoptotic intrinsic pathway takes place in the mitochondria and is governed by the stress-mediated release of cytochrome c to the cytosol that culminates in the formation of the apoptosome. The apoptosome activates initiator caspase 9 which leads to the activation of executioner caspase 3. The active caspase 3 is responsible for the cleavage of death substrates that lead to well-known hallmarks of an apoptotic cell including DNA fragmentation, nuclear fragmentation and membrane blebbing.⁸⁴

B-cell lymphoma 2 (Bcl-2) family proteins are responsible for the homeostasis that dictates cell survival or cell death, through the intrinsic pathway. Its founding member is the Bcl-2 protein, which allowed the association between t(14;18) chromosomal translocations and human follicular lymphoma by Fukuhara and Rowley⁸⁵. In t(14;18) translocations, the BCL-2 gene from chromosome 18 becomes fused with the immunoglobulin heavy-chain *locus* on chromosome 14, becoming regulated by the immunoglobulin enhancer, thus dysregulating Bcl-2 expression at the transcriptional level⁸⁶. The remaining Bcl-2 family members were then included by sequence homology identification, which enabled the postulation of four different domains, named Bcl-2 homology domains (BH1, BH2, BH3, BH4)^{87,88}. The Bcl-2 family proteins can be divided into two groups, according to their function: anti-apoptotic and pro-apoptotic (Figure 1.20)⁸⁷. The anti-apoptotic members are Bcl-2, Bcl-XL and Mcl-1, which possess multi-region structures, including the four BH domains and a membrane-binding domain (MBD). While the pro-apoptotic members comprise multi-region and BH3-only proteins. Bax, Bak and Bid are the multi-region proteins having the same homology domains as the anti-apoptotic members, except for Bid which besides the MBD, has only the BH3 and BH4 domains. The BH3-only proteins, Bim/Bad and Noxa also incorporate the MBD and are usually termed as activators or sensitizers of the apoptotic process.⁸⁷

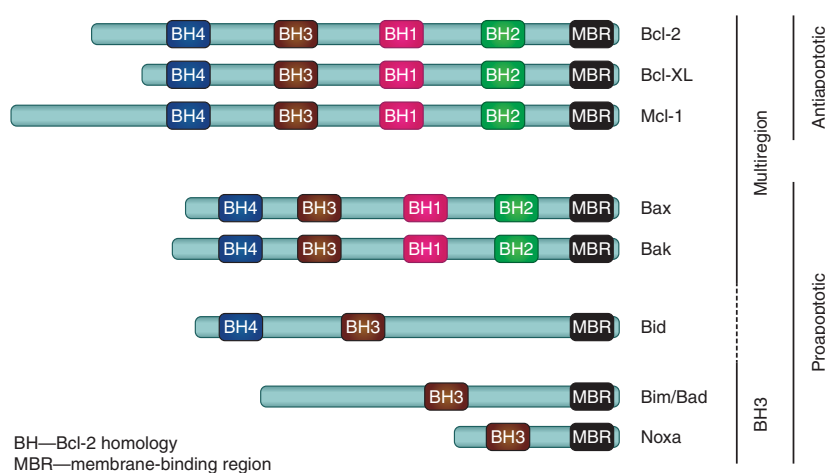


Figure 1.20 - Schematic representation of the sequence homology present in Bcl-2 family members, highlighting the BH domains and the MBR domain. The Bcl-2 family is divided by function: anti-apoptotic or pro-apoptotic, and the latter can be distinguished by structure: multi-region or BH3-only proteins.⁸⁷

The role of the Bcl-2 family proteins in the apoptotic intrinsic pathway may be described by the embedded together model (Figure 1.21), which takes place in the mitochondrial outer membrane (MOM)⁸⁷. As the cell senses a death signal, the activator Bcl-2 proteins, Bim and Noxa, migrate and insert into the MOM, where they recruit cytoplasmic pro-apoptotic protein, Bax (or Bak). The active Bax protein experiences conformational changes which allow its oligomerization, leading to MOM permeabilization. Meanwhile, the activator proteins also recruit anti-apoptotic proteins, Bcl-2 or Bcl-XL, to the MOM, where they become sequestered, thus unable to inhibit their pro-apoptotic partners.⁸⁷ This phenomenon allows cytochrome c displacement from the mitochondria to the cytosol which leads to the apoptotic events described before. One may consider this mechanism as a population-equilibrium, since the higher availability of pro-apoptotic members causes the progression of apoptosis, while an excess of anti-apoptotic members avoids the apoptotic pathway, through the inhibition of the pro-apoptotic proteins.

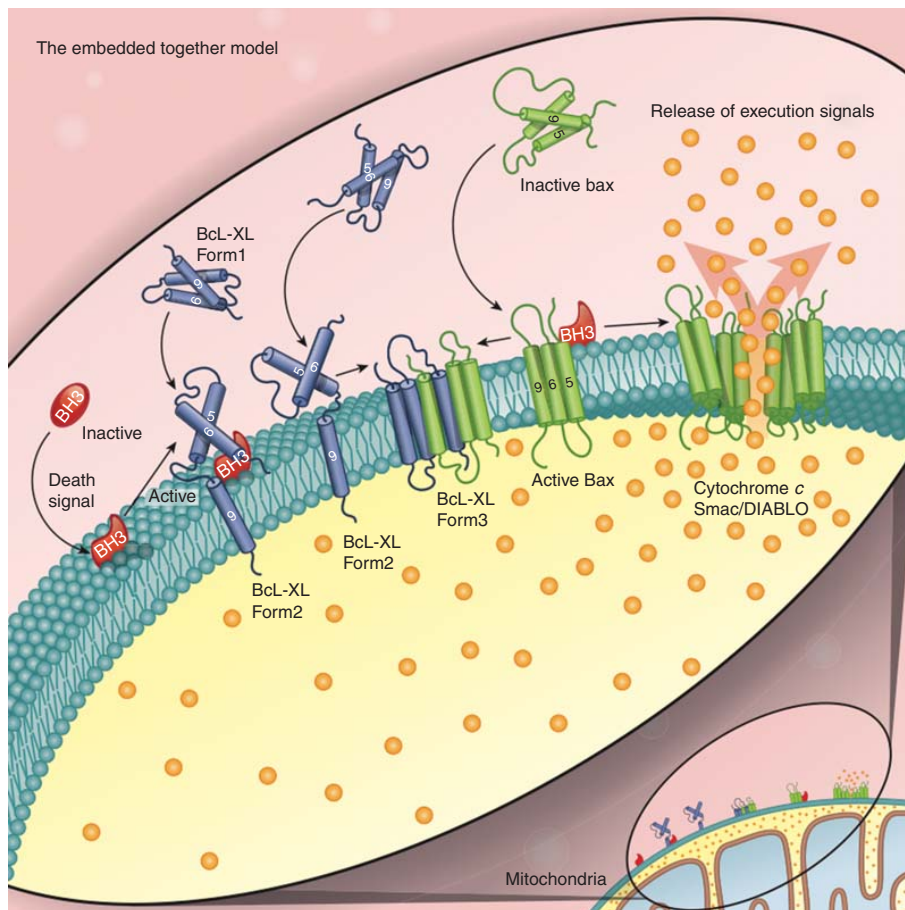


Figure 1.21 - Schematic representation of the embedded together model, which might explain the underlying mechanism of the Bcl-2-induced apoptotic pathway.⁸⁷

Bcl-2 protein is estimated to be overexpressed in almost half of all human cancers⁸⁶. Chromosomal translocation as a mechanism of BCL-2 gene activation is associated with non-Hodgkin's lymphomas, while loss of endogenous miRNA, that represses Bcl-2 expression, and gene hypomethylation were reported in chronic lymphocytic leukaemia^{89,90}. Moreover, Bcl-2

overexpression is correlated with, not only non-Hodgkin's lymphomas and chronic lymphocytic leukaemia, but also small cell lung and breast cancers^{91,92}.

The Bcl-2 role in the mentioned pathologies entails the dysregulation of the pro-apoptotic and anti-apoptotic equilibrium, present in healthy cells. In cancer, this equilibrium is shifted toward the anti-apoptotic members that, in excess, inhibit the pro-apoptotic proteins. When faced with a death signal, the tumor cell cannot activate its apoptotic mechanism, which results in cell survival and proliferation and extensive therapeutic problems, since these cells become resistant to chemotherapeutics⁹³⁻⁹⁵.

The human Bcl-2 protein has two reported isoforms, alpha and beta, with structures solved through NMR by Petros *et al*, in 2001⁹⁶. The most common alpha isoform has 239 amino acids and comprises the four BH regions and the MBD, in a helical bundle fold (Figure 1.22A)⁸⁸. Its function, as discussed previously, is to inhibit its physiological partner pro-apoptotic protein, Bax, through protein-protein interactions with Bax's BH3 domain (Figure 1.22B)⁹⁷. Therefore, while Bcl-2 poses as a potential pharmacological target for inhibition, BH3 mimetics are the main category of promising therapeutic agents explored so far^{88,98-100}.

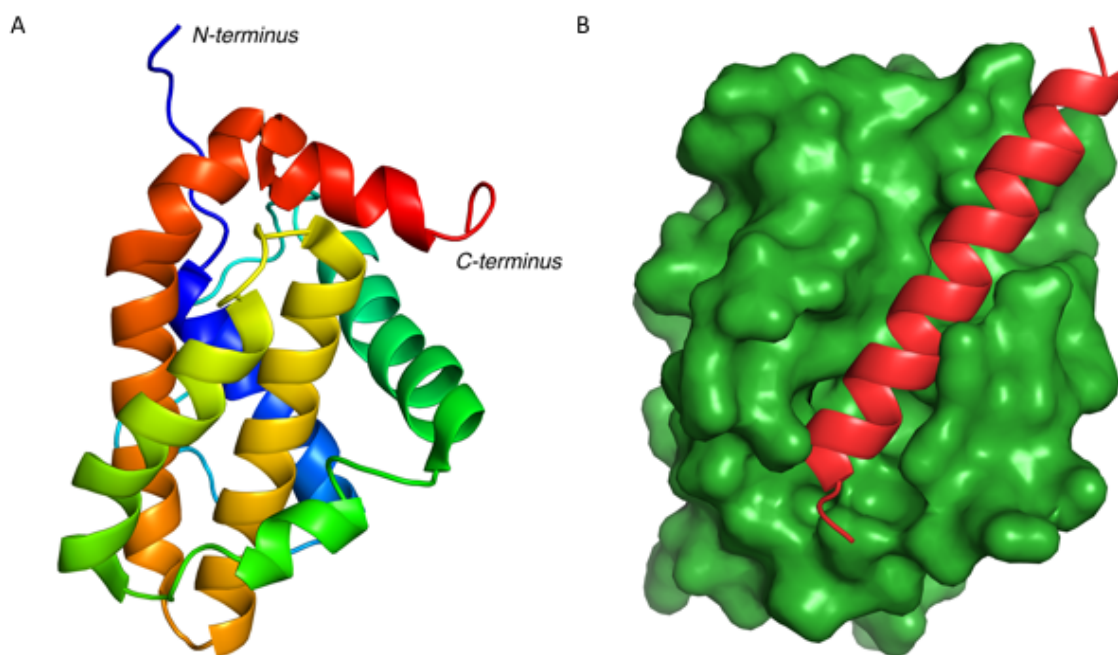


Figure 1.22 – A. NMR structure of human Bcl-2 protein alpha isoform (PDB code: 1G5M)⁹⁶. B. Crystal structure of Bcl-2 (surface in green) and Bax's BH3 domain (in red) complex at 2.7 Å (PDB code: 2XA0)⁹⁷.

The main representative of the BH3 mimetic drug molecules, which displayed significant therapeutic activity in clinical trials, is the arylsulfonamide-based compound, venetoclax, also known as ABT-199 (Figure 1.23)¹⁰¹⁻¹⁰⁴. It was discovered through the fragment-based NMR method by Abbott Laboratories, in a ten-year research effort. The hit-to-lead process started with ABT-737 (Figure 1.23), which exhibited nanomolar affinity to Bcl-2. However, its prospects of becoming a therapeutic agent were hampered by its poor physicochemical and pharmaceutical properties, namely its low oral bioavailability¹⁰³. ABT-263 (Figure 1.23), also known as navitoclax,

was derived from ABT-737 in order to overcome these hurdles, yielding a lead compound that retained its extremely high affinity and also displayed significant oral bioavailability¹⁰³. Despite this success, navitoclax revealed selectivity issues toward Bcl-2 when compared to its homologue Bcl-XL. This problem is particularly important, since Bcl-XL is the primary survival factor in platelets¹⁰⁴. A lead optimization phase allowed the discovery of venetoclax which is characterized by a sub-nanomolar affinity to Bcl-2 ($K_i < 0.01$ nM) and over three orders of magnitude inferior affinity to Bcl-XL ($K_i = 48$ nM)¹⁰⁴. This improvement was achieved through SBDD by exploiting the structural information provided by a co-crystal Bcl-2-navitoclax structure (PDB, Protein Data Bank¹⁰⁵ - www.rcsb.org, code: 4LVT)¹⁰⁴. The derivatization was performed by adding an azaindole moiety to the navitoclax structure, indicating this chemical family as promising Bcl-2 specific binders. In 2016, the FDA (United States Food and Drug Administration) approved venetoclax for the treatment of patients with chronic lymphocytic leukaemia. Meanwhile, venetoclax has been associated with undesired side-effects as neutropenia¹⁰⁶.

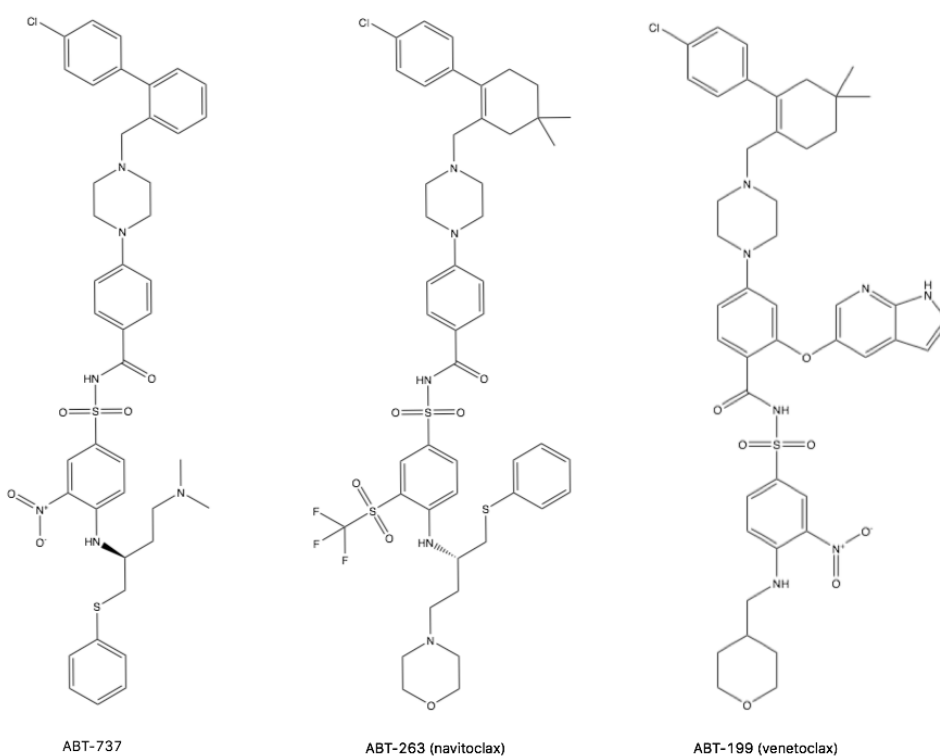


Figure 1.23 - Chemical structures of the three generations of arylsulfonamides' most representative examples. Fragment-based and SBDD approaches allowed the progression from ABT-737, through ABT-263, to ABT-199 as the most potent selective inhibitor of Bcl-2 protein.

In the present thesis, two alternative families of compounds were studied for Bcl-2 inhibition: 4H-chromene and indole-based compounds.

The 4H-chromene family emerged in a structure-based virtual screening approach, where the molecule HA14-1 was identified as a candidate binder of Bcl-2 (Figure 1.24)¹⁰⁷. Molecular docking and *in vitro* studies showed promising affinity toward Bcl-2 (IC₅₀ = 9 μM, this parameter corresponds to the compound concentration necessary for inhibition of 50% of the target molecules) and induced apoptosis of human promyelocytic leukemia (HL-60) cells¹⁰⁷. Subsequently, derivatization of this lead compound was performed to enhance its binding affinity to Bcl-2 and its anti-proliferative properties. Both optimizations were achieved in the case of its activemethine derivative, while a hypothetical improvement of the anti-proliferative activity of the methoxy and ethoxy derivatives was not confirmed (Figure 1.25)^{108,109}.

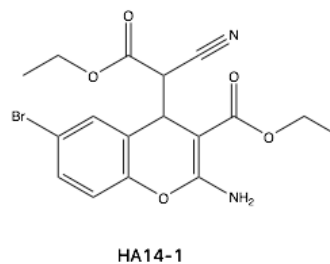


Figure 1.24 - Chemical structure of the reference 4H-chromene molecule, HA14-1.

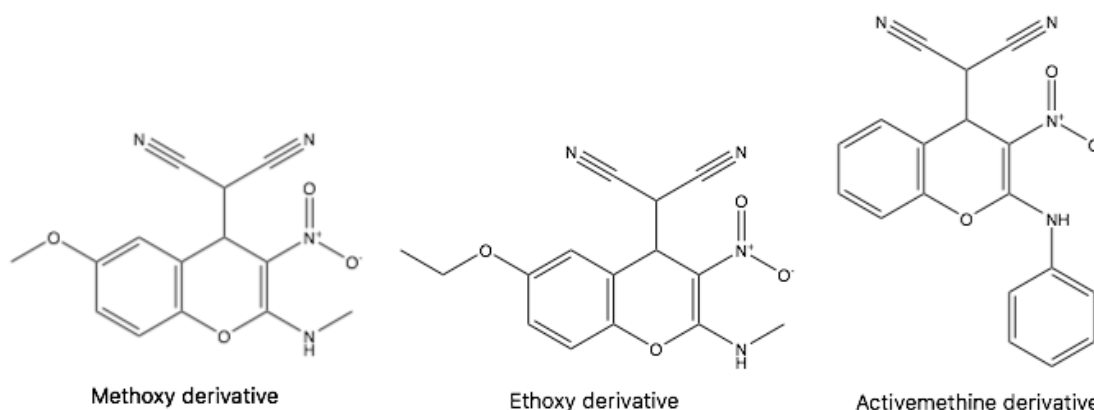
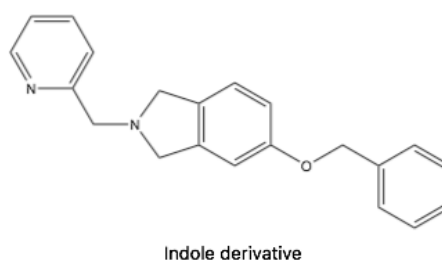


Figure 1.25 - Chemical structures of the three 4H-chromene derivatives derived from HA14-1, which are considered as candidate inhibitors of Bcl-2.

As mentioned in the lead optimization process that yielded venetoclax, azaindole moieties pose as promising binders of Bcl-2. This observation led to a collaboration with Dr. Maria M. B. Marques (LAQV, REQUIMTE, FCT-UNL, Portugal), who provided indole derivatives with promising scaffolds for derivatization, which may provide BH3 mimetics with improved activity, selectivity and fewer side-effects (Figure 1.26). Compounds with indole moieties have already been reported as Bcl-2 inhibitors through *in silico* and *in vitro* studies¹¹⁰.



In the present work, the 4H-chromene derivatives, ethoxy and activemethine, and an indole derivative are investigated regarding their

Figure 1.26 - Chemical structure of the indole derivative under study.

potential BH3 mimetic interaction with human Bcl-2 protein. The commercially available drug molecule, venetoclax, is used to validate the techniques employed for characterizing protein-ligand interactions. Structural determination of protein-ligand complexes is pursued through X-ray crystallography. Furthermore, *in silico* methods are applied in order to study the impacts of nsSNP in protein stability and structure, and the effects on ligand binding.

1.3.2. *Streptococcus dysgalactiae* LytR

Bacterial growth is characterized by two phenotypes, single cells (planktonic) and cell aggregates (biofilm)¹¹¹. Biofilms are commonly described as a cluster of cells enclosed in a self-produced matrix¹¹² (Figure 1.27). Regarding numerous bacterial infectious diseases, the biofilm phenotype is pivotal, since it confers to bacteria several properties that not only promote their proliferation, but also ensure protection against the host immune system and the treatment with antibiotics. Actually, in the biofilm phenotype, microorganisms may become 100-1000 times less susceptible to conventional antibiotics when compared to their planktonic phenotype¹¹³. There are three main characteristics of the biofilm phenotype: decreased antimicrobial susceptibility, phagocytic tolerance and quorum sensing¹¹¹. The decreased susceptibility to antibiotics has two aspects: tolerance and resistance¹¹¹. On one hand, tolerance represents the survival of the bacteria although their proliferation is limited. On the other hand, resistance means that bacteria are capable of proliferating despite the presence of antibiotics. Upon infection, the first responses to invading bacteria are from the host innate immune system, namely through neutrophils¹¹⁴. Regarding the planktonic phenotype, neutrophils phagocytosis is a common and easy response. However, this has no parallel concerning the biofilm phenotype, since the biofilm seems to confer bacteria protection toward phagocytosis, which is related to the chronic persistence of infections.¹¹⁵⁻¹¹⁷

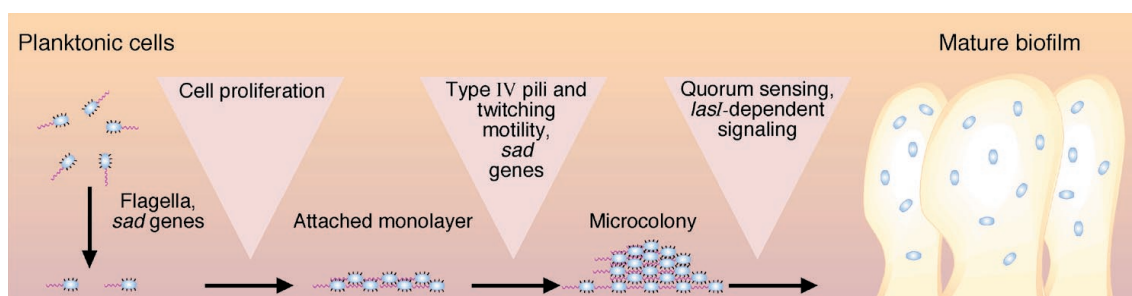


Figure 1.27 - Model of the development of mature biofilm from planktonic cells (of *P. aeruginosa*). This process starts with cell proliferation and attachment, followed by the expression of characteristic genes. The quorum sensing phenomenon is pivotal for the production of the extracellular matrix that encloses the cells.¹¹⁷

Streptococcus dysgalactiae has been associated with numerous diseases in several animal species: toxic shock-like syndrome in cattle¹¹⁸; bovine mastitis¹¹⁹; suppurative polyarthritis in lambs¹²⁰; bacteraemia in dogs¹²¹; severe septicaemia in fish¹²²; and cellulitis in humans¹²³. This microorganism is a Gram-positive bacterium, which is a group that encompasses many

pathogens and possesses a very characteristic cell wall. This cell wall has two major components, the peptidoglycan (PG) and the PG-attached anionic polymers (AP), which include wall teichoic acids (WTA)¹²⁴. WTA have important cellular roles, as reviewed by Weidenmaier and Peschel¹²⁵, such as: control of autolytic activity; antigenicity and innate immune recognition; biofilm formation; release of proteins into the culture medium; cation homeostasis; antibiotic resistance; and cell division. Thus, the understanding of the underlying mechanism of WTA's attachment to the PG is of great interest and can lead to new pharmaceutical therapies for antibiotic resistant pathologies.

The proteins involved in the attachment of WTA to the cell wall belong to the LytR-Cps2A-Psr (LCP) family. LCP proteins were reported to be associated with MreB cytoskeleton, which suggests that MreB proteins are involved in PG and AP insertion into the cell wall, while LCP proteins carry out the last stage of AP attachment to PG¹²⁴ (Figure 1.28). This mechanism starts with the synthesis of the AP, as WTA,

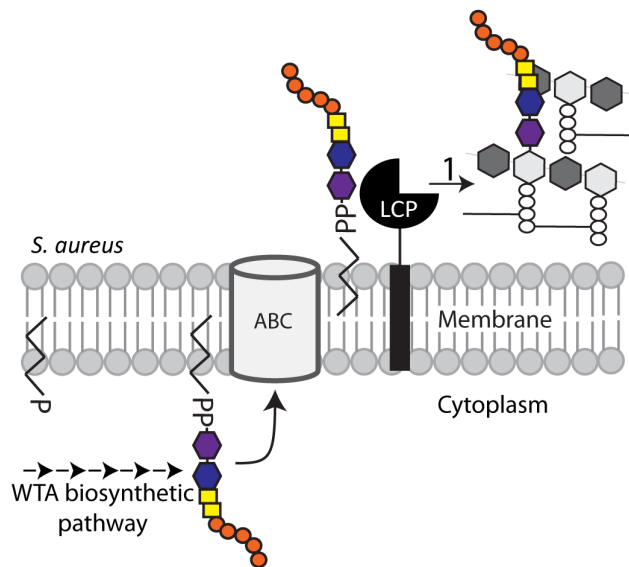


Figure 1.28 - Schematic representation of the attachment of wall teichoic acids to the peptidoglycan cell wall by LCP proteins. WTA molecules are colored while the PG are in grey scales.¹⁵⁸

inside the cell, which are attached to a carrier lipid, undecaprenyl-phosphate, and translocated (flipped) to the outside of the cell. The final stage, in which the LCP proteins are thought to be pivotal, is the covalent attachment to the PG. This phosphotransferase reaction links the phospho-teichoic acid chain to the C-6 hydroxyl of N-acetylmuramic acid within the PG.¹²⁴

LCP proteins are located in the cell membrane, specially enriched in the mid-cell region, confirming the predictions of their transmembrane domains, and have an extracellular domain, called LCP domain, which is believed to have a catalytic function¹²⁶. This membrane localization dismisses previous reports of LCP proteins as transcription regulators, mainly from phenotypical evidences^{127,128}. Meanwhile, the catalytic function is corroborated by structural and biochemical information. In 2011, the structure determination, by X-ray crystallography, of Cps2A from serotype 2 *S. pneumoniae* D39 revealed the presence of decaprenyl-phosphate in a hydrophobic pocket of the protein¹²⁴ (Figure 1.29). This showed that this protein has indeed a catalytic domain that accommodates lipidic substrates.

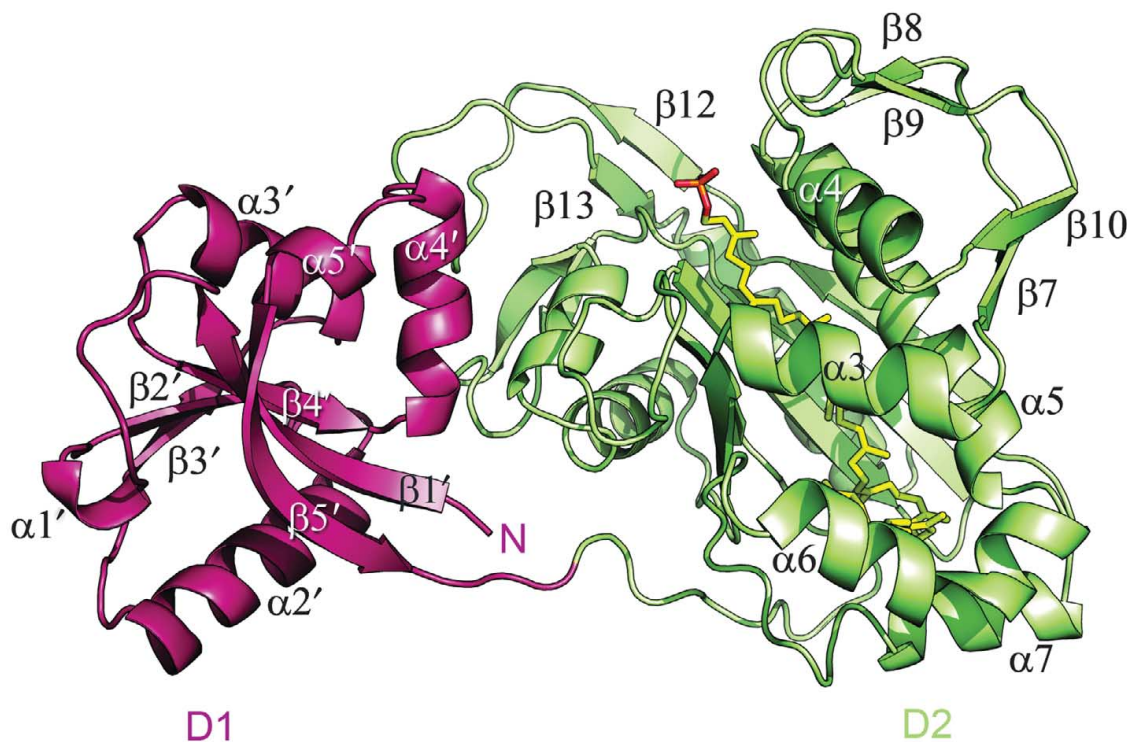


Figure 1.29 - Crystal structure of the extracellular domains of *Cps2A* from serotype 2 *S. pneumoniae* D39. An accessory domain is represented in pink (D1), while the LCP domain is in green (D2), with named secondary structures. The lipidic substrate, decaprenyl-phosphate, is represented in yellow and orange, in the hydrophobic pocket.¹²⁴

Furthermore, the hydrophobic character of the residues in the observed binding pocket and the charged residues that stabilize the phosphate head-group are completely conserved across the LCP family of proteins¹²⁴ (Figure 1.30). Schaefer *et al*¹²⁹, in 2017, proved by *in vitro* reconstitution the ligase activity of LCP proteins, by observing the attachment of small WTA precursors to PG.

Streptococcus dysgalactiae LytR is one member of the LCP family of proteins that has 422 residues, which encompass one transmembrane region and an extracellular catalytic domain (Figure 1.31). LytR is predicted to have an α/β mixed structure, similar to all proteins from the LCP family. Its function, as mentioned previously, is the attachment of WTA to PG in Gram-positive bacteria's cell wall, which confers them resistance to the host's innate immune system and to conventional antibiotics. Therefore, LytR poses as a relevant therapeutic target for inhibition, which would allow the treatment of numerous intricate infectious pathologies.

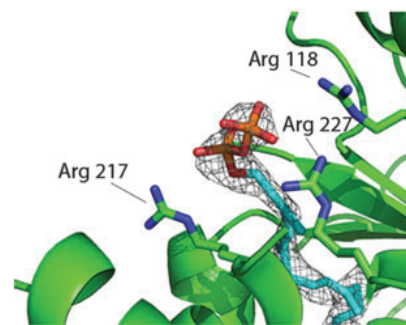


Figure 1.30 - Crystal structure at 2.49 Å of *YwtF* from *Bacillus subtilis* (PDB code: 3MEJ), depicting the conserved charged residues that stabilize the phosphate head-group of octaprenyl-phosphate.¹²⁶

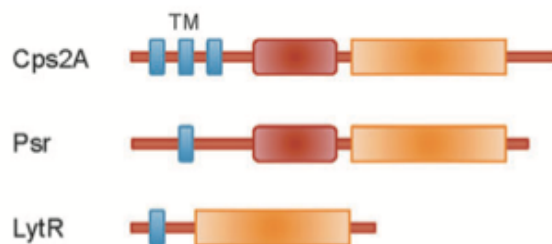


Figure 1.31 - Schematic representation of the sequence homology between the proteins of the LCP family. The transmembrane domain (TM) is depicted in blue, the accessory domain in red and the catalytic domain in orange.¹²⁶

Conventional antibiotics, such as β -lactams and glycopeptides, act through inhibition of the pathways responsible for PG synthesis¹²⁴. However, as they do not demonstrate activity against some bacteria in the form of biofilms, new compounds are examined for their anti-biofilm properties. Ellagic acid is a natural product (present in green tea) that was reported as having anti-biofilm activity. Thus, a virtual screening approach was employed aiming to discover novel compounds with improved properties. Two generations of ligand-based virtual screening against the Chinese Natural Product Database led, firstly, to esculetin and then to fisetin, as promising inhibitors of biofilm formation in *Staphylococcus aureus* and *Streptococcus dysgalactiae*.¹³⁰

In the present thesis, the first step is to determine the currently unknown three-dimensional structure of *Streptococcus dysgalactiae* LytR, followed by validation of ellagic acid and fisetin as promising inhibitors of this protein (Figure 1.32), which may explain their anti-biofilm activity.

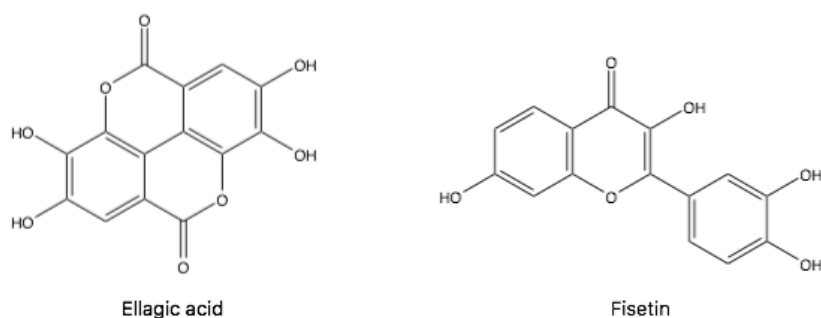


Figure 1.32 - Chemical structures of the candidate inhibitors under study for *Streptococcus dysgalactiae* LytR, ellagic acid and fisetin.

2. Materials and Methods

2.1. Target structural determination

2.1.1. Recombinant protein expression and purification

2.1.1.1. Human Bcl-2 protein

The studied Bcl-2 construct, purchased from NZYTech, lacks loop 2 from native Bcl-2 (residues 35-92). Instead, the protein has a derivative loop of the homologue anti-apoptotic protein, Bcl-XL (residues 35-50). This modification was made to improve protein stability, due to the intrinsic disordered nature and flexibility of loop 2 in native form, as reported by Petros *et al*⁹⁶. The recombinant protein studied (171 residues, M.W. = 21628 Da) has a His tag in its N-terminus, formed by six His residues, enabling protein purification through immobilized metal affinity chromatography (IMAC) in a fast protein liquid chromatography (FPLC) system (ÄKTA Start, GE Healthcare Life Sciences, 29-0220-94). Afterwards, this His tag was removed by incubation with thrombin from bovine plasma (Sigma-Aldrich, CAS: 9002-04-4), which cleaved Arg-1 and Gly+1 peptidic bond.

Recombinant protein expression was achieved using a pET28a(+) vector in transformed BL21 (DE3) *Escherichia coli* cells. These were transformed with approximately 0.4 nM of pDNA using a thermal shock method, consisting of 15 minutes in ice, followed by 45 seconds at 42° C and then again 15 minutes in ice. Afterwards, lysogeny broth (LB) medium was added to the cells suspension and incubated for 1 hour at 37° C. Then, the cells were recovered by centrifugation at 1000 x *g* during 3 minutes, followed by the removal of the excess LB medium and the pellets resuspended. The resulting cell suspension was spread onto solid LB medium supplemented with 50 µg/mL kanamycin (NZYtech, CAS: 25389-94-0) and incubated overnight at 37° C.

BL21 cells pre-inoculum was performed in 15 mL of LB medium with 50 µg/mL kanamycin, where an individualized colony was transferred to. The suspension was incubated overnight at 180 rpm and 37° C. Afterwards, the culture was transferred to 2 L non-baffled flasks with 500 mL of LB medium supplemented with 50 µg/mL kanamycin. The cell suspensions were incubated at 180 rpm and 37° C, until their optical density at 600 nm (OD₆₀₀) reached a value between 0.5 and 0.8. Subsequently, the flasks were cooled at 4° C, prior to induction with 0.5 mM of IPTG (Isopropyl-β-D-thiogalactoside) and then incubated overnight, at 180 rpm and 19° C. The grown and induced BL21 cultures were recovered by centrifugation in an Avanti J-26 XPI centrifuge (Beckman Coulter) with a JA-10 rotor (Beckman Coulter) at 6500 rpm for 15 minutes and the obtained pellets were stored at - 20° C.

The pellets were resuspended in buffer solution A, consisting of 50 mM HEPES pH 9.0, 300 mM NaCl, 10 mM imidazole and 1 mM β-mercaptoethanol. The suspension was subjected to 10 cycles of 1 minute of ultra-sounds at 0.5 cycles and 80% amplitude (UP100H, Hielscher Ultrasound Technology). The resulting extract was centrifuged for 1 hour at 10000 rpm and 4° C. The supernatant was recovered for the purification steps.

The first step of purification performed was done by IMAC, using a FPLC system coupled with a 5 mL Ni²⁺ column (HisTrap™ HP, GE Healthcare Life Sciences, 17-5247-01). The elution buffer, solution B, consisted in 50 mM HEPES pH 9.0, 300 mM NaCl and 1 M imidazole. After

equilibrating the system with buffer A, the sample was injected and the column was washed with 8 column volumes (CV) for removal of proteins with no affinity to the Ni²⁺. When clear flow-through was obtained, a 0.3 M imidazole elution was established. After a 4 CV elution of proteins with weak Ni²⁺ affinity, a 0.5 M imidazole gradient of 6 CV was performed, allowing the elution of several peaks, collected individually. Finally, the last fraction was collected with 1 M imidazole gradient of 1 CV. In order to assess which fraction contained the recombinant protein, 5 µL of each fraction were loaded in a 10% SDS-PAGE.

Afterwards, the selected protein fraction was dialyzed overnight using a 12000-14000 Da cut-off membrane (Dialysis tubes Membra-Cel™, Roth, 0654.1) against a buffer solution of 50 mM HEPES pH 9.0, 300 mM NaCl and 1 mM DTT (1,4-Dithiothreitol). Along with the protein fraction, 1 mL of 1 mg/mL thrombin from bovine plasma was added to guaranty proteolysis of the His tag fragment. A 10% SDS-PAGE was performed to verify the effectiveness of proteolysis.

The final purification step was achieved by size exclusion chromatography (SEC), using a high-performance liquid chromatography (HPLC) system (Prominence, Shimadzu) coupled with a superdex-75 10/300 column (GE Healthcare Life Sciences, 17517401). After system equilibration with buffer containing 50 mM HEPES pH 9.0, 500 mM NaCl and 1 mM DTT, Bcl-2 samples were centrifuged at 10000 rpm for 5 minutes at 4° C, to remove insoluble protein, and then injected in the HPLC system, 1.8 mg of protein/run. The flow rate was set to 0.6 mL/minute and 0.5 mL fractions were collected between 6.6 and 21.6 mL of elution volume. A 10% SDS-PAGE was carried out to select the pure fractions of Bcl-2 that were later retrieved for concentration. Pure Bcl-2 was concentrated in an Amicon® Ultra-15 Centrifugal Filter Unit with UltraCel-10 Membrane, using centrifugation cycles of 2500 x g, at 4° C and in an Amicon® Ultra-2 Centrifugal Filter Unit with UltraCel-10 Membrane, with cycles of 4000 x g, at 4° C.

2.1.1.2. *Streptococcus dysgalactiae* LytR

The LytR construct included only the extracellular domain of LytR from *Streptococcus dysgalactiae* (residues 48-342, 304 residues, M.W. = 33493 Da), having a His tag in its C-terminus, formed by six His residues, to enable protein purification through IMAC.

Recombinant protein expression was performed using a pET21c vector, in transformed BL21 (DE3) *E. coli* cells. The adopted transformation protocol was the same as mentioned earlier for Bcl-2 expression, section 2.1.1.1., with the difference that cells were incubated in solid LB with 100 µg/mL of ampicillin (Sigma-Aldrich, CAS: 69-52-3).

BL21 *E. coli* cells were inoculated in 15 mL of LB medium with 100 µg/mL of ampicillin and incubated overnight at 190 rpm and 37° C. The resulting cultures were transferred to 2 L non-baffled flasks containing 500 mL of LB medium with 100 µg/mL of ampicillin and incubated at 190 rpm and 37° C, until the suspension OD₆₀₀ reached values between 0.5 and 0.8. Then, LytR expression was induced by addition of 1 mM IPTG and incubated for 5 hours at 190 rpm and 30° C. The cell recovery and preservation followed the same steps as described before for Bcl-2 expression.

Cell lysis was achieved through ultra-sounds, similarly to the Bcl-2 procedure. In this case, the pellet was resuspended in 10 mM Na₃PO₄ pH 7.2, 100 mM NaCl, 5 mM MgCl₂ and 10 mM imidazole.

The first step of purification consisted of an IMAC and the elution protocol was the same as for Bcl-2. However, the used buffer solutions were different, since buffer A had 10 mM Na₃PO₄ pH 7.2, 100 mM NaCl, 5 mM MgCl₂ and 10 mM imidazole, and buffer B contained 10 mM Na₃PO₄ pH 7.2, 100 mM NaCl, 5 mM MgCl₂ and 1 M imidazole. In this case, the proteins with low affinity to the column were eluted in 4 CV, while the fraction of protein with high affinity was eluted in 5 CV.

Size exclusion chromatography was used to exchange protein buffer for imidazole removal. Three coupled HiTrap desalting columns (GE Healthcare Life Sciences, 11-0003-29) were equilibrated with 10 mM Na₃PO₄ pH 7.2, 500 mM NaCl and 5 mM MgCl₂ allowing 4.5 mL protein solution injections. UV-Visible detection at 280 nm and conductivity measurement enabled protein collection in the exchanged buffer.

The last purification step involved a SEC in a superdex-75 10/300 column. The system was equilibrated in 10 mM Na₃PO₄ pH 7.2, 500 mM NaCl and 5 mM MgCl₂ and 10 mg injections were performed in each run. Defined flow-rate was 0.6 mL/minute and 0.250 mL fractions were collected between 6.6 and 18 mL of elution volume. Pure LytR was selected according the 10% SDS-PAGE results obtained for the peak fractions. LytR was concentrated in an Amicon® Ultra-15 Centrifugal Filter Unit with UltraCel-10 Membrane, using centrifugation cycles of 2500 x g, at 4° C and in an Amicon® Ultra-2 Centrifugal Filter Unit with UltraCel-10 Membrane, with cycles of 4000 x g, at 4° C.

2.1.2. X-ray crystallography

2.1.2.1. Protein crystallization

2.1.2.1.1. Human Bcl-2 protein

Ligand-free Bcl-2 crystallization was attempted with 10 mg/mL of protein in 50 mM HEPES pH 8.0, 500 mM NaCl and 10 mM DTT at 4° C and 20° C, using seven reported precipitant solutions (PDB codes: 4AQ3, 4LVT, 4IEH, 4LXD, 4MAN, 4B4S, 2XA0)^{97,101,104,131,132}, with slight pH variations, as described in Table 2.1. The method employed was vapor diffusion through the hanging drop technique, using 600 µL of precipitant in the reservoir. Two different protein-precipitant ratios were tried, consisting of 1 µL of protein with 1 µL of precipitant and 2 µL of protein mixed with 1 µL of precipitant. Optimization of these conditions was attempted by decreasing and increasing the precipitating agents' concentration.

Table 2.1 – Precipitant solutions derived from published structures for Bcl-2 crystallization.

PDB code	Derived condition
4AQ3	0.1 M Tris-HCl pH 8.5, 1.25 M ammonium sulphate, 7% glycerol
4LVT	0.1 M sodium acetate pH 4.5, 0.2 M sodium chloride, 0.5 M ammonium sulphate
4IEH	0.1 M Tris-HCl pH 8.5, 0.05 M succinic acid, 0.25 M sodium malonate, 12% PEG 3350
4LXD	0.1 M MES pH 6.5, 1.5 M ammonium sulphate, 8% PEG 400
4MAN	0.1 M Tris-HCl pH 8.5, 1.5 M lithium sulphate
4B4S	0.1 M Tris-HCl pH 8.0, 0.1 M magnesium chloride, 20% PEG 400, 20% PEG 3350
2XA0	0.1 M sodium acetate pH 4.5, 3 M sodium formate

2.1.2.1.2. *Streptococcus dysgalactiae* LytR

The first approach to LytR crystallization was carried by Dr. Filipe Freire (currently at IBET/Merck Healthcare Laboratory), succeeding in obtaining protein crystals that diffracted poorly or didn't diffract when exposed to synchrotron X-ray radiation. These crystals were obtained through drops performed by a crystallization robot (Oryx8, Douglas Instruments) in 96-well plates using commercial screens (Jena Biosciences Classic 6, CS-106L, and JCSG+™, Molecular Dimensions, MD1-37). Protein concentration was at 100 mg/mL and 200 mg/mL, in drops with 0.6 µL of protein and 0.6 µL of precipitant, at 20° C. Two precipitant solutions were used: 2.2 M (NH₄)₂SO₄, 20% (v/v) glycerol and 25% (v/v) 1,3-propanediol, 10% (v/v) glycerol.

Optimization of these conditions was attempted through variation of precipitants concentrations (1.2-1.6 M (NH₄)₂SO₄, 5-25% (v/v) glycerol and 20-30% (v/v) 1,3-propanediol, 5% (v/v) glycerol) in scaled-up conditions with 1 µL protein and 1 µL precipitant drops, at 4° C and 20° C. Regarding the first condition, several different approaches were followed: besides varying precipitant concentration, different protein-precipitant drop ratios (1 or 2 µL protein and 1 or 2 µL precipitant drops), several ammonium salts (at 1.4 M: chloride, iodide, fluoride, phosphate and citrate), different PEG molecules (at 10% (w/v): 400, 2000, 8000, 20000), octyl-beta-glucoside (β-OG) as additive 1 and 2% (w/v); two distinct techniques (sitting and hanging drop) and microseeding approach were tested.

Furthermore, a different method, developed in the Max Planck Institute (Munich, Germany) in which (NH₄)₂SO₄ concentration is increased gradually by the addition of precipitant solution to the reservoir, was also tested. This allowed a screening of pH, ranging from 4 to 7.

Since no crystals were obtained, new crystallization conditions for LytR at 80, 100 and 150 mg/mL were searched using a crystallization robot, with drops of 0.67 µL of protein and 0.33 µL of precipitant solution and using different commercially available screens, namely JBS Classic 1-8 (Jena Biosciences, CS-101L-CS-108L). A condition with 0.1 M Tris-HCl 8.5, 0.2 M NH₄H₂PO₄, 50% (v/v) MPD (2-Methyl-2,4-pentanediol) was further optimized by changing the solution pH (8.0-9.0), NH₄H₂PO₄ (0.05-0.2 M) and MPD (30-50%, w/v) concentrations and by including additives (1% (v/v) β-OG). These variations were attempted in hand-made drops of 1 or 2 µL

protein and 1 or 2 μL precipitant, with 700 μL of precipitant in the reservoir and protein concentration between 100 and 130 mg/mL, at 20° C.

An *in house* screen was tested by adaptation of several conditions with MPD, other alcohols and organic solvents from different commercial screens available in the laboratory, namely 80!¹³³ (in house, sparse matrix), MacrosolTM (Molecular Dimensions, MD1-22), JCSG+TM (Molecular Dimensions, MD1-37), Crystal Screen 1 & 2 (Hampton Research, HR2-110 & HR2-112) and EWI & II (Wizard Classic 1 & 2, Molecular Dimensions, MD15-W12-B).

Two different commercial screens were also tested, the 80!¹³³ (in house, sparse matrix) and JCSG+TM (Molecular Dimensions, MD1-37) in hand-made drops of 1 μL protein and 1 μL precipitant with 700 μL of precipitant in the reservoir and LytR concentration at 60 mg/mL in 10 mM Na₃PO₄ pH 7.2, 250 mM NaCl and 5 mM MgCl₂, at 20° C.

The microbatch technique was also employed to attempt LytR crystallization using several promising conditions previously found. These conditions were 2.2 M (NH₄)₂SO₄ and 20% (v/v) glycerol; 25% (v/v) 1,3-propanediol and 10% (v/v) glycerol; 0.1 M Tris-HCl 8.5, 0.2 M NH₄H₂PO₄ and 50% (v/v) MPD; 0.1 M sodium cacodylate pH 6.5, 0.2 M (NH₄)₂SO₄ and 30% (w/v) PEG 8000. Drops with 2 μL protein and 2 μL precipitant were performed, using 700 μL of mineral oil in the reservoir (protein at 90 mg/mL in 10 mM Na₃PO₄ pH 7.2, 500 mM NaCl and 5 mM MgCl₂) at 20° C.

2.1.2.2. X-ray diffraction

Obtained crystals were flash frozen in liquid nitrogen using either glycerol or paratone as cryoprotectants. The crystals were exposed to synchrotron X-ray radiation at beamline I24, Diamond Light Source (DLS - Didcot, Oxfordshire, United Kingdom), beamlines ID29, ID30A3 and ID30B, European Synchrotron Radiation Facility (ESRF - Grenoble, France) and at beamline P13, PetralIII (Hamburg, Germany).

2.1.3. Homology Modeling

2.1.3.1. Human Bcl-2 protein

Homology modeling of Bcl-2 protein was achieved for 72% of the structure with more than 90% confidence. Three templates were selected to model the structure, corresponding to human Bcl-2 structures determined by X-ray crystallography (PDB code: 2XA0)⁹⁷ and NMR methods (PDB codes: 1G5M and 2O2F)⁹⁶. However, these templates didn't display homology regarding the missing loop 2 from Bcl-2, so this portion was derived from *ab initio* methods by the Phyre2¹³⁴ server.

2.1.3.2. *Streptococcus dysgalactiae* LytR

The studied extracellular domain of the *Streptococcus dysgalactiae* LytR protein was modeled using the ModWeb⁴⁵ server. The template selected for modeling is from the crystal

structure of *Streptococcus agalactiae*, gbs0355 (PDB code: 3OKZ)¹³⁵, which has 72% sequence similarity to the query sequence. The model covered 70% of the submitted LytR protein sequence.

2.2. Virtual Screening and Molecular Docking

2.2.1. Ligand-based

2.2.1.1. Human Bcl-2 protein

The compounds under study, including the drug molecule, venetoclax, the 4H-chromene reference molecule HA14-1, its derivatives (methoxy, ethoxy and activemethine) and the 4H-chromene moiety alone, were screened against a total of twelve small molecules databases: PubChem (compounds and substances), Drug Bank, Zinc, ChEMBL, eMolecules, ChEBI, Chem ID Plus, Super Natural II, Chem Bank, Cambridge Crystallographic Data Centre (CCDC), ChemBridge and National Cancer Institute (NCI). This ligand-based screening was performed for at least 90% structure similarity, except for CCDC since, in this case, it is not possible to define a similarity cut-off. The resulting compounds were selected for docking calculations with Bcl-2 protein as target, using AutoDock Vina version 1.1.2¹³⁶. The cubic box size was $x = 43.5$, $y = 33.0$, $z = 31.5$ Å, centred at $x = 6.25$, $y = 19.66$, $z = 19.60$ Å. Exhaustiveness was set to 10 and the maximum number of binding modes to output to 20.

2.2.2. Structure-based

2.2.2.1. Human Bcl-2 protein

Structure-based screening was performed by the Texas advanced computer centre server (TACC), from Texas University, USA, using Bcl-2 protein as target against four different databases: LigSearch, Traditional Chinese Medicine Database, Zinc and Cancer PPD (Database of anti-cancer peptides & proteins). The docking calculations were executed with AutoDock Vina version 1.1.2¹³⁶, similarly to the ligand-based screening procedure.

From the ligand-based and the structure-based screenings, the highest scored docked compounds of each database were chosen for further study. These ligands were visually inspected with PyMOL (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) which allowed to withdraw defected structures or several similar structures with the same binding affinity value from the top results list. Afterwards, the selected compounds were docked with Bcl-2, using AutoDock 4.2⁵⁹, in order to retrieve more accurate, thus more significant, binding affinity scores from the interaction between the ligands and the target protein.

2.2.3. Molecular Docking

2.2.3.1. Human Bcl-2 protein

Molecular docking studies were performed to assess possible interactions of several small molecule inhibitor candidates toward Bcl-2. Docking calculations regarding Bcl-2 were

completed using AutoDock version 4.2⁵⁹ for all studied ligands (4H-chromene derivatives, indole derivative and venetoclax). Docking parameters applied in the calculations were the program's default, apart from the number of genetic algorithm runs, which were set to 1000. The cubic box was defined with $x = 116$, $y = 88$, $z = 84$ Å number of points, grid spacing of 0.375 Å and grid centred at $x = 6.25$, $y = 19.66$, $z = 19.60$ Å. Bcl-2 studies included venetoclax, which is the most potent human Bcl-2 protein inhibitor commercially available, as reported by Souers *et al*¹⁰⁴. These results were compared to the crystal structure of Bcl-2 bound to the BH3 domain peptide derived from Bax (PDB code: 2XA0)⁹⁷.

2.2.3.2. *Streptococcus dysgalactiae* LytR

In the case of LytR, docking calculations were executed in AutoDock Vina version 1.1.2¹³⁶ with previously reported anti-biofilm small molecules, ellagic acid and fisetin¹³⁰. Some known lipidic substrates of homologue Cps2A proteins^{124,126}, such as decaprenyl and octaprenyl phosphates, were also docked, to establish comparisons with the studied ligands' binding. LytR docking results were analysed considering the crystal structure of gbs0355 from *Streptococcus agalactiae* (PDB code: 3OKZ)¹³⁵.

2.3. Target-ligand complex characterization

2.3.1. Thermal Shift Assay (TSA)

TSA were performed to assess Bcl-2 protein stability in a screening of buffer (Appendix 6.1) and additive (Appendix 6.2) solutions. These results provided valuable information regarding in which conditions the protein is more stable and influenced the choice of buffer solutions for the purification, crystallization, ITC and NMR experiments.

TSA were also used to probe ligand binding to Bcl-2 (ethoxy and activemethine 4H-chromene derivatives, indole derivative and venetoclax) and to LytR (ellagic acid and fisetin). A first approach consisted on using increasing concentrations of ligand solubilized in DMSO (dimethyl sulfoxide), corresponding to 0.5:1, 0.75:1, 1:1, 1.5:1, 2:1, 5:1 and 10:1 ligand-protein molar ratios. These assays included a measurement of the effect of DMSO concentration on Bcl-2 stability, as a control. A second approach, was meant to allow K_D determination of protein-ligand binding, which consisted on the above mentioned concentration proportions of ligand to protein, however, with a fixed percentage of DMSO of 2%, as reported by Iyer *et al*⁶⁵.

A third approach was performed, exclusively for Bcl-2, considering that either the presence of DMSO or the excess of ligand might interfere with protein stability and also with the probe fluorescence. Thus, the Bcl-2 samples used were previously incubated with indole derivative and venetoclax in 5% DMSO, after the last step of purification (SEC). Then, each sample was concentrated, allowing the removal of DMSO and excess of ligand.

The above-mentioned assays were completed in MicroAmp® fast 96-well reaction plates (Applied Biosystems™, ThermoFisher Scientific), using a total sample volume of 20 µL, containing 2 µL of protein with a final concentration of 10 or 5 µM, 10 µL of buffer, additive or

ligand solution, 3 μL of protein thermal shift dye (ROXTM) and 5 μL of dye buffer solution. The TSA were performed in 2 minute cycles of 1% increments between 25° C and 95° C in a StepOnePlusTM Real-Time PCR System (Applied BiosystemsTM, ThermoFisher Scientific, 4376600). Data processing and analysis was performed with Protein Thermal ShiftTM software from Applied BiosystemsTM, ThermoFisher Scientific.

2.3.2. Urea-Polyacrylamide Gel Electrophoresis

2.3.2.1. Human Bcl-2 protein

Urea-polyacrylamide gel electrophoresis was performed using a Novex 6% Tris-TBE urea minigel in a XCell SureLockTM Mini-Cell from Invitrogen. Running buffer was diluted 1:5 from 89 mM Tris base and 89 mM boric acid, while the sample buffer used contained 45 mM Tris base and 45 mM boric acid, 6% FicollR Type 400, 3.5 M urea and 0.005% bromophenol blue.

Bcl-2 was incubated with each ligand (ethoxy and activemethine 4H-chromene derivatives and indole derivative), for 1 hour, at a concentration of 50 μM (1 mg/mL) and 1:10 protein-ligand molar ratio. Then, the excess of ligand was removed by desalting in a PD MiniTrap G-25 column (GE Healthcare Life Sciences, 28-9180-07) which diluted the protein to 0.5 mg/mL. Afterwards, Bcl-2 was concentrated to 1 mg/mL and samples for the urea gel were prepared, using 5 μL of protein-ligand solution and 5 μL of sample buffer. The samples loaded to the gel included a control of Bcl-2 without ligands at 1 mg/mL as well, and were kept at 4° C overnight before loading. Finally, the samples were loaded into the urea gel and it was subjected to 180 V and 40 mA for 2 hours and 30 minutes.

A similar approach to the one described above for the TSA was attempted, where in the final purification step, protein was incubated with each ligand and then concentrated. The loaded samples had 5 μL of ligand incubated protein at 15 mg/mL and 5 μL of sample buffer. In this case, the urea gel was under 180 V and 40 mA for 1 hour.

2.3.3. Isothermal Titration Calorimetry (ITC)

In both case studies, the ITC experiments were performed in a Nano ITC Standard Volume (TA Instruments) and data was processed and analysed through the NanoanalyzeTM software from TA Instruments.

2.3.3.1. Human Bcl-2 protein

Before the ITC experiments, the protein solution was dialyzed in two steps, in a buffer containing 50 mM HEPES pH 9.0, 500 mM NaCl and 1 mM DTT. The first dialysis was performed for 2 hours at 4° C, in 500 mL of buffer. The second one occurred overnight at 4° C, in also 500 mL of buffer. The ITC instrument cell was filled with protein solution at 50 μM and the syringe with ligand solution in the dialysis buffer. For the experiment with indole derivative, ligand

concentration was 125 μM in 2.5% DMSO. In the case of venetoclax, the ligand was at 100 μM in 3.5% DMSO. Both experiments were performed at 20° C with 10 μL ligand injections.

2.3.3.2. *Streptococcus dysgalactiae* LytR

Similar to what was described before, the protein solution was dialyzed in two steps. In this case, the buffer had 10 mM Na_3PO_4 pH 7.2, 500 mM NaCl and 5 mM MgCl_2 . Protein concentration in the instrument cell was at 50 μM , while the injections were of 250 μM fisetin in 5% DMSO. The experiment was performed at 25° C, with 10 μL injections. In the case of LytR, a blank experiment was performed, where the same conditions were applied in the absence of ligand.

2.3.4. Circular Dichroism (CD)

2.3.4.1. *Streptococcus dysgalactiae* LytR

Fisetin binding to LytR was investigated through CD. Protein was at 0.5 mg/mL in 10 mM Na_3PO_4 pH 7.2, 150 mM NaCl, 5 mM MgCl_2 and 0.15% DMSO, to facilitate ligand solubilisation. Ligand was added to protein solution to a final concentration of 17 μM , which corresponds to a 1:1 protein-ligand molar ratio. Spectra were acquired between 190 and 260 nm in triplicate with 5 s/nm measurements, at 20° C, in a Chirascan-plus qCD spectrometer (Applied Photophysics).

2.4. Target-ligand structural determination

2.4.1. X-ray crystallography

2.4.1.1. Human Bcl-2 protein

Following the methodology reported by Perez *et al*¹³¹, 1 mg/mL of pure Bcl-2 was incubated overnight, at 4° C, with 1 mM of indole derivative, in 7.5% DMSO to facilitate its solubilisation. Protein was concentrated to 15 mg/mL in 50 mM HEPES pH 9.0, 500 mM NaCl and 1 mM TCEP (Tris(2-carboxyethyl)phosphine), allowing the removal of DMSO and excess ligand, and drops were made using the same parameters as before, including the variations in precipitant concentration of the published conditions. Several commercially available screens were also tested, namely JBS Classic 1-8 (Jena Biosciences, CS-101L-CS-110L), for Bcl-2 and indole co-crystallization at 5 mg/mL. A crystallization robot was used for testing different screenings, using 0.67 μL of protein and 0.33 μL of precipitant drops and the sitting drop technique. The plates were stored at 20° C. Some promising results were obtained in the presence of alcohols and a customized screen was prepared based on several conditions containing MPD, other alcohols and organic solvents from different commercial screens available in the laboratory, namely 80¹³³ (*in house*, sparse matrix), MacrosolTM (Molecular Dimensions, MD1-22), JCSG+TM (Molecular Dimensions, MD1-37), Crystal Screen 1 & 2 (Hampton Research, HR2-110 & HR2-112) and EWI & II (Wizard Classic 1 & 2, Molecular Dimensions, MD15-W12-B).

Co-crystallization was also attempted by ligand incubation with 10:1 protein-ligand molar ratio and 5% (v/v) DMSO. Bcl-2 was incubated with the indole derivative and venetoclax and concentrated to 15 mg/mL. Crystallization trials were carried out using the same crystallization conditions previously mentioned.

Since interesting results were achieved with MPD, co-crystallization experiments with both ligands were performed using three different values of pH (4.5, 6.0 and 7.5) and a different range of this precipitant concentration (20-50%, v/v). The vapor diffusion method (hanging drop technique) was used, with drops containing 1 μ L of protein and 1 μ L of precipitant, at 20° C.

At the same time, a preliminary condition found in an initial commercial screen was also optimized for the Bcl-2-venetoclax complex, varying both concentrations of MPD (5-25%, w/v) and ethanol (5-35%, v/v), in 0.1 M Tris-HCl pH 8.5 and 0.1 M MES pH 6.0. Drops with 1 μ L of 15 mg/mL protein and 1 μ L of precipitant were made at both 4° and 20° C. Optimization of these conditions, at pH 6.0, was further attempted by varying protein concentration (7.5 and 10 mg/mL), drop ratio (1 μ L or 2 μ L of protein and 1 μ L of precipitant), reservoir volume (600 and 700 μ L) and by replacing MPD for PEG 400.

Three different crystallization additive screens were tried, namely JBScreen Plus volatiles (Jena Biosciences, CS-505), salts (Jena Biosciences, CS-503) and additives (Jena Biosciences, CS-504), using the parameters recommended by Jena Biosciences, in 2 μ L drops, at 20° C. Optimization of the additives screens results was carried by performing slight variations in precipitant concentrations and, in the case of the volatile additive, by increasing its concentration.

2.4.2. Saturation Transfer Difference NMR (STD-NMR)

For both Bcl-2 and LytR binding studies through STD-NMR, the protocol followed was developed by Viegas *et al*⁷⁷. The protein-ligand mixtures contained protein at 50 μ M, 500 μ M of ligand, 10% D₂O and 5% DMSO-d₆.

Bcl-2 was in 50 mM Na₃PO₄ pH 8.0, 500 mM NaCl and 10 mM DTT and only the indole derivative was probed for binding. Apart from the mixture described previously, a second experiment was performed where protein and ligand concentrations were reduced to 25 μ M and 250 μ M, respectively. LytR buffer had 10 mM Na₃PO₄ pH 7.2, 500 mM NaCl and 5 mM MgCl₂ and both ligands, ellagic acid and fisetin were tested for protein binding.

¹H NMR spectra were acquired in a Bruker Avance 600 MHz NMR spectrometer and the software used for experiment preparation and data processing was Bruker TopSpinTM. Relevant Bcl-2 and LytR experimental details are presented in Table 2.2, including the pulse program, number of scans and dummy scans, increments, central frequency of the spectrum, ¹H sweep width, number of averages and number of irradiation frequencies.

Table 2.2 – Experiment details of the STD-NMR spectra of Bcl-2 and LytR.

Experiment parameters	Bcl-2	LytR
Pulse programs	zgespgp / stddiffesgp	
Number of scans / dummy scans	32 and 48 / 4	
Increments (TD)	4096 and 131072	
Central frequency (O1)	2809 Hz	
¹ H sweep width	12626 Hz / 21.03 ppm and 12335 Hz / 20.55 ppm	
Number of averages (L4)	40	
Number of irradiation frequencies (NBL)	2	

2.4.3. Molecular Dynamics (MD)

2.4.3.1. Bcl-2 protein-ligand complexes

MD simulations were performed using the Gromacs software version 5.0.5¹³⁷, for native Bcl-2 protein, for the studied compounds individually and, finally, for the complexes between Bcl-2 and each ligand (ethoxy derivative, indole derivative and venetoclax). All MD simulations were executed according to the procedures provided in *Gromacs Tutorials* website by Dr. Justin A. Lemkul of the University of Maryland, Baltimore, USA.

The MD simulations for the native Bcl-2 protein were performed using the OPLS-AA/L all-atom force field¹³⁸ and the default cubic box parameters. 15049 explicit flexible SPC (Single Point Charge) water molecules¹³⁹ were generated and one was replaced by a sodium ion to balance the global net charge of -1.00 e. After energy minimization and temperature and pressure equilibration of the system, the MD simulation was carried out for 30 ns, at physiological temperature (37° C) and atmospheric pressure (1.0 bar). Afterwards, the results were analysed, through built-in functions of Gromacs software, with respect to the following parameters: root mean-square deviation (rmsd, function: g_rms), root mean-square fluctuation (rmsf, function: g_rmsf), gyration radius (function: g_gyrate), hydrogen bonds (function: g_hbond), potential energy (function: g_energy), solvent accessible surface (function: g_sas) and essential dynamics analysis (E.D., function: g_covar and g_anaeig).

In order to perform the MD simulations of the protein-ligand complexes, the ligand topology files were generated by the automated topology builder (ATB version 2.0)¹⁴⁰. To corroborate the obtained topologies, MD simulations of each compound were carried out, using the GROMOS96 54a7 force field¹⁴¹ and the cubic box parameters. After solvent generation of explicit flexible SPC water¹³⁹, the systems global net charges were zero, so electrostatic balancing by ion replacement was not required. After the systems energy minimizations and temperature and pressure equilibrations, the MD simulations were completed for 50 ns, at default temperature (27° C) and pressure (1.0 bar). The results were analysed and compared among the different compounds leading to the assessment of their relative hydrophobicity and flexibility. This analysis was performed through the built-in Gromacs software¹³⁷ functions for the following parameters:

rmsd, potential energy, surface accessible surface and essential dynamics analysis. The successful MD simulation of these compounds confirmed that the topology files generated by ATB¹⁴⁰ were correct and could be used in protein-ligand complexes MD simulations.

MD simulations of the protein-ligand complexes were executed using the GROMOS96 54a7 force field¹⁴¹, the ligands topology files generated by ATB¹⁴⁰ and the default cubic box. The ligands involved in these simulations were the reference drug, venetoclax, and the ethoxy and indole derivatives. The specific parameters concerning the number of SPC water molecules¹³⁹ added to the box, the global system net charge and the type and number of ions used for electrostatic equilibration are displayed in Table 2.3. After energy minimization and system equilibration, the MD simulations were performed for 30 ns, at 37° C and 1.0 bar.

Comparative analysis was performed for the three protein-ligand complexes with each other and also relative to ligand-free Bcl-2 form. The analysed parameters are the ones above mentioned for ligand-free Bcl-2 with the exception of residue occupancy probability, which was completed only for the MD simulations of the complexes. The residue occupancy probability is a method that provides information regarding the frequency of contacts between the ligand atoms and the protein residues, which are significantly close to the ligand, equal or less than 3 Å apart. This analysis allows the retrieval of valuable information regarding the influence of each ligand on protein structure, stability and specific interactions with given small molecules.

Table 2.3 – Specific input parameters for MD simulations of each protein-ligand complex studied for Bcl-2.

Ligand	SPC water molecules (#)	System net charge (e)	Ion replacement (# Na ⁺)
Native	15049	-1.00	1
4H-chromene ethoxy derivative	10239	-5.00	5
Indole derivative	11466	-5.00	5
Venetoclax	13010	-5.00	5

2.4.3.2. Structural impact of non-synonymous single nucleotide polymorphisms (nsSNP) in Human Bcl-2 protein

For an exhaustive analysis of the structural effects of non-synonymous single nucleotide polymorphisms (nsSNP) on Bcl-2 (full-length) and its putative implications in diseases, a bioinformatics search was performed for all validated non-synonymous mutations of both isoforms of human Bcl-2 protein, alpha (UniProt: P10415) and beta (P10415-2).

The nsSNP were retrieved from dbSNP (NCBI) database, filtering the search with the following criteria: *Homo sapiens* (Organism), SNP (Variation Class), missense (Function Class), by-1000 Genomes, by-cluster and by-frequency (Validation Status). This search resulted in 36 SNP retrieved for isoform alpha and 3 non-redundant additional SNP for isoform beta. After

gathering 39 different reported Bcl-2 SNP, analysis was performed using eight different programs, in order to assess the potential phenotypic effect of these mutations: I-Mutant¹⁴², Panther¹⁴³, SNP&Go¹⁴⁴, SIFT¹⁴⁵, Provean¹⁴⁶, Polyphen 2.0¹⁴⁷, nsSNPAnalyzer¹⁴⁸ and PhD-SNP¹⁴⁹. These programs attempt to predict the SNP impact on protein stability and function, using bioinformatic methods, such as machine learning and hidden Markov models. The 39 SNP lists were filtered manually by eliminating SNP with less than five *deleterious* predictions from the eight programs used, yielding a group of 12 SNP. Afterwards, the effect of the 12 nsSNP on structure stability was predicted through seven different bioinformatics programs: I-Mutant¹⁴², mCSM¹⁵⁰, SDM¹⁵¹, DUET¹⁵², MuPro¹⁵³, INPS-MD¹⁵⁴ and iStable¹⁵⁵. These programs allowed the stability effect prediction through sequence or structure submission, or both.

The MD simulation for the alpha native form was carried out using the OPLS-AA/L all-atom force field¹³⁸ and the default cubic box parameters. 15049 explicit flexible SPC water molecules were added to the box and electrostatic balance was accomplished by replacement of one water molecule for a sodium ion. After system energy minimization and temperature and pressure equilibration were completed, the MD simulation was performed for 30 ns, at 37° C and 1.0 bar. The comparison between native and nsSNP forms was achieved through the above mentioned built-in functions of Gromacs software¹³⁷, including the rmsd, rmsf, gyration radius, hydrogen bonds, potential energy, solvent accessible surface and essential dynamics analysis parameters.

The 12 identified and filtered nsSNP of isoform alpha were submitted to MD simulations. For that purpose, their PDB files were generated through mutation of the native form using Coot software version 0.8.2¹⁵⁶. The MD simulations were performed as in the native form procedure and the results analysis included the same parameters for comparison between native and nsSNP forms. The parameters used in the MD simulations specific for each nsSNP namely the number of SPC water molecules¹³⁹ in the box, the system global net charge and the type and number of ions added for electrostatic equilibration are presented in Table 2.4.

Table 2.4 – Specific input parameters for MD simulations of each Bcl-2 nsSNP studied.

nsSNP	SPC water molecules (#)	System net charge (e)	Ion replacement (# Na ⁺)
Native	15049	-1.00	1
G8E	15411	-2.00	2
D34Y	15419	0.00	0
A43T	15419	0.00	0
H94P	15364	-1.00	1
L97P	15364	-1.00	1
F104S	15423	-1.00	1
S105F	15416	-1.00	1
S105P	15366	-1.00	1
R129C	15418	-2.00	2

G203S	15414	-1.00	1
R207W	15415	-2.00	2
G233D	15415	-2.00	2

3. Results and Discussion

3.1. Target structural determination

3.1.1. Recombinant protein purification

3.1.1.1. Human Bcl-2 protein

The first stage of human Bcl-2 protein purification comprised an immobilized metal affinity chromatography (IMAC), where the recombinant protein is separated from the *E. coli* BL21 endogenous proteins due to the high affinity of its His tag for the Ni²⁺ in the column matrix. As seen in Figure 3.1A, the soluble fraction from the cell lysis has a high content of proteins that don't show affinity for the metal, being eluted first. Afterwards, proteins that display low affinity toward Ni²⁺ are removed from the column matrix by addition of imidazole, while the recombinant protein maintains adsorbed. Finally, Bcl-2 is eluted by increasing the imidazole concentration.

SDS-PAGE is a technique that allows the separation of proteins according to their molecular weight, exclusively. Therefore, the application of this technique to a purification process is of great value, since it is possible to identify the fractions that contain the protein of interest and assess their purity, following each step of the purification protocol. In this case, the 10% SDS-PAGE with each corresponding fraction of the IMAC identifies peak 5 as containing Bcl-2 protein (Figure 3.1B). It is noteworthy that the remaining fractions don't have the Bcl-2 band, which indicates that the column's capacity wasn't exceeded.

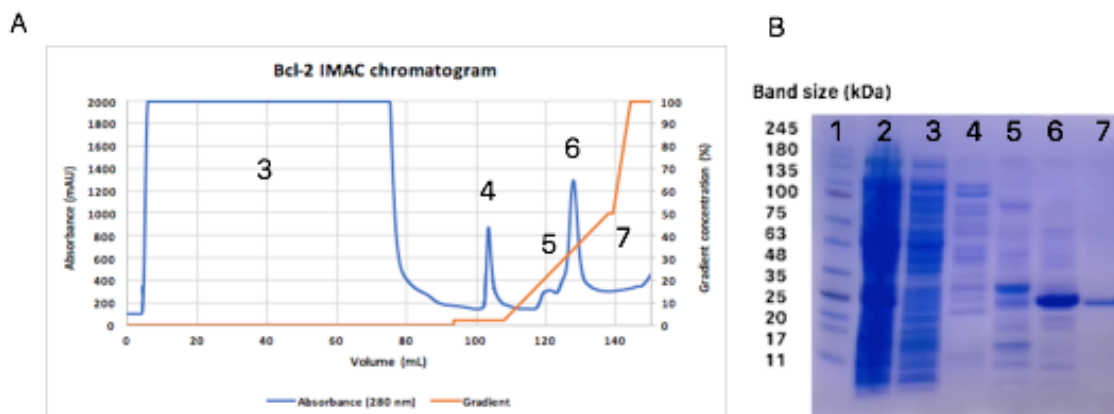


Figure 3.1 - A. Bcl-2 IMAC chromatogram with corresponding fractions noted on the SDS-PAGE gel (B). Gel wells legend: 1. NZYTech protein marker II; 2. Soluble fraction from cell lysis; 3. Proteins with no affinity to the column (first chromatogram peak); 4. Proteins with low affinity to the column (second chromatogram peak); 5. High affinity contaminant fraction (shoulder of the third chromatogram peak); 6. Bcl-2 fraction; 7. Column's washing step fraction.

After IMAC, it is necessary to remove imidazole from the Bcl-2 solution, since it is a destabilizer of the protein intramolecular interactions. Also, the cleavage of the His tag is accomplished by incubation with thrombin. Figure 3.1B shows, in lane 7, a single band corresponding to Bcl-2 with lower M.W. compared to the band derived from the IMAC fraction, indicating a successful enzymatic proteolysis.

The final purification step is a size exclusion chromatography (SEC), which is necessary to obtain pure Bcl-2. The chromatogram (Figure 3.2A) shows thrombin being eluted first, since it has a higher molecular weight (around 36 kDa) and pure Bcl-2 fractions collected afterwards. The 10% SDS-PAGE gel shows that the first peak doesn't contain Bcl-2 and also which fractions can be considered pure (Figure 3.2B). The protein purification and concentration process resulted in a final yield of 4 mg of pure Bcl-2 per liter of cell culture.

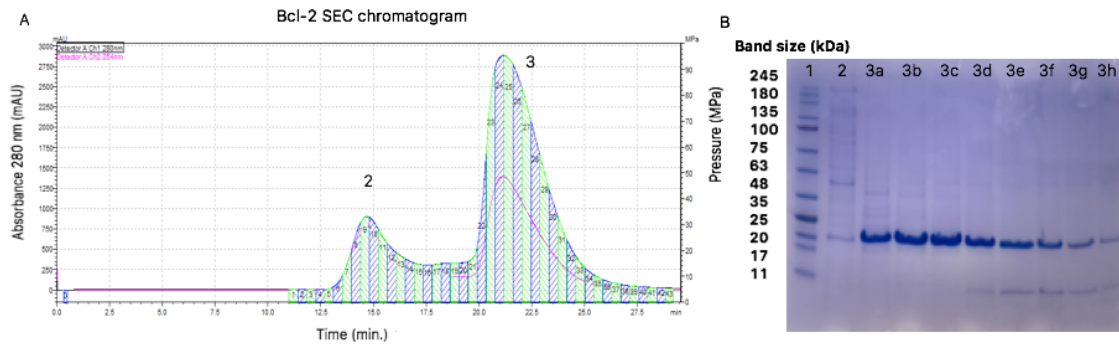


Figure 3.2 - A. Bcl-2 SEC chromatogram with corresponding peaks noted on SDS-PAGE gel (B). Gel samples: 1. NZYTech protein marker II; 2. Thrombin from bovine plasma (first chromatogram peak); 3. Different fractions of the Bcl-2 peak named from 3a to 3h (second chromatogram peak).

3.1.1.2. *Streptococcus dysgalactiae* LytR

Streptococcus dysgalactiae LytR purification starts with an IMAC, isolating the recombinant protein through its His tag high affinity toward the immobilized Ni^{2+} (Figure 3.3A). LytR is eluted as a single peak, fraction 5, as observed in the 10% SDS-PAGE (Figure 3.3B).

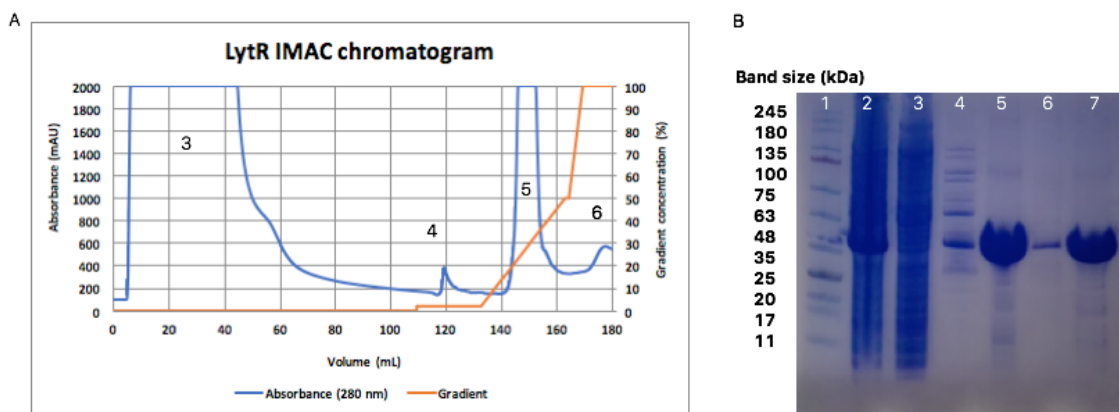


Figure 3.3 - A. LytR IMAC chromatogram with corresponding peaks noted on SDS-PAGE gel (B). Gel samples: 1. NZYTech protein marker II; 2. Soluble fraction from cell lysis; 3. Proteins with no affinity to the column (first chromatogram peak); 4. Proteins with low affinity to the column (second chromatogram peak); 5. Fraction containing LytR; 6. Column's washing step fraction; 7. Fraction containing LytR in exchanged buffer (desalting).

Similar to the Bcl-2 purification process, the imidazole in the protein solution must be removed. Therefore, a desalting step is performed by SEC (Figure 3.4). After the IMAC LytR

fraction injection, the protein is rapidly eluted in the buffer of interest. Then, the salt and imidazole pass through the column as it is observed by the change in conductivity and absorbance, respectively. The 10% SDS-PAGE shows that the fraction in the new buffer maintains the same electrophoretic profile, indicating that there wasn't a significant protein loss and the new buffer is suitable for the protein (Figure 3.3B).

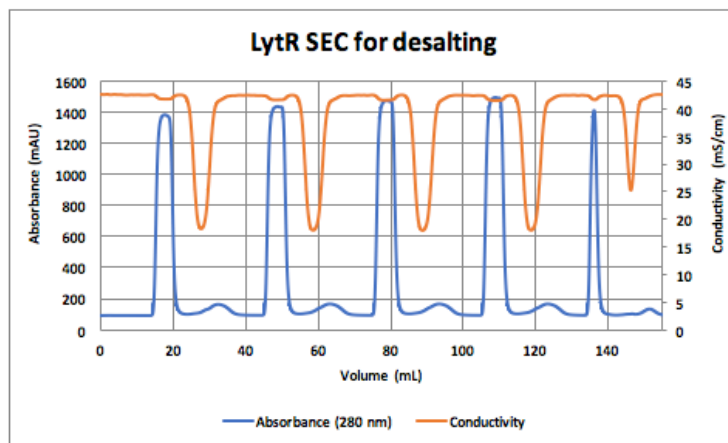


Figure 3.4 - LytR SEC chromatogram for the buffer exchange. Each peak in blue corresponds to a different injection.

Finally, the last stage of purification is a SEC in which the LytR protein is separated from possible contaminants with different molecular weight. The chromatogram shows that a contaminant was present, by the elution of two distinct peaks, with relatively close molecular weights (Figure 3.5A). Ideally, more resolution would be preferable to allow a clear separation of the two entities. However, with smaller fractions collected (250 μ L each), it is possible to obtain a reasonable amount of pure protein. The 10% SDS-PAGE shows that the second peak corresponds to the protein of interest (Figure 3.5B). The overall purification and concentration protocol of LytR enables a significant yield of 24 mg of pure protein per liter of culture.

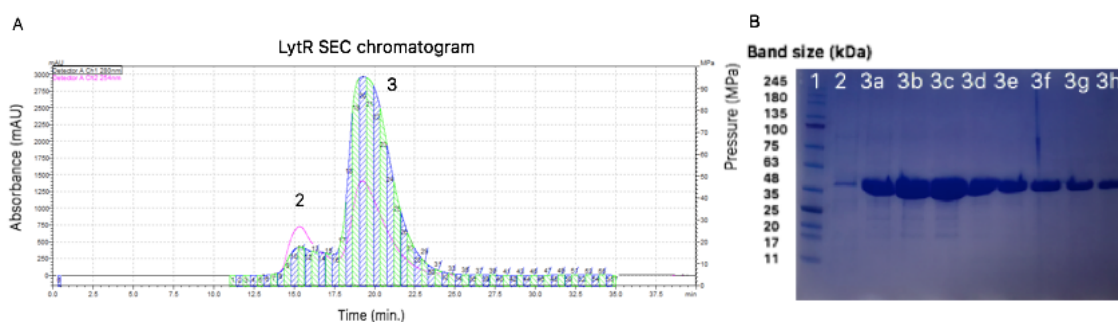


Figure 3.5 - A. LytR SEC chromatogram with corresponding peaks noted on SDS-PAGE gel (B). Gel samples: 1. NZYTech protein marker II; 2. Contaminant from the LytR containing fraction of IMAC (first chromatogram peak); 3. Different fractions of the LytR peak named from 3a to 3h (second chromatogram peak).

3.1.2. X-ray crystallography

3.1.2.1. Human Bcl-2 protein

Among the seventeen Bcl-2 structures deposited in the PDB to date, six of them were solved by NMR and eleven by X-ray crystallography. Only two structures correspond to ligand-free Bcl-2 protein and the remaining fifteen are from Bcl-2 complexes with inhibitors, either small molecules or peptides. These very few ligand-free Bcl-2 structures were determined by NMR methods and this fact stands as proof of the complex and difficult endeavour of crystallizing ligand-free Bcl-2 protein.

As a first approach, seven different crystallization conditions were adopted from already published work (PDB codes: 4AQ3, 4LVT, 4IEH, 4LXD, 4MAN, 4B4S, 2XA0)^{97,101,104,131,132}. Some very relevant information wasn't available in the corresponding papers, such as protein concentration, protein buffer, technique implemented for vapor diffusion, drop protein-precipitant ratio or conditions of ligand incubation. Crystallization of ligand-free Bcl-2, with the precipitants reported in the publications previously mentioned, yielded no crystals. Protein precipitation was observed in a wide range of conditions tested. In order to avoid protein precipitation to this great extent, these conditions were used, but with lower concentrations of precipitant. Once again, the main observable outcome was protein amorphous precipitation. However, in one condition, using PEG 400 and PEG 3350 as precipitant agents, crystalline material was observed through polarized light. These crystals couldn't be collected successfully and further crystallization attempts of ligand-free Bcl-2 were dismissed. The difficulty in obtaining Bcl-2 crystals, even in the reported conditions in the literature, led to the acknowledgement that perhaps Bcl-2 cannot be crystallized in the absence of antagonists, which promote its structural stabilization.

Structure determination of Bcl-2 was not accomplished through X-ray crystallography, however, as mentioned previously, it is already available on PDB and the entry 4AQ3¹³¹, which has the same protein sequence as the studied recombinant protein, was selected for the virtual screening approach.

3.1.2.2. *Streptococcus dysgalactiae* LytR

Regarding LytR from *Streptococcus dysgalactiae*, no structures are deposited on the PDB. One can only find homologous proteins that belong to the LCP family, from different organisms. Also, structures with substrate molecules bound are for example PDB entries 2XXQ and 4DE9, which have decaprenyl-phosphate and octaprenyl-phosphate present, respectively^{124,126}.

Apo LytR was previously crystallized in our group, however, the crystals weren't optimized and diffracted poorly when exposed to X-ray radiation from a synchrotron source. These crystals were obtained through drops performed by a crystallization robot using commercial screens.

This crystallization hit condition was obtained in drops with 0.6 μL of protein and 0.6 μL of precipitant, at 20° C and the precipitant solution had 2.2 M ammonium sulphate and 20% (v/v) glycerol. Protein concentration was varied between 100 mg/mL and 200 mg/mL, resulting in crystals that diffracted between 3 and 3.5 Å. Optimization of these conditions was attempted in scaled-up conditions providing crystals at 20° C, but with small size and multiple nature (Figure 3.6). The crystals were not reproducible and variation of numerous parameters, such as precipitant concentration, drop ratio, and incorporation of additives, didn't yield any protein crystals. Furthermore, a different method, developed in the Max Planck Institute, in which ammonium sulphate concentration was increased in increments only produced salt crystals, in the condition with pH 7. At lower pH values, protein precipitation was observed.

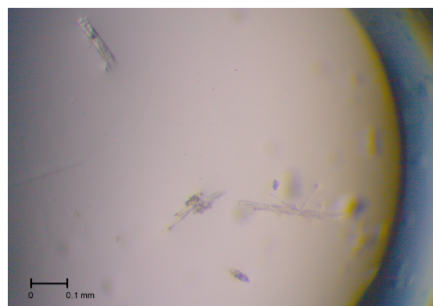


Figure 3.6 - Crystals of multiple nature obtained in a derived condition with ammonium sulphate and glycerol as precipitant agents.

Another promising condition retrieved from the tested commercial screens had 25% (v/v) 1,3-propanediol and 10% (v/v) glycerol in the precipitant solution. Similar to the approach adopted for the previously mentioned crystallization condition, optimization in scaled-up conditions was pursued, however, no crystals were obtained.

In order to find new crystallization conditions for LytR, different commercial screens were tested using a crystallization robot. The results showed MPD as promising precipitation agent, since crystals were obtained in a condition with 0.1 M Tris-HCl pH 8.5, 0.2 M ammonium phosphate and 50% MPD. The crystals were tested for X-ray diffraction at the I24 beamline, DLS, which yielded no diffraction, confirming that they were protein crystals. In order to obtain well-ordered crystals, variation of precipitant concentration and buffer pH was implemented in scaled-up conditions. However, this resulted in amorphous protein precipitate several minutes after the drops were made. Needle-shaped crystals were observed a few minutes after mixing protein and precipitant solutions (Figure 3.7A) and large crystalline agglomerates formed in the next days (Figure 3.7B). The fact that so much protein had precipitated and the rapid appearance of crystals that formed large irregular aggregates led to the realization that these were salt crystals.

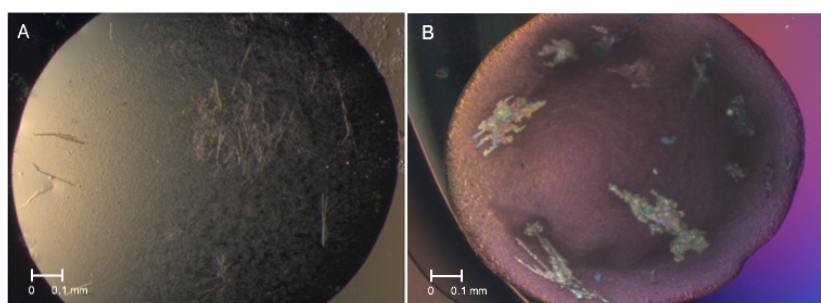


Figure 3.7 - Crystals obtained from a condition with 50% MPD as precipitant agent. A. Few minutes after the drop was made. B. Several hours upon protein and precipitant mixing.

Two different crystallization screens, namely 80!¹³³ and JCSG+TM, were used to search for new preliminary conditions. Several conditions produced crystals that were collected and submitted to synchrotron X-rays, at P13 beamline, PetralIII. Among the fifteen crystals, from six different conditions, only one did not display characteristic salt diffraction (Figure 3.8A). The precipitant solution in this condition had 0.1 M Bis-Tris pH 5.5, 0.2 M magnesium chloride and 25% (w/v) PEG 3350. Another promising condition was found to produce crystals, however, upon collection, the crystals dissolved (Figure 3.8B). This fragility indicates that they should be protein crystals, adding this condition with 0.1 M cacodylate pH 6.5, 0.2 M ammonium sulphate and 30% (w/v) PEG 8000, to the list of preliminary crystallization conditions of LytR.

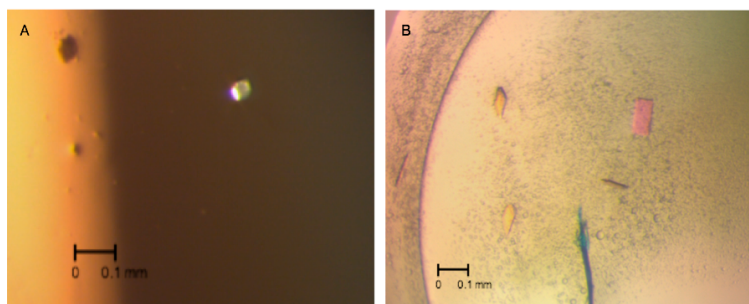


Figure 3.8 - Crystals obtained from two different conditions. A. Precipitant solution had 0.1 M Bis-Tris pH 5.5, 0.2 M magnesium chloride and 25% (w/v) PEG 3350. B. Precipitant solution consisted of 0.1 M cacodylate pH 6.5, 0.2 M ammonium sulphate and 30% (w/v) PEG 8000.

As a final attempt to reproduce the crystallization conditions found before, a crystallization robot was used to perform drops with the same drop protein-precipitant ratio, varying only the protein concentration. This procedure was also applied for two novel preliminary conditions, that had MPD and PEG 8000 as precipitant agents. Furthermore, microbatch under oil experiments with these four promising conditions were tried. Only the drops performed by the crystallization robot yielded crystals. Five crystals were collected and tested in the ID30A3 beamline, ESRF, however, four of them didn't diffract, while one resulted in salt diffraction. These results indicate that in fact these conditions can yield protein crystals, although they must be optimized to generate well-ordered crystals that can diffract X-rays at high resolution, which would enable structure determination.

In the light of the presented results, homology modeling emerges as an alternative to achieve a three-dimensional model of LytR that may allow the implementation of the virtual screening strategy.

3.1.3. Homology Modeling

3.1.3.1. Human Bcl-2 protein

As mentioned previously, for the virtual screening approach the human Bcl-2 structure used was retrieved from PDB entry 4AQ3. However, the protein's highly-flexible loop 2 was replaced by a loop derived from the homologous protein Bcl-XL to promote protein stability⁹⁶, therefore, there's no structural information regarding that region of Bcl-2. This issue does not affect significantly the virtual screening for candidate inhibitors, since the Bcl-2 binding region of interest is not related to this loop, being near the BH3 domain. However, for the prediction of the impact of nsSNP on Bcl-2 structure and stability, through MD, it is fundamental to have its full-length model. This was carried out before the present work by Dr. Jayaraman Muthukumaran using the Phyre2¹³⁴ server to perform homology modeling, combined with *ab initio* methods.

The human Bcl-2 protein sequence was submitted to the Phyre2¹³⁴ server using the intensive mode protocol. The three-dimensional structure was modeled through three different templates (PDB code: 2XA0, 1G5M, 2O2F)^{96,97}. The sequence identity between the query and the chosen templates varies between 45 and 88%, which indicates a high probability of predicting a reliable model, being the standard sequence homology inferior limit of 30%⁴³. The robustness of the prediction is accentuated by the fact that 72% of the residues from the query were modeled with more than 90% confidence, which usually results in deviations of 2 to 4 Å between the model predicted and the *real* structure, in the core regions¹³⁴. As there is no template available for the loop 2 region of Bcl-2, this portion was modeled by *ab initio* methods, which entailed energy minimization iterations to build the three-dimensional arrangement of the loop's main-chain. The full-length model of Bcl-2 protein predicted by Phyre2¹³⁴ is depicted in Figure 3.9.

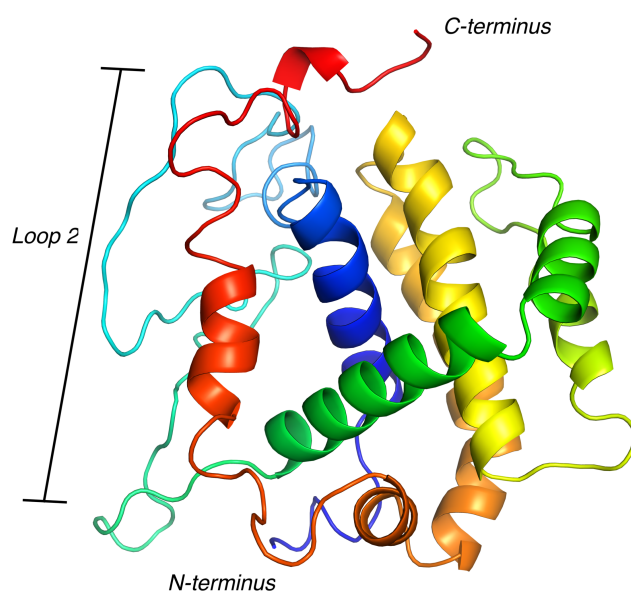


Figure 3.9 - Bcl-2 full-length model predicted by the Phyre2¹³⁴ server, through homology modeling and *ab initio* methods. The highly-flexible Bcl-2 loop 2 is highlighted. Representation was performed using the PyMOL software (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

3.1.3.2. *Streptococcus dysgalactiae* LytR

Streptococcus dysgalactiae LytR three-dimensional structure is currently unknown, therefore, homology modeling was employed to predict a model that could be used for the virtual screening phase. In order to allow a correspondence between the candidate inhibitors predicted and the biophysical evidence obtained in the protein-ligand complex characterization process, only the protein sequence of the recombinant protein produced was modeled.

Previously to the present work, Dr. Jayaraman Muthukumaran submitted the protein sequence to the ModWeb⁴⁵ server, which uses a similar approach to model structures by homology as the one used by Phyre2¹³⁴. This methodology also comprises sequence-sequence, sequence-profile and profile-profile searches, in this case for homologues deposited on the PDB⁴⁵. The template structure selected for modeling was gbs0355 from *Streptococcus agalactiae* (PDB code: 3OKZ)¹³⁵, which has a sequence identity of 72% compared to the query sequence. *Streptococcus dysgalactiae* LytR modeled structure is depicted in Figure 3.10, from which 70% of the residues were covered by the chosen template.

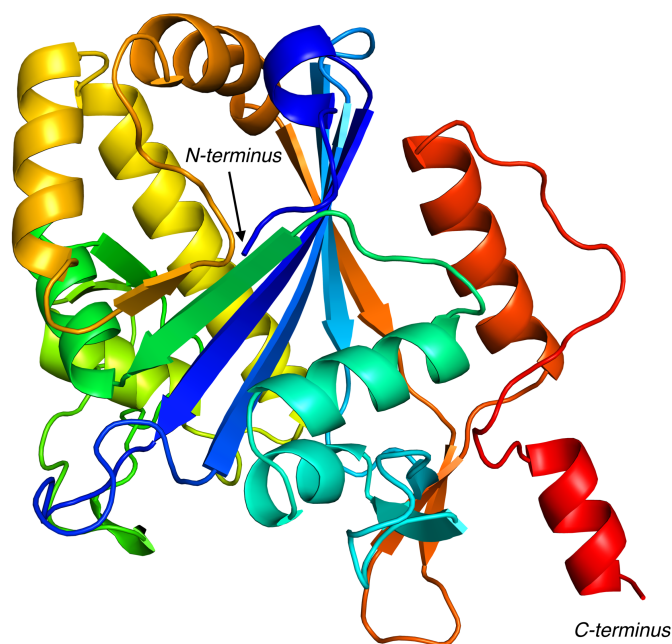


Figure 3.10 - Three-dimensional structure of LytR LCP domain, predicted by homology modeling through the ModWeb⁴⁵ server. Representation was performed using the PyMOL software (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

2.5. Virtual Screening and Molecular Docking

3.2.1. Human Bcl-2 protein

Wang *et al*¹⁰⁷ reported the discovery of a 4H-chromene-based compound with significant affinity for Bcl-2, named HA14-1. This active compound was found through a structure-based virtual screening approach and validated by *in vitro* target- and cell-based techniques. A competitive binding assay, based on fluorescence polarization, entailed the substitution in Bcl-2 binding site of a Bak BH3-derived peptide with a fluorescent probe, by the HA14-1. The binding affinity of HA14-1 for Bcl-2 was determined by its IC₅₀, corresponding to approximately 9 μM. The capability of the active compound to induce tumor death was assessed through a cell viability assay. HL-60 cells were incubated with the HA14-1 and cell viability was found to decrease in a concentration-dependent manner, with more than 90% of cells losing viability at 50 μM of compound. These studies revealed a promising 4H-chromene-based compound which could be used as a reference for a hit-to-lead optimization process that yields Bcl-2 inhibitors with improved binding affinities.

Recently, Parthiban and collaborators¹⁰⁸ explored the 4H-chromene family of compounds for cancer treatment, through Bcl-2 inhibition. Cell-based assays against two different cancer cell-lines were performed to evaluate the 4H-chromene derivatives' induction of apoptosis. Two compounds, a methoxy and an ethoxy derivatives, were found to have considerable activity against the tumor cells. The methoxy derivative showed anti-proliferative potential against HeLa (cervical cancer) and HEP-2 (epidermoid laryngeal carcinoma) cells, with IC₅₀ values of 115 and 86 μM, respectively. The ethoxy derivative had a more significant activity against both cell-lines, with IC₅₀ values of 18 μM for HeLa and 25 μM for HEP-2. Docking studies of these compounds targeting Bcl-2 suggested ligand binding in the physiological BH3 pocket, similarly to the reference molecule HA14-1. Further studies, also conducted by Parthiban *et al*¹⁰⁹, revealed a third 4H-chromene-based compound with improved affinity for Bcl-2 and anti-proliferative properties. This active compound, named activemethine derivative, exhibited extensive activity against four different tumor cell-lines, with IC₅₀ below 1 μM. Similar to the previously reported derivatives, the activemethine derivative showed BH3 mimetic behaviour, in docking studies.

Molecular docking studies were performed to corroborate the reported information regarding the 4H-chromene derivatives binding to Bcl-2 protein, except for the methoxy derivative. AutoDock Vina¹³⁶ program was used for the virtual screening approach, therefore, it was desirable to predict these compounds binding using the same software. However, this program encompasses several simplifications in order to allow faster computation times. To achieve more realistic and significant results, AutoDock 4⁵⁹ was also used. The best ligand pose from each molecule is depicted on Figure 3.11 and binding free energies (ΔG), with corresponding inhibitory constants (K_i) are presented in Table 3.1. Venetoclax is also presented, serving as a reference, because of its picomolar binding affinity (<0.01 nM) to Bcl-2 and extensive selectivity compared to Bcl-XL.

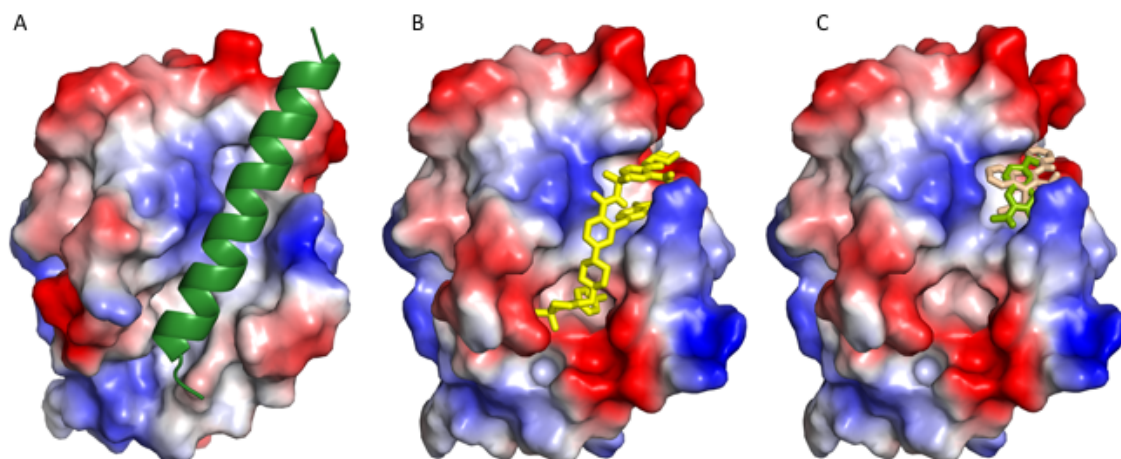


Figure 3.11 - Representation of each ligand best pose from the molecular docking results. A. Crystal structure of Bcl-2 (electrostatic surface) complexed with Bax BH3 domain (in green), PDB code: 2XA0⁹⁷. B. Docking of Bcl-2 (electrostatic surface) with venetoclax (in yellow). C. Docking of Bcl-2 (electrostatic surface) with ethoxy (in light green) and activemethine (in gold) 4H-chomene derivatives. Representations were performed using the PyMOL software (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

Table 3.1 – Docking results of Bcl-2 and the 4H-chromene derivatives, including venetoclax as a reference. All values were predicted by AutoDock⁵⁹ and AutoDock Vina¹³⁶ except for the AutoDock Vina's K_i , which was calculated using Boltzmann equation, from the ΔG prediction.

Compound	AutoDock ⁵⁹ ΔG ; K_i	AutoDock Vina ¹³⁶ ΔG ; K_i
Venetoclax (ABT-199)	- 15.7 kcal/mol ; 3.3 pM	- 11.2 kcal/mol ; 6.2 nM
HA14-1	- 7.1 kcal/mol ; 6.4 μ M	- 5.9 kcal/mol ; 47.3 μ M
Ethoxy derivative	- 6.3 kcal/mol ; 24.4 μ M	- 6.2 kcal/mol ; 28.5 μ M
Activemethine derivative	- 8.8 kcal/mol ; 0.38 μ M	- 7.7 kcal/mol ; 2.27 μ M

As reported by Parthiban *et al*^{108,109}, the activemethine derivative exhibits better binding affinity to Bcl-2 compared to the ethoxy derivative. While the activemethine derivative shows sub-micromolar inhibition of Bcl-2 ($K_i = 0.38 \mu\text{M}$), the ethoxy derivative demonstrates higher K_i value of 24 μM . However, when the activemethine derivative is compared to the drug molecule, venetoclax, its noticeable a significant difference in inhibitory potency, with K_i of 3 pM. This value is in agreement with the time-resolved Förster resonance energy transfer (TR-FRET) binding assay results from Souers *et al*¹⁰⁴, since their method didn't have enough sensibility to detect picomolar interactions, establishing venetoclax K_i toward Bcl-2 as $< 0.01 \text{ nM}$. The difference between the predicted binding energies of the 4H-chromene derivatives and venetoclax is explained by the mapping of protein ligand-interactions (Figure 3.12), obtained with LigPlot¹⁵⁷. In the case of venetoclax, there is an extensive network of protein residues involved in mainly hydrophobic interactions, and also a hydrogen bond with Arg66. Meanwhile, regarding the 4H-chromene derivatives the number of residues contributing for ligand binding is much lower. From

a visual perspective it is also obvious that venetoclax is capable of extending its aromatic chain along Bcl-2 binding pocket, while the 4H-chromene derivatives cover a smaller region of this site. Hence, the elongation of the 4H-chromene derivatives poses as an attractive approach to extend the residue network involved in protein-ligand interactions, increasing their binding free energy toward Bcl-2.

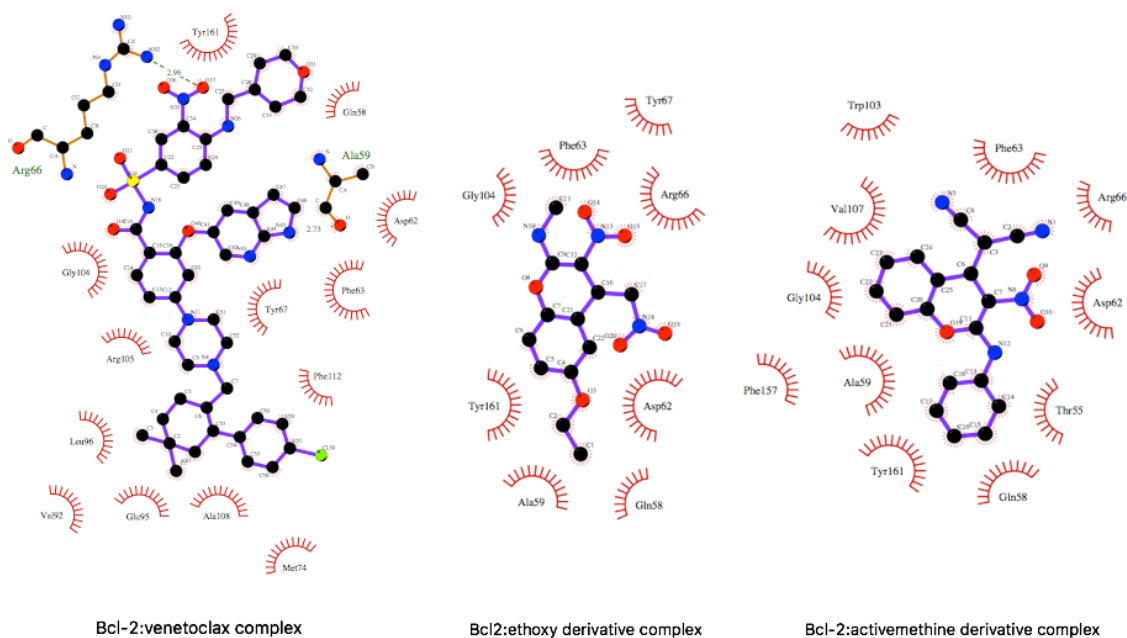
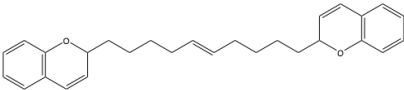
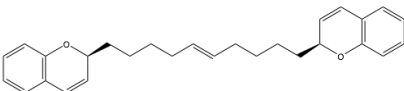
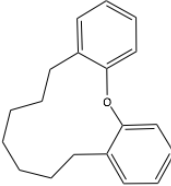
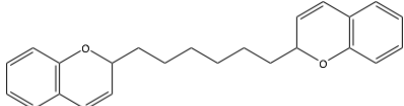
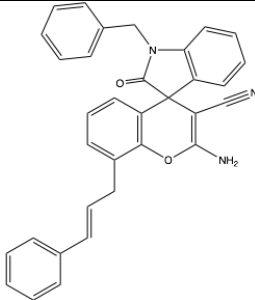
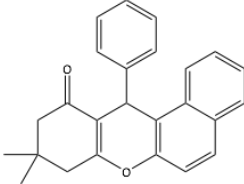
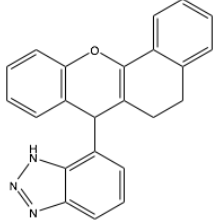


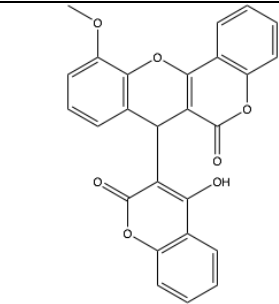
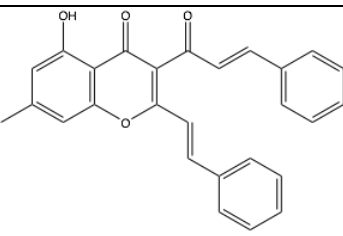
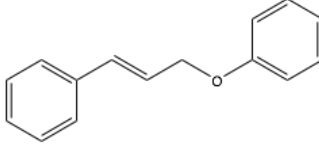
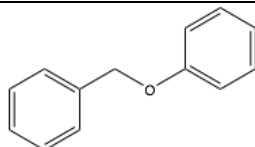
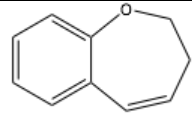
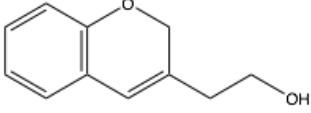
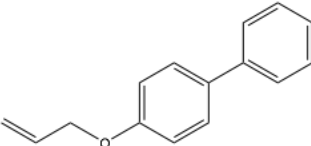
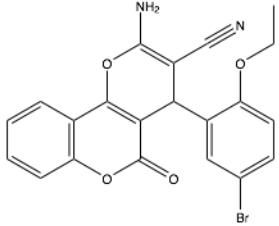
Figure 3.12 - Protein-ligand interaction mapping of the docking results for venetoclax, ethoxy and activemethine derivatives, generated through LigPlot¹⁵⁷. Residues circled in red represent hydrophobic interactions, while the dashed green lines correspond to hydrogen bonds with the ligands' atoms.

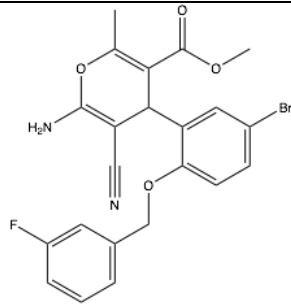
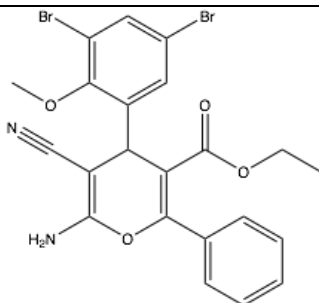
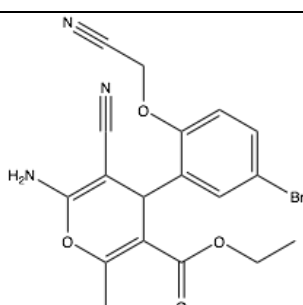
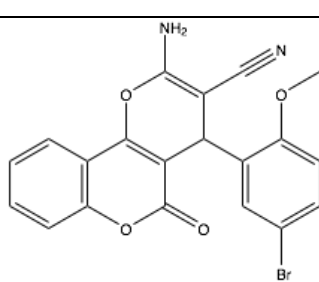
In order to find new potential inhibitor candidates of Bcl-2 with improved binding affinity, a ligand-based virtual screening approach was implemented, using the 4H-chromene family as reference. Thus, the 4H-chromene moiety and the HA14-1 molecule were used in similarity searches against twelve different drug-like molecules libraries. The best hits from each library, which possess at least 90% structure similarity with the query compounds, are presented in the following Table 3.2.

Table 3.2 – List of the best hits from the ligand-based virtual screening of the 4H-chromene moiety and the HA14-1 molecule.

Query compound	Database	Hit ID	Chemical structure	AutoDock Vina ¹³⁶ ΔG (kcal/mol)
4H-chromene moiety	PubChem	69680591		- 9.8

		15274151		- 9.7
		15274152		- 9.6
		57065988		- 9.5
		86009822		- 9.3
	CCDC	KUYCUE		- 10.6
		AJETIT		- 10.2
		QONBAW		- 10.0

		HAKJEK		- 9.7
		MADLOU		- 9.7
	NCI	86646		- 7.1
		77971		- 6.8
		84068		- 6.6
		84595		- 6.5
		86618		- 6.4
HA14-1	PubChem	3311964		- 8.3

		4767155		- 7.4
		5068673		- 7.3
		4529244		- 7.2
		2838790		- 7.1

The compounds that showed improved binding affinities compared to the three 4H-chromene derivatives were from PubChem and CCDC databases, when the 4H-chromene moiety was used for the virtual screening. The best hits demonstrate low binding free energies which correspond to nanomolar affinities toward Bcl-2 and pose as promising active compounds for subsequent studies.

As stated previously, the capability of synthetically grow the aromatic chain of the inhibitor candidates of Bcl-2 is of great interest, as exemplified by the fitting of venetoclax onto the Bcl-2 binding pocket. A collaboration established with an organic synthesis laboratory allowed the exploration of indole derivatives in the Bcl-2 inhibition studies. The drug molecule, venetoclax, was discovered by derivatization of navitoclax, which had selectivity issues toward Bcl-2 compared to the homologue Bcl-XL. This derivatization proved to be essential in increasing drug

specificity, avoiding inhibition of Bcl-XL which is associated with thrombocytopenia. One of the chemical modifications performed was the addition of an azaindole group. A library of available azaindole derivatives was docked to Bcl-2 protein, using AutoDock Vina¹³⁶. The best pose of the best hit is depicted in Figure 3.13A, and the predicted binding free energies of every compound are presented in Table 3.3. In comparison with the 4H-chromene derivative activemethine, the indole molecule reveals similar binding affinity toward Bcl-2, which is associated with the few interacting residues involved (Figure 3.13B). Once again, the more extensive network of residues that interact with venetoclax is pivotal for its stronger binding affinity.

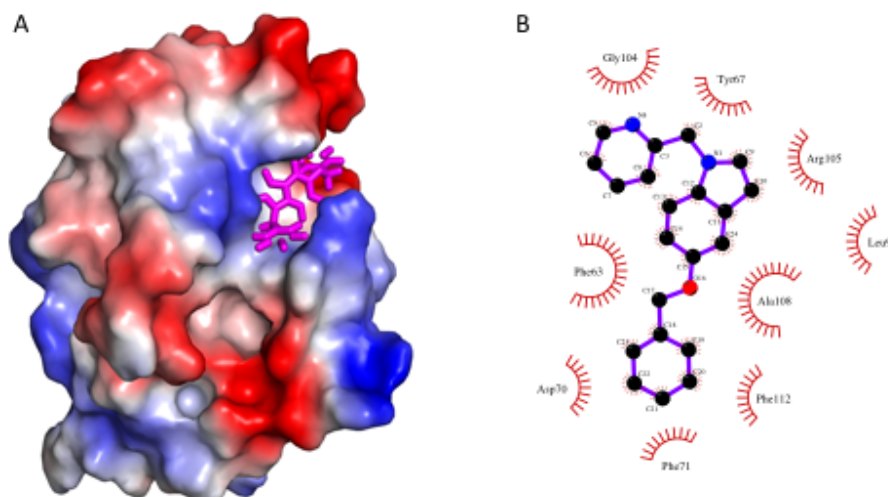
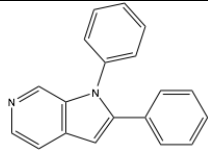
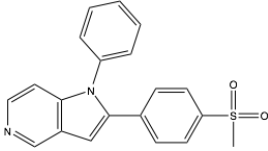
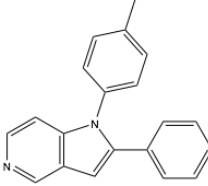
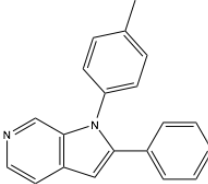
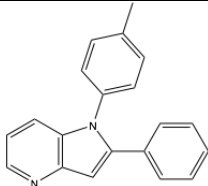
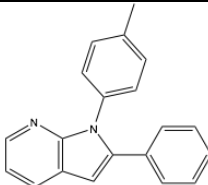
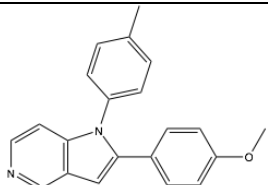
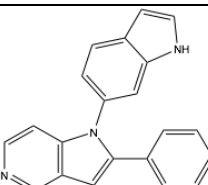
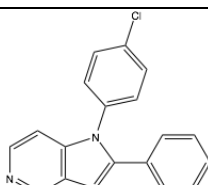
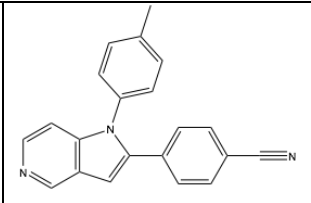
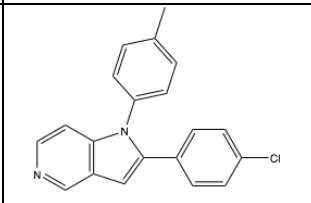
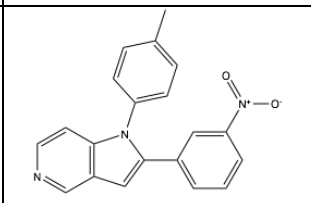
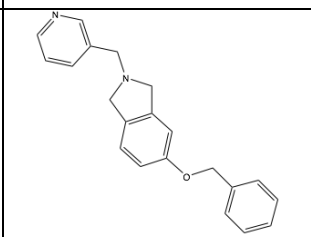


Figure 3.13 - Best pose docked for the indole 15 molecule to Bcl-2. A. Representation, using the PyMOL software (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC), of the three-dimensional complex structure between the Bcl-2 protein (electrostatic surface) and the indole derivative (in magenta). B. Ligplot¹⁵⁷ representation of the interaction network involving the protein residues and the indole molecule atoms. Residues circled in red represent hydrophobic interactions with the ligands' atoms.

Table 3.3 – List of docking results of Bcl-2 with the azaindole derivatives' library, using AutoDock Vina¹³⁶.

Compound ID number	Chemical structure	AutoDock Vina ¹³⁶ ΔG (kcal/mol)
1		- 7.3
2		- 7.2

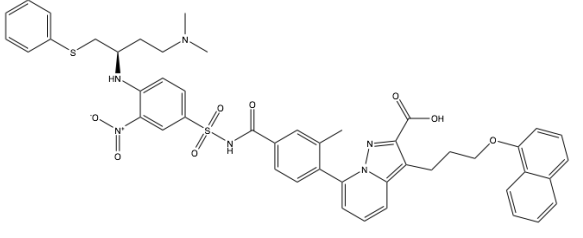
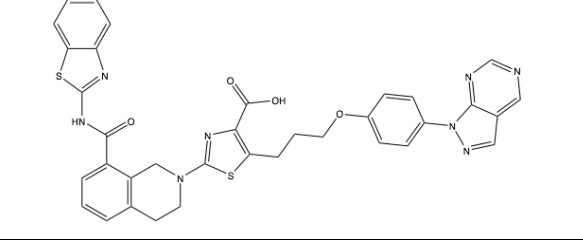
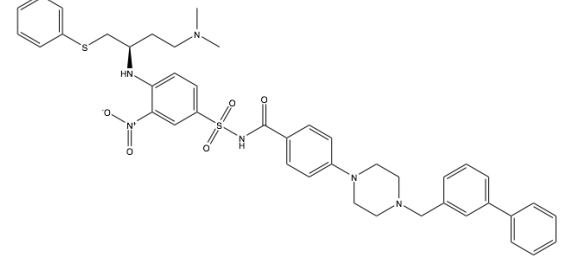
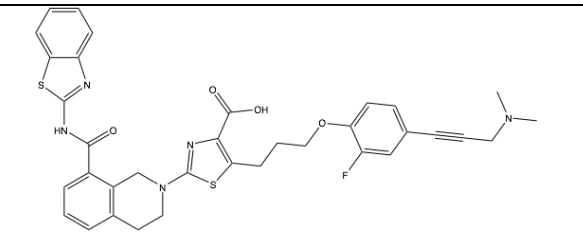
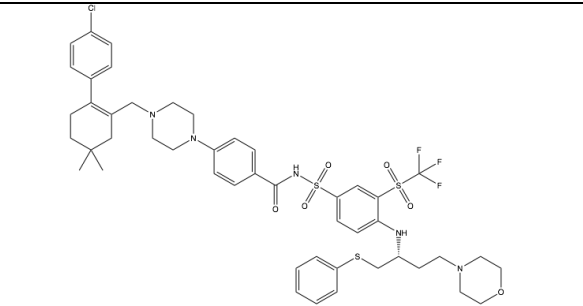
3		- 7.3
4		- 7.7
5		- 7.4
6		- 7.4
7		- 7.3
8		- 7.4
9		- 7.5
10		- 8.5
11		- 7.3

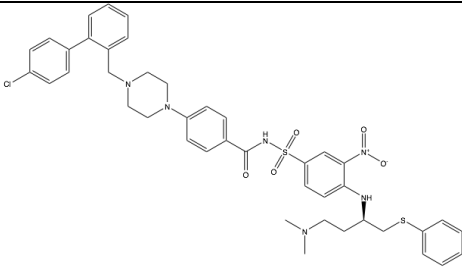
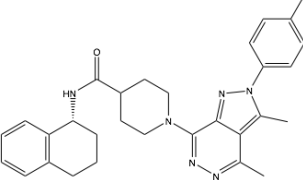
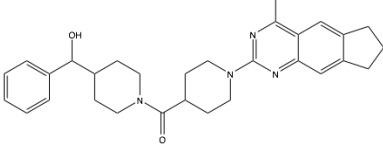
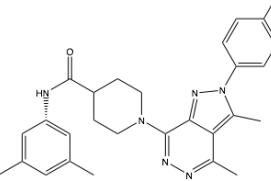
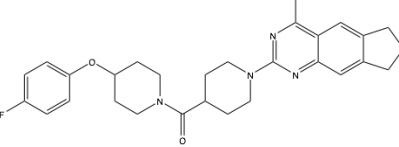
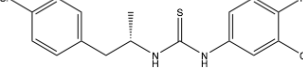
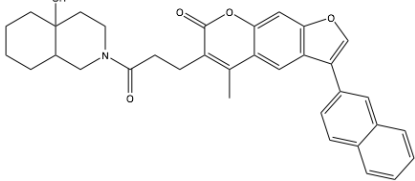
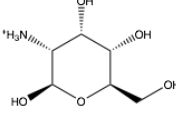
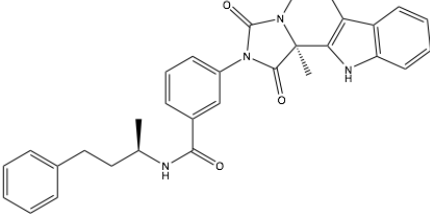
12		- 8.0
13		- 7.8
14		- 7.9
15		- 8.4

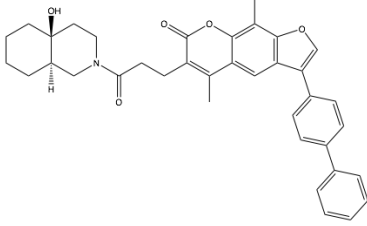
Similarly to the 4H-chromene compounds' results, the indole derivatives docking shows binding in the physiological pocket of Bcl-2, which accommodates its partner Bax protein. The best hit corresponds to the indole derivative 15, which has a predicted binding free energy of - 8.4 kcal/mol, corresponding to an inhibitory constant of 0.54 μ M. This value is very close to the activemethine derivative's K_i , although working with the indole derivative presents several advantages, such as easy access to the compound (the indole collaboration is within our faculty department, while the 4H-chromene is with a group in the Pondicherry University, India), possibility of chemical modification and improved compound solubility.

An alternative approach performed to search for novel potential active compounds was structure-based virtual screening. In this case, the target protein structure is required to dock the libraries' compounds, providing a ranking of the best hits with the lowest binding affinities. Three different databases were probed, from which the best hits are presented in the following Table 3.4.

Table 3.4 – List of docking results from structure-based virtual screening of Bcl-2 against LigSearch, Zinc and Zinc Natural and TCMD databases.

Database	Compound ID	Chemical structure	AutoDock Vina ¹³⁶ ΔG (kcal/mol)
LigSearch (PDB ID)	3WIY		- 10.9
	4TUH		- 10.7
	3INQ		- 10.6
	4QVX		- 10.5
	4LVT		- 10.4

	2YXJ		- 10.3
Zinc	ZINC20564307		- 12.1
	ZINC11881899		- 11.9
	ZINC09642808		- 11.8
	ZINC12201884		- 11.7
Zinc Natural and TCMD	ZINC43060005		- 12.1
	ZINC70687417		- 12.1
	ZINC02380010		- 12.0
	ZINC08918423		- 12.0

	ZINC01900624		- 11.9
--	--------------	--	--------

Comparing these results with the affinities predicted for the ligands under study (4H-chromene and indole derivatives) and for the compounds from the ligand-based virtual screening, the binding free energies are considerably lower. The structure-based virtual screening hits display predicted binding energies corresponding to nanomolar and sub-nanomolar inhibitory constants, which are close to the values from venetoclax (from AutoDock Vina¹³⁶).

It was expected that the structure-based approach would return compounds with improved affinity toward Bcl-2, since in the case of the ligand-based virtual screening, the search is restricted to similar compounds to the query molecules. Hence, the structure-based virtual screening is not limited by the compounds' chemical properties, providing a wider range of possible hits. However, it may comprise the disadvantage of retrieving compounds with undesired chemical and pharmaceutical characteristics, implying an additional filtering effort.

3.2.3. *Streptococcus dysgalactiae* LytR

In 2010, Dürig *et al*¹³⁰ reported a series of active compounds with anti-biofilm properties against *Streptococcus dysgalactiae*. The first promising compound was ellagic acid, which is present in green tea, a drink commonly used in the treatment of infectious diseases. Ellagic acid was shown to effectively reduce biofilm formation in *Streptococcus dysgalactiae*, at low concentrations (4 µg/mL). This compound displayed specificity toward this bacterium, since at the same concentration it didn't inhibit biofilm formation of *Staphylococcus aureus*. A ligand-based virtual screening approach was pursued in order to find novel promising anti-biofilm active compounds. This process entailed a 90% similarity search against the Chinese Natural Product Database, and yielded esculetin. This compound demonstrated improved activity toward *Staphylococcus aureus*, which wasn't affected by the ellagic acid. A second virtual screening was performed, resulting in the discovery of fisetin, which possessed higher activity compared to the previous two compounds. Fisetin showed significant reduction in biofilm formation of both bacteria at 25 µg/mL, while esculetin at the same concentration only affected the biofilm maturation process.

Molecular docking studies were performed for the commercially available anti-biofilm compounds, ellagic acid and fisetin. The goal was to assess whether the compounds interact with *Streptococcus dysgalactiae* LytR protein, which could explain their anti-biofilm properties. The best pose of each molecule is depicted in Figure 3.14 and their respective predicted binding free

energies are presented in Table 3.5. In the case of LytR, both reported lipidic substrates of LCP proteins, decaprenyl- and octaprenyl-phosphate, were docked and used as references.

Table 3.5 – Docking results of LytR with the reference lipidic substrates and the candidate inhibitors under study. Predicted binding free energies were calculated by AutoDock Vina¹³⁶ and the inhibitory constants were derived from Boltzmann equation.

Compound	AutoDock Vina ¹³⁶ ΔG ; K_i
Decaprenyl-phosphate	- 8.5 kcal/mol
Octaprenyl-phosphate	- 6.0 kcal/mol
Ellagic acid	- 6.8 kcal/mol ; 10.4 μ M
Fisetin	- 7.7 kcal/mol ; 2.3 μ M

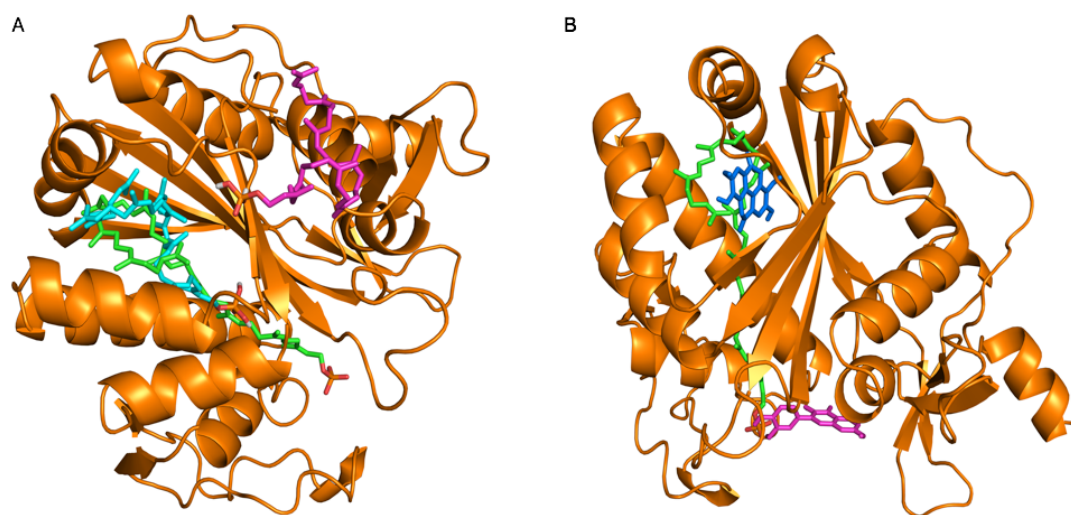
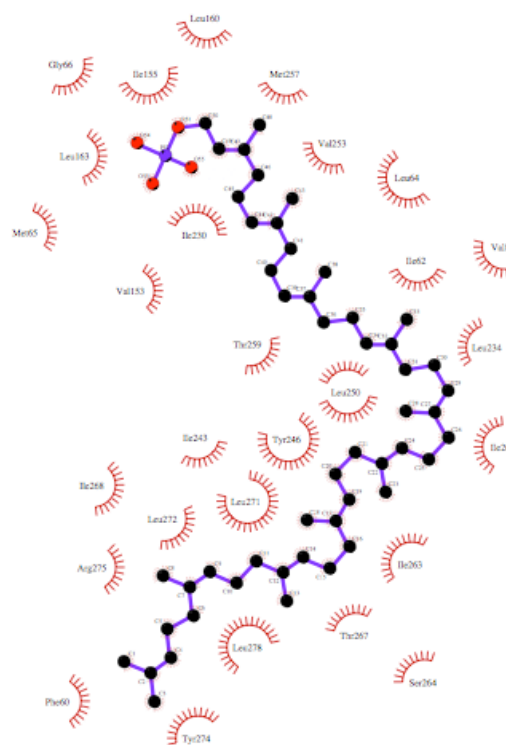


Figure 3.14 – Representation, using the PyMOL software (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC), of the best pose from each ligand on the LytR modeled structure. A. Docking results of LytR (in orange) with the lipidic substrates, decaprenyl-phosphate in blue and octaprenyl-phosphate in magenta. The superimposition of the LCP homologue *gbs0355* from *Streptococcus agalactiae* provided the position of its lipidic substrate (in green) on the LytR structure, allowing the interpretation of a possible binding pocket. B. Docking results of LytR (in orange) with the candidate inhibitors under study, ellagic acid (in blue) and fisetin (in magenta). The *gbs0355* from *Streptococcus agalactiae* superimposed lipidic substrate is included as reference of a possible binding pocket.

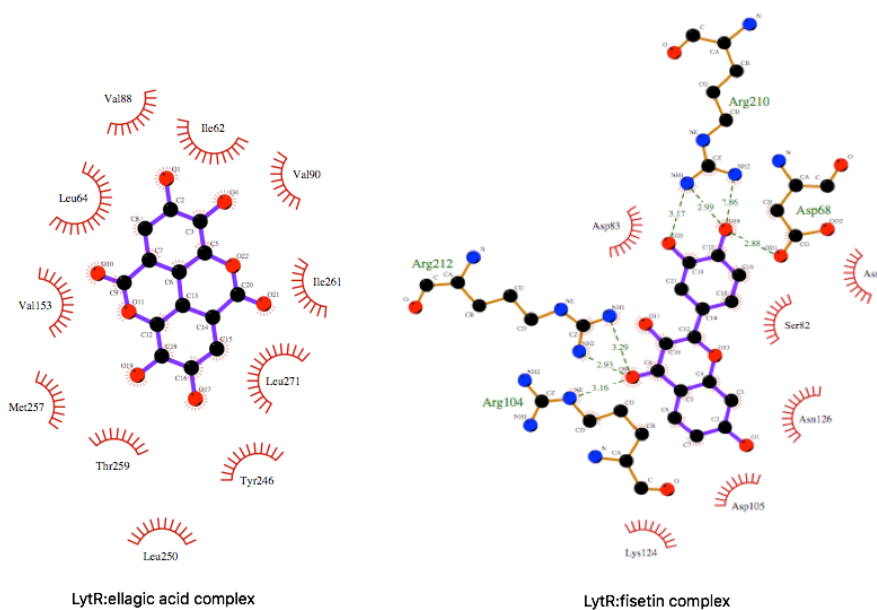
The results suggest that both ellagic acid and fisetin bind close to the hypothetical enzymatic binding pocket of LytR, in which the lipidic substrate, decaprenyl-phosphate, was docked as well. The conserved hydrophobic character of the residues in this pocket in the LCP family support these observations. Previous report indicated that three conserved arginine residues are pivotal in the stabilization of the phosphate head-group of the lipidic substrate¹²⁶, although the mapping of the substrate-interacting residues in the docking results doesn't corroborate that (Figure 3.15). The presence of both active compounds in this pocket and their

interaction with crucial residues for substrate-binding (for example, Arg104, Arg210, Arg212, Leu64 and Val63, Figure 3.16) suggest their inhibitory activity of the hypothetical enzymatic reaction catalysed by LytR. The predicted binding free energies of ellagic acid and fisetin are very similar to the energies of the interactions of LytR with the lipidic molecules. Despite having longer carbon chains and covering a higher protein surface, the lack of flexibility of the lipidic substrates may be unfavourable to allow a more energetically desirable binding. Fisetin, which exhibits better binding affinity compared to the ellagic acid, benefits from a more flexible structure allowing more contacts with protein residues. Therefore, fisetin seems to be the most promising active molecule for inhibiting *Streptococcus dysgalactiae* LytR role in biofilm formation.



LytR:decaprenyl-phosphate complex

Figure 3.15 - Mapping of the LytR residue network interacting with the substrate, decaprenyl-phosphate, using LigPlot¹⁵⁷. Residues circled in red represent hydrophobic interactions, with the ligands' atoms.



LytR:ellagic acid complex

LytR:fisetin complex

Figure 3.16 - Mapping of protein-ligand interactions of LytR with ellagic acid and fisetin, derived from the docking results, using LigPlot¹⁵⁷. Residues circled in red represent hydrophobic interactions, while the dashed green lines correspond to hydrogen bonds with the ligands' atoms.

3.3. Target-ligand complex characterization

3.3.1. Thermal Shift Assay (TSA)

TSA were performed, for both Bcl-2 and LytR, assuming that the existence of binding of the respective small molecule inhibitor candidates would increase protein stability and thus augment their resistance to thermal denaturation. In the case of protein-ligand interactions, a shift in protein melting temperature (T_m) toward higher values should occur, indicating the mentioned increase in stability.

Due to ligands' poor solubility in aqueous solutions, such as protein buffers, it is required the use of organic solvents such as DMSO. However, organic solvents induce protein denaturation by exposing internal hydrophobic amino acids toward the surface, in order to make contact with the organic solvent molecules. Thus, a compromise must be reached to enable both ligand solubility and protein stability. In the TSA experiments performed, DMSO was used to solubilize the ligands under study.

3.3.1.1. Human Bcl-2 protein

Bcl-2 showed poor stability in the purification buffer used during the first stages of the present thesis, resulting in a low yield upon concentration and precipitation at higher concentrations. TSA was used to find a more suitable protein buffer, testing protein stability against a wide screen of buffers (Appendix 6.3). Bcl-2 was found to be more stable in buffers with more basic pH and higher salt concentrations, being the best hit 0.1 M CAPS pH 11 and 500 mM NaCl, with a ΔT_m of 3.99° C. A screen of additives was also tested (Appendix 6.3). However, the ΔT_m observed were inferior to 2° C, which indicated that there weren't any additives that greatly increased protein stability. These results led to the use of a protein buffer containing 50 mM Hepes pH 9.0, 500 mM NaCl, and 1 mM DTT, as reported by Petros and collaborators⁹⁶. In the case of protein crystallization experiments, the reducing agent DTT was replaced by the less volatile and more stable TCEP.

The first approach attempted to observe binding of the ligands under study (4H-chromene) toward Bcl-2, and comprised the testing of increasing concentrations of ligand, which meant ligand:protein molar ratios of 0.5:1, 0.75:1, 1:1, 1.5:1, 2:1, 5:1 and 10:1. To facilitate ligand solubilisation, DMSO concentration was increased in accordance with ligand concentration, corresponding to 0.25, 0.38, 0.50, 0.75, 1, 5, 10 and 25% (v/v).

Bcl-2 candidate inhibitors, including the 4H-chromene derivatives, ethoxy and activemethine, and the indole derivative, didn't display a shift in the protein T_m to higher temperatures, compared to the control sample, *i.e.* the protein without any ligand (Figure 3.17). However, a decrease in T_m is observed, specially with higher concentrations of ligand and, subsequently, higher percentage of DMSO. For the 4H-chromene derivatives, ethoxy and activemethine, the average ΔT_m were - 2.47° and - 2.24° C, respectively. In the case of the indole

derivative, the average ΔT_m was -2.24°C . These results show that the temperature at which Bcl-2 unfolds in the presence of the studied ligands is lower compared to their absence.

The same experimental conditions were applied to control samples, where the effect of each percentage of DMSO in the protein's unfolding was evaluated, in order to establish that the previous results were independent from the presence of organic solvent. The protein's T_m shifts were much less significant (Figure 3.18), the average ΔT_m being 0.18°C , which means that DMSO presence does not affect protein stability extensively, non the chromophore fluorescence. The explanation for the decrease in T_m exhibited at higher ligand concentrations might be that the excess ligand interacts with either or both the protein and the chromophore, resulting in a fluorescence phenomenon at lower temperatures than the protein's actual T_m .

A second approach was designed to allow the calculation of the dissociation constant (K_D) of possible protein-ligand binding. To avoid DMSO interference in the obtained results, its concentration was maintained at 2% (v/v) for all ligand concentrations, which were similar to the ones described in the previous experiment. In this case, venetoclax was also included to enable validation of the procedure.

A clear increase in Bcl-2 T_m was observed upon incubation with venetoclax in close to 1:1 molar ratio (Figure 3.19A). This interaction corroborates the literature and the docking studies performed, since the ΔT_m equal to 18°C demonstrates a strong binding. This result stands also as proof that the experiment allows observation of at least

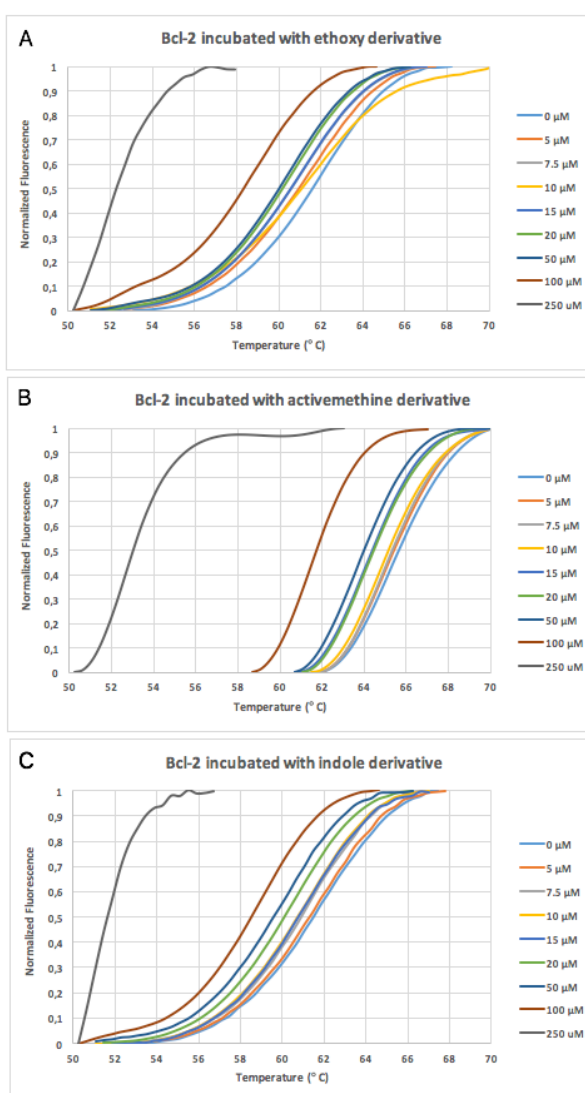


Figure 3.17 - TSA results of Bcl-2 incubated with the 4H-chromene derivatives, ethoxy (A) and activemethine (B), and the indole derivative (C).

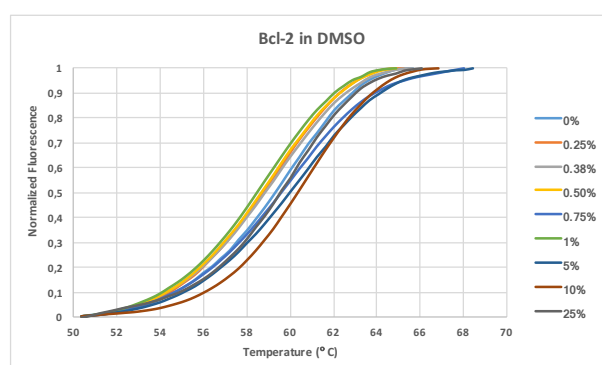


Figure 3.18 - TSA of Bcl-2 in different percentages of DMSO corresponding to the concentrations used in the previous ligand binding experiment.

strong protein-ligand binding. In the case of the indole derivative, which at this stage was the main focus among the ligands under study, no significant alterations in protein stability were observed (Figure 3.19B). Small decreases of T_m were noticed, but these minor variations can't be considered as significant, with an average ΔT_m of - 0.11° C.

K_D determination based on these results was not possible, since in the case of venetoclax there are only two plateau observed and, thus, insufficient data. In the case of the indole derivative there isn't evidence of binding.

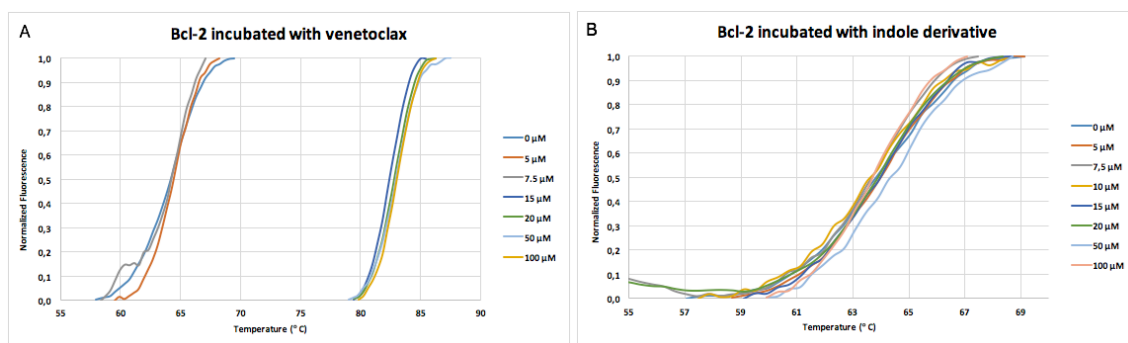


Figure 3.19 - TSA of Bcl-2 incubated with venetoclax (A) and the indole derivative (B) in 2% (v/v) DMSO. In figure A, the concentration of 10 μM of venetoclax is not represented.

A final experiment was carried out to observe possible binding of the indole derivative to Bcl-2, where venetoclax was also used as reference (Figure 3.20). After SEC, while the protein was in a diluted state, the ligand was incubated overnight in a 1:5 protein-ligand ratio, and a final concentration of 5% (v/v) DMSO to facilitate solubilisation. Then, the protein solution was concentrated to 15 mg/mL, allowing the excess ligand and organic solvent to be removed. The final solution should contain protein and ligand in a 1:1 molar ratio, free from DMSO, which allows the exclusion of two possible factors that may interfere with the TSA results.

Once again, a significant T_m shift is observed in the presence of venetoclax, with a similar ΔT_m of 18° C. In the case of the indole derivative, a small T_m shift occurs, however it doesn't seem to be significant enough to prove binding, considering the instrument sensitivity and the approximations implied in the Boltzmann fitting performed by the Protein Thermal Shift™ software. These results are in agreement with the computational studies, where the docking predicted low-binding affinity of the indole derivative for Bcl-2.

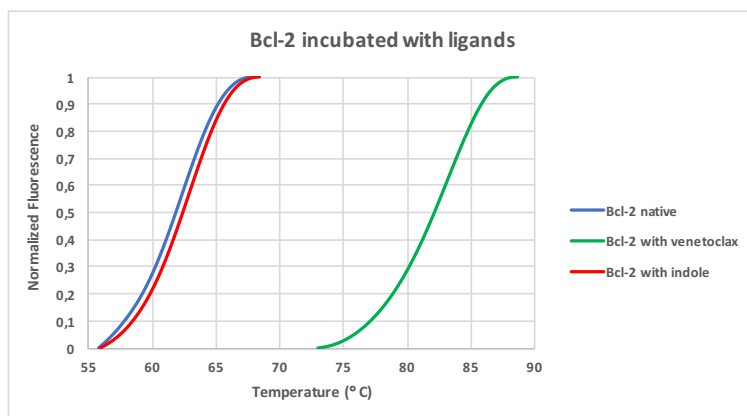


Figure 3.20 - TSA results of Bcl-2 with venetoclax and the indole derivative, where excess ligand and DMSO were removed from the protein solution.

3.3.1.2. *Streptococcus dysgalactiae* LytR

TSA experiments of LytR protein followed the same conditions as the first two approaches in the study of Bcl-2-ligand interaction. Firstly, ellagic acid and fisetin concentrations were varied, regarding the ligand-protein molar ratio (0.5:1, 0.75:1, 1:1, 1.5:1, 2:1, 5:1 and 10:1) with increasing concentration of DMSO (v/v - 0.25, 0.38, 0.50, 0.75, 1, 5, 10 and 25%).

LytR inhibitor candidates didn't show significant variation in protein stability (Figure 3.21), displaying concentration-independent ΔT_m up to 1.21° and 0.40° C, for ellagic acid and fisetin, respectively. These incoherent variations of T_m regarding ligand concentration suggest a lack of sensitivity of the instrument used, thus having no significant meaning. In comparison with the results observed of Bcl-2, LytR demonstrates higher sensitivity toward the presence of DMSO, since at concentrations above 10% (v/v), it is not capable of maintaining its folding.

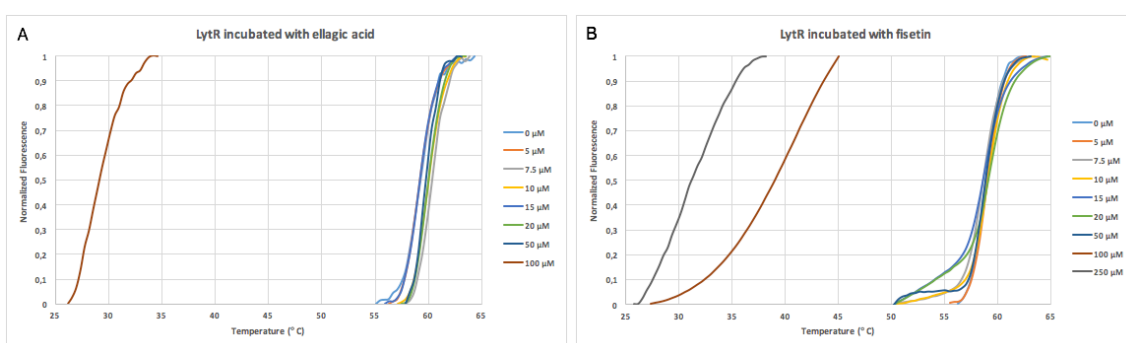


Figure 3.21 - TSA results of LytR incubated with ellagic acid (A) and fisetin (B). In figure A, the 250 µM concentration of ellagic acid is not represented, since LytR was completely unfolded from the beginning of the assay.

In order to avoid subjecting LytR to high concentrations of organic solvent, the second approach used for Bcl-2 TSA was implemented, where DMSO concentration was kept at 2% (v/v) for all ligand concentrations, except 250 µM.

The results demonstrate a small shift in T_m toward higher temperatures in agreement with the increase of ligand concentration (Figure 3.22). However, these variations don't exceed the 2° C accepted for the indication of binding in similar assays⁶⁴, which implicates that it is not possible to confirm ellagic acid and fisetin binding to LytR, through these TSA experiments.

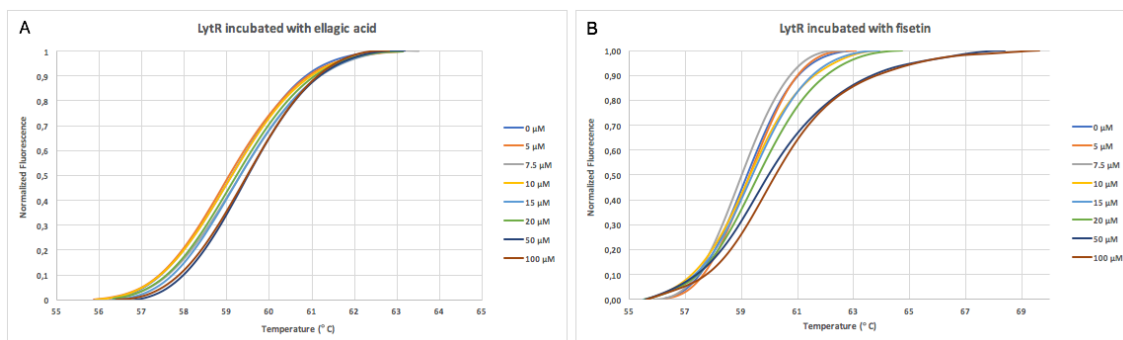


Figure 3.22 - TSA results of LytR incubated with ellagic acid (A) and fisetin (B) in 2% (v/v) DMSO.

3.3.2. Urea-Polyacrylamide Gel Electrophoresis

3.3.2.1. Human Bcl-2 protein

4H-chromene derivatives, ethoxy and activemethine, and the indole derivative were incubated for 1 hour with Bcl-2 in a protein-ligand molar ratio of 1:10. Excess ligand was removed by size exclusion chromatography and samples were loaded into the gel, alongside a sample of native unbound protein.

The electrophoretic mobility of the ligand-free and ligand-bound Bcl-2 was identical (Figure 3.23), suggesting that neither ligand interacted with the protein, altering its conformation to a more compact form. These results confirm the previous TSA experiments, that also failed to indicate significant protein-ligand interactions.

An alternative approach was tested in order to corroborate the previous results, where venetoclax was included to validate the assay. In this case, significant electrophoretic mobility was observed for venetoclax, as expected, however for the indole derivative, the profile remained identical to the unbound Bcl-2 sample (Figure 3.24).

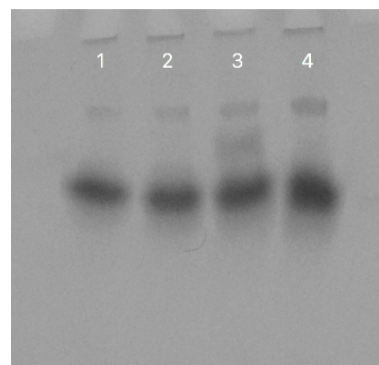


Figure 3.23 - Urea gel showing identical electrophoretic mobility of unbound Bcl-2 (4) and upon incubation with the ethoxy (1), activemethine (2) and indole (3) derivatives.

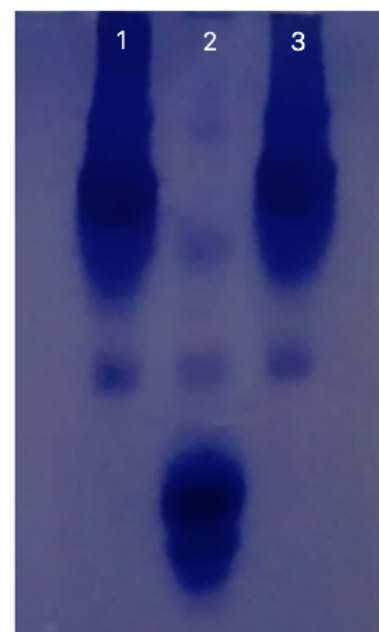


Figure 3.24 - Urea gel of displaying similar mobility for unbound Bcl-2 (1) and upon incubation with the indole derivative (3). Higher mobility is observed when Bcl-2 is incubated with venetoclax (2).

3.3.3. Isothermal Titration Calorimetry (ITC)

ITC experiments were conducted to assess protein-ligand binding and quantify the respective dissociation constants, through the measurement of each reaction's change in heat or enthalpy. For Bcl-2, the ligands tested were venetoclax and the indole derivative. In the case of LytR, only fisetin was used, due to the very poor solubility of ellagic acid. A major issue in these experiments was the strict balance between ligand solubility and protein stability, implying a constant optimization of the concentration of DMSO used.

3.3.3.1. Human Bcl-2 protein

In both cases, with venetoclax and the indole derivative, Bcl-2 was found precipitated at the end of the experiments, due to the presence of 3.5% and 2.5% DMSO, respectively. For the indole reaction, no binding to the protein is observed, since the heat released appears to be related with the dilution effect upon titration (Figure 3.25A). The results with venetoclax suggest a weak binding to Bcl-2 (Figure 3.25B). Despite the higher concentration of organic solvent compared with the indole trial, a clear binding tendency is depicted. However, determination of the reaction's parameters, including K_D , was not possible due to the reduced ligand concentration used in the experiment. Optimization of the experimental conditions is fundamental, although in this case because of venetoclax high hydrophobicity is challenging to find a concentration of DMSO suitable also to maintain protein stability. Attractive alternatives would be the titration of protein into the ligand solution or a competitive assay including the Bax protein or a Bax's BH3 derived peptide.

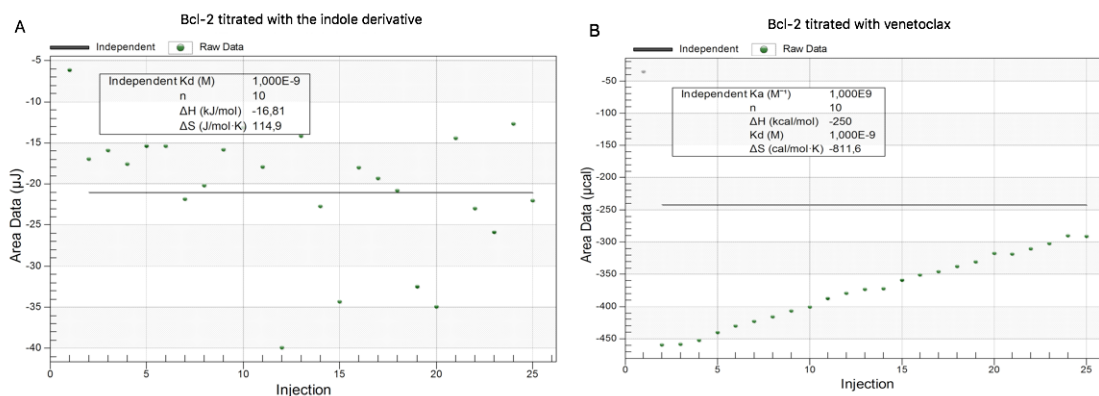


Figure 3.25 - Results of the ITC experiments of Bcl-2 with the indole derivative (A) and venetoclax (B). A non-linear regression fitting of the independent model to the data was attempted unsuccessfully.

3.3.3.2. *Streptococcus dysgalactiae* LytR

LytR titration with fisetin did not reveal protein-ligand interactions (Figure 3.26) and resulted in protein precipitation at the end of the trial, due to the 5% (v/v) DMSO in solution. A weak tendency was noted, so a blank experiment was performed, where protein was not present. Comparing both results, it seems that the ligand dilution during titration is responsible for the

observed small heat changes. These results are in agreement with the TSA experiments, which also failed in demonstrating significant interactions between fisetin and LytR.

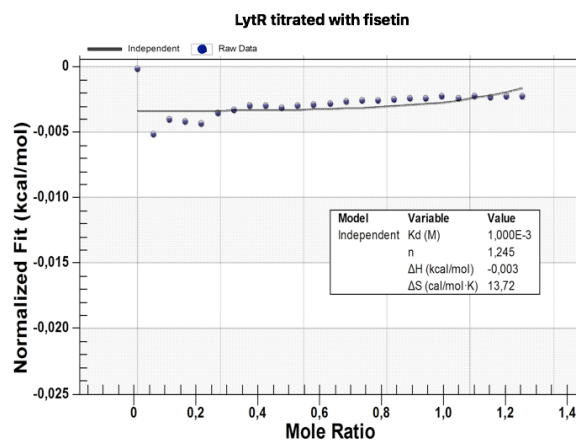


Figure 3.26 - ITC experiment of LytR with fisetin. A non-linear regression fitting of the independent model was attempted unsuccessfully.

3.3.4. Circular Dichroism (CD)

Circular Dichroism allows the observation of variations in protein secondary structure, since at certain wavelengths there is a distinct absorption of polarized light by α -helices, β -sheets and random coils. This technique might add some insight regarding ligand binding to proteins that involve conformational changes through interactions with residues located in specific pockets. This could be the case of LytR interactions with fisetin and ellagic acid. Since ellagic acid is poorly soluble in DMSO, only fisetin's possible binding to LytR was tested through CD.

3.3.4.1. *Streptococcus dysgalactiae* LytR

CD spectra were acquired before and after ligand addition to the 0.5 mg/mL protein solution to reach a 1:1 protein-ligand ratio in 0.15% (v/v) DMSO. A spectrum in the absence of DMSO was also measured in order to establish the implications of the used concentration of organic solvent in protein secondary structure.

LytR incubation with fisetin resulted in a decrease in absorption at the characteristic wavelengths of the α -helices (Figure 3.27), when compared to the control spectrum of LytR in 0.15% DMSO. However, there's no evident variation in light absorption in comparison to the native LytR. In fact, an increase in absorption at the α -helices' wavelengths is noticeable only when the protein is in DMSO without the ligand. The combination of these results suggests that DMSO has a possible role in the variation of the protein secondary structure, which is masked upon incubation with fisetin. Thus, a protein-ligand binding may be possible or the ligand molecules in solution interact with DMSO avoiding its interference with LytR structure. However, these marginal variations of absorption upon ligand incubation cannot represent a significant evidence of binding.

Thus, the CD binding assay's results of LytR with fisetin allow similar interpretations as the other techniques employed, TSA and ITC, which also didn't reveal significant protein-ligand interactions.

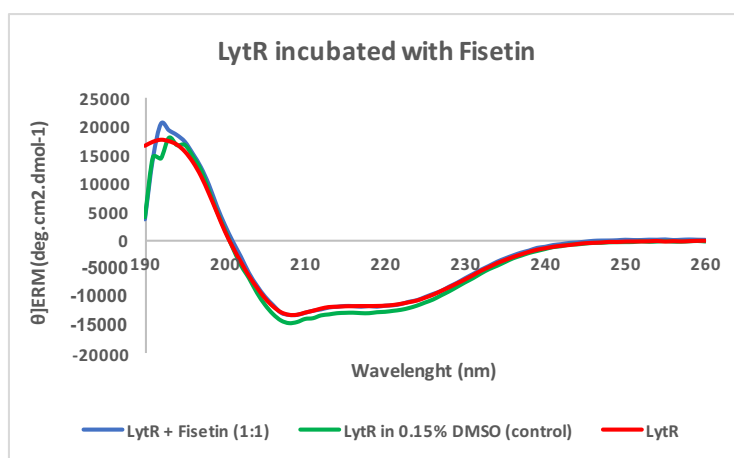


Figure 3.27 - CD spectra of LytR in the absence and presence of fisetin.

3.4. Target-ligand structural determination

After a target hit validation through biophysical methods, structural characterization is pursued in order to unravel with atomic resolution the underlying protein-ligand interactions. Although neither of the techniques previously described was able to reveal significant proof of interaction between the two proteins under study and their respective inhibitor candidates, X-ray crystallography, STD-NMR and molecular dynamics studies were conducted. In the case of Bcl-2 several structures with inhibitors deposited on PDB indicate that co-crystallization experiments can be successful. Furthermore, the availability of venetoclax poses a great opportunity to attempt co-crystallization with Bcl-2, since the co-crystal structures with inhibitors of the same family were already published (PDB code: 4LVT, 4LXD, 4LXE, 4MAN)¹⁰⁴. On the other hand, the LytR structure is currently unknown and the information available on PDB is from homologues of the LCP family (for example, PDB code: 2XXP, 3MEJ)^{124,126}. The unsuccessful crystallization attempts of apo LytR and the realization that lipidic substrates are commonly co-expressed and crystallized in LCP family proteins, which could occur also in LytR, frustrating any interaction with candidate inhibitors, led to the discarding of further structural studies on LytR, except by STD-NMR.

3.4.1. X-ray crystallography

3.4.1.1. Human Bcl-2 protein

Co-crystallization of Bcl-2 with the indole derivative and venetoclax was attempted through a methodology adapted from Perez *et al*¹³¹. Pure protein was incubated with the ligand

at 1:10 protein-ligand molar ratio in 5% DMSO. Then, protein was concentrated to 15 mg/mL, implying the removal of DMSO and excess ligand.

A protein crystal of Bcl-2 previously incubated with the indole derivative was found in a condition, with 0.1 M Tris-HCl pH 8.5, 0.2 M sodium chloride, 5% (v/v) MPD and 5% (v/v) ethanol, from a commercially available screen (Figure 3.28). The crystal was tested in beamline I24, at DLS, and revealed spots at low resolution. This condition was optimized for Bcl-2 and venetoclax co-crystallization, by increasing the precipitant agents' concentrations (MPD and ethanol) and varying the solution's pH. Crystals were found in several conditions at pH 6.0 and between 15-25% (v/v) MPD and 25-35% (v/v) ethanol. These crystals displayed a yellow color (Figure 3.29), which finds correspondence with venetoclax color in powder and solution, suggesting the presence of the ligand in the crystals. The crystal morphology consisted of superimposed thin plates with multiple orientations, which indicates that the crystals' internal order is not consistent, jeopardizing their potential diffraction properties. Data collection of the most promising crystals was performed at beamline ID30A3, in the ESRF, yielding no diffraction.

Preliminary conditions were found for both ligands' co-crystallization with Bcl-2, however further optimization of the crystals packing is required to yield well-ordered crystals capable of diffracting at high resolution. The fact that the protein solution remained colored, greyish for the indole and yellow for venetoclax, after protein concentration suggests that there was binding from the ligands. In the case of venetoclax this is more evident since protein crystals also displayed the yellow color. Several commercially available additive screens were also tested in order to increase protein stability and enable the formation of ordered crystals. However, these attempts weren't successful, since crystals were obtained only with 3-pentanone, displaying similar morphology.

3.4.1.2. *Streptococcus dysgalactiae* LytR

As previously mentioned, the unsuccessful crystallization attempts of apo LytR jeopardized the following co-crystallization experiments with its candidate inhibitors. As described in section 1.4.3.1. of the present thesis' Introduction, there are two available approaches for protein-ligand crystallization, soaking and co-crystallization. Crystallization of apo LytR would provide crucial information regarding both methods, either by offering protein crystals that could be incubated in ligand solutions or by indicating preliminary crystallization conditions that could be optimized for co-crystallization attempts.

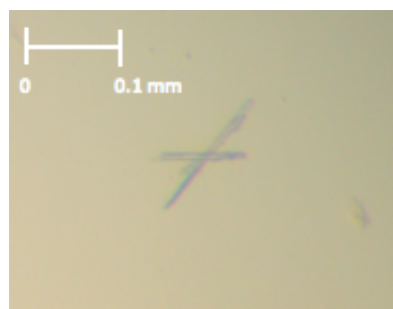


Figure 3.28 - Co-crystals of Bcl-2 with the indole derivative found in a preliminary crystallization condition.

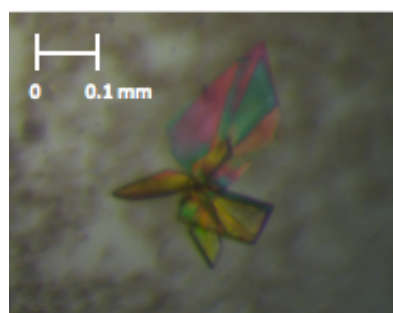


Figure 3.29 - Co-crystal of Bcl-2 with venetoclax.

Despite the mentioned perspectives for protein-ligand structure determination, there are evidences that may prevent protein interactions with candidate inhibitors. PDB entries 2XXP and 3MEJ are from LCP family proteins that contain bound lipidic substrates, which were co-expressed and co-crystallized unknowingly^{124,126}. If the same phenomenon occurs with the recombinant LytR protein under study, which was similarly expressed by an *E. coli* strain, it is plausible that the inhibitor candidates can't interact due to the occupation of the binding pocket by the substrate. In order to confirm the presence of the lipidic substrate, mass spectrometry studies will be pursued which would provide the mass of the molecule present in the sample corresponding either to the apo LytR or to a superior value due to the presence of a ligand^{124,126}.

3.4.2. Saturation Transfer Difference (STD-NMR)

The saturation transfer difference NMR technique can provide different structural information from the obtained by X-ray crystallography, since only the interacting epitope of the ligand is characterized.

A common procedure was adopted from Viegas *et al*⁷⁷ for the binding studies of Bcl-2 and LytR with their respective inhibitor candidates. Prior to the experiments, 1D-¹H spectra of the proteins and the ligands were acquired in order to select a protein frequency to be saturated that didn't include the ligand signal. The ligands were incubated with the protein in 5% (v/v) DMSO to enable ligand solubilisation. STD-NMR spectra were acquired at four different times to assess whether incubation time would influence ligand binding.

3.4.2.1. Human Bcl-2 protein

The indole derivative was the only inhibitor candidate tested for Bcl-2 binding, through STD-NMR. A first approach where protein concentration was 50 μ M and ligand was at 500 μ M was performed. After subtraction of the spectral intensities of the protein solution from the protein-ligand mixture, no signals were found (Figure 3.30). This indicates that there was no saturation transfer phenomenon between the protein's and the ligand's ¹H. A second experiment was attempted where protein and ligand concentrations were reduced to half. The purpose of this change was to decrease the impact of the DMSO on protein stability. However, once again, no interaction was visible in the STD-NMR spectra. These results suggest that either the experimental conditions weren't suitable for protein-ligand interaction detection or the dissociation constant associated is above the limit of detection of the technique, which is 100 mM.

At the time the STD-NMR experiments were performed, venetoclax wasn't available, however, further studies would certainly benefit from the inclusion of venetoclax as a form of validation. The sub-nanomolar affinity of venetoclax for Bcl-2 might represent a hurdle for these studies, since the limit of detection is 0.1 nM.

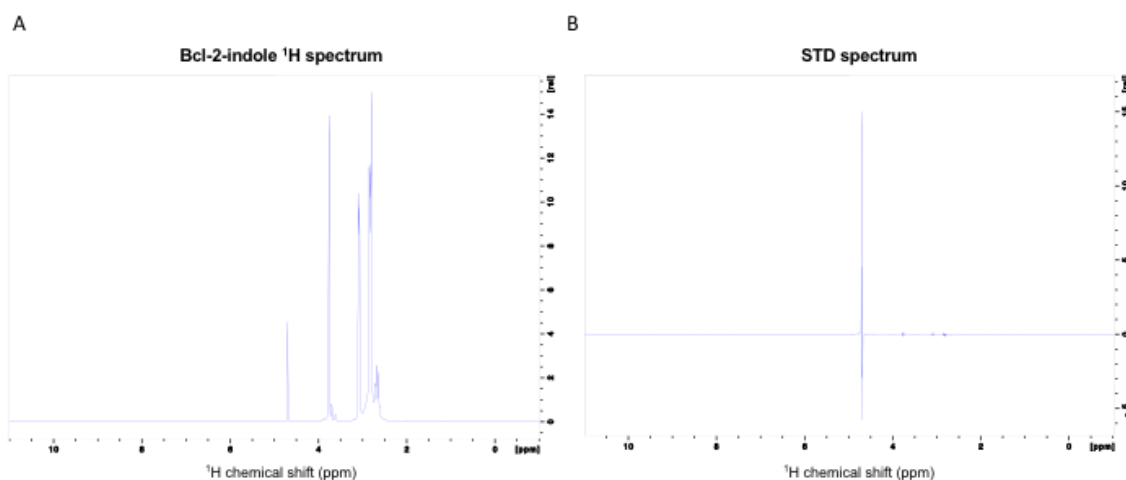


Figure 3.30 - A. ^1H NMR spectrum of Bcl-2 incubated with the indole derivative. B. Difference (STD) spectrum, with 1 second of saturation time, showing no signal from ligand binding. Spectra acquired in a 600 Hz NMR equipped with cryoprobe, at 20°C , with protein at $50\ \mu\text{M}$ and ligand at $500\ \mu\text{M}$.

3.4.2.2. *Streptococcus dysgalactiae* LytR

In the case of LytR's STD-NMR studies, both ligands, ellagic acid and fisetin, were probed for binding. Protein and ligand concentrations were $50\ \mu\text{M}$ and $500\ \mu\text{M}$, respectively, in both experiments. STD-NMR spectra acquired in the experiment with ellagic acid didn't show signal intensity from saturation transfer to the ligand (Figure 3.31). Similarly, the incubation of LytR with fisetin didn't reveal significant protein-ligand interactions (Figure 3.32). As in the case of the Bcl-2 experiment with the indole derivative, these results suggest that either the experimental conditions require further optimization or the protein-ligand binding reaction has a dissociation constant above to the techniques' detection limit.

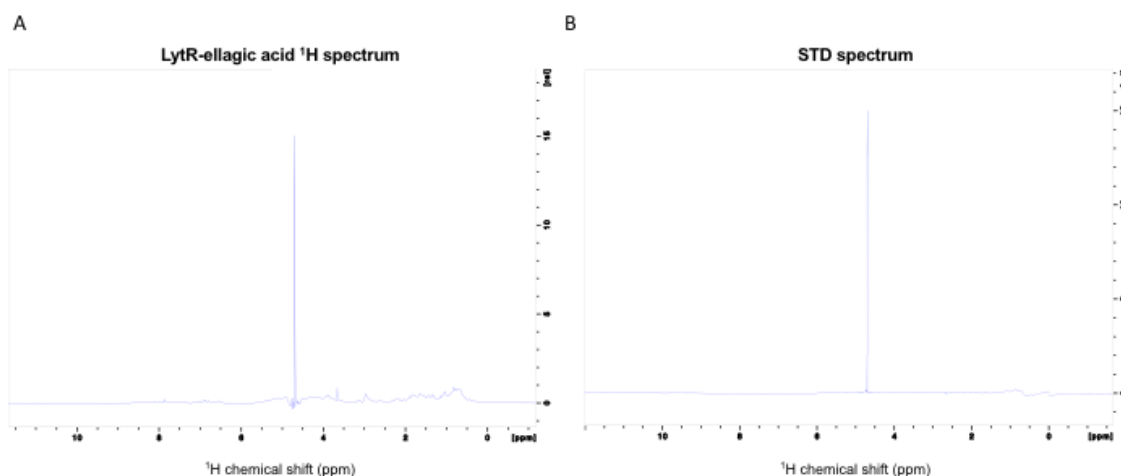


Figure 3.31 - A. ^1H NMR spectrum of LytR incubated with ellagic acid. B. Difference (STD) spectrum, with 1 second of saturation time, showing no signal from ligand binding. Spectra acquired in a 600 Hz NMR equipped with cryoprobe, at 20°C , with protein at $50\ \mu\text{M}$ and ligand at $500\ \mu\text{M}$.

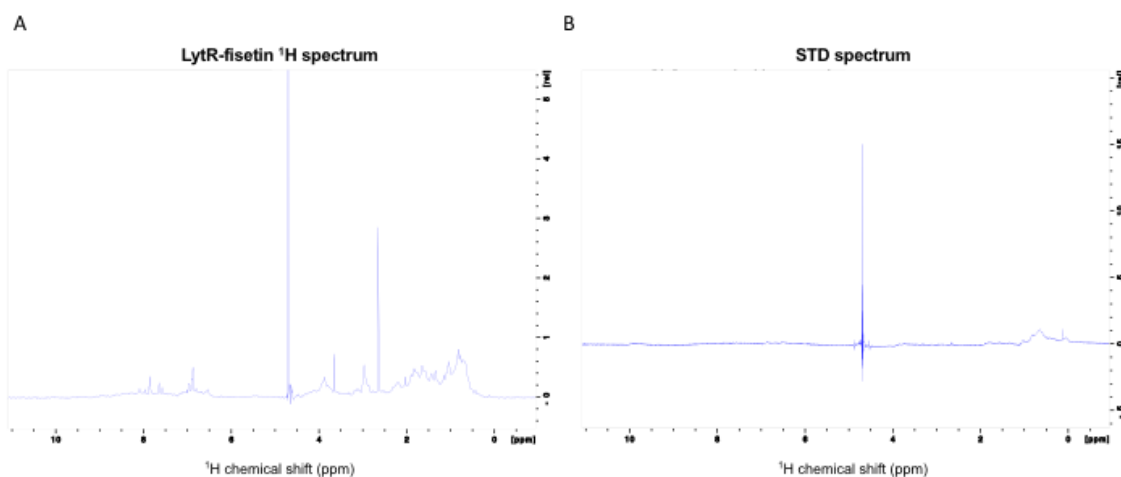


Figure 3.32 - A. ^1H NMR spectrum of LytR incubated with fisetin. B. Difference (STD) spectrum, with 1 second of saturation time, showing no signal from ligand binding. Spectra acquired in a 600 Hz NMR equipped with cryoprobe, at 20°C , with protein at $50\ \mu\text{M}$ and ligand at $500\ \mu\text{M}$.

3.4.3. Molecular Dynamics (MD)

3.4.3.1. Protein-ligand complexes of Human Bcl-2

Docking programs consider the target macromolecule as a rigid entity, which is an unrealistic concept and an artefact to simplify the calculations required. Thus, it is important to study the protein-ligand complex considering the flexibility of the two molecules. Molecular dynamics aims to provide structural information regarding a protein's behaviour in solution and its interactions with other molecules, such as inhibitors. In MD simulations, information such as binding site accessibility and ligand-residue specific interactions can be retrieved.

Following the molecular docking studies of the 4H-chromene ethoxy derivative, the indole derivative, and venetoclax, MD simulations were performed to validate the previously predicted interactions. The results suggest that the compounds interact with the protein, through the 30 ns simulations, in the same regions as predicted in the docking studies (Figure 3.33). Residue interaction probabilities show a similar binding profile for each ligand, the 4H-chromene and the indole derivatives having comparable affinity, while venetoclax has stronger interactions.

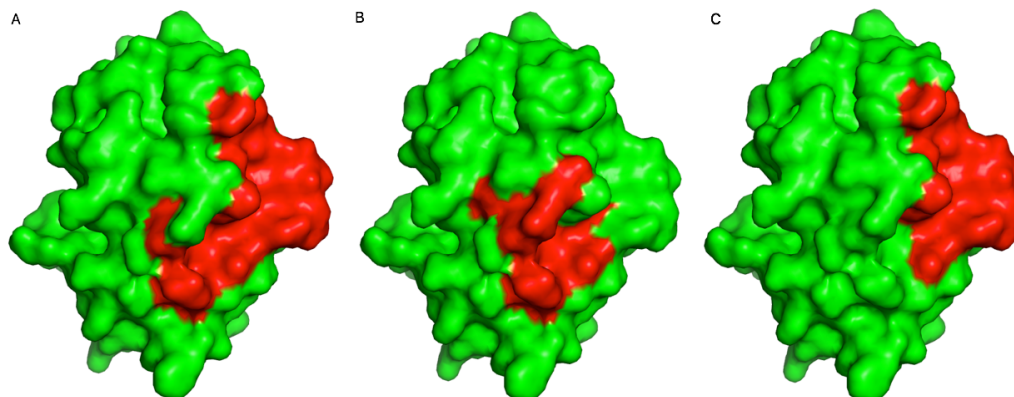


Figure 3.33 - Mapping of the Bcl-2 (surface in green) residue occupancy probability of venetoclax (A), ethoxy (B) and indole (C) derivatives (in red), throughout the MD simulations. Representations were performed using the PyMOL software (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

Parameters like the rmsd indicate that the presence of the compounds enhance protein atom's displacement (Figure 3.34A), which was expected, since ligand accommodation in the binding site implies structural rearrangement of protein residues. Unexpectedly, venetoclax promotes higher protein flexibility than the indole derivative, suggesting that its interactions with many residues implicate the displacement of neighbouring atoms. While in the case of weaker interactions established by the ethoxy and indole derivatives, the shorter length of the ethoxy derivative molecule allows higher fluctuations compared to the indole derivative. The essential dynamics (ED) analysis corroborates these results (Figure 3.34B), since the area in the virtual box experienced by the protein with each ligand varies accordingly (Table 3.6).

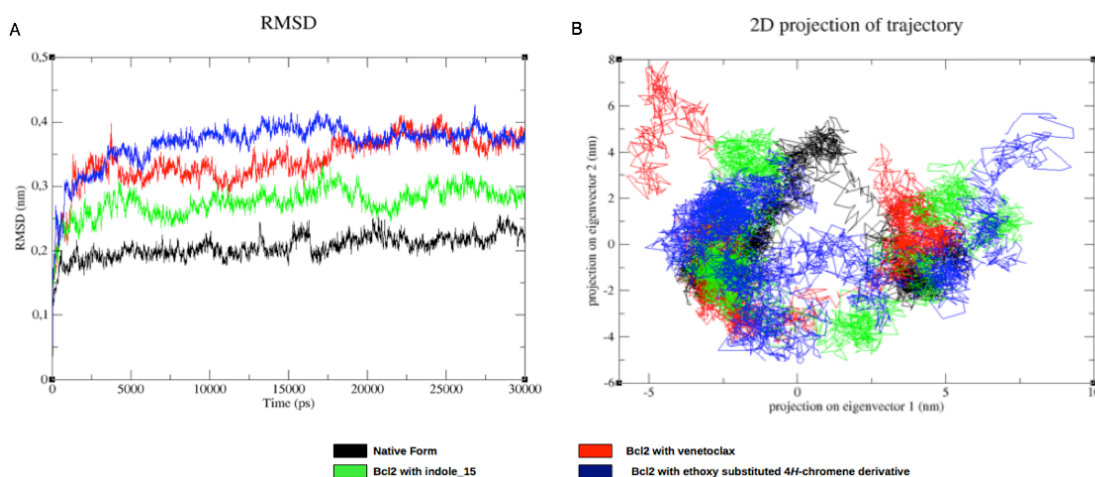


Figure 3.34 - Rmsd (A) and essential dynamics (B) analysis of the MD simulations of native form Bcl-2 and in the presence of the ligands under study (ethoxy and indole derivatives and venetoclax).

The protein stability promoted by each ligand is a consequence of the binding affinity associated with the reaction. Venetoclax, having a stronger interaction with Bcl-2, has a lower potential energy associated, while the indole derivative is followed by the ethoxy derivative (Table 3.6). In the case of the ethoxy and indole derivatives, they have higher potential energies associated when compared with the native unbound Bcl-2. This indicates that those interactions are unfavourable, destabilizing the protein.

Table 3.6 – Potential energy and essential dynamics (ED) analysis of the MD simulations of the Bcl-2 complexes under study.

Bcl-2 complex	Potential energy (kJ/mol)	ED 2D projection (nm ²)
Native form	- 505092	44.64
Ethoxy derivative	- 452348	55.76
Indole derivative	- 502071	51.44
Venetoclax	- 569035	61.48

3.4.3.2. Structural and functional impact of non-synonymous single nucleotide polymorphisms (nsSNP) on Human Bcl-2 protein

Single nucleotide polymorphisms are point-mutations that can result in alterations of a protein sequence if the different codon corresponds to a distinct amino acid (non-synonymous). These modifications can be associated with several diseases, since protein stability and function may be disturbed. Thus, in the case of Bcl-2, which is frequently associated with human cancers, the study of the impact of nsSNP in the protein structure and function is of great interest.

A literature-based search for human Bcl-2 nsSNP against the dbSNP (NCBI) database generated 36 results for the isoform alpha and 3 for the isoform beta. The phenotypic effect of the reported nsSNP was predicted by eight bioinformatic programs. This analysis yielded a list of 12 nsSNP with a predicted deleterious phenotype. The 12 nsSNP are presented in Table 3.7 alongside the stability prediction performed by seven different bioinformatic programs. The predictions' tendency indicates that the nsSNP promote a decrease in stability, resulting in a destabilization of the protein structure.

Table 3.7 – List of Bcl-2 nsSNP with deleterious phenotype prediction. The nsSNP localization in Bcl-2 sequence is described and the correspondent predicted stabilities presented.

nsSNP	Bcl-2 domain	Stability prediction
G8E	BH4	Decrease
D34Y	Disordered	Increase
A43T	Disordered	Decrease
H94P	BH3	Decrease
L97P		Decrease
F104S		Decrease
S105F		Inconclusive
S105P		Decrease
R129C	Disordered	Decrease
G203S	Disordered	Decrease
R207W	Disordered	Decrease
G233S	Disordered	Decrease

50 ns MD simulations of full-length Bcl-2 protein with each nsSNP were performed in order to study the effect of these mutations in protein dynamics. Table 3.8 displays two important parameters for this analysis, the system's potential energy and essential dynamics. Unexpectedly, the potential energy results suggest that the mutations are favourable for protein folding and stabilization, since in comparison with the native form, the values are lower. In the case of the essential dynamics' analysis, the results indicate that the structures with the nsSNP generally are more flexible than the native form. This contradicts the potential energy results,

however Bcl-2 possesses an extremely flexible loop which may interfere with the analysis concerning protein flexibility.

Table 3.8 – List of Bcl-2 nsSNP under study and correspondent potential energy and essential dynamics analysis from the MD simulations.

nsSNP	Potential energy (kJ/mol)	ED 2D projection (nm ²)
Native form	- 739988	184.58
G8E	- 757365	206.09
D34Y	- 736337	159.06
A43T	- 757057	156.19
H94P	- 753942	110.99
L97P	- 754350	210.22
F104S	- 757091	199.99
S105F	- 756523	243.53
S105P	- 754278	155.57
R129C	- 757121	167.43
G203S	- 756615	169.39
R207W	- 756785	166.20
G233S	- 757958	177.83

Regarding the 12 nsSNP under study, it is important to determine their impact in inhibitor binding, since a specific phenotype may interfere with the effectiveness of a standard treatment. Both venetoclax and the indole derivative were docked to the nsSNP Bcl-2 structures using AutoDock Vina¹³⁶. The binding free energies predicted (Table 3.9) show that venetoclax is generally favoured regardless of the presence of the mutations, while the indole derivative marginally loses affinity. Thus, in patients with these mutations, venetoclax should still be considered as an effective drug for treatment.

Table 3.9 – Docking results of Bcl-2 nsSNP with venetoclax and the indole derivative obtained from AutoDock Vina¹³⁶.

nsSNP	Venetoclax ΔG (kcal/mol)	Indole derivative ΔG (kcal/mol)
Native form	- 9.1	- 7.8
G8E	- 9.4	- 7.7
D34Y	- 9.9	- 7.7
A43T	- 10.2	- 7.7
H94P	- 9.0	- 7.7
L97P	- 8.1	- 7.7
F104S	- 10.5	- 7.7
S105F	- 10.5	- 7.7

S105P	- 9.1	- 7.6
R129C	- 9.1	- 7.7
G203S	- 9.2	- 7.7
R207W	- 9.5	- 7.7
G233S	- 9.3	- 7.8

4. Conclusions and future perspectives

Drug discovery has evolved immensely since the first developments in pharmacology in the 19th century. The areas of chemistry and medicine were the major contributors in the initial stages, while, more recently, gene technology, protein science and molecular biology have become pivotal in this process. The advances regarding these areas resulted in the advent of the concept of rational drug design, in the form of structure-based drug design (SBDD). This approach has become the main focus in both academia and pharmaceutical industry for the discovery of novel active compounds. In the past decades, computational methods have influenced the search for promising molecules that modulate a target's activity more efficiently and with fewer costs. Further efforts attempt to characterize through biophysical, biochemical and structural techniques the underlying protein-ligand interactions. The information retrieved is fundamental for the rational design of a lead compound that possesses the therapeutic, biophysical and clinical properties of a successful drug molecule.

In the present thesis, the structure-based drug design method was implemented, associated with the virtual screening approach, in order to find and study promising active compounds for the Human Bcl-2 and the *Streptococcus dysgalactiae* LytR proteins. The SBDD methodology comprised several stages: target structure determination; virtual screening for inhibitor candidates; biophysical characterization of protein-ligand interactions; and structure determination of protein-ligand complexes.

The human Bcl-2 protein expression and purification was successful providing a reasonable yield of 4 mg of protein per liter of culture. This production of pure protein enabled the application of several techniques aiming to determine the target's structure and to characterize biophysically and structurally the protein complexes with promising inhibitor candidates. Structure determination was attempted through X-ray crystallography, however due to the protein's poor stability, no crystals were obtained in the already reported co-crystallization conditions of Bcl-2 with inhibitors.

Homology modeling was implemented to retrieve the Bcl-2 full-length structure, which was pivotal for the molecular dynamics simulations of the impact of nsSNP in protein stability, structure and interactions with ligands. A model was generated using three templates of the human Bcl-2 protein deposited on PDB and through *ab initio* methods, in the case of the loop 2 region, which doesn't have sequence homology with known structures.

The 4H-chromene derivatives, ethoxy and activemethine, revealed micromolar binding affinities toward Bcl-2, 24.4 μM and 0.38 μM respectively, estimated through molecular docking studies. The predicted binding site corresponds to the same physiological binding groove that interacts with Bax's BH3 domain. These observations indicate that the 4H-chromene molecules under study are potential BH3 mimetics and inhibitors of the Bcl-2 protein. Comparing these predictions with the docked reference drug, venetoclax, a significant difference is noticeable regarding the extension of the residue network involved in the protein-ligand interactions. In the case of the 4H-chromene derivatives, a fewer number of residues interact with the compounds, which reflects in lower binding energies of these compounds when compared with venetoclax.

The virtual screening approach was adopted to find novel active compounds for Bcl-2 with improved binding properties at a faster rate. A ligand-based similarity search for 4H-chromene-related molecules against several databases yielded numerous compounds with great predicted affinities for Bcl-2. Furthermore, a library of fifteen compounds from the indole family, which is chemically-related to venetoclax, was docked to Bcl-2 and the most promising binder was the indole derivative 15, with a predicted K_i of 0.54 μM . A structure-based search was performed using the Bcl-2 structure 4AQ3 deposited on the PDB¹³¹, resulted in the finding of several active compounds with improved affinities compared to the ligand-based search.

Biophysical characterization of possible protein-ligand interactions was performed through several techniques: TSA; urea-polyacrylamide gel electrophoresis; and ITC.

Three different approaches were followed in order to assess the binding of the 4H-chromene and/or the indole derivatives toward Bcl-2. However, no significant evidence was found of protein-ligand interactions regarding either family of compounds. The opposite was found for venetoclax, which revealed significant protein T_m shifts of 18° C upon ligand incubation. This fact stands as proof that strong interactions are observable through TSA in the used experimental conditions.

Urea-polyacrylamide gel electrophoresis binding assays showed similar results compared to the TSA for both 4H-chromene and indole derivatives. In this case also, venetoclax displayed strong binding to Bcl-2 which corresponded to a great electrophoretic mobility observed. While the ligands under study showed similar electrophoretic profiles when compared to the native unbound protein.

The ITC experiments corroborated the observations from both TSA and electrophoresis, since the indole derivative doesn't display significant interactions with Bcl-2, while the venetoclax exhibits a noticeable binding affinity. In this case, venetoclax's binding to Bcl-2 is not as significant as in the other techniques, however the inability to generate a binding model results from the non-optimal experimental conditions used. Higher concentrations of venetoclax are required to obtain more significant heat changes, however the poor solubility of the compound in DMSO is an important obstacle. To surpass this hurdle two approaches can be implemented: i) titration of the protein into the ligand solution, which would allow higher ligand concentration and minimize protein denaturation; ii) a competitive assay including the Bax protein or its BH3 domain peptide, which would stabilize the Bcl-2 molecules in solution and prevent their unfolding caused by the presence of DMSO.

Structural characterization of Bcl-2 complexes with the indole derivative and venetoclax was pursued through X-ray crystallography, STD-NMR and MD.

Preliminary co-crystallization conditions were found, containing ethanol and MPD as precipitant agents. Diffraction experiments of the obtained crystals with X-rays using a synchrotron source revealed weak diffraction, as expected by the crystal's morphology. Further optimization of the co-crystallization conditions is required for obtaining well-ordered crystals that can diffract at high resolutions and enable investigation of the protein-ligand interactions at the atomic level. An attractive approach that may provide better crystals is the use of additives, such

as 3-pentanone, with which crystals were already obtained. The use of additives could provide further stability to the protein-ligand complex and facilitate the crystal's nucleation and growth phases.

STD-NMR studies with the 4H-chromene and indole derivatives weren't able to demonstrate evident ligand binding to Bcl-2. No saturation was transferred from the protein to the ligands, which may have different explanations: i) the experimental conditions weren't optimal to detect binding, namely because ligand solubilisation and protein stability were two hurdles encountered; ii) the dissociation constants associated to the protein binding of these ligands is lower than the techniques' detection limit of 100 mM. Future STD-NMR studies should include venetoclax as validation of the procedure, however its extremely high affinity toward Bcl-2 might be below the detection limit of 0.1 nM.

Molecular dynamics simulations corroborated the molecular docking analysis, suggesting similar binding magnitudes of the 4H-chromene and the indole derivatives, while venetoclax displayed stronger binding. The simulations also revealed that the Bcl-2 complexes with the ligands under study have an unfavourable potential energy when compared to the unbound Bcl-2 or to its complex with venetoclax. This observation may explain the failure of the biophysical techniques in providing evidence of protein-ligand interactions with the 4H-chromene and indole derivatives.

The deleterious nsSNP studied unexpectedly improved protein stability, considering their global potential energies. Simulations of the Bcl-2 complexes with venetoclax and the indole derivative revealed that these nsSNP promote venetoclax binding while marginally decreasing the affinity of the indole derivative. Thus, these nsSNP do not jeopardize treatments with the drug venetoclax.

A combined virtual screening and organic synthesis approach could enable ligand modification with promising moieties for Bcl-2 binding. The main focus should be the extension of the aromatic chain of the ligands to cover a larger area of contact with Bcl-2 residues, thus improving their binding affinity. These interactions, as in the case of venetoclax, are mainly hydrophobic and this must be a criterion for the selection of the ligand's chemical groups to modify.

Streptococcus dysgalactiae LytR protein expression and purification was achieved with a yield of 24 mg of protein per liter of culture. Pure protein was employed in crystallization attempts of LytR in order to determine its three-dimensional structure for the molecular docking studies. However, LytR structure determination wasn't successful, and only preliminary conditions were found. Further optimization is required to obtain well-ordered crystals that diffract to high resolution. Thus, homology modeling and *ab initio* methods were employed to obtain a LytR three-dimensional structure. The model was obtained using a template from a LCP family homologue with 72% sequence identity.

Superimposition of LytR with the homology modeling template, gbs0355 from *Streptococcus agalactiae*, revealed a hypothetical binding pocket where the lipidic substrate of the LCP proteins might be accommodated. Molecular docking of LytR with the same lipidic substrate

of gbs0355, decaprenyl-phosphate, showed binding in this identical region, reinforcing this hypothesis. Studies regarding the candidate inhibitors, ellagic acid and fisetin, demonstrated binding close to this pocket. Ellagic acid binds at an internal hydrophobic region, while fisetin binds the three LCP conserved arginine residues pivotal for substrate stabilization. The predicted binding affinities are in the micromolar range, 10.4 μM for ellagic acid and 2.3 μM for fisetin, and with similar magnitudes compared to the lipidic substrates docked. Both ellagic acid and fisetin pose as promising inhibitors of LytR, although, the latter seems to possess greater specificity due to the hydrogen-bonds established with the relevant arginine residues.

The biophysical techniques employed to validate the protein-ligand interactions didn't reveal evidence of binding of either ligand. TSA exhibits marginal protein T_m shifts toward higher temperatures, not significant to suggest ligand binding. The ITC experiment with fisetin corroborated the previous results, since there aren't observable heat changes from ligand binding. Circular dichroism also doesn't reveal definitive proof of protein-ligand interactions, although a minimal difference in light absorption between the native LytR and fisetin incubated spectra is noticeable in the α -helical region. The unsuccessful outcome of these techniques may be related with the co-purification of a lipidic substrate molecule, which also occurred unknowingly with several LCP proteins. This substrate accommodation in the binding pocket may jeopardize the interactions with the candidate inhibitors, hence no evidence of binding is observed. To confirm the presence of the lipid, the technique of mass spectrometry should be considered.

Structure characterization of protein-ligand interactions was attempted through STD-NMR. However, similarly to the previous studies, no evidence was found of either ellagic acid or fisetin binding to LytR. This could be explained by the same reasons why the STD-NMR experiments with Bcl-2 failed to prove binding.

In summary, the structure-based drug design approach enabled the study of candidate inhibitors for Human Bcl-2 and *Streptococcus dysgalactiae* LytR proteins. Neither small molecules' binding was evident in the biophysical characterization stage, where numerous techniques were employed. One of the main challenges concerned the ligand solubility in aqueous solutions, which entailed the use of the organic solvent DMSO. This jeopardizes protein stability and a significant concentration of DMSO can cause protein denaturation and avoid the realization and quantification of binding phenomena. In the case of the Bcl-2 protein, the structural determination of protein-ligand complexes by X-ray crystallography showed promising preliminary results, which should be pursued in the future. This would be an important step in the SBDD approach, since it would allow the confirmation of protein-ligand binding and the rational modification of the ligands to improve binding affinity.

5. Bibliography

1. Drews, J. *In Quest of Tomorrow's Medicines*. (1999).
2. Drews, J. Drug Discovery : A Historical Perspective. *Science (80-)*. **287**, 1960–1964 (2000).
3. Hudson, J. *The History of Chemistry*. (1992).
4. Himmelweit, F. *The Collected Papers of Paul Ehrlich*. (1960).
5. Sneader, W. *Drug Discovery: A History*. (2005).
6. Lounnas, V. *et al*. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput. Struct. Biotechnol. J.* **6**, e201302011 (2013).
7. Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **10**, 787–797 (2003).
8. Roberts, N. A. *et al*. Rational Design of Peptide-Based HIV Proteinase Inhibitors. *Science (80-)*. **248**, 358–361 (1990).
9. Erickson, J. *et al*. Design, activity and 2.8 Å crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. *Science (80-)*. **249**, 527–533 (1990).
10. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).
11. Kurosawa, G. *et al*. Comprehensive screening for antigens overexpressed on carcinomas via isolation of human mAbs that may be therapeutic. *PNAS* **105**, 7287–7292 (2008).
12. Yang, Y., Adelstein, S. J. & Kassis, A. I. Target discovery from data mining approaches. *Drug Discov. Today* **14**, 147–154 (2009).
13. Abell, A. N. *et al*. Ablation of MEKK4 Kinase Activity Causes Neurulation and Skeletal Patterning Defects in the Mouse Embryo. *Mol. Cell. Biol.* **25**, 8948–8959 (2005).
14. Castanotto, D. & Rossi, J. J. The promises and pitfalls of RNA-interference-based techniques. *Nature* **457**, 426–433 (2009).
15. Fox, S. *et al*. High-Throughput Screening: Update on Practices and Success. *J. Biomol. Screen.* **11**, 864–869 (2006).
16. Noah, J. New developments and emerging trends in high-throughput screening methods for lead compound identification. *Int. J. High Throughput Screen.* **2010**, 141–149 (2010).
17. Mayr, L. M. & Fuerst, P. The Future of High-Throughput Screening. *J. Biomol. Screen.* **13**, 443–448 (2008).
18. Law, R. *et al*. The multiple roles of computational chemistry in fragment-based drug design. *J Comput Aided Mol Des* **23**, 459–473 (2009).
19. Boppana, K., Dubey, P. K., Jagarlapudi, S. A. R. P., Vadivelan, S. & Rambabu, G. Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models. *Eur. J. Med. Chem.* **44**, 3584–3590 (2009).
20. Jacoby, E. *et al*. in *Chemogenics: Methods in Molecular Biology* **575**, 320 (2009).
21. Lionta, E., Spyrou, G., Vassilatis, D. K. & Cournia, Z. Structure-Based Virtual Screening for Drug Discovery : Principles , Applications and Recent Advances. *Curr. Top. Med. Chem.* **14**, 1923–1938 (2014).
22. Moore, K. & Rees, S. Cell-Based Versus Isolated Target Screening: How Lucky Do You Fell? *J.*

- Biomol. Screen.* **6**, 69–74 (2001).
23. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
 24. Nienhaus, G. U. *Protein-Ligand Interactions: Methods and Applications*. (2005).
 25. Du, X. *et al.* Insights into Protein-Ligand Interactions : Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **17**, 1–34 (2016).
 26. Stauch, B. H. *Methods for the Investigation of Protein-Ligand Complexes*. (2013).
 27. Grøftehauge, M. K., Hajizadeh, N. R., Marcus, J. & Pohl, E. Protein – ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr.* **D71**, 36–44 (2015).
 28. Meyer, B. & Peters, T. NMR Spectroscopy of Proteins: NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors *Angewandte. Angew. Chem. Int. Ed.* **42**, 864–890 (2003).
 29. Freyer, M. W. & Lewis, E. A. in *Methods in Cell Biology* **84**, 79–113 (2008).
 30. Tuukkanen, A. T. & Svergun, D. I. Weak protein – ligand interactions studied by small-angle X-ray scattering. *FEBS J.* **281**, 1974–1987 (2014).
 31. Evans, B. R. W. & Williams, J. Studies of the Binding of Different Iron Donors to Human Serum Transferrin and Isolation of Iron-Binding Fragments from the N- and C-Terminal Regions of the Protein. *Biochem J* **173**, 543–552 (1978).
 32. Obach, R. S., Huynh, P., Allen, M. C. & Beedham, C. Human Liver Aldehyde Oxidase: Inhibition by 239 Drugs. *J Clin Pharmacol* **44**, 7–19 (2004).
 33. Coelho, C. *et al.* Structural insights into xenobiotic and inhibitor binding to human aldehyde oxidase. *Nat. Chem. Biol.* **11**, 779–783 (2015).
 34. Mortelmans, K. & Zeiger, E. The Ames Salmonella / microsome mutagenicity assay. *Mutat. Res.* **455**, 29–60 (2000).
 35. Roux, S., Sablé, E. & Porsolt, R. D. in *Current Protocols in Pharmacology* **Chapter 10**, 23 (2004).
 36. Rhodes, G. *Crystallography Made Crystal Clear*. (Elsevier, 2006).
 37. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr.* **F70**, 2–20 (2014).
 38. Bergfors, T. M. *Protein Crystallization (IUL Bioechnology Series)*. (1999).
 39. Chao, W. A., Mess, K. H., Tigner, M. & Zimmermann, F. *Handbook of Accelerated Physics and Engineering*. (2013).
 40. Carvalho, A. L., Trincão, J. & Romão, M. J. in *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols* **572**, 297 (2010).
 41. Boggon, T. J. & Shapiro, L. Screening for phasing atoms in protein crystallography. *Structure* **8**, R143–R149 (2000).
 42. Koonin, E. V., Wolf, Y. I. & Karev, G. P. The structure of the protein universe and genome evolution.

- Nature* **420**, 218–223 (2002).
43. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **12**, 85–94 (1999).
 44. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
 45. Pieper, U. *et al.* ModBase, a database of annotated comparative protein structure models, and associated resources. *Nuc* **39**, D465–D474 (2011).
 46. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
 47. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
 48. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. **16**, 404–405 (2000).
 49. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
 50. Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
 51. Canutescu, A. A. & Dunbrack, R. L. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci.* **12**, 963–972 (2003).
 52. Xie, W. & Sahinidis, N. V. Residue-rotamer-reduction algorithm for the protein side-chain conformation problem. *Bioinformatics* **22**, 188–194 (2006).
 53. Jefferys, B. R., Kelley, L. A. & Sternberg, M. J. E. Protein Folding Requires Crowd Control in a Simulated Cell. *J. Mol. Biol.* **397**, 1329–1338 (2010).
 54. Rotkiewicz, P. & Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J Comput Chem* **29**, 1460–1465 (2008).
 55. Kortagere, S. *In Silico Models for Drug Discovery.* (2013).
 56. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
 57. Rush, T., Grant, J. A., Mosyak, L. & Nicholls, L. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J Med Chem* **48**, 1489–1495 (2005).
 58. McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **11**, 494–502 (2007).
 59. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility Garrett. *J. Comput. Chem.* **30**, 2785–2791 (2009).
 60. Jones, G., Willett, P., Glen, R. & Al, E. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727–748 (1997).
 61. Vivoli, M., Novak, H. R., Littlechild, J. A. & Harmer, N. J. Determination of Protein-ligand Interactions Using Differential Scanning Fluorimetry. *J. Vis. Exp.* 1–13 (2014). doi:10.3791/51809
 62. Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J. & Weiss, M. S. Optimization of

- protein buffer cocktails using Thermofluor. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69**, 209–214 (2013).
63. Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. protein Sci.* **79**, 28.9.1-14 (2015).
 64. Luan, C.-H., Light, S., Dunne, S. F. & Anderson, W. F. in *Structural Genomics and Drug Discovery* 344 (2014). doi:10.1007/978-1-4939-0354-2
 65. Iyer, D. *et al.* Identification of a novel BCL2-specific inhibitor that binds predominantly to the BH1 domain. *FEBS J.* **283**, 3408–3437 (2016).
 66. Makey, D. G. & Seal, U. S. The detection of four molecular forms of human transferrin during the iron binding process. *Biochim. Biophys. Acta* **453**, 250–256 (1976).
 67. Mehtab, S. *et al.* Interaction of vanadium (IV) with human serum apo-transferrin. *J. Inorg. Biochem.* **121**, 187–195 (2013).
 68. Costa, J. *et al.* New insights on vanadium binding to human serum transferrin. *Inorganica Chim. Acta* **420**, 60–68 (2014).
 69. Otrelo-cardoso, A. R. *et al.* Highly selective tungstate transporter protein TupA from *Desulfovibrio alaskensis* G20. *Sci. Rep.* **7**, 5798 (2017).
 70. Ladbury, J. E. Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochem. Soc. Trans.* **38**, 888–893 (2010).
 71. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–139 (2005).
 72. Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673 (2004).
 73. Whitmore, L. & Wallace, B. A. Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. *Biopolymers* **89**, 392–400 (2007).
 74. Mathews, B. W. Solvent Content of Protein. *J Mol Biol* **33**, 491–497 (1968).
 75. Danley, D. E. Crystallization to obtain protein–ligand complexes for structure-aided drug design. *Acta Crystallogr.* **D62**, 569–575 (2006).
 76. Hassell, A. M. *et al.* Crystallization of protein–ligand complexes research papers. *Acta Crystallogr.* **D63**, 72–79 (2007).
 77. Viegas, A., Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **88**, 990–994 (2011).
 78. Durrant, J. D. & Mccammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **9:71**, (2011).
 79. Chitrala, N. K. & Yeguvapalli, S. Computational Screening and Molecular Dynamic Simulation of Breast Cancer Associated Deleterious Non- Synonymous Single Nucleotide Polymorphisms in TP53. *PLoS One* **9**, e104242 (2014).
 80. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).

81. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
82. Lockshin, R. A. & Zakeri, Z. Cell death in health and disease. *J Cell Mol Med* **11**, 1214–1224 (2007).
83. Fulda, S., Gorman, A. M., Hori, O. & Samali, A. Cellular Stress Responses: Cell Survival and Cell Death. *Internatiol J. Cell Biol.* **2010**, 214074 (2010).
84. Portt, L., Norman, G., Clapp, C., Greenwood, M. & Greenwood, M. T. Anti-apoptosis and cell survival : A review. *Biochim. Biophys. Acta* **1813**, 238–259 (2011).
85. Fukuhara, S. & Rowley, J. D. Chromosome 14 translocations in non-burkitt lymphomas. *Int J Cancer* **22**, 14–21 (1978).
86. Yip, K. W. & Reed, J. C. Bcl-2 family proteins and cancer. *Oncogene* **27**, 6398–6406 (2008).
87. Shamas-din, A., Kale, J., Leber, B. & Andrews, D. W. Mechanisms of Action of Bcl-2 Family Proteins. *Cold Spring Harb Perspect Biol* **5**:a008714, (2013).
88. Delbridge, A. R. D. & Strasser, A. The BCL-2 protein family, BH3-mimetics and cancer therapy. *Cell Death Differ.* **22**, 1071–1080 (2015).
89. Cimmino, A. *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *PNAS* **102**, 13944–13949 (2005).
90. Hanada, B. M., Delia, D., Aiello, A., Stadtmauer, E. & Reed, J. C. BCL-2 Gene Hypomethylation and High-Level Expression in B-Cell Chronic Lymphocytic Leukemia. *Blood* **82**, 1820–1828 (1993).
91. Ikegaki, N., Katsumata, M., Minna, J. & Tsujimoto, Y. Expression of bcl-2 in Small Cell Lung Carcinoma Cells. *Cancer Res.* **54**, 6–8 (1994).
92. Rochaix, P. *et al.* In vivo Patterns of BCL-2 Family Protein Expression in Breast Carcinomas in Relation to Apoptosis. *J. Pathol.* **187**, 410–415 (1999).
93. Fulda, S. Tumor resistance to apoptosis. *Int J Cancer* **124**, 511–515 (2009).
94. Strasser, A., Harris, A. W. & Cory, S. BCL-2 Transgene Inhibits T Cell Death and Perturbs Thymic Self-Censorship. *Cell* **67**, 889–899 (1991).
95. Strasser, A., Harris, A. W., Jacks, T. & Cory, S. DNA Damage Can Induce Apoptosis in Proliferating Lymphoid Cells via p53-Independent Mechanisms Inhibitable by Bcl-2. *Cell* **79**, 329–339 (1994).
96. Petros, A. M. *et al.* Solution structure of the antiapoptotic protein bcl-2. *Proc. Natl. Acad. Sci.* **98**, 3012–3017 (2001).
97. Ku, B., Liang, C., Jung, J. U. & Oh, B. Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Res.* **21**, 627–641 (2011).
98. Besbes, S., Mirshahi, M., Pocard, M. & Billard, C. New dimension in therapeutic targeting of BCL-2 family proteins. *Oncotarget* **6**, 12862–12871 (2015).
99. Cory, S., Roberts, A. W., Colman, P. M. & Adams, J. M. Targeting BCL-2-like Proteins to Kill Cancer Cells. *TRENDS in CANCER* **2**, 443–460 (2016).
100. Correia, C. *et al.* Emerging understanding of Bcl-2 biology: Implications for neoplastic progression and treatment. *Biochim. Biophys. Acta* **1853**, 1658–1671 (2015).
101. Toure, B. B. *et al.* The Role of the Acidity of N-Heteroaryl Sulfonamides as Inhibitors of Bcl-2

- Family Protein–Protein Interactions. *ACS Med. Chem. Lett.* **4**, 186–190 (2013).
102. Oltersdorf, T. *et al.* An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **435**, 677–681 (2005).
 103. Tse, C. *et al.* ABT-263: A Potent and Orally Bioavailable Bcl-2 Family Inhibitor. *Cancer Res.* **68**, 3421–3428 (2008).
 104. Souers, A. J. *et al.* ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **19**, 202–208 (2013).
 105. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
 106. Cang, S., Iragavarapu, C., Savooji, J., Song, Y. & Liu, D. ABT-199 (venetoclax) and BCL-2 inhibitors in clinical development. *J. Hematol. Oncol.* **8:129**, (2015).
 107. Wang, J. *et al.* Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *PNAS* **97**, 7124–7129 (2000).
 108. Parthiban, A. *et al.* Design , synthesis , molecular docking , and biological evaluation of N -methyl-3-nitro-4- (nitromethyl) -4 H -chromen-2-amine derivatives as potential anti-cancer agents. *Med Chem Res* **23**, 642–659 (2014).
 109. Parthiban, A. *et al.* Design , synthesis , in vitro and in silico anti-cancer activity of 4 H -chromenes with C4-active methine groups. *Med Chem Res* **24**, 1226–1240 (2015).
 110. Kamath, P. R., Sunil, D., Ajees, A. A., Pai, K. S. R. & Das, S. Some new indole-coumarin hybrids ; Synthesis, anticancer and Bcl-2 docking studies. *Bioorg. Chem.* **63**, 101–109 (2015).
 111. Bjarnsholt, T. The Role of Bacterial Biofilms in Chronic Infections. *APMIS* **121**, 1–51 (2013).
 112. Gomes, F., Saavedra, M. J. & Henriques, M. Bovine mastitis disease / pathogenicity : evidence of the potential role of microbial biofilms. *Pathog. Dis.* **74**, (2016).
 113. Donlan, R. M. Role of biofilms in antimicrobial resistance. *ASAIO J* **46**, S47-52 (2000).
 114. Nussler, A. K., Wittel, U. A., Nussler, N. C. & Beger, H. G. Leukocytes, the Janus cells in inflammatory disease Leukocytes. *Langenbeck's Arch Surg* **384**, 222–232 (1999).
 115. Leid, J. G., Shirliff, M. E., Costerton, J. W. & Stoodley, P. Human Leukocytes Adhere to, Penetrate, and Respond to Staphylococcus aureus Biofilms. *Infect. Immun.* **70**, 6339–6345 (2002).
 116. Matz, C. *et al.* Marine Biofilm Bacteria Evade Eukaryotic Predation by Targeted Chemical Defense. *PLoS One* **3**, e2744 (2008).
 117. Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial Biofilms: A Common Cause of Persistent Infections. *Science (80-.)*. **284**, 1318–1322 (1999).
 118. Chénier, S., Leclèner, M., Messier, S. & Fecteau, G. Streptococcus dysgalactiae cellulitis and toxic shock like syndrome in a Brown Swiss cow. *J Vet Diagon Invest* **20**, 99–103 (2008).
 119. Olson, M. E., Ceri, H., Morck, D. W., Buret, A. G. & Read, R. R. Biofilm bacteria: formation and comparative susceptibility to antibiotics. *Can J Vet Res* **66**, 86–92 (2002).
 120. Lacasta, D., Ferrer, L. M., Ramos, J. J., Loste, A. & Bueso, J. P. Digestive pathway of infection in Streptococcus dysgalactiae polyarthritis in lambs. *Small Rumin Res* **78**, 202–205 (2008).
 121. Vela, A. I. *et al.* Neonatal mortality in puppies due to bacteremia by Streptococcus dysgalactiae

- subsp. dysgalactiae. *J Clin Microbiol* **44**, 666–668 (2006).
122. Hagiwara, H., Takano, R., Noguchi, M. & Narita, M. A study of the lesions induced in *Seriola dumerili* by intradermal or intraperitoneal injection of *Streptococcus dysgalactiae*. *J. Comp. Pathol.* **140**, 25–30 (2009).
 123. Koh, T. H. *et al.* Streptococcal cellulitis following preparation of fresh raw seafood. *Zoonoses Public Heal.* **56**, 206–208 (2009).
 124. Kawai, Y. *et al.* A widespread family of bacterial cell wall assembly proteins. *EMBO J.* **30**, 4931–4941 (2011).
 125. Weidenmaier, C. & Peschel, A. Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. *Nat. Rev. Microbiol.* **6**, 276–287 (2008).
 126. Eberhardt, A. *et al.* Attachment of Capsular Polysaccharide to the Cell Wall in *Streptococcus pneumoniae*. *Microb. Drug Resist.* **18**, 240–255 (2012).
 127. Hanson, B. R., Lowe, B. A. & Neely, M. N. Membrane Topology and DNA-Binding Ability of the Streptococcal CpsA Protein. *J. Bacteriol.* **193**, 411–420 (2011).
 128. Lazarevic, V., Margot, P., Soldo, B. & Karamata, D. Sequencing and analysis of the *Bacillus subtilis* lytRABC divergon: a regulatory unit encompassing the structural genes of the N-acetylmuramoyl-L-alanine amidase and its modifier. *J. Gen. Microbiol.* **138**, 1949–1961 (1992).
 129. Schaefer, K., Matano, L. M., Qiao, Y., Kahne, D. & Walker, S. In vitro reconstitution demonstrates the cell wall ligase activity of LCP proteins. *Nat. Chem. Biol.* **13**, 396–401 (2017).
 130. Dürig, A. & Kouskoumvekaki, I. Chemoinformatics-assisted development of new anti-biofilm compounds. *Appl Microbiol Biotechnol* **87**, 309–317 (2010).
 131. Perez, H. L. *et al.* Identification of a phenylacetylsulfonamide series of dual Bcl-2 / Bcl-xL antagonists. *Bioorg. Med. Chem. Lett.* **22**, 3946–3950 (2012).
 132. Rautureau, G. J. P. *et al.* The restricted binding repertoire of Bcl-B leaves Bim as the universal BH3-only prosurvival Bcl-2 protein antagonist. *Cell Death Dis.* **3**, e443-9 (2012).
 133. Jancarik, J. & Kim, S.-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409–411 (1991).
 134. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
 135. Kuzin, A. *et al.* Crystal Structure of protein gbs0355 from *Streptococcus agalactiae*, Northeast Structural Genomics Consortium Target SaR127.
 136. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
 137. Spoel, D. Vander, Lindahl, E., Hess, B. & Groenhof, G. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
 138. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).

139. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. Van & Hermans, J. in *Intermolecular Forces* 331–338 (1981). doi:10.1007/978-94-015-7658-1
140. Koziara, K. B., Stroet, M., Malde, A. K. & Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0 : prediction of hydration free enthalpies. *J Comput Aided Mol Des* **28**, 221–233 (2014).
141. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* **40**, 843–856 (2011).
142. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2 . 0 : predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**, 306–310 (2005).
143. Thomas, P. D. *et al.* PANTHER : A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res.* **13**, 2129–2141 (2003).
144. Capriotti, E. *et al.* WS-SNPs & GO : a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics* **14**, S6 (2013).
145. Ng, P. C. & Henikoff, S. SIFT : predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
146. Choi, Y. & Chan, A. P. Sequence analysis PROVEAN web server : a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
147. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20.1-7.20.41 (2013).
148. Bao, L., Zhou, M. & Cui, Y. nsSNPAnalyzer : identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* **33**, 480–482 (2005).
149. Capriotti, E., Calabrese, R. & Casadio, R. Sequence analysis Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **22**, 2729–2734 (2006).
150. Pires, D. E. V, Ascher, D. B. & Blundell, T. L. mCSM : predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **30**, 335–342 (2014).
151. Worth, C. L., Preissner, R. & Blundell, T. L. SDM - a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **39**, 215–222 (2011).
152. Pires, D. E. V, Ascher, D. B. & Blundell, T. L. DUET : a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* **42**, 314–319 (2014).
153. Cheng, J., Randall, A. & Baldi, P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins Struct. Funct. Bioinforma.* **1132**, 1125–1132 (2006).
154. Savojardo, C., Fariselli, P. & Martelli, P. L. INPS-MD : a web server to predict stability of protein variants from sequence and structure. *Bioinformatics* **32**, 2542–2544 (2016).
155. Chen, C., Lin, J. & Chu, Y. iStable : off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics* **14**, S5 (2013).
156. Emsley, P. & Lohkamp, B. Features and development of Coot. *Acta Crystallogr.* **D66**, 486–501

(2010).

157. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
158. Amer, B. R. & Clubb, R. T. A sweet new role for LCP enzymes in protein glycosylation. *Mol. Microbiol.* **94**, 1197–1200 (2014).

6. Appendix

6.1. TSA buffer screening

Table 6.1 – List of buffer solutions present in the buffer screening performed for Bcl-2.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Citric acid pH 4.0	Sodium acetate pH 4.5	Sodium citrate pH 4.7	Sodium acetate pH 5.0	Potassium phosphate pH 5.0	Sodium phosphate pH 5.5	Sodium citrate pH 5.5	MES pH 5.8	Potassium phosphate pH 6.0	MES pH 6.2	Sodium phosphate pH 6.5	Sodium cacodylate pH 6.5
B	MES pH 6.5	PIPES pH 6.7	MOPS pH 7.0	HEPES pH 7.0	Ammonium acetate pH 7.3	Sodium phosphate pH 7.5	Tris pH 7.5	Imidazole pH 8.0	Hepes pH 8.0	Tris pH 8.0	Bicine pH 8.5	Tris pH 8.5
C	Bicine pH 9.0	CAPS pH 9.0	Glycine-HCl pH 9.5	Glycine-HCl pH 10	CAPS pH 10.5	CAPS pH 11	Citric acid pH 4.0 + 300 mM NaCl	Sodium acetate pH 4.5 + 300 mM NaCl	Sodium citrate pH 4.7 + 300 mM NaCl	Sodium acetate pH 5.0 + 300 mM NaCl	Potassium phosphate pH 5.0 + 300 mM NaCl	Sodium phosphate pH 5.5 + 300 mM NaCl
D	Sodium citrate pH 5.5 + 300 mM NaCl	MES pH 5.8 + 300 mM NaCl	Potassium phosphate pH 6.0 + 300 mM NaCl	MES pH 6.2 + 300 mM NaCl	Sodium phosphate pH 6.5 + 300 mM NaCl	Sodium cacodylate pH 6.5 + 300 mM NaCl	MES pH 6.5 + 300 mM NaCl	PIPES pH 6.7 + 300 mM NaCl	MOPS pH 7.0 + 300 mM NaCl	HEPES pH 7.0 + 300 mM NaCl	Ammonium acetate pH 7.3 + 300 mM NaCl	Sodium phosphate pH 7.5 + 300 mM NaCl
E	Tris pH 7.5 + 300 mM NaCl	Imidazole pH 8.0 + 300 mM NaCl	Hepes pH 8.0 + 300 mM NaCl	Tris pH 8.0 + 300 mM NaCl	Bicine pH 8.5 + 300 mM NaCl	Tris pH 8.5 + 300 mM NaCl	Bicine pH 9.0 + 300 mM NaCl	CAPS pH 9.0 + 300 mM NaCl	Glycine HCl pH 9.5 + 300 mM NaCl	Glycine HCl pH 10 + 300 mM NaCl	CAPS pH 10.5 + 300 mM NaCl	CAPS pH 11 + 300 mM NaCl
F	Citric acid pH 4.0 + 1 M NaCl	Sodium acetate pH 4.5 + 1 M NaCl	Sodium citrate pH 4.7 + 1 M NaCl	Sodium acetate pH 5.0 + 1 M NaCl	Potassium phosphate pH 5.0 + 1 M NaCl	Sodium phosphate pH 5.5 + 1 M NaCl	Sodium citrate pH 5.5 + 1 M NaCl	MES pH 5.8 + 1 M NaCl	Potassium phosphate pH 6.0 + 1 M NaCl	MES pH 6.2 + 1 M NaCl	Sodium phosphate pH 6.5 + 1 M NaCl	Sodium cacodylate pH 6.5 + 1 M NaCl
G	MES pH 6.5 + 1 M NaCl	PIPES pH 6.7 + 1 M NaCl	MOPS pH 7.0 + 1 M NaCl	HEPES pH 7.0 + 1 M NaCl	Ammonium acetate pH 7.3 + 1 M NaCl	Sodium phosphate pH 7.5 + 1 M NaCl	Tris pH 7.5 + 1 M NaCl	Imidazole pH 8.0 + 1 M NaCl	Hepes pH 8.0 + 1 M NaCl	Tris pH 8.0 + 1 M NaCl	Bicine pH 8.5 + 1 M NaCl	Tris pH 8.5 + 1 M NaCl
H	Bicine pH 9.0 + 1 M NaCl	CAPS pH 9.0 + 1 M NaCl	Glycine-HCl pH 9.5 + 1 M NaCl	Glycine-HCl pH 10 + 1 M NaCl	CAPS pH 10.5 + 1 M NaCl	CAPS pH 11 + 1 M NaCl						

6.2. TSA additive screening

Table 6.2 – List of additive solutions present in the additive screening performed for Bcl-2. Ionic Liquid: [1] 1-Butyl-3-methyl-imidazolium chloride; [2] 1-Ethanol-3-methyl-imidazolium hexafluorophosphate; [3] 1-Octyl-3-methyl-imidazolium hexafluorophosphate; [4] 1-Butyl-3-methyl-imidazolium hexafluorophosphate; [5] 1,3-Dimethyl-imidazolium dimethylphosphate.

	1	2	3	4	5	6	7	8	9	10	11	12
A	10 mM β -mercaptoEtOH	20 mM β -mercaptoEtOH	10 mM DTT	20 mM DTT	10 mM TCEP	20 mM TCEP	5% (w/v) Glycerol	10% (v/v) Glycerol	20% (v/v) Glycerol	30% (v/v) Glycerol	10% (v/v) PEG 400	10% (w/v) PEG 1K
B	10% (w/v) PEG 3350	10% (w/v) PEG 8K	10% Ethylene glycol	50 mM NaCl	100 mM NaCl	200 mM NaCl	400 mM NaCl	600 mM NaCl	800 mM NaCl	1 M NaCl	1.5 M NaCl	20mM Imidazole pH 7.6
C	100mM Imidazole pH 7.6	300mM Imidazole pH 7.6	500mM Imidazole pH 7.6	20 mM MgCl ₂	20 mM CaCl ₂	20 mM NiCl ₂	20 mM MnCl ₂	20 mM ZnCl ₂	20 mM Sodium Acetate	20 mM Lithium Acetate	20 mM Zinc Acetate	20 mM Ammonium Acetate
D	0.2M Urea	300mM GdnHCl	5% (v/v) DMSO	10% (v/v) propan-2-ol	10% (v/v) ethanol	10% (w/v) Octyl glucoside	10 mM CHAPS	1% Triton X-100	10 mM EDTA	3% D-glucose	3% Sucrose	3% Maltose
E	10% (v/v) acetaldehyde	10mM ADP	10mM ATP	10mM Arginine	10mM Glutamate	10mM Glutamine	10mM Cysteine	10mM L-lysine	0.04% azide	0.2M Ionic liquid [1]	0.4M Ionic liquid [1]	0.2M Ionic liquid [2]
F	0.4M Ionic liquid [2]	0.2M Ionic liquid [3]	0.4M Ionic liquid [3]	0.2M Ionic liquid [4]	0.4M Ionic liquid [4]	0.2M Ionic liquid [5]	0.4M Ionic liquid [5]					

6.3. TSA results of buffer and additive screenings with Bcl-2

Table 6.3 - List of melting temperatures of Bcl-2 in each condition from the buffer and additive screenings performed. FP: False positive result (protein sample was denatured from the start of the assay).

Condition	Buffer screen T_m (° C)	Additive screen T_m (° C)
Control	62.64	63.12
A1	49.03	61.90
A2	52.63	61.90
A3	59.45	62.32
A4	62.64	61.90
A5	62.64	55.82
A6	63.04	67.91
A7	64.23	62.71
A8	62.26	63.12
A9	62.64	63.51
A10	62.26	63.91
A11	63.04	62.32
A12	49.83	70.30
B1	63.04	62.32
B2	64.23	61.90
B3	62.64	62.32
B4	62.26	62.32
B5	63.44	62.32
B6	63.44	63.12
B7	62.64	63.51
B8	61.46	63.91
B9	63.44	63.91
B10	62.64	63.91
B11	62.64	64.31
B12	63.04	61.90
C1	63.04	61.90
C2	63.04	61.41
C3	63.04	60.22
C4	63.44	62.32
C5	64.23	62.71
C6	63.84	67.10 ^{FP}
C7	48.63	61.90
C8	60.66	66.70

C9	59.85	61.90
C10	63.44	61.90
C11	62.64	60.63
C12	63.44	61.90
D1	63.04	61.41
D2	62.64	61.90
D3	63.04	61.90
D4	63.44	61.90
D5	63.84	62.32
D6	50.23	28.58
D7	63.84	28.98
D8	65.03	28.20
D9	63.44	61.41
D10	63.44	57.83
D11	63.44	62.32
D12	63.84	57.42
E1	64.23	70.30 ^{FP}
E2	63.04	62.71
E3	64.23	61.90
E4	64.23	62.32
E5	64.23	62.71
E6	64.23	61.90
E7	64.23	61.90
E8	63.84	62.32
E9	63.84	62.71
E10	63.84	59.83
E11	65.03	59.03
E12	65.03	63.51
F1	50.23	63.51
F2	59.45	61.03
F3	66.24	28.98
F4	64.23	46.61
F5	63.84	63.12
F6	63.44	61.90
F7	64.23	61.03
F8	63.44	
F9	63.84	
F10	64.23	
F11	64.63	
F12	51.03	
G1	65.03	
G2	64.63	

G3	64.63	
G4	64.63	
G5	65.03	
G6	65.03	
G7	64.23	
G8	63.84	
G9	65.83	
G10	65.03	
G11	65.43	
G12	67.83	
H1	66.24	
H2	65.43	
H3	65.43	
H4	65.83	
H5	66.63	
H6	66.63	

6.4. ^1H NMR spectra of the ligands under study

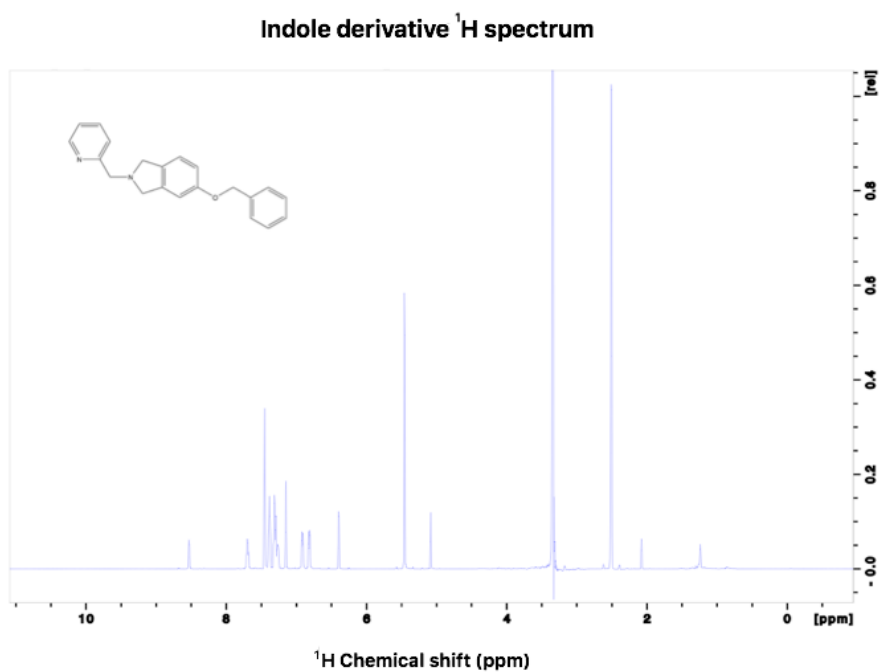


Figure 6.1 - ^1H NMR spectrum of the indole derivative.

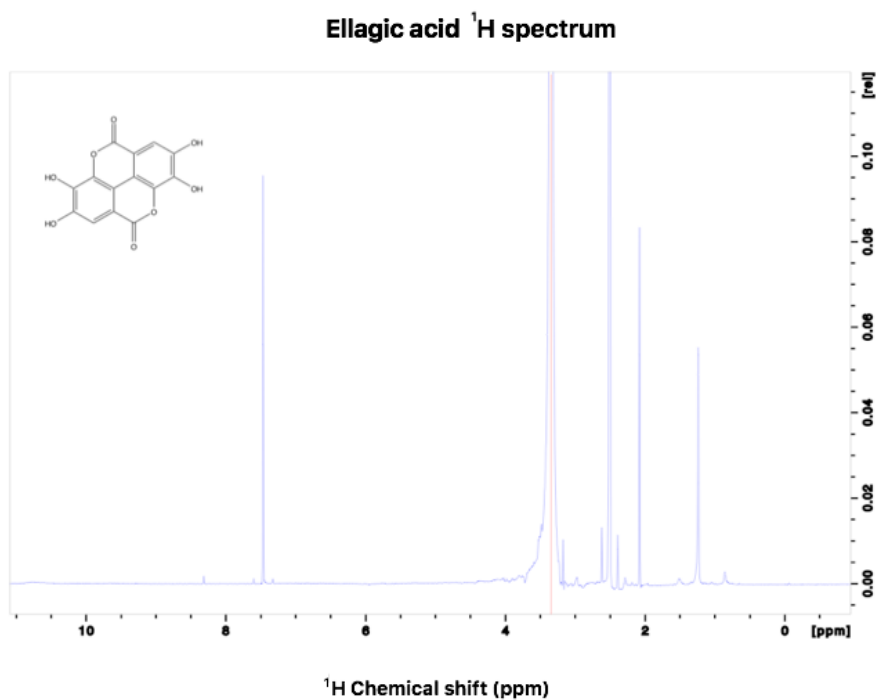


Figure 6.2 - ^1H NMR spectrum of ellagic acid.

Fisetin ¹H spectrum

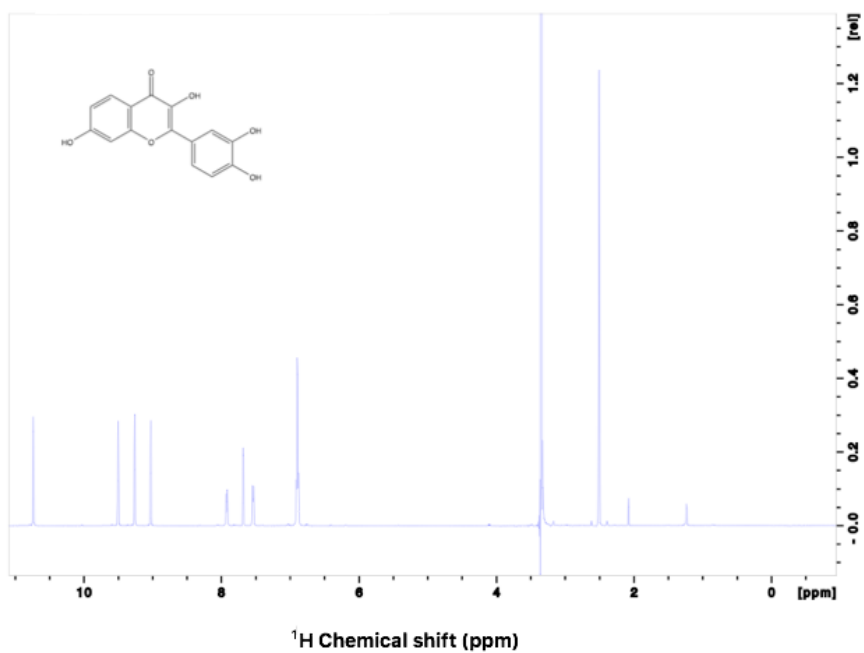


Figure 6.3 - ¹H NMR spectrum of fisetin.