Universidade de Lisboa

Faculdade de Farmácia

Dissertation

Design of a Multitarget Strategy for Proteasome and HDAC Inhibition

Pedro de Sousa Monteiro

Dissertation supervised by Professor Rita Alexandra do Nascimento Cardoso Guedes

Masters in Pharmaceutical and Therapeutic Chemistry

2019

Abstract

There are two main pathways through which cells dispose of damaged/misfolded proteins in order to ensure cell viability and health: the proteasome and aggresome pathways. It has been reported that by inhibiting both pathways a synergistic therapeutic effect is achieved in Multiple Myeloma. The immunoproteasome is a proteasome isoform of interest for cancer, inflammatory and autoimmune therapies. The purpose of this work is to find potential dual-inhibitors for the human immunoproteasome and HDAC6 through *in silico* studies.

Docking-based Virtual Screening was done on both intended targets with a database of approved compounds for therapy from DrugBank and ligand-protein interactions were analysed as a way to narrow down the Screening results. This led to four compounds of interest for further study: Abarelix, Ritonavir, Alprostadil and Paritaprevir.

Resumo

Existem duas principais vias de eliminação de proteínas malformadas ou danificadas de forma a promover a longevidade celular: a via proteassomal e a via aggresomal. Já se verificou que inibindo ambas as vias, se obtém níveis sinérgicos de efeitos terapêuticos em pacientes com Mieloma Múltiplo. O imunoproteassoma é uma isoforma do complexo proteico proteassoma com interesse para terapias de cancro, doenças inflamatórias e autoimunes. O objetivo deste trabalho é encontrar compostos com potencial de dupla inibição do imunoproteassoma e HDAC6 humanos através de estudos *in silico*.

Screening Virtual à base de Docking Molecular foi feito para ambos os alvos pretendidos utilizando a base de dados de compostos aprovados para terapia do DrugBank e as interações entre os ligandos e as proteínas foram analisadas para filtrar os resultados dos Screenings Virtuais feitos. Estas análises levaram a quatro compostos de interesse para estudos futuros: Abarelix, Ritonavir, Alprostadil e Paritaprevir.

Acknowledgements

Gratitude must be expressed to several people without whom this thesis would have been infinitely harder to do. Firstly, I want to thank my supervisor Professor Rita Guedes who taught and showed me how to tackle investigation work and how to direct it accordingly with the results obtained along the way.

Secondly, I want to thank the entire Computational Chemistry Group where I did most of my work this past year, since everyone there was always ready and available to clear any doubts on procedure and or any other matters.

I want to thank my friends Ruben Coutinho and Carla Santos for helping me endure and face difficulties with a different perspective.

From my family I must thank my mother Sameiro and father Augusto, who have struggled in order for me to get the opportunity to take this course and get a Master's degree.

Lastly, I want to thank my girlfriend Cláudia who has helped me keep a levelled head and mindset while never letting me dwell on the hardships of the work.

Figure Index

Table Index

Abbreviation Index

ATP – Adenosine Triphosphate

FDA – Food and Drug Administrations

MHC-I – Major Histocompatibility Complex class I

LMP-2 – Low Molecular Mass Protein 2

LMP-7 – Low Molecular Mass Protein 7

PA18α – Proteasome Activator α

PA18β – Proteasome Activator β

h CP – Human Constitutive Proteasome

h IP – Human Immunoproteasome

HDAC – Histone Deacetylase

DNA – Deoxyribonucleic Acid

HAT – Histone Acetyltransferases

NAD – Nicotinamide Adenine Dinucleotide

Hsp90 – Heat Shock Protein 90

PPP3CA – Protein Phosphatase 3 Catalytic Subunit Alpha Isozyme

CADD – Computer-Aided Drug Design

QSAR – Quantitative Structure-Activity Relationship

PDB – Protein Data Base

MOE - Molecular Operating Environment

RMSD – Root-Mean-Square Deviation

GOLD – Genetic Optimisation for Ligand Docking

PLP – Piecewise Linear Potential

ASP – Astex Statistical Potential

GA – Genetic Algorythm

PLIP – Protein-Ligand Interaction Profiler

HIV – Human Immunodeficiency Virus

HCV – Hepatitis C Virus

Aminaocid List

Arginine – ARG-R

Histidine – HIS-H

Lysine – LYS-K

Serine – SER-S

Threonine – THR-T

Glutamine – GLN-Q

Cysteine – CYS-C

Glycine – GLY-G

Proline – PRO-P

Alanine – ALA-A

Leucine – LEU-L

Methionine – MET-M

Phenylalanine – PHE-F

Tyrosine – TYR-Y

Valine – VAL-V

Index

Chapter 1: Introduction

1. The Biologic Targets in this Thesis

1.1.1 The Proteasome and the Ubiquitin-Proteasome System

The proteasome is a protein complex tasked with regulating abnormal, damaged or misfolded protein levels inside the cell through selective proteolysis. Depending on the substrate, or client proteins the proteasome can influence several biological processes including cell-cycle control, metabolic adaptation and cell differentiation by degrading cyclins, metabolic enzymes and transcription factors, respectively. The proteasome comprises two sub-complexes; a catalytic core responsible for the protein cleaving designated the 20S proteasome, and one or two 19S regulator proteins tasked with substrate identification and 20S core activation. Depending on how many 19S proteins are connected to the 20S core, one or two, the complex is designated as 26S or 30S proteasome respectively. The proteasome complex may undergo changes in its catalytic core and regulator proteins to create specialized variants of the complex, one of which will be expanded upon further in this section. With the purpose of differentiating the different proteasome isoforms the proteasome is referred as constitutive proteasome as its other forms are dubbed immuno- and thymo- proteasomes. [1–4]

Protein degradation by the proteasome may occur through two methods, a ubiquitin-dependent process (Figure 1) and a ubiquitin-independent one. For the former to occur, ATP is necessary as an energy source to jumpstart the activation of ubiquitin by the E1 enzyme. The activated ubiquitin is then transferred to the ubiquitin-conjugating E2 and lastly the ubiquitin-ligase E3 transfers the ubiquitin from E2 to the target protein. This process is repeated to form a chain of ubiquitin molecules connected through the lysine 48 residue in each ubiquitin. The chain serves to tag the target protein for degradation via the proteasome-ubiquitin system. The proteasome is then capable of identifying the client proteins through the 19S regulator proteins which remove the ubiquitin chain, open and activate the 20S core. [1, 5] For the latter degradation method to occur, the substrate proteins are required to possess an unfolded section in their structure whether be it a part of their natural structure or a result of oxidative stress, heat or any other denaturising factor. [6-7]

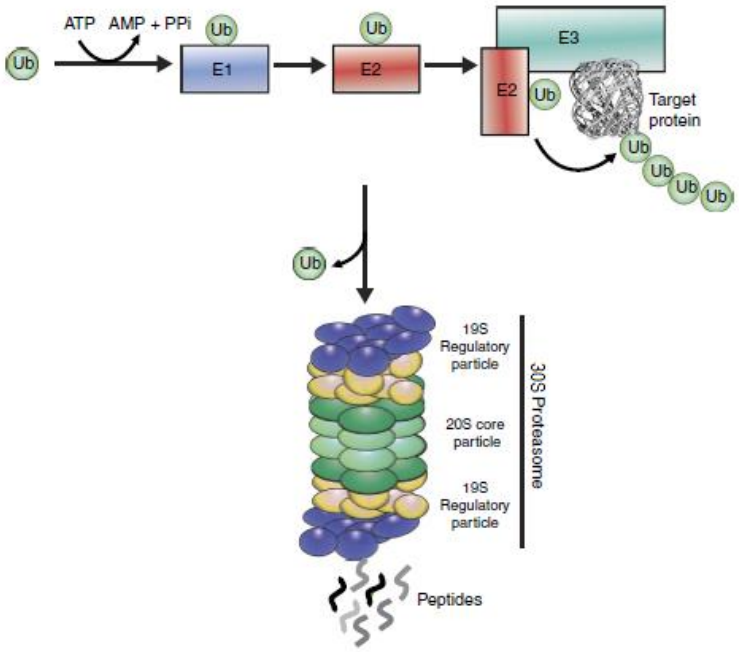


Figure 1: Schematic of the ubiquitin-proteasome degradation method. [5]

1.1.2 20S Proteasome Structure

The proteasome is a complex with a well-known structure through X-ray crystallography characterization. Furthermore, through these studies, it has been established that eukaryotic proteasomes share the same structural architecture, despite organism differences, whereas prokaryotic proteasomes possess a less complex structure. [2. 8-9]

As mentioned above, the proteasome comprises two sub-complexes. The eukaryotic 20S catalytic core has a cylindrical structure formed by the stacking of four protein chains, two outer α chains and two inner β chains, which in turn are themselves divided in seven different sub-units each, creating a $\alpha_{1\text{-}7}\beta_{1\text{-}7}\beta_{1\text{-}7}\alpha_{1\text{-}7}$ structure. Each β chain provides three distinct specific catalytic centres; β_1 sub-unit with caspase-like, β_2 sub-unit with trypsin-like and β_5 sub-unit with chymotrypsin-like activities meaning they cleave proteins after acidic, basic and hydrophobic residues respectively. All three catalytic sub-unit types have N-terminal threonine residues acting as nucleophiles in peptide bond hydrolysis. The α chains provide a physical barrier for the β inner chains blocking proteins which haven't been identified the 19S sub-complex for ubiquitin-dependent degradation or

folded proteins that are too bulky to enter the catalytic centre for ubiquitin-independent degradation. [1, 5, 9-10]

The 19S sub-complex is formed from two sub-structures, a "base" connected to the 20S α chains and a "lid" on top of the "base" structure. The base comprises nine sub-units, six ATPases, Rpt1-6, and three non ATPases, Rpn1-2 and Rpn10 while the lid is formed by eight sub-units Rpn3, Rpn5-9 and Rpn11-12. The 19S base is critical for structural assembly of the proteasome 26S and 30S complexes since the C-termini of the Rpt proteins are inserted in the α chains' sub-units, the Rpn proteins are responsible for the 19S sub-complex integrity (Rpn1-2) and ubiquitin recognition (Rpn10). The lid's primary role is deubiquitylation and unfolding of the substrate proteins. [11]

Though the proteasome's best-known form is the 26S/30S, catalytic 20S conjugated with one or two 19S regulators, oxidative stress induces 26S/30S disassembly towards 20S proteasome and induces the activity of another proteasome regulator protein, the 11S sub-complex. This heptameric complex binds to the 20S proteasome through their C-termini residues that insert themselves in the pockets between α sub-units. This conjugation induces a conformational shift in the α ring, forcing the α_2, α_3, α_4 and α_5 sub-units to change from the "closed gate" conformation, enabling unfolded proteins to reach the catalytic core suffer ubiquitin-independent degradation. [12]

1.1.3 Peptide Hydrolysis in the 20S Proteasome

Since the N-terminal threonine is the residue responsible for the nucleophilic attack in peptide hydrolysis, the catalytic sub-units of the proteasome are Threonine Proteases. For the hydrolytic activity of these type of enzymes it is paramount that the threonine is in the N-terminal of the protein so that its amino group is not involved in an amide bond to other amino acids.

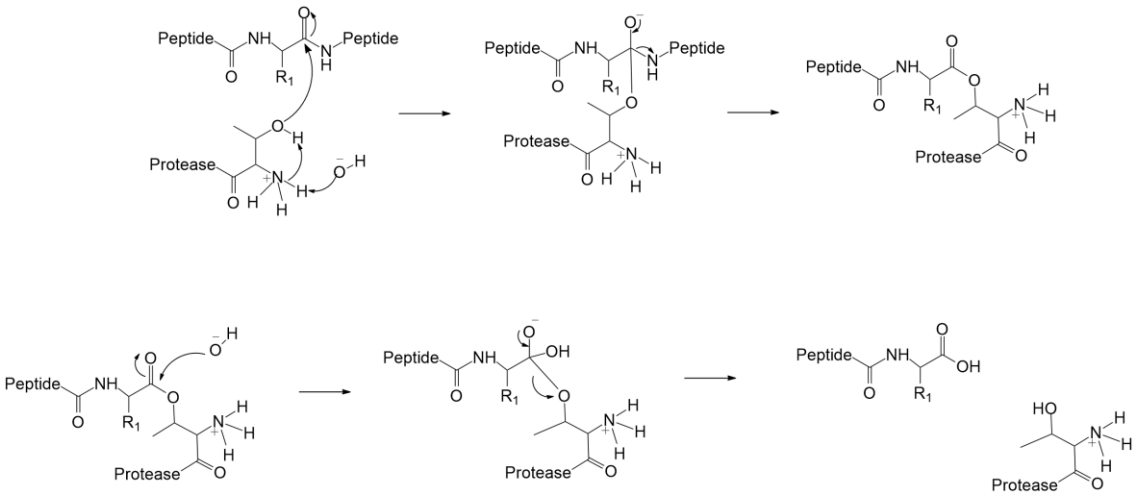


Figure 2: Peptide hydrolysis by Threonine Proteases. [13]

The amino group is necessary for the hydrolysis because it's part of a water mediated hydrogen transfer that enables the hydroxyl of the secondary alcohol to attack the carbonyl group of peptides forming an acyl-enzyme. A water molecule then attacks the ester's carbonyl in order to restore the threonine's hydroxyl group (Figure 2). [13]

1.1.4 20S Proteasome Inhibitors

The proteasome has been an interesting therapeutic target in oncology ever since 2003 when the FDA approved Bortezomib for the treatment of relapsed or refractory multiple myeloma. Since then several other proteasome inhibitors with different types of chemical structures were developed or discovered, though many are peptide derivatives. Bortezomib is a peptide boronate acid and reversibly inhibits the proteasome (Figure 3) by forming tetrahedral adducts with the threonine residues in the catalytic sub-units. These adducts are then stabilized by a hydrogen bond between the N-terminal of the threonine residues and one of the boronic acid's hydroxyl groups. [14]

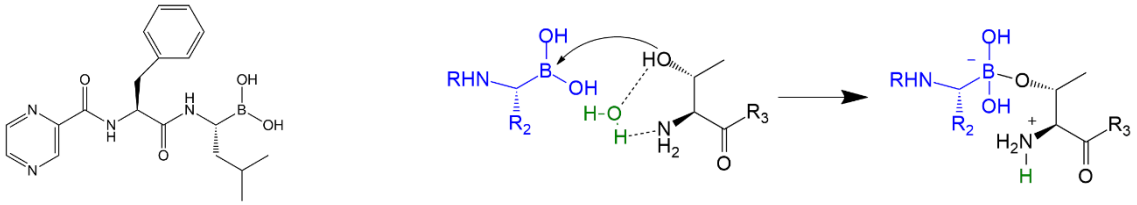


Figure 3: Structure of Bortezomib [15] and proteasome inhibition mechanism of boronate acids. [14]

The second proteasome inhibitor approved for multiple myeloma therapy was Carfilzomib in 2012. Unlike its predecessor, Carfilzomib (Figure 4) is a peptide epoxyketone derived from the natural product epoxomicin with an inhibitory specificity towards the chymotrypsin-like activity of β5 sub-units over the other catalytic sites of the proteasome. The catalytic threonine residue of the proteasome attacks the carbonyl forming a hemiacetal intermediate. The N-terminal then opens the epoxide moiety by attacking the least substituted carbon and establishing a seven-member morpholine ring. Unlike the boronate inhibition, epoxyketones inhibit the proteasome irreversibly. [14, 16]

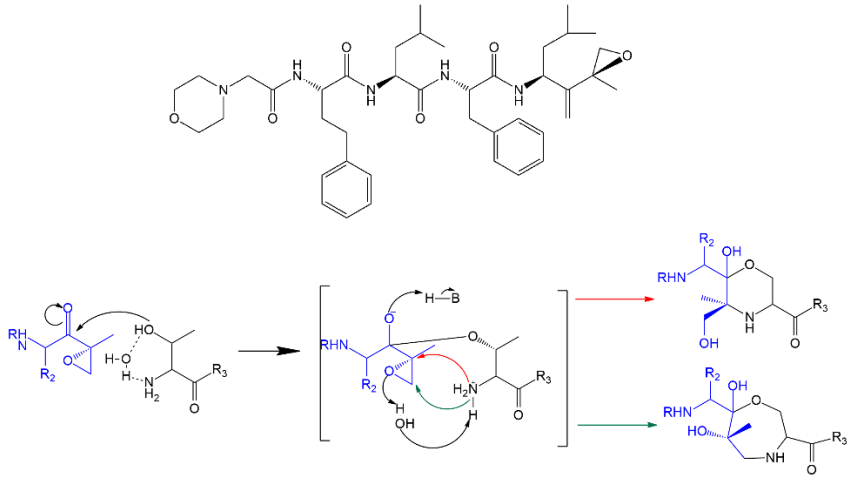


Figure 4: Structure of Carfilzomib [17] and proteasome inhibition mechanism of epoxyketones. [14]

Ixazomib (Figure 5) is the third and most recent proteasome inhibitor approved by the FDA in 2015. As a peptide boronate, its mode of action and specificity are identical to Bortezomib's. The difference between the two inhibitors is the pocket retention time, as both compounds inhibit reversibly, in time they both leave the β5 active pocket, however Ixazomib does this approximately six times faster. [18-19]

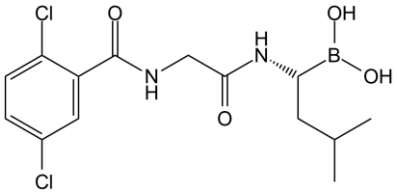


Figure 5: Structure of Ixazomib. [20]

Many other compounds with different chemical structures are currently in clinical trials, meaning there is a potential for more structural diversity in approved proteasome inhibitors in the future. Worthy of note is Marizomib, a β-lactone derived from the marine microorganism *Salinispora tropica*, capable of irreversibly inhibiting the proteasome catalytic trio (Figure 6). The catalytic threonine residue attacks the C from the carbonyl moiety opening the lactone followed by the formation of a tetrahydrofuran ring resulting of a displacement of the chloride atom from the inhibitor. Clinical trials for Marizomib are entering Phase III for the treatment of newly diagnosed glioblastoma (NCT03345095) [21] and Phase II for the treatment of multiple myeloma has already been completed (NCT00461045). [14, 22-24]

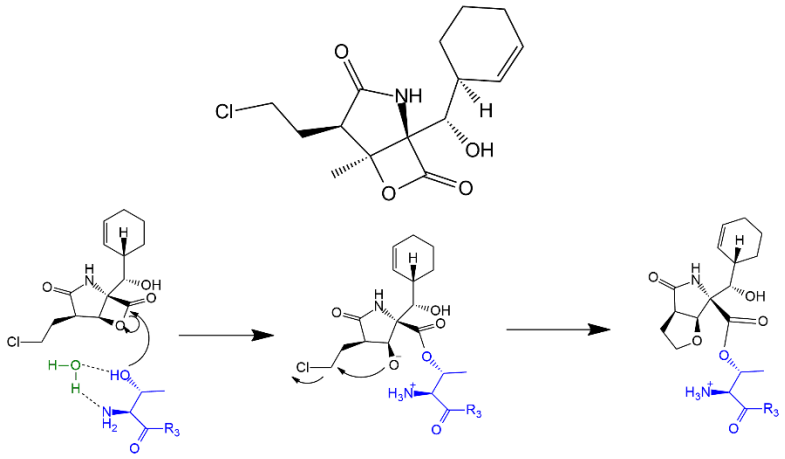


Figure 6: Marizomib Structure[24] and proteasome inhibition mechanism of β-lactones. [14]

1.2 The Immunoproteasome

The immune system possesses a diverse array of mechanisms to defend the organism from most threats. Of interest to this work is the interferon-γ's capabilities to up-regulate the expression of several proteins important for antigen presentation on the major histocompatibility complex class I (MHC-I) including three different catalytic sub-units and two non-enzymatic activators for the proteasome proteolytic complex. [25–27]

The catalytic β1, β2 and β5 are replaced by their immuno-counterparts iβ1(LMP-2), iβ2(MECL-1) and iβ5(LMP-7) as for the 19S regulatory complex is replaced with the 11S regulator, comprising the proteasome activators α and β (PA18α and PA18β). [28–30]

1.2.1 The Immunoproteasome Function

For both constitutive and immune proteasomes the catalytic β1, β2 and β5 sub-units the catalytic threonine in unaltered, making each of the sub-units Threonine Proteases, however they possess affinity to different substrate peptide bonds to hydrolyse. Though sub-units β2 and β5 keep their trypsin and chymotrypsin-like activities respectively, the same cannot be said for β1 sub-units. Constitutive β1 sub-units presents a preference for cleavage after acidic residues and as for iβ1 there is a change to chymotrypsin-like activity creating a second catalytic site generating peptides with C-terminal hydrophobic residues for antigen presentation on MHC-I. [31]

In normal cases peptides for antigen presentation originate from the cell's own proteins, but in cells infected with viral proteins presentation of peptides derived from the foreign proteins will signal virus specific cytotoxic T lymphocytes for elimination of infected cells. [31-32]

Though its ability to generate peptides for antigen presentation for defence against viral organisms the immunoproteasome is also capable of cleaving cell native misfolded proteins in likeness of the constitutive proteasome. [31]

1.2.2 Structural Differences Between Human CP and IP

Proteasome β1

```
Residue          10          20          30          40          50          60
h CP     TTIMAVQFDG GVVLGADSRT TTGSYIANRV TDKLTPIHDR IFCCRSGSAA DTQAVADAVT
h IP     TTIMAVEFDG GVVMGSDSRV SAGEAVVNRV FDKLSPLHER IYCALSGSAA DAQAVADMAA

Residue          70          80          90         100         110         120
h CP     YQLGFHSIEL NEPPLVHTAA SLFKEMCYRY REDLMAGIII AGWDPQEGGQ VYSVPMGGMM
h IP     YQLELHGIEL EEPPLVLAAA NVVRNISYKY REDLSAHLMV AGWDQREGGQ VYGTLGGMLT

Residue         130         140         150         160         170         180
h CP     VRQSFAIGGS GSSYIYGYVD ATYREGMTKE ECLQFTANAL ALAMERDGSS GGVIRLAAIA
h IP     RQPFAIGGSG STFIYGYVDA AYKPGMSPEE CRRFTTDAIA LAMSRDGSSG GVIYLVTITA

Residue         190         200
h CP     ESGVERQVLL GDQIPKFAVA TLPPA
h IP     AGVDHRVILG NELPKFYDE-------
```

Proteasome β2

```
Residue          10          20          30          40          50          60
h CP     TTIAGVVYKD GIVLGADTRA TEGMVVADKN CSKIHFISPN IYCCGAGTAA DTDMTTQLIS
h IP     TTIAGLVFQD GVILGADTRA TNDSVVADKS CEKIHFIAPK IYCCGAGVAA DAEMTTRMVA

Residue          70          80          90         100         110         120
h CP     SNLELHSLST GRLPRVVTAN RMLKQMLFRY QGYIGAALVL GGVDVTGPHL YSIYPHGSTD
h IP     SKMELHALST GREPRVATVT RILRQTLFRY QGHVGASLIV GGVDLTGPQL YGVHPHGSYS

Residue         130         140         150         160         170         180
h CP     KLPYVTMGSG SLAAMAVFED KFRPDMEEEE AKNLVSEAIA AGIFNDLGSG SNIDLCVISK
h IP     RLPFTALGSG QDAALAVLED RFQPNMTLEA AQGLLVEAVT AGILGDLGSG GNVDACVITK

Residue         190         200         210         220         230
h CP     NKLDFLRPYT VPNKKGTRLG RYRCEKGTTA VLTEKITPLE IEVLEETVQT MDTS
h IP     TGAKLLRTLS SPTEPVKRSG RYHFVPGTTA VLTQTVKPLT LELVEETVQA MEVE
```

Proteasome β5

```
Residue          10          20          30          40          50          60
h CP     TTTLAFKFRH GVIVAADSRA TAGAYIASQT VKKVIEINPY LLGTMAGGAA DCSFWERLLA
h IP     TTTLAFKFQH GVIAAVDSRA SAGSYISALR VNKVIEINPY LLGTMSGCAA DCQYWERLLA

Residue          70          80          90         100         110         120
h CP     RQCRIYELRN KERISVAAAS KLLANMVYQY KGMGLSMGTM ICGWDKRGPG LYYVDSEGNR
h IP     KECRLYYLRN GERISVSAAS KLLSNMMCQY RGMGLSMGSM ICGWDKKGPG LYYVDEHGTR

Residue         130         140         150         160         170         180
h CP     ISGATFSVGS GSVYAYGVMD RGYSYDLEVE QAYDLARRAI YQATYRDAYS GGAVNLYHVR
h IP     LSGNMFSTGS GNTYAYGVMD SGYRPNLSPE EAYDLGRRAI AYATHRDSYS GGVVNMYHMK

Residue         190         200
h CP     EDGWIRVSSD NVADLHEKYS GSTP
h IP     EDGWVKVEST DVSDLLHQYR EANQ
```

Figure 7: Residue alignment for human constitutive and immunoproteasome β sub-units. A green highlight represents equal residue sequences. A yellow highlight represents a section where iβ1 is shortened by one residue causing following sequences to be misplaced otherwise. A blue highlight represents residue changes important for isoform specificity. Residues listed in a red colour are responsible for the chymotrypsin-like activity of β5 sub-units. Residues listed in bold create a H-bond network that stabilizes proteolytic activity.

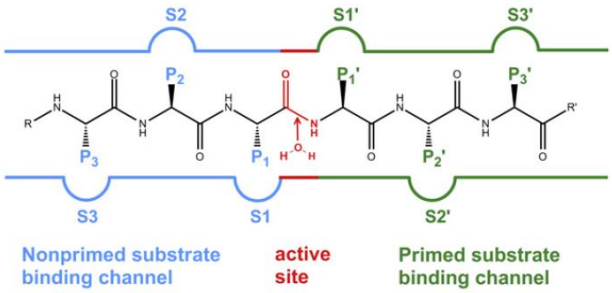


Figure 8: Representation of a peptide chain in the proteasome pocket. [33]

Though both isoforms are very similar there are a few important structural differences between them that justify the difference in inhibitor specificity (Figure 7). From the three catalytic sub-units β2 is the one with least significant changes to its overall structure and binding channels with the exception of Y114H which affects a primed pocket of the adjacent iβ1.

Looking at the β1 amino acid comparison a few changes are of importance towards proteasome specificity, namely the T20V, T31F, R45L, T52A substitutions which diminish the size and increase the hydrophobic nature of the S1 pocket. Also important are the T22A and A27V changes which together with the above-mentioned change in β2 which reduce the size and polarize the S3 pocket of iβ1.

Though the iβ5 sub-unit retains the same amino acids responsible for the chymotrypsin-like activity in A20. M45, A49 and C52. the binding pocket changes are quite significant. The change A27S makes the S3 pocket more hydrophilic, A46S and V128T increase the size of the oxyanion hole and alongside G47 a H-bond network is formed to stabilize the intermediate product of the proteolytic activity. Affecting the size of the pockets, G48C diminishes the depth of the S2 pocket and S53Q helps stabilizing M45 making it adopt a conformation which widens the S1 bonding pocket of the sub-unit.

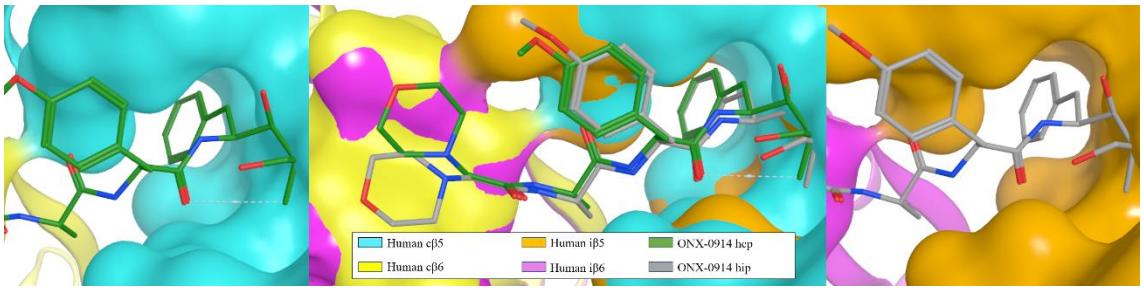


Figure 9: Human constitutive and immune β5 pockets with crystallographic ligand ONX-914.

Analysing the surfaces of both constitutive and immune β5 pockets (Figure 9) becomes clear the size change between both isoforms. The constitutive pocket is in most regions, smaller than the immune counterpart with exception noteworthy to the surface area created by C48 in the immune β5 sub-unit where the superficial surface in the immune proteasome "bridges" the sub-units that form the pocket.

1.2.3 Immunoproteasome in Pathophysiology

The immunoproteasome has been connected to a plethora of diseases. In a study on induced inflammatory colitis, immunoproteasome deficiency led to reduced symptoms when compared to wild type mice and treatment with ONX-0914 led to reduced pathological symptoms. [34] This leads to the conclusion that in inflammatory pathologies the increase of cytokines leads to an upregulation of the immune isoform of the proteasome and therefore specific immunoproteasome inhibition is a potential therapy for these conditions.

Another area of therapeutic interest on the immunoproteasome is its connection with the pathogenesis of autoimmune diseases through modulation of the differentiation of $CD4^{+}$T cells. Inhibition of the iβ5 has been shown to supress the differentiation of pro-inflammatory T cells (Th1 and Th17) while enhancing the generation of anti-inflammatory regulatory T cells. [35]

In likelihood to the constitutive proteasome, the immunoproteasome has also been found to be overexpressed in cancer cells, however the upregulation of the immunoproteasome genes is regulated by extrinsic and intrinsic cell processes in different types of cancer

cells. In breast cancer the upregulation on the presence of IFN-γ-secreting tumor-infiltrating lymphocytes to induce immunoproteasome overexpression. In acute myeloid lymphoma the upregulation is regulated autonomously by the cancer cell since they are reliant on oxidative phosphorylation and therefore more susceptible to oxidative stress. [36]

1.2.4 Immunoproteasome Inhibitors

Though several immunoproteasome inhibitors have been discovered and studied, in terms of chemical families they fall on the same categories as the constitutive proteasome's inhibitors. Since the nucleophilic threonine residue remains unchanged in constitutive and immunoproteasomes the inhibition reactional mechanism will be the same in either case. It is known now that specificity between the two proteasome isoforms comes from the non-covalent interactions between the inhibitor and the substrate-binding channels of the proteins. [32. 38-39] That however, is not to say that there are no promising molecules when it comes to specific immunoproteasome inhibitors.

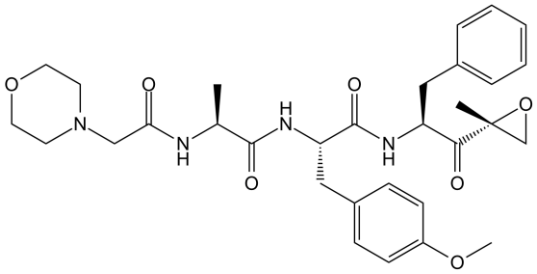


Figure 10: Structure of immunoproteasome selective inhibitor ONX-0914. [39]

There have been made efforts to discover immunoproteasome selective inhibitors in order to closing in on the prospects of selective immunoproteasome inhibition approved for therapy. Onyx Pharmaceutical not only owns the already approved Carfilzomib (Kyprolis™) proteasome inhibitor, also one of the best known selective iβ5 inhibitors, ONX-0914 (Figure 10). ONX-0914 has been studied thoroughly and has shown higher inhibitory activity for the iβ5 (IC_{50} = 0.083 μM) than for the other immune sub-units

(IC_{50} = 0.211 μM for iβ1 and IC_{50} = 0.675 μM iβ2) [40] for promising results in viral-mediated heart inflammation in mice. [41]

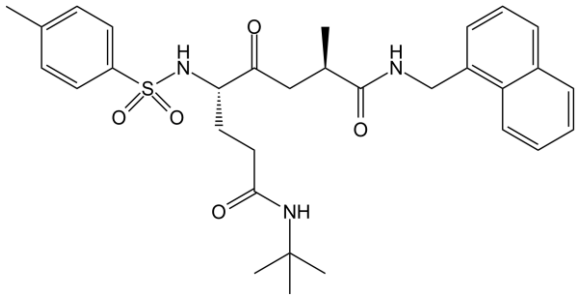


Figure 11: Structure of immunoproteasome inhibitor from Cornell University PKS2256. [42]

In a patent review by Ogorevc *et al.* published in 2018 [42] several immunoproteasome inhibitors are presented. PKS2256 (Figure 11), a compound developed by Cornell University (WO2015106200) has the lowest iβ5 IC_{50} in the review at IC_{50} = 0.09 nM.

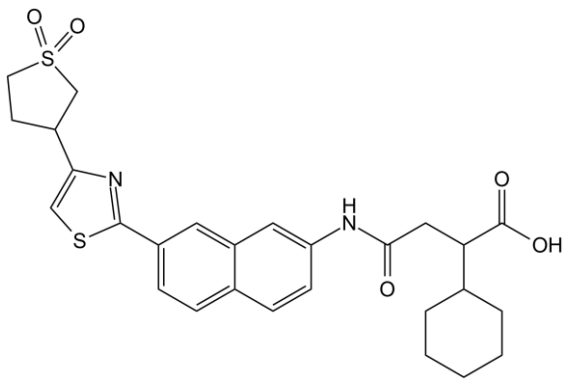


Figure 12: Structure of immunoproteasome inhibitor from Roche "example12". [42]

Roche is in possession of several high potential inhibitors for the immunoproteasome and one has been used in X-ray crystallography and is one of the reference inhibitors featured in this thesis (PDB ID 5M2B), however in the review by Ogorevc *et al*. there is one inhibitor with lower values of IC_{50}, "example 12" (Figure 12). Roche "example 12" has shown higher affinity for iβ5 over constitutive β5 sub-unit (IC_{50} = 0.021 μM for the immune isoform and IC_{50} = 1.075 μM for the constitutive). [42]

1.3.1 Histone Deacetylase (HDAC)

In eukaryotic cells genetic information is kept in molecules of DNA packed inside the cell nucleus in chromatin complexes. These complexes are created from the arrangement and agglomeration of nucleosomes, an octameric core comprising two copies of four histones (H2A, H2B, H3 and H4) and DNA (146-147 base-pairs) tightly wrapped around them. Nucleosome cores are connected through linker DNA of variable length and a linker histone (H1). [44-46]

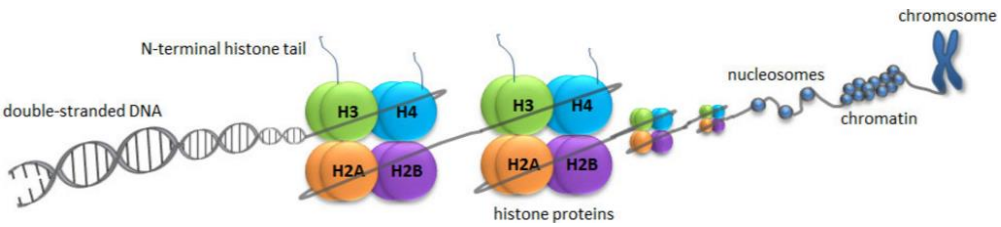


Figure 13: Histone role in DNA packaging and storage. [46]

Chromatin structure has been found to be highly influenced by interactions between nucleosomes and non-histone proteins and even nucleosomal histones and DNA chains. Such interactions are possible through the histones H3 and H4 N-terminal histone tails (Figure 13) though they may be rather short (26 and 19 amino acids respectively) the interactions in which they intervene are reported to possess central roles in chromatin structure and function. [47]

Scientific works dating as early as 1964 [48] have established that histones H3 and H4 suffer post-translational modifications on the N-terminal tails. Due to their highly basic nature the N-terminal tails are able to interact with neighbouring nucleosomes influencing the packaging of nucleosomes and consequently chromatin structure. It stands to reason that modifications to these amino acid chains will have their own effects on chromatin organization.

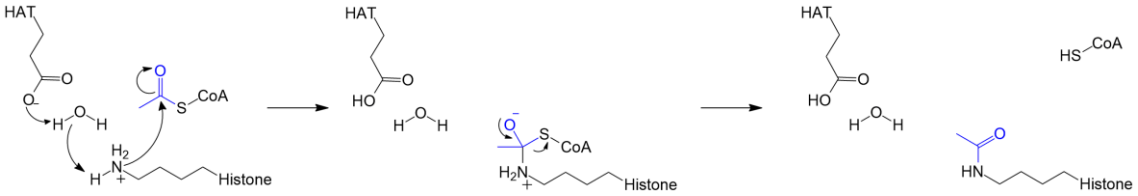


Figure 14: Lysine ε-amino group acetylation mechanism of HATs. [49]

Histone acetylation was one of the first post-translational modifications to be reported (alongside histone methylation) in the 60's. This transformation is now known to be regulated by two enzymatic families, histone acetyltransferases (HATs) and histone deacetylases (HDACs). Histone acetylation is a specific case of protein lysine acetylation which refers to the transfer of an acetyl moiety to the ε-amino group of the amino acid (Figure 14). [50]

1.3.2 The HDAC Superfamily of Enzymes

HDAC is a superfamily of enzymes which are divided in four classes according to homology with yeast deacetylases. Class I HDACs comprises HDACs 1, 2. 3 and 8 which are ubiquitously expressed and present mainly in the cell nucleus. Class II HDAC is itself divided in two subclasses IIa and IIb. The former consists of HDACs 4, 5, 7, 9 and unlike other members of the HDAC superfamily, class IIa HDACs present a restricted expression pattern. HDACs 5 and 9 are highly expressed in muscle tissue and the brain. HDAC4 is mainly present in the brain and skeleton growth plates and HDAC7 in endothelial cells and thymocytes. Class IIb HDAC is formed by HDAC 6 and 10. HDAC10 has been recently reported to a viable therapeutic target due to its ability to overcome chemoresistance in neuroblastoma [51] as well as having a crucial role in homologous recombination. [52] HDAC6 has been reported to be the main cytoplasmatic deacetylase in mammalian cells [53] and to serve as an important regulator in the proteasome-ubiquitin system by binding to the complex protein-ubiquitin and transport the misfolded proteins through microtubules to the perinuclear aggresome (this latter role of HDAC6 will be expanded upon further in this section). [54]

Class III of the HDAC family, also designated as the Sirtuin family, comprises seven (SIRT1-7) enzymes located in different parts of the cell. SIRT 1 and 2 can be found on the cytoplasm, SIRT1 is also active inside the cell's nucleus alongside SIRT 6 and 7, SIRT 3, 4 and 5 operate inside the mitochondria. Unlike the other HDAC classes, sirtuins are nicotinamide adenine dinucleotide (NAD^+) dependent for their deacetylation mechanism. while the other HDAC classes employ a catalytic network centred around a zinc atom, also called a zinc finger in the literature. [56-57]

The sole member of HDAC class IV is HDAC11. Though it is the least known member of the family, HDAC11 has been discovered to have tremendously more efficient (10,000-fold) lysine defatty-acylase than deacetylase activity. Additionally HDAC11 has been reported to possess the ability to restore the responsiveness of $CD4^+$ T cells by inducing inflammatory antigen-presenting cells. [57-58]

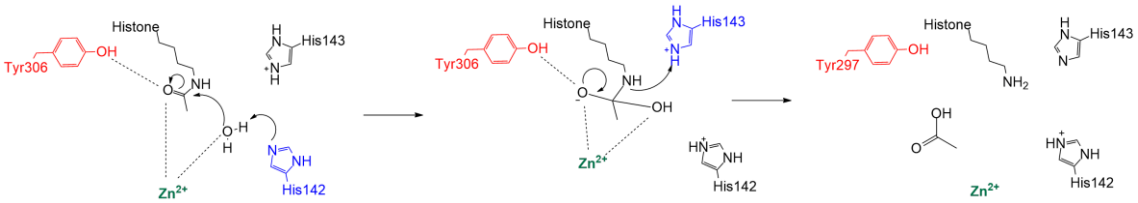


Figure 15: Zn-depend HDAC enzymatic mechanism based on HDAC8 pocket. [59]

HDAC classes I, II and IV employ the same enzymatic mechanism revolving a zinc ion. In the active pocket of these enzymes, two histidine residues are of paramount importance. One histidine residue begins the reaction by activating a zinc-coordinated water molecule, enabling the nucleophilic attack on the carbonyl moiety of the acetylated histone. The other is necessary for hydrogen replenishing on the leaving amino group of the amide hydrolysis (Figure 15).

1.3.3 HDAC Inhibitors

Though there are several known chemical structures with HDAC inhibitory activity, they all follow the same "architecture": a Zinc-chelating group for interactions inside the

active site; a “cap” group for surface recognition and interactions with the outer regions of the HDAC pocket; and an aliphatic Linker group connecting the other two moieties of the molecule.

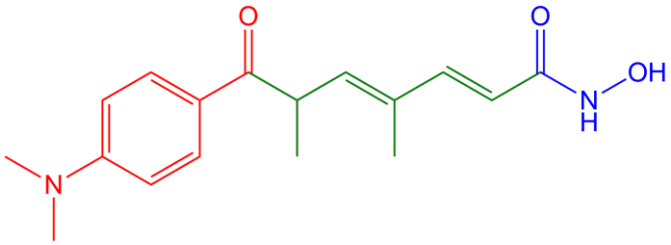


Figure 16: Structure of HDAC inhibitor Trichostatin A. [60]

The first reported HDAC inhibitor was Trichostatin A (Figure 16), a natural compound isolated from a strain from the bacteria *Streptomyces* and was firstly identified as an antifungal antibiotic. [60] Trichostatin A is a Zinc dependent HDAC pan-inhibitor (inhibits several HDAC isoforms without great specificity) with IC_{50} for HDAC6 in the magnitude of the nanomolar (0.4 nM). [61] In Figure 16 the division of the structure in the three components of the HDAC inhibitor architecture is highlighted in different colours: Blue highlights the hydroxamic acid group that chelates with the Zn ion on the HDAC active site; Red highlights the “cap” moiety of the molecule; Green highlights the aliphatic linker between the “cap” and the Zn-binding group.

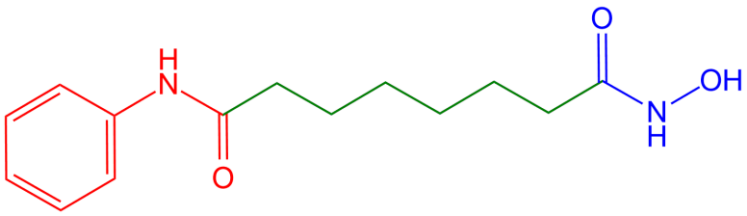


Figure 17: Structure of HDAC inhibitor Vorinostat (SAHA). [62]

Vorinostat (also known as SAHA) is another Zinc dependant HDAC pan-inhibitor with low value of IC_{50} for the human HDAC6 (1.4 nM). [63] Like Trichostatin A, Vorinostat is a compound using the quite popular hydroxamic acid group as a Zinc-binding group.

In Figure 17 the structure of Vorinostat appears with the same colour scheme as with Trichostatin A (Figure 16).

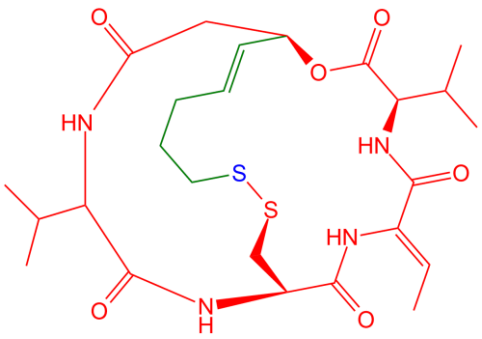


Figure 18: Structure of the depsipetide HDAC inhibitor Romidepsin. [64]

Romidepsin is quite different from Trichostatin A or Vorinostat. Unlike the previous inhibitors described, Romidepsin possesses a more specific HDAC inhibition profile since it has been shown that it mainly inhibits class I HDACs and has higher activity (IC_{50} = 4 nM in HDAC1) than for class II isoforms (IC_{50} = 281 nM in HDAC6). [65] Another difference is that Romidepsin is a pro-drug, meaning it will open the disulfide bridge once inside the cell and the sulfur atom on the longest “leg” will be used as the Zinc-binding moiety. [66]

1.3.4 Proteasome and HDAC6 Synergy

In a previous section of this report, HDAC6 was mentioned to serve as a regulator of protein degradation by transporting misfolded proteins to the perinuclear aggresome. In more detail, HDAC6 is found mainly in the cytoplasm its substrates there is tubulin which can polymerize to form microtubules, platforms for intracellular transport of proteins to be degraded by the lysosome. HDAC6 has been reported to regulate the acetylation of Hsp90 (heat shock protein 90) in multiple myeloma cells and therefore have an impact on the degradation of PPP3CA (Protein Phosphatase 3 Catalytic Sub-unit Alpha Isozyme), an enzyme highly expressed in multiple myeloma cells.[53, 67]

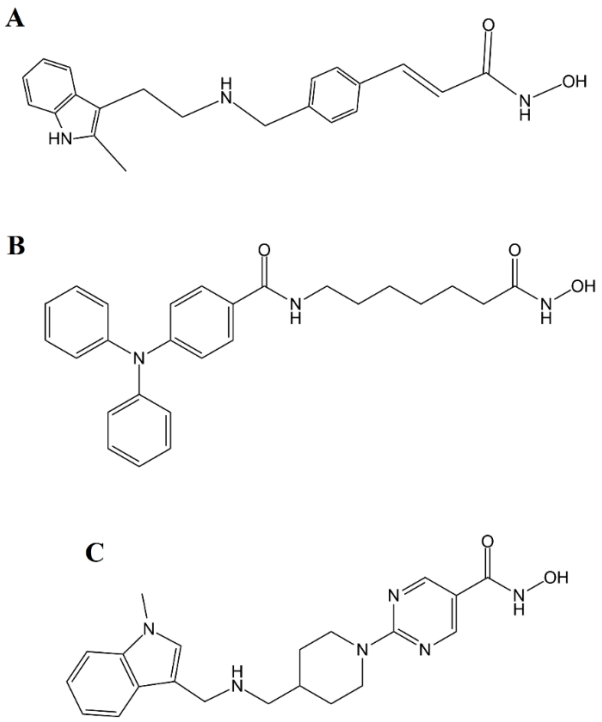


Figure 19: Structures of three HDAC inhibitors used in combination with Bortezomib: A) Panobinostat [67]; B) Ricolinostat [68]; C) Quisinostat [69]

HDAC inhibitors have been paired with Bortezomib in combinatory therapies as a method for overcoming drug resistance. The combinatory inhibition of the two major cellular mechanism for disposal of damaged/misfolded proteins has been shown to lead to a synergistic level of apoptosis. By inhibiting both proteasome and aggresome pathways, an increase in poly-ubiquitinated proteins results in increased cellular death. [70]

1.4 Objectives of this Investigation Work

The immunoproteasome is a possible therapeutic target for several pathologies of interest, as well as being a target in cases of Constitutive Proteasome inhibition resistance and it's interesting to investigate if the synergy with HDAC6 could lead to new and powerful therapies.

The purpose of the investigation described in this thesis is to ascertain the existence (or not) of potential dual-inhibitors for the Immunoproteasome and HDAC6 among already known and therapy-approved compounds. On this work Virtual Screenings were conducted to look for possible dual-inhibitors of the human Immunoproteasome and human HDAC6.

Chapter 2: Methodology

2.1 Computer-Aided Drug Design in Drug Discovery

The process of discovering new and effective drugs is complex, requiring interdisciplinary knowledge and immense investments of time and economic funds, in order to yield a drug design with therapeutic effectiveness and commercial viability. In order to help in the effort of drug designing, computational tools were developed in order to reduce the time and investment needed. The collective of these procedures is known as Computer-Aided Drug Design (CADD) or *in silico* drug designing.

The fundamental basis for CADD is that any pharmacologically active molecule acts through interactions with biological molecules like proteins or nucleic acids as an example. Therefore, it stands to reason that developments in techniques like X-ray crystallography and Nuclear Magnetic Resonance spectrometry will empower the viability of CADD methods. [71]

In silico methods allow for detailed studies of the biologic target for active site identification to understand how will any potential drug interact with it, if at all, homology studies to search for homologue models that may be economically more viable to acquire for posterior *in vitro* or *in vivo* assays. Likewise, it's also possible to create mathematical models from known active molecules in order to predict if any other compound being studied shares any important physical or chemical characteristics with the reference molecules. It's also possible to study a large number of molecules and trim the list of possible drugs through ranking systems based on the free energy of the compound-target complex. [71-72]

However, CADD's major flaw is that all the information these methods may provide, and how reliable they may be, it must be considered that they are all predictions and on their own, they are not sufficient evidence for the activity of any given compound. *In silico* predictions are required biological confirmation before the molecule can be considered for drug production.

2.2 Introduction to Virtual Screening

Virtual Screening is an ample term that simply put refers to any computational processing of a large database of molecule data in order to trim said molecular database to a "handful" of lead compounds. As it is a strictly computational procedure, Virtual Screening's strengths lie in saving both time and funds for research groups prior to laboratorial investment, while its weaknesses are defined by the amount of information gathered (crystallographic structures, compound characterization etc.) and computational power available.

As said above, the term Virtual Screening is applicable to a myriad of procedures. Virtual Screening can be divided into two major groups of approaches: Ligand-Based and Structure-Based. Ligand-Based Virtual Screening approaches revolve around knowing the molecular structure of active molecules for the intended macromolecular target like Quantitative Structure-Activity Relationships (QSAR) that relies on calculating molecular descriptors of known bioactive molecules to build a model that will then be used to predict bioactivity of molecules of interest. Structure-Based Virtual Screening methodologies centre around the knowledge of a specific biologic target's structure in order to determine possible lead compounds. [73] Though several approaches could be taken, in this work a Docking-Based Virtual Screening approach was employed, meaning Molecular Docking and its scoring and ranking were the main filter to identify any possible dual Immunoproteasome-HDAC6 inhibitors. [74]

2.3 Introduction to Molecular Docking

Molecular docking has been established has an important part of the bioinformatic arsenal employed in CADD as it allows for predicting the binding modes of small molecules in a biological macromolecule's binding sites (Figure 20). This allow to further study biomolecules of therapeutic interest and determine the interactions established between the protein and small molecules that may lead to target modulation. For any docking protocol a protein three dimensional structure, whether ligand-bound or not, as well as any ligand's structure and a computational framework capable of running docking and scoring procedures are needed. [75]

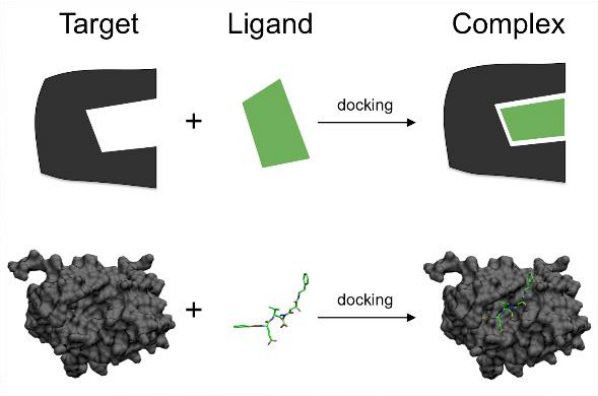


Figure 20: Demonstrative scheme of molecular docking. [76]

2.4 Crystallographic Structures

Table 1: Data collection on Human Immunoproteasome β5 sub-unit crystallographic structures from PDB.

Macromolecule	PDB	Ligand	Ligand ID	Resolution (Å)	DOI
Yeast 20S proteasome with human beta5i and human beta6	5M2B	Ro19	7DX	2.7	10.1002/cbic.201700021
Yeast 20S proteasome with human beta5i and human beta6	5L5D	ONX-914	04C	2.8	10.15252/embj.201695222
Yeast 20S proteasome with human beta5i and human beta6	5L5E	Carfilzomib	3BV	2.9	10.15252/embj.201695222
Yeast 20S proteasome with human beta5i and human beta6	5L5F	Bortezomib	BO2	2.5	10.15252/embj.201695222
Yeast 20S proteasome with human beta5i and human beta6	5L5H	PR-924	39V	2.6	10.15252/embj.201695222
Yeast 20S proteasome with human beta5i and human beta6	5L5I	Epoxyketone9	38X	2.9	10.15252/embj.201695222
Yeast 20S proteasome with human beta5i and human beta6	5L5J	Epoxyketone14	6NS	2.9	10.15252/embj.201695222

Table 1(cont): Data collection on Human Immunoproteasome β5 sub-unit crystallographic structures from PDB.

Macromolecule	PDB	Ligand	Ligand ID	Resolution (Å)	Reference
Yeast 20S proteasome with human beta5i and human beta6	5L5O	Epoxyketone 16	79P	2.6	10.15252/embj.201695222
Yeast 20S proteasome with human beta5i and human beta6	5L5P	Epoxyketone 17	79L	2.8	10.15252/embj.201695222
Yeast 20S proteasome with human beta5i and human beta6	5L5Q	Epoxyketone 18	6NV	2.8	10.15252/embj.201695222

In Protein Data Base (PDB) there are ten available structures for a yeast-human hybrid proteasome (20S constitutive yeast proteasome with human immunoproteasome iβ5 and β6 sub-units) [77–86].

From these crystallographic structures were selected through a series of small and fast docking procedures to select the best structures out of the list since if even in small docking calculations, the crystallographic ligands' conformation can be replicated, it would stand to reason that in more extensive and exhaustive calculations the crystallographic conformations will be replicated as well. Five out of the ten were selected: 5M2B, 5L5F, 5L5H, 5L5I and 5L5J.

Table 2: Data collection on Human and Danio Rerio HDAC6 crystallographic structures from PDB.

Macromolecule	PDB	Ligand	Ligand ID	Resolution (Å)	DOI
Human HDAC6	5EDU	Trichostatin A	TSN	2.79	10.2210/pdb5EDU/pdb
Danio Rerio HDAC6	533I	Vorinostat	SHH	1.32	10.2210/pdb5EEI/pdb
Danio Rerio HDAC6	5WGL	Ricolinostat	AH4	1.70	10.2210/pdb5WGL/pdb
Danio Rerio HDAC6	5WGM	ACY-1083	AH7	1.75	10.1073/pnas.1718823114
Danio Rerio HDAC6	6R0K	SS208	JNN	1.15	10.2210/pdb6R0K/pdb

Table 2(cont.): Data collection on Human and Danio Rerio HDAC6 crystallographic structures from PDB.

Macromolecule	PDB	Ligand	Ligand ID	Resolution (Å)	DOI
Danio Rerio HDAC6	6CSP	Cyclohexenylhydroxamate	FBM	1.28	10.2210/pdb6CSP/pdb
Danio Rerio HDAC6	6CW8	RTS-V5	FGY	1.9	10.2210/pdb6CW8/pdb
Danio Rerio HDAC6	6DVM	DDK-122	HBJ	1.47	10.2210/pdb6DVM/pdb
Danio Rerio HDAC6	6DVO	Bavarostat	HBV	1.98	10.2210/pdb6DVO/pdb
Danio Rerio HDAC6	6MR5	Sulfanylacetamide	W45	1.85	10.2210/pdb6MR5/pdb

Though PDB provides a larger number of crystallographic structures for the human HDAC6 than for the Immunoproteasome, most were considered improper to be used since the ligand was not positioned near the zinc cation inside the pocket. This would make it hard to analyse pose replication since the centre of the active site would always be defined in the docking validation step to be that zinc atom. In alternative, HDAC6 from *Danio Rerio* provides a larger selection of structures with ligands positioned near the zinc cation. Nevertheless, this work is to be focused on inhibition on Human macromolecules and so a homology study had to be done. On Molecular Operating Environment (MOE) after aligning and superposing the ten structures for HDAC6 (Table 2) it's possible to retrieve a plot of the average RMSD (Root-Mean-Square Deviation) between the structures.

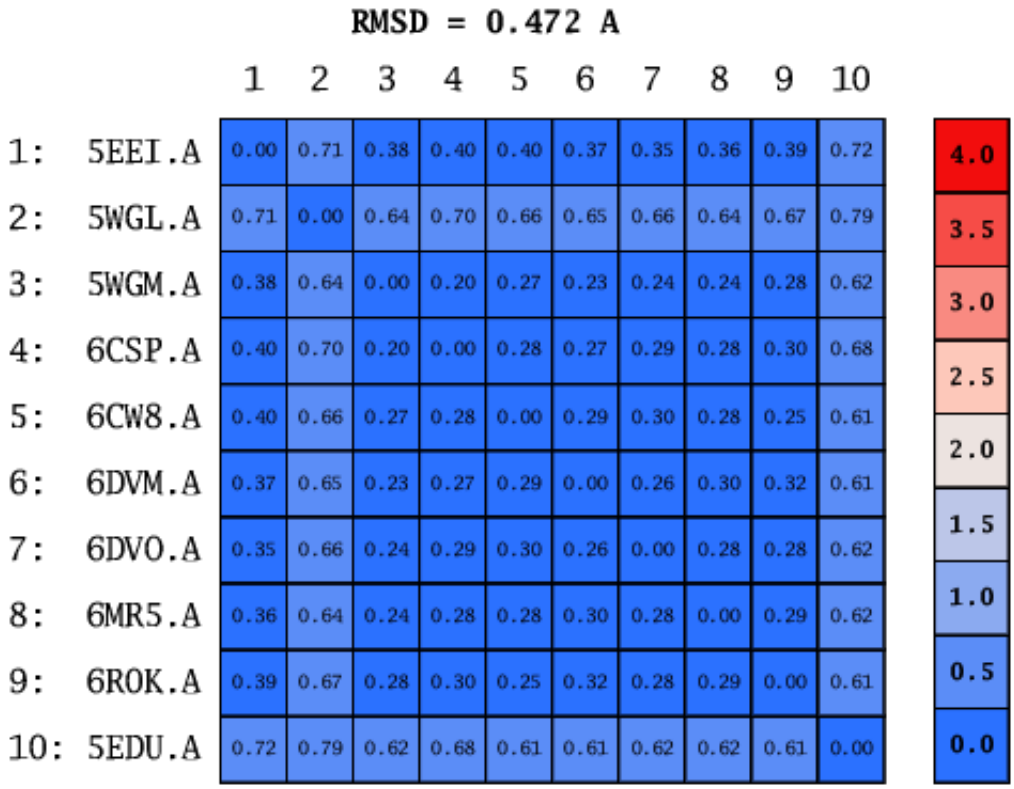


Figure 21: RMSD (Å) plot produced by MOE of the ten HDAC6 crystallographic structures. The RMSD value appearing as a header (0.472 Å) for the plot is the average deviation value between all ten structures.

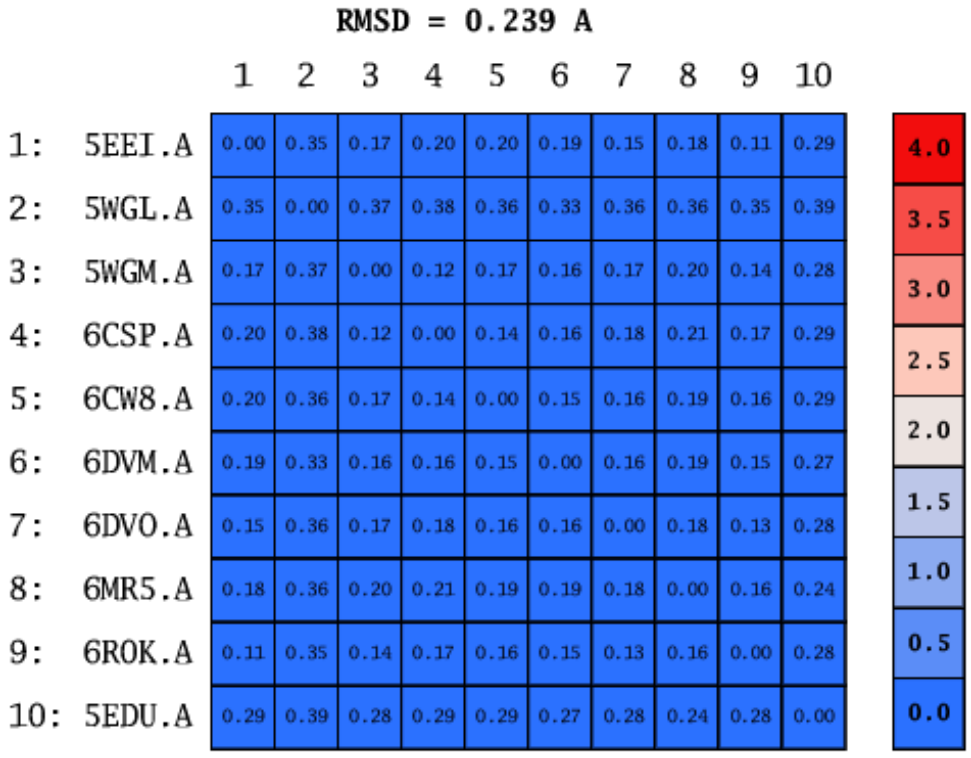


Figure 22: RMSD (Å) plot produced by MOE of the ten HDAC6 crystallographic structures considering only residues within a 10 Å radius from the zinc cation. The RMSD value appearing as a header (0.239 Å) for the plot is the average deviation value between all ten structures.

In (Figure 21) a plot comparing the deviation between the ten crystallographic structures (351 residues) shows that the differences between the Human structure and the *Danio Rerio* structures is minimal (0.61 to 0.79 Å). If considering only residues closer to the zinc cation of the active site (46 residues), as seen in Figure 22, the deviations registered are even smaller (0.24 to 0.39 Å) to a point it can be assumed the deviation comes from the different resolutions of each structure and not from actual differences in the residue sequence. Therefore, the structures from *Danio Rerio* will be considered equivalent to the human structure and will be used in this work as such.

2.5 Scoring Functions

For the docking calculations several scoring functions were available for study. From Genetic Optimisation for Ligand Docking (GOLD) four functions were used, the empirical Piecewise Linear Potential (ChemPLP), GoldScore, ChemScore and Astex Statistical Potential (ASP). [87] From Molecular Operating Environment (MOE) the GBVI/WSA dG was the only one selected, since it has been reported to be by far the best scoring function present in MOE. [88] From AutoDock both functions were tested, AutoDock 4.2 and AutoDock Vina for the enormous availability it provides being a cost-free program.

This makes for seven scoring functions in total from three different programs. Scoring functions can be divided into classes depending on what parameters they consider to calculate the protein-ligand bonding energy. [89] Score functions can be physics-based if their calculations are based in measuring the non-covalent components of the system's force-field such as van der Waals, electrostatic energy (H-bond contribution can be included in electrostatic energy contributions or be a term of its own) and solvation energy terms. From the seven soring functions to be studied, GoldScore and GBVI/WSA dG fit in this category. [88-89]

Some scoring functions calculate several energetic factors important in the protein-ligand binding and then rely on multiple linear regression or partial least-squares to determine the weight of each individual parameter. Empirical scoring functions like ChemScore,

PLP and AutoDock Vina have a scoring method in which they reward scores for protein-ligand interactions like H-bonds and coordination with metal atoms and penalize for frozen rotatable bonds, steric clashes and covalent bonds established between the protein and the ligand. [88–90]

Scoring functions can also be knowledge-based and functions falling into this category, like ASP, may differ between them in technical aspects but the principal behind them remains the same. They sum statistical potentials between the protein and ligand. This means they compute the occurrence frequency of a interaction between a pair of atoms from the protein and ligand, and take it a measurement of that interaction's contribution to the binding mode. [88-89]

There are some functions like AutoDock 4.2. that are not categorized as any of these types of scoring functions since they go a step further and combine multiple principles from the previous classes and should therefore be classified as hybrid scoring functions. In the case of AutoDock 4.2. it computes the force-field components as a physics-based function and takes it a step ahead by applying an empirical regression to determine the weight of each component in the overall binding in likeness of an empirical scoring function. [88-89, 91]

2.6 Docking Protocol Validation

It is of paramount importance to validate any computational method or protocol before actually using it for the intended objective in likeness to calibrating a scale or any other device. In this case for molecular docking there are a few topics that need to be shed light upon before confidently applying the docking protocol for virtual screening. How many scoring functions are available for use? Which of them will better describe the system? On which crystallographic structure should the screening be run? This section will answer these and other crucial questions.

2.5.1. Self-Docking

The first step is to look at the available scoring functions and establish which will be the one that can better describe the protein-ligand interactions for the immunoproteasome. Self-Docking consists in docking a crystallographic ligand on its native structure and analyse if the docking and scoring procedures are able to replicate the binding mode and ligand conformation recorded in the crystallographic structure. In order to evaluate the similarity between the docked conformations and the crystallographic reference the root-mean-square deviation (RMSD) is calculated. This allows for a measurement of the average distance between the atoms and will be the main parameter for determining the accuracy of the docking procedure. [75]

For the GOLD fitness functions most settings were kept as default, defining the centre of the binding site on the catalytic oxygen in Thr1 of the iβ5 sub-unit with a 10 Å radius. The calculations were set to dock the ligands 500 times (GA runs) and save the structure of the best 10 scoring conformations. When considering water/solvent presence in the pocket, the water molecules were enabled to move within a radius of 1 Å of their position in the crystallographic structure if the fitness function would consider this to be optimal for the docking procedure. Since in four of the five structures studied the reference ligand was a covalent inhibitor, the covalent bonds between ligands and proteins were also defined according to the GOLD User Guide. This means that in the protein input file the bond with the link atom of the ligand is preserved and the same is done for the ligand input. [87]

For the GBVI/WSA dG scoring function from MOE the covalent docking calculations were set using the chemical reactions already defined. For the epoxyketone inhibitors the ketalization reaction forming a hemi-ketal was employed and in the case of Bortezomib the transesterification reaction was used. The docking protocol was defined to place the ligands in 500 different poses refining the best 10 hits using the Rigid Receptor refinement method. For the non-covalent inhibitor, the placement used was the Triangle Matcher method (500 poses) and the refinement employed was the Rigid Receptor method to refine the 10 best hits from the placement protocol.

Docking using AutoDock 4.2 is parameterized in a different way from the GOLD calculations. Instead of defining an atom as the centre of a spherical search space, the space takes a box-like shape centred in a spatial point without any need for atoms to be

present. The search space, designated as grid box in AutoDock 4.2. is defined differently for each molecule docked since its size and centre should be optimized to be as small as possible and still accommodate the small molecules in the pocket. Another step required clearing before docking is running AutoGrid to calculate grid maps for the ligand, this will create a three-dimensional grid of regularly spaced points surrounding the grid box defined previously. In the case of covalent docking, is necessary to set up a covalent map which allows the scoring function to identify which protein atom is to establish covalent bonds with the ligand, the reactive oxygen of the catalytic threonine was chosen. Once AutoGrid has completed the creation of the grid maps everything is set to define the docking parameters. AutoDock provides four different docking algorithms which are employed to optimize the conformational space provided by the ligands. When using the AutoDock 4.2 scoring function the algorithm used was Genetic Algorithm (GA), the number of runs was set to 500 so that the calculation could be comparable to the already defined MOE and GOLD dockings, with the remaining settings left as default.

For docking validation in HDAC6, the docking parameters used were the same as for the immunoproteasome, the sole exception was in defining the centre of the search area. For HDAC6 the active site was defined centred on the zinc atom present int the pocket with a radius of 10 Å. Since the crystallographic structures of HDAC6 collected for this work did not have covalent ligands, the search parameters used were the same as the non-covalent immunoproteasome inhibitor Ro19.

This series of docking calculations will define which of the scoring functions being studied can describe the crystallographic ligands' poses with greater accuracy. Nevertheless, more information is needed before the docking protocol can be applied to virtual screening.

2.5.2 Cross-Docking

Having access to several crystallographic structures of the human immunoproteasome with different crystallographic ligands raises the question "where?" regarding a docking protocol, meaning there is a need to choose a proper structure where to run the screening. To this end Cross-Docking will be employed. Cross-Docking assays consist on docking crystallographic ligands on non-native crystallographic structures of the same macromolecule. When incubating protein with ligands in order to get a crystallographic

image, the presence of the small molecule on the active sites of the protein may change the protein three-dimensional structure in order for the pocket to better accommodate the ligand. This could induce some error on Self-Docking calculations and by performing Cross-Docking it's possible to check if the docked positions obtained from the previous test can be replicated without pocket shape influence.

From the Self-Docking assay two fitness functions were considered equivalent on how well they describe the interactions between the ligands and the immunoproteasome, ChemScore and GoldScore, therefore both will be used for Cross-Docking to see which, if any, of them stands out as more reliable fitness function to use for a docking protocol meant for virtual screening.

When it comes to docking parameterization, the settings were the same used as the settings for the Self-Docking assays.

2.5.3 Score *vs* IC_{50} Curve

This assay will provide insight on whether the docking protocol defined by the previous steps of validation can differentiate between ligands of higher and lower activities. For this assay, activities (IC_{50}) and the respective molecule structures will be taken from the ChEMBL database. For HDAC6 the activity assay used was ChEMBL1113906 with data for 53 molecules and IC_{50} values ranging from 20 to 630 nM.

For the immunoproteasome information from several assays had to be combined since each individual activity assay was too small (number of molecules) for a curve to be built (5 compounds for the largest). The assays ChEMBL3266189, ChEMBL2050572. ChEMBL3388364, ChEMBL3369588, ChEMBL3240089, ChEMBL3240088 and ChEMBL4003677 were combined for a total of 18 compounds with values of IC_{50} ranging from 1.2 to 100,000 nM.

For both proteins the docking protocols established before were applied with the databases retrieved from ChEMBL in order to plot a Score *vs* IC_{50} curve.

2.7 Virtual Screening

For docking-based Virtual Screening the docking calculations were programmed in similarity to the docking protocols of the validation step. The only differences were in the setting of GA runs which were decreased from the 500 used in the validation step to 50 and the Search Efficiency parameter that was changed from 100% to the pre-set for Virtual Screening of 30%. These parameter changes allow to run calculations on a larger number of compounds in more time efficient way.

2.8 Interaction Analysis

In order to analyse the interactions between the crystallographic ligands or docked compounds and the proteins the Protein-Ligand Interaction Profiler (PLIP) was used. [92] This tool allows for interaction analysis with little to no effort since no structure preparation is required for the input. PLIP's input is a Protein Data Bank structure, a protein or ligand name or any custom protein-ligand complex in PDB format file.

Table 3: Distance parameters for interaction detection by the PLIP algorithm. [92]

Variable	Description	Value
HBOND_DIST_MAX	Maximum distance between acceptor and donor in hydrogen bonds	4.1 Å
HYDROPH_DIST_MAX	Maximum distance between carbon atoms for hydrophobic interaction	4.0 Å
PISTACK_DIST_MAX	Maximum distance between aromatic ring centres for stacking	7.5 Å
PICATION_DIST_MAX	Maximum distance between charge and aromatic ring centres	6.0 Å
SALTBRIDGE_DIST_MAX	Maximum distance between two centres of charges in salt bridges	5.5 Å

The PLIP algorithm is able to identify or predict seven type of interactions based on geometric parameters like distance and angle between atoms or charges. Since the analysis of interaction on this work will be based on the nature and distance of the

interactions, this will be the prime parameter being considered for the interaction types detected on the crystallographic ligands and screened compounds (Table 3).

Chapter 3: Results and Discussion

3.1. Docking Validation: Immunoproteasome

3.1.1 Self-Docking

Starting with the analysis of the docking protocol validation. The first process was the Self-Docking assay to ascertain which scoring function would better describe the immunoproteasome and its crystallographic ligands and to also establish if there is any importance to the presence of water molecules in the active pocket.

Table 4: RMSD (Å) of the 5 best scoring poses of each crystallographic ligand on all the scoring functions for study. Grey highlights RMSD values below 2.3 Å. The empty cells represent a calculation not performed.

Top Score Pose		Score Function						
		ASP	PLP	Chem Score	Gold Score	GBVI/ WSA	AutoDock 4.2	AD Vina
Bortezomib	1	6.137	0.971	0.458	1.404	5.828		
	2	6.177	0.993	0.986	1.375	0.040		
	3	6.098	0.957	1.060	1.231	5.539		
	4	6.123	0.972	0.580	1.406	5.960		
	5	6.013	0.995	1.013	1.149	5.098		
Epoxiketone 9	1	1.316	1.455	8.281	1.879	2.673	2.525	7.248
	2	1.943	1.555	1.802	1.927	1.899	4.437	3.601
	3	1.308	1.739	1.542	1.874	2.557	4.513	4.345
	4	7.946	1.766	1.624	1.930	2.178	4.196	7.415
	5	7.899	1.375	8.160	2.061	3.169	4.156	8.914
Epoxiketone 14	1	1.630	2.056	8.007	1.424	12.391	6.146	2.245
	2	1.455	6.999	2.124	1.373	4.803	4.764	1.937
	3	1.578	2.202	1.977	1.398	5.096	7.315	5.317
	4	1.917	6.730	2.069	1.736	6.638	10.069	3.794
	5	1.499	2.016	1.883	1.255	7.234	10.017	4.939
PR-924	1	8.124	1.353	0.956	1.310	6.476	6.834	3.104
	2	8.099	1.340	1.383	1.396	6.395	2.557	4.952
	3	8.129	1.059	1.047	1.438	6.207	5.487	9.675
	4	8.287	1.405	1.635	1.941	6.927	12.672	5.929
	5	8.313	1.481	2.469	1.014	7.684	10.716	9.642

Table 4(continued): RMSD (Å) of the 5 best scoring poses of each crystallographic ligand on all the scoring functions for study. Grey highlights RMSD values below 2.3 Å.

Top Score Pose		Score Function						
		ASP	**PLP**	**Chem Score**	**Gold Score**	**GBVI/ WSA**	**AutoDock 4.2**	**AD Vina**
Ro19	1	2.350	2.483	2.220	2.688	1.703	2.496	1.324
	2	2.383	2.470	1.909	2.695	1.474	2.493	1.951
	3	1.755	2.477	2.078	2.832	11.050	2.493	1.999
	4	2.101	2.438	1.878	2.806	10.949	2.316	5.127
	5	2.339	2.471	1.856	2.674	10.436	2.116	9.756

From the results presented on Table for the Self-Docking assays without water molecules in the active sites, the scoring functions that stand out are ChemScore and GoldScore. Though GoldScore showed to be ineffective in reproducing the reference pose for the ligand Ro19, the same can't be said for the remaining ligands since the deviation of the docked poses are under the 2 Å threshold (the fifth best scoring pose for the Epoxyketone9 has a deviation above 2 Å, though since it is so lightly superior, it will be considered a good representation nonetheless).

Though ChemScore was capable in reproducing the poses of Ro19, in the cases of Epoxyketones 9 and 14 the top scoring pose presents far too high deviation (8.281 Å and 8.007 Å respectively) from the reference pose. Even if the best five scoring poses are being analysed at the moment, the best scored position has additional importance since for further validation assays and Virtual Screening only the best scoring pose is going to be analysed.

Table 5: RMSD (Å) of the 5 best scoring poses of each crystallographic ligand on all the scoring functions for study with water molecules left in the active pocket. Grey highlights RMSD values below 2.3 Å. The empty cells represent a calculation not performed.

Top Score Pose		Score Function						
		H2O ASP	**H2O PLP**	**H2O Chem Score**	**H2O Gold Score**	**H2O GBVI/ WSA**	**H2O AutoDock 4.2**	**H2O AD Vina**
Bortezomib	1	0.862	1.280	1.237	1.321	7.387		
	2	0.622	1.286	0.828	1.113	5.769		
	3	1.163	1.282	1.359	1.066	5.680		
	4	1.203	1.283	0.871	1.021	2.150		
	5	1.157	1.283	1.195	0.767	5.614		
Epoxiketone 9	1	1.518	1.596	2.017	1.877	3.234	10.553	3.176
	2	1.964	1.474	1.670	1.890	3.629	3.198	2.821
	3	3.426	1.555	1.906	1.881	4.295	9.325	6.251
	4	1.481	1.361	1.579	1.894	3.779	2.775	5.693
	5	1.859	1.360	1.831	1.892	2.470	10.974	3.527
Epoxiketone 14	1	1.566	1.747	1.461	1.672	8.275	4.841	2.225
	2	1.295	1.677	1.778	1.672	6.235	5.475	5.507
	3	1.595	1.623	2.014	1.640	2.710	9.865	8.650
	4	1.294	2.361	1.521	1.152	6.435	10.483	5.300
	5	0.922	1.704	1.585	1.284	5.216	10.086	6.960
PR-924	1	8.129	1.348	0.772	1.430	1.800	2.421	3.114
	2	1.362	1.330	2.621	1.347	3.965	3.112	4.965
	3	8.426	1.406	1.430	1.319	4.226	6.363	10.092
	4	1.851	0.932	0.863	1.301	2.259	3.599	5.951
	5	1.915	1.364	1.808	1.318	5.227	3.354	5.966

From the results presented in Table 5 it will be possible to understand the importance of the water molecules for the representation of the crystallographic poses in the Self-Docking assay.

The main difference brought by the presence of the water molecules in the active pocket lies in the representation of Epoxyketone14. The position and quantity of solvent molecules vary from one crystallographic structure to another (the crystallographic structure 5M2B, with Ro19 as the ligand, presents no water molecules near the active site and that's why its missing from Table 5) and as such it stands to reason that it may have some effect on the pose taken by the reference ligand.

The presence of water molecules allows for the most of the crystallographic poses to be better replicated by all scoring functions from GOLD with greatly increased accuracy by

the ASP scoring function. PLP and ChemScore have also improved in Epoxyketone14 pose replication, again establishing the effect the water molecules have on the docked positions. On another note the GoldScore function results were not affected by the presence of water molecules in the active site, which means that GoldScore is able to replicate the reference poses without the need of water molecules working somewhat like a "crutch".

All in all, from the docking assay the scoring functions that stand out from the seven studied are GoldScore and ChemScore. GoldScore has shown to be able to replicate ligand poses in four out of five crystallographic structures with little deviation. ChemScore though presenting some difficulty in reproducing the reference pose in its best scoring position in two structures, was able to produce decent poses in the complex where GoldScore fails to do so.

3.1.2 Cross-Docking

The Cross-Docking assay allows to determine on which crystallographic structure the docking-based Virtual Screening should be done. When considering that the protein may undergo structural change during the process of X-ray crystallography due to the presence of ligands and solvent, this assay can help determine which protein structure can reproduce the crystallographic poses of the ligands best.

During the Self-Docking assay two scoring functions (ChemScore and GoldScore) were considered noteworthy and so, this assay will also serve the purpose of deciding on which of these two functions should be used for the Virtual Screening.

Table 6: Cross-Docking with the ChemScore function. A table presenting the RMSD (Å) of the best scoring pose of each ligand on each crystallographic structure. Grey highlights RMSD values below 2.3 Å; numbers in bold are used for RMSD values from the Self-Docking assay.

ChemScore					
Protein PDB ID	**Crystallographic Ligand**				
	Bortezomib	Epoxiketone9	Epoxiketone14	PR-924	Ro19
5L5F	**0.458**	1.563	5.387	0.855	3.573
5L5I	1.239	**8.281**	5.401	0.846	1.897
5L5J	1.126	1.640	**8.007**	8.686	1.855
5L5H	1.131	1.594	5.018	**0.956**	1.514
5M2B	6.241	2.180	5.166	1.628	**2.220**

The ideal result in this assay is a crystallographic structure that can replicate reference poses with a RMSD lower than 2 Å on the best scoring position. When analysing the deviations in Table the ligand Epoxyketone14 stands out has being particularly complicated to reproduce its reference pose since in all structures the deviation is far greater than the 2 Å threshold. That said, with the ChemScore function two structures (5L5I and 5L5H) can be considered optimal due to only failing to replicate the troublesome Epoxyketone14.

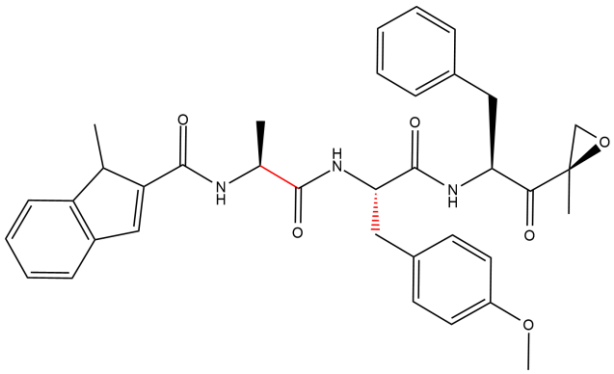


Figure 23: Structure of crystallographic ligand Epoxyketone14. [93] The red highlight represents two bonds important for discussion.

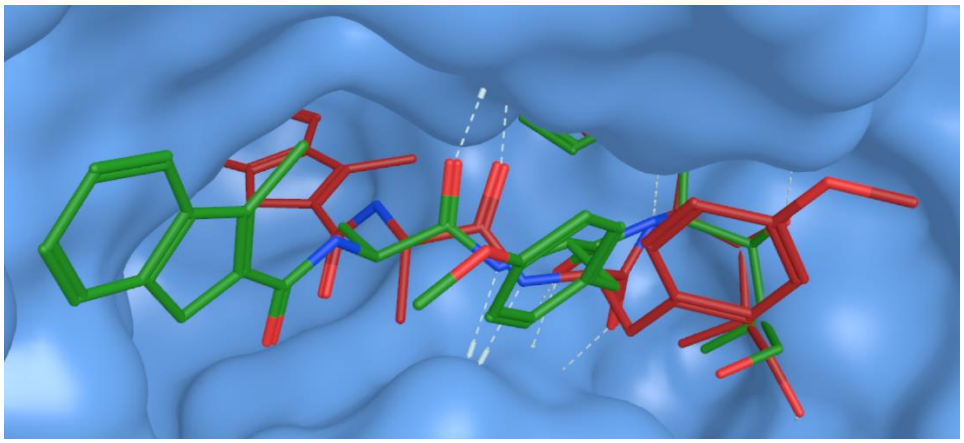


Figure 24: Epoxyketone14 in 5L5I for Cross-Docking (red) and Epoxyketone14 reference pose (green).

The difficulty in reproducing the reference pose of Epoxyketone14 lies in the two highlighted bonds on Figure 23. Single bonds are free to rotate and if the active pocket allows for the molecule to fit in different areas (Figure 24), the docking protocol is likely to fit them in several ways and then rank them according to the scoring function. In the case of the ChemScore function, there appears to be no reason for the docked pose to have a lower score, since the different position does not create any new "penalizing" occurrence for ChemScore, like new covalent bonds between ligand and protein.

Table 7: RMSD (Å) of the best scoring pose of each ligand on each crystallographic structure. Grey highlights RMSD values below 2.3 Å; numbers in bold are used for RMSD values from the Self-Docking assay.

	GoldScore				
Protein PDB ID	**Crystallographic Ligand**				
	Bortezomib	Epoxiketone9	Epoxiketone14	PR-924	Ro19
5L5F	**1.404**	1.747	4.935	1.472	1.345
5L5I	1.499	**1.879**	5.111	1.336	2.229
5L5J	1.074	1.528	**1.424**	1.852	2.188
5L5H	1.470	1.914	5.108	**1.310**	2.019
5M2B	3.710	1.678	6.030	1.451	**2.688**

Unlike what could be observed in the results obtained with ChemScore, with GoldScore there is a structure where all the crystallographic ligands were able to have their reference poses replicated with an acceptable level of deviation.

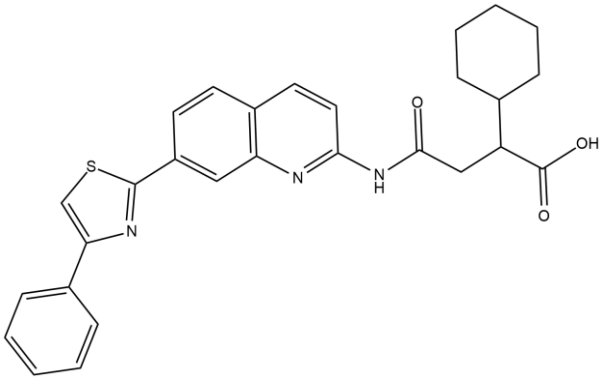


Figure 25: Structure of crystallographic ligand Ro19. [94]

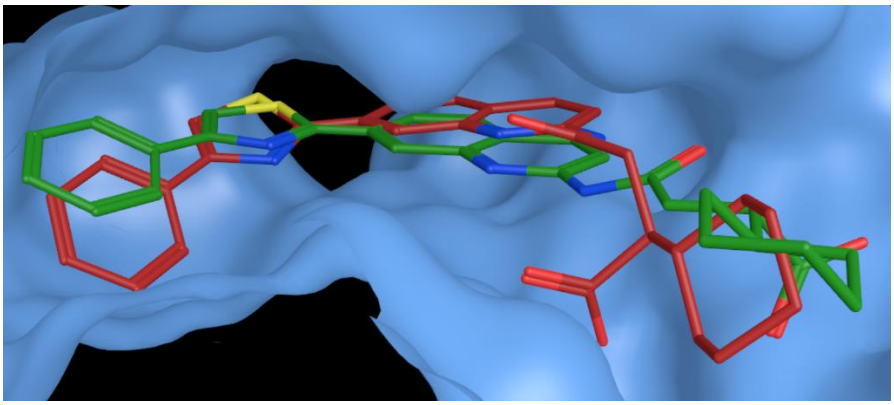


Figure 26: Ro19 in 5L5J for Cross-Docking (red) and Ro19 reference pose (green).

Though the inhibitor Ro19 (Figure 25) had its pose replicated with a deviation surpassing the 2 Å threshold (2.188 Å) it is noteworthy that it was not by a large margin. Besides that, when comparing the best scored position of Ro19 with the crystallographic position, it becomes clear that most atoms are somewhat deviated from the reference pose and apart from the carboxyl moiety that are facing opposite directions, each part of the docked molecule occupies the same section of the pocket as the reference pose, as shown in Figure 26.

From the Cross-Docking assay it can be concluded that the most appropriate scoring function for using in a docking-based Virtual Screening is GoldScore and the better structure where to run it is 5L5J.

3.1.3 Score *vs* IC_{50}

This assay allows to determine if the scoring function is able to distinguish inhibitors of the immunoproteasome based on their activity. The ideal correlation is for the higher the score given by the scoring function to correspond to the lowest value of IC_{50}. The molecules used and their respective values of IC_{50} were obtained from the ChEMBL database. The scores were plotted against the values of $\ln(IC_{50})$ in order to better get a linear correlation between the data sets.

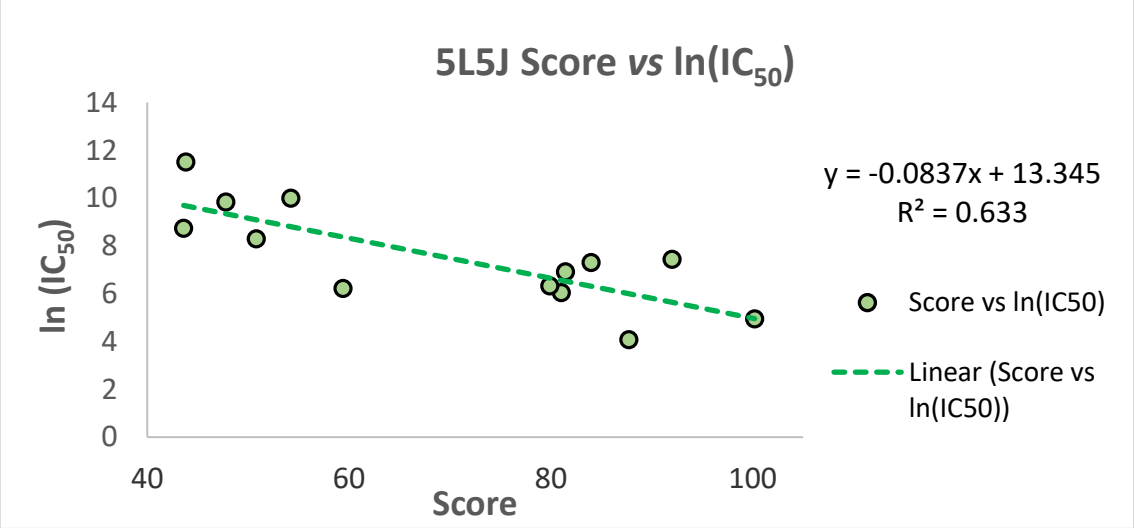


Figure 27: Scatter-plot of Score *vs* $\ln(IC_{50})$ from inhibitors and activities from ChEMBL.

Looking at the Score *vs* $\ln(IC_{50})$ plot, the first thing noticeable is that the prefered correlation type between docking score and activity has been verified. The tendency line equation translates this to the negative value of the x multiplying factor (-0.0837). Though some of the data points may not follow the tendency line equation ($R^2 = 0.63$), there are clear point agglomerates of lower score and higher $\ln(IC_{50})$ and higher score and lower $\ln(IC_{50})$. These point agglomerates lead to the conclusion that even if the correlation between docking score and activity is not precisely described by the linear regression obtained, the docking scores are able to differentiate between low and high activity inhibitors.

3.2. Docking Validation: HDAC6

3.2.1 Self-Docking

In likeness to the previous section, the first step in establishing validated docking protocol for the HDAC6 will be to determine which of the scoring functions will best describe the crystallographic pose.

Table 8: RMSD (Å) of the 5 best scoring poses on all the scoring functions for study on Human HDAC6 structure (5EDU) without water molecules considered. Grey highlights a deviation under 2.3 Å.

Top Score Pose	Score Function						
	ASP	**PLP**	**ChemScore**	**GoldScore**	**GBVI/ WSA**	**AutoDock 4.2**	**AD Vina**
1	5.638	1.373	5.509	6.099	1.603	2.848	2.757
2	6.291	1.184	5.558	6.097	2.105	2.078	2.690
3	6.238	1.187	3.301	5.563	1.516	1.923	2.102
4	5.668	1.211	3.141	6.081	1.394	2.032	0.983
5	5.581	1.149	5.355	5.727	3.549	3.594	3.029

The results of this assay are a perfect example of the reason why docking validation needs to be done on every different structure. The self-docking assay on the immunoproteasome distinguished the ChemScore and GoldScore functions from the seven functions in study. When repeating the assay for HDAC6 the scoring function which stands out is PLP. Though GBVI/WSA replicated the reference pose of Trichostatin A to an acceptable degree, unlike the immunoproteasome crystallographic ligands, the deviation of the PLP docked poses as undeniably lower.

Table 9: RMSD (Å) of the 5 best scoring poses on all the scoring functions for study with water molecules left in the active pocket. Gray highlights RMSD values below 2.3 Å.

Top Score Pose	Score Function						
	H2O ASP	**H2O PLP**	**H2O ChemScore**	**H2O GoldScore**	**H2O GBVI/ WSA**	**H2O AutoDock 4.2**	**H2O AD Vina**
1	5.649	2.939	1.814	6.070	1.636	2.649	2.735
2	5.604	1.531	5.663	6.070	3.609	3.317	2.177
3	6.163	1.422	2.992	5.594	1.026	2.026	2.162
4	5.738	0.856	5.755	6.040	1.168	2.380	1.848
5	5.543	1.463	3.056	6.066	2.568	1.665	3.150

Analysing the self-docking with the water molecules on the active pocket, PLP failed to replicate the reference pose on its best scoring position while ChemScore's deviation for the best scoring pose is under the 2 Å threshold.

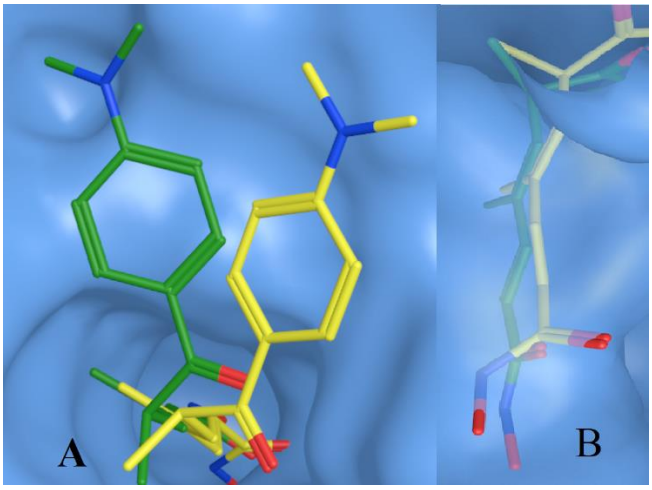


Figure 28: Trichostatin A pose in HDAC6 active site. A) view from outside the oxanion hole; B) view of the inside of the oxanion hole. The yellow molecule is the reference position and the green molecule is the best scoring position by PLP with water in the active site.

Despite the best scored position deviation in the presence of water molecules, PLP is still the most accurate of the seven scoring functions on this assay, with GBVI/WSA as the runner-up. Though the deviation of the top scoring position to the reference ligand pose is above the 2 Å threshold, the poses are not very different from one another (Figure 28). For the aromatic moiety of Trichostatin A, the PLP docked pose is not overlapped with the crystallographic pose, nevertheless they are not very far from each other.

The inside of the oxanion hole shows that though the zinc-binding hydroxamic moiety is facing the same way in both poses, suggesting that the docked position would still be able to coordinate to the zinc cation in the same manner than the crystallographic pose. This stands to support PLP as the scoring function that best describes the HDAC6-Trichostatin A complex and why that is the scoring function to be used when using molecular docking on the HDAC6.

Table 10: RMSD (Å) of the 5 best scoring poses on all scoring functions from GOLD on *Danio Rerio* structures.

		Score Function			
Top5 Poses		**ASP**	**PLP**	**ChemScore**	**GoldScore**
Acy-1083	1	3.700	7.694	6.205	3.793
	2	3.970	6.466	6.242	3.825
	3	3.780	7.491	6.304	3.779
	4	3.977	7.508	6.187	3.791
	5	3.602	7.481	6.300	3.703
Vorinostat	1	4.955	3.276	8.418	8.052
	2	4.051	4.971	8.353	7.999
	3	4.758	3.209	8.491	7.979
	4	4.802	4.898	8.033	7.977
	5	2.597	3.391	8.337	8.018
Ricolinostat	1	7.808	7.591	9.772	10.511
	2	7.793	8.507	9.319	7.988
	3	7.453	7.332	9.244	6.948
	4	8.369	7.080	9.099	7.457
	5	7.767	7.516	9.817	7.706
SS208	1	5.857	5.523	5.665	7.836
	2	6.064	5.536	8.838	7.979
	3	5.788	5.536	5.330	8.007
	4	6.120	5.476	5.670	8.096
	5	5.810	5.517	5.684	8.144

As established before, *Danio Rerio* HDAC6 shares high homology with its Human counterpart and so some crystallographic structures from the zebrafish will be used as "proxies" of sorts for a largest pool of crystallographic information. The Self-Docking calculations were firstly done on the GOLD scoring functions since it has been shown that these four functions were better suited for replicating the crystallographic poses for Immunoproteasome and Human HDAC6 ligands.

As seen in Table , neither of the scoring functions was able to even remotely replicate the reference pose from the crystallographic structure. This could be due to the HDAC6 pocket being wide with a considerable amount of water molecules that could be stabilising the reference pose.

Table 11: RMSD (Å) of the 5 best scoring poses on all scoring functions from GOLD on *Danio Rerio* structures with water molecules on the pocket. Grey highlights RMSD values lower than 2.3 Å.

	Top5 Score	Score Function			
		H2O ASP	H2O PLP	H2O ChemScore	H2O GoldScore
Acy-1083	1	3.782	7.471	6.231	4.148
	2	4.094	7.456	6.395	4.094
	3	3.573	7.387	6.527	4.021
	4	3.715	7.319	6.208	3.998
	5	3.527	7.340	6.228	4.148
Vorinostat	1	5.003	4.978	7.965	8.004
	2	4.991	5.051	8.072	8.001
	3	4.776	4.964	8.006	8.043
	4	4.804	4.934	8.077	8.001
	5	4.811	4.823	8.165	8.021
Ricolinostat	1	7.879	3.775	9.679	7.708
	2	7.536	8.096	9.488	7.795
	3	7.625	7.864	9.431	8.089
	4	7.911	10.453	9.861	8.329
	5	8.242	9.888	9.729	7.840
SS208	1	5.964	5.100	7.818	7.973
	2	5.955	1.232	2.186	7.988
	3	5.788	5.139	2.194	7.929
	4	5.793	2.086	7.666	7.998
	5	5.767	2.191	7.600	8.021

In the case of crystallographic inhibitor SS208 the presence of solvent molecules in the pocket had some effect as seen in Table , seen as for the PLP and ChemScore functions, 3 out of the 5 best scored poses had acceptable deviation from the reference pose. For these calculations with water molecules in the pocket not all water molecules were kept, the molecules interacting with the crystallographic ligand or that were the closest were the ones left for the docking calculations, and all preserved waters were allowed some flexibility of movement. This leads to the conclusion that not only is the presence of water

molecules in the pocket that help the ligand take the crystallographic pose but the amount of water molecules as well that allowed for the crystallographic ligand to take its pose.

3.2.2 Cross-Docking

Since Self-Docking with the structures from *Danio Rerio* was improper for validation of the docking protocol with Cross-Docking was different. The crystallographic ligands from the *Danio Rerio* structures were docked on the Human HDAC6 to see if it's possible to replicate the reference poses on the Human macromolecule using ChemPLP, the scoring function used

Table 3: RMSD (Å) of the best five scoring poses of Danio Rerio crystallographic ligands on Human HDAC6 (5EDU) crystallographic structure without water molecules considered.

PDB ID - Ligand	Top5 poses	RMSD (Å)	PDB ID - Ligand	Top5 poses	RMSD (Å)
5EEI - Vorinostat	1	5.022	6R0K - SS208	1	5.721
	2	5.270		2	5.613
	3	5.032		3	5.653
	4	5.362		4	5.500
	5	5.158		5	5.300
5WGL - Ricolinostat	1	8.748	5WGM - ACY-1083	1	4.070
	2	9.226		2	4.044
	3	9.174		3	3.986
	4	9.142		4	3.974
	5	9.176		5	3.899

From the results presented in Table 3 what can be seen is that even in the Human HDAC6 the crystallographic poses from Danio Rerio structures were not replicated. This leads to the conclusion that the macromolecule itself is not the reason for the deviations observed but the reason lies in the presence and amount of water molecules in the pocket.
As a result, from these validation steps, the docking protocol to be applied with Human HDAC6 should be on the 5EDU-Trichostatin A structure and run on ChemPLP scoring function.

3.2.3 Score *vs* IC_{50}

The same analysis must be done for the HDAC6 as was done with the immunoproteasome. Since the best scoring function has been determined, it must be tested with reference inhibitors of the human HDAC6 to see if it can distinguish between inhibitors with higher or lower activity through higher or lower scoring.

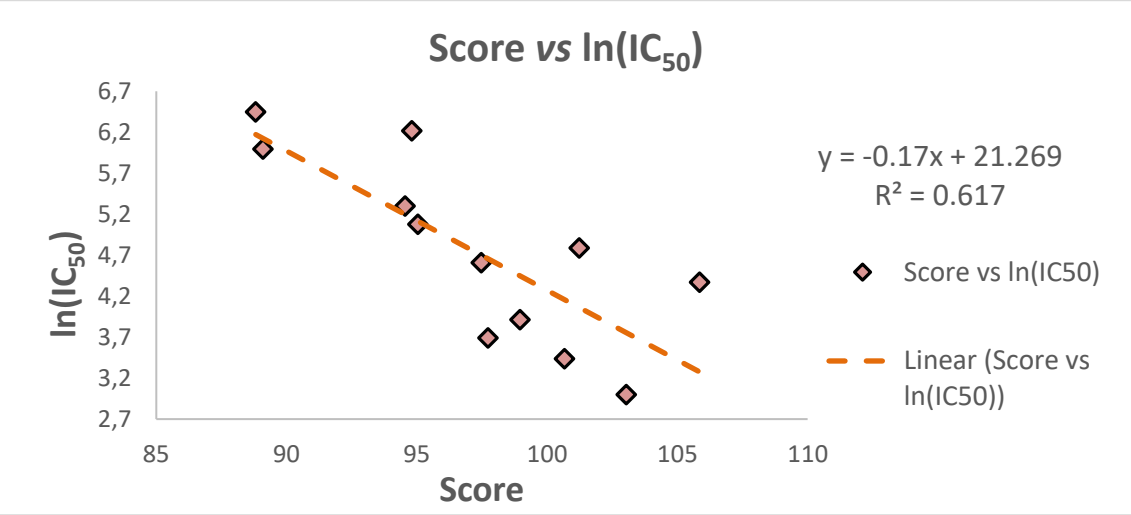


Figure 29: Scatter-plot of Score *vs* ln(IC_{50}) from inhibitors and activities from ChEMBL1113906.

Unlike with the immunoproteasome, the plot for score vs IC_{50} as presented in Figure 29 doesn't show agglomerates of data points. This can be the result of the array of registered activities being smaller for the HDAC6 (20-630 nM) than for the immunoproteasome (59-22.000 nM).

That said, the docking protocol was still able to wield the same sort of relation between score and IC_{50} (lower IC_{50} corresponding to higher value of score) shown through the x multiplying factor (-0.17). The linear trendline equation describes the data set to the same degree as with the immunoproteasome (R^2 = 0.617) leading to the conclusion that this docking protocol is able to describe the HDAC6-Inhibitors system to an agreeable degree and will therefore be used for Virtual Screening.

3.3. Virtual Screening

Since docking protocols have been established for the Immunoproteasome and HDAC6, Virtual Screening of a compound database can be done on both targets. The DrugBank approved molecule database was used for this purpose.

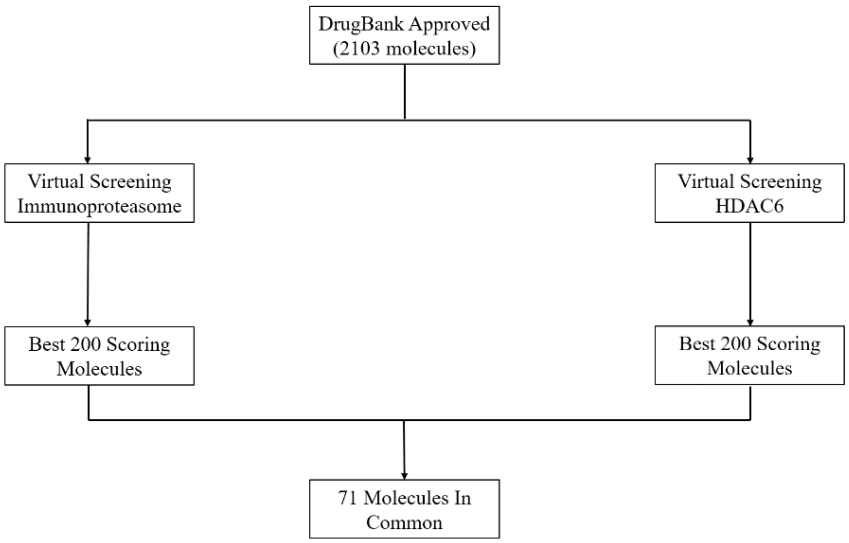


Figure 30: Virtual Screening analysis workflow.

A docking-based Virtual Screening was run on both therapeutic targets according to the docking protocols established during the validation process. The 200 best scoring molecules on each protein were selected with scores ranging from 116.46 to 78.56 for the immunoproteasome and 128.30 to 79.83 for HDAC6. Before analysing the interactions between the screened molecules to better ascertain their possible inhibitory abilities, the 200 best scoring lists were checked for molecules in common between them and 71 were found. Since the purpose of this work is to investigate the possibility of dual inhibition of immunoproteasome and HDAC6, compounds with a high score on both proteins are of great interest.

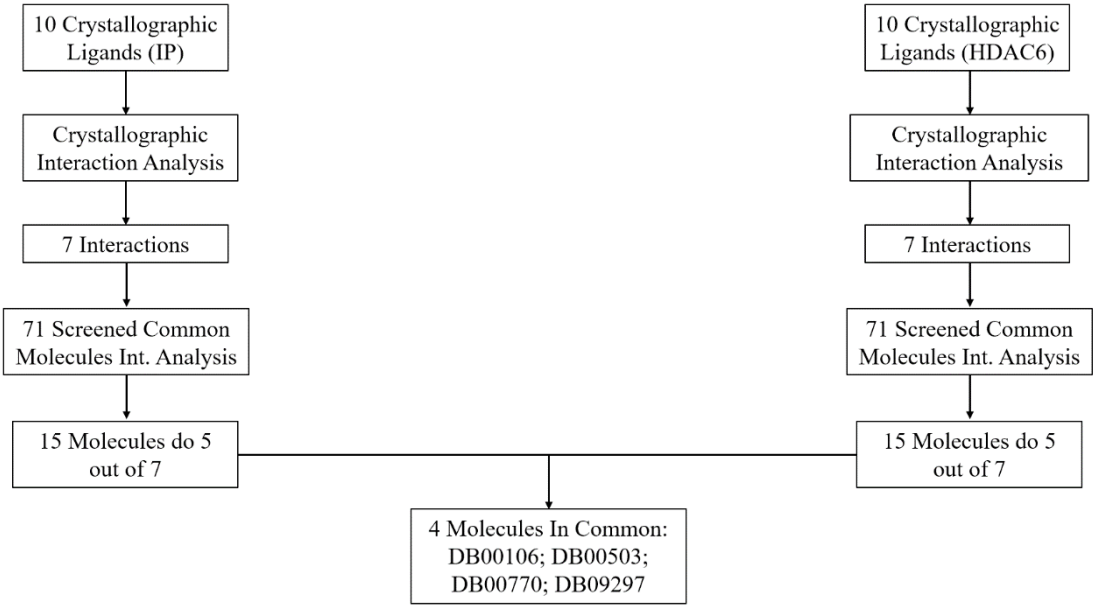


Figure 31: Ligand-Protein Interaction workflow.

Before jumping into analysis of the screened compounds, the crystallographic ligands need to be investigated to see which interactions may be of greater importance for inhibition. The 10 crystallographic structures used in this work were analysed and interactions with any residue seen in 8 out of the 10 reference ligands was deemed important to analyse the screening results. 7 residues in each protein were found according to these parameters. Of the 71 screened molecules that were common in the best 200 scores of both screenings, 15 molecules were found that established interactions with 5 out of the 7 main residues for immunoproteasome and for HDAC6. On those 15 compound lists, 4 were common between them.

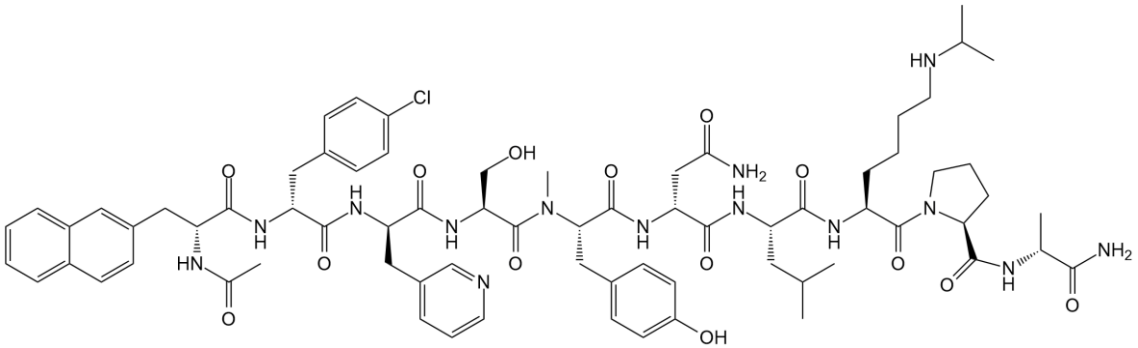


Figure 32: Structure of DB00106 (Abarelix). [95]

Compound DB00106 is Abarelix, sold under the name Plenaxis™ in Germany and the Netherlands is a gonadotropin-releasing hormone antagonist and is primarily used in oncology as it reduces the levels of testosterone produced in patients with advanced prostate cancer. Abarelix was discontinued from the United States market due to being a commercial let-down with high incidence of allergic reactions. [96]

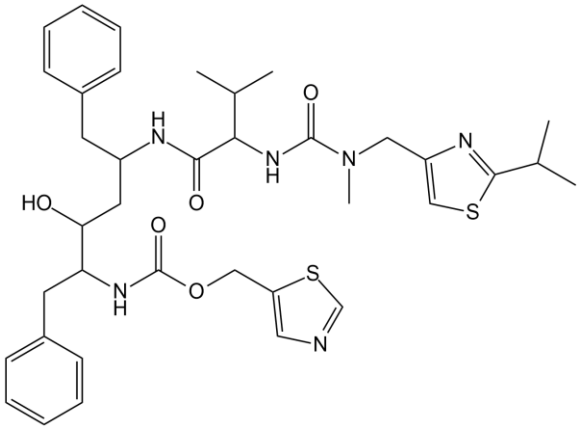


Figure 33: Structure of DB00503 (Ritonavir). [97]

Compound DB00503 is Ritonavir, sold under the name Norvir™, is an HIV protease inhibitor that modulates the reproductive cycle of HIV. Though firstly it was designed as an independent antiviral, its more commonly used as a booster of other protease inhibitors. It has also been reported that even though Ritonavir is not active against the Hepatitis C virus (HCV) is used in combination with HCV antivirals since it is also a CYP3A inhibitor, increasing plasma drug concentration. [97-98]

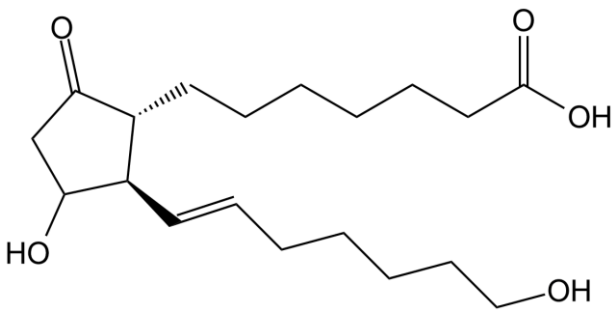


Figure 34: Structure of DB00770 (Alprostadil). [99]

Compound DB00770 is Alprostadil, a vasodilator used in palliative therapy increasing pulmonary blood flow in infants until surgery can be performed in neonates with congenital heart defects. In adults is used for the treatment of erectile dysfunction. [99-100]

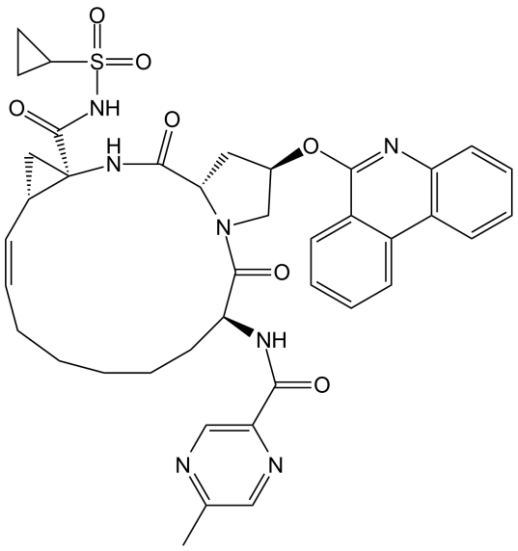


Figure 35: Structure of DB09297 (Paritaprevir). [101]

Compound DB09207 is Paritaprevir, a component of the drug cocktail sold as Viekirax™ in Europe for the treatment of HCV. Paritaprevir is a viral protease inhibitor that targets the viral Nonstructural Protein 3 and its activating cofactor Nonstructural Protein 4A (NS3/NS4A). By inhibiting these targets, Paritaprevir hinders the viral protease function of generating non-structural proteins required for virus maturation and therefore, prevents viral replication and function. [101-102]

3.3.1 Crystallographic Interactions

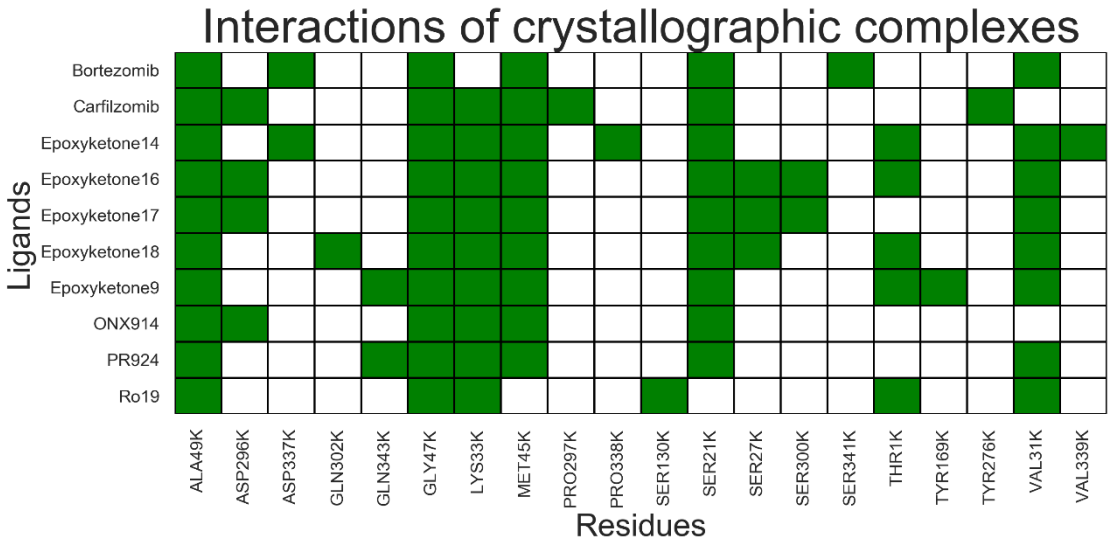


Figure 36: Heatmap representing all interactions of crystallographic ligands with Immunoproteasome's β5 sub-unit. Green is used to mark the interactions established between the protein and the ligands.

Through Figure 36 is simple so see which residue interactions are more recurrent in immunoproteasome β5 crystallographic structures. The residues that stand out the most are ALA49, GLY47, LYS33, MET45. SER21 and VAL31 since at least 8 out of the 10 interact with these residues. The catalytic THR1 will also be considered a residue to consider important despite only half of the reference ligands establishing interactions with this residue. This can be justified due to most of these ligands (only exception is Ro19) are covalent ligands and therefore are bound to the catalytic residue and covalent bonds are not picked up as interactions by the PLIP algorithm at its current version.

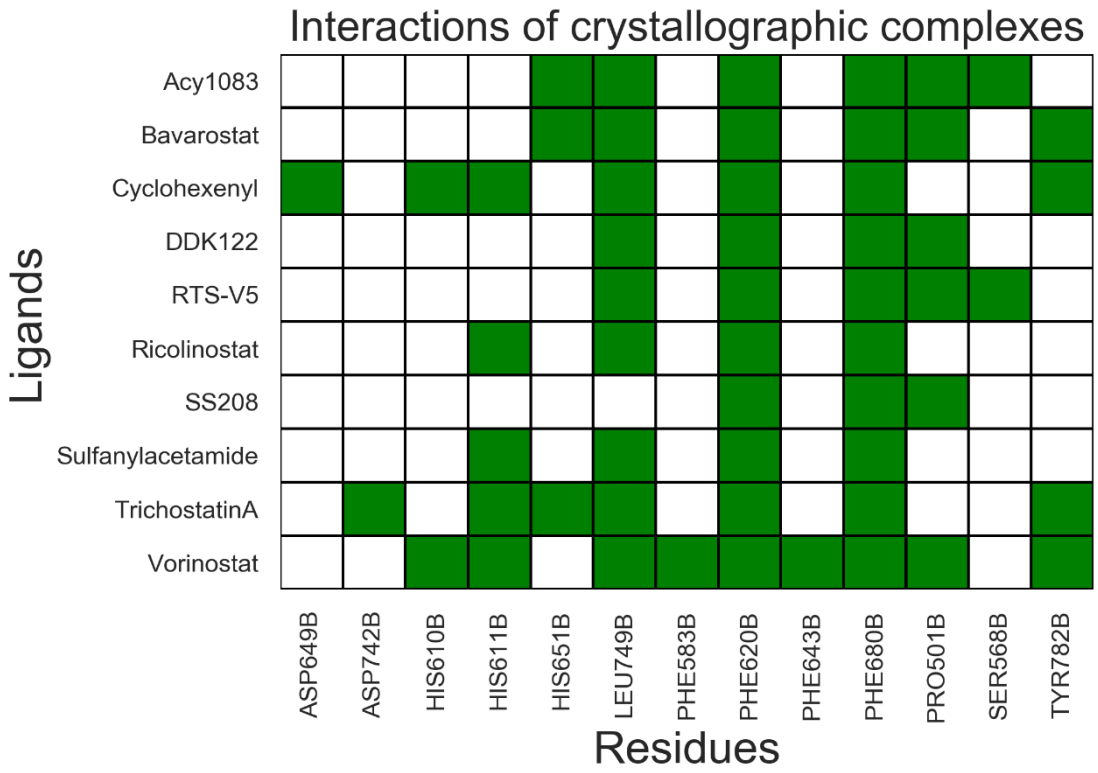


Figure 37: Heatmap representing all interactions of crystallographic ligands with HDAC6. Green is used to mark the interactions established between the protein and the ligands.

Analysing Figure 37 allows to easily identify which residues are most important in HDAC6 inhibition. The residues that stand out are LEU749, PHE620 and PHE680 since at least 9 out of the 10 crystallographic ligands are shown to interact with them. Zn-dependant HDACs exert their catalytic activity through a pair of HIS residues that promote the de-acetylation of acetyllysine groups inside their pocket (shown previously in Figure 15). For this reason, the residue pair HIS610 and HIS611 will also be considered important for analysis of docked molecules. Though HIS652 is not present in the crystallographic interactions, its neighbour HIS651 is, and so, these two residues will be useful for analysing the results of the virtual screening. An interaction that would be of interest to analyse is with the Zn cation in HDAC6, however this was not possible since metal coordination, like covalent bonds, is not parameterized in the PLIP algorithm and therefore the Zn atom was listed as a ligand and the only interactions registered were between the metal atom and the protein.

3.3.2 Screened Compound Interactions: Immunoproteasome

Since the most recurring residue interactions in the crystallographic ligands have been established, a comparison between them and the screened compounds is now possible. The four compounds that exhibited interactions with most of the important residues in both proteins are the DrugBank entries DB00106 (Abarelix), DB00503 (Ritonavir), DB00770 (Alprostadil) and DB09297 (Paritaprevir).

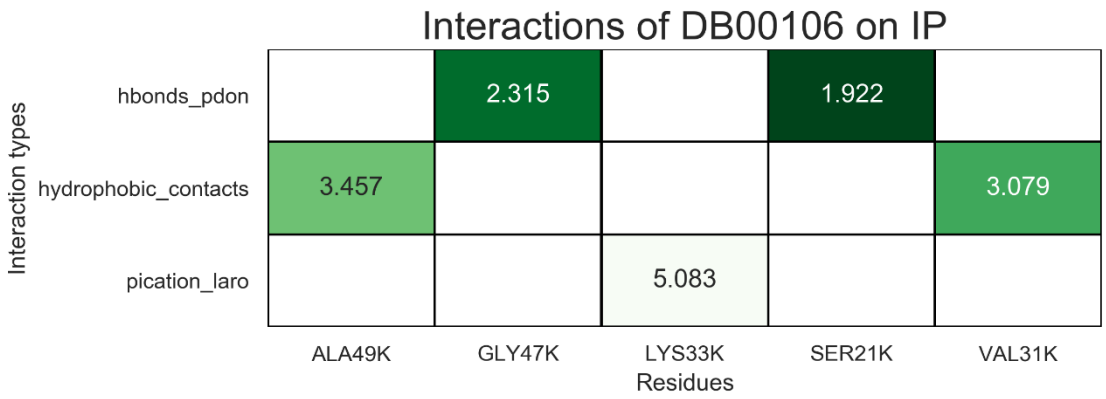


Figure 38: Heatmap representing the important interactions of DB00106 (Abarelix) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligands. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

Analysing the heatmap presented in Figure 38 with the heatmaps representing the crystallographic ligand's interaction with the immunoproteasome in Anex I what can be established is that the H-bonds between the ligand and residues GLY47 and SER21 are of similar distance as the reference interactions of the same type and with the same residues. Though reference ligands all show a pair of H-bonds with these residues (protein donating H and the ligand donating the H) DB00106 only establishes one interaction with the protein being the H donor. The π-cation interaction with LYS33 is not in the list of reference interactions, the crystallographic interactions mainly interact with LYS33 through hydrophobic contacts.

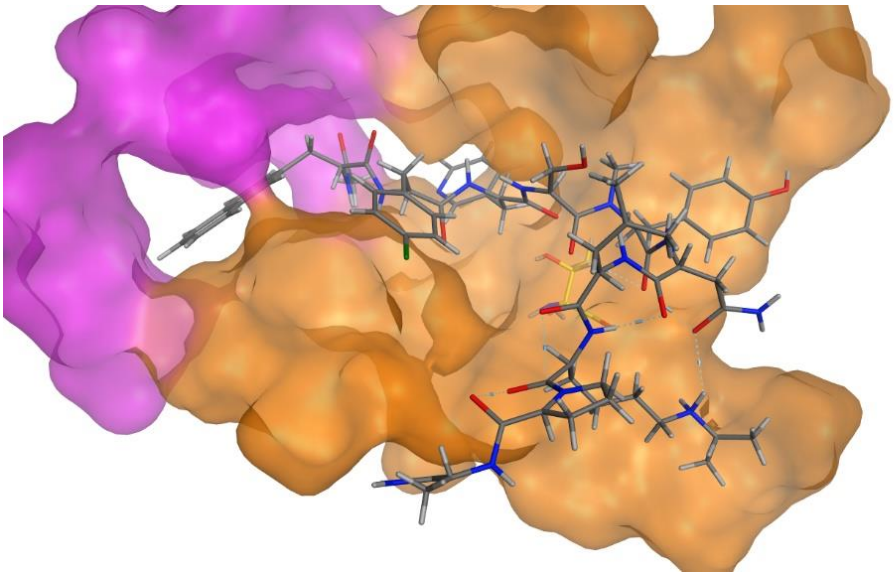


Figure 39: DB00106 (Abarelix) in immunoproteasome β5 pocket. The orange section of the pocket is the β5 sub-unit; the purple region is the β6 sub-unit; the yellow residue in the β5 is the catalytic threonine residue.

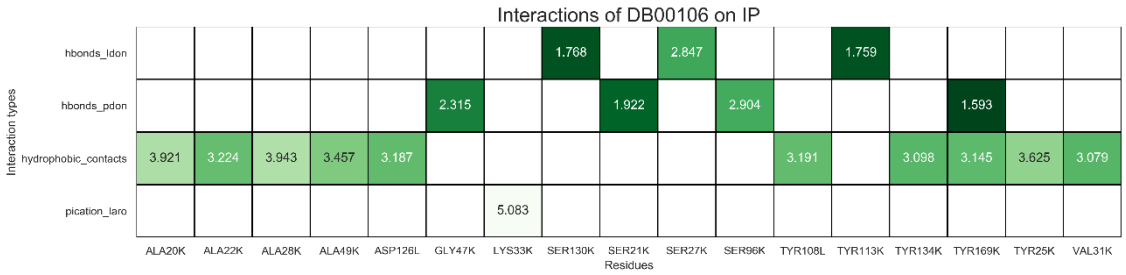


Figure 40: Heatmap representing all interactions of DB00106 (Abarelix) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligands. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

When seeing DB00106 in the immunoproteasome β5 pocket (Figure 39) its easy to see that the molecule occupies a large portion of the actvive site encompassing both sub-units. In order to provide some context to the interactions established with the important residues, all the interactions between DB00106 and the immunoproteasome (Figure 40) are being analysed.

DB00106 being a large molecule (molecular weight =1416.09 $g \cdot mol^{-1}$) is expected to fill up the active site and interact with a plethora of residues, this does not necessarily mean that it will be a proper immunoproteasome inhibitor. On the docking-based Virtual

Screening Abarelix had a score of 89.01 and consequently a ligand efficency of 0.88 which in turn mean the high score is due to its size and sheer number of interactions with the protein. Figure 40 allows to see all the interactions between DB00106 and the immunoproteasome and the major difference between the screened compound and the reference is the number of interactions. DB00106 interacts with 17 residues for a total of 18 interactions established with β5 and β6 sub-units of the immunoproteasome as for the reference ligand with the most interactions (Epoxyketone14) establishes 14 interactions over 10 residues.

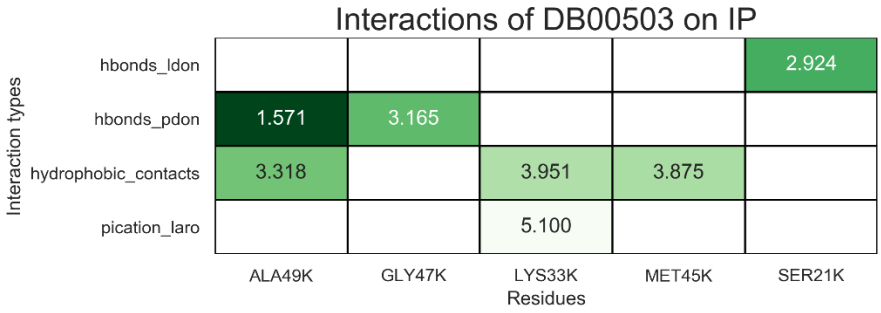


Figure 41: Heatmap representing the important interactions of DB00503 (Ritonavir) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

The interactions established by DB00503 (Ritonavir) with the focal residues in immunoproteasome β5 sub-unit inhibition are very similar to the crystallographic interactions. Firstly, the H-bond with ALA49 is shorter with DB00503 than with the crystallographic residues. On another note, the H-bond with SER21 though longer than in most reference ligands, is similar to Carfilzomib's, the longest of the reference ligands. The H-bond with GLY47 is longer than the reference interactions of the same type with the same residue. The π-cation interaction with LYS33K is not registered in the data from the crystallographic ligands and despite the possible importance of this interaction for inhibition, an additional interaction likely to help stabilise and strengthen the hydrophobic interactions with LYS33.

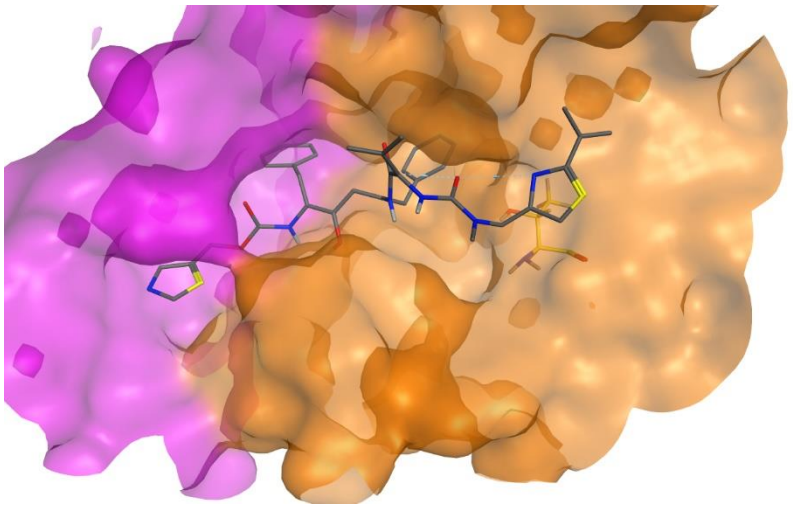


Figure 42: DB00503 (Ritonavir) in immunoproteasome β5 pocket. The orange section of the pocket is the β5 sub-unit; the purple region is the β6 sub-unit; the yellow residue in the β5 is the catalytic threonine residue.

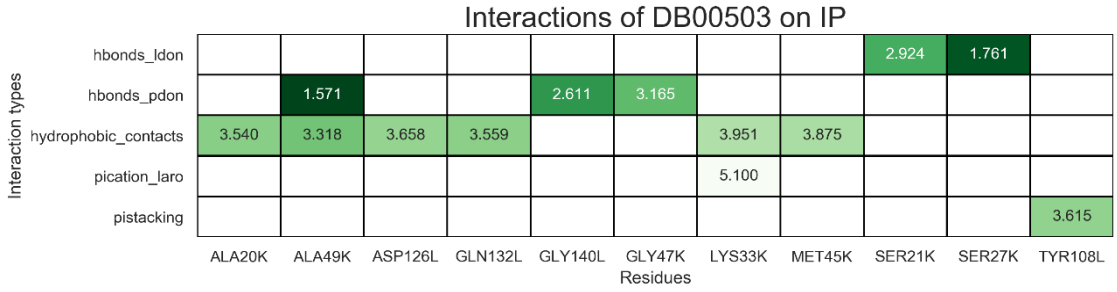


Figure 43: Heatmap representing all interactions of DB00503 (Ritonavir) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligands. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

From Figure 42 what can be seen is that, unlike DB00106 above, DB00503 does not occupy a large portion of the solvent accessible area of the active site. On the virtual screening DB00503 had a score of 93.50 with a ligand efficiency of 1.87 which verifies that DB00503 is a more efficient ligand than DB00106 for the immunoproteasome. Consequently, the number of residues it interacts with is much closer to the crystallographic ligands since DB00503 interacts with 11 residues for a total of 13 interactions.

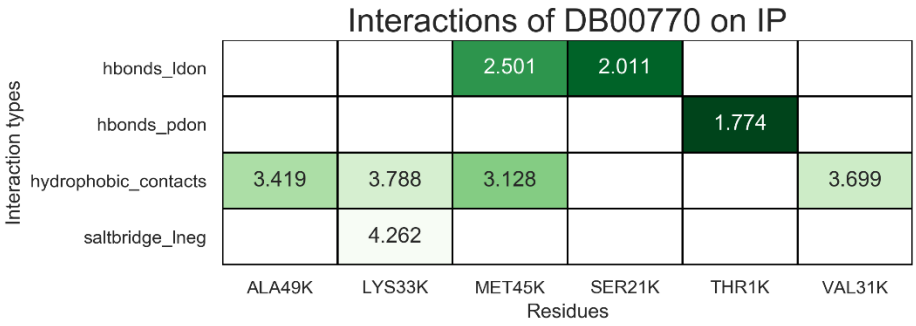


Figure 44: Heatmap representing the important interactions of DB00770 (Alprostadil) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

DB00770 is shown in Figure 44 to interact with the catalytic THR1 through a H-bond with shorter distance than reference ligands that interact with this residue through the same interaction type (Epooxyketone9, Epoxyketone16, Epoxyketone18 and Ro19). The H-bond with SER21 is shorter for DB00770 than for the crystallographic ligands, however reference ligands interacting with this residue do it with two H-bonds (protein donating H and the ligand donating the H) and DB00770 is only interacts by being the H-donor. Two other interactions that may be relevant for inhibition are the H-bond with MET45 and salt bridge with LYS33 since neither is present in reference interactions. This could be importance since additional interactions could help stabilise the hydrophobic interactions with the residues.

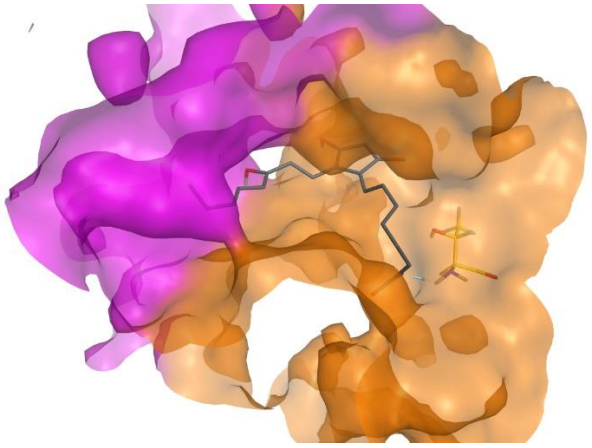


Figure 45: DB00770 (Alprostadil) in immunoproteasome β5 pocket. The orange section of the pocket is the β5 sub-unit; the purple region is the β6 sub-unit; the yellow residue in the β5 is the catalytic threonine residue.

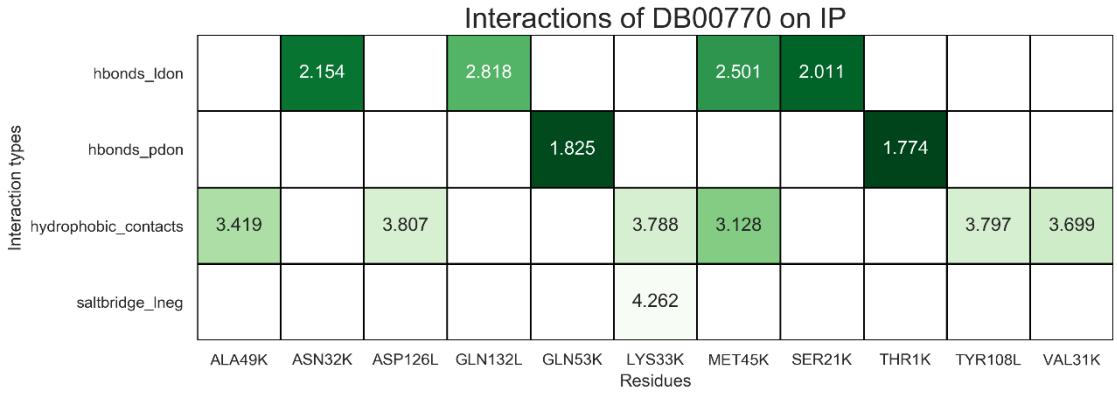


Figure 46: Heatmap representing all the interactions of DB00770 (Alprostadil) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

DB00770 (Alprostadil) is seen in Figure 44 to be completely out of the solvent accessible area of the active pocket and in the Virtual Screening it scored 78.15 with a Ligand Efficiency of 3.15. In accordance with its Ligand Efficiency, DB00770 shows an overall interaction profile similar in number of interactions with the larger crystallographic ligands interacting with 11 residues for a total of 13 interactions.

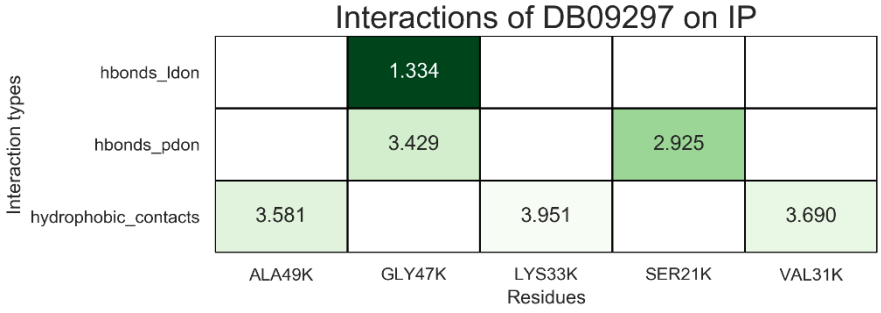


Figure 47: Heatmap representing the important interactions of DB09297 (Paritaprevir) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

Noteworthy in DB09297's (Paritaprevir) interactions with the most important residues are the interactions with GLY47. Unlike the other screened compounds DB09297 is able to interact with GLY47 with a pair of H-bonds and when Paritaprevir is the H donor the

interaction is considerably shorter than the one established by the reference ligands. On another note, alike the other screened compounds, interacting with SER21 only one H-bond is seen while reference ligands do so with a pair of H-bonds and DB09297's H-bond with SER21 is similar to Carfilzomib's interaction in distance.

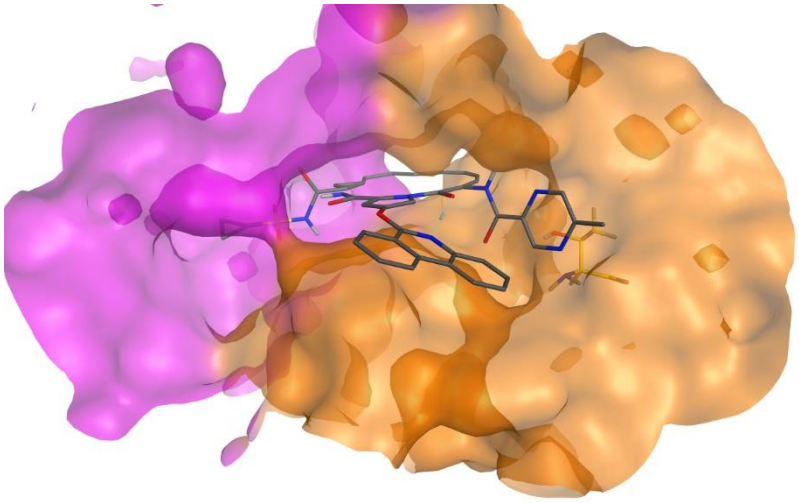


Figure 48: DB09297 (Paritaprevir) in immunoproteasome β5 pocket. The orange section of the pocket is the β5 sub-unit; the purple region is the β6 sub-unit; the yellow residue in the β5 is the catalytic threonine residue.

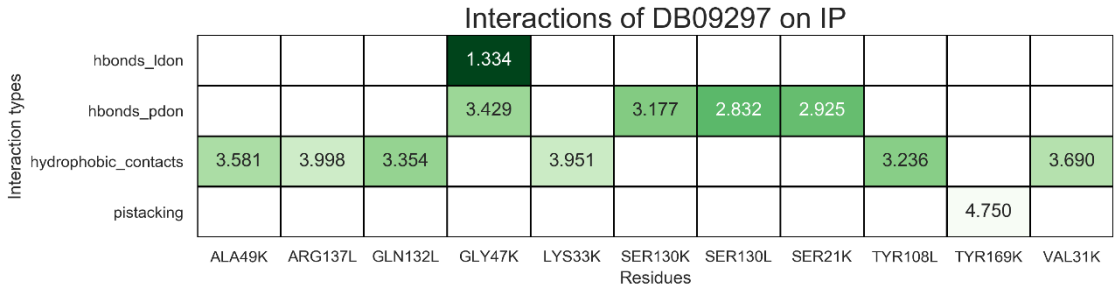


Figure 49: Heatmap representing all the interactions of DB09297 (Paritaprevir) with immunoproteasome. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

Through Figure 48 is seen that DB09297 is inside the pocket in a way so that it's not covering a large section of the solvent accessible portion of the pocket. In the Virtual Screening on the immunoproteasome DB09297 scored 99.32 with a Ligand Efficiency of

1.32. Though not the highest value of Ligand Efficiency among the four ligands analysed, it is the highest scoring for the immunoproteasome of the four. The complete interaction profile is similar in dimension to the most interactive reference ligands since DB09297 interacts with 11 residues for a total of 12 interactions.

3.3.3 Screened Compound Interactions: HDAC6

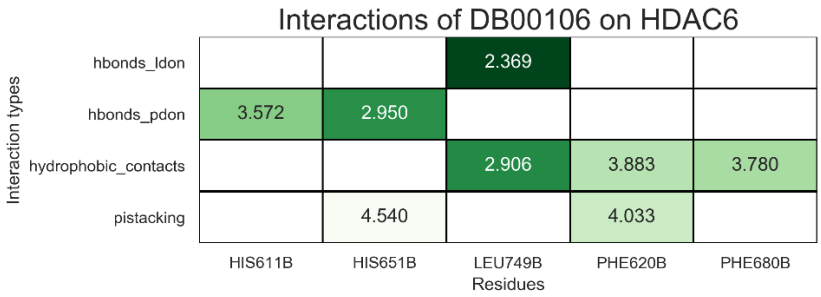


Figure 50: Heatmap representing the important interactions of DB00106 (Abarelix) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

Analysing the residues whose interactions were considered crucial for HDAC6 inhibition, DB00106 is able to have a shorter hydrophobic contact with LEU749 than any of the reference ligands (crystallographic ligands' interactions in Anex II). The H-bond with HIS651 is similar in distance with Trichostatin A's interaction of the same type. The π stacking interaction with PHE620 is longer than in most crystallographic residues that do this interaction (Acy-1083, DDK-122. RTS-V5 and SS208) but is not much different from Bavarostat's distance (3.930 Å) meaning it can still be considered similar to the data from the reference ligands. The π stacking with HIS651 is only seen in one of the reference ligands and is of similar distance to Bavarostat's interaction.

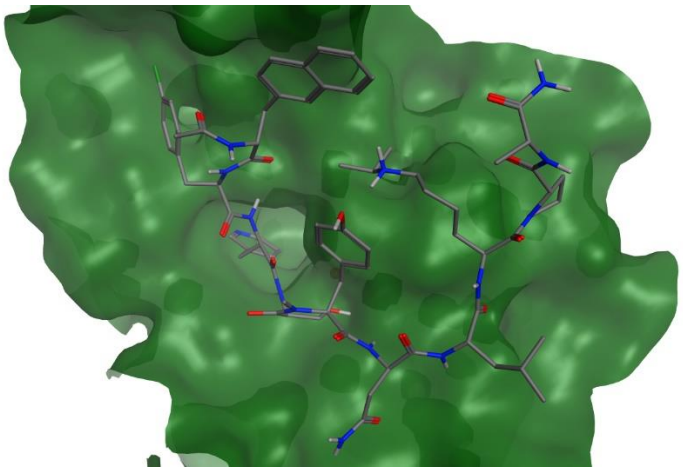


Figure 51: DB00106 (Abarelix) in HDAC6 pocket. Green is the pocket of HDAC6; the red sphere inside the protein is the zinc atom in the active pocket.

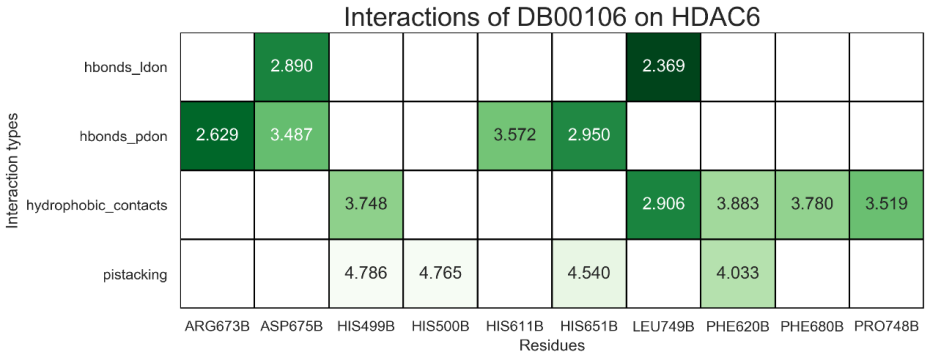


Figure 52: Heatmap representing all the interactions of DB00106 (Abarelix) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

In Figure 51 DB00106 is shown to once again occupy a large solvent accessible area of the active site, though it is expected in the case of HDAC6's pocket with its tunnelling shape. In the Virtual Screening on HDAC6 DB00106 scored 92.49 with a Ligand Efficiency of 0.92. DB00106's interaction profile with HDAC6 is larger than any crystallographic ligands since it interacts with eleven residues while Vorinostat (crystallographic ligand interacting with the most residues) interacts with ten.

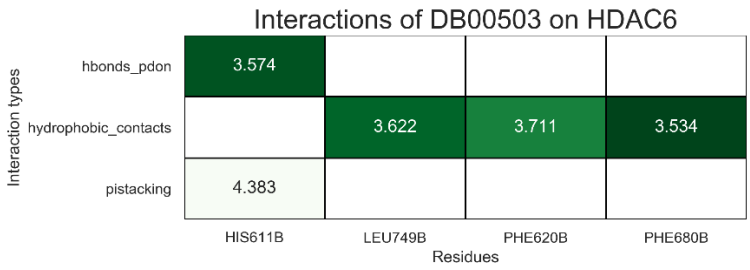


Figure 53: Heatmap representing the important interactions of DB00503 (Ritonavir) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

Noteworthy among the interactions between DB00503 and HDAC6 are the two different interactions with HIS611. Vorinostat is the only reference ligand establishing H-bonds with HIS611 and the protein being the H donor and does so with a shorter distance. From the reference ligands none interacts with HIS611 through π stacking, however Bavarostat interacts with HIS651 through π stacking with a similar distance to the interaction between the screened compound and HIS611.

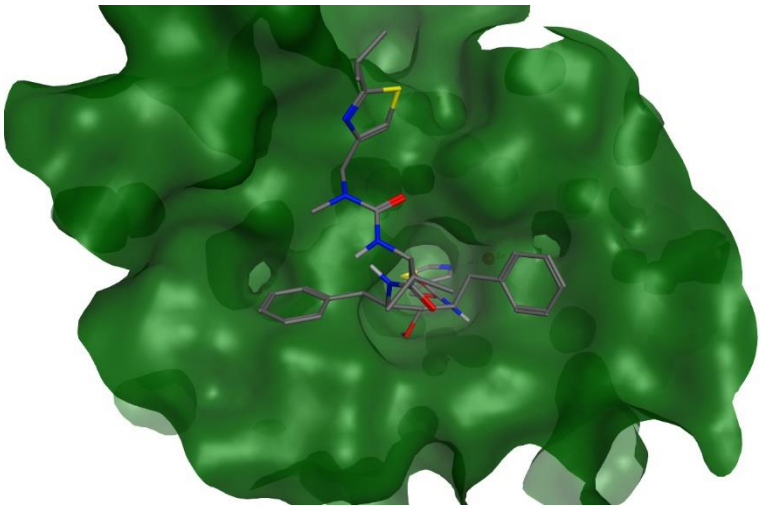


Figure 54: DB00503 (Ritonavir) in HDAC6 pocket. Green is the pocket of HDAC6; the red sphere inside the protein is the zinc atom in the active pocket.

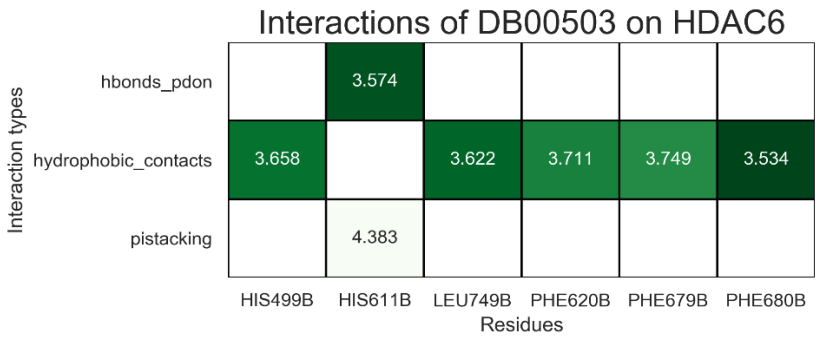


Figure 55: Heatmap representing all interactions of DB00503 (Ritonavir) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

In Figure 54 is possible to see how DB00503 fits the pocket of HDAC6, as expected though a part is inside the "tunnel" part of the pocket, most of the molecule is in the solvent accessible section. In the Virtual Screening DB00503 scored 93.91 with a Ligand Efficiency of 1.88. Most interactions between DB00503 and HDAC6 are with residues that were considered paramount for HDAC inhibition with the exception of the hydrophobic contacts established with HIS499.

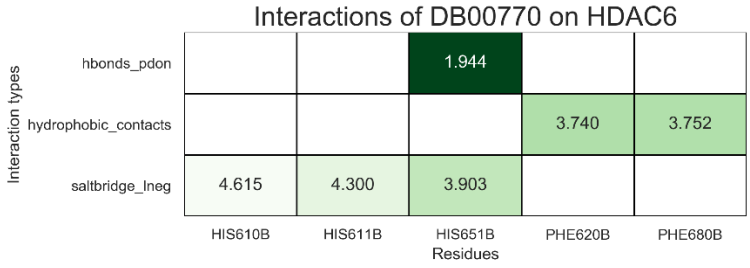


Figure 56: Heatmap representing the important interactions of DB00770 (Alprostadil) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

DB00770 interacts with HDAC6 in a unique way through three salt bridge interactions as the aniotic component of the interaction with HIS610, HIS611 and HIS651. Though these interactions are not seen in the reference ligands interaction profiles, salt bridge

interactions have two components to them, electrostatic interaction between acidic and basic groups and a H-bond. The PLIP algorithm differentiates H-bonds form salt bridges through two parameters: H-bond minimal angle (no angle between charge centres in salt bridge is required) and atom distance. For an interaction to be considered an H-bond the H donor and receptor need to be at a maximum distance of 4.1 Å as for salt bridge interactions, the maximum distance between the charge centres is 5.5 Å. In short, a salt bridge predicted by the PLIP algorithm has the potential to be an H-bond if the ligand is closer to the residue in question as long as the angle between the "would be" H donor and receptor is acceptable (minimum 100º).

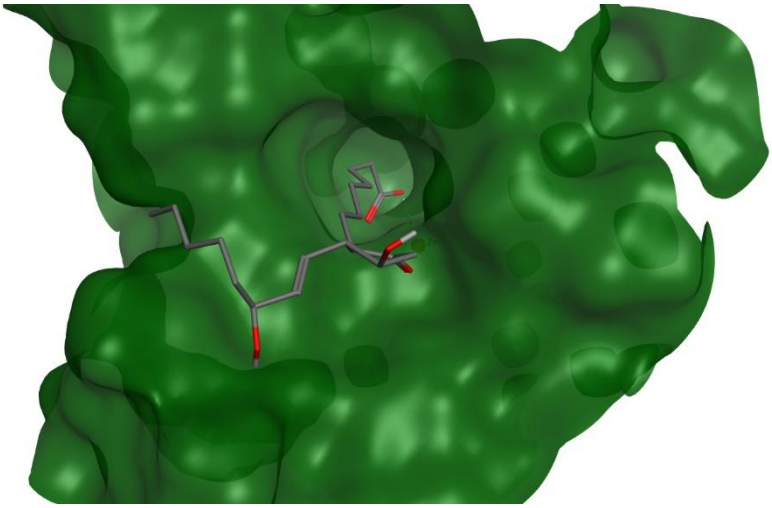


Figure 57: DB00770 (Alprostadil) in HDAC6 pocket. Green is the pocket of HDAC6; the red sphere inside the protein is the zinc atom in the active pocket.

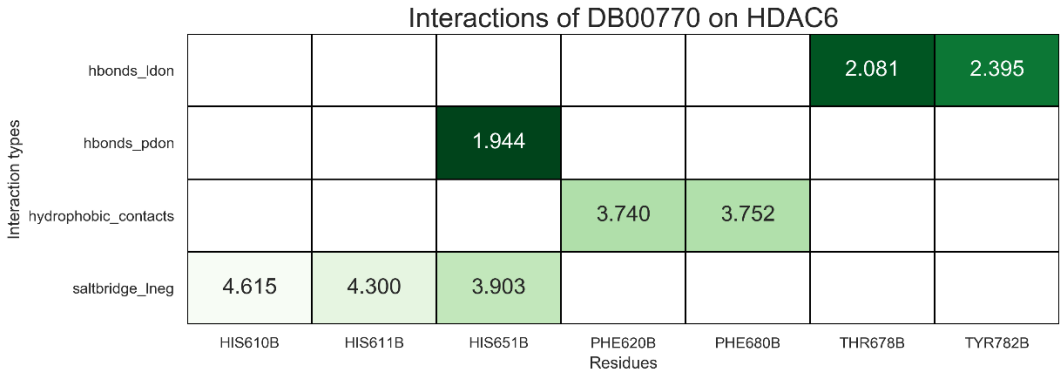


Figure 58: Heatmap representing all interactions of DB00770 (Alprostadil) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

Through the complete interaction profile of DB00770 (Figure 58) is shown that two additional interactions are established with HDAC6 besides the ones more common among reference ligands. These H-bond interactions are understandable when analysing Figure 57 where it can be seen that DB00770 has two -OH groups in the solvent accessible region of the pocket.

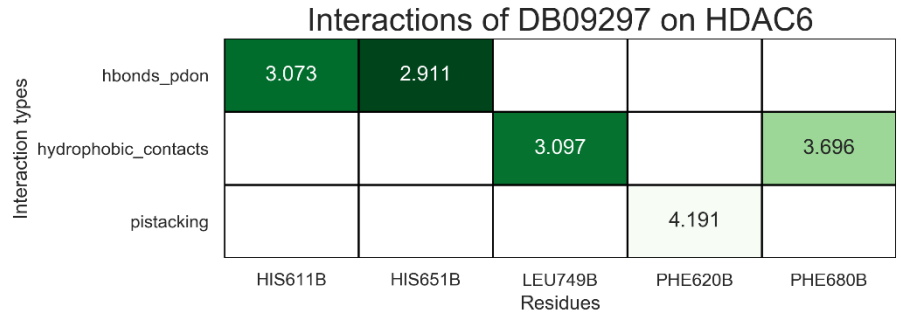


Figure 59: Heatmap representing the important interactions of DB009297 (Paritaprevir) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

Analysing the interactions established by DB09297 with the residues considered central for HDAC6 inhibition the major differences are the π stacking interaction with PHE620 and the H-bonds with HIS611 and HIS651. The π stacking is longer than reference interactions of the same type with the same residue. As for the H-bonds, Vorinostat is the only crystallographic ligand interacting with HIS611 through H-bond as an H receptor and the interactions is shorter than the one established by DB09297. Regarding the H-bond with HIS651, the only reference ligand where this interaction is found is with Trichostatin A and the distance is similar between the crysllographic ligand and DB09297.

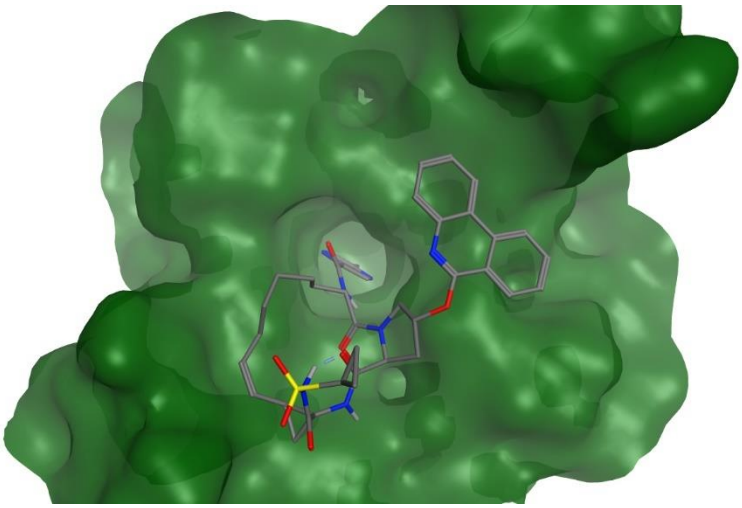


Figure 60: DB09297 (Paritaprevir) in HDAC6 pocket. Green is the pocket of HDAC6; the red sphere inside the protein is the zinc atom in the active pocket.

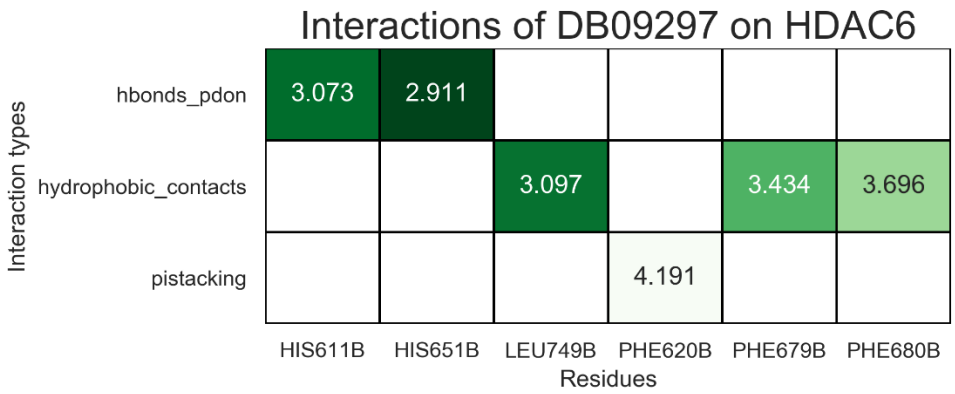


Figure 61: Heatmap representing all interactions of DB009297 (Paritaprevir) with HDAC6. Green is used to mark the interactions established between the protein and the ligand. The numbers are the distance between the atoms interacting and a colour gradient is used to represent the different values of distance (the shorter the interaction the stronger it will be and darker the colour).

DB09297's interaction profile with HDAC6 doesn't suffer much difference from the "important residue" version to the entirety of its interactions. Only an interaction with one residue is added, the hydrophobic contacts with PHE679 with a distance similar to the hydrophobic contacts established with the neighbouring PHE680. DB09297 in the Virtual Screening scored 88.54 with a Ligand Efficiency of 1.61.

3.4 Concluding Remarks

The Virtual Screening and Interaction Analysis assays yield four compounds with potential human immunoproteasome and HDAC6 dual-inhibition: Abarelix, an antagonist to gonadotropin releasing hormone; Ritonavir, an HIV protease used as a booster for other therapeutic protease inhibitors as it also inhibits CYP3A increasing plasma drug concentrations; Alprostadil, a vasodilator used to increase pulmonary activity in infants until corrective or palliative surgery can be performed; Paritaprevir an Hepatitis C serine protease inhibitor that prevents viral replication and function.

Though the result of this investigation lead to these four compounds, that is not to say that they are definitely dual-inhibitors for the intended targets but instead they should be seen as hit compounds. This work managed to only scratch the surface of what is a drug discovery endeavour, more studies can and should be conducted to verify the results presented in this thesis.

References and Appendix

1.References

[1] K. Tanaka, “Review The proteasome : Overview of structure and functions,” vol. 85. pp. 12–36, 2009.

[2] M. Groll, L. Ditzel, J. Löwe, D. Stock, M. Bochtler, and H. D. Bartunik, “Structure of 20S proteasome from yeast at 2.4A resolution,” *Nature*, no. 386, pp. 463–471, 1997.

[3] M. Groll, Y. Koguchi, R. Huber, and J. Kohno, “Crystal Structure of the 20 S Proteasome : TMC-95A Complex : A Non-covalent Proteasome Inhibitor,” pp. 1–6, 2001.

[4] G. Ben-Nissan and M. Sharon, “Regulating the 20S Proteasome Ubiquitin-Independent Degradation Pathway,” pp. 862–884, 2014.

[5] D. Vilchez, I. Saez, and A. Dillin, “The role of protein clearance mechanisms in organismal ageing and age-related diseases,” *Nature Communications*. 2014.

[6] G. Ben-Nissan and M. Sharon, “Regulating the 20S proteasome ubiquitin-independent degradation pathway,” *Biomolecules*. 2014.

[7] J. M. Baugh, E. G. Viktorova, and E. V. Pilipenko, “Proteasomes Can Degrade a Significant Proportion of Cellular Proteins Independent of Ubiquitination,” *J. Mol. Biol.*, vol. 386, no. 3, pp. 814–827, 2009.

[8] M. Unno *et al.*, “The Structure of the Mammalian 20S ˚ Resolution Proteasome at 2 . 75 A,” vol. 10, no. 02. pp. 609–618, 2002.

[9] W. Harshbarger *et al.*, “Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib Short Article Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib,” *Struct. Des.*, vol. 23, no. 2. pp. 418–424, 2015.

[10] G. N. DeMartino and C. A. Slaughter, “The proteasome, a novel protease regulated by multiple mechanisms,” *Journal of Biological Chemistry*. 1999.

[11] D. Finley, “Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome,” *Annu. Rev. Biochem.*, vol. 78, no. 1, pp. 477–513, 2009.

[12] F. G. Whitby *et al.*, “Structural basis for the activation of 20S proteasomes by 11S regulators,” *Nature*, vol. 408, no. 6808, pp. 115–120, 2000.

[13] A. F. Kisselev, Z. Songyang, and A. L. Goldberg, “Why Does Threonine , and Not Serine , Function as the Active Site Nucleophile in Proteasomes?,” *J. Biol. Chem.*, vol. 275. no. 20, pp. 14831–14837, 2000.

[14] A. F. Kisselev, W. A. Van Der Linden, and H. S. Overkleeft, “Proteasome inhibitors: An expanding army attacking a unique target,” *Chem. Biol.*, vol. 19, no. 1, pp. 99–115. 2012.

[15] National Institutes of Health, “Bortezomib Structure,” *PubChem*. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/Bortezomib. [Accessed: 04-Mar-2019].

[16] J. Schrader, T. R. Schneider, H. Stark, G. Bourenkov, and A. Chari, “The inhibition mechanism of human 20,” *Science (80-.).*, vol. 353, no. 6299, pp. 1–6, 2016.

[17] National Institutes of Health, “Carfilzomib Structure,” *PubChem*. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/Carfilzomib#section=2D-Structure. [Accessed: 05-Apr-2019].

[18] S. Gandolfi, J. P. Laubach, T. Hideshima, D. Chauhan, K. C. Anderson, and P. G. Richardson, “The proteasome and proteasome inhibitors in multiple myeloma,” *Cancer Metastasis Rev.*, vol. 36, no. 4, pp. 561–584, 2017.

[19] E. Kupperman *et al.*, “Evaluation of the proteasome inhibitor MLN9708 in preclinical models of human cancer,” *Cancer Res.*, vol. 70, no. 5. pp. 1970–1980, 2010.

[20] National Institutes of Health, “Ixazomib Structure,” *PubChem*. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/mln2238#section=Structures. [Accessed: 05-Apr-2019].

[21] ClinicalTrials.gov, “A Phase III Trial of With Marizomib in Patients With Newly Diagnosed Glioblastoma,” 2019. [Online]. Available: A Phase III Trial of With Marizomib in Patients With Newly Diagnosed Glioblastoma. [Accessed: 13-Mar-2019].

[22] M. Stirling *et al.*, “Comparison of biochemical and biological effects of ML858 (salinosporamide A) and bortezomib,” *Mol. Cancer Ther.*, vol. 5. no. 12. pp. 3052–3061, 2006.

[23] J. Chandra *et al.*, *Marizomib, a Proteasome Inhibitor for All Seasons: Preclinical Profile and a Framework for Clinical Trials*, vol. 11, no. 3. 2011.

[24] National Institutes of Health, “Marizomib Structure,” *PubChem*. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/marizomib#section=2D-Structure. [Accessed: 05-Apr-2019].

[25] A. Rouette *et al.*, “Expression of immunoproteasome genes is regulated by cell-intrinsic and -extrinsic factors in human cancers,” *Sci. Rep.*, 2016.

[26] F. Wang *et al.*, “Interferon Gamma Induces Reversible Metabolic Reprogramming of M1 Macrophages to Sustain Cell Viability and Pro-Inflammatory Activity,” *EBioMedicine*, vol. 30, pp. 303–316, 2018.

[27] K. Schroder, P. J. Hertzog, T. Ravasi, and D. A. Hume, “Interferon gamma orchestrates : an overview of signals , mechanisms and functions,” *J. Leukoc. Biol.*, vol. 75. no. February, pp. 163–189, 2004.

[28] Z. Miller, L. Ao, K. B. Kim, and W. Lee, “Inhibitors of the Immunoproteasome: Current Status and Future Directions.”

[29] H. Kimura, P. Caturegli, M. Takahashi, and K. Suzuki, “New Insights into the Function of the Immunoproteasome in Immune and Nonimmune Cells,” *Journal of Immunology Research*. 2015.

[30] M. B. Winter *et al.*, “Immunoproteasome functions explained by divergence in cleavage specificity and regulation,” *Elife*, vol. 6, pp. 1–23, 2017.

[31] D. A. Ferrington and D. S. Gregerson, “Immunoproteasomes: Structure, function, and

antigen presentation," *Prog. Mol. Biol. Transl. Sci.*, 2012.

[32] E. W. Hewitt and M. Biology, "The MHC class I antigen presentation pathway : strategies for viral immune evasion," no. 110, pp. 163–169, 2003.

[33] E. M. Huber *et al.*, "Immuno- and Constitutive Proteasome Crystal Structures Reveal Differences in Substrate and Inhibitor Specificity," *Cell*, vol. 148, no. 4, pp. 727–738, 2012.

[34] M. Basler *et al.*, "Prevention of Experimental Colitis by a Selective Inhibitor of the Immunoproteasome," 2019.

[35] M. Basler, C. J. Kirk, and M. Groettrup, "The immunoproteasome in antigen processing and other immunological functions," *Curr. Opin. Immunol.*, vol. 25. no. 1, pp. 74–80, 2013.

[36] A. Rouette, A. Trofimov, D. Haberl, and G. Boucher, "Expression of immunoproteasome genes is regulated by cell-intrinsic and – extrinsic factors in human cancers," *Nat. Publ. Gr.*, no. September, pp. 1–14, 2016.

[37] E. M. Huber *et al.*, "Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity," *Cell*, 2012.

[38] D. J. Kuhn and R. Z. Orlowski, "The Immunoproteasome as a Target in Hematologic Malignancies," vol. 49, no. 3, 2012.

[39] B. Incorporated, "ONX-0914 Structure." [Online]. Available: https://www.biovision.com/onx-0914.html. [Accessed: 02-Mar-2019].

[40] H. W. B. Johnson *et al.*, "The Discovery of Highly Selective Inhibitors of the Immunoproteasome The Discovery of Highly Selective Inhibitors of the Immunoproteasome Low Molecular Mass Polypeptide 2 (LMP2) Subunit .," *ACS Med. Chem. Lett.*, vol. 2. 2017.

[41] N. Althof *et al.*, "The immunoproteasome-specific inhibitor ONX 0914 reverses susceptibility to acute viral myocarditis," vol. 10, no. 2. pp. 200–218, 2018.

[42] E. Ogorevc, E. S. Schiffrer, I. Sosič, and S. Gobec, "A patent review of immunoproteasome inhibitors," *Expert Opin. Ther. Pat.*, vol. 0, no. 0, p. 1, 2018.

[43] H. J. Szerlong and J. C. Hansen, "Nucleosome Distribution and Linker DNA: Connecting Nuclear Function to Dynamic Chromatin Structure," *Biochem. Cell Biol.*, vol. 89, no. 1, pp. 24–34, 2011.

[44] X. Ye, C. Feng, T. Gao, G. Mu, W. Zhu, and Y. Yang, "Linker Histone in Diseases," vol. 13, 2017.

[45] L. Mariño-Ramirez, M. G. Kann, B. A. Shoemaker, and D. Landsman, "Histone Structure and Nucleosome Stability," vol. 2. no. 5. pp. 719–729, 2005.

[46] C. Diagnostics, "DNA Packaging." [Online]. Available: https://www.creative-diagnostics.com/blog/index.php/what-are-histones/. [Accessed: 03-Sep-2019].

[47] A. Eberharter and P. B. Becker, "Histone acetylation : a switch between repressive and permissive chromatin Second in review series on chromatin dynamics," vol. 3, no. 3, pp. 224–229, 2002.

[48] G. Allfrey, R. Faulkner, and A. E. Mirsky, “POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS *,” vol. 315. no. 1938, pp. 786–794, 1964.

[49] H. Yuan and R. Marmorstein, “Histone Acetyltransferases: Rising Ancient Counterparts to Protein Kinases,” vol. 99, no. 2. pp. 98–111, 2013.

[50] X. Yang and E. Seto, “HATs and HDACs : from structure , function and regulation to novel strategies for therapy and prevention,” pp. 5310–5318, 2007.

[51] J. Ridinger *et al.*, “Dual role of HDAC10 in lysosomal exocytosis and DNA repair promotes neuroblastoma chemoresistance,” no. June, pp. 1–17, 2018.

[52] S. Kotian, S. Liyanarachchi, and A. Zelent, “Histone Deacetylases 9 and 10 Are Required for Homologous Recombination,” vol. 286, no. 10, pp. 7722–7726, 2011.

[53] Y. Zhang *et al.*, “Mice Lacking Histone Deacetylase 6 Have Hyperacetylated Tubulin but Are Viable and Develop Normally,” *Mol. ans Cell. Biol.*, vol. 28, no. 5. pp. 1688–1701, 2008.

[54] T. Li *et al.*, “Histone deacetylase 6 in cancer,” vol. 1, pp. 1–10, 2018.

[55] M. Haberland, R. L. Montgomery, and E. N. Olson, “The many roles of histone deacetylases in development and physiology: implications for disease and therapy Michael,” *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 32–42. 2011.

[56] A. E. Kane and D. A. Sinclair, “Cardiovascular Aging Compendium Sirtuins and NAD + in the Development and Treatment of Metabolic and Cardiovascular Diseases,” pp. 868–885. 2018.

[57] A. Villagra *et al.*, “The histone deacetylase HDAC11 regulates the expression of interleukin 10 and immune tolerance,” vol. 10, no. 1, pp. 92–100, 2009.

[58] J. Cao, L. Sun, P. Aramsangtienchai, N. A. Spiegelman, X. Zhang, and W. Huang, “HDAC11 regulates type I interferon signaling through defatty-acylation of SHMT2.” vol. 116, no. 12. pp. 5487–5492. 2019.

[59] E. Seto and M. Yoshida, “Erasers of Histone Acetylation : The Histone Deacetylase Enzymes,” *Cold Spring Harb. Perspect. Biol.*, pp. 1–26, 2014.

[60] M. Yoshidas, M. Kishima, M. Akita, and T. Beppu, “Potent and Specific Inhibition of Mammalian Histone Deacetylase Both in Vivo and in Vitro by Trichostatin A,” *J. Biol. Chem.*, vol. 265. no. 28, pp. 17174–17179, 1990.

[61] G. Giannini *et al.*, “N -Hydroxy- (4-oxime) -cinnamide : A versatile scaffold for the synthesis of novel histone deacetilase (HDAC) inhibitors,” *Bioorg. Med. Chem. Lett.*, vol. 19, no. 8, pp. 2346–2349, 2009.

[62] V. M. Richon *et al.*, “A class of hybrid polar inducers of transformed cell differentiation inhibits histone deacetylases,” *Cell Biol.*, vol. 95. no. March, pp. 3003–3007, 1998.

[63] D. R. Nandarapu, F. Ballante, T. Chuang, A. Pirolli, B. Marrocco, and G. R. Marshall, “Design and Synthesis of Simplified Largazole Analogs as Design and Synthesis of Simplified Largazole Analogs as Isoform- Selective Human Lysine Deacetylase Inhibitors .,” *J. Med. Chem.*, 2015.

[64] H. Nakajima, Y. B. Kim, H. Terano, M. Yoshida, and S. Horinouchi, “FR901228 , a Potent

Antitumor Antibiotic , Is a Novel Histone Deacetylase Inhibitor," vol. 133, no. 241, pp. 126–133, 1998.

[65] M. Ni *et al.*, "New macrocyclic analogs of the natural histone deacetylase inhibitor FK228 ; design , synthesis and preliminary biological evaluation," *Bioorganic Med. Chem.*, 2015.

[66] E. M. Bertino and G. A. Otterson, "Romidepsin : a novel histone deacetylase inhibitor for cancer," pp. 1151–1158, 2011.

[67] Y. Imai, M. Hirano, M. Kobayashi, and M. Futami, "HDAC Inhibitors Exert Anti-Myeloma Effects through Multiple Modes of Action," *Cancers (Basel).*, vol. 11, no. 475. pp. 1–15. 2019.

[68] PubChem CID: 53340666, "Ricolinostat Structure." [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/acy-1215. [Accessed: 12-Oct-2019].

[69] PubChem CID: 11538455. "Quisinostat Structure." [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/Quisinostat. [Accessed: 12-Oct-2019].

[70] T. Hideshima, P. G. Richardson, and K. C. Anderson, "Mechanism of Action of Proteasome Inhibitors and Deacetylase Inhibitors and the Biological Basis of Synergy in Multiple Myeloma," vol. 10, no. November, pp. 2034–2043, 2011.

[71] B. K. Singh and N. Bisht, "Role of Computer Aided Drug Design in Drug Development and Drug Discovery," *Int. J. Pharm. Sci. Res.*, vol. 9, no. 4, pp. 1405–1415. 2018.

[72] I. Hoque, A. Chatterjee, S. Bhattacharya, and R. Biswas, "An Approach of Computer-Aided Drug Design (CADD) Tools for In Silico Pharmaceutical Drug Design and Development," *Int. J. Adv. Res. Biol. Sci.*, vol. 4, pp. 60–71, 2017.

[73] B. J. Neves *et al.*, "QSAR-Based Virtual Screening : Advances and Applications in Drug Discovery," vol. 9, no. November, pp. 1–7, 2018.

[74] M. Kontoyianni, "Docking and Virtual Screening in Drug Discovery," vol. 1647, pp. 255–266, 2017.

[75] D. Ramirez and J. Caballero, "Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data ?," pp. 1–17, 2018.

[76] "Docking Representation," 2005. [Online]. Available: https://commons.wikimedia.org/wiki/File:Docking_representation_2.png. [Accessed: 03-Jun-2019].

[77] M. Groll, "5M2B - PDB iCP structure," *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5m2b. [Accessed: 17-Feb-2019].

[78] M. Groll and E. M. Huber, "5L5D - PDB iCP structure," *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5l5d. [Accessed: 13-Feb-2019].

[79] M. Groll and E. M. Huber, "5L5E - PDB iCP structure," *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5L5E. [Accessed: 13-Feb-2019].

[80] M. Groll and E. M. Huber, "5L5F - PDB iCP structure," *Protein Data Bank*, 2016. .

[81] M. Groll and E. M. Huber, “5L5H - PDB iCP structure,” *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5l5h. [Accessed: 13-Feb-2019].

[82] M. Groll and E. M. Huber, “5L5I - PDB iCP structure,” *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5l5i. [Accessed: 13-Feb-2019].

[83] M. Groll and E. M. Huber, “5L5J - PDB iCP structure,” *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5l5j. [Accessed: 13-Feb-2019].

[84] M. Groll and E. M. Huber, “5L5O - PDB iCP structure,” *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5l5o. [Accessed: 13-Feb-2019].

[85] M. Groll and E. M. Huber, “5L5P - PDB iCP structure,” *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5l5p. [Accessed: 13-Feb-2019].

[86] M. Groll and E. M. Huber, “5L5Q - PDB iCP structure,” *Protein Data Bank*, 2016. [Online]. Available: https://www.rcsb.org/structure/5l5q. [Accessed: 13-Feb-2019].

[87] C. S. D. R. Update, “GOLD User Guide A Component of the CSD-Discovery Suite 2019 CSD Release Update 1 To access our new format tutorials please visit the GOLD web page,” no. 800579, 2019.

[88] M. Su, Q. Yang, Y. Du, G. Feng, Z. Liu, and Y. Li, “Comparative Assessment of Scoring Functions : The CASF-2016 Update,” 2018.

[89] J. Liu and R. Wang, “Classification of current scoring functions,” *J. Chem. Inf. Model.*, vol. 55. no. 3, pp. 475–482. 2015.

[90] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, “Improved Protein–Ligand Docking Using GOLD,” vol. 623, no. January, pp. 609–623, 2003.

[91] R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell, “A Semiempirical Free Energy Force Field with Charge-Based Desolvation,” *J. Comput. Chem.*, vol. 28, pp. 2111–2121, 2007.

[92] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder, “PLIP : fully automated protein – ligand interaction profiler,” vol. 43, no. April, pp. 443–447, 2015.

[93] E. M. Huber, W. Heinemeyer, G. De Bruin, H. S. Overkleeft, and M. Groll, “A humanized yeast proteasome identifies unique binding modes of inhibitors for the immunosubunit b 5 i,” *Embo J.*, vol. 35. no. 23, pp. 2602–2613, 2016.

[94] H. Cui *et al.*, “Structural Elucidation of a Nonpeptidic Inhibitor Specific for the Human Immunoproteasome,” *ChemBioChem*, vol. 18, pp. 523–526, 2017.

[95] DrugBank, “DB00106 - Abarelix.” [Online]. Available: https://www.drugbank.ca/drugs/DB00106. [Accessed: 21-Nov-2019].

[96] Drugs.com, “Aberalix.” [Online]. Available: https://www.drugs.com/international/abarelix.html. [Accessed: 25-Nov-2019].

[97] DrugBank, “DB00503 - Ritonavir.” [Online]. Available: https://www.drugbank.ca/drugs/DB00503. [Accessed: 21-Nov-2019].

[98] Drugs.com, “Ritonavir.” [Online]. Available: https://www.drugs.com/monograph/ritonavir.html. [Accessed: 25-Nov-2019].

[99] DrugBank, “DB00770 - Alprostadil.” [Online]. Available: https://www.drugbank.ca/drugs/DB00770.

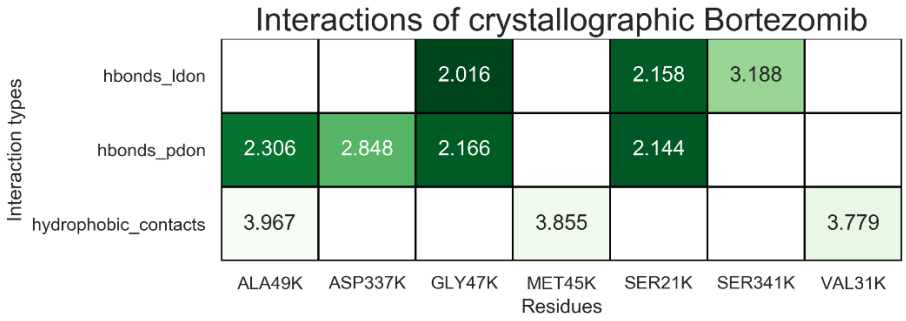
[100] Drugs.com, “Alprostadil.” [Online]. Available: https://www.drugs.com/international/alprostadil.html. [Accessed: 25-Nov-2019].

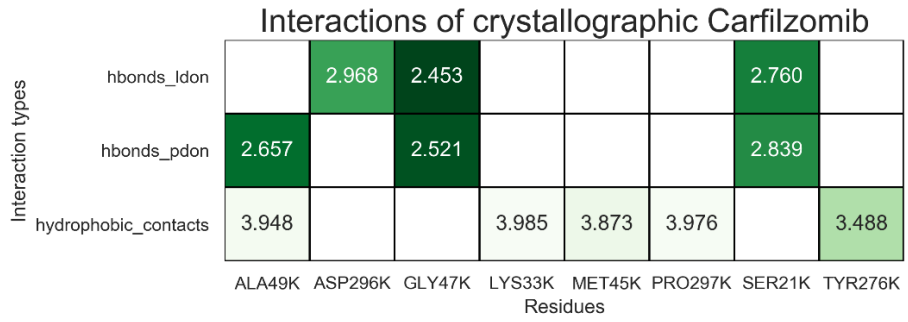
[101] Drugbank, “DB09297 - Paritaprevir.”

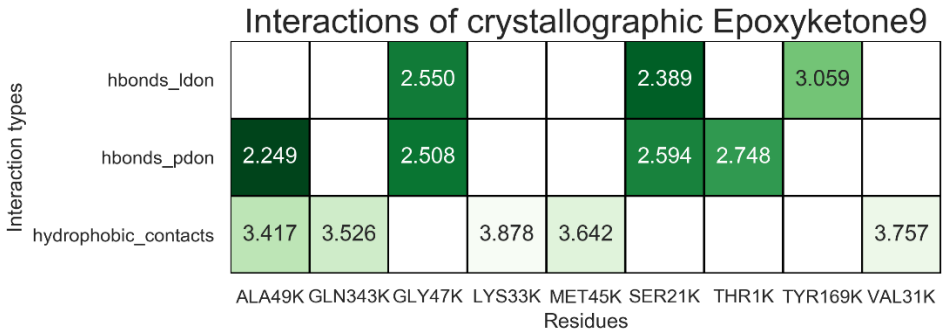
[102] Drugs.com, “Paritaprevir.” [Online]. Available: https://www.drugs.com/international/paritaprevir.html. [Accessed: 25-Nov-2019].

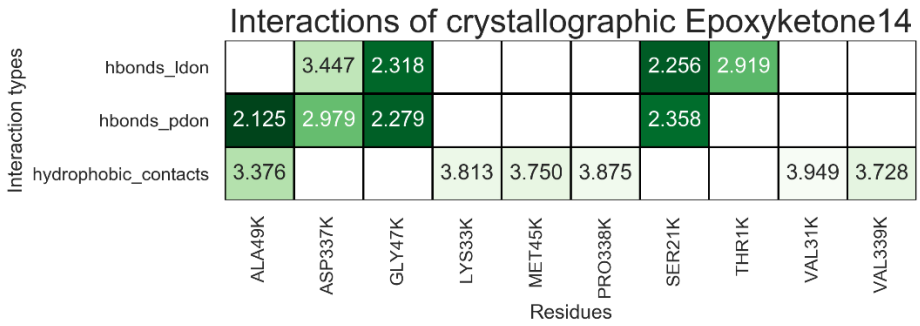
2.Appendix

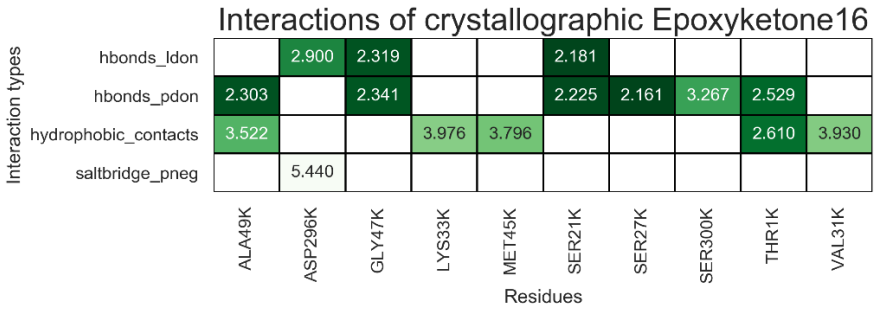
Appendix I: Crystallographic ligands of the immunoproteasome, their interaction types and distances with their native protein

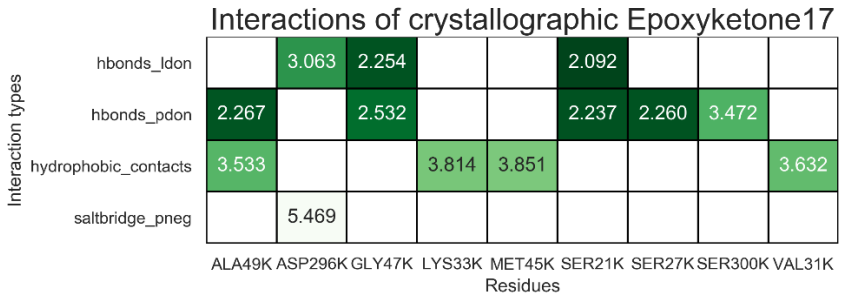


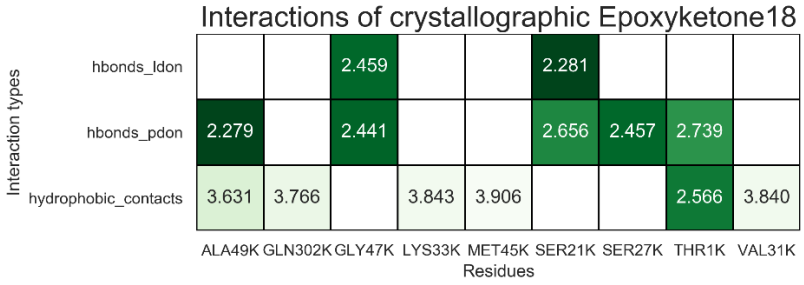


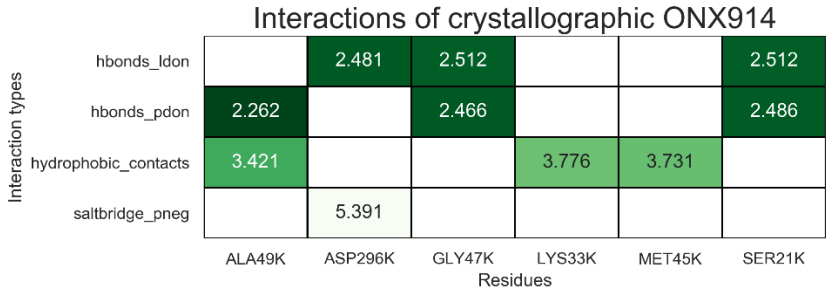


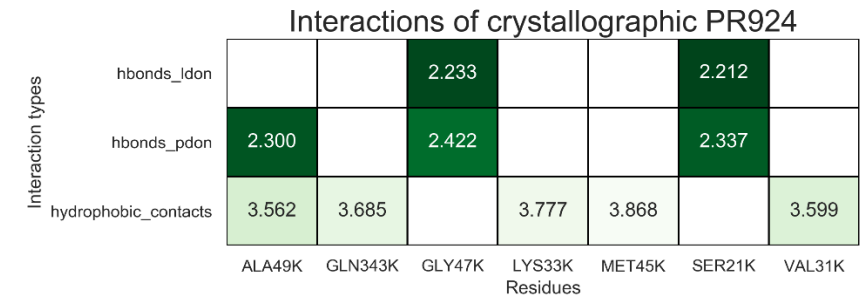

Interactions of crystallographic Epoxyketone14
Interaction types
hbonds_ldon
hbonds_pdon
hydrophobic_contacts
3.447 2.318 2.256 2.919
2.125 2.979 2.279 2.358
3.376 3.813 3.750 3.875 3.949 3.728
ALA49K ASP337K GLY47K LYS33K MET45K PRO338K SER21K THR1K VAL31K VAL339K
Residues

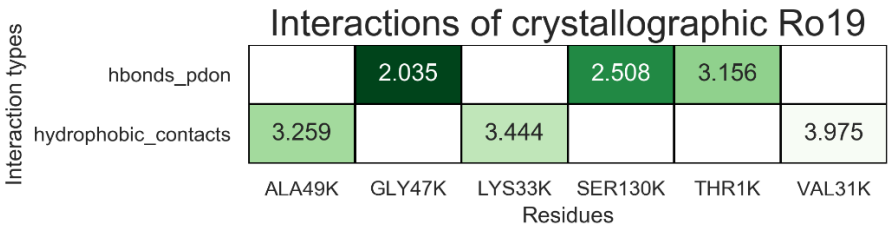

Interactions of crystallographic Epoxyketone16
Interaction types
hbonds_ldon
hbonds_pdon
hydrophobic_contacts
saltbridge_pneg
2.900 2.319 2.181
2.303 2.341 2.225 2.161 3.267 2.529
3.522 3.976 3.796 2.610 3.930
5.440
ALA49K ASP296K GLY47K LYS33K MET45K SER21K SER27K SER300K THR1K VAL31K
Residues


Interactions of crystallographic Epoxyketone17
Interaction types
hbonds_ldon
hbonds_pdon
hydrophobic_contacts
saltbridge_pneg
3.063 2.254 2.092
2.267 2.532 2.237 2.260 3.472
3.533 3.814 3.851 3.632
5.469
ALA49K ASP296K GLY47K LYS33K MET45K SER21K SER27K SER300K VAL31K
Residues

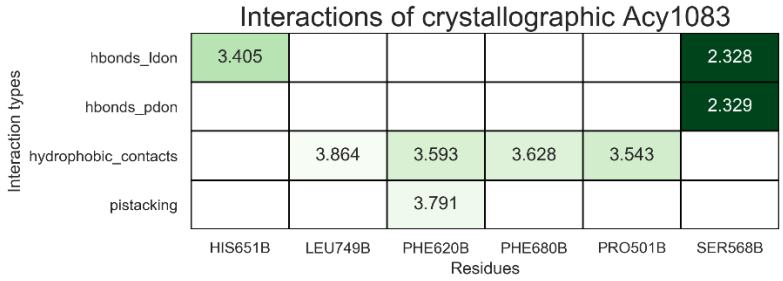

Interactions of crystallographic Epoxyketone18
Interaction types
hbonds_ldon
hbonds_pdon
hydrophobic_contacts
2.459 2.281
2.279 2.441 2.656 2.457 2.739
3.631 3.766 3.843 3.906 2.566 3.840
ALA49K GLN302K GLY47K LYS33K MET45K SER21K SER27K THR1K VAL31K
Residues

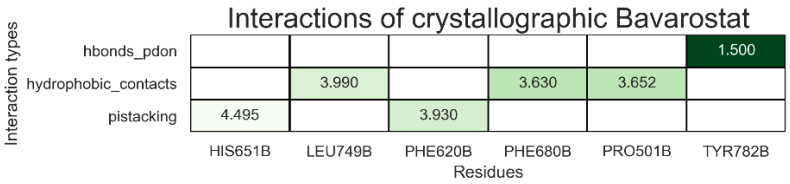

Interactions of crystallographic ONX914
Interaction types
hbonds_ldon
hbonds_pdon
hydrophobic_contacts
saltbridge_pneg
2.481 2.512 2.512
2.262 2.466 2.486
3.421 3.776 3.731
5.391
ALA49K ASP296K GLY47K LYS33K MET45K SER21K
Residues

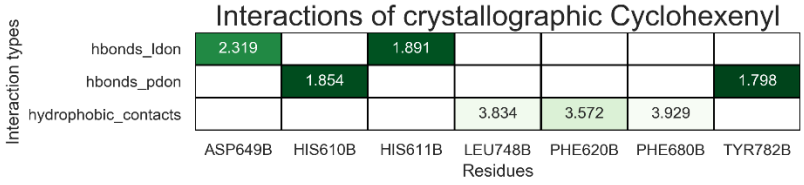


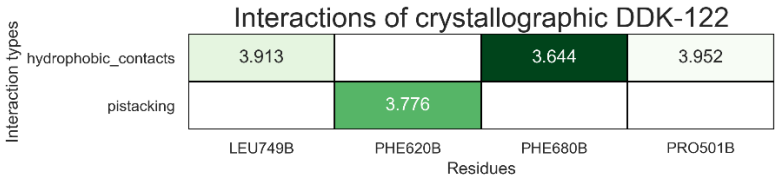


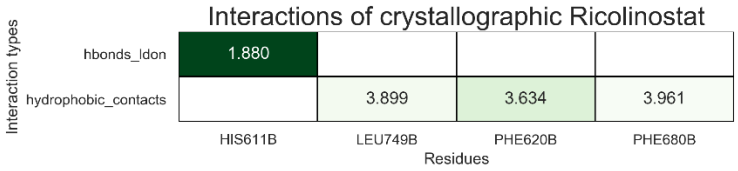
Appendix II: Crystallographic ligands of HDAC6, their interaction types and distances with their native protein

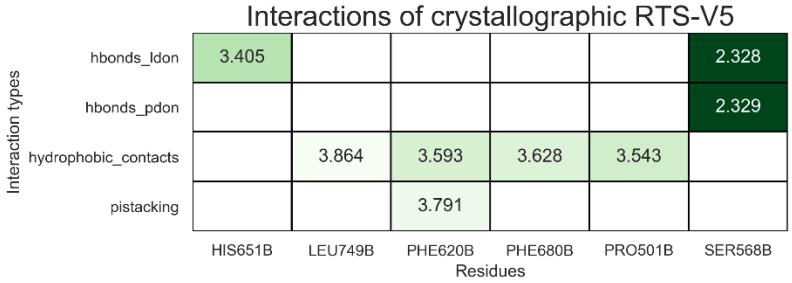


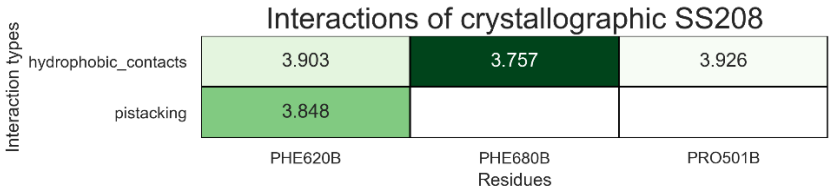


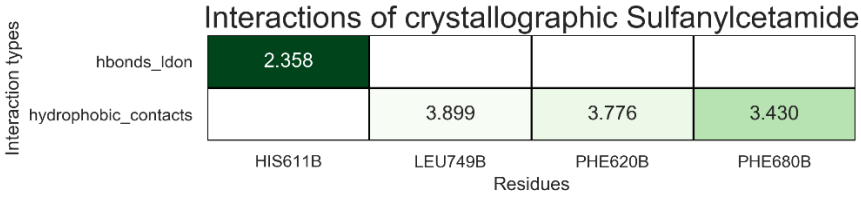


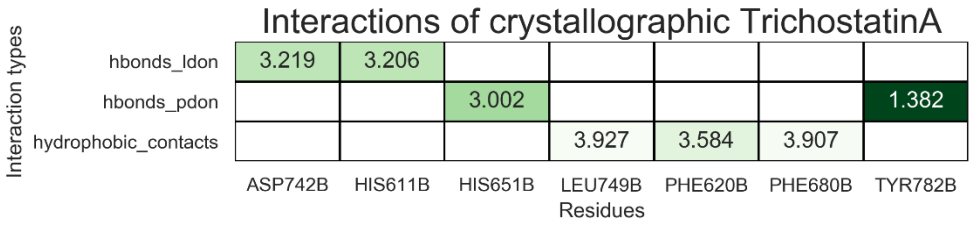

Interactions of crystallographic DDK-122
Interaction types
hydrophobic_contacts
pistacking
3.913
3.644
3.952
3.776
LEU749B
PHE620B
PHE680B
PRO501B
Residues

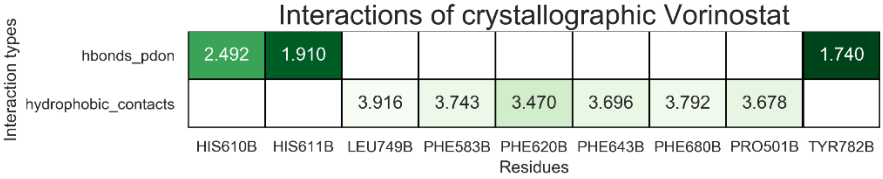

Interactions of crystallographic Ricolinostat
Interaction types
hbonds_ldon
hydrophobic_contacts
1.880
3.899
3.634
3.961
HIS611B
LEU749B
PHE620B
PHE680B
Residues


Interactions of crystallographic RTS-V5
Interaction types
hbonds_ldon
hbonds_pdon
hydrophobic_contacts
pistacking
3.405
2.328
2.329
3.864
3.593
3.628
3.543
3.791
HIS651B
LEU749B
PHE620B
PHE680B
PRO501B
SER568B
Residues


Interactions of crystallographic SS208
Interaction types
hydrophobic_contacts
pistacking
3.903
3.757
3.926
3.848
PHE620B
PHE680B
PRO501B
Residues


Interactions of crystallographic Sulfanylcetamide
Interaction types
hbonds_ldon
hydrophobic_contacts
2.358
3.899
3.776
3.430
HIS611B
LEU749B
PHE620B
PHE680B
Residues


Interactions of crystallographic TrichostatinA
Interaction types
hbonds_ldon
hbonds_pdon
hydrophobic_contacts
3.219
3.206
3.002
1.382
3.927
3.584
3.907
ASP742B
HIS611B
HIS651B
LEU749B
PHE620B
PHE680B
TYR782B
Residues


Interactions of crystallographic Vorinostat
Interaction types
hbonds_pdon
hydrophobic_contacts
2.492
1.910
1.740
3.916
3.743
3.470
3.696
3.792
3.678
HIS610B
HIS611B
LEU749B
PHE583B
PHE620B
PHE643B
PHE680B
PRO501B
TYR782B
Residues