



Università degli Studi di Cagliari e Università degli Studi di Sassari

**PhD School in Chemical Science and Technology**

Cycle XXXII

**Antiviral drug discovery: from synthesis of virus-  
targeting molecules to fragment-based lead discovery  
for a novel host target**

Scientific disciplinary sector: CHIM/08

PhD Student

Roberta Ibba

Coordinator of the PhD Programme

Prof. Stefano Enzo

Supervisor

Prof. Antonio Carta

Final exam. Academic Year 2018 – 2019  
Thesis defence: January-February 2020 Session



# Acknowledgements

Roberta Ibba gratefully acknowledges Sardinian Regional Government for the financial support of her/his PhD scholarship (P.O.R. Sardegna F.S.E. - Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014-2020 - Axis III Education and training, Thematic goal 10, Investment Priority 10ii), Specific goal 10.5.

During the last three years, I had the pleasure and the honour of being part of two groups and work with so many different persons that individually have taught me so much. It has been the hardest and best time of my life.

I would like to thank my Supervisor Prof. Antonio Carta for the ongoing support along the last three years, for have succeeded in dealing with not-the-simplest PhD student, as I used to be. I thank him for always support my ideas, plans, projects even when we disagreed on them. I would like to thank Dr. Sandra Piras for being so motherly and nice when I needed, but at the same time critical and objective during all our long conversations. I owe a great acknowledgement to all the students I supervised along those years, I had the fortune to learn a lot from anybody of them and I also find some friend in them. I thank my lab colleague Federico for supporting me and for the patience in dealing with a stressed,

third-year PhD student; during forced cohabitation in the lab, we became friends besides colleagues.

I am profoundly grateful to Prof. Nicole Zitzmann and all the Zitzmann Lab members. I spent the best time I could ever desire, I learnt so much from all of you, I broadened my outreach experiencing applied biochemistry for the very first time and I shared a great time in Oxford with you all. I would like to thank Nicole for welcoming me in her amazing group, for ever being so supportive from several points of view and for being so inspirational to me. I want to thank Dr. Pietro Roversi for mentoring me from the very first time we met until today, for being the most supportive, optimist, inspiring, funny, passionate mentor, colleague and friend I could ever desire. I am very grateful to Snezana, Dom, Michelle, Steve, Jo and J.L. for always being there for me whenever I needed. And last but not least I would like to thank all the PhD students and students I had the honour to spend so much time with. I thank Mario, Johan, July, Nilanka, Re'em, Maria, Johnny, James, and Jack.

# Abstract

Antiviral Drug Discovery is still a great challenge due to the virus's ability to mutate in order to survive and the continuous discovery of new pathogen viruses. Viral infection is caused by the presence of a virus in the body. Antivirals are compounds, either chemically synthesised or produced by a living organism, that inhibit viral infection and replication. Antivirals interfere with one or more of the virus replication cycle stage. Based on their inhibitory mechanism, antivirals can be divided into two groups: inhibitors that target the viruses (VTAs) or inhibitors that target host cell factors (HTAs). In this thesis newly synthesised VTAs are presented, a host cell target has been studied and assessed as potential antiviral target and the results of Fragment-Based Lead Discovery project is presented, and in the last part of the thesis, synthesis of small antiviral molecules and investigation of the target are described.

# Table of content

a) Viral infections .....	1
b) Antivirals .....	2
1.1. Introduction.....	5
1.2. Project aim.....	7
1.2.1. Series 1: naphthoimidazole derivatives .....	8
1.2.2. Series 2: benzimidazole derivatives .....	10
1.3. Materials and methods.....	11
1.3.1. Synthesis.....	11
1.3.1.1. General synthetic strategies .....	11
1.3.2. Chemical characterisation .....	12
1.3.2.1. Nuclear Magnetic Resonance .....	12
1.3.2.2. Mass spectroscopy .....	13
1.3.2.3. Melting Point measurements.....	14
1.3.2.4. Retention factors .....	14
1.3.3. Biological assays .....	15
1.3.3.1. Cell lines and viruses .....	15
1.3.3.2. Cytotoxicity assay .....	15

1.3.3.3. Antiviral assay.....	16
1.4. Chemistry .....	17
1.4.1. Synthetic routes .....	17
1.4.1.1. Series 1: naphthoimidazole derivatives .....	17
1.4.1.2. Series 2: benzimidazole derivatives .....	17
1.4.2. Analytical section .....	19
1.4.2.1. Series 1: naphthoimidazole derivatives .....	19
1.4.2.2. Series 2: benzimidazole derivatives .....	24
1.5. Results and discussion.....	37
1.6. Conclusions .....	41
2.1. Introduction.....	43
2.1.1. An overview of ERQC machinery .....	45
2.1.2. UDP-glucose: glycoprotein glucosyltransferase (UGGT) .....	49
2.2. Project Aim.....	53
2.3. Experimental section .....	56
2.3.1. Cloning .....	56
2.3.1.1. <i>Ct</i> UGGT <sup>G177C/A786C</sup> .....	56
2.3.1.2. <i>Ct</i> UGGT <sup>G179C/T742C</sup> .....	61
2.3.1.3. <i>Ct</i> UGGT <sup>S180C/T742C</sup> .....	64
2.3.1.4. <i>Ct</i> UGGT- $\Delta$ TRXL1.....	66

2.3.1.5. <i>Ct</i> UGGT- $\Delta$ TRXL3 .....	67
2.3.2. Protein expression and purification .....	68
2.3.3. Crystallisation .....	69
2.3.3.1. <i>Ct</i> UGGT <sup>S180C/T742C</sup> .....	69
2.3.3.2. <i>Ct</i> UGGT <sup>G177C/A786C</sup> .....	71
2.3.4. Urea-denatured thyroglobulin (UDT) preparation .....	72
2.3.5. Mass spectroscopy detection of disulphide bonds .....	73
2.3.6. Re-glucosylation assay .....	76
2.3.6.1. HPLC Re-glucosylation assay .....	77
2.4. Materials and methods .....	79
2.4.1. Cloning .....	79
2.4.1.1. Primer design .....	79
2.4.1.2. Vectors .....	79
2.4.1.3. Base-directed mutagenesis .....	79
2.4.1.4. Gibson Assembly .....	80
2.4.1.5. Transformation .....	81
2.4.1.6. Colony PCR .....	81
2.4.2. Expression in HEK293F cells .....	82
2.4.3. Protein purification .....	83
2.4.3.1. Sample preparation .....	83



2.4.3.2. Nickel affinity chromatography (NAC) .....	83
2.4.3.3. Size exclusion chromatography (SEC) .....	84
2.4.3.4. Anion exchange.....	84
2.4.3.5. Protein concentration .....	85
2.4.4. Gel electrophoresis .....	85
2.4.4.1. SDS-PAGE .....	85
2.4.4.2. DNA gels .....	86
2.4.5. Crystallisation .....	86
2.4.5.1. Crystallisation screens .....	86
2.4.5.2. X-ray Data Collection .....	87
2.4.5.3. X-ray Data Processing .....	88
2.4.6. Mass spectroscopy detection of disulphide bonds .....	88
2.5. Results and discussion.....	90
2.6. Fragment Based Lead Discovery .....	93
2.6.1. Introduction.....	93
2.6.2. Project aim .....	95
2.6.3. Experimental section .....	98
2.6.3.1. Crystal growth.....	99
2.6.3.2. Crystal soaking .....	101
2.6.3.3. X-ray diffraction .....	101

2.6.3.4. X-ray Data Processing .....	102
2.6.4. Results and discussion .....	103
2.7. Conclusions .....	106
3.1. Introduction.....	111
3.2. Materials and methods.....	113
3.2.1. Synthesis.....	113
3.2.1.1. General synthetic approach .....	113
3.2.1.2. Chemical characterisation .....	114
3.2.1.2.1. Nuclear Magnetic Resonance .....	114
3.2.1.2.2. Mass spectroscopy .....	114
3.2.1.2.4. Retention factors.....	114
3.2.2. Biological assays .....	115
3.2.2.1. Cell lines and viruses .....	115
3.2.2.2. Cytotoxicity assay .....	116
3.2.2.3. Antiviral assay.....	116
3.2.2.4. Yield reduction assay .....	117
3.2.2.5. Virucidal activity assay .....	118
3.2.2.6. Cell pre-treatment assay .....	118
3.2.2.7. Time of addition assay for 6h derivative .....	119
3.2.2.8. Time of addition assay for 7c derivative .....	119

3.2.2.9. Absorption assay .....	120
3.2.2.10. Penetration assay .....	121
3.3. Chemistry .....	122
3.3.1. Synthetic routes .....	122
3.3.1.1. Anti-VV benzimidazole derivatives .....	122
3.3.1.2. Anti-EVs benzimidazole derivatives .....	123
3.3.2. Analytical section .....	124
3.3.2.1. Synthesis, purification and characterisation of derivatives 4a- d, 5a, 5-6c, 5-11m, 7a-f, 7i-l, 8-13c .....	124
3.3.2.2. Synthesis, purification and characterisation of derivatives 5-9 g and 5-9h .....	137
3.4. Results and discussion.....	142
3.5. Conclusions .....	156
Appendix.....	157
References .....	158

# A. Introduction

## a) Viral infections

Viral infection is caused by the presence of a virus in the body. Viruses are small infectious obligate parasites that need a host cell to reproduce. Over the years, viruses have developed many strategies to be able to elude the immune response of the host and ensure viral survival, replication and proliferation, so that, notwithstanding the ever-improving antiviral drug development, viruses continue causing many diseases in humans. The first viruses were discovered at the end of the nineteenth century, but only 70 years later, it was possible to start treating viral infections. About thirty years ago most of antiviral compounds were discovered by chance (fortuitously) by testing for antiviral activity libraries of molecules originally developed for other purposes. Using an empirical approach, it resulted in a series of antiviral drugs with an unclear mechanism of action. For instance, in 1959 the 5-iodo-2-deoxyuridine, originally designed as anticancer agent, showed its antiviral activity against the Herpes Virus. From this

discovery many antiviral molecules were proposed for treating viral infections<sup>1</sup>.

Nowadays, more than 90 antiviral drugs are approved for the treatment of several human infectious diseases and many more virus inhibitors are reported in the literature.

The prevalence of chronic viral infectious diseases, the emergence of new viral infections such as picornaviruses, enteroviruses and increasing resistance to currently used antiviral drugs keep the antiviral drug discovery still a great challenge. Technical breakthroughs in virus cultivation in laboratories, identification of viral enzymes, and determination of target molecular biology led to enhanced knowledge on viral mechanism of infection and implied that targets are nowadays more directly targetable.

## b) Antivirals

Antivirals are compounds, either chemically synthesised or produced by a living organism, that inhibit viral infection and replication. Antivirals interfere with one or more of the virus replication cycle stage, this includes cell binding, cell penetration, viral uncoating, viral genome (DNA/RNA) replication, maturation, assembly of progenies, virions release. The antiviral treatment's aim is to

complement the action of vaccines or provide a treatment when the vaccine is still wanting. The antivirals need to be selective, whereby active against viruses and not cytotoxic at once.

Based on their inhibitory mechanism, antivirals can be divided into two groups: inhibitors that target the viruses (virus structural proteins, viral envelope, viral enzymes, etc.) or inhibitors that target host cell factors. Virus targeting antivirals (VTAs) can directly bind the virus, such as structural components of the virus, or they can interfere with the replication process. Host targeting antivirals (HTAs) are inhibitors which target host proteins that are necessary for the viral replication cycle, or they regulate host cell processes (e.g. inhibition of apoptosis signalling which would lead to virions release, inhibition of signalling for cell lysis, and so on), drugs which stimulate the function of the host immune system can be classified as HTAs as well.

This Ph.D. project aimed to explore different approaches to the antiviral drug discovery and for this reason, it is composed of three different sections. In Section I, newly synthesised virus targeting molecules are presented. Section II explores a host cell target that has been studied and assessed as potential antiviral target and the results

of Fragment-Based Lead Discovery project is presented. In Section III synthesis of small antivirals molecules and investigation of the target are described.

Experiments and results presented in Section II were carried out at the Zitzmann lab by the Biochemistry Department of the University of Oxford during a 9-months internship funded by Professor Zitzmann group, 'Regione Autonoma della Sardegna' and Erasmus+ Mobility programme, under supervision of Prof. Nicole Zitzmann, Dr. Pietro Roversi and Dr. Dominic Alonzi.

# Section I

## 1. Virus targeting antivirals

Virus targeting antivirals (VTAs) are drugs that directly acting on viral proteins, viral genome or viral envelope inhibit virus replication and release of virions. VTAs are very selective if compared to HTAs.

### 1.1. Introduction

Bovine viral diarrhoea virus (BVDV) is the prototype of the *Pestivirus* genus of the *Flaviviridae* family. The *Pestivirus* genus also contains other animal pathogens such as the border disease virus (BDV) and



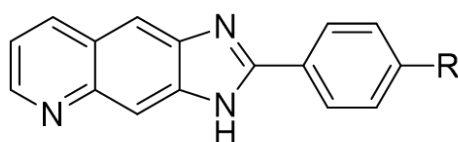
the classical swine fever virus (CSFV). BVDV causes teratogenesis, early embryonic death, abortion, respiratory problems and immune system disorder on cows which results in acute infections of immunocompetent cattle causing a mortality rate of 17 to 32%<sup>2</sup> and an estimated loss between 10 and 40 million \$ per million calvings. The virus is also considered to be a valuable surrogate for the hepatitis C virus (HCV) in antiviral drug studies<sup>3</sup>. BVDV can also be a problematic contaminant in the laboratory, BVDV strains have been identified in commercially available lots of foetal bovine serum and cell lines<sup>4</sup>, and therefore in interferons and vaccines for medical use<sup>5,6</sup>. No antivirals are currently available for controlling BVDV infections in laboratories or in farms, so BVDV remains an agronomical burden, hence the antiviral research on molecules that can specifically inhibit the virus replication could be useful to control outbreaks in farms.

Recently, a certain number of selective anti-BVDV compounds has been reported, this includes virus targeting and host targeting derivatives, e.g. polymerase inhibitors<sup>3,7-9</sup>, protease inhibitors<sup>10</sup> and human cellular enzymes inhibitors<sup>11-14</sup>. The antiviral research against animal pathogens of *Pestivirus* genus is still a challenge. In the framework of a long-lasting antiviral research program, Carta group and their collaborators, designed and synthesised a series of angular and linear *N*-polycyclic derivatives active against BVDV, HCV and other related viruses<sup>15-17</sup>. The molecular target for BVDV inhibition

was identified in the viral RNA dependent RNA polymerase (RdRp)<sup>18,19</sup>. Aiming to select the best scaffold related to antiviral activity, Carta et al. synthesised several polycyclic heteroaromatic derivatives improving selectivity and potency<sup>17,20,21</sup>.

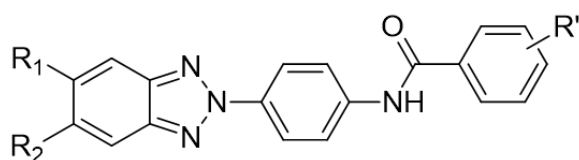
## 1.2. Project aim

In this context, considering the background research of Carta group, I selected two type of derivatives that are proved to be active against BVDV in screening assays: imidazo[4,5-*g*]quinoline derivatives<sup>17,20</sup> (Figure 1), 5-chlorobenzotriazole derivatives<sup>21</sup> and 5,6-dichlorobenzotriazole derivatives whose anti-BVDV activity is not published yet (Figure 2), and designed two series of isostere derivatives: Series 1, naphtho[2,3-*d*]imidazole: **4a-f**, **4i-l**; Series 2, 5,6-variously substituted- benzimidazole derivatives: **5-11n**, **5-11o**, **5-11p**.



R	BVDV EC <sub>50</sub>
H	10 μM
Cl	18 μM
NO <sub>2</sub>	20 μM
CN	4 μM
CF <sub>3</sub>	7 μM

Figure 1 Imidazo[4,5-*g*]quinolines and their anti-BVDV activity<sup>17,20</sup>



<b>R<sub>1</sub></b>	<b>R<sub>2</sub></b>	<b>R'</b>	<b>BVDV EC<sub>50</sub></b>
Cl	H	3,4,5-tri-OCH <sub>3</sub>	3 μM
Cl	H	4-NO <sub>2</sub>	>100 μM
Cl	Cl	3,4,5-tri-OCH <sub>3</sub>	35 μM
Cl	Cl	4-NO <sub>2</sub>	4 μM
Cl	Cl	4-Cl	70 μM

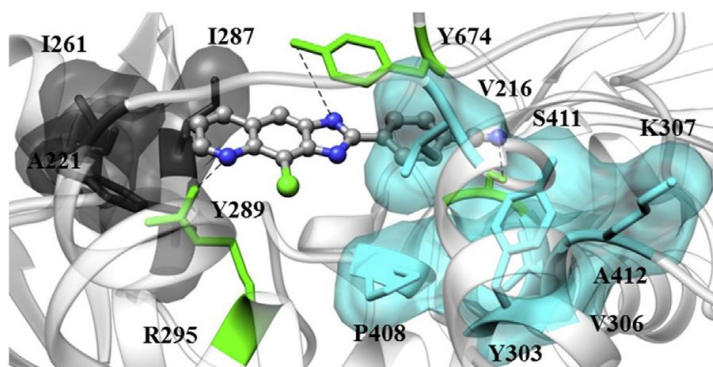
Figure 2 5-chlorobenzotriazole and 5,6-dichlorobenzotriazole derivatives and their anti-BVDV activity <sup>21</sup>

An isostere is a molecule resulted from the exchange of an atom or a group of atoms with similar biochemical properties. The aim of an isosteric replacement is to generate a new molecule with similar or optimised biological properties, lower toxicity or improved pharmacokinetics characteristics when compared to parental compound. The isosteric replacement I carried out consisted of substitution of a nitrogen atom with a carbon atom, as depicted in Figures 4 and 5.

### 1.2.1. Series 1: naphthoimidazole derivatives

Imidazo[4,5-*g*]quinoline derivatives were published as interesting BVDV inhibitors and the most active derivative of that series, 4-(4-chloro-3*H*-imidazo[4,5-*g*]quinolin-2-yl)benzonitrile<sup>20</sup> was docked

and scored for affinity towards the binding site of the BVDV RdRp, the resulting receptor/ligand complex was relaxed by energy minimisation, followed by MD simulation depicted in Figure 3. The naphthoimidazole core showed significant H-bond interaction between nitrogen atoms and key aminoacids R295, Y674 and S411. They also predicted the decrease of affinity of this molecule with the substrate when R295 and Y674 are mutated into alanine<sup>20</sup>.



*Figure 3 Equilibrated MD snapshot of the RpRd of BVDV in complex with 4-(4-chloro-3H-imidazo[4,5-g]quinolin-2-yl)benzonitrile <sup>20</sup>*

The isosteric replacement of the nitrogen atom in position 5 of imidazo[4,5-*g*]quinoline derivatives resulted in naphtho[2,3-*d*]imidazole derivatives (**4a-f**, **4i-l**) designed with the aim of confirming the essential interaction between N5 and the arginine 295. The 2-phenyl moiety was modified in order to assess the relevance of interactions with serine 411.

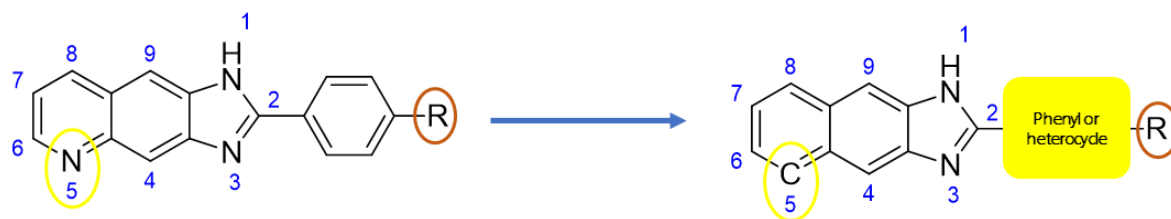


Figure 4 Projected rational modifications

### 1.2.2. Series 2: benzimidazole derivatives

Drawing on molecular modeling experiments published by Chai *et al.*<sup>22</sup>, where the benzimidazole moiety is proved to own anti-BVDV activity binding the RdRp as allosteric inhibitor, I designed an isosteric substitution of the N2 of benzotriazole derivatives depicted in Figure 5, with a carbon atom. This substitution limits the free rotation of the phenyl moiety bound to the atom in position 2 and transform the nitrogen in position 1 from a H-bond acceptor to a H-bond donor. This resulted in the series 2 of benzimidazole derivatives (5-11n, 5-11o, 5-11p) wherein substitution on benzimidazole moiety is evaluated, meanwhile substitution of R'= 4-Cl, 4-NO<sub>2</sub> or 3,4,5-trimethoxy is conserved.

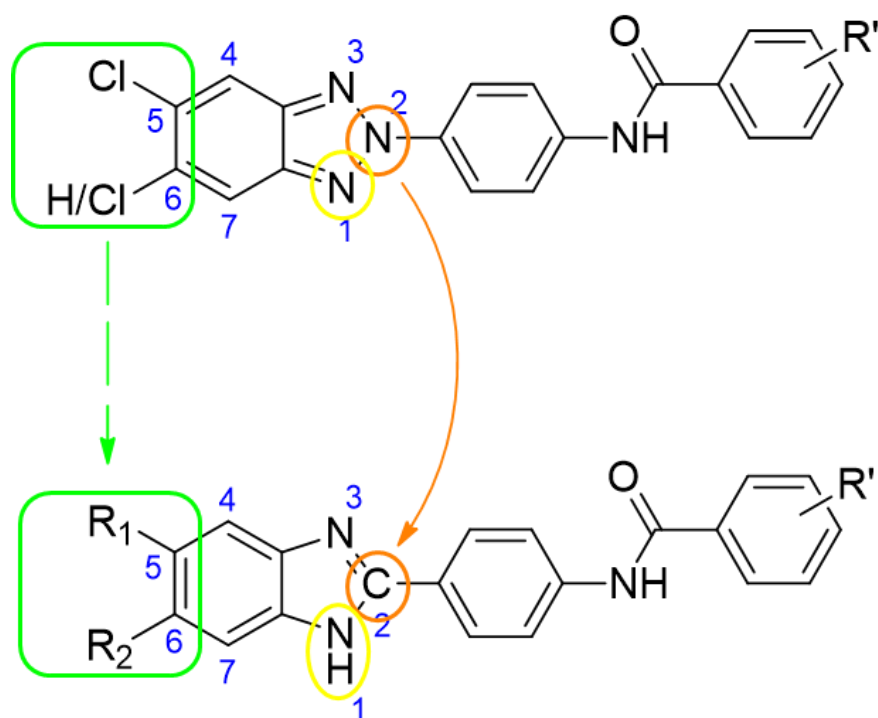


Figure 5 Projected rational modifications

## 1.3. Materials and methods

### 1.3.1. Synthesis

#### 1.3.1.1. General synthetic strategies

All starting materials were purchased by Sigma-Aldrich, Across Organics and Carlo Erba producers. Nitrogen groups reduction to amine derivatives were performed in ethanol by three different routes as follows: chlorine derivatives reductions were made with

methylhydrazine, as bland reducing agent in ethanol using the autoclave at 100°C. All the other reductions were conducted in ethanol with Pd/C as catalyst, and H<sub>2</sub> at room temperature, or with Pd/C and hydrazine at 80°C. Benzimidazole ring closure has been carried out as described by Bahrami *et al.*<sup>23</sup> by mixing o-phenylenediamines and aldehydes in a ratio 1:1 in acetonitrile, added with H<sub>2</sub>O<sub>2</sub> 30% (ratio 1:7) and HCl 37% (ratio 1:3.5) at room temperature for a few hours to some days. Amide groups were obtained from the reaction of aniline derivatives with benzoyl chloride derivative in a ratio 1:1.2 in DMF at 80°C till reaction was completed.

Products were purified by crystallization from ethanol or methanol, or by flash chromatography using proper elution by mixing suitable chosen solvents among petroleum spirit (PS), diethyl ether (DE), ethyl acetate (EA), chloroform and methanol.

### **1.3.2. Chemical characterisation**

#### **1.3.2.1. Nuclear Magnetic Resonance**

Nuclear Magnetic Resonance (NMR) spectra were registered in solutions in deuterated acetone or DMSO and were recorded with a Bruker Avance III 400 NanoBay (400 MHz) instrument. <sup>1</sup>H NMR chemical shifts are reported in parts per million (ppm) downfield

from tetramethylsilane (TMS) used as internal standard. Chemical shift values are reported in ppm ( $\delta$ ) and coupling constants (J) are reported in Hertz (Hz). Signal multiplicities are represented as s (singlet), ws (wide singlet), d (doublet), dd (doublet of doublets), ddd (doublet of doublet of doublets), t (triplet), td (triplet of doublets), q (quadruplet) and m (multiplet). The assignment of exchangeable protons (OH and NH) was confirmed by the addition of D<sub>2</sub>O. <sup>13</sup>C NMR chemical shifts are reported in downfield from tetramethylsilane (TMS) used as internal standard. Suitable method among APT (Attached Proton Test), jmod (J-modulated spin-echo for X-nuclei coupled to H-1 to determine number of attached protons) and ZGPG30 (1D sequence with power-gated decoupling using 30-degree flip angle) was selected for each compound. Two-dimensional NMR experiments HSQC (Heteronuclear Single Quantum Coherence) and HMBC (Heteronuclear Multiple Bond Correlation) were performed to correctly assign the peaks.

#### **1.3.2.2. Mass spectroscopy**

The solutions for ESI-MS measurements were prepared by dissolving the solid compounds in HPLC acetonitrile to obtain a concentration of 1.0-2.0 ppm. Mass spectra in the positive-ion mode were obtained on a Q Exactive Plus Hybrid Quadrupole-Orbitrap (Thermo Fisher Scientific) mass spectrometer. The solutions were infused at a flow rate of 5.00  $\mu$ l/min into the ESI chamber. The spectra were recorded



in the  $m/z$  range 150–800 at a resolution of 140 000 and accumulated for at least 2 min in order to increase the signal-to-noise ratio. The instrumental conditions used for the measurements were as follows: spray voltage 2300 V, capillary temperature 250 °C, sheath gas 10 (arbitrary units), auxiliary gas 3 (arbitrary units), sweep gas 0 (arbitrary units), and probe heater temperature 50 °C. ESI-MS spectra were analysed by using Thermo Xcalibur 3.0.63 software (Thermo Fisher Scientific), and the average deconvoluted monoisotopic masses were obtained through the Xtract tool integrated in the software.

#### **1.3.2.3. Melting Point measurements**

Compound melting points (m.p.) were taken in open capillaries in a Köfler hot stage or Digital Electrothermal melting point apparatus and are uncorrected.

#### **1.3.2.4. Retention factors**

Retention factors ( $R_f$ ) were measured by Thin Layer Chromatography (TLC) using Merck F-254 commercial plates.

### **1.3.3. Biological assays ♦**

#### **1.3.3.1. Cell lines and viruses**

Cell lines were purchased from American Type Culture Collection (ATCC). The absence of mycoplasma contamination was checked periodically by the Hoechst staining method. Cell line supporting the multiplication of BVDV was the following: Madin-Darby Bovine Kidney (MDBK) [ATCC CCL 22 (NBL-1) Bos Taurus]. Virus was purchased from American Type Culture Collection (ATCC), Bovine Viral Diarrhoea Virus (BVDV) [strain NADL (ATCC VR-534)].

#### **1.3.3.2. Cytotoxicity assay**

Cytotoxicity assays were run in parallel with antiviral assays. MDBK cells were seeded at an initial density of  $6 \times 10^5$  cells/cm<sup>3</sup> in 96-well plates, in culture medium [Minimum Essential Medium with Earle's salts (MEM-E) with L-glutamine, supplemented with 10% horse serum and 1 mM sodium pyruvate, and 0.025 g/dm<sup>3</sup> kanamycin]. Cell cultures were then incubated at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere in the absence or presence of serial dilutions of test compounds. Cell viability was determined after 48–96 h at 37 °C by

---

♦ Biological assays were carried out by Prof. R. Loddo and collaborators at the Department of Biomedical Science, Section of Microbiology and Virology, University of Cagliari (Italy)

the 3-(4,5-dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide (MTT) method<sup>24</sup>.

#### 1.3.3.3. Antiviral assay

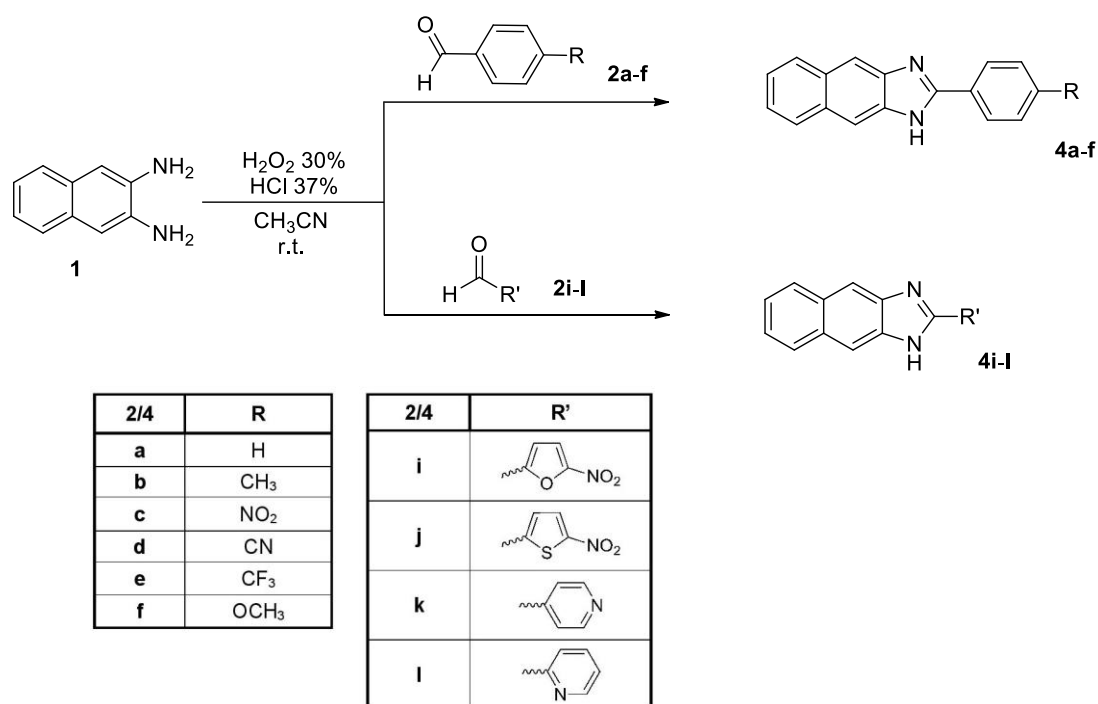
Compounds activity against BVDV was based on inhibition of virus induced cytopathogenicity in MDBK cells acutely infected with an m.o.i. of 0.01. Briefly, MDBK cells were seeded in 96-well plates at a density of  $3 \times 10^4$  cells/well and were allowed to form confluent monolayers by incubating overnight in growth medium at 37 °C in a humidified CO<sub>2</sub> (5%) atmosphere. Cell monolayers were then infected with 50 mm<sup>3</sup> of virus dilution in maintenance medium (MEM-E with L-glutamine, supplemented with 0.5% inactivated FBS, 1 mM sodium pyruvate and 0.025 g/dm<sup>3</sup> kanamycin) to give an m.o.i of 0.01. After 2 h, 50 mm<sup>3</sup> of maintenance medium, without or with serial dilutions of test compounds, were added. After a 3 days incubation at 37 °C, cell viability was determined by the MTT method. Linear regression analysis: viral and cell growth at each drug concentration was expressed as percentage of untreated controls and concentrations resulting in 50% (EC<sub>50</sub> and CC<sub>50</sub>) growth inhibition. NM 108 (2'-C-methylguanosine) and ribavirin are used as positive controls.

## 1.4. Chemistry

### 1.4.1. Synthetic routes

#### 1.4.1.1. Series 1: naphthoimidazole derivatives

Derivatives **4a-f**, **i-l** were obtained for reaction of naphthalene-2,3-diamine (**1**) with the corresponding aldehydes **2a-f**, **i-l** both dissolved in acetonitrile (CH<sub>3</sub>CN) and added of H<sub>2</sub>O<sub>2</sub> 30% in a ratio 1:7 and HCl 37% in a ratio 1:3.5.

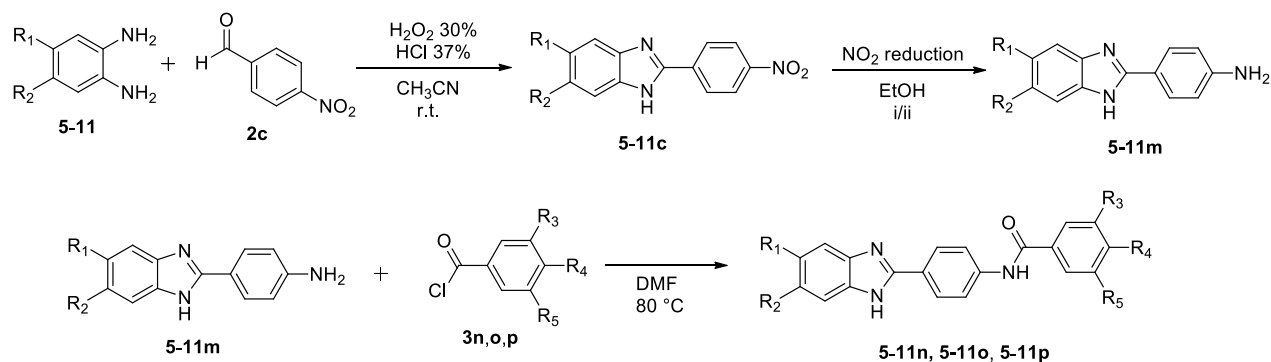


*Scheme 1 Synthetic route for derivatives **4a-f**, **4i-l**.*

#### 1.4.1.2. Series 2: benzimidazole derivatives

Intermediate o-phenyldiamines **5-11** (commercially purchased) were mixed with 4-nitrobenzaldehyde **2c** in a ratio 1:1 to obtain nitro-

derivatives **5-11c** that were appropriately reduced to aniline derivatives **5-11m**. Synthetic details will be described more thoroughly in Section III (paragraph 3.3.1.1). Derivatives **5-11n**, **5-11o**, **5-11p** were obtained for reaction of aniline derivatives **5-11m** with the corresponding benzoyl chloride derivative (**3n,o,p**) in a ratio 1:1.2 in DMF at 80°C till reaction was completed.



i: when  $R_1$  or both  $R_1$  and  $R_2 = \text{Cl}$ , reaction conditions:  $\text{CH}_3\text{NHNH}_2$  in autoclave at 100°C for 48 hours  
 ii: all the other derivatives' reaction conditions:  $\text{NH}_2\text{NH}_2$  and Pd/C at 80°C for 1 hour

	$R_1$	$R_2$
<b>5</b>	H	H
<b>6</b>	Cl	H
<b>7</b>	Cl	Cl
<b>8</b>	F	H
<b>9</b>	F	F
<b>10</b>	$\text{CH}_3$	H
<b>11</b>	$\text{CH}_3$	$\text{CH}_3$

<b>3,5-11</b>	$R_3$	$R_4$	$R_5$
<b>n</b>	$\text{OCH}_3$	$\text{OCH}_3$	$\text{OCH}_3$
<b>o</b>	H	$\text{NO}_2$	H
<b>p</b>	H	Cl	H

*Scheme 2 Synthetic route for derivatives **5-11n**, **5-11o**, **5-11p***

## 1.4.2. Analytical section

### 1.4.2.1. Series 1: Synthesis, purification and characterisation of naphthoimidazole derivatives

Synthesis of derivatives **4a-f**, **4i-l** was carried on as described in general synthetic procedure for benzimidazole ring closure: mixing *o*-phenylenediamine (**1**) and corresponding aldehyde (**2a-f**, **i-l**) in a ratio 1:1 dissolved in acetonitrile (CH<sub>3</sub>CN), added with H<sub>2</sub>O<sub>2</sub> 30% (ratio 1:7) and HCl 37% (ratio 1:3.5) at room temperature overnight, solid is obtained. The solid has been filtered off with vacuum, washed with acetonitrile and with water till neutral pH of filtrate. The crude solid has been stored and dried in the oven overnight and were obtained pure by crystallisation from ethanol.

#### 1.4.2.1.1. Characterisation of 2-phenyl-1*H*-naphtho[2,3-*d*]imidazole (**4a**)

Compound **4a** (C<sub>17</sub>H<sub>12</sub>N<sub>2</sub>, MW 244.291) was obtained in total yield 95%; m.p. 254.6-255 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.55. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.49 (2H, d, J= 6.8 Hz, H-2',6'); 8.37 (2H, s, H-4,9); 8.19 (2H, dd, J= 3.2 Hz, H-5,8); 7.77 (3H, m, H-3',4',5'); 7.56 (2H, dd, J= 3.2 Hz, H-6,7). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 152.93 (C), 133.65 (CH), 132.37 (C), 130.94 (3C), 129.59 (3CH), 128.55 (2CH), 128.05 (2CH), 125.44 (2CH), 123.81 (C), 110.68 (CH). ESI-MS (*m/z*): calcd. for C<sub>17</sub>H<sub>12</sub>N<sub>2</sub> 245.10732, found 245.10721 [M+H]<sup>+</sup>.

**1.4.2.1.2. Characterisation of 2-(*p*-tolyl)-1*H*-naphtho[2,3-*d*]imidazole (4b)**

Compound **4b** (C<sub>18</sub>H<sub>14</sub>N<sub>2</sub>, MW 258.317) was obtained in total yield 74%; m.p. 276.6-278 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.62. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.35 (2H, s, H-4,9); 8.26 (2H, d, J= 8 Hz, H-2',6'); 8.19 (2H, dd, J= 3.2 Hz, H-5,8); 7.57 (4H, m, H-6,7,3',5'); 2.47 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 152.94 (C), 144.64 (C), 131.89 (C), 130.94 (2C), 130.21 (2CH), 128.49 (2CH), 128.03 (2CH), 126.31 (C), 125.52 (2CH), 120.57 (C), 110.50 (2CH), 21.27 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>18</sub>H<sub>14</sub>N<sub>3</sub> 259.12298, found 259.12311 [M+H]<sup>+</sup>.

**1.4.2.1.3. Characterisation of 2-(4-nitrophenyl)-1*H*-naphtho[2,3-*d*]imidazole (4c)**

Compound **4c** (C<sub>17</sub>H<sub>11</sub>N<sub>3</sub>O<sub>2</sub>) was obtained in total yield 72%; m.p. 236-238 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.63. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.57 (2H, d, J= 8.8 Hz, H-3',5'); 8.51 (2H, d, J= 8.8 Hz, H-2',6'); 8.28 (2H, s, H-4,9); 8.10 (2H, q, J= 3.6 Hz, H-5,8); 7.47 (2H, q, J= 3.2 Hz, H-6,7). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 152.52 (C), 148.75 (C), 137.82 (C), 133.94 (C), 130.57 (2C), 128.39 (2CH), 127.89 (2CH), 127.27 (C), 124.35 (2CH), 124.29 (2CH), 111.46 (2CH). ESI-MS (*m/z*): calcd. for C<sub>17</sub>H<sub>11</sub>N<sub>3</sub>O<sub>2</sub> 290.09240, found 290.09219 [M+H]<sup>+</sup>.

**1.4.2.1.4. Characterisation of 4-(1*H*-naphtho[2,3-*d*]imidazol-2-yl)benzonitrile (4d)**

Compound **4d** (C<sub>18</sub>H<sub>11</sub>N<sub>3</sub>, MW 269.300) was obtained in total yield 85%; m.p. 237.7-239.1 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.58. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.53 (2H, d, J= 8.4 Hz, H-3',5'); 8.29 (2H, s, H-4,9); 8.17 (2H, d, J= 8 Hz, H-2',6'); 8.12 (2H, dd, J= 3.2 Hz, H-5,8); 7.49 (2H, dd, J= 3.2 Hz, H-6,7). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 152.51 (C), 136.59 (C), 133.15 (2CH), 131.32 (C), 130.66 (2C), 128.30 (2CH), 127.93 (2CH), 124.54 (2CH), 118.24 (C), 113.79 (C), 111.29 (2CH). ESI-MS (*m/z*): calcd. for C<sub>18</sub>H<sub>11</sub>N<sub>3</sub> 270.10257, found 270.10245 [M+H]<sup>+</sup>.

**1.4.2.1.5. Characterisation of 2-(4-(trifluoromethyl)phenyl)-1*H*-naphtho[2,3-*d*]imidazole (4e)**

Compound **4e** (C<sub>18</sub>H<sub>11</sub>F<sub>3</sub>N<sub>2</sub>, MW 312.289) was obtained in total yield 89%; m.p. 211.3-212 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.78. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.66 (2H, d, J= 8 Hz, H-2',6'); 8.38 (2H, s, H-4,9); 8.17 (2H, dd, J= 3.2 Hz, H-5,9); 8.14 (2H, d, J= 8.4 Hz, H-3',5'); 7.54 (2H, dd, J= 3.2 Hz, H-6,7). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 151.60 (C), 132.96 (C), 132.44 (1C, m, CF<sub>3</sub>), 131.00 (2C), 129.26 (2CH), 128.04 (2CH), 127.13 (C), 126.41 (CH), 125.39 (2CH), 124.97 (C), 123.38 (CH), 122.26 (C), 111.03 (2CH). ESI-MS (*m/z*): calcd. for C<sub>18</sub>H<sub>11</sub>F<sub>3</sub>N<sub>2</sub> 313.09471, found 313.09464 [M+H]<sup>+</sup>.



**1.4.2.1.6. Characterisation of 2-(4-methoxyphenyl)-1*H*-naphtho[2,3-*d*]imidazole (4f)**

Compound **4f** (C<sub>18</sub>H<sub>14</sub>N<sub>2</sub>O, MW 274.317) was obtained in total yield 91%; m.p. 251-252.2 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.40. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.40 (2H, d, J= 8.8 Hz, H-2',6'); 8.32 (2H, s, H-4,9); 8.19 (2H, dd, J= 3.2 Hz, H-5,8); 7.57 (2H, dd, J= 3.2 Hz, H-6,7); 7.35 (2H, d, J= 8.8 Hz, H-3',5'); 3.94 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 163.65 (C), 152.93 (C), 132.06 (2C), 130.88 (3C), 130.59 (2CH), 128.00 (2CH), 125.43 (2CH), 115.26 (2CH), 110.24 (2CH), 55.89 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>18</sub>H<sub>14</sub>N<sub>2</sub>O 275.11789, found 275.11758 [M+H]<sup>+</sup>.

**1.4.2.1.7. Characterisation of 2-(5-nitrofuranyl)-1*H*-naphtho[2,3-*d*]imidazole (4i)**

Compound **4i** (C<sub>15</sub>H<sub>9</sub>N<sub>3</sub>O<sub>3</sub>, MW 279.250) was obtained in total yield 79%; m.p. 217-217.4 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.46. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.19 (2H, s, H-4,9); 8.05 (2H, dd, J= 3.2 Hz, H-5,8); 7.93 (1H, d, J= 3.6 Hz, H-4'); 7.68 (1H, d, J= 3.6, H-3'); 7.42 (2H, dd, J= 3.2 Hz, H-6,7). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 152.10 (C), 146.64 (C), 144.70 (C), 138.98 (2C), 130.55 (2C), 127.83 (2CH), 124.04 (2CH), 114.78 (2CH), 111.80 (2CH). ESI-MS (*m/z*): calcd. for C<sub>15</sub>H<sub>9</sub>N<sub>3</sub>O<sub>3</sub> 280.07167, found 280.07166 [M+H]<sup>+</sup>.

**1.4.2.1.8. Characterisation of 2-(5-nitrothiophen-2-yl)-1*H*-naphtho[2,3-*d*]imidazole (4j)**

Compound **4j** (C<sub>15</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub>S, MW 295.316) was obtained in total yield 71%; m.p. 253-254.1 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.54. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.28 (1H, d, J= 4.4 Hz, H-4'); 8.18 (2H, s, H-4,9); 8.12 (1H, d, J= 4.4 Hz, H-3'); 8.04 (2H, dd, J= 3.2 Hz, H-5,8); 7.42 (2H, dd, J= 3.2 Hz, H-6,7). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+TFA-*d*): δ 153.09 (C), 149.18 (C), 147.44 (C), 138.99 (C), 136.40 (C), 130.82 (2C), 130.41 (CH), 129.24 (CH), 1270.84 (2CH), 124.58 (2CH), 111.26 (2CH). ESI-MS (*m/z*): calcd. for C<sub>15</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub>S 296.04882, found 296.04886 [M+H]<sup>+</sup>.

**1.4.2.1.9. Characterisation of 2-(pyridin-4-yl)-1*H*-naphtho[2,3-*d*]imidazole (4k)**

Compound **4k** (C<sub>16</sub>H<sub>11</sub>N<sub>3</sub>, MW 245.279) was obtained in total yield 85%; m.p. 250.9-251.6 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.11. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 9.03 (2H, d, J= 6 Hz, H-3',5'); 8.60 (2H, d, J= 5.6 Hz, H-2',6'); 8.31 (2H, s, H-4,9); 8.09 (2H, dd, J= 3.2 Hz, H-5,8); 7.46 (2H, dd, J= 3.2 Hz, H-6,7). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 150.95 (C), 146.59 (2CH), 139.73 (C), 138.01 (C), 130.78 (2C), 127.97 (2CH), 127.45 (C), 124.45 (2CH), 122.62 (2CH), 112.02 (2CH). ESI-MS (*m/z*): calcd. for C<sub>16</sub>H<sub>11</sub>N<sub>3</sub> 246.10257, found 246.10246 [M+H]<sup>+</sup>.

**1.4.2.1.10. Characterisation of 2-(pyridin-2-yl)-1*H*-naphtho[2,3-*d*]imidazole (4l)**

Compound **4l** (C<sub>16</sub>H<sub>11</sub>N<sub>3</sub>, MW 245.279) was obtained in total yield 65%; m.p. 238.3-239 °C; TLC (PS/EA 6/4): R<sub>f</sub> 0.20. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.79 (1H, d, J= 4.8 Hz, H-3'); 8.46 (1H, d, J= 8 Hz, H-6'); 8.16 (2H, s, H-4,8); 8.05 (1H, td, <sup>1</sup>J= 7.6 Hz, <sup>2</sup>J= 1.6 Hz, H-5'); 8.01 (2H, dd, J= 3.2 Hz, H-5,8); 7.59 (1H, ddd, <sup>1</sup>J= 7.6 Hz, <sup>2</sup>J= 4.8 Hz, <sup>3</sup>J= 1.2 Hz, H-4'); 7.38 (2H, dd, J= 3.2 Hz). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 155.15 (C), 150.05 (CH), 148.35 (C), 138.17 (CH), 137.07 (C), 130.70 (2C), 129.05 (C), 128.27 (2CH), 126.00 (CH), 124.14 (2CH), 122.17 (2CH), 108.72 (CH). ESI-MS (*m/z*): calcd. for C<sub>16</sub>H<sub>11</sub>N<sub>3</sub> 246.10257, found 246.10259 [M+H]<sup>+</sup>.

**1.4.2.2. Series 2: Synthesis, purification and characterisation of benzimidazole derivatives**

Intermediated **5-11c** and **5-11m** synthesis, purification and characterisation are reported in Section III, paragraph 3.3.1.1.

**1.4.2.2.1. Synthesis, purification and characterisation of *N*-(4-(1*H*-benzo[*d*]imidazol-2-yl)phenyl)-3,4,5-trimethoxybenzamide (5n), *N*-(4-(1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-nitrobenzamide (5o), *N*-(4-(1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-chlorobenzamide (5p)**

A mixture of 230 mg of 4-(1*H*-benzo[*d*]imidazol-2-yl)aniline (1.09 mmol) (**5m**) and 276 mg of 3,4,5-trimethoxybenzoyl chloride (1.20 mmol) (**3n**), or 222 mg of 4-nitrobenzoyl chloride (1.20 mmol) (**3o**), or 210 mg of 4-nitrobenzoyl chloride (1.20 mmol) (**3p**) [ratio 1:1.1] in 10 ml of DMF was stirred at 80 °C, to obtain **5n**, **5o**, **5p** respectively. The reactions were completed after 24 hours. The reaction mixtures were poured into cold water, obtaining precipitation of the solid compounds. The precipitates were filtered off, washed with water and dried in oven overnight. The pure products (**5n**, **5o** and **5p**) were obtained by crystallisation from ethanol. Compound **5n** (C<sub>23</sub>H<sub>21</sub>N<sub>3</sub>O<sub>4</sub>, MW 403.43) was obtained in total yield 20%; m.p. 164 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.58. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.61 (1H, s, NH), 8.30 (2H, d, J= 8.8 Hz, H-2',6'), 8.09 (2H, d, J= 8.8 Hz, H-3',5'), 7.75 (2H, m, H-4,7), 7.43 (2H, m, H-5,6), 8.09 (2H, d, J= 8.8, H-3',5'), 7.35 (2H, s, H-2'',6''), 3.90 (6H, s, 2CH<sub>3</sub>-C3'',5''), 3.75 (3H, s, CH<sub>3</sub>-C4''). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 165.34 (C), 152.64 (2C), 149.52 (C), 142.59 (C), 140.59 (C), 134.57 (C), 129.52 (2C), 128.15 (2CH), 124.35 (2CH), 120.52 (2CH), 120.39 (C), 114.18 (2CH), 105.54 (2CH), 60.12 (CH<sub>3</sub>), 56.16 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd for C<sub>23</sub>H<sub>21</sub>N<sub>3</sub>O<sub>4</sub> 404.16048, found 404.16101 [M+H]<sup>+</sup>. Compound **5o** (C<sub>20</sub>H<sub>14</sub>N<sub>4</sub>O<sub>3</sub>, MW 358.35) was obtained in total yield 56%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5):

$R_f$  0.49.  $^1\text{H-NMR}$  ( $\text{DMSO-}d_6$ ):  $\delta$  10.79 (1H, s, NH), 8.41 (2H, d,  $J$ = 8.8 Hz, H-3'',5''), 8.21 (2H, d,  $J$ = 2.8 Hz, H-2'',6''), 8.20 (2H, d,  $J$ = 8.8 Hz, H-3',5'), 7.98 (2H, d,  $J$ = 8.8 Hz, H-2',6'), 7.65 (1H, s, H-7), 7.53 (1H, s, H-4), 7.20 (2H, d,  $J$ = 5.2 Hz, H-5,6).  $^{13}\text{C-NMR}$  (jmod,  $\text{DMSO-}d_6$ ):  $\delta$  164.09 (C=O), 150.98 (C), 149.22 (C), 143.82 (C), 140.39 (C), 140.11 (C), 135.07 (C), 129.26 (2CH), 127.00 (2CH), 125.80 (C), 123.58 (2CH), 122.35 (CH), 121.59 (CH), 120.42 (2CH), 118.65 (CH), 111.15 (CH). ESI-MS ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{14}\text{N}_4\text{O}_3$  359.11387, found 359.11386  $[\text{M}+\text{H}]^+$ . Compound **5p** ( $\text{C}_{20}\text{H}_{14}\text{ClN}_3\text{O}$ , MW 347.79) was obtained in total yield 62%; m.p. 170.5 °C; TLC (PS/EA 6/4):  $R_f$  0.60.  $^1\text{H-NMR}$  ( $\text{DMSO-}d_6$ ):  $\delta$  8.36 (1H, s, NH), 8.20 (2H, d,  $J$ = 8.8 Hz, H-2',6'), 8.02 (2H, m, H-3',5'), 7.94 (2H, d,  $J$ = 8.4 Hz, H-2'',6''), 7.63 (2H, m, H-4,7), 7.56 (2H, d,  $J$ = 8.4 Hz, H-3'',5''), 7.24 (1H, m, H-5,6).  $^{13}\text{C-NMR}$  (jmod,  $\text{DMSO-}d_6$ ):  $\delta$  166.42 (C), 164.63 (C), 150.72 (C), 140.90 (C), 137.76 (C), 136.61 (C), 133.35 (C), 131.10 (2CH), 129.69 (2CH), 129.61 (C), 129.35 (CH), 129.24 (CH), 128.69 (CH), 128.49 (2CH), 127.23 (CH), 122.49 (CH), 120.33 (CH). ESI-MS ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{14}\text{ClN}_3\text{O}$  348.08982, found 348.08972  $[\text{M}+\text{H}]^+$ .

**1.4.2.2.2. Synthesis, purification and characterisation of *N*-(4-(5-chloro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-3,4,5-trimethoxybenzamide (6n), *N*-(4-(5-chloro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-nitrobenzamide (6o), *N*-(4-(5-chloro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-chlorobenzamide (6p)**

A mixture of 400 mg of 4-(5-chloro-1*H*-benzo[*d*]imidazol-2-yl)aniline (1.64 mmol) (**6m**) and 416 mg of 3,4,5-trimethoxybenzoyl chloride (1.80 mmol) (**3n**), or 334 mg of 4-nitrobenzoyl chloride (1.80 mmol) (**3o**), or 315 mg of

4-nitrobenzoyl chloride (1.80 mmol) (**3p**) [ratio 1:1.1] in 10 ml of DMF was stirred at 80 °C, to obtain **6n**, **6o**, **6p** respectively. The reactions were completed after 24 hours. The reaction mixtures were poured into cold water, obtaining precipitation of the solid compounds. The precipitates were filtered off, washed with water and dried in oven overnight. The crudes obtained were all purified by flash chromatography (CHCl<sub>3</sub>/CH<sub>3</sub>OH in ratio 96/4). Compound **6n** (C<sub>23</sub>H<sub>20</sub>ClN<sub>3</sub>O<sub>4</sub>, MW 437.88) was obtained in total yield 15%; m.p. 286-287 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.53. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.35 (1H, s, NH), 8.17 (2H, d, J= 8 Hz, H-2',6'), 7.95 (2H, d, J= 8.4 Hz, H-3',5'), 7.70-7.64 (1H, m, H-6), 7.53 (1H, d, J= 8.4 Hz, H-7), 7.31 (2H, s, H-2'',6''), 7.24-7.19 (1H, m, H-4), 3.89 (6H, s, 2OCH<sub>3</sub>), 3.75 (3H, s, OCH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+TFA-*d*): δ 165.66 (C=O), 152.63 (2C), 149.98 (C), 143.85 (C), 140.68 (C), 132.57 (C), 130.66 (C), 130.25 (C), 129.24 (C), 128.81 (2CH), 126.17 (CH), 120.64 (2CH), 117.19 (C), 115.22 (CH), 113.48 (CH), 105.38 (2CH), 60.03 (OCH<sub>3</sub>), 55.96 (2OCH<sub>3</sub>). ESI-MS (*m/z*): calcd for C<sub>23</sub>H<sub>20</sub>ClN<sub>3</sub>O<sub>4</sub> 438.12151, found 438.12173 [M+H]<sup>+</sup>. Compound **6o** (C<sub>20</sub>H<sub>13</sub>ClN<sub>4</sub>O<sub>3</sub>, MW 392.80) was obtained in total yield 25%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.45. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.81 (1H, s, NH), 8.40 (2H, d, J= 8.8 Hz, H-3'',5''), 8.22 (2H, d, J= 8.8 Hz, H-2'',6''), 8.18 (2H, d, J= 8 Hz, H-3',5'), 7.99 (2H, d, J= 8.8 Hz, H-2',6'), 7.71-7.65 (1H, m, H-7), 7.54 (1H, d, J= 8.8 Hz, H-4), 7.25-7.20 (1H, m, H-6). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+ TFA-*d*): δ 164.67 (C=O), 149.88 (C), 149.31 (C), 143.37 (C), 139.81 (C), 132.50 (C), 130.70 (C), 130.33 (C), 129.24 (2CH), 128.88 (2CH), 126.19 (CH), 123.49 (2CH), 120.64 (2CH), 117.61 (C), 115.25 (CH), 113.50 (CH). ESI-MS (*m/z*):

calcd for C<sub>20</sub>H<sub>13</sub>ClN<sub>4</sub>O<sub>3</sub> 393.07489, found 393.07480 [M+H]<sup>+</sup>. Compound **6p** (C<sub>20</sub>H<sub>13</sub>Cl<sub>2</sub>N<sub>3</sub>O, MW 382.24) was obtained in total yield 34%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.62. <sup>1</sup>H-NMR (APT, DMSO-*d*<sub>6</sub>+ TFA-*d*): δ 10.55 (1H, s, NH), 8.16 (2H, d, J= 8.8 Hz, H-3',5'), 8.02 (2H, d, J= 8.8 Hz, H-2'',6''), 7.97 (2H, d, J= 8.8 Hz, H-2',6'), 7.65 (2H, d, J= 8.4 Hz, H-3'',5''), 7.59 (2H, ws, H-4,7), 7.22 (1H, dd, J<sub>ortho</sub>= 8.6 Hz, J<sub>meta</sub>= 2 Hz, H-6). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+ TFA-*d*): δ 165.16 (C=O), 149.93 (C), 143.81 (C), 136.97 (C), 132.88 (C), 132.50 (C), 130.67 (C), 130.30 (C), 129.67 (2CH), 128.88 (2CH), 128.49 (CH), 126.16 (CH), 120.47 (2CH), 119.25 (CH), 116.38 (C), 115.24 (CH), 113.54 (CH). ESI-MS (*m/z*): calcd for C<sub>20</sub>H<sub>13</sub>Cl<sub>2</sub>N<sub>3</sub>O 382.05084, found 382.05090 [M+H]<sup>+</sup>.

**1.4.2.2.3. Synthesis, purification and characterisation of *N*-(4-(5,6-dichloro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-3,4,5-trimethoxybenzamide (7n), *N*-(4-(5,6-dichloro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-nitrobenzamide (7o), *N*-(4-(5,6-dichloro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-chlorobenzamide (7p)**

A mixture of 300 mg of 4-(5,6-dichloro-1*H*-benzo[*d*]imidazol-2-yl)aniline (1.08 mmol) (**7m**) and 274 mg of 3,4,5-trimethoxybenzoyl chloride (1.19 mmol) (**3n**), or 220 mg of 4-nitrobenzoyl chloride (1.19 mmol) (**3o**), or 208 mg of 4-nitrobenzoyl chloride (1.19 mmol) (**3p**) [ratio 1:1.1] in 10 ml of DMF was stirred at 80 °C, to obtain **7n**, **7o**, **7p** respectively. The reactions were completed after 24 hours. The reaction mixtures were poured into cold water, obtaining precipitation of the solid compounds. The precipitates were filtered off, washed with water and dried in oven

overnight. The pure products were obtained by crystallisation from ethanol (**7n** and **7p**) or from methanol (**7o**). Compound **7n** ( $\text{C}_{23}\text{H}_{19}\text{Cl}_2\text{N}_3\text{O}_4$ , MW 472.32) was obtained in total yield 32%; m.p. 150.6 °C; TLC (PS/EA 6/4):  $R_f$  0.34.  $^1\text{H}$ -NMR ( $\text{DMSO}-d_6$ ):  $\delta$  10.43 (1H, s, NH), 8.13 (2H, d,  $J$  = 8.8 Hz), 7.95 (2H, d,  $J$  = 8.4 Hz), 7.83 (2H, s, H-4,7), 7.29 (2H, s, H-2',6'), 3.87 (3H, s,  $\text{CH}_3$ ), 3.81 (6H, ws,  $2\text{CH}_3$ ).  $^{13}\text{C}$ -NMR (jmod,  $\text{DMSO}-d_6$ ):  $\delta$  166.94 (C), 165.29 (C), 153.61 (C), 152.62 (C), 152.59 (C), 141.32 (C), 141.20 (C), 140.45 (C), 129.63 (C), 127.32 (2CH), 125.81 (C), 124.50 (C), 124.00 (C), 120.58 (2CH), 113.75 (2CH), 105.30 (2CH), 60.10 ( $\text{CH}_3$ ), 55.88 ( $2\text{CH}_3$ ). ESI-MS ( $m/z$ ): calcd for  $\text{C}_{23}\text{H}_{19}\text{Cl}_2\text{N}_3\text{O}_4$  472.08254, found 472.08221  $[\text{M}+\text{H}]^+$ . Compound **7o** ( $\text{C}_{20}\text{H}_{12}\text{Cl}_2\text{N}_4\text{O}_3$ , MW 427.24) was obtained in total yield 10%; m.p. >300 °C; TLC ( $\text{CHCl}_3/\text{CH}_3\text{OH}$  9.5/0.5):  $R_f$  0.53.  $^1\text{H}$ -NMR ( $\text{DMSO}-d_6$ ):  $\delta$  10.83 (1H, s, NH), 8.41 (2H, d,  $J$  = 8.8 Hz, H-3'',5''), 8.22 (2H, d,  $J$  = 8.8 Hz, H-2'',6''), 8.19 (2H, d,  $J$  = 8.8 Hz, H-3',5'), 8.00 (2H, d,  $J$  = 8.8 Hz, H-2',6'), 7.93 (1H, s, H-4), 7.76 (1H, s, H-7).  $^{13}\text{C}$ -NMR (APT,  $\text{DMSO}-d_6$ , TFA- $d$ ):  $\delta$  164.57 (C=O), 153.64 (C), 151.60 (C), 149.32 (C), 142.96 (C), 139.94 (C), 133.20 (C), 131.48 (C), 129.92 (CH), 129.29 (2CH), 128.73 (CH), 127.62 (C), 123.53 (2CH), 120.56 (CH), 119.21 (C), 115.54 (CH), 114.56 (CH), 114.11 (CH). ESI-MS ( $m/z$ ): calcd for  $\text{C}_{20}\text{H}_{12}\text{Cl}_2\text{N}_4\text{O}_3$  427.03592, found 427.03629  $[\text{M}+\text{H}]^+$ . Compound **7p** ( $\text{C}_{20}\text{H}_{12}\text{Cl}_3\text{N}_3\text{O}$ , MW 416.69) was obtained in total yield 23%; m.p. 161-162 °C; TLC (PS/EA 6/4):  $R_f$  0.69.  $^1\text{H}$ -NMR ( $\text{DMSO}-d_6$ ):  $\delta$  10.63 (1H, s, NH), 8.17 (2H, d,  $J$  = 8.4 Hz, H-2',6'), 7.86 (2H, s, H-4,7), 7.81 (2H, d,  $J$  = 8.4 Hz, H-3',5'), 7.62 (2H, d,  $J$  = 8.4 Hz, H-2''-6''), 7.55 (2H, d,  $J$  = 8.4 Hz, H-3'',5'').  $^{13}\text{C}$ -NMR (jmod,  $\text{DMSO}-d_6$ ):  $\delta$  166.47 (C), 164.82 (C), 153.22 (C), 141.47 (C), 137.78 (2C),



136.68 (C), 133.20 (C), 131.10 (2CH), 129.66 (CH), 129.53 (C), 128.69 (2CH), 128.50 (CH), 127.56 (CH), 124.97 (C), 120.34 (CH), 115.94 (CH), 113.90 (CH). ESI-MS ( $m/z$ ): calcd for  $C_{20}H_{12}Cl_3N_3O$  416.01181, found 416.01187  $[M+H]^+$ .

**1.4.2.2.4. Synthesis, purification and characterisation of *N*-(4-(5-fluoro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-3,4,5-trimethoxybenzamide (8n), *N*-(4-(5-fluoro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-nitrobenzamide (8o), *N*-(4-(5-fluoro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-chlorobenzamide (8p)**

A mixture of 500 mg of 4-(5-fluoro-1*H*-benzo[*d*]imidazol-2-yl)aniline (2.20 mmol) (**8m**) and 558 mg of 3,4,5-trimethoxybenzoyl chloride (2.42 mmol) (**3n**), or 449 mg of 4-nitrobenzoyl chloride (2.42 mmol) (**3o**), or 424 mg of 4-chlorobenzoyl chloride (2.42 mmol) (**3p**) [ratio 1:1.1] in 10 ml of DMF was stirred at 80 °C, to obtain **8n**, **8o**, **8p** respectively. The reactions were completed after 24 hours. The reaction mixtures were poured into cold water, obtaining precipitation of the solid compounds. The precipitates were filtered off, washed with water and dried in oven overnight. The crudes obtained were purified by flash chromatography ( $CHCl_3/CH_3OH$  in ratio 95/5) for **8o**. Pure **8n** and **8p** were obtained by crystallisation from ethanol. Compound **8n** ( $C_{23}H_{20}FN_3O_4$ , MW 421.42) was obtained in total yield 13%; m.p. 171.8 °C; TLC ( $CHCl_3/CH_3OH$  9.5/0.5):  $R_f$  0.46.  $^1H$ -NMR ( $DMSO-d_6$ ):  $\delta$  10.60 (1H, s, NH), 8.37 (2H, s, H-2'',6''), 8.16 (2H, d,  $J$ = 8.4 Hz, H-2'6'), 7.80 (2H, d,  $J$ = 8.4, H-3',5'), 7.63 (1H, m, H-6), 7.43 (1H, m, H-4), 7.14 (1H, m, H-7).  $^{13}C$ -NMR ( $j_{mod}$ ,  $DMSO-d_6$ ):  $\delta$  165.19 (C), 158.94 (1C,  $^1J_{C-F}$ = 235 Hz, C-F), 151.67 (2C), 140.91 (C), 140.52 (2C), 137.86 (C),

133.78 (C), 128.21 (CH), 127.75 (2CH), 122.96 (2C), 122.80 (C), 120.48 (2CH), 117.20 (CH), 115.29 (1C,  $^3J_{\text{C-F}} = 10$  Hz, CH), 111.05 (1C,  $^2J_{\text{C-F}} = 25$  Hz, CH), 100.85 (1C,  $^2J_{\text{C-F}} = 26$  Hz, CH), 60.10 (CH<sub>3</sub>), 56.13 (2CH<sub>3</sub>). ESI-MS ( $m/z$ ): calcd for C<sub>23</sub>H<sub>20</sub>FN<sub>3</sub>O<sub>4</sub> 422.15106, found 422.15108 [M+H]<sup>+</sup>. Compound **8o** (C<sub>20</sub>H<sub>13</sub>FN<sub>4</sub>O<sub>3</sub>, MW 376.34) was obtained in total yield 29%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.35. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.79 (1H, s, NH), 8.4 (2H, d, J = 8.8 Hz, H-3'',5''), 8.22 (2H, d, J = 8.8 Hz, H-2'',6''), 8.17 (2H, d, J = 8.4 Hz, H-3',5'), 7.98 (2H, d, J = 8.8 Hz, H-2',6'), 7.58 (1H, ws, H-7), 7.38 (1H, ws, H-4), 7.06 (1H, m, H-6). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+TFA-*d*): δ 164.59 (C=O), 160.11 (C, d,  $^1J_{\text{C-F}} = 240$  Hz, C-F), 149.93 (C), 149.32 (C), 143.29 (C), 139.84 (C), 132.23 (C, m, C-F), 129.28 (2CH), 128.79 (2CH), 128.36 (C), 123.52 (2CH), 120.60 (2CH), 115.34 (C, d,  $^3J_{\text{C-F}} = 10$  Hz, CH-7), 114.41 (C, d,  $^2J_{\text{C-F}} = 25$  Hz, CH-6), 100.47 (C, d,  $^2J_{\text{C-F}} = 29$  Hz, CH-4). ESI-MS ( $m/z$ ): Calcolato per C<sub>20</sub>H<sub>13</sub>FN<sub>4</sub>O<sub>3</sub> 377.10445, Misurato 377.10449 [M+H]<sup>+</sup>. Compound **8p** (C<sub>20</sub>H<sub>13</sub>ClFN<sub>3</sub>O, MW 365.79) was obtained in total yield 37%; m.p. 138.6 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.66. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.55 (1H, s, NH), 8.15 (1H, m, H-6), 7.98-7.93 (4H, m, H-3',5',3'',5''), 7.65-7.56 (4H, m, H-2',6',2'',6''), 7.39 (1H, m, H-4), 7.07 (1H, m, H-7). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 165.55 (C), 159.49 (1C, d,  $^1J_{\text{C-F}} = 234$  Hz, C-F), 153.39 (C), 141.51 (2C), 137.49 (2C), 134.31 (C), 132.02 (CH), 130.58 (2CH), 129.62 (2CH), 129.41 (2CH), 126.00 (C), 121.23 (2CH), 110.88 (1C, d,  $^2J_{\text{C-F}} = 25$  Hz, CH). ESI-MS ( $m/z$ ): calcd for C<sub>20</sub>H<sub>13</sub>ClFN<sub>3</sub>O 366.08039, found 366.08069 [M+H]<sup>+</sup>.

**1.4.2.2.5. Synthesis, purification and characterisation of *N*-(4-(5,6-difluoro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-3,4,5-trimethoxybenzamide (9n), *N*-(4-(5,6-difluoro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-nitrobenzamide (9o), *N*-(4-(5,6-difluoro-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-chlorobenzamide (9p)**

A mixture of 300 mg of 4-(5,6-difluoro-1*H*-benzo[*d*]imidazol-2-yl)aniline (1.22 mmol) (**9m**) and 309 mg of 3,4,5-trimethoxybenzoyl chloride (1.34 mmol) (**3n**), or 249 mg of 4-nitrobenzoyl chloride (1.34 mmol) (**3o**), or 235 mg of 4-nitrobenzoyl chloride (1.34 mmol) (**3p**) [ratio 1:1.1] in 10 ml of DMF was stirred at 80 °C, to obtain **9n**, **9o**, **9p** respectively. The reactions were completed after 24 hours. The reaction mixtures were poured into cold water, obtaining precipitation of the solid compounds. The precipitates were filtered off, washed with water and dried in oven overnight. The crudes obtained were purified by flash chromatography (CHCl<sub>3</sub>/CH<sub>3</sub>OH in ratio 98/2) for **9o** and **9p**. Pure **9n** was obtained by crystallisation from methanol. Compound **9n** (C<sub>23</sub>H<sub>19</sub>F<sub>2</sub>N<sub>3</sub>O<sub>4</sub>, MW 439.41) was obtained in total yield 28%; m.p. 285-287 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.50. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.34 (1H, s, NH), 8.14 (2H, d, J= 8.4 Hz, H-2',6'), 7.94 (2H, d, J= 8.4 Hz, H-3',5'), 7.69 (1H, m, H-7), 7.55 (1H, m, H-4), 7.30 (2H, s, H-2'',6''), 3.89 (6H, s, C-3'',5''-OCH<sub>3</sub>), 3.75 (3H, s, C-4''-OCH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 165.10 (C=O), 153.17 (C), 152.63 (2C), 146.64 (2C, dd, <sup>1</sup>J<sub>C-F</sub>= 251 Hz, <sup>2</sup>J<sub>C-F</sub>= 14 Hz, C-F), 140.78 (C), 140.45 (C), 139.28 (C, d, J<sub>C-F</sub>= 12 Hz, C), 130.34 (C, d, J<sub>C-F</sub>= 12 Hz, C), 129.78 (C), 126.94 (2CH), 124.76 (C), 120.47 (2CH), 105.87 (C, d, J<sub>C-F</sub>= 19

Hz, CH-F), 105.37 (2CH), 99.17 (C, d,  $J_{\text{C-F}} = 22$  Hz, CH-F), 60.12 (CH<sub>3</sub>), 56.11 (2CH<sub>3</sub>). ESI-MS ( $m/z$ ): calcd for C<sub>23</sub>H<sub>19</sub>F<sub>2</sub>N<sub>3</sub>O<sub>4</sub> 440.14164, found 440.14157 [M+H]<sup>+</sup>. Compound **9o** (C<sub>20</sub>H<sub>12</sub>F<sub>2</sub>N<sub>4</sub>O<sub>3</sub>, MW 394.33) was obtained in total yield 48%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.37. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.79 (1H, s, NH), 8.40 (2H, d,  $J = 8.8$  Hz, H-3'',5''), 8.22 (2H, d,  $J = 8.8$  Hz, H-2'',6''), 8.16 (2H, d,  $J = 8.8$  Hz, H-3',5'), 7.98 (2H, d,  $J = 8.8$  Hz, H-2',6'), 7.64 (2H, ws, H-4,7). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+ TFA-*d*): δ 164.55 (C=O), 150.62 (C), 149.34 (C), 148.71 (2C, dd,  $^1J_{\text{C-F}} = 245$  Hz,  $^2J_{\text{C-F}} = 17$  Hz, C-F), 143.34 (C), 139.88 (C), 138.42 (C), 129.25 (2CH), 128.78 (2CH), 127.76 (C, m, C), 123.44 (2CH), 120.59 (2CH), 117.87 (C, m, C), 102.50 (2C, dd,  $^2J = 15.5$  Hz,  $^3J = 7$  Hz, CH). ESI-MS ( $m/z$ ): calcd for C<sub>20</sub>H<sub>12</sub>F<sub>2</sub>N<sub>4</sub>O<sub>3</sub> 395.09502, found 395.09509 [M+H]<sup>+</sup>. Compound **9p** (C<sub>20</sub>H<sub>12</sub>ClF<sub>2</sub>N<sub>3</sub>O, MW 383.78 °C) was obtained in total yield 47%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.41. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.54(1H, s, NH), 8.14( 1H, d,  $J = 8.8$  Hz, H-7), 8.02 (2H, d,  $J = 8.4$  Hz, H-2',6'), 7.954 (2H, m, H-4, 2'',6''), 7.64 (2H, d,  $J = 8.4$  Hz, H-3',5'), 7.57 (2H, d,  $J = 8.4$  Hz, H-3'',5''). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 166.49 (C=O), 164.70 (C) 153.18 (C), 146.76 (2C, dd,  $^1J_{\text{C-F}} = 237$  Hz,  $^2J_{\text{C-F}} = 14$  Hz, C-F), 140.72 (C), 137.77 (C), 136.63 (C), 133.43 (C), 131.15 (2CH), 129.73 (2CH), 128.74 (2CH), 128.55 (2CH), 127.01 (CH), 124.93 (C), 120.37 (CH). ESI-MS ( $m/z$ ): calcd for C<sub>20</sub>H<sub>12</sub>ClF<sub>2</sub>N<sub>3</sub>O 384.07097, found 384.07104 [M+H]<sup>+</sup>.

**1.4.2.2.6. Synthesis, purification and characterisation of 3,4,5-trimethoxy-*N*-(4-(5-methyl-1*H*-benzo[*d*]imidazol-2-yl)phenyl)benzamide (10n), *N*-(4-(5-methyl-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-nitrobenzamide (10o), *N*-(4-(5-methyl-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-chlorobenzamide (10p)**

A mixture of 350 mg of 4-(5-methyl-1*H*-benzo[*d*]imidazol-2-yl)aniline (1.57 mmol) (**10m**) and 399 mg of 3,4,5-trimethoxybenzoyl chloride (1.73 mmol) (**3n**), or 321 mg of 4-nitrobenzoyl chloride (1.73 mmol) (**3o**), or 303 mg of 4-chlorobenzoyl chloride (1.73 mmol) (**3p**) [ratio 1:1.1] in 10 ml of DMF was stirred at 80 °C, to obtain **10n**, **10o**, **10p** respectively. The reactions were completed after 24 hours. The reaction mixtures were poured into cold water, obtaining precipitation of the solid compounds. The precipitates were filtered off, washed with water and dried in oven overnight. The solid obtained were pure for derivatives **10o** and **10p**. Crude of **10n** was purified by flash chromatography (CHCl<sub>3</sub>/CH<sub>3</sub>OH in ratio 95/5). Compound **10n** (C<sub>24</sub>H<sub>23</sub>N<sub>3</sub>O<sub>4</sub>, MW 417.46) was obtained in total yield 35%; m.p. 287-279 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.50. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.32 (1H, s, NH), 8.15 (2H, d, J= 8.8 Hz, H-3',5'), 7.93 (2H, d, J= 8.8 Hz, H-2',6'), 7.52-7.50 (1H, m, H-7), 7.43-7.39 (1H, m, H-4), 7.31 (2H, s, H-2'',6''), 7.02-7.00 (1H, m, H-6), 3.90 (6H, s, 2OCH<sub>3</sub>), 3.75 (3H, s, OCH<sub>3</sub>), 2.44 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+ TFA-*d*): δ 165.49 (C=O), 152.65 (2C), 148.24 (C), 143.68 (C), 140.73 (C), 136.11 (C), 131.69 (C), 129.50 (C), 129.28 (C), 128.62 (2CH), 127.40 (CH), 120.57 (2CH), 120.47 (CH), 117.28 (C), 113.06 (CH), 105.44 (2CH), 59.99 (OCH<sub>3</sub>), 55.97 (2OCH<sub>3</sub>), 20.94 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd for

$C_{24}H_{23}N_3O_4$  418.17613, found 418.17618  $[M+H]^+$ . Compound **10o** ( $C_{21}H_{16}N_4O_3$ , MW 372.38) was obtained in total yield 80%; m.p. 297-298 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.45. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.77 (1H, s, NH), 8.40 (2H, d, J= 8.8 Hz, H-3'',5''), 8.22 (2H, d, J= 8.8 Hz, H-2'',6''), 8.17 (2H, d, J= 8.8 Hz, H-3',5'), 7.97 (2H, d, J= 8.8 Hz, H-2',6'), 7.47 (1H, d, J= 8 Hz, H-7), 7.37 (1H, s, H-4), 7.03 (1H, d, J= 8 Hz, H-6), 2.44 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+ TFA-*d*): δ 164.48 (C=O), 149.35 (C), 148.11 (C), 143.23 (C), 139.87 (C), 136.15 (C), 131.69 (C), 129.50 (C), 129.29 (2CH), 128.72 (2CH), 127.44 (CH), 123.49 (2CH), 120.58 (2CH), 117.80 (C), 113.29 (CH), 113.10 (CH), 20.93 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd for  $C_{21}H_{16}N_4O_3$  373.12952, found 373.12961  $[M+H]^+$ . Compound **10p** ( $C_{21}H_{16}ClN_3O$ , MW 361.82) was obtained in total yield 60%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.82. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.52 (1H, s, NH), 8.14 (2H, d, J= 8.4 Hz, H-2'',6''), 8.02 (2H, d, J= 8 Hz, H-2',6'), 7.95 (2H, d, J= 8.4 Hz, H-3',5'), 7.64 (2H, d, J= 8.4 Hz, H-3'',5''), 7.46 (1H, d, J= 7.6 Hz, H-7), 7.36 (1H, s, H-4), 7.02 (1H, d, J= 8 Hz, H-6), 2.43 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+TFA-*d*): δ 165.08 (C=O), 148.21 (C), 143.52 (C), 136.94 (C), 136.11 (C), 132.93 (C), 131.71 (C), 129.68 (2CH), 129.52 (C), 128.65 (2CH), 128.49 (2CH), 127.41 (CH), 120.46 (CH), 117.44 (C), 113.27 (CH), 113.08 (CH), 20.95 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd for  $C_{21}H_{16}ClN_3O$  362.10547, found 362.10547  $[M+H]^+$ .

**1.4.2.2.7. Synthesis, purification and characterisation of *N*-(4-(5,6-dimethyl-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-3,4,5-trimethoxybenzamide (11n), *N*-(4-(5,6-dimethyl-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-nitrobenzamide (11o), *N*-(4-(5,6-dimethyl-1*H*-benzo[*d*]imidazol-2-yl)phenyl)-4-chlorobenzamide (11p)**

A mixture of 300 mg of 4-(5,6-dimethyl-1*H*-benzo[*d*]imidazol-2-yl)aniline (1.26 mmol) (**11m**) and 321 mg of 3,4,5-trimethoxybenzoyl chloride (1.39 mmol) (**3n**), or 258 mg of 4-nitrobenzoyl chloride (1.39 mmol) (**3o**), or 243 mg of 4-chlorobenzoyl chloride (1.39 mmol) (**3p**) [ratio 1:1.1] in 10 ml of DMF was stirred at 80 °C, to obtain **11n**, **11o**, **11p** respectively. The reactions were completed after 24 hours. The reaction mixtures were poured into cold water, obtaining precipitation of the solid compounds. The precipitates were filtered off, washed with water and dried in oven overnight. The pure solid products **11n**, **11o**, **11p** were obtained by washing the crudes with few millilitres of diethyl ether. Compound **11n** (C<sub>25</sub>H<sub>25</sub>N<sub>3</sub>O<sub>4</sub>, MW 431.48) was obtained in total yield 40%; m.p. 272-274 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.47. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.43 (1H, s, NH), 8.16 (2H, d, J= 8.4 Hz, H-2',6'), 8.00 (2H, d, J= 8.4 Hz, H-3',5'), 7.46 (2H, s, H-4,7), 7.32 (2H, s, H-2'',6''), 3.89 (6H, s, 2OCH<sub>3</sub>), 3.75 (3H, s, OCH<sub>3</sub>), 2.37 (6H, s, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 165.20 (C=O), 152.65 (2C), 149.05 (C), 141.66 (C), 140.56 (C), 134.42 (C), 132.63 (2C), 129.62 (2C), 127.51 (2CH), 122.04 (C), 120.50 (2CH), 114.24 (2CH), 105.42 (2CH), 60.13 (OCH<sub>3</sub>), 56.13 (2OCH<sub>3</sub>), 19.94 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd for C<sub>25</sub>H<sub>25</sub>N<sub>3</sub>O<sub>4</sub> 432.19178, found 432.19174 [M+H]<sup>+</sup>. Compound **11o** (C<sub>22</sub>H<sub>18</sub>N<sub>4</sub>O<sub>3</sub>, MW 386.40) was obtained in total yield 18%; m.p. >300 °C;

TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.40. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.77 (1H, s, NH), 8.40 (2H, d, J= 8.8 Hz, H-3'',5''), 8.34 (2H, d, J= 2 Hz, H-2'',6''), 8.23-8.14 (4H, m, H-2',3',5',6'), 7.96 (1H, d, J= 8.8 Hz, H-7), 7.35 (1H, s, H-4), 2.33 (6H, s, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 165.77 (C), 164.03 (C), 150.05 (C), 149.21 (C), 140.41 (C), 139.84 (C), 136.36 (C), 132.77 (C), 130.67 (2CH), 130.48 (C), 129.25 (2CH), 126.77 (CH), 125.97 (C), 123.71 (2CH), 123.57 (2CH), 120.38 (C), 19.99 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd for C<sub>22</sub>H<sub>18</sub>N<sub>4</sub>O<sub>3</sub> 387.14517, found 387.14487 [M+H]<sup>+</sup>. Compound **11p** (C<sub>22</sub>H<sub>18</sub>ClN<sub>3</sub>O, MW 375.85) was obtained in total yield 19%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 9.5/0.5): R<sub>f</sub> 0.32. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 10.54 (1H, s, NH), 8.14-7.93 (6H, m, H-2', 3', 5', 6', 2'', 6''), 7.65- 7.56 (3H, m, H-3'', 5'', 7), 7.37 (1H, s, H-4), 2.34 (6H, s, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 166.43 (C), 164.63 (C), 149.89 (C), 140.43 (C), 137.76 (C), 136.58 (C), 133.39 (C), 131.11 (2CH), 130.92 (C), 129.66 (2CH), 129.59 (C), 128.72 (2CH), 128.51 (2CH), 126.88 (CH), 124.86 (C), 120.30 (CH), 19.97 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd for C<sub>22</sub>H<sub>18</sub>ClN<sub>3</sub>O 376.12112, found 376.12097 [M+H]<sup>+</sup>.

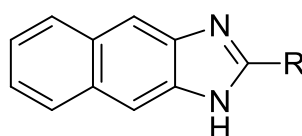
## 1.5. Results and discussion

As described above, the driven idea for this project is the isosteric substitution of a nitrogen atom with a carbon atom and evaluate the antiviral activity of the resulted molecules compared with the parental compounds. As far as naphthoimidazole derivatives (**4a-f**, **i-l**) which were designed and synthesised in order to confirm the relevance of nitrogen



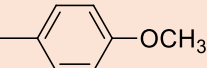
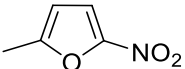
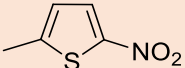
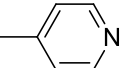
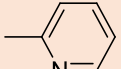
atom in position 5 of the imidazoquinoline derivatives for the interaction with the target, these molecules confirmed the predicted affinity data previously published since they showed a consistent loss of antiviral activity with respect of parental compounds. Derivative **4a** still maintains a weak antiviral activity, EC<sub>50</sub> at 16 µM, but associated with cytotoxicity CC<sub>50</sub> at 37 µM. Cytotoxicity is consistent in the greater part of analysed compounds, **4b**, **4e**, **4f**, **4j** presented CC<sub>50</sub> values ranging between 2.8 and 11.8 µM, see Table 1. Clearly, the interactions between the chemical moiety here labelled as R, and serine 411 of the target RdRp are not responsible for the anti-BVDV activity.

*Table 1 Antiviral activity and cytotoxicity of naphthoimidazole derivatives (4a-f, i-l), NM 108 and ribavirin were used as positive controls.*



**4a-f, i-l**

Compound	R	MDBK <sup>a</sup> CC <sub>50</sub>	BVDV <sup>b</sup> EC <sub>50</sub>
<b>4a</b>		<b>37.5±3.5</b>	<b>16.2±2.6</b>
<b>4b</b>		<b>11.6</b>	>11.6
<b>4c</b>		>100	>100
<b>4d</b>		>100	>100
<b>4e</b>		<b>11.8</b>	>11.8

<b>4f</b>		<b>2.8</b>	<b>&gt;2.8</b>
<b>4i</b>		<b>&gt;100</b>	<b>&gt;100</b>
<b>4j</b>		<b>4.7</b>	<b>&gt;4.7</b>
<b>4k</b>		<b>&gt;100</b>	<b>&gt;100</b>
<b>4l</b>		<b>&gt;100</b>	<b>&gt;100</b>
<b>NM 108</b>		<b>&gt;100</b>	<b>1.5±0.2</b>
<b>ribavirin</b>		<b>&gt;100</b>	<b>18±2</b>

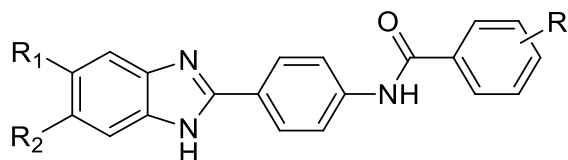
<sup>a</sup>Compound concentration ( $\mu\text{M}$ ) required to reduce the viability of mock-infected MDBK cells by 50%, as determined by the MTT method.

<sup>b</sup>Compound concentration ( $\mu\text{M}$ ) required to achieve 50% protection of MDBK cells from the BVDV-induced cytopathogenicity, as determined by the MTT method.

Concerning the benzimidazole derivatives (**5-11n**, **5-11o**, **5-11p**) which were synthesised and evaluated for activity and selectivity against BVDV, the results are presented in Table 2. It results very clear that when R' is a chlorine atom (**5-11p**), corresponding compounds have no relevant antiviral activity ( $\text{EC}_{50}$  values always higher than  $\text{CC}_{50}$  values). All trimethoxy-derivatives (**5-11n**) turned out as highly potent BVDV inhibitors with  $\text{EC}_{50}$  values ranging between 0.09 and 41  $\mu\text{M}$ , whereas only some nitro-derivatives (**6-9o**) showed interesting  $\text{EC}_{50}$  values: 1.9, 53, 7.9 and 3.6  $\mu\text{M}$  respectively. Selecting the most active trimethoxy-derivatives (**5-11n**), preliminary SARs highlight that the substitution on the benzimidazole scaffold, in position 5 or both 5 and 6 generally increases the activity, the presence of one (**6n**) or two chlorine atoms (**7n**) strengthen the activity but is enhanced with cytotoxicity,  $\text{CC}_{50}$  values: 45 and 28  $\mu\text{M}$  respectively. The presence of one or two smaller atoms such as fluorine (**8n** and **9n**)

augmented the activity ( $EC_{50}$ : 41 and 1.4  $\mu$ M) cytotoxicity values higher than 100  $\mu$ M. The best anti-BVDV activity is obtained when there is a methyl group in position 5 or both 5 and 6, resulting in derivatives **10n** and **11n**, with  $EC_{50}$  values of 0.23 and 03  $\mu$ M respectively. This SAR analysis was confirmed by the selectivity index (S.I.), which is a parameter of preferential antiviral activity of a compound in relation to its cytotoxicity ( $CC_{50}/EC_{50}$ ), and S.I. values are also tabulated in Table 2, and several new compounds present a better score than positive control compounds, NM108 (S.I. 66.7) and the gold-standard reference ribavirin for which S.I. is 5.6.

Table 2 Antiviral activity and cytotoxicity of benzimidazole derivatives (**5-11n**, **5-11o**, **5-11p**), NM108 and ribavirin were used as positive controls.



**5-11n, 5-11o, 5-11p**

Compound	R <sub>1</sub>	R <sub>2</sub>	R'	MDBK <sup>a</sup> CC <sub>50</sub>	BVDV <sup>b</sup> EC <sub>50</sub>	S.I.
<b>5n</b>	H	H	3,4,5-tri-OCH <sub>3</sub>	<b>28.5±7</b>	<b>2.2±0.4</b>	12.9
<b>5o</b>	H	H	4-NO <sub>2</sub>	>100	>100	1
<b>5p</b>	H	H	4-Cl	>100	>100	1
<b>6n</b>	Cl	H	3,4,5-tri-OCH <sub>3</sub>	<b>45</b>	<b>0.09±0.01</b>	<b>500</b>
<b>6o</b>	Cl	H	4-NO <sub>2</sub>	>100	<b>1.9±0.2</b>	>52.6
<b>6p</b>	Cl	H	4-Cl	>100	>100	1
<b>7n</b>	Cl	Cl	3,4,5-tri-OCH <sub>3</sub>	<b>28</b>	<b>1.3±0.1</b>	21.5
<b>7o</b>	Cl	Cl	4-NO <sub>2</sub>	>100	<b>53±7</b>	>1.9
<b>7p</b>	Cl	Cl	4-Cl	60	>60	1
<b>8n</b>	F	H	3,4,5-tri-OCH <sub>3</sub>	>100	<b>41±9</b>	>2.4

<b>8o</b>	F	H	4-NO <sub>2</sub>	>100	<b>7.9±2.2</b>	>12.7
<b>8p</b>	F	H	4-Cl	>100	>100	1
<b>9n</b>	F	F	3,4,5-tri-OCH <sub>3</sub>	>100	<b>1.4±0.3</b>	<b>&gt;71.4</b>
<b>9o</b>	F	F	4-NO <sub>2</sub>	>100	<b>3.6±0.1</b>	>27.8
<b>9p</b>	F	F	4-Cl	>100	>100	1
<b>10n</b>	CH <sub>3</sub>	H	3,4,5-tri-OCH <sub>3</sub>	>100	<b>0.23±0.04</b>	<b>&gt;434.8</b>
<b>10o</b>	CH <sub>3</sub>	H	4-NO <sub>2</sub>	>100	>100	1
<b>10p</b>	CH <sub>3</sub>	H	4-Cl	>100	>100	1
<b>11n</b>	CH <sub>3</sub>	CH <sub>3</sub>	3,4,5-tri-OCH <sub>3</sub>	>100	<b>0.3±0.1</b>	<b>&gt;333.3</b>
<b>11o</b>	CH <sub>3</sub>	CH <sub>3</sub>	4-NO <sub>2</sub>	>100	>100	1
<b>11p</b>	CH <sub>3</sub>	CH <sub>3</sub>	4-Cl	>100	>100	1
<b>NM 108</b>				>100	<b>1.5±0.2</b>	<b>&gt;66.7</b>
<b>ribavirin</b>				>100	<b>18±2</b>	<b>&gt;5.6</b>

<sup>a</sup>Compound concentration (μM) required to reduce the viability of mock-infected MDBK cells by 50%, as determined by the MTT method.

<sup>b</sup>Compound concentration (μM) required to achieve 50% protection of MDBK cells from the BVDV-induced cytopathogenicity, as determined by the MTT method.

## 1.6. Conclusions

The naphthoimidazole derivatives designed and synthesised as isostere of imidazoquinoline derivatives (previously reported) in order to evaluate the interactions between atoms of the chemical scaffold of these derivatives and the key aminoacids of the BVDV RdRp. The quinoline nitrogen atom was substituted with a carbon atom, which is not able to make H-bond interactions with aminoacid R295 and this resulted in loss of activity, as showed by the antiviral assay results. So, we can confirm the predicted relevance of N - R295 interaction from the docking studies published. This isosteric substitution led to not active derivatives against BVDV. At the same time, it is also clear that the biochemical interactions between the aromatic

moiety, here labelled as R, and serine 411 of the RdRp target are not responsible, but only bearing for the anti-BVDV activity.

Concerning the benzimidazole derivatives which were synthesised and evaluated for activity and selectivity against BVDV, several compounds showed interesting  $EC_{50}$  values and selectivity index, emerging as interesting agents, not toxic, against *in cellula* virus infection. If compared with benzotriazole parental compounds, the isostere benzimidazoles synthesised turned out much more active.

# Section II

## 2. Host targeting antivirals

### 2.1. Introduction

Inhibitors of specific host functions on which many different viruses depend have the additional potential to treat a wide range of viral infections (broad-spectrum antivirals)<sup>25</sup>. Most enveloped viruses use the endoplasmic reticulum (ER) glycoprotein quality control (QC) machinery for the biosynthesis of the glycoproteins needed for the assembly of their envelope<sup>26</sup>. The endoplasmic reticulum is the largest membrane-bound

organelle in the cell; it is the site of many cellular processes, including protein synthesis, modification and transport, lipid synthesis and calcium homeostasis regulation by calcium storage and secretion<sup>27-30</sup>. The ER forms a contiguous structure of interconnected sheets and tubules - with a common luminal space - that spreads from the nuclear envelope (NE) to the cell cortex. At the morphological level, it can be classified into two types, smooth ER (SER) and rough ER (RER). The RER has a sheet-like morphology and is characterised by the presence of ribosomes associated with the cytosolic face of the ER. The biosynthesis of secretory and membrane proteins begins with translation of mRNA in the cytosol by ribosomes; then the nascent polypeptide enters the ER lumen through the translocon<sup>27,31,32</sup>. Once inside the ER, the newly synthesized protein is destined to start proper folding and acquire post-translational modifications such as N-glycosylation. Proteins undergo folding in the ER with the support of chaperones and folding enzymes, such as glycosyltransferases, glycosidases, disulphide isomerases, oxidoreductases, and peptidyl-prolyl isomerases, to name only a few. When the protein is properly folded, if it is destined for secretion, it leaves towards the Golgi and from here it carries on to its final destination<sup>28</sup>. Despite the many ER enzymes and complexes dedicated to proper folding of proteins, a fraction of these proteins never achieves their native and functional form and remains misfolded or it aggregates. Misfolded proteins can either be retained in the ER, entering the calnexin/calreticulin cycle or be shunted to the ER-associated

degradation (ERAD) pathway, ensuring that aberrant polypeptides do not progress down the secretory pathway<sup>33</sup>. Interestingly, there are several links between activation of ER stress response pathways and human pathology. One such example is the viral infection of a cell, consisting of a series of events such as viral entry, viral mRNA synthesis and processing, viral polypeptide synthesis and modification, viral genome replication and maturation. Similarly, to some intracellular parasites, viruses rely on the host's cellular machinery to complete their replication. In particular, viral glycoproteins are processed in the host's ER, which is oversaturated during viral infections. The resulting ER stress in turn activates the unfolded protein response (UPR), potentially limiting the viral life cycle. For this reason, viruses have evolved mechanisms to cope with ER stress and UPR<sup>34</sup>. Preventing the maturation of infectious viruses by inhibiting ERQC provides an interesting target for the development of broad-spectrum antivirals. It has been demonstrated that inhibition of ER  $\alpha$ -glucosidase II ( $\alpha$ -GluII) by iminosugars, glycomimetics originally isolated from plants<sup>35,36</sup> resulted in antiviral activity against many virus families *in cellula* and *in vivo*<sup>37-40</sup>.

### **2.1.1. An overview of ERQC machinery**

The ER hosts several resident proteins that are involved in the “folding process”. The newly synthesized protein enters the ER and undergoes a series of modifications “helped” by molecular



chaperones and folding enzymes that work all together with the aim of ensuring proper folding of glycoproteins destined for the secretory pathway. Modification of a glycoprotein begins as soon as it enters the ER and continues till the very last moment before the glycoprotein is released from the ER. Chaperones and folding enzymes are very versatile proteins that can interact with many different clients. Briefly, the synthesis, maturation and release of glycoproteins from the ER can be described starting with the translation process in ribosomes seated on translocons on the cytosolic surface of the RER. The nascent glycoprotein is translocated inside the ER lumen where an oligosaccharyltransferase (OST) attaches pre-assembled glycans (comprised of three glucoses, nine mannoses – three for each arm – and two N-acetyl glucosamines, Glc<sub>3</sub>Man<sub>9</sub>-GlcNAc<sub>2</sub>) to an asparagine (Asn) residue of an N-glycosylation motif on the nascent polypeptide. Disulphide bonds formation and removal of the amino-terminal signal sequence by a peptidase also take place at this stage. ER  $\alpha$ -glucosidase I ( $\alpha$ -GluI) removes the terminal glucose residue from the N-linked glycan, generating a diglucosylated species<sup>28</sup>. Diglucosylated proteins are processed by ER  $\alpha$ -glucosidase II ( $\alpha$ -GluII), the main ERQC glycosyl hydrolase, admitting folding client glycoproteins into ERQC and releasing them from it. ER  $\alpha$ -GluII first cleaves the middle glucose from the glycan, competing with malectin, an ER lectin that has a

high specificity for diglucosylated glycans<sup>41</sup>. The resulting monoglucosylated glycan is recognised by ER lectins calnexin (CNX) and calreticulin (CRT). Calnexin is a membrane-associated protein and calreticulin its soluble paralogue; they both contain a globular carbohydrate-binding domain that is specific for an Asn-bound GlcMan<sub>9</sub>-GlcNAc<sub>2</sub> glycan. CNX and CRT promote proper folding of glycoproteins mostly by associating with a number of chaperones and foldases, thus preventing aggregation, stabilizing the folding process and retaining misfolded proteins in the ER. Once the same  $\alpha$ -GluII removes the final glucose residue from the N-linked glycan, the glycoprotein is released by the lectins and associated chaperones. The resulting deglucosylated glycoprotein is free to complete its folding process and leave the ER, but not before being checked by the so-called ER glycoprotein folding checkpoint enzyme, UDP-Glucose glycoprotein glucosyltransferase (UGGT) that controls if any misfolded state persist. If so, UGGT transfers a glucose (from UDP-Glucose) to a glycan on the misfolded glycoprotein, regenerating a mono-glucosylated glycan that can bind the ER lectins and undergo further attempts of assisted folding. The shuttling of glycoproteins between the ER lectins and UGGT, with  $\alpha$ -GluII and UGGT removing and adding a glucose on the Man<sub>9</sub>-glycan respectively, is called the calnexin/calreticulin cycle<sup>28</sup>. A misfolded glycoprotein that

fails to attain its native fold has one destiny: the endogenous ER quality control system called ER-associated degradation (ERAD).

As iminosugars<sup>40</sup> are active against viral replication by inhibiting ER  $\alpha$ -GluI and ER  $\alpha$ -GluII it can be hypothesised that UGGT inhibition would also potentially have antiviral potential. Whereas UGGT activity is beneficial to avoid premature secretion of misfolded glycoproteins under physiological conditions, during a viral infection, when the ER is stressed by a big amount of glycoproteins that need the ERQC machinery to proper fold glycoproteins and assembly mature virions, UGGT inhibition could lead to antiviral effects, just as ER  $\alpha$ -GluI and  $\alpha$ -GluII inhibition. This idea is supported by the data published in 2017, proving that heterozygous knock-out of UGGT1 in mice reduced viral pathogenicity<sup>42</sup>.

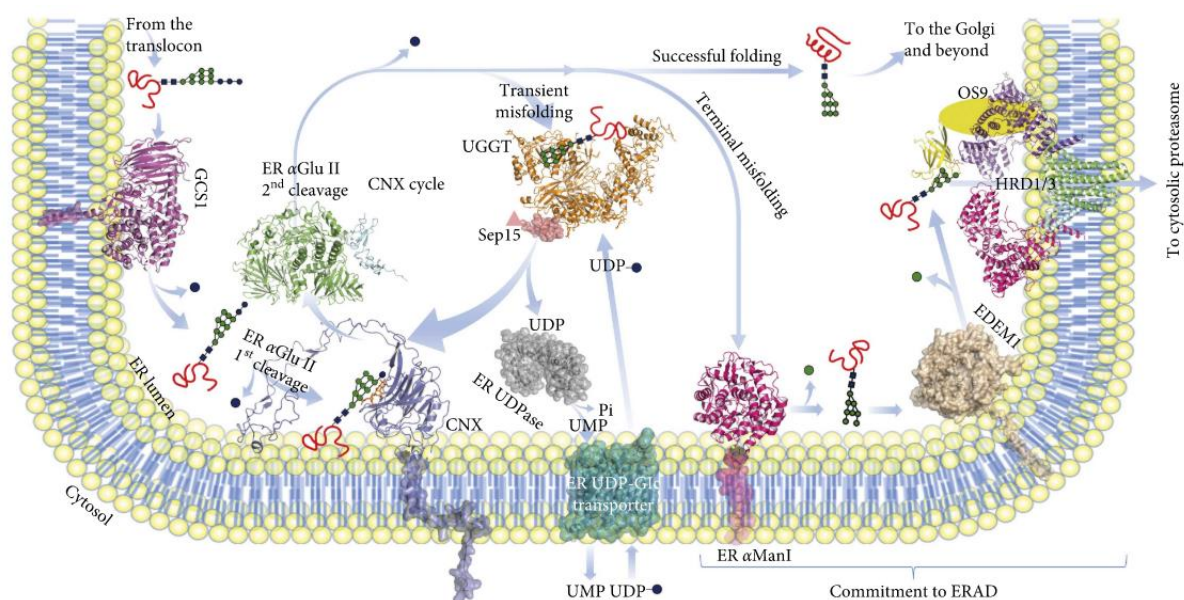


Figure 6 Structural view of ERQC/ERAD <sup>43</sup>

### 2.1.2. UDP-glucose: glycoprotein glucosyltransferase (UGGT)

UDP-glucose: glycoprotein glucosyltransferase is the glycoprotein misfold sensor in the ER. This 170-kDa,  $\text{Ca}^{2+}$ -dependent enzyme has for a long time resisted structural determination<sup>44</sup>, until the Zitzmann group published in 2017 the first crystal structures of a full-length eukaryotic protein (*Chaetomium thermophilum* UGGT), showing the intrinsic interdomain flexibility of the protein<sup>45</sup>. UGGT structure is organised as follows: at the N-terminal portion, ~1,200 residues comprise the enzyme's misfold sensor domain, whereas at the C-terminal portion ~300 residues encode the catalytic domain: a glucosyltransferase 24 family (GT24) A-type domain.

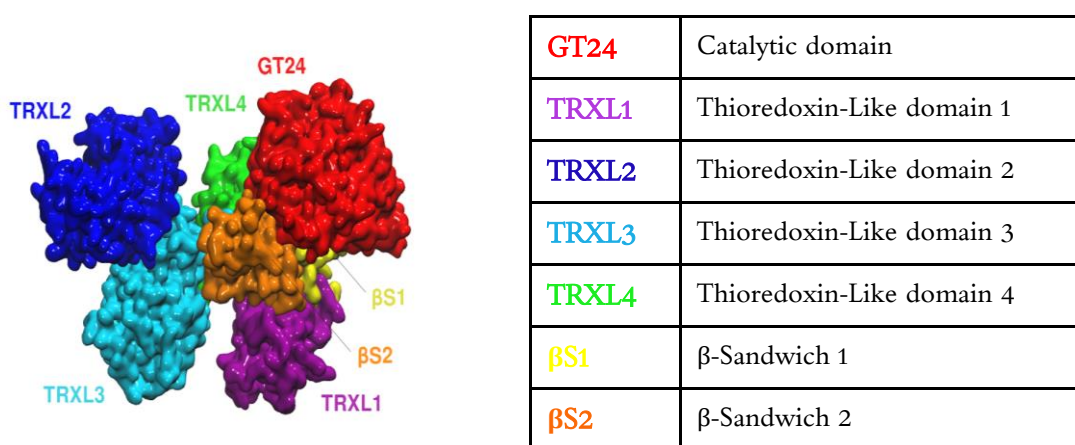


Figure 7 UGGT structural composition

More in detail, Roversi *et al.* first described the UGGT domain architecture: four thioredoxin-like (TRXL) domains in the N-terminal portion, followed by two seven-stranded  $\beta$ -sandwiches ( $\beta$ S1 and  $\beta$ S2) that clasp the C-terminal catalytic glucosyltransferase (GT) domain (Figure 8). Unusually, UGGT folds in a non-sequential order<sup>45</sup>.

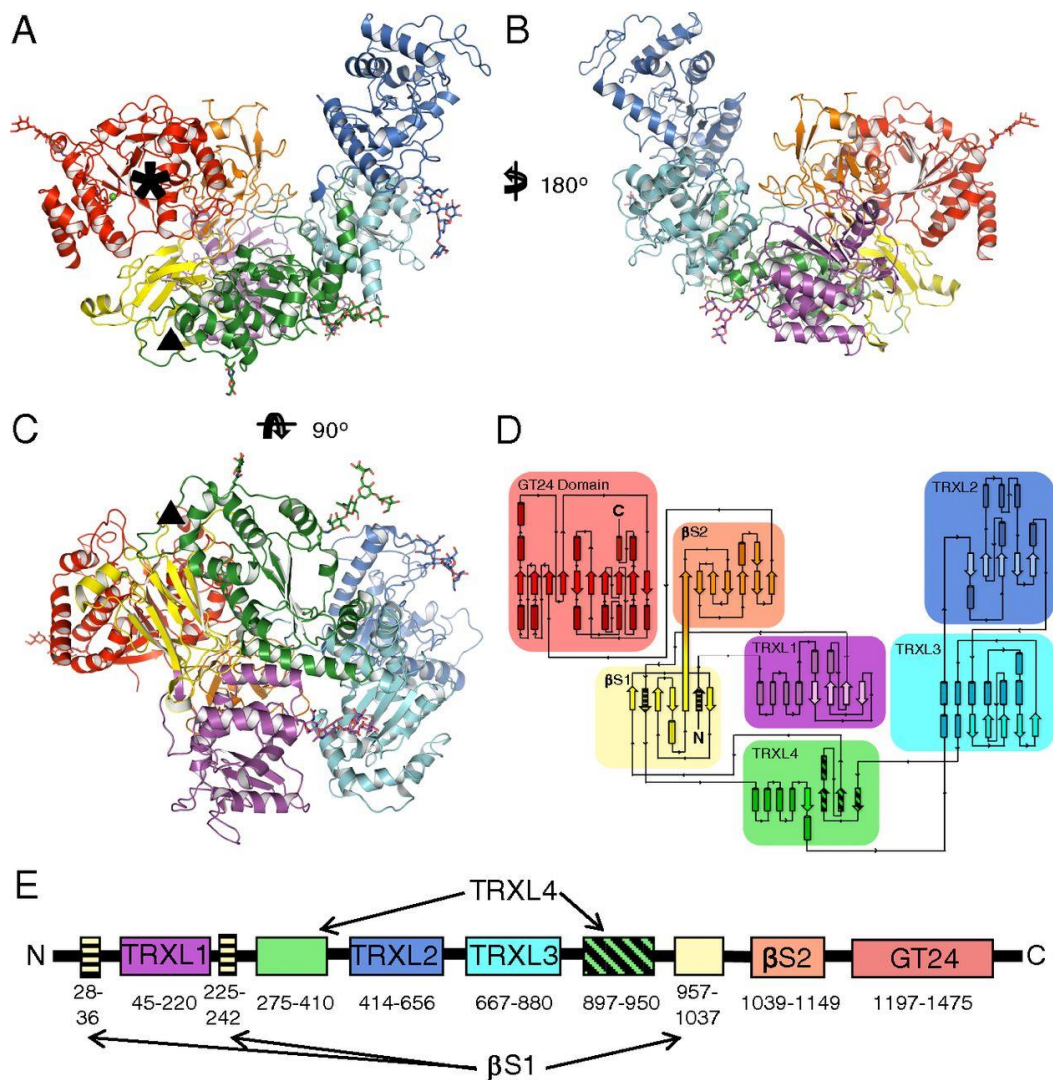


Figure 8 Structure and topology of CtUGGT. (A–C) Three orthogonal views of crystal structure of the CtUGGT (PDB ID 5MU1). The seven domains are in cartoon representation: TRXL1 (purple), TRXL2 (blue), TRXL3 (cyan), TRXL4 (green),  $\beta$ S1 (yellow),  $\beta$ S2 (orange), and GT (red). The five N-linked glycans

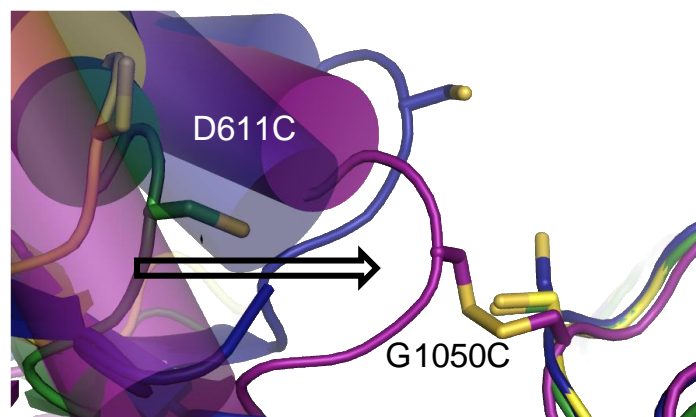
(at N56, N329, N638, N894, and N1227) are in stick representation. Six conserved cysteine residues form three disulfide bonds, one in the TRXL1 domain (CtUGGT C138–C150) and two in the GT domain (CtUGGT C1330–C1423 and C1419–C1437). The  $\text{Ca}^{2+}$  ion bound at the conserved nucleotide-sugar coordinating <sup>1302</sup>DAD<sup>1304</sup> motif is represented as a green sphere. A black asterisk marks the catalytic site. A black triangle marks the dangling ends around the disordered loop (CtUGGT residues 243–285, between the second strand of the  $\beta$ S1 sandwich and the N-terminal part of the TRXL4 domain). (D and E) The 2- and 1-dimensional topological diagrams of CtUGGT, respectively. The first and second strands of  $\beta$ -sandwich  $\beta$ S1 (residues 28–36 and 225–242, respectively, striped yellow) flank the part of sequence encoding TRXL1 (residues 45–220, magenta), with the rest of  $\beta$ S1 encoded by a portion of sequence (residues 957–1,037, yellow) more than 700 residues downstream of the  $\beta$ S1 sandwich's second strand. TRXL1 adopts a noncanonical subdomain structure, different from all structures of Pfam family PF01323 members, in which an N-terminal  $\alpha$ -helical subdomain is followed by a C-terminal thioredoxin subdomain. The N- and C-terminal halves of TRXL4 (residues 275–410 and 897–950, the latter portion in striped green) occur in the standard order but are separated in sequence by TRXL2 and TRXL3 <sup>45</sup>.

UGGT activity has been studied using both native and misfolded glycoprotein substrates<sup>44,46–48</sup>, glycopeptides or small synthetic clients<sup>49,50</sup>. These studies have proven UGGT as highly selective for re-glucosylation of misfolded over folded substrates. The enzyme re-glucosylates only glycans in misfolded proteins, and it is able to cope with a variety of distances between the acceptor  $\text{Man}_9\text{GlcNAc}_2$  glycan and the misfold site. No UGGT inhibitors are known except for UDP, its own reaction product<sup>51</sup>.

Across UGGT sequences in higher eukaryotes, the GT catalytic domain surface shows high conservation, especially in the binding regions for UDP-Glc. TRXL domains at the N-terminus of UGGT are conserved in other ER chaperones<sup>52–54</sup>, which all present domains of the same fold. All these proteins have in common the ability of binding misfolded proteins<sup>45</sup>. The mechanism by which UGGT recognises and reglucosylates misfolded glycoproteins, which differ



considerably in size and shape remains unknown, but it has been postulated that the enzyme's interdomain conformational flexibility can explain UGGT versatility. A model in which UGGT can adapt to substrate glycoproteins of different size and fold is supported by three X-ray *Ct*UGGT (wild type protein) structures, labelled: 'open' (PDB ID 5MZO), 'intermediate' (PDB ID 5MU1) and 'closed' (PDB ID 5N2J). Additionally, the double Cysteine mutant *Ct*UGGT<sup>D611C/G1050C</sup>, especially engineered to form an extra disulphide bridge between the TRXL2 and  $\beta$ S2 domains, was trapped in a 'closed-like' conformation (PDB ID 5NV4). In activity assays of UGGT-mediated re-glucosylation of a misfolded glycoprotein substrate, the *Ct*UGGT<sup>D611C/G1050C</sup> mutant had lower activity than wild-type *Ct*UGGT, suggesting that the mobility of the TRXL2 domain is important for re-glucosylation of some substrates<sup>45</sup>.



*Figure 9 Zoom in to the interface between the TRXL2 and  $\beta$ S2 domains, with the Cys residues introduced by the mutations D611C and G1050C modeled in stick representation for the WT structures, and the observed disulfide bridge in the structure of the *Ct*UGGT<sup>D611C/G1050C</sup> double mutant in magenta (PDB ID 5nv4). The double mutant *Ct*UGGT<sup>D611C/G1050C</sup> prevents *Ct*UGGT from acquiring the open/intermediate conformations.<sup>45</sup>*

## 2.2. Project Aim

The published *CtUGGT* conformations with PDB ID 5MZO, 5MU1, 5N2J and 5NV4 mainly differ in the spatial organization of domains TRXL2 and TRXL3 (Figure 10 A). Across these structures, the TRXL2 domain is rotated by different amounts with respect to the rest of the protein and adopts different degrees of proximity to it. The TRXL3 domain instead appears in the same relative conformation in all structures except for the ‘open’ one (Figure 10, cyan in A and green in B and C), in which the TRXL3 and TRXL1 (Figure 10, respectively cyan and purple in A) domains move apart, leading to the opening of a large cleft between them.

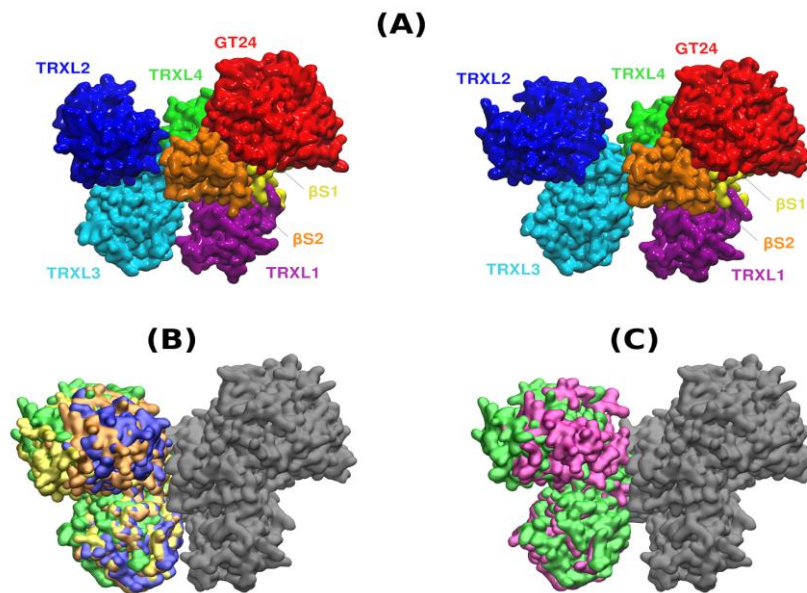


Figure 10 *CtUGGT* conformations

In order to better describe these conformations, extensive Molecular Dynamics (MD) experiments on the previously described crystal structure of *CtUGGT* (5MZO, 5MU1, 5N2J, 5NV4) were run by Argentinian



collaborators<sup>♦</sup>. The majority of UGGT's overall inter-domain motions can be described in terms of three types of main movements of the N-terminal misfold recognition domains: bending, twisting and clamping.

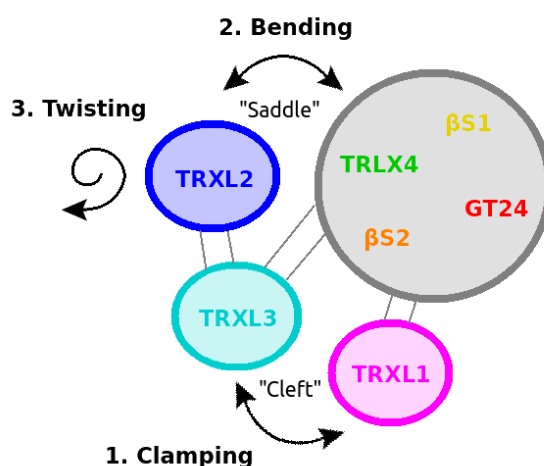


Figure 11 Simplified representation of CtUGGT overall movements. 'Clamping' movement between domains TRXL3 and TRXL1; bending' movement between TRXL2 and the core comprised of domains 'GT24-βS1-βS2-TRXL4'; 'twisting' movement of TRXL2 with respect to TRXL3. The grey area represents the strong structural inter-domain orientation invariance of the TRXL4-βS1-βS2-GT24 domains.

‘Clamping’ describes the changes in the distance between the centres of mass of the TRXL1 and TRXL3 domains and measures the openness of the cleft between them. ‘Bending’ describes the changes in the angle between the centres of mass of the TRXL1, TRXL2 and TRXL3 domains, and measures the proximity of the TRXL2 and GT24 domains. Lastly, ‘twisting’ describes the changes in the dihedral angle between the Cα atoms of

<sup>♦</sup> Carlos P. Modenutti Juan I. Blanco Capurro, Marcelo A. Martí. Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pab. II (CE1428EHA), Buenos Aires, Argentina.

residues Y518, F466, T863 and I735 (the first two residues in the TRXL2 and the last two in the TRXL3 domain), thus informing on the relative orientation between the TRXL2 and TRXL3 domains.

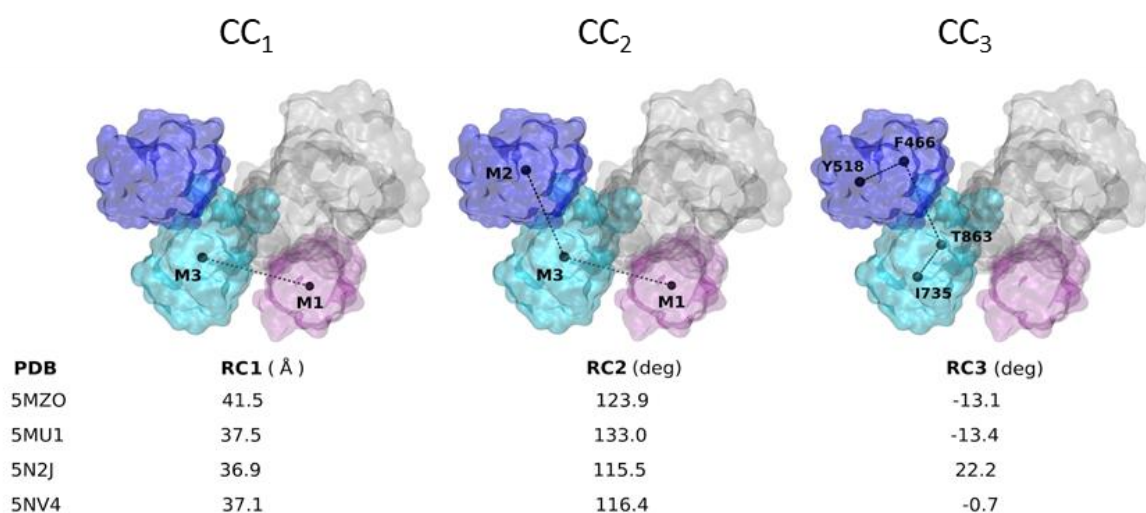


Figure 12 Reaction coordinates for describing the CtUGGT conformational states observed in the X-ray structures; domain colour code is the same as in figure 1: TRXL1 (magenta), TRXL2 (blue) and TRXL3 (cyan). Along 'CC1', the 'clamp' coordinate measures the openness of the cleft between the TRXL1 and TRXL3 domains; along 'CC2', the 'bend' coordinate measures the distance between the TRXL2 and GT24 domains across the central saddle; along 'CC3' the 'twist' coordinate changes with the relative orientation of the TRXL2 and TRXL3 domains.

The first objective of this project was to probe the functional role of the 'clamping' motion. To achieve this, three novel double cysteine CtUGGT mutants were engineered: CtUGGT<sup>G177C/A786C</sup>, CtUGGT<sup>G179C/T742C</sup> and CtUGGT<sup>S180C/T742C</sup>, designed to form disulphide bridges across the interfaces of the TRXL1-TRXL3 domains, in order to verify the contribution of the TRXL1/TRXL3 inter-domain clamping movement to UGGT re-glucosylation activity. We also set out to clone, express and purify these three mutants of CtUGGT lacking domains TRXL1, TRXL2 and TRXL3:

the mutants designed are *CtUGGT-ΔTRXL1*, lacking residues 42-224; *CtUGGT-ΔTRXL2*, lacking residues 417-650 (cloned, expressed, purified and crystallized by Jonathan Rushton<sup>♦</sup>); *CtUGGT-ΔTRXL3*, lacking residues 666-870. These mutants were used to probe the contributions of individual UGGT domains to the enzyme's activity.

The second objective of this project was to find new small-molecule ligands for UGGT by a fragment-based lead discovery (FBLD) approach; these ligands would then be the starting points for medicinal chemistry to develop the first class of UGGT inhibitors; this will be described in paragraph 2.6.

## 2.3. Experimental section

### 2.3.1. Cloning

#### 2.3.1.1. *CtUGGT*<sup>G177C/A786C</sup>

First, mutation of the *CtUGGT* into *CtUGGT*<sup>A786C</sup> was effected starting from the gene of *CtUGGT* inserted in Litmus28i (an optimal vector for mutagenesis experiments) as follows: 12.5 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs) were added to 1.25 µl of each forward and reverse primers (1\_F and 1\_R) at

---

<sup>♦</sup> Jonathan Rushton, master student at Department of Biochemistry – University of Oxford, Oxford OX1 3QU, England, United Kingdom.

10  $\mu$ M (see Appendix), 1  $\mu$ l of *Ct*UGGT:Litmus28i DNA 1 ng/ $\mu$ L and 9  $\mu$ l of nuclease-free water, obtaining a 25  $\mu$ l final volume. PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 60 °C for 20 seconds; step 4: 72 °C for 135 seconds. Steps 2-4 were repeated 25 times; step 5: 72 °C for 2 minutes. Kinase, Ligase & DpnI (KLD) treatment was made to delete the parental fragments present in the mixture: 1  $\mu$ L of PCR product was mixed with 5  $\mu$ L of 2X KLD Reaction Buffer, 1  $\mu$ L of 10X KLD Enzyme Mix (both from New England Biolabs) and 3  $\mu$ L of nuclease-free water. The mixture was incubated at room temperature for 5 minutes. The KLD reaction mixture was used to transform *E. coli* DH-5 $\alpha$  chemically competent cells using the following protocol: 5  $\mu$ l of KLD reaction mix were added to a tube of thawed New England BioLabs DH-5 $\alpha$  competent *E. coli* cells on ice, and mixed gently for a few seconds; after the transformation, the bacteria were incubated on ice for 30 minutes, heat-shocked at 42 °C for 30 seconds and incubated on ice again for 5 minutes. 950  $\mu$ l of SOC media was added to a final volume of 1 ml and the mixture was incubated for 1 hour at 37 °C with gentle shaking at 200/300 rpm. 100  $\mu$ l of the bacteria were spread onto a pre-warmed (37 °C) LB agar culture plate containing carbenicillin (0.1 mg/ml). The plate was incubated at 37 °C overnight. Colony-PCR (see protocol in 3.4.2.7) was performed on DNA from various colonies (using T7\_F and T7\_R primers, see Appendix) and the DNA

obtained was loaded on a 1% w/v agarose gel and run for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were used to inoculate 5 ml LB supplemented with 0.1 mg/mL carbenicillin. This was used to make glycerol stocks (mixing 16% glycerol with 84% bacteria and freezing and storing at -80 °C). The DNA was mini-prepped (Qiagen) obtaining 50 µl at 400 ng/µl concentration and after sequencing was used as template for introduction of second mutation. To obtain the double mutant *CtUGGT*<sup>G177C/A786C</sup>, the second mutation G177C was introduced starting from the gene of *CtUGGT*<sup>A786C</sup> in Litmus28i as follows: 12.5 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs) were added to 1.25 µl of each forward and reverse primers (2\_F and 2\_R, see Appendix) at 10 µM, 1 µl of *CtUGGT*<sup>A786C</sup>:Litmus28i DNA at 1 ng/µl and 9 µl of nuclease-free water, obtaining a 25 µl final volume. PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 66 °C for 30 seconds; step 4: 72 °C for 135 seconds; steps 2-4 were repeated 25 times. Step 5: 72 °C for 2 minutes. After KLD treatment (see above) *E. coli* DH-5α chemically competent cells were transformed with the DNA as described previously. Colony-PCR was performed on DNA from various colonies (using T7\_F and T7\_R primers) and the DNA obtained was loaded on a 1% w/v agarose gel and run for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with

amplified DNA of the appropriate size, which were used to inoculate 5 ml LB supplemented with 0.1 mg/ml carbenicillin. This was used to make glycerol stocks (see above). The DNA was mini-prepped (Qiagen), obtaining 50 µl at 700 ng/µl concentration and sequenced to confirm the mutations. The *CtUGGT*<sup>G177C/A786C</sup> insert in Litmus 28i was cloned into the pHL-sec vector<sup>55</sup> to contain a hexa-His Tag at the C-terminus. DNA for pHL-sec was linearised using AgeI and KpnI restriction enzymes at 37 °C for 16 hours in Cutsmart™ buffer (New England Biolabs). The restriction digest was then run on a 0.8% w/v agarose gel at 150 V for 1 hour. The linearised vector was cut from the gel and purified with a QIAquick gel extraction kit (Qiagen). PCR was performed on the *CtUGGT*<sup>G177C/A786C</sup> insert in Litmus 28i, 1 µl (1 ng/µl) added at 25 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs), 2.5 µl of each forward and reverse primers (pHL-sec\_*CtUGGT*\_F and pHL-sec\_*CtUGGT*\_R, see Appendix) and 19 µl of nuclease-free water. PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 62 °C for 30 seconds; step 4: 70 °C for 150 seconds; step 2-4 were repeated 35 times; step 5: 72 °C for 2 minutes. The PCR products were run on a 1% w/v agarose gel at 150 V for 1 hour and the amplified insert was cut from the gel and purified with the same QIAquick gel extraction kit. A Gibson Assembly was then performed using the gel-purified PCR-amplified *CtUGGT*<sup>G177C/A786C</sup> insert mixed with gel-purified

linearised pHL-sec at a ratio of 3:1 with NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) using suggested protocol, for 1 hour at 50 °C. 2 µl of this ligation product was added to 50 µl XL10-Gold Ultracompetent cells (Agilent), following the transformation guideline protocol. The cells were then plated on 0.1 mg/ml carbenicillin agar plates and incubated overnight at 37 °C. Colony-PCR was performed on DNA from various colonies (using pHL-sec\_*CtUGGT\_F* and pHL-sec\_*CtUGGT\_R* primers, see Appendix) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size. The DNA was mini-prepped (Qiagen) and after sequencing, a glycerol stock of one of the previously described colonies was used to inoculate 5 ml LB supplemented with 0.1 mg/ml carbenicillin and incubated overnight at 37 °C. This culture was then used to inoculate 200 ml LB supplemented with 0.1 mg/ml carbenicillin and left to incubate at 37 °C, shaking at 110 rpm. Upon reaching an OD<sub>600nm</sub> of 2.0, the cells were spun down at 3320xg for 18 minutes. The pellets were resuspended, and the DNA purified using a Maxiprep kit (Qiagen), following the recommended protocol to obtain 0.5 ml of *CtUGGT*<sup>G177C/A786C</sup>:pHL-sec plasmid DNA at 605 ng/µl.

### 2.3.1.2. *CtUGGT*<sup>G179C/T742C</sup>

First mutation of the *CtUGGT* into *CtUGGT*<sup>T742C</sup> was effected starting from the gene of *CtUGGT* in Litmus28i as follows: 12.5 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs) were added to 1.25 µl of each forward and reverse primers at 10 µM (3\_F and 3\_R, see Appendix), 1 µl of *CtUGGT* DNA 1 ng/µL and 9 µl of nuclease-free water, obtaining a 25 µl final volume; PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 59 °C for 20 seconds; step 4: 72 °C for 135 seconds; step 2-4 were repeated 25 times; step 5: 72 °C for 2 minutes. Kinase, Ligase & DpnI (KLD) treatment was made as described above. The KLD reaction mixture was used to transform *E. coli* DH-5α chemically competent cells using the same protocol previously described. 100 µl of the mixture was spread onto a pre-warmed (37 °C) LB agar culture plate containing Carbenicillin (0.1 mg/ml). The plate was incubated at 37 °C overnight. Colony-PCR was performed on DNA from various colonies (using T7\_F and T7\_R primers) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were used to inoculate 5 ml LB supplemented with 0.1 mg/mL carbenicillin. This was used to make glycerol stocks. The DNA was mini-prepped (Qiagen) obtaining 50 µl at 500 ng/µl concentration and after sequencing was used as



template for introduction of second mutation. To obtain double mutant *CtUGGT*<sup>G179C/T742C</sup>, second mutation G179C was introduced starting from the gene of *CtUGGT*<sup>T742C</sup> in Litmus28i as follows: 12.5 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs) were added to 1.25 µl of each forward and reverse primers (4\_F and 4\_R, see Appendix) at 10 µM, 1 µl of *CtUGGT* DNA 1 ng/µL and 9 µl of nuclease-free water, obtaining a 25 µl final volume; PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 63 °C for 30 seconds; step 4: 72 °C for 135 seconds; step 2-4 were repeated 25 times; step 5: 72 °C for 2 minutes. After KLD treatment (see above) *E. coli* DH-5α chemically competent cells were transformed with the DNA as described previously. Colony-PCR was performed on DNA from various colonies (using T7\_F and T7\_R primers) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were used to inoculate 5 ml LB supplemented with 0.1 mg/ml carbenicillin. This was used to make glycerol stocks. The DNA was mini-prepped (Qiagen) obtaining 50 µl at 650 ng/µl concentration and sequenced to confirm the mutations. The *CtUGGT*<sup>G179C/T742C</sup> insert in Litmus 28i was cloned into the pHL-sec vector to contain a hexa-His Tag at the C-terminus. Linearised pHL-sec DNA vector was obtained as previously described. PCR was performed on the

*CtUGGT*<sup>V178C/A786C</sup> insert in Litmus 28i, 1 µl (1 ng/µl) added at 25 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs), 2.5 µl each forward and reverse primer (pHL-sec\_*CtUGGT*\_F and pHL-sec\_*CtUGGT*\_R) and 19 µl of nuclease-free water, using the same PCR protocol as above. The PCR products were run on a 1% w/v agarose gel at 150 V for 1 hour and the amplified insert was cut from the gel and purified with the same QIAquick gel extraction kit. A Gibson Assembly was then performed using the gel-purified PCR-amplified *CtUGGT*<sup>G179C/T742C</sup> insert mixed with gel-purified linearised pHL-sec at a ratio of 3:1 with NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) using suggested protocol, for 1 hour at 50 °C. 2 µl of this ligation product was added to 50 µl XL10-Gold Ultracompetent cells (Agilent), following transformation guideline protocol. The cells were then plated on 0.1 mg/mL carbenicillin agar plates and incubated overnight at 37 °C. Colony-PCR was performed on DNA from various colonies (using pHL-sec\_*CtUGGT*\_F and pHL-sec\_*CtUGGT*\_R primers) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were mini-prepped, glycerol-stocked, sequenced and maxi-prepped as described earlier, to obtain 0.5 ml of *CtUGGT*<sup>G179C/T742C</sup>:pHL-sec plasmid DNA at 1830 ng/µl.

### 2.3.1.3. *CtUGGT*<sup>S180C/T742C</sup>

Using the previously obtained *CtUGGT*<sup>T742C</sup> in Litmus28i, second mutation S180C was introduced to obtain double mutant *CtUGGT*<sup>S180C/T742C</sup>. Mutation was carried out as follows: 12.5 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs) were added to 1.25 µl of each forward and reverse primer (5\_F and 5\_R, see Appendix) at 10 µM, 1 µl of *CtUGGT* DNA 1 ng/µl and 9 µl of nuclease-free water, obtaining a 25 µl final volume; PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 68 °C for 30 seconds; step 4: 72 °C for 135 seconds; step 2-4 were repeated 25 times; step 5: 72 °C for 2 minutes. After KLD treatment (see above) *E. coli* DH-5α chemically competent cells were transformed with the DNA as described previously. Colony-PCR was performed on DNA from various colonies (using T7\_F and T7\_R primers) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were used to inoculate 5 ml LB supplemented with 0.1 mg/ml carbenicillin. This was used to make glycerol stocks. The DNA was mini-prepped (Qiagen) obtaining 50 µl at 650 ng/µl concentration and sequenced to confirm the mutations. The *CtUGGT*<sup>S180C/T742C</sup> insert in Litmus 28i was cloned into the pHL-sec vector to contain a hexa-His Tag at the C-terminus. Linearised pHL-sec DNA vector was

obtained as previously described. PCR was performed on the *CtUGGT*<sup>S180C/T742C</sup> insert in Litmus 28i, 1 µl (1 ng/µl) added at 25 µl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs), 2.5 µl each forward and reverse primer (pHL-sec\_*CtUGGT*\_F and pHL-sec\_*CtUGGT*\_R) and 19 µl of nuclease-free water, using the same PCR protocol as above. The PCR products were run on a 1% w/v agarose gel at 150 V for 1 hour and the amplified insert was cut from the gel and purified with the same QIAquick gel extraction kit. A Gibson Assembly was then performed using the gel-purified PCR-amplified *CtUGGT*<sup>S180C/T742C</sup> insert mixed with gel-purified linearised pHL-sec at a ratio of 3:1 with NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) using suggested protocol, for 1 hour at 50 °C. 2 µl of this ligation product was added to 50 µl XL10-Gold Ultracompetent cells (Agilent), following transformation guideline protocol. The cells were then plated on 0.1 mg/mL carbenicillin agar plates and incubated overnight at 37 °C. Colony-PCR was performed on DNA from various colonies (using pHL-sec\_*CtUGGT*\_F and pHL-sec\_*CtUGGT*\_R primers, see Appendix) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were mini-prepped, glycerol-stocked,

sequenced and maxi-prepped as described earlier, to obtain 0.5 ml of *CtUGGT*<sup>S180C/T742C</sup>:pHL-sec plasmid DNA at 3650 ng/μl.

#### 2.3.1.4. *CtUGGT-ΔTRXL1*

*CtUGGT-ΔTRXL1* construct lacks residue *CtUGGT* 42-224. The deletion of the *CtUGGT* TRXL1 domain was performed starting from the gene of *CtUGGT* in Litmus28i as follows: 12.5 μl of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs) were added to 1.25 μl of each forward and reverse primer at 10 μM (6\_F and 6\_R, see Appendix), 1 μl of *CtUGGT* DNA 1 ng/μl and 9 μl of nuclease-free water, obtaining a 25 μl final volume; PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 65 °C for 20 seconds; step 4: 72 °C for 130 seconds; step 2-4 were repeated 25 times; step 5: 72 °C for 2 minutes. KPN treatment and *E. coli* XL10-Gold Ultracompetent cells (Agilent) transformation and plating as described before Colony-PCR was performed on DNA from various colonies (using T7\_F and T7\_R primers) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were miniprepped obtaining 50 μl of *CtUGGT-ΔTRXL1*:Litmus28i plasmid DNA at 400 ng/μl, and sequenced to check if deletion was obtained. The *CtUGGT-ΔTRXL1* insert in Litmus 28i was cloned into the pHL-sec vector to contain a hexa-His Tag at the C-terminus as described before for the double

Cys mutants to obtain 0.5 ml CtUGGT- $\Delta$ TRXL1:pHL-sec plasmid DNA at 600 ng/ $\mu$ l.

#### 2.3.1.5. *CtUGGT- $\Delta$ TRXL3*

*CtUGGT- $\Delta$ TRXL3* construct lacks residue CtUGGT 666-870. The deletion of the *CtUGGT* TRXL3 domain was performed starting from the gene of *CtUGGT* in Litmus28i as follows: 12.5  $\mu$ l of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs) were added to 1.25  $\mu$ l of each forward and reverse primer (7\_F and 7\_R, see Appendix) at 10  $\mu$ M, 1  $\mu$ l of *CtUGGT* DNA 1 ng/ $\mu$ l and 9  $\mu$ l of nuclease-free water, obtaining a 25  $\mu$ l final volume; PCR protocol: step 1: 98 °C for 30 seconds; step 2: 98 °C for 10 seconds; step 3: 62 °C for 20 seconds; step 4: 72 °C for 130 seconds; step 2-4 were repeated 25 times; step 5: 72 °C for 2 minutes. KPN treatment and *E. coli* XL10-Gold Ultracompetent cells (Agilent) transformation and plating as described before Colony-PCR was performed on DNA from various colonies (using T7\_F and T7\_R primers) and the DNA obtained was run on a 1% w/v agarose gel for 50 minutes at 150 V. Analysis of this gel allowed identification of colonies with amplified DNA of the appropriate size, which were minipreped obtaining 50  $\mu$ l of *CtUGGT- $\Delta$ TRXL3*:Litmus28i plasmid DNA at 450 ng/ $\mu$ l, and sequenced to check if deletion was obtained. The *CtUGGT- $\Delta$ TRXL3* insert in Litmus 28i was cloned into the pHL-sec vector to contain a hexa-His Tag at the C-terminus as described before for the double

Cys mutants to obtain 0.5 ml *CtUGGT-ΔTRXL3:pHL-sec* plasmid DNA at 690 ng/μl.

### **2.3.2. Protein expression and purification**

A 200 ml volume (400 ml for *CtUGGT<sup>G179C/T742C</sup>*) of HEK293F cells culture was transfected and incubated for 4 days, supernatant processed as described in the methods, and run on a 5 ml HisTrap HP column. Eluted fractions containing the right size protein (detected by SDS-PAGE) were pooled and concentrated in a 50 KDa MWCO centrifugal concentrator before loading on a Superdex 200 16/60 column for size exclusion chromatography. Eluted fractions were analysed by SDS-PAGE. Protein containing fractions were pooled and concentrated as before, flash-frozen in liquid nitrogen in 100 μl aliquots and stored at -80 °C. 0.8 ml at 7.91 mg/ml of *CtUGGT<sup>G177C/A786C</sup>* were obtained, 0.8 ml at 6.28 mg/ml of *CtUGGT<sup>G179C/T742C</sup>* were obtained, 0.1 ml at 0.4 mg/ml of *CtUGGT-ΔTRXL1* were obtained and 0.2 ml at 12.80 mg/ml of *CtUGGT-ΔTRXL3* were obtained.

A 400 ml volume of HEK293F cells culture was transfected with *CtUGGT<sup>S180C/T742C</sup>* DNA and incubated for 4 days, supernatant processed as described in the methods, and run on a 5 ml HisTrap HP column, but failed to bind to HisTrap column. The HisTrap flow-

through (FT), 600 ml at about 120 nM NaCl, was collected and the protein checked by SDS-PAGE with positive result. It was three times filtered through 1  $\mu$ m filter, then 0.45  $\mu$ m filter, then a 0.22  $\mu$ m filter. It was diluted to 1 litre with H<sub>2</sub>O and titrated till pH 8.5 with NaOH, then loaded onto a HiPrep Q HP 16/60 anion exchange column and purified as described in methods. The SDS-PAGE gel showed *CtUGGT*<sup>S180C/T742C</sup> in fractions A3 and A4 which were pooled, concentrated and dialysed to 20 mM MES pH 6.5, 50 mM NaCl into 150 KDa MWCO spin concentrators.

The 5 ml final volume was loaded onto a Superdex S200 16/600 column for size exclusion chromatography. Protein containing fractions were pooled and concentrated as before, flash-frozen in liquid nitrogen in 100  $\mu$ l aliquots and stored at -80 °C. 0.8 ml at 19.00 mg/ml of *CtUGGT*<sup>S180C/T742C</sup> were obtained.

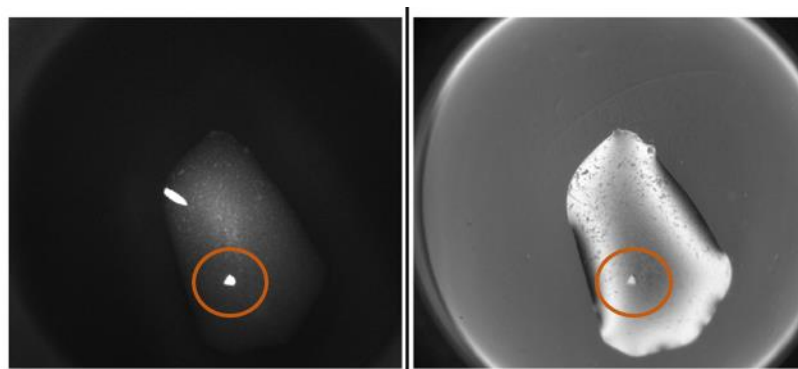
### 2.3.3. Crystallisation

#### 2.3.3.1. *CtUGGT*<sup>S180C/T742C</sup>

The trigonal P<sub>3</sub><sub>2</sub><sub>1</sub><sub>2</sub> *CtUGGT*<sup>S180C/T742C</sup> crystal grew from protein at OD<sub>280</sub>=7.29 in HEPES 20 mM pH 6.5, 50 mM NaCl, 5 mM UDP-Glc, 1 mM CaCl<sub>2</sub> mixed in protein:mother liquor ratio 100 nl:100 nl with condition 57 of the MORPHEUS II screen<sup>56</sup> (2 mM Lanthanides, 0.1 M Buffer System 6 (1.0M, pH 8.5 at 20 °C, Gly-Gly, AMPD), pH



8.5, 50 % v/v Precipitant Mix 5 (30% w/v PEG 3000, 40% v/v 1, 2, 4- Butanetriol, 2% w/v NDSB 256)). The crystal grew between day 57 and day 71, at 18 °C. The crystal was flash-cooled in liquid N<sub>2</sub>. The orthorhombic P<sub>2</sub><sub>1</sub>2<sub>1</sub>2<sub>1</sub> CtUGGT<sup>S180C/T742C</sup> crystal comes from the same protein sample mixed in protein:mother liquor ratio 133 nl:66 nl with condition 14 of the JCSG+ screen (0.2 M Sodium thiocyanate, 20% w/v PEG 3350). The crystal grew in two days at 18 °C and broke into smaller pieces upon fishing. The crystal was cooled down in liquid N<sub>2</sub> in a cryoprotecting solution made by mixing 2 µl of ethylene glycol in 8 µL of mother liquor (i.e. 20% EG). X-ray diffraction data from the trigonal P<sub>3</sub><sub>2</sub>12 and the orthorhombic P<sub>2</sub><sub>1</sub>2<sub>1</sub>2<sub>1</sub> CtUGGT<sup>S180C/T742C</sup> crystals were collected on beamlines I24 and I03 @Diamond, respectively.



*Figure 13 The trigonal P<sub>3</sub><sub>2</sub>12 CtUGGT<sup>S180C/T742C</sup> crystal grew in 57-71 days at 18°C, in condition 57 of MORPHEUS II screen in ratio 100 nl protein:100 nl ML. Left hand side: UV fluorescence. Right hand side: normal lighting.*

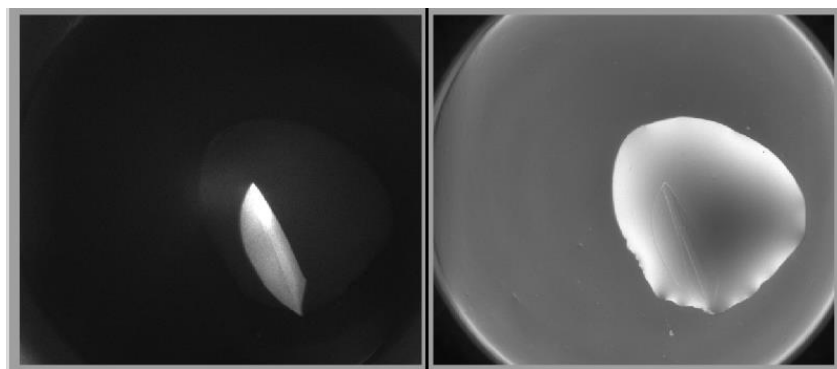


Figure 14 The orthorhombic  $P2_12_12_1$  CtUGGT<sup>S180C/T742C</sup> crystal grew in two days at 18°C, in condition 14 of JCSG+ screen in ratio 133 nl protein:67 nl ML. Left hand side: UV fluorescence. Right hand side: normal lighting.

### 2.3.3.2. CtUGGT<sup>G177C/A786C</sup>

Data was collected on I04@DLS, with a PILATUS 6M detector unit, at wavelength  $\lambda=0.97949$  Å, beam size 19×10 μm, 0.10° oscillation, 1800 images, 0.10 s/image and T=100%; plate set at 4.5 Å max resolution.

Table 3 X-ray diffraction Data collection statistics. Values in parentheses refer to the highest resolution shell.

CtUGGT Structure	G177C/A786C	S180C/T742C	
Beamline, date	I04@DLS, 08.10.2018	I24@DLS, 08.08.2018	I03@DLS, 05.05.2018
Space group (Z)	P <sub>4</sub> <sub>3</sub> 2 <sub>1</sub> 2 (8)	P <sub>3</sub> 2 <sub>1</sub> 2 (6)	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> (4)
Wavelength (Å)	0.97949	0.96861	0.97622
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	<i>a</i> = <i>b</i> =138.602	<i>a</i> = <i>b</i> =148.91	<i>a</i> =76.90,
	<i>c</i> =175.75	<i>c</i> =235.58	<i>b</i> =139.34,
$\alpha$ , $\beta$ , $\gamma$ (°)			<i>c</i> =184.03

	$\alpha=90.0,$ $\beta=90.0, \gamma=90.0$	$\alpha=90.0,$ $\beta=90.0, \gamma=120.0$	$\alpha=90.0,$ $\beta=90.0, \gamma=90.0$
<b>Resolution range(Å)</b>	108.8-6.1 (6.2-6.1)	129.0-5.5 (5.6-5.5)	92.0-4.4 (4.5-4.4)
<b><math>R_{\text{merge}}</math></b>	0.260 (1.433)	0.115 (2.303)	0.268 (1.198)
<b><math>R_{\text{meas}}</math></b>	0.283 (1.669)	0.129 (2.573)	0.317 (1.639)
<b><math>CC_{1/2}</math></b>	0.989 (0.467)	0.998 (0.571)	0.978 (0.415)
<b><math>I / \sigma(I)</math></b>	5.9 (0.8)	10.5 (1.0)	2.3 (0.9)
<b>Completeness (%)</b>	98.6 (100.0)	99.9 (100.0)	83.6 (73.2)
<b>Redundancy</b>	6.4 (3.6)	9.4 (9.8)	3.2 (2.1)

#### 2.3.4. Urea-denatured thyroglobulin (UDT)

##### preparation

Bovine thyroglobulin (Sigma-Aldrich) was denatured with urea following the protocol by Trombetta *et al.*<sup>57</sup>, 20mg of thyroglobulin (TG) powder was dissolved in 1 ml of ‘Column Buffer’ using a vortex mixer and spun at 12000xg for 1 min to ensure that mixing occurred. The thyroglobulin was then injected onto a Superose™ 6 10/300 Increase gel filtration column (GE Life Sciences) which had been equilibrated with the ‘Column Buffer’. SDS-PAGE analysis of the elution profile was used to select fractions containing pure protein, which were pooled and spin-concentrated to 2 ml. The thyroglobulin was then dialysed against 200 ml of ‘Denaturation Buffer’ at room

temperature for 36 hours (with fresh buffer changes at 12, 70 and 24 h) to achieve a dialysis exchange factor of 1,000,000x. After 36 h, the thyroglobulin was dialysed against 200 ml of ‘Renaturation Buffer’ for 96 h at 4 °C (with fresh buffer changes at 48, 60 and 84 h). The urea-treated thyroglobulin was diluted to 1 ml with ‘Column Buffer’, spin-filtered and run down a Superose™ 6 10/300 gel filtration column equilibrated with ‘Column Buffer’. After SDS-PAGE analysis, fractions containing pure protein were pooled and concentrated. The pooled fractions gave the UDT used in the activity assays.

### 2.3.5. Mass spectroscopy detection of disulphide bonds ♦

In peptide mass spectrometry (performed as described in 2.4.6) fragment ions that appear to extend from the amino- or carboxy-terminus of a peptide are termed “b” or “y” ions, respectively.

In Figure 15 (A, B), mass spectrometry detection of ions derived from fragmentation of the disulphide-bridged tryptic peptides <sup>766</sup>FLDLETALETGEFEPDVAYDCSLANFLASSNMK<sup>798</sup> and <sup>176</sup>FCVGSR<sup>181</sup> in the double mutant *CtUGGT*<sup>G177C/A786C</sup>. The ions confirm the establishment of the engineered disulphide bridge at

---

♦ Performed by Dr. Abhinav Kumar at Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, England, United Kingdom.

positions 177-786 between the TRXL1 and TRXL3 domains. No peptides containing free Cys at either position 177 or 786 were detected.

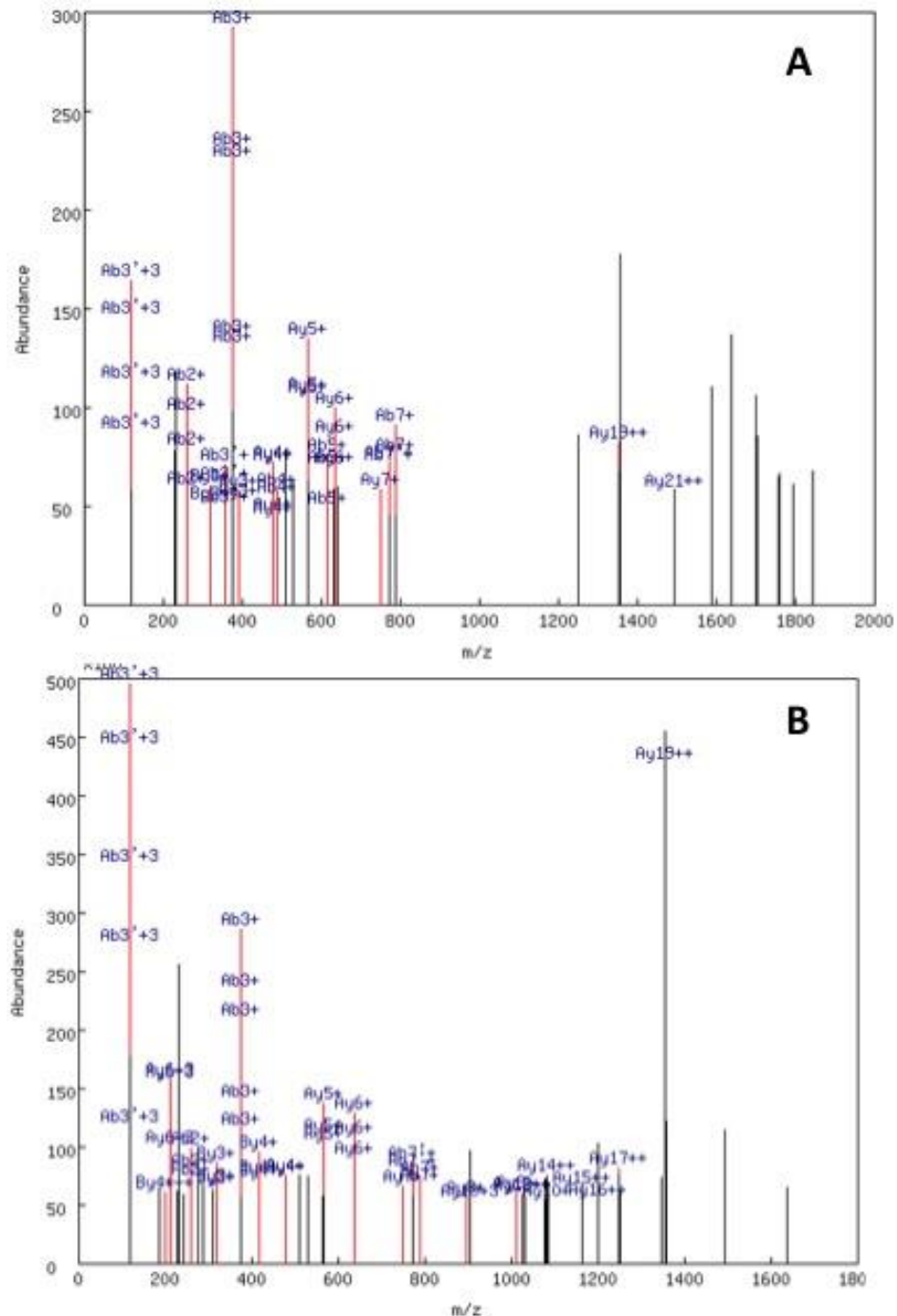


Figure 15 Mass spectrometry detection of ions derived from fragmentation of the disulphide-bridged tryptic peptides  $^{766}\text{FLDLETALETGEFEPDVA YDCSLANFLASSNMK}^{798}$  and  $^{176}\text{FCVGS R}^{181}$  in the double mutant  $\text{CtUGGT}^{\text{G177C/A786C}}$ .

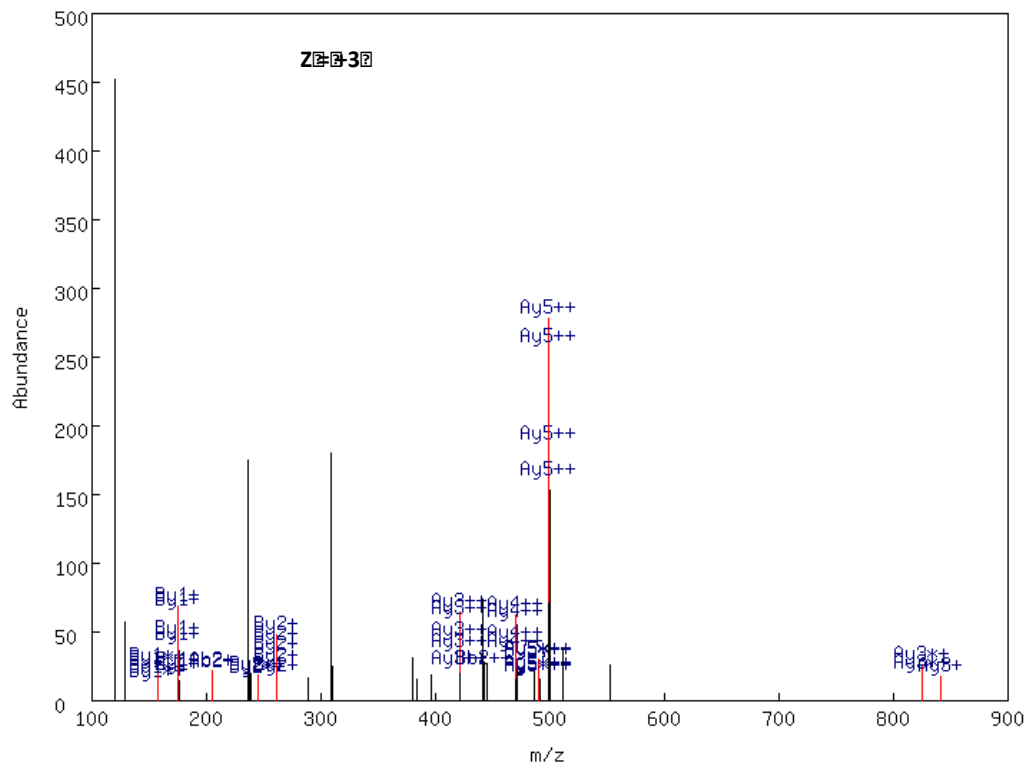


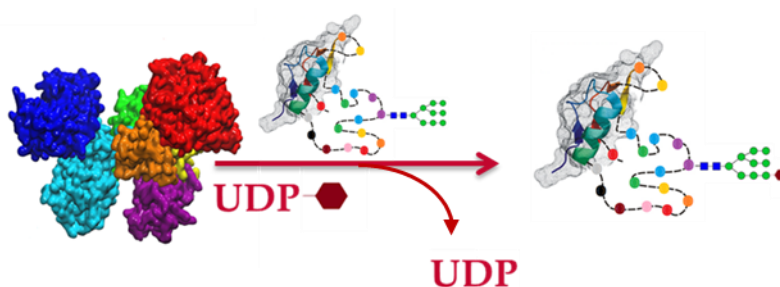
Figure 16 Mass spectrometry detection of ions derived from fragmentation of the disulphide-bridged tryptic peptides  $^{741}\text{DCSR}^{744}$  and  $^{176}\text{FGVCSR}^{181}$  in the double mutant  $\text{CtUGGT}^{\text{G179C/T742C}}$ .

In Figure 16, mass spectrometry detection of ions derived from fragmentation of the disulphide-bridged tryptic peptides  $^{741}\text{DCSR}^{744}$  and  $^{176}\text{FGVCSR}^{181}$  in the double mutant  $\text{CtUGGT}^{\text{G179C/T742C}}$ . The ions confirm the establishment of the engineered disulphide bridge at positions 179-742 between the TRXL1 and TRXL3 domains. No peptides containing free Cys at either position 179 or 742 were detected.

In Figure 17, mass spectrometry detection of ions derived from fragmentation of the disulphide-bridged tryptic peptides  $^{741}\text{DCSR}^{744}$



protein glycans. The product results of UGGT reaction are two: the monoglucosylated glycan on the misfolded protein and the UDP, as depicted in Figure 18. Both reaction products can be detected to evaluate and measure the enzyme activity, UDP released during the reaction can be detected using the commercial kit “UDP-Glo™ Glycosyltransferase Assay” from Promega, whereas the monoglucosylated glycans can be labelled and detected as already reported<sup>58</sup>. The UDP detection with the commercial kit “UDP-Glo™ Glycosyltransferase Assay” was carried out with several false positives and false negatives, so results are not here reported nor commented.



*Figure 18 UGGT re-glucosylation of misfolded protein representation.*

#### 2.3.6.1. HPLC Re-glucosylation assay

Urea-denatured thyroglobulin (UDT) was used as a glucosyl acceptor in the activity assay and was thus used to assess the enzymatic activity of the purified *Ct*UGGT mutants, as previously described<sup>45</sup>. UGGT re-glucosylates substrates using UDP-Glc, and therefore the experiment was set as follow: 100 µg of UDT, 86 µM UDP-Glucose,



8.6 mM CaCl<sub>2</sub>, 8.6 mM Tris-HCl pH 8.0 and 45 pmol of C $\alpha$ UGGT enzyme. The reaction mixtures were set up at 37 °C. Each reaction was 70  $\mu$ l to start with, in triplicate. 10  $\mu$ l aliquots were taken at each time point (5', 15', 30', 1 h, 2h), and the glucosylation quenched by addition to each 10  $\mu$ L aliquot of 1  $\mu$ l of PNGaseF denaturing buffer, then heating for 10 min at 90 °C. Then 5  $\mu$ l of 10X PNGase glycobuffer 2 (NEB), 5  $\mu$ l of NP40 10%, 1  $\mu$ l of PNGase F at 1 mg/ml and 27  $\mu$ l of water were added to each sample for the overnight digestion with PNGase F. The N-linked glycan were labelled with anthranilic acid (2-AA) (Sigma-Aldrich), purified by adsorption to Speed-amide SPE columns and detected by normal-phase high-performance liquid chromatography, as detailed described by Caputo *et al*<sup>58</sup>. The amount of glucosylation was measured in comparison to control by measuring the peak area of the PNGase F released 2-AA-labelled species Man<sub>9</sub>GlcNAc<sub>2</sub> and Glc<sub>1</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> using Waters Empower software. This allows the % of glucosylation to be determined as the % of Glc<sub>1</sub>-species (Peak Area Glc<sub>1</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>) as a total of potential glucosylation species (Peak Area of Glc<sub>1</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> + Man<sub>9</sub>GlcNAc<sub>2</sub>).

## 2.4. Materials and methods

All chemicals used were obtained from Sigma-Aldrich and all enzymes were obtained from New England BioLabs. All solutions were made using ultrapure Milli-Q® water, Millipore. For DNA or RNA solution were used nuclease free water, not diethylpyrocarbonate (DEPC) treated (Ambion). All 0.22 µm filtering of solutions was performed with Millipore Express PLUS 0.22 µm PES membranes (Millipore).

### 2.4.1. Cloning

#### 2.4.1.1. Primer design

The primers were designed using the NEBaseChanger® tool from New England BioLabs website (see Appendix).

#### 2.4.1.2. Vectors

Cloning vector Litmus28i was used for cloning and mammalian expression vector pHL-sec<sup>55</sup> was used for transfection of HEK293F cells.

#### 2.4.1.3. Base-directed mutagenesis

*Ct*UGGT construct inserted in Litmus28i was used for base-directed mutation. Cloning was performed with the Q5® Site Directed Mutagenesis Kit from New England BioLabs according to the

protocol. Amplification and insertion of mutation was performed by PCR using 12.5 µl of Q5® Hot Start High-Fidelity 2X Master Mix (NEB), added to 1.25 µl of each forward and reverse primers at 10 µM concentration, 1 µl of template DNA (1 ng/µl) and 9 µl of nuclease-free water, obtaining a 25 µl final volume. Thermocycling protocol as the manufacturer's suggestion, varying the annealing temperature and the extension time to comply with the amplification primers and length of the template DNA. Circularisation Kinase, Ligase & DpnI (KLD) treatment was made to delete the parental fragments present in the mixture, as follows: 1 µl of PCR product was mixed with 5 µl of 2X KLD Reaction Buffer, 1 µl of 10X KLD Enzyme Mix and 3 µl of nuclease-free water. The mixture was incubated at room temperature for 5 minutes. The KLD reaction mixture was used to transform *E. coli* DH-5α chemically competent cells.

#### **2.4.1.4. Gibson Assembly**

Gibson Assembly was performed to insert the gene into pHL-sec vector using the Gibson Assembly® Master Mix NEBuilder® HiFi DNA Assembly (New England BioLabs) according to NEB's protocol, using linearised vector (digested overnight with AgeI and KpnI restriction enzymes at 37 °C) and the amplified gene purified from DNA gel. PCR was carried out using properly designed primers with some base pairs that overlap with pHL-sec vector to both 5' ends of the gene sequence.

#### **2.4.1.5. Transformation**

Transformation of New England BioLabs DH-5 $\alpha$  competent *E. coli* cells were carried out following manufacturer's protocol. Briefly, 5  $\mu$ l of KLD reaction mix or Gibson Assembly product were added to a tube of thawed NEB DH-5 $\alpha$  competent *E. coli* cells on ice, mixed gently for a few seconds, incubated on ice for 30 minutes, heat-shocked at 42 °C for 30 seconds and incubated on ice again for 5 minutes. 950  $\mu$ l of SOC media was added for a final volume of 1 ml and the mixture was incubated for 1 hour at 37 °C with gentle shaking at 200/300 rpm. An aliquot of the mixture was spread onto a pre-warmed (37 °C) LB agar culture plate containing carbenicillin (0.1 mg/ml).

#### **2.4.1.6. Colony PCR**

The protocol used to perform a colony PCR is the following: 10  $\mu$ l of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs), 1  $\mu$ l of each forward and reverse primer (scheduled in Appendix), 8  $\mu$ l of nuclease-free water, few cells picked from one isolated colony from the agar plate. PCR: step 1: 94 °C for 5 minutes; step 2: 94 °C for 1 minute; step 3: 68 °C for X minutes (X = about 1-minute per Kb of DNA); step 4: 72 °C for 1 minute. Steps 2-4 repeated 34 times; step 5: 72 °C for 1 minute.

### **2.4.2. Expression in HEK293F cells**

Human epithelial kidney FreeStyle™ 293-F (HEK293-F) cells (ThermoFisher Scientific) were grown in suspension cultures and passaged in GIBCO® FreeStyle™ Expression Media (ThermoFisher Scientific) according to the protocols, maintaining the density within the limits of  $0.1\text{--}2.0 \times 10^6$  cells/ml. Cells were transfected using the FreeStyle™ 293 Expression System (ThermoFisher Scientific) and protocol as follow: each 400 ml of culture was added of 20 ml transfection mixture. Transfection mixture was freshly prepared before transfection and consisted of 400 ng of DNA insert in expression vector dissolved in 10 ml of Gibco™ OptiPRO™ SFM (serum-free medium) and 400 µl of FreeStyle™ MAX Reagent diluted in 10 ml of Gibco™ OptiPRO™ SFM, gently mixed and incubated at room temperature for 10 minutes. Cells were cultured in sterile Erlenmeyer flasks with 0.2 µm vent caps (Corning), filling not more than 40% of the flask volume, in a shaking incubator at 150 (rpm), 37 °C kept at 8% CO<sub>2</sub> for 4 days, monitoring the protein expression by harvesting an aliquot every day and analysing the supernatants on an SDS-PAGE gel in order to appoint when maximal protein expression resulted.

### **2.4.3. Protein purification**

#### **2.4.3.1. Sample preparation**

Once the expression of the desired protein was at maximum peak, the culture was harvested using a 1500xg spin for 10 minutes. 10x PBS was added to the supernatant so that a final concentration of 1x PBS was achieved, the pH was adjusted to 7.6 with NaOH and it was filtered down with a 0.22 µm filter.

#### **2.4.3.2. Nickel affinity chromatography (NAC)**

Nickel affinity chromatography (NAC) was carried out with a 5 ml nickel column HisTrap™ Excel (GE Healthcare Life Sciences) stored in 20% ethanol. Eluted with two different buffers, all 0.22 µm filter-sterilised and degassed. Binding Buffer: 1x Phosphate buffered saline (PBS), pH 7.6. Elution Buffer: 1xPBS, 500 nM imidazole, pH 7.6. HisTrap™ Excel (GE Healthcare Life Sciences). HisTrap™ Excel (GE Healthcare Life Sciences) nickel column was washed out of ethanol with 5 column volumes (CV) of MilliQ water, then washed with 5 CV of each binding buffer, elution buffer and binding buffer again. The column was equilibrated with 15 CV of binding buffer, the supernatant was applied ÄKTA Start chromatography system in a 4 °C cabinet at a flow rate of 1 ml/minute. Once the sample was loaded the column was washed with 20 CV of binding buffer and eluted at

the end using a gradient elution of 0% to 100% elution buffer over 20 CV. Fractions of all steps were collected.

#### **2.4.3.3. Size exclusion chromatography (SEC)**

All size exclusion chromatography (SEC) columns were obtained from GE Life Sciences. All columns were stored in 20% ethanol and run using an ÄKTA pure FPLC chromatography system at room temperature. For all the constructed purification, a Superdex 200 16/600 was used. Before equilibrating into running buffer, columns were equilibrated with 1.5CV MilliQ. Buffer used for SEC was degassed and 0.22 µm filter sterilised.

For all UGGT constructs, running buffer composed of 20 mM HEPES, 50 mM NaCl, pH 6.5 was used. 1.2CV of buffer was used for each run at a flow rate of 1 ml/min and fractions were collected after 0.2 CV.

For Thyroglobulin purification a Superose™ 6 10/300 Increase gel filtration column was used. ‘Column Buffer’: 20 mM HEPES, pH 7.4, 150 mM NaCl. ‘Denaturation Buffer’: 20 mM HEPES pH 8.0, 150 mM NaCl, 8 M Urea. ‘Renaturation Buffer’: 20 mM HEPES pH 8.0, NaCl 150 mM.

#### **2.4.3.4. Anion exchange**

An HiPrep Q® HP 16/60 column (GE health care) was used for anion exchange using an ÄKTA pure FPLC chromatography system at room

temperature. Buffer was degassed and 0.22 µm filtered before use. The column was stored in 20% ethanol.

Binding Buffer:  $\text{K}_2\text{HPO}_4 + \text{KH}_2\text{PO}_4$  20 mM, pH 8.5

Elution Buffer:  $\text{K}_2\text{HPO}_4 + \text{KH}_2\text{PO}_4$  20 mM, pH 8.5, 1 M NaCl

The column was equilibrated with MilliQ water before and binding buffer. Sample was diluted to 1 litre with  $\text{H}_2\text{O}$  and titrated till pH 8.5 with NaOH, then loaded onto the column. Elution was performed in three steps: 3.5 CV of 25% B, 3.5 CV of 50% B, 3.5 CV of 100% B, collecting 15 ml fractions.

#### **2.4.3.5. Protein concentration**

Before SEC purification or long-term storage of protein, it was necessary to concentrate the protein. For proteins derived from *Chaetomium thermophilum*, Sartorius Vivaspin 20, PES membrane centrifugal filter units (Generon) were used for concentration with an appropriate molecular weight cut-off (MWCO) for the sample, generally 100 000 or 50 000 Daltons.

### **2.4.4. Gel electrophoresis**

#### **2.4.4.1. SDS-PAGE**

SDS-PAGE was carried out using the Thermo Fisher Scientific Novex® systems following the protocols. For SDS-PAGE, pre-cast 1



mm NuPAGE™ 4-12% Bis-Tris Protein gels were used with 10, 12 or 17 wells. Gels were run in NuPAGE® MOPS buffer for 50 minutes at 200V. Gels were stained using InstantBlue protein stain (Expedeon). Samples were prepared by adding 4x LDS sample buffer (Thermo Fisher Scientific), plus 50 mM DTT if required, heated to 80°C for 1 minute, and then loaded onto the gel.

#### **2.4.4.2. DNA gels**

DNA agarose gels were prepared using agarose (a4-0700, geneflow) in TAE buffer (1% agarose w/v). GelRed nucleic acid gel stain (Biotium) was added to molten agar, before pouring into a mould, to allow for visualisation of DNA bands under UV excitation. Once set, gels were run at 150V for 50 minutes in TAE buffer. Samples were prepared by the addition of 6x purple loading dye (New England BioLabs).

### **2.4.5. Crystallisation**

#### **2.4.5.1. Crystallisation screens**

All crystallisation plates were set with crystallisation screens and protein using the crystallisation facility at the Department of Biochemistry. Commercial crystallisation screens were obtained from Hampton Research and Molecular Dimensions. Both commercial sparse screens and custom optimisation screens were dispensed into

MRC 2-Well crystallization plates (Swissci) UVXPO, 50 µl into each reservoir. Commercial screen used were: MORPHEUS®<sup>56</sup>, JCSG-*plus*<sup>TM</sup> <sup>59</sup>, STURA<sup>TM</sup> <sup>60</sup>, MACROSOL<sup>TM</sup> <sup>61</sup>, ProPlex<sup>TM</sup>.

Sitting drops with protein were set up with a Mosquito® robot (TTP labtech) in two well-drops plates as follow: one drop with 100 nl protein and 100 nl mother liquor and one drop with 133 nl protein and 67 ml mother liquor. All crystallisation plates were stored at 18 °C in a Rigaku Minstrel<sup>TM</sup> HT crystallisation hotel and imager at 18 °C, with images of wells being taken at both visible and ultraviolet wavelengths at regular intervals over the course of 3 months.

#### **2.4.5.2. X-ray Data Collection**

Crystals were picked using 50-100 µm nylon loops mounted on pins appropriate for the synchrotron to be used. If the crystallisation condition did not contain an appropriate cryoprotectant, loops containing crystals were passed through a droplet containing cryoprotectant. Crystals were cryocooled in liquid nitrogen prior to shipping to the synchrotron. All data collection took place under a gaseous nitrogen stream at 100 K. Diamond Light Source (DLS) at Harwell, UK, beamlines I03, I04, I04-1 and I24 were used to collect data from crystals, setting proper parameters for each crystal.

#### 2.4.5.3. X-ray Data Processing

Diffraction data were indexed, integrated, scaled, merged and an anisotropic cut-off imposed using the autoPROC toolbox (Global Phasing)<sup>62–64</sup>. Molecular replacement was completed using the programme Phaser<sup>65</sup>. Structure manipulation was executed using Coot<sup>66</sup> hence refinement was carried out using BUSTER<sup>67,68</sup>. Figures were made using PyMOL<sup>69</sup>.

#### 2.4.6. Mass spectroscopy detection of disulphide bonds

Tryptic peptides of the double mutants *Ct*UGGT<sup>G177C/A786C</sup>, *Ct*UGGT<sup>G179C/T742C</sup> and *Ct*UGGT<sup>S180C/T742C</sup> were separately run on an EASY-nLC43 1000 UHPLC system (Proxeon) and electrosprayed directly into a Q Exactive mass spectrometer (ThermoFischer Scientific) through an EASY-Spray nano-electrospray ion source (ThermoFischer Scientific). Peptides were trapped on an in-house packed guard column (75 µm i.d. x 20 mM, reposit C18, 3µm, 120 Å) using solvent A (0.1% Formic Acid in water) at a pressure of 500 bar and then separated using an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. x 500 mm, RSLC C18, 2 µm, 100 Å) eluted with a linear gradient (7 % to 31 % solvent B (0.1% formic acid in acetonitrile) in 60 min) at a flow rate of 200 nL/min. Full scan

MS spectra were acquired in the Orbitrap (scan range 350-2000  $m/z$ , resolution 70000, AGC target  $3e6$ , maximum injection time 100 ms). After the MS scans, the 10 most intense peaks were selected for HCD fragmentation at 30 % of normalized collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17500, AGC target  $5e4$ , maximum injection time 120 ms) with first fixed mass at 120  $m/z$ .

**Data analysis (crosslinking).** MS data were converted into .mgf format using MSconvert from the ProteoWizard toolbox and searched using the pLink software. The database contained *CtUGGT* only. Search parameters were as follows: maximum number of missed cleavages = 2, fixed modification = none, variable modification 1=Oxidation-Met, variable modification 2=carbamidomethyl-Cys, disulphide bonds were considered as the crosslink (Cys-Cys, -2.015650 Da), mass accuracy filter = 20 ppm for precursor ions with consideration of the first 5 isotopic peaks, MS2 tolerance = 20 ppm. Data were filtered by E-value ( $E < 1.0 \times 10^{-7}$ ) and a manual inspection was performed.

**Data analysis (protein identification).** The raw data files generated were processed using MaxQuant software (Version 1.5.0.35), integrated with Andromeda search engine as previously described<sup>70</sup>. Raw MS data were searched against a database containing proteins of interest as well as list of common contaminants by Andromeda. Trypsin with a maximum number of 2 missed cleavages was selected

as the protease. Acetylation (Protein N-term), Carbamidomethylation (Cysteine) and Oxidation (Methionine) were used as variable modifications. Protein and PSM false discovery rate (FDR) were set at 0.01.

## 2.5. Results and discussion

In order to probe the functional role of the 'clamping' motion observed in the Molecular Dynamics simulations, we engineered three novel double cysteine *CtUGGT* mutants: *CtUGGT*<sup>G177C/A786C</sup>, *CtUGGT*<sup>G179C/T742C</sup> and *CtUGGT*<sup>S180C/T742C</sup>, designed to form disulfide bridges across the interfaces of the TRXL1-TRXL3 domains. All three mutants were expressed and purified from the supernatant of mammalian HEK293F cells. The presence of the engineered disulfide bridges was confirmed by mass spectrometry (Figures 15, 16 and 17). The crystal structures of *CtUGGT*<sup>G177C/A786C</sup> and *CtUGGT*<sup>S180C/T742C</sup> were determined respectively to 6.1 Å and 4.4 Å resolution. Both crystal structures show the TRXL3 domain tethered to the TRXL1 domain (Figure 19A). Figure 7B shows the activity of the three double Cys mutants (in addition to the activity of the WT, as control, and the already published *CtUGGT*<sup>D611C/G1050C</sup>). It appears that, despite their structural similarity, the *CtUGGT*<sup>S180C/T742C</sup> and *CtUGGT*<sup>G177C/A786C</sup> mutants differ significantly in their ability to re-glucosylate urea-treated bovine thyroglobulin, with the former more active than, and the latter having similar activity to, WT *CtUGGT*.

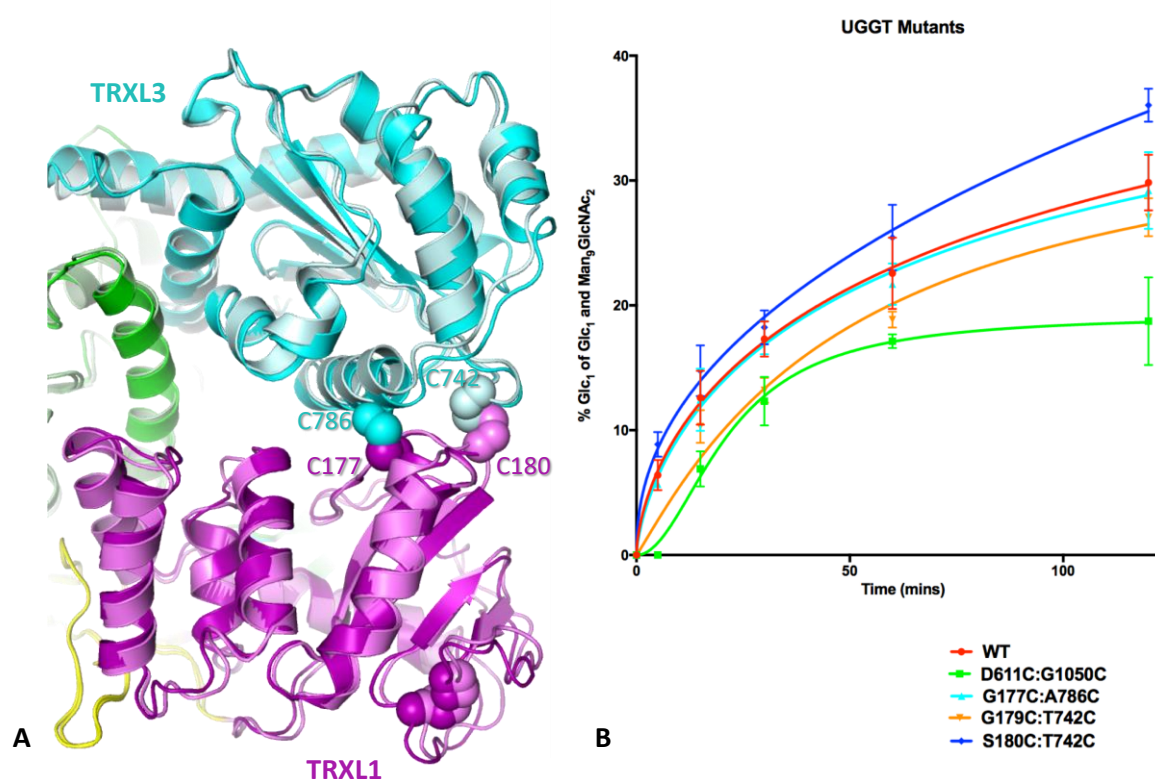


Figure 19 (A) The TRXL1 (magenta) and TRXL3 (cyan) domains in the crystal structures of *CtUGGT*<sup>G177C/A786C</sup> (dark colours) and *CtUGGT*<sup>S180C/T742C</sup> (lighter colours). The disulphide bonds are in spheres representation. (B) Activity of *CtUGGT* double Cys mutants and WT *CtUGGT* against urea-denatured bovine thyroglobulin.

In order to probe the contributions of individual UGGT TRXL domain to UGGT re-glucosylating activity, we cloned three mutants of *CtUGGT* that lack domains TRXL1, TRXL2 and TRXL3: *CtUGGT*- $\Delta$ TRXL1, lacking residues 42-224; *CtUGGT*- $\Delta$ TRXL2, lacking residues 417-650; and *CtUGGT*- $\Delta$ TRXL3, lacking residues 666-870. Only the latter two mutants expressed and were purified, and *CtUGGT*- $\Delta$ TRXL2 yielded crystals that enabled structure determination to 5.5 Å resolution. The *CtUGGT*- $\Delta$ TRXL2 crystal

structure resembles the WT structure apart from a minor rearrangement of the TRXL3 domain (Figure 19 A). UGGT-mediated re-glucosylation activity assays of *CtUGGT-ΔTRXL2* and *CtUGGT-ΔTRXL3* against urea-misfolded bovine thyroglobulin detect impaired re-glucosylation activity upon deletion of TRXL3 and complete loss of activity upon deletion of TRXL2 (Figure 20 B).

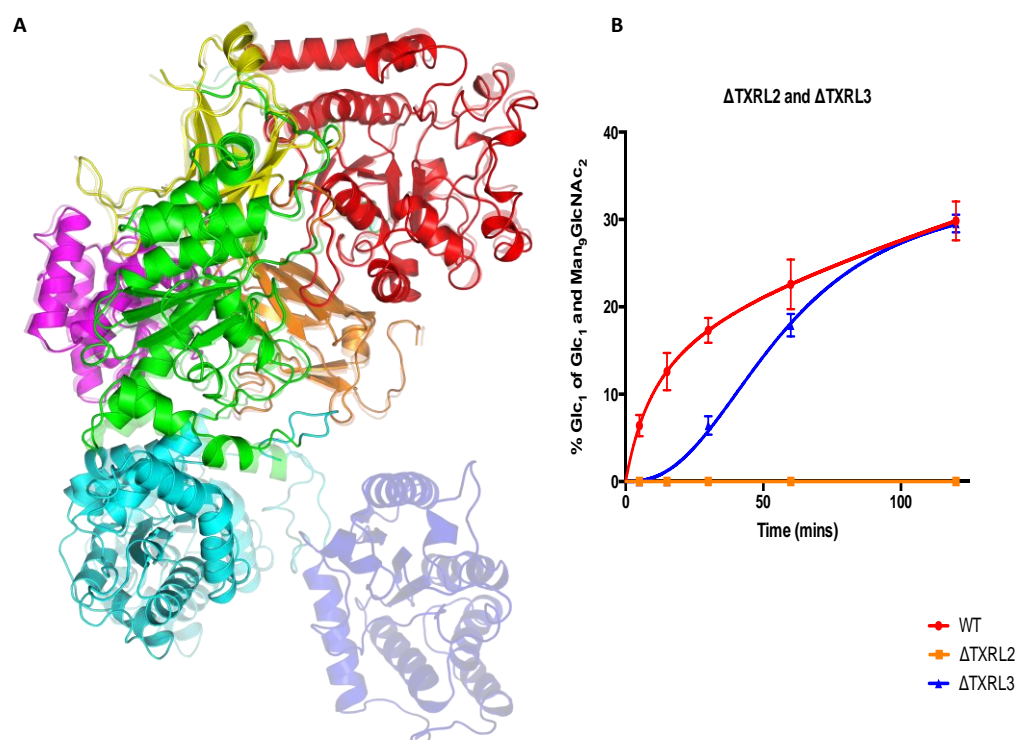


Figure 20 **A)** Crystal structure of *CtUGGT-ΔTRXL2* (copy "A", solid colours) overlaid onto WT *CtUGGT* ('open' conformation (PDB ID: 5MZO), semi-transparent). Domain coloured as in Figure 1A. **(B)** Activity of *CtUGGT-ΔTRXL2* and *CtUGGT-ΔTRXL3* against urea-denatured bovine thyroglobulin, compared to WT *CtUGGT*.

## 2.6. Fragment Based Lead Discovery

### 2.6.1. Introduction

Fragment-based Lead Discovery (FBLD) is a widely applied method for the identification of small, low complexity chemical fragments (typically approximately 200 Da), that bind to the biological molecule of interest (target) with low-affinity. It is an important tool in the hit-to-lead drug development process; the hits obtained from a FBLD screen are usually optimised through structure-based drug design (SBDD), ideally yielding higher affinity compounds that can be used to study the mode of binding or as a starting point (lead) for drug discovery<sup>71</sup>. The latter approach identifies functional moieties that match the structure and chemical properties required to bind the target. Usually, fragments bind to the most targetable binding sites on the surface of the “target”, so fragment binding is typically of low affinity (millimolar or sub-millimolar  $K_D$ ). Due to this low affinity between target and the fragments, sensitive detection methods are required to detect binding of such fragments<sup>72</sup>. A common approach to library screening has been the use of biophysical methods such as X-ray crystallography<sup>73,74</sup> or NMR (protein-observed NMR and ligand-observed NMR)<sup>75</sup>, which are highly sensitive techniques, but a big amount of protein is needed. Surface plasmon resonance (SPR)<sup>76,77</sup>, microscale thermophoresis (MST)<sup>78</sup>, high-concentration



screening using HTS assays<sup>79</sup>, thermal shift analysis (TSA)<sup>80</sup>, mass spectrometry (MS)<sup>81,82</sup> and biochemical assays<sup>83</sup> are also widely used techniques to study fragment binding. Although these techniques are less sensitive than NMR or X-ray crystallography, lesser amounts of protein are required.

On the computational front, a fragment library screening is challenging due to the few interactions between each fragment and the target, and the binding affinity is much smaller than the one of larger, more specific ligands or drugs. A novel approach exploits the idea that the energetics of binding are dominated by the activation energy for dissociation (off-rate): Dynamic Undocking (DUck)<sup>84</sup> is used to screen *in silico* libraries of fragments on known protein targets. *In silico* predictive methods provide tools to explore flexibility of the target and identify functional groups that are able to bind the protein surface or that could be used to grow and optimize a fragment.

The effectiveness of FBLD depends on the “quality” of each step of the process. An appropriate library must be selected. A desirable fragment library covers a wide range of molecular weights, shapes and pharmacophore moieties. The screening technique selected, and its sensitiveness highly influence the results of the experiment<sup>71</sup>.

**Fragment to lead optimization.** As one or more fragments are identified, the following strategies can be used to move from fragment to lead optimization: *fragment growing*, *fragment merging*, and *fragment linking*.

*Fragment growing* entails fragment optimization by structure-based design introducing moieties or functional groups to attain the desired selectivity and drug-like properties. *Fragment merging* is an approach that combines information from multiple chemical hits together. *Fragment linking* consist in the design of a lead by linking various hits that sit in the same extended binding site but interact with neighbouring regions of it<sup>71</sup>.

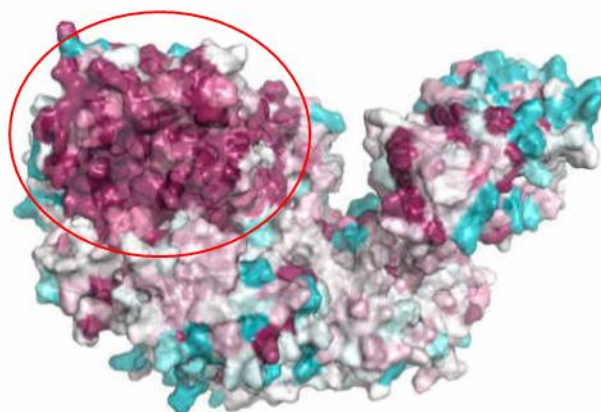
### 2.6.2. Project aim

As carbohydrate mimics, iminosugar inhibitors of the ER alpha glucosidases have some undesired off-target effects, as they also inhibit certain other carbohydrate processing enzymes within the human host. A new class of molecules inhibiting host glycoprotein folding enzymes that viruses depend upon would have great potential for antiviral drug discovery. To complement the medicinal chemistry programme ongoing in the laboratory of Prof. Zitzmann and centred around ER alpha-glucosidase inhibitors with fewer off-target effects,

we aim here to investigate the potential of UGGT1, the major calnexin cycle misfold sensor, as an antiviral target.

UGGT1 would constitute a novel antiviral target, but no inhibitors are known other than the product UDP-glucose. Deletion of this enzyme leads to antiviral effects during Zika virus infection in cellula, demonstrating that UGGT1 is a viable target for antiviral drug discovery. The recent availability of recombinant full length eukaryotic *HsUGGT1* provided us with a unique opportunity to conduct fragment-based lead discovery (FBLD) screens of chemical compound libraries against the enzyme, hoping to develop UGGT inhibitors as antiviral drugs, aimed at a whole range of viruses that rely on the calnexin cycle (and hence on UGGT1) for correct folding, including dengue, Ebola, HIV, HCV, HBV, and Zika virus. These ligand compounds would constitute leads in the development of UGGT1-specific inhibitors.

Roversi *et al.*<sup>45</sup> studied UGGT1 sequence homology across higher eukaryotes and mapped UGGT1 sequence conservation on the structure of *CtUGGT*. The sequence conservation mapping was coloured with the scale turquoise – maroon = variable – conserved, and highlight that the catalytic domain is the most conserved domain across eukaryotic UGGT1s (Figure 21).



*Figure 21 Conservation Map of HsUGGT1 structurally aligned to the structure of CtUGGT. Conservation was mapped with the colour scale turquoise – maroon = variable – conserved.*

The Zitzmann group succeeded in the structure determination of *CtUGGT* full length (PDB IDs: 5MU1, 5MZO, 5N2J) and of the *CtUGGT* catalytic domain (PDB ID: 6FSN) at a high resolution (1.19 Å), but no crystals of *HsUGGT1* were obtained. As FBLD requires crystals that are grown reliably and diffract to resolutions equal to or better than 2.0 Å, and taking into account the high degree of conservation between GT24 sequences across eukaryotes, the *CtUGGT* catalytic domain (GT24) was selected to perform a FBLD screening that would yield the first fragments to be tested against the human UGGT1 protein. Within the molecules, in our crystals there is more than one excellent target for UGGT1 inhibitors. In this proposed FBLD screen we aim to identify fragments that bind to the UGGT1 catalytic site and compounds that may interfere with glycoprotein recognition. We chose to identify small molecule

ligands of UGGT1 that bind to its catalytic domain. The potential long-term outcome of this work would be either an allosteric inhibitor or a catalytic site inhibitor of UGGT1.

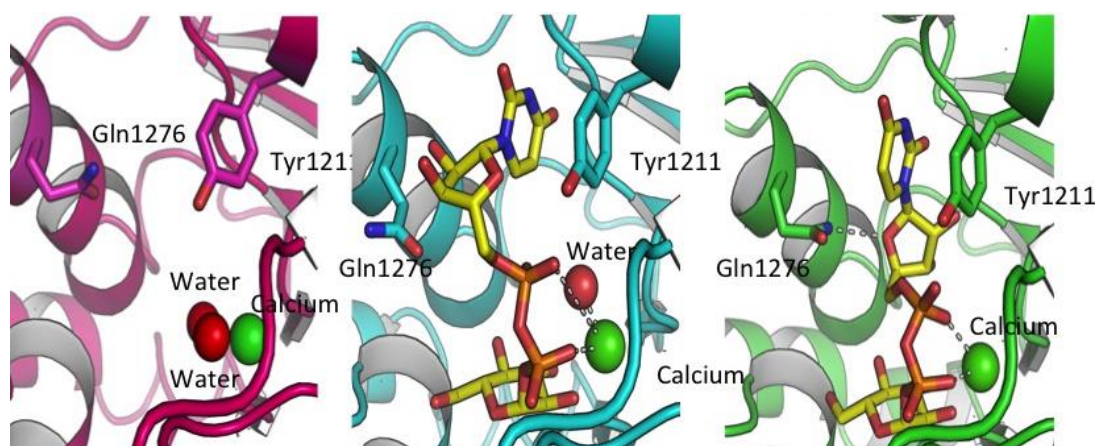


Figure 22 *CtUGGT-GT24* crystal structures. Apo form, magenta. UDP-Glc conformation 1 (PDB ID 6FSN, cyan). UDP-Glc conformation 2 (green).

### 2.6.3. Experimental section

Crystals of the GT24 catalytic domain of *CtUGGT* (residues 1187-1473, plus a C-terminal linker and 6xHis tag of sequence GTKHHHHHH) grow at 18 °C when mixed with a crystallisation solution containing 20.6% v/v PEG 500 MME; 10.3 % w/v PEG 20000, 51.5 mM Bicine 48.5 mM Tris (this solution was optimised from an initial hit from the commercially available Morpheus crystallisation screen<sup>56</sup>. The space group prior to soaking is H3 (the centred hexagonal setting of rhombohedral space group R3), one molecule per asymmetric unit,  $a=b=118.8$  Å  $c=68.7$  Å. Figure Y shows

the hexagonal centred cell in blue and its rhombohedral equivalent in orange. Prior to the structure-based lead discovery effort, fifty of these crystals were soaked in 0, 5, 10, 15, 20, and 30% DMSO and tested for diffraction; no significant deterioration of the diffraction quality was observed.

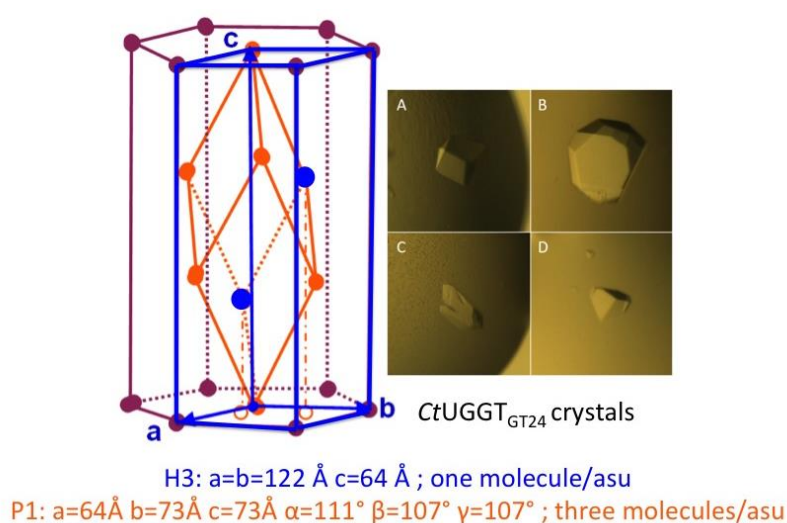
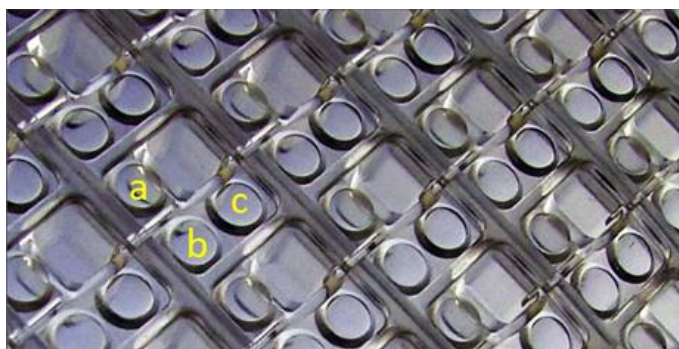


Figure 23 CtUGGT-GT24 crystals. In blue, the H3 cell. In orange, the P1 cell equivalent to the R3 cell.

### 2.6.3.1. Crystal growth

To grow hundreds of crystals of the CtUGGT GT24 catalytic domain, a crystallisation solution was prepared in a 96-wells deep well block, with 500  $\mu\text{l}$  of the same solution in each well: 257.5 $\mu\text{l}$  of Morpheus Precipitant mix 1 (40% v/v PEG 500 MME; 20 % w/v PEG 20000); 25.8 $\mu\text{l}$  of 1M Bicine (Morpheus Buffer system 3 acid component);

24.3  $\mu$ l of 1M Tris (Morpheus Buffer system 3 basic component); 192.4  $\mu$ l of MilliQ Water. The Hydra robot (Art Robbins Instruments) at the Research Complex in Harwell was first used to transfer 25  $\mu$ l of the crystallisation solution from all deep well block 96 wells to all mother liquor reservoirs of six MRC three-lens 96-wells crystallisation plates (see Figure 24). A number of 6x96x3=1728 vapour diffusion sitting drop experiments were set up at the Research Complex in Harwell in MRC three-drop 96-wells crystallisation plates - using a Mosquito robot equipped with an anti-evaporation cover and 60% controlled humidity, in order to avoid drying up of the crystallisation drops during deposition. The CUGGT-GT24 protein at a concentration of 6.5 mg/ml, in presence of 1 mM CaCl<sub>2</sub> and 5 mM UDP-glucose, was mixed in protein:mother liquor proportions 1.35:1 (drops a, c) and 2:1 (drop d) in drops of total volume 200 nl, and the crystals left to grow at 18 °C.



*Figure 24 MRC Three-lens 96-wells sitting drop crystallisation plate.*

### 2.6.3.2. Crystal soaking

In four days, about two-thirds of the sitting drops experiments yielded crystals (see Figure YB). All crystal drops were imaged, and the best crystals were marked for soaking with the Echo robot at the Xchem facility attached to beamline I04-1 at the Diamond Light Source<sup>85</sup>.

The 768 compounds of the DSI-Poised Library (Diamond-SGC-iNext, ex DSPL, each compound 500 mM in deuterated DMSO) (<https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening/Fragment-Libraries/DSI-poised-library.html>) were soaked into as many crystal drops: 40 nl of compound was dispensed by the Echo robot onto each 200 nl drop, making the final compound concentration 100 mM in 20% DMSO.

The soaked crystals were left incubating for a variable time between two and four hours and were fished and cryocooled with the aid of the SGC Shifter Robot.

### 2.6.3.3. X-ray diffraction

Automated data collection was carried out on 692 soaked crystals, with the following conditions: wavelength: 0.91587 Å; 1,000 image frames, 0.06 sec/frame and 0.18° oscillation/frame; 100% transmission, plate set for max resolution ~ 1.8 Å. Automated loop centring failed



about 6% of the time and about 50 crystals were re-measured with manual optical centring.

The symmetry is sometimes lowered by soaking, the crystals can in this case index in space group P1 with three molecules in the asymmetric unit of a pseudo-rhombohedral cell of dimensions  $a=68.5$  Å,  $b=72.2$  Å,  $c=72.9$  Å,  $\alpha=110.4^\circ$ ,  $\beta=108.5^\circ$ ,  $\gamma=108.4^\circ$  (in orange in Figure YA).

#### 2.6.3.4. X-ray Data Processing

Data processing, model refinement and ligand fitting were carried out with the purposely written shell script pipeline COaLLA ("Crystal Oligomorph and Ligand Likelihood Assignment"). In order to decide on the correct symmetry, each dataset was indexed and integrated both in H3 and in P1 using autoPROC<sup>64</sup> suite of programs; for each dataset, refinement of the protein model was carried out in both the H3 and P1 form in autoBUSTER<sup>67</sup>, with the space group giving rise to the lower  $R_{\text{free}}$  being chosen as the correct one for the calculation of the Fo-Fc map. The smile string for each soaked compound was fed together with the refined model and phases in order to attempt docking the ligand using rhofit<sup>86</sup>. The best hits were listed by ranking the rhofit score and/or CC, and the hits inspected in Coot<sup>66</sup>.

## 2.6.4. Results and discussion

Table 4 lists three compounds that were located bound to the *Ct*UGGT-GT24 domain crystal in which they were soaked.

Compound x0763's binding site is not far from the putative binding site for the C-branch of the client glycoprotein glycan *Ct*UGGT Y1339 (*Hs*UGGT1 Y1395). Its binding site is a shallow surface pocket and it is likely difficult to drug, although fairly well conserved between *Ct*UGGT and *Hs*UGGT1.

Table 4 Compounds that were found bound to *Ct*UGGT-GT24 domain.

Compound	SMILE string	Space Group	Resolution (Å)	Binding Site on <i>Ct</i> UGGT (equivalent residue in <i>Hs</i> UGGT1)
x0441	<chem>CCN1N=CC=C1C</chem> <chem>NC</chem>	P1	2.44	Y1350(H1406) H1402(N1458) Q1381(Q1437) M1403(M1459)
x0763	<chem>CC(C)NC(=O)C=1</chem> <chem>C=C(C)N(C)N1</chem>	P1	2.37	T1492(T1498) M1491(M1497) R1333(R1389) R1452(R1508) E1444(E1500)
x0248	<chem>OC=1C=CC(CN2C</chem> <chem>COCC2)=C3C=CC</chem> <chem>=NC13</chem>	P1	2.43	Y1346(Y1401) W1347(W1402) H1402 (N1458)

Compound x0441 is somewhat less interesting because it binds at a crystal contact, and the binding site is not particularly conserved.

Also, the pose is ambiguous at this resolution (2.44 Å). The most interesting aspect about this compound is that this site faces the UGGT saddle and is proximal to the putative binding site of the Asn that carries the glycan, *Ct*UGGT Y1346 (*Hs*UGGT1 Y1402).

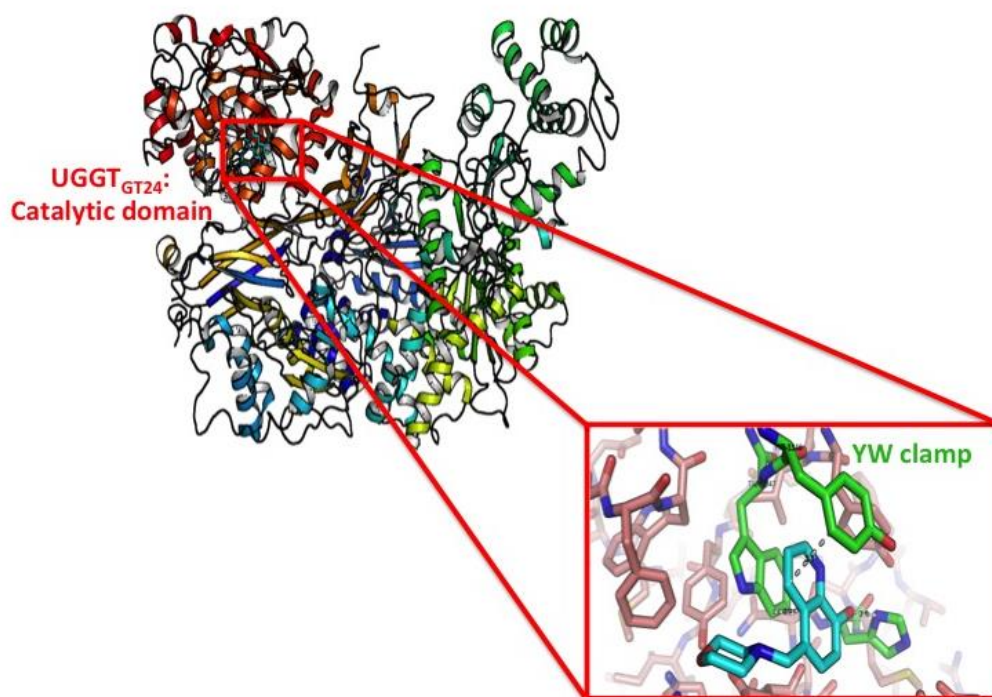


Figure 25 Compound x0248 binding site in the context of the full-length enzyme. The inset shows the YW clamp.

The most promising compound is x0248, an 8-OH-quinoline derivative that also binds to the putative glycan-binding face of the GT24 domain, but in close proximity of the UDP-Glc binding site. The compound's morpholine ring is not ordered in the crystal but the quinoline pose is unambiguously clear in the electron density. Figure 25 shows the x0248 ligand superposed onto the full-length *Ct*UGGT structure.

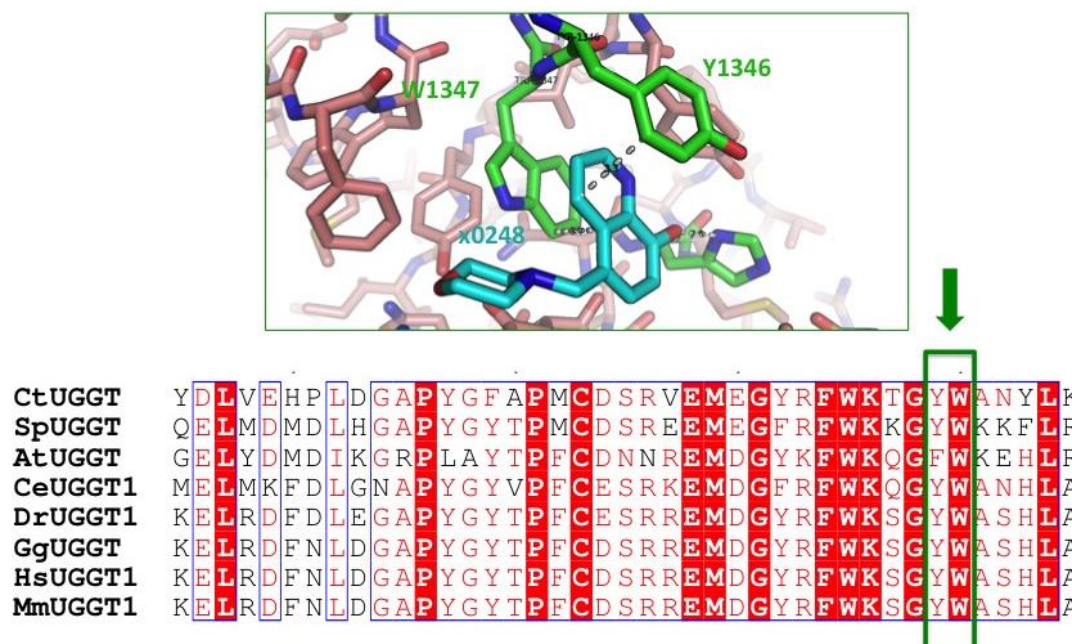


Figure 26 Detail of the YW clamp and sequence conservation of the residues involved in binding x0248.

Figure 26 zooms into the two residues sandwiching the quinoline ring ("the WY clamp"). Both residues are conserved across UGGTs of various species. The WY clamp sits nearby the conserved DQD motif which is involved in UDP-Glc binding (see Figure 27). The compound is now undergoing a first round of medicinal chemistry, including *in vitro* HsUGGT1-mediated re-glucosylation assays to assay inhibitory activity. Derivatives that may inhibit human UGGT1 *in vitro* will be tested in antiviral assays *in cellula*.

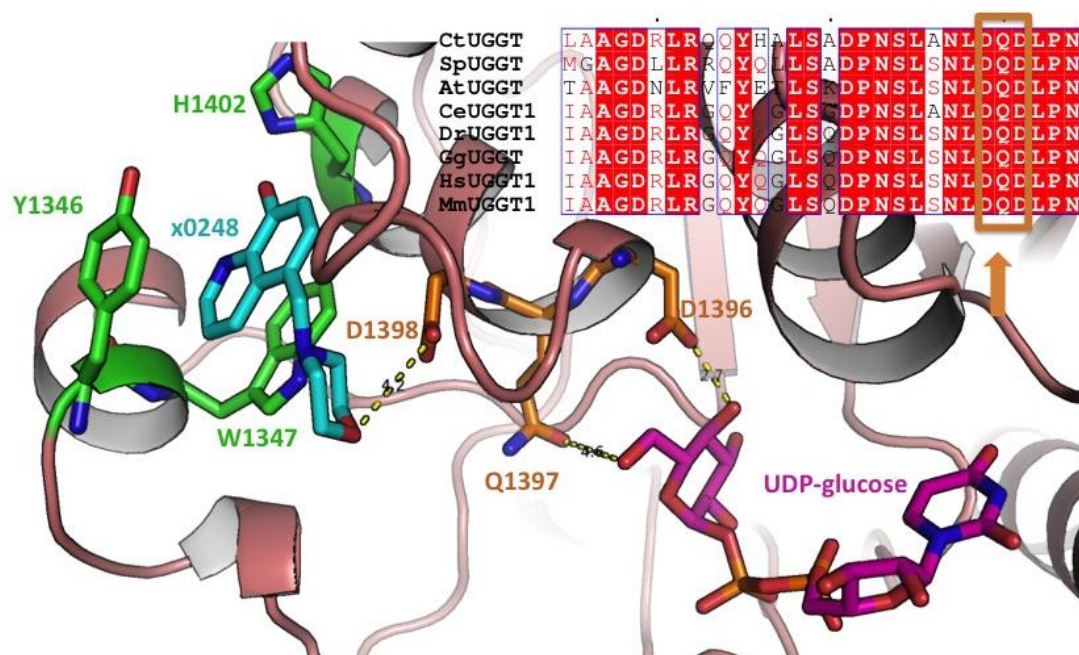


Figure 27 The x0248 binding site is in close proximity of the conserved DQD motif that is involved in binding UDP-Glc.

## 2.7. Conclusions

The *CtUGGT* mutants cloned, purified, expressed and assayed in this thesis have enhanced our understanding of the molecule and will be instrumental in designing host-targeting antiviral that are either directly targeting the enzyme's active site or acting on the portions of the molecule that effect misfold recognition and therefore may be relied upon by viruses to help viral glycoproteins' folding.

When it comes to correlating UGGT interdomain conformational mobility with its activity, among the *CtUGGT* double cysteine mutants tested so far,

the *CtUGGT*<sup>D611C/G1050C</sup> mutant described in ref. 45 is the least active in re-glucosylating urea-treated bovine thyroglobulin, compatible with our observation that its MD trajectory is the most severely limited one. The extra disulphide bridge engineered in *CtUGGT*<sup>D611C/G1050C</sup> joins the  $\beta$ S2-TRXL2 domains across the molecule's central cleft above the saddle, so this mutant's impaired activity suggests that during the enzyme:substrate encounter the central saddle may accommodate a portion of misfolded thyroglobulin. Indeed, the collective evidence of *CtUGGT* crystal structures, the cryo-EM structure<sup>45</sup> and our MD simulations all highlight the TRXL2 domain as the most mobile in the structure. The total loss of activity of the *CtUGGT*- $\Delta$ TRXL2 construct in the re-glucosylation of urea-treated bovine thyroglobulin (Figure 20) also supports a critical role for the TRXL2 domain in adjusting the size of the central cleft / varying the surface area of the saddle, making the bending motion a crucial one for UGGT activity. Interestingly, two *a priori* rather similar double cysteine mutants, *CtUGGT*<sup>S180C/T742C</sup> and *CtUGGT*<sup>G177C/A786C</sup>, both designed to constrain the TRXL1-TRXL3 clamping motion shut, differ significantly in their ability to re-glucosylate urea-treated bovine thyroglobulin, with the former mutant more active than, and the latter mutant having similar activity to WT *CtUGGT*. These observations point to the possibility that each misfolded glycoprotein substrate may depend on a different degree on a different subset of UGGT inter-domain conformational degrees of freedom. In the light of these data, the extent to which various portions of the UGGT

structure and its motions are critical to its activity will need further testing with re-glucosylation assays using the same set of UGGT mutants on different glycoprotein substrates.

In the light of our data, involvement of TRXL2 and placement of substrates between this domain and the  $\beta$ S2-GT24 rigid unit is the simplest hypothesis supported by the data so far. Substrate recruitment via TRXL2 movements would not require complete burial of the misfolded glycoprotein into the central cleft of the molecule: UGGT would minimally need to establish contact with the portion of the substrate containing the misfold. This is likely the case for relatively big substrates like transferrin (77 KDa, radius of gyration ( $r_{\text{og}}$ ) = 29.7 Å) or urea-treated bovine thyroglobulin (670 KDa, a long snake-like chain of eleven 60 aminoacids compact domains, no structure available). For smaller substrates, like glycopeptides or synthetic fluorescent probes, closer UGGT conformations, bringing TRXL2 towards  $\beta$ S2 and the GT24 domain across the central cleft may be needed. Apart from TRXL2, other untested UGGT regions potentially harbouring exposed hydrophobics are the *Ct*UGGT TRXL4 disordered region 243–285; the flexible linker around the endo-proteolysis site between the  $\beta$ S2 and GT24 domains (*Ct*UGGT 1153-1195), and the residues between the last helix and the ER retrieval motif at the C-terminus (*Ct*UGGT 1474-1510)<sup>45</sup>. Again, structural and functional data from a range of UGGT mutants and glycoprotein substrates will be required to further test these hypotheses and fully dissect UGGT structure-function relationship.

The first FBLD study hunting for UGGT fragment ligands by X-ray crystallography has yielded a fragment (the 8-OH-quinoline, fragment x0248) which when tested against full-length human UGGT1 has confirmed a sub-millimolar dissociation constant. The study constitutes a valid example of how molecules discovered by FBLD as ligands of fungal enzymes can prove to be ligands of the human orthologue enzyme as well. Moreover, the high degree of sequence conservation at the site of binding of x0248 and its proximity to the UDP-Glc binding site are very promising in terms of the potential of derivatives of this compound to evolve into the first class of UGGT inhibitors.



## Section III

### **3. Under-investigation target for newly synthesised antivirals**

The aim of my whole Ph.D. project was to study different approaches to the antiviral drug discovery as presented in Section I and II of this thesis. In Section III I will present a series of small benzimidazole derivatives which showed interesting activity against different viruses in a wide antiviral screening on a panel of representative viruses of several families. They were sorted by structure and antiviral activity and were investigated more deeply in order to better apprehend the mechanism of action. Only some evidence

and preliminary data can be reported here, further analysis is still in progress.

Target viruses are Vaccinia Virus (VV) and some Enteroviruses (EVs), such as Coxsackieviruses, EV-A71 and EV-D68. The synthesised compounds showed a few common features, and this is why I pooled them here together.

### 3.1. Introduction

Vaccinia virus (VV) is a double-stranded DNA virus, best known for playing a key role in the eradication of the smallpox virus. VV belongs to the *Poxvirus* family and to the *Orthopoxvirus* genus. It is a large virus that can infect different types of mammalian cells and it is widely studied as a representative of its genus<sup>87</sup>. It is characterised by a cytoplasmic replication cycle<sup>88</sup> that produces two different types of infectious virions, called mature virions and extracellular virions. These are morphologically different, because mature virions have a single envelope containing viral particles and can leave cells as they are after cell lysis, and mediate host-to-host infections. Alternatively, they receive an additional membrane before egress from the cells (*via* exocytosis), generating the so-called, extracellular virions which spread within a tissue or from tissue to tissue<sup>89,90</sup>. The additional membrane of extracellular virions must be removed before subsequent cell entry, so

cell entry process is almost the same for both extracellular and mature virions. It has been demonstrated that virions enter the cells by endocytosis and penetration occurs by membrane fusion in intracellular vacuoles, it includes micropinocytosis process<sup>91–93</sup>. Several viral proteins involved in binding to the cell surface have been identified so far<sup>94–96</sup> and they are all involved in endocytosis processes. Vaccinia virus is actually used for vaccination for smallpox in selected high-risk individuals, but some of them develop severe reactions to inoculation with VV<sup>97</sup>, so antiviral drug research could lead to the discovery of molecules that can prevent or treat VV skin infections.

Enterovirus genus of the family *Picornaviridae* includes poliovirus, coxsackieviruses echoviruses, numbered enteroviruses and rhinoviruses. These viruses, although quite different, can all cause a variety of diseases, from mild to severe disorders, including respiratory diseases, non-specific febrile illnesses, hand-foot-and-mouth disease, aseptic meningitis, encephalitis, paralysis, myocarditis, neonatal sepsis-like disease and also some chronic diseases, like insulin-dependent diabetes mellitus<sup>98–100</sup>. Enteroviral infections can be disseminated via faecal-oral and respiratory routes and in the last years several outbreaks of Enteroviruses (EVs) have been reported, mostly caused by Coxsackieviruses<sup>101–103</sup>, EV-A71<sup>104–106</sup> and EV-D68<sup>107–109</sup>, increasingly linked to severe neurological disease such as acute flaccid myelitis (EV-D68) and viral encephalitis (EV-A71). \*\*\*

Enteroviruses are non-enveloped viruses with positive-sense, single-stranded RNA viruses (ssRNA<sup>+</sup>), and their life cycle begins with binding to one or multiple cell surface receptors, resulting in receptor-mediated endocytosis. Enteroviruses can use different mechanism of endocytosis, depending on the serotype and cell type, but generally, the mechanism of entry involves binding to receptors such as transmembrane protein of epithelial tight junctions<sup>110</sup> and co-receptors that activates intracellular signalling pathways that lead to final endocytosis of the viruses. Nowadays there are no approved vaccines or effective drug treatment for Coxsackieviruses or EV-D68, two inactivated EV-A71 vaccines were recently developed and marketed only in China. This motivated our research of new potent anti-Enteroviruses antivirals.

## 3.2. Materials and methods

### 3.2.1. Synthesis

#### 3.2.1.1. General synthetic approach

All starting materials were purchased by Sigma-Aldrich, Across Organics and Carlo Erba producers. Benzimidazole ring closure has been carried out as described by Bahrami *et al.*<sup>23</sup> by mixing *o*-phenylenediamine and aldehydes in a ratio 1:1 in acetonitrile (CH<sub>3</sub>CN), added with H<sub>2</sub>O<sub>2</sub> 30% (ratio 1:7) and HCl 37% (ratio

1:3.5) at room temperature for a proper time, generally few hours and maximum to overnight time. Nitrogen groups reductions to obtain amine derivatives were performed in ethanol by three different routes as follow:

- in autoclave at 100°C with methylhydrazine in case of chlorine derivatives, to avoid displacement of chlorine;
- at room temperature with H<sub>2</sub> and Pd/C as catalyst;
- at 80°C with hydrazine and Pd/C as catalyst.

Products were purified by crystallization from ethanol or methanol, or by flash chromatography using proper elution by mixing suitable chosen solvents among petroleum spirit (PS), diethyl ether (DE), ethyl acetate (EA), chloroform and methanol.

### **3.2.1.2. Chemical characterisation**

#### **3.2.1.2.1. Nuclear Magnetic Resonance**

See 1.3.2.1.

#### **3.2.1.2.2. Mass spectroscopy**

See 1.3.2.2.

#### **3.2.1.2.3. Melting Point measurements**

See 1.3.2.3.

#### **3.2.1.2.4. Retention factors**

See 1.3.2.4.

### 3.2.2. Biological assays ♦

#### 3.2.2.1. Cell lines and viruses

Cell lines were purchased from American Type Culture Collection® (ATCC). The absence of mycoplasma contamination was checked periodically by the Hoechst staining method. Cell lines supporting the multiplication of Enteroviruses were the following: Monkey kidney (Vero-76) [ATCC® CRL-1587™ *Cercopithecus Aethiops*], Human cervix adenocarcinoma (Hela cells) [ATCC® CCL-2™] and Macaca mulatta, monkey, rhesus (LLC-MK2) [ATCC® CCL-7™]. Viruses were purchased from American Type Culture Collection® (ATCC), Viruses representative of positive-sense, single-stranded RNAs (ssRNA+) belong to *Picornaviridae* family: human enterovirus B [coxsackie type B4 (CV-B4), strain J.V.B. (ATCC® VR-184™), coxsackie type B5 (CV-B5), strain Faulkner (ATCC® VR-185™), coxsackie type B3 (CV-B3), strain Nancy (ATCC® VR-30™)], human enterovirus B [echovirus 9] human enterovirus A71 strain BrCr (ATCC® VR-1775™) and human enterovirus D 68 strain Fermon (ATCC® VR-1826™)]. Viruses representatives of DNA

---

♦ Biological assays were carried out by Dr. G. Sanna, Prof. R. Loddo and collaborators at the Department of Biomedical Science, Section of Microbiology and Virology, University of Cagliari (Italy)

group belong to *Poxviridae* family: Vaccinia Virus (VV) [strain Elstree- Lister Vaccine (ATCC® VR-1549™)] was used.

#### **3.2.2.2. Cytotoxicity assay**

Hela cells were seeded in 96-well plates at an initial density of  $5 \times 10^5$  cells/mL, in Minimum Essential Medium with Earle's salts (MEM-E), L-glutamine, 1mM sodium pyruvate and 25 mg/L kanamycin, supplemented with 10% foetal bovine serum (FBS). Vero-76 and LLC-MK2 cells were seeded in 96-well plates at an initial density of  $5 \times 10^5$  cells/mL, in Dulbecco's Modified Eagle Medium (D-MEM) with L-glutamine and 25 mg/L kanamycin, supplemented with 10% FBS. Cell cultures were then incubated at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere, in the absence or presence of serial dilutions of test compounds. The test medium used for the cytotoxic assay as well as for antiviral assay contained 1% of the appropriate serum. Cell viability was determined after 72-120 hrs at 37 °C by MTT method<sup>78</sup>.

#### **3.2.2.3. Antiviral assay**

Compound's activity against EV-A71, D68 was based on inhibition of virus-induced cytopathogenicity in Vero-76 and Hela cells respectively, acutely infected with at an m.o.i. of 0.01. After a 3 or 4-day incubation at 37 °C, cell viability was determined by the MTT method. Compound's activity against CV-B3, CV-B4, CV-B5 and Echo9 was determined by plaque reduction assays in infected cell

monolayers. Briefly, monolayer of Vero-76 cells or LLC-MK2 (Echo-9) was grown overnight on 24-well plate. The cells were then infected for 2 hrs with 250  $\mu$ L of proper virus dilutions to give 50-100 PFU/well (PFU = plaque forming unit). Following removal of unadsorbed virus, 500  $\mu$ L of medium [D-MEM with L-glutamine and 4500 mg/L D-glucose, supplemented with 1% inactivated FBS] containing 0.75% methyl-cellulose, with serial dilutions of test compounds, were added. The overlaid medium was also added to not treated wells as non-infection controls. Cultures were incubated at 37°C for 3 (CV-B3, CV-B4, CV-B5) or 5 days (E-9) and then fixed with PBS containing 50% ethanol and 0.8% crystal violet, washed and air-dried. Plaques in the control (no inhibitor) and experimental wells were then counted. NM 107 (2'-*C*-methylcytidine) and pleconaril were used as reference compounds.

#### **3.2.2.4. Yield reduction assay**

Vero-76 cells were inoculated with EV-A71 at a m.o.i. of 0.1 in maintenance medium and tested compounds at not cytotoxic concentrations. Following 2 hours adsorption period at 37 °C and 5% CO<sub>2</sub>, the inoculum was removed and replaced with fresh medium containing the same concentration of compounds. After 72 hours at 37 °C and 5% CO<sub>2</sub>, each sample was harvested and diluted with serial passages, starting from 10<sup>-1</sup> up to 10<sup>-8</sup>. The titer of the serial dilutions of the virus-containing supernatant was determined by standard



plaque assay, counting the number of obtained plaques in at least two different dilutions for each concentration. NM 107 (2'-*C*-methylcytidine) was used as reference compound.

#### **3.2.2.5. Virucidal activity assay**

**6h** (20  $\mu$ M) was incubated with  $1 \times 10^5$  PFU/ml of EV-A71 at either 4 or 37 °C for 1 hour. The mixture without test sample was used as control. At the end of the incubation period, samples were serially diluted in media and titres were determined on Vero-76 cells at high dilutions, at which compound was not active. Virus titres were determined by plaque assay in Vero-76 cells.

#### **3.2.2.6. Cell pre-treatment assay**

Vero-76 cell monolayers in 24-well plates were incubated with 20  $\mu$ M concentration of the **6h** or NM 107 (50  $\mu$ M) for 2 hrs at 4 °C. After removal of the compounds and two gentle washes, cells were infected with EVA71. After virus adsorption to cells, the inoculum was removed and the cells were then overlaid with medium, incubated for 4 days at 37 °C, and then virus titres were determined by plaque assay.

### 3.2.2.7. Time of addition assay for 6h derivative ♦

The confluent monolayers of Vero-76 cells in 24-well tissue culture plates were infected for 1 hrs at room temperature with EV-A71 dilutions to give final m.o.i. of 1-5. After adsorption, the monolayers were washed two times with D-MEM medium with L-glutamine, supplemented with 1% inactivated FBS, 1mM sodium pyruvate and 0.025 g/l kanamycin (Maintenance Medium) and incubated with the same medium at 5% CO<sub>2</sub> and 37 °C (time zero). Vero-76 cells were treated with compound **6h** (20 µM, approximately 10 times higher than the IC<sub>50</sub>) or reference for 1 hour during infection period (at -1 to 0) and at specific time point, 0 to 2, 2 to 4, 4 to 6, 6 to 8, and 8 to 10 hrs post-infection. After each incubation period, the monolayers were washed two times with maintenance medium and incubated with fresh medium until 12 hrs post-infection. Then, the monolayers were frozen at -80°C and the viral titres were determined by plaque assay.

### 3.2.2.8. Time of addition assay for 7c derivative ♦♦

A time-of-addition experiment was carried out with Vero-76 cells. The confluent monolayers of Vero-76 cells, seed in 24-well tissue

---

♦ TOA for 6h was carried out by Dr. G. Sanna at the Department of Biomedical Science, Section of Microbiology and Virology, University of Cagliari (Italy)

♦♦ TOA for 7c was carried out by Prof. R. Loddo and collaborators at the Department of Biomedical Science, Section of Microbiology and Virology, University of Cagliari (Italy)

culture plates, were inoculated at room temperature with 6650 PFU of CVB-5, corresponding to a multiplicity of infection of 0.03 PFU/cell. After adsorption for 60 min, the monolayers were washed two times with maintenance medium in the presence of FBS inactivated and incubated with the same medium at 5% CO<sub>2</sub> and 37 °C. The test medium containing 10 × EC<sub>50</sub> compound concentration, was added at -1 to 0 (adsorption), 0 to 2 (during infection), 2 to 4, 4 to 6, 6 to 8, 8 to 10. After each incubation period, the monolayers were washed two times with maintenance medium and incubated with fresh medium until 20 hrs post-infection. Then, after 24–36 hrs post-infection the CPE (cytopathic effect) was evaluated and the monolayers were collected, centrifuged and frozen at -80 °C. The viral titre was determined by a plaque reduction assay.

#### **3.2.2.9. Absorption assay**

Vero-76 cells grown in 24-well plate were infected with EVA71, with a m.o.i. of 0.1, in the presence or absence of compound **6h**. Multiwell plates were incubated for 120 min at 4 °C. Medium containing unadsorbed virus was then removed, cells were washed twice with PBS and overlayed with medium. Plaques were counted after 72 hours of incubation at 37 °C.

### 3.2.2.10. Penetration assay

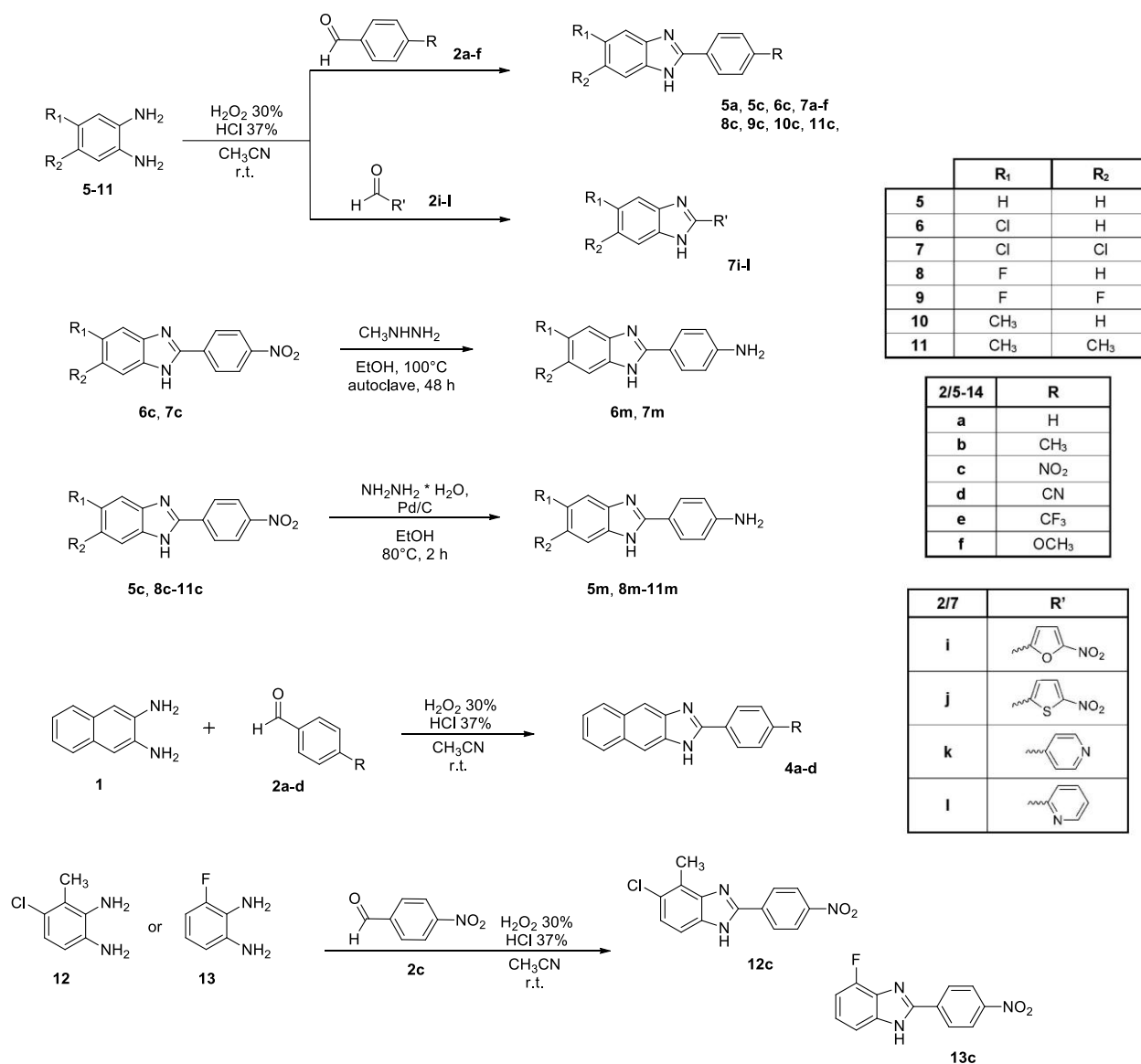
A 24-well tissue culture plate was seeded with Vero-76 cells ( $4 \times 10^5$  cells/well), which were then incubated overnight at 37 °C under 5% CO<sub>2</sub>. The cells were chilled on ice for 1 h, and the medium was removed. The cells were infected with 100 PFU (200ul, 1:80) of EV-A71 on ice for 2 h. The medium containing unbound virus was then removed; various concentrations of **6h** (100-0.8 µM) in medium were added, and the cells were incubated at 37 °C for 1 h to trigger endocytosis of the virus. The infected cells were then treated with alkaline phosphate-buffered saline (PBS; pH 11) for 1 min to inactivate any viruses that had not penetrated the cells, and then acidic PBS (pH 3) was immediately added to neutralize the mix. The neutralized medium was removed, cells were overlayed with 0.75% methylcellulose in media and then were incubated at 37 °C. After 72 hours incubation at 37°C cells were stained, and plaques determined by counting. Penetration assay was carried out according to the method<sup>111</sup> with some modifications.

## 3.3. Chemistry

### 3.3.1. Synthetic routes

#### 3.3.1.1. Anti-VV benzimidazole derivatives

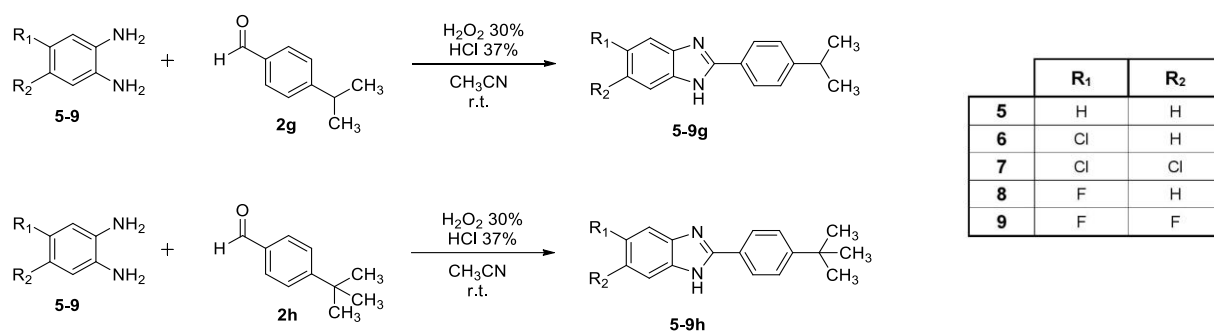
Derivatives **4a-d**, **5a**, **5-13c**, **7a-b**, **7d-f**, **7i-l** were obtained for reaction of proper o-phenylenediamine (**1**, **5-14**) with the corresponding aldehydes (**2a-f**, **i-l**) both dissolved in acetonitrile ( $\text{CH}_3\text{CN}$ ) and added of  $\text{H}_2\text{O}_2$  30% in a ratio 1:7 and HCl 37% in a ratio 1:3.5. Aniline derivatives **6-11m** were obtained by suitably reduction of corresponding nitro-derivatives **6-11c**. When  $\text{R}_1$  is a chlorine atom (**6c**), or when both  $\text{R}_1$  and  $\text{R}_2$  are chlorine atoms (**7c**) the nitrogen is reduced to amine group with methylhydrazine, as bland reducing agent in ethanol using the autoclave at  $100^\circ\text{C}$  for 48 hours. All the other reductions for nitro-derivatives **5c**, **8c-11c** were conducted in ethanol Pd/C as catalyst and hydrazine at  $80^\circ\text{C}$  for 2 hours.



Scheme 3 Synthetic route for derivatives **4a-d**, **5a**, **5-6c**, **5-11m**, **7a-f**, **7i-l**, **8-13c**.

### 3.3.1.2. Anti-EVs benzimidazole derivatives

Benzimidazole ring closure has been carried out by mixing *o*-phenylenediamine (**5-9**) and corresponding aldehyde (**2g**, **2h**) in a ratio 1:1 dissolved in acetonitrile ( $\text{CH}_3\text{CN}$ ), added with  $\text{H}_2\text{O}_2$  30% (ratio 1:7) and  $\text{HCl}$  37% (ratio 1:3.5) at room temperature overnight, obtaining derivatives **5-9g** and **5-9h**.



*Scheme 4 Synthetic route for derivatives 5-9g, 5-9h.*

### 3.3.2. Analytical section

#### 3.3.2.1. Synthesis, purification and characterisation of derivatives 4a-d, 5a, 5-6c, 5-11m, 7a-f, 7i-l, 8-13c

Synthesis of derivatives **4a-d**, **5a**, **5-6c**, **7a-f**, **7i-l**, **8-13c** was carried on as described in general synthetic procedure for benzimidazole ring closure: mixing o-phenylenediamine (**1**, **5-13**) and corresponding aldehyde (**2a-f**, **i-l**) in a ratio 1:1 dissolved in acetonitrile (CH<sub>3</sub>CN), added with H<sub>2</sub>O<sub>2</sub> 30% (ratio 1:7) and HCl 37% (ratio 1:3.5) at room temperature overnight, solid is obtained. The solid has been filtered off with vacuum, washed with acetonitrile and with water till neutral pH of filtrate. The crude solid has been stored and dried in the oven overnight. Compounds **5-11m** were obtained by reduction of nitro-derivatives **5-11c** as previously described.

#### 3.3.2.1.1. Characterization of derivatives 4a-d

See Section I

#### 3.3.2.1.2. Characterisation of derivative 2-phenyl-1*H*-benzo[*d*]imidazole (5a)

Compound **5a** (C<sub>13</sub>H<sub>10</sub>N<sub>2</sub>, MW 194.232) was obtained in total yield 76%; m.p. 290-292 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.35. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.31 (2H, m, H-2',6'), 7.74 (2H, m, H-4,7), 7.64 (3H, m, H-3',4',5'), 7.40 (2H, m, H-5,6). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>): δ 150.32 (C), 135.79 (2C), 132.19, (CH), 129.78 (2CH), 127.82 (2CH), 126.86 (C), 124.63 (2CH), 114.98 (2CH). ESI-MS (*m/z*): calcd. for C<sub>12</sub>H<sub>10</sub>N<sub>2</sub> 195.09167, found 195.09151 [M+H]<sup>+</sup>.

#### 3.3.2.1.3. Characterisation of 2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (5c)

Compound **5c** (C<sub>13</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub>, MW 239.229) was obtained in total yield 44%; m.p. 214-215 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.72. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.49 (4H, s, H-2',3',5',6'), 7.76 (2H, m, H-4,7), 7.41 (2H, m, H-5,6). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 148.45 (C), 148.24 (C), 136.98 (2C), 133.63 (C), 128.10 (2CH), 124.41 (2CH), 124.18 (2CH), 115.18 (2CH). ESI-MS (*m/z*): calcd. for C<sub>13</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub> 240.07675, found 240.07655 [M+H]<sup>+</sup>.



#### 3.3.2.1.4. Characterisation of 4-(1*H*-benzo[*d*]imidazol-2-yl)aniline (5m)

Compound **5m** (C<sub>13</sub>H<sub>11</sub>N<sub>3</sub>, MW 209.247) was obtained in total yield 63%; m.p. >300 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.38. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 7.86 (2H, d, J= 8.4 Hz, H-2',6'), 7.50 (2H, m, H-4,7), 7.14 (2H, m, H-5,6), 6.68 (2H, d, J= 8.4 Hz, H-3',5'). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 152.38 (C), 150.76 (2C), 138.88 (C), 127.85 (2CH), 121.46 (2CH), 116.63 (C), 113.54 (2CH), 113.22 (2CH). ESI-MS (*m/z*): calcd. for C<sub>13</sub>H<sub>11</sub>N<sub>3</sub> 210.10257, found 210.10257 [M+H]<sup>+</sup>.

#### 3.3.2.1.5. Characterisation of 5-chloro-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (6c)

Compound **6c** (C<sub>13</sub>H<sub>8</sub>ClN<sub>3</sub>O<sub>2</sub>, MW 273.674) was obtained in total yield 97%; m.p. 256-258 °C; TLC (PS/ES 7/3): R<sub>f</sub> 0.44. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.43 (4H, s, H-2',3',5',6'), 7.73 (1H, d, J= 1.6 Hz, H-4), 7.69 (1H, d, J= 8.4 Hz, H-7), 7.308 (1H, dd, <sup>1</sup>J= 10.4 Hz, <sup>2</sup>J= 2 Hz, H-6), 3.68 (1H, s, NH). <sup>13</sup>C-NMR (DMSO-*d*<sub>6</sub>): δ 150.30 (C), 148.07 (C), 140.27 (C), 137.78 (C), 135.25 (C), 127.62 (2CH), 127.36 (C), 124.31 (2CH), 123.39 (CH), 116.70 (CH), 115.27 (CH). ESI-MS (*m/z*): calcd for C<sub>13</sub>H<sub>8</sub>ClN<sub>3</sub>O<sub>2</sub> 274.03778, found 274.03760 [M+H]<sup>+</sup>.

#### 3.3.2.1.6. Characterisation of 4-(5-chloro-1*H*-benzo[*d*]imidazol-2-yl)aniline (6m)

Compound **6m** (C<sub>13</sub>H<sub>10</sub>ClN<sub>3</sub>, MW 243.692) was obtained in total yield 86%; m.p. 133-135 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.29. <sup>1</sup>H-

NMR (DMSO- $d_6$ ):  $\delta$  7.83 (2H, d,  $J$ = 8.4 Hz, H-2',6'), 7.54 (1H, m, H-4), 7.43 (1H, ws, H-7), 7.13 (1H, d,  $J$ = 8.4 Hz, H-6), 6.66 (2H, d,  $J$ = 8.4 Hz), 5.66 (2H, s, NH<sub>2</sub>). <sup>13</sup>C-NMR (APT, DMSO- $d_6$ +TFA- $d$ ):  $\delta$  153.63 (C), 150.93 (C), 132.17 (C), 130.24 (C), 129.65 (C), 129.60 (2CH), 125.52 (CH), 114.42 (CH), 114.13 (2CH), 112.76 (CH), 107.87 (C). ESI-MS ( $m/z$ ): calcd. for C<sub>13</sub>H<sub>10</sub>ClN<sub>3</sub> 244.06360, found 244.06346 [M+H]<sup>+</sup>.

#### 3.3.2.1.7. Characterisation of 5,6-dichloro-2-phenyl-1*H*-benzo[*d*]imidazole (7a)

Compound **7a** (C<sub>13</sub>H<sub>8</sub>Cl<sub>2</sub>N<sub>2</sub>, MW 263.122) was obtained in total yield 65%; m.p. >300 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.15. <sup>1</sup>H-NMR (DMSO- $d_6$ ):  $\delta$  8.19 (2H, d,  $J$ = 6.4 Hz, H-2',6'), 7.90 (2H, s, H-4,7), 7.60 (3H, m, H-3',4',5'). <sup>13</sup>C-NMR (jmod, DMSO- $d_6$ ):  $\delta$  153.23 (C), 137.36 (C), 131.28 (CH), 129.20 (3CH), 127.79 (C), 127.62 (C), 127.06 (2CH), 125.48 (C), 125.37 (C), 116.13 (CH). ESI-MS ( $m/z$ ): calcd. for C<sub>13</sub>H<sub>8</sub>Cl<sub>2</sub>N<sub>2</sub> 263.01373, found 263.01422 [M+H]<sup>+</sup>.

#### 3.3.2.1.8. Characterisation of 5,6-dichloro-2-(*p*-tolyl)-1*H*-benzo[*d*]imidazole (7b)

Compound **7b** (C<sub>14</sub>H<sub>10</sub>Cl<sub>2</sub>N<sub>2</sub>, MW 277.149) was obtained in total yield 53%; m.p. 286 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.45. <sup>1</sup>H-NMR (DMSO- $d_6$ ):  $\delta$  8.54 (2H, d,  $J$ = 8 Hz, H-2',6'), 8.19(2H, s, H-4,7), 7.27 (2H, d,  $J$ = 8 Hz, H-3',5'), 2.30 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (ATP, DMSO- $d_6$ +TFA):  $\delta$  150.73 (C), 140.81 (C), 140.00 (C), 137.73 (C), 130.71 (C),

128.38 (2CH), 127.08 (C), 126.41 (2CH), 125.88 (C), 118.25 (2CH), 23.58 (CH<sub>3</sub>). ESI-MS ( $m/z$ ): calcd. for C<sub>14</sub>H<sub>10</sub>Cl<sub>2</sub>N<sub>2</sub> 277.02938, found 277.02945 [M+H]<sup>+</sup>.

**3.3.2.1.9. Characterisation of 5,6-dichloro-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (7c)**

Compound **7c** (C<sub>13</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub>, MW 308.120) was obtained in total yield 69%; m.p. 226 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.58. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.40 (4H, ws, H-2',3',5',6'), 7.91 (2H, s, H-4,7). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 151.50 (C), 148.25 (C), 188.80 (2C), 134.88 (C), 127.42 (2CH), 125.47 (2C), 124.32 (2CH), 116.83 (2CH). ESI-MS ( $m/z$ ): calcd. for C<sub>13</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub> 307.99881, found 307.99899 [M+H]<sup>+112</sup>.

**3.3.2.1.10. Characterisation of 4-(5,6-dichloro-1*H*-benzo[*d*]imidazol-2-yl)benzonitrile (7d)**

Compound **7d** (C<sub>14</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub>, MW 288.132) was obtained in total yield 79%; m.p. >300 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.47. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.30 (2H, d, J= 8.4 Hz, H-2',6'), 8.03 (2H, d, J= 8.4 Hz, H-3',5'), 7.89 (2H, s, H-4,7). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 151.89 (C), 140.63 (C), 133.32 (C), 132.99 (2CH), 131.33 (C), 127.94 (CH), 127.54 (2CH), 126.16 (CH), 125.21 (C), 124.82 (C), 118.42 (C), 112.51 (C). ESI-MS ( $m/z$ ): calcd. for C<sub>14</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub> 288.00898, found 288.00891 [M+H]<sup>+</sup>.

**3.3.2.1.11. Characterisation of 5,6-dichloro-2-(4-(trifluoromethyl)phenyl)-1*H*-benzo[*d*]imidazole (7e)**

Compound **7e** (C<sub>14</sub>H<sub>7</sub>Cl<sub>2</sub>F<sub>3</sub>N<sub>2</sub>, MW 331.120) was obtained in total yield 68%; m.p. 225 °C; TLC (PS/EA 9/1): R<sub>f</sub> 0.29. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.40 (2H, d, J= 8 Hz, H-2',6'), 8.14 (2H, d, J= 8 Hz, H-3',5'), 7.98 (2H, s, H-4,7). <sup>13</sup>C-NMR (ATP, DMSO-*d*<sub>6</sub>+TFA): δ 150.92 (C), 135.52 (C), 131.47 (1C, q, <sup>1</sup>J<sub>C-F</sub>= 30 Hz, CF<sub>3</sub>), 130.11 (C), 128.11 (2CH), 126.87 (C), 126.82 (C), 126.08 (2C, <sup>3</sup>J<sub>C-F</sub>= 4 Hz, 2CH), 125.02 (C), 122.32 (C), 116.19 (2CH). ESI-MS (*m/z*): calcd. for C<sub>14</sub>H<sub>7</sub>Cl<sub>2</sub>F<sub>3</sub>N<sub>2</sub> 331.00111, found 331.00125 [M+H]<sup>+</sup>.

**3.3.2.1.12. Characterisation of 5,6-dichloro-2-(4-methoxyphenyl)-1*H*-benzo[*d*]imidazole (7f)**

Compound **7f** (C<sub>14</sub>H<sub>10</sub>Cl<sub>2</sub>N<sub>2</sub>O, MW 293.148) was obtained in total yield 51%; m.p. 225 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.34. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.16 (2H, d, J= 8.8 Hz, H-2',6'), 7.92 (1H, s, H-7), 7.76 (1H, s, H-4), 7.17 (2H, d, J= 8.8 Hz, H-3',5'), 3.90 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+TFA-*d*): δ 163.52 (C), 151.20 (C), 131.65 (2C), 130.04 (2CH), 128.15 (2C), 115.18 (2CH), 115.10 (2CH), 114.50 (C), 55.62 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>14</sub>H<sub>10</sub>Cl<sub>2</sub>N<sub>2</sub>O 293.02429, found 293.02429 [M+H]<sup>+</sup>.

**3.3.2.1.13. Characterisation of 5,6-dichloro-2-(5-nitrofuran-2-yl)-1*H*-benzo[*d*]imidazole (7i)**

Compound **7i** (C<sub>11</sub>H<sub>5</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>3</sub>, MW 298.082) was obtained in total yield 81%; m.p. 225 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.44. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 7.92 (2H, s, H-4,7), 7.90 (1H, d, J= 4 Hz, H-4'), 7.55 (1H, d, J= 4 Hz, H-3'). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+TFA-*d*): δ 151.92 (C), 145.94 (C), 142.90 (C), 137.99 (2C), 126.08 (2C), 116.74 (2CH), 114.43 (CH), 114.12 (CH). ESI-MS (*m/z*): calcd. for C<sub>11</sub>H<sub>5</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>3</sub> 297.97807, found 297.97775 [M+H]<sup>+</sup>.

**3.3.2.1.14. Characterisation of 5,6-dichloro-2-(5-nitrothiophen-2-yl)-1*H*-benzo[*d*]imidazole (7j)**

Compound **7j** (C<sub>11</sub>H<sub>5</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub>S, MW 314.147) was obtained in total yield 58%; m.p. >300 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.35. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.18 (1H, d, J= 4.4 Hz, H-4'), 7.86 (2H, s, H-4,7), 7.82 (1H, d, J= 4.4 Hz, H-3'). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 151.55 (C), 147.01 (C), 146.93 (C), 139.36 (2C), 130.89 (2CH), 126.79 (2CH), 125.77 (2C). ESI-MS (*m/z*): calcd. for C<sub>11</sub>H<sub>5</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub>S 313.95523, found 313.95499 [M+H]<sup>+</sup>.

**3.3.2.1.15. Characterisation of 5,6-dichloro-2-(pyridin-4-yl)-1*H*-benzo[*d*]imidazole (7k)**

Compound **7k** (C<sub>12</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub>, MW 264.110) was obtained in total yield 58%; m.p. >300 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.44. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.99 (2H, d, J= 5.6 Hz, H-2',6'), 8.55 (2H, d, J= 5.6 Hz, H-3',5'), 7.96 (2H, s, H-4,7). <sup>13</sup>C-NMR (ATP, DMSO-*d*<sub>6</sub>+TFA-*d*): δ 148.14 (C), 144.31 (C), 142.72 (2CH), 138.77 (2C), 126.85 (2CH), 123.34 (2CH), 117.46 (2CH). ESI-MS (*m/z*): calcd. for C<sub>12</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub> 264.00898, found 264.00928 [M+H]<sup>+</sup>.

**3.3.2.1.16. Characterisation of 5,6-dichloro-2-(pyridin-2-yl)-1*H*-benzo[*d*]imidazole (7l)**

Compound **7l** (C<sub>12</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub>, MW 264.110) was obtained in total yield 58%; m.p. >300 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.64. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.77 (1H, d, J= 3.6 Hz, H-6'), 8.33 (1H, d, J= 7.6 Hz, H-3'), 8.04 (1H, t, J= 7 Hz, H-4'), 7.86 (2H, s, H-4,7), 7.59 (1H, t, J= 5.6 Hz, H-5'). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 152.76 (C), 149.57 (CH), 147.02 (C), 138.29 (2C), 137.82 (2CH), 125.54 (CH), 125.25 (2C), 121.96 (CH), 116.75 (CH). ESI-MS (*m/z*): calcd. for C<sub>12</sub>H<sub>7</sub>Cl<sub>2</sub>N<sub>3</sub> 264.00898, found 264.00906 [M+H]<sup>+</sup>.

**3.3.2.1.17. Characterisation of 4-(5,6-dichloro-1*H*-benzo[*d*]imidazol-2-yl)aniline (7m)**

Compound **7m** (C<sub>13</sub>H<sub>9</sub>Cl<sub>2</sub>N<sub>3</sub>, MW 278.137) was obtained in total yield 97%; m.p. 222-223 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.28.

$^1\text{H-NMR}$  ( $\text{DMSO-}d_6$ ):  $\delta$  7.83 (2H, d,  $J$  = 7.6 Hz, H-2'6'), 7.78 (1H, s, H-4), 7.61 (1H, s, H-7), 6.67 (2H, d,  $J$  = 7.6 Hz, H-3'5'), 5.72 (2H, s,  $\text{NH}_2$ ).  $^{13}\text{C-NMR}$  (jmod,  $\text{DMSO-}d_6$ ):  $\delta$  155.21 (C), 151.22 (C), 143.89 (C), 134.56 (C), 128.12 (2CH), 123.46 (C), 123.28 (C), 118.78 (CH), 113.48 (2CH), 116.05 (C), 111.76 (CH). ESI-MS ( $m/z$ ): calcd. for  $\text{C}_{13}\text{H}_9\text{Cl}_2\text{N}_3$  278.02463, found 278.02478  $[\text{M}+\text{H}]^+$ .

#### 3.3.2.1.18. Characterisation of 5-fluoro-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (8c)

Compound **8c** ( $\text{C}_{13}\text{H}_8\text{FN}_3\text{O}_2$ , MW 257.220) was obtained in total yield 87%; m.p. 226.3 °C; TLC (PS/EA 7/3):  $R_f$  0.45.  $^1\text{H-NMR}$  ( $\text{DMSO-}d_6$ ):  $\delta$  8.45 (4H, d,  $J$  = 1,6 Hz, H-2',3',5',6'), 7.74 (1H, m, H-7'), 7.45 (1H, m, H-4'), 7.23 (1H, m, H-6').  $^{13}\text{C-NMR}$  (jmod,  $\text{DMSO-}d_6$ ):  $\delta$  160.54 (C), 158.18 (C), 149.72 (C), 148.35 (C), 137.91 (C), 133.97 (C), 127.95 (2CH), 124.36 (2CH), 116.44 (CH), 112.26 (CH), 101.29 (CH). ESI-MS ( $m/z$ ): calcd. for  $\text{C}_{13}\text{H}_8\text{FN}_3\text{O}_2$  258.06733, found 258.06732  $[\text{M}+\text{H}]^+$ .

#### 3.3.2.1.19. Characterisation of 4-(5-fluoro-1*H*-benzo[*d*]imidazol-2-yl)aniline (8m)

Compound **8m** ( $\text{C}_{13}\text{H}_{10}\text{FN}_3$ , MW 227.237) was obtained in total yield 63%; m.p. 271 °C; TLC (PS/EA 7/3):  $R_f$  0.12.  $^1\text{H-NMR}$  ( $\text{DMSO-}d_6$ ):  $\delta$  8.05 (2H, d,  $J$  = 8.8 Hz, H-2',6'), 7.72 (1H, m, H-7), 7.54 (1H, m, H-4), 7.40 (1H, m, H-6), 6.78 (2H, d,  $J$  = 8.8 Hz, H-3',5').  $^{13}\text{C-NMR}$  (jmod,  $\text{DMSO-}d_6$ ):  $\delta$  160.81 (C), 158.42 (C), 153.90 (C), 151.15 (C),

132.42 (C), 129.67 (2CH), 128.48 (C), 114.43 (CH), 113.66 (2CH), 113.02 (CH), 99.95 (CH). ESI-MS ( $m/z$ ): calcd. for  $C_{13}H_{10}FN_3$  228.09315, found 228.09319  $[M+H]^+$ .

**3.3.2.1.20. Characterisation of 5,6-difluoro-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (9c)**

Compound **9c** ( $C_{13}H_7F_2N_3O_2$ , MW 275.210) was obtained in total yield 52%; m.p. 292 °C; TLC (PS/EA 7/3):  $R_f$  0.40.  $^1H$ -NMR (DMSO- $d_6$ ):  $\delta$  8.39 (4H, m, H-2',3',5',6'), 7.715 (2H, t, H-4,7).  $^{13}C$ -NMR (jmod, DMSO- $d_6$ ):  $\delta$  149.70 (C), 148.88 (C), 148.39 (2C, dd,  $^1J_{C-F}$  = 245 Hz,  $^2J_{C-F}$  = 19 Hz, 2C-F), 131.84 (C, m), 131.66 (C), 131.25 (C, m), 128.35 (2CH), 124.28 (2CH), 102.90 (2C, dd,  $^2J_{C-F}$  = 15 Hz,  $^3J_{C-F}$  = 8 Hz, CH-4,7). ESI-MS ( $m/z$ ): calcd. for  $C_{13}H_7F_2N_3O_2$  276.05791, found 276.05856  $[M+H]^+$ .

**3.3.2.1.21. Characterisation of 4-(5,6-difluoro-1*H*-benzo[*d*]imidazol-2-yl)aniline (9m)**

Compound **9m** ( $C_{13}H_7F_2N_3O_2$ , MW 245.227) was obtained in total yield 87%; m.p. 246-247 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5):  $R_f$  0.40.  $^1H$ -NMR (DMSO- $d_6$ ):  $\delta$  7.81 (2H, d,  $J$  = 8.4 Hz, H-2',6'), 7.54 (1H, s, H-4), 7.448 (1H, s, H-7), 6.67 (2H, d,  $J$  = 8.8 Hz, H-3',5'), 5.64 (2H, s, NH<sub>2</sub>).  $^{13}C$ -NMR (jmod, DMSO- $d_6$ ):  $\delta$  154.62 (C), 150.85 (2C), 146.21 (2C, dd,  $^1J_{C-F}$  = 236 Hz,  $^2J_{C-F}$  = 16 Hz, C-F), 127.75 (2CH), 116.53 (2C), 113.51 (2CH), 102.45-101.30 (2C, m, CH-F). ESI-MS ( $m/z$ ): calcd. for  $C_{13}H_9F_2N_3$  246.08373, found 246.08331  $[M+H]^+$ .



**3.3.2.1.22. Characterisation of 5-methyl-2-(4-nitrophenyl)-  
1*H*-benzo[*d*]imidazole (10c)**

Compound **10c** (C<sub>14</sub>H<sub>11</sub>N<sub>3</sub>O<sub>2</sub>, MW 253.256) was obtained in total yield 42%; m.p. 94-95 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.74. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.57 (2H, d, J= 9.2 Hz, H-3',5'), 8.52 (2H, d, J= 8.8 Hz, H-2',6'), 7.73 (1H, d, J= 8.4 Hz, H-7), 7.63 (1H, s, H-4), 7.37 (1H, d, J= 8.4 Hz), 2.50 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 149.13 (C), 146.57 (C), 135.75 (C), 133.82 (C), 131.98 (C), 130.62 (C), 128.91 (2CH), 127.32 (CH), 124.48 (2CH), 114.28 (CH), 113.76 (CH), 21.23 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>14</sub>H<sub>11</sub>N<sub>3</sub>O<sub>2</sub> 254.09240, found 254.09219 [M+H]<sup>+</sup>.

**3.3.2.1.23. Characterisation of 4-(5-methyl-1*H*-  
benzo[*d*]imidazol-2-yl)aniline (10m)**

Compound **10m** (C<sub>14</sub>H<sub>13</sub>N<sub>3</sub>, MW 223.273) was obtained in total yield 31%; m.p. 110-112 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.26. <sup>1</sup>H-NMR (acetone-*d*<sub>6</sub>): δ 7.94 (2H, d, J= 8.8 Hz, H-2',6'), 7.39 (1H, d, J= 8 Hz, H-7), 7.31 (1H, s, H-4), 6.97 (1H, d, J= 8 Hz, H-6), 6.78 (2H, d, J= 8.4 Hz, H-3',5'), 5.09 (2H, s, NH<sub>2</sub>), 2.42 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (APT, DMSO-*d*<sub>6</sub>+ TFA-*d*): δ 153.69 (C), 149.49 (C), 135.10 (C), 131.54 (C), 129.33 (2CH), 129.11 (C), 126.45 (CH), 114.07 (2CH), 113.42 (2CH), 108.04 (C), 20.97 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>14</sub>H<sub>13</sub>N<sub>3</sub> 224.11822, found 224.11761 [M+H]<sup>+</sup>.

**3.3.2.1.24. Characterisation of 5,6-dimethyl-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (11c)**

Compound **11c** (C<sub>15</sub>H<sub>13</sub>N<sub>3</sub>O<sub>2</sub>, MW 267.283) was obtained in total yield 62%; m.p. 260-262 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.43. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.38 (4H, s, H-2',3',5',6'), 7.42 (2H, s, H-4,7), 2.34 (6H, s, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (DMSO- *d*<sub>6</sub>): δ 147.92 (C), 147.52 (C), 137.75 (C), 135.79 (2C), 132.07 (2C), 128.85 (2CH), 127.69 (2CH), 124.19 (2CH), 19.96 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>15</sub>H<sub>13</sub>N<sub>3</sub>O<sub>2</sub> 268.10805, found 268.10797 [M+H]<sup>+</sup>.

**3.3.2.1.25. Characterisation of 4-(5,6-dimethyl-1*H*-benzo[*d*]imidazol-2-yl)aniline (11m)**

Compound **11m** (C<sub>15</sub>H<sub>15</sub>N<sub>3</sub>, MW 237.299) was obtained in total yield 88%; m.p. 109-111 °C; TLC (CHCl<sub>3</sub>/CH<sub>3</sub>OH 95/5): R<sub>f</sub> 0.17. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 7.80 (2H, d, J= 8.4 Hz, H-2',6'), 7.31 (1H, s, H-7), 7.18 (1H, s, H-4), 6.65 (2H, d, J= 8.8 Hz, H-3',5'), 5.55 (2H, s, NH<sub>2</sub>), 2.29 (6H, d, J= 7.2 Hz). <sup>13</sup>C-NMR (DMSO-*d*<sub>6</sub>): δ 151.62 (C), 150.21 (C), 137.32 (2C), 130.01 (2C), 127.52 (2CH), 117.13 (2C), 114.37 (C), 113.78 (2CH), 19.84 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>15</sub>H<sub>15</sub>N<sub>3</sub> 238.13387, found 238.13365 [M+H]<sup>+</sup>.

**3.3.2.1.26. Characterisation of 6-chloro-7-methyl-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (12c)**

Compound **12c** (C<sub>14</sub>H<sub>10</sub>ClN<sub>3</sub>O<sub>2</sub>, MW 287.701) was obtained in total yield 78%; m.p. >300 °C; TLC (PS/EA 8/2): R<sub>f</sub> 0.52. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.37 (4H, s, H-2',3',5',6'), 7.52 (2H, d, J<sub>ortho</sub> = 8.8 Hz, H-4), 7.36 (2H, d, J<sub>ortho</sub> = 8.4 Hz, H-5), 2.59 (3H, s, CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 149.10 (C), 148.42 (C), 138.49 (C), 135.02 (C), 133.61 (C), 128.44 (2CH), 127.92 (C), 124.83 (CH), 124.19 (2CH), 123.75 (C), 113.12 (CH), 14.53 (CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>14</sub>H<sub>10</sub>ClN<sub>3</sub>O<sub>2</sub> 288.05343, found 288.05396 [M+H]<sup>+</sup>.

**3.3.2.1.27. Characterisation of 4-fluoro-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole (13c)**

Compound **13c** (C<sub>13</sub>H<sub>8</sub>FN<sub>3</sub>O<sub>2</sub>, MW 257.220) was obtained in total yield 65%; m.p. >300 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.81. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.45 (4H, dd, J = 9.2 Hz, H-2',3',5',6'), 7.47 (1H, d, J<sub>ortho</sub> = 8 Hz, H-7), 7.28 (1H, td, J<sub>ortho</sub> = 8.1 Hz, J<sub>meta</sub> H-F = 5.1 Hz, H-6), 7.09 (1H, m, H-5). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 153.46 (C), 150.98 (C), 149.64 (C), 148.08 (C), 139.71 (1C, m, C-F), 135.17 (C), 127.71 (2CH), 124.29 (2CH), 123.99 (1C, d, <sup>3</sup>J<sub>C-F</sub> = 7 Hz, CH-6), 109.81 (CH-7), 107.83 (1C, d, <sup>2</sup>J<sub>C-F</sub> = 16 Hz, CH-5). ESI-MS (*m/z*): calcd. for C<sub>13</sub>H<sub>8</sub>FN<sub>3</sub>O<sub>2</sub> 258.06733, found 258.06766 [M+H]<sup>+</sup>.

### 3.3.2.2. Synthesis, purification and characterisation of derivatives 5-9 g and 5-9h

Synthesis of derivatives 5-9g and 5-9h was carried on as described in general synthetic procedure for benzimidazole ring closure: mixing o-phenylenediamine (5-9) and corresponding aldehyde (2g, 2h) in a ratio 1:1 dissolved in acetonitrile ( $\text{CH}_3\text{CN}$ ), added with  $\text{H}_2\text{O}_2$  30% (ratio 1:7) and HCl 37% (ratio 1:3.5) at room temperature overnight, solid is obtained. The solid has been filtered off with vacuum, washed with acetonitrile and with water till neutral pH of filtrate. The crude solid has been stored and dried in oven overnight. All the below listed compounds were obtained as pure solids after crystallisation from ethanol, and then characterised.

#### 3.3.2.2.1. Characterisation of derivative 2-(4-isopropylphenyl)-1H-benzo[d]imidazole (5g)

Compound 5g ( $\text{C}_{16}\text{H}_{16}\text{N}_2$ , MW 236.312) was obtained in total yield 88%; m.p. 206-207 °C; TLC (PS/EA 7/3):  $R_f$  0.62.  $^1\text{H}$ -NMR ( $\text{DMSO}-d_6$ ):  $\delta$  8.11 (2H, d,  $J$  = 8.6 Hz, H-2',6'), 7.66 (1H, d,  $J$  = 7 Hz, H-4), 7.61 (2H, d,  $J$  = 8.6 Hz, H-3',5'), 7.54 (1H, d,  $J$  = 7 Hz, H-7), 7.21 (2H, m, H-5,6), 1.34 (9H, s, 3 $\text{CH}_3$ ).  $^{13}\text{C}$ -NMR (jmod,  $\text{DMSO}-d_6$ ):  $\delta$  152.49 (C), 151.18 (C), 142.98 (C), 136.94 (C), 128.42 (C), 127.19 (2CH), 125.79 (2CH), 123.55 (CH), 122.50 (CH), 117.68 (CH), 111.16 (CH),

33.46 (CH), 22.99 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd for C<sub>20</sub>H<sub>14</sub>N<sub>4</sub>O<sub>3</sub> 359.11387, found 359.11386 [M+H]<sup>+</sup>.

**3.3.2.2.2. Characterisation of derivative 5-chloro-2-(4-isopropylphenyl)-1*H*-benzo[*d*]imidazole (6g)**

Compound **6g** (C<sub>16</sub>H<sub>15</sub>ClN<sub>2</sub>, MW 270.76) was obtained in total yield 75%; m.p. 228-230 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.75. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.12 (2H, d, J= 8.4 Hz, H-2',6'), 7.66 (1H, d, J= 1.6 Hz, H-4), 7.61 (1H, d, J= 8.8 Hz, H-7), 7.40 (2H, d, J= 8.4 Hz, H-3',5'), 7.19 (1H, dd, J<sub>ortho</sub>= 8.6 Hz, J<sub>meta</sub>= 1.6 Hz, H-6), 2.95 (1H, q, aliphatic CH), 1.24 (6H, d, J= 6.8 Hz, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 152.90 (C), 150.61 (C), 127.46 (2C), 126.86 (2CH), 126.67 (2CH), 126.06 (2C), 121.97 (CH), 116.01 (CH), 114.85 (CH), 33.32 (CH), 15.18 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>16</sub>H<sub>15</sub>ClN<sub>2</sub> 271.09965, found 271.10001 [M+H]<sup>+</sup>.

**3.3.2.2.3. Characterisation of derivative 5,6-dichloro-2-(4-isopropylphenyl)-1*H*-benzo[*d*]imidazole (7g)**

Compound **7g** (C<sub>16</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>2</sub>, MW 305.202) was obtained in total yield 87%; m.p. 257-259 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.75. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.19 (d, 2H, J= 8 Hz, H-2',6'), 7.96 (s, 2H, H-4,7), 7.53 (d, 2H, J= 8 Hz, H-3',5'), 3.01 (q, 1H, J= 6.8 Hz, CH), 1.26 (d, 6H, J= 6.8 Hz, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 153.19 (C), 152.38 (C), 135.09 (2C), 127.70 (2CH), 127.60 (2CH), 126.43 (2C), 123.42 (C),

115.73 (2CH), 33.47 (CH), 23.46 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>16</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>2</sub> 305.06068, found 305.06085 [M+H]<sup>+</sup>.

#### 3.3.2.2.4. Characterisation of derivative 5-fluoro-2-(4-isopropylphenyl)-1*H*-benzo[*d*]imidazole (8g)

Compound **8g** (C<sub>16</sub>H<sub>15</sub>FN<sub>2</sub>, MW 254.302) was obtained in total yield 76%; m.p. 255 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.56. <sup>1</sup>H-NMR (acetone-*d*<sub>6</sub>): δ 8.14 (2H, d, *J* = 8 Hz, H-2',6'), 7.58 (1H, m, H-4), 7.44 (2H, d, *J* = 8 Hz, H-3',5'), 7.32 (1H, d, *J*<sub>ortho</sub> = 8 Hz), 7.02 (1H, ddd, *J*<sub>para</sub> = 1.8 Hz, *J*<sub>ortho</sub> = 8 Hz, <sup>3</sup>*J*<sub>H-F</sub> = 11.4 Hz, H-6), 3.01 (1H, q, *J* = 7 Hz, aliphatic-CH), 1.30 (6H, d, *J* = 7.2 Hz, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (100 MHz DMSO-*d*<sub>6</sub>): δ 159.78 (C), 157.44 (C), 152.72 (C), 150.80 (C), 137.47 (1C, d, <sup>1</sup>*J*<sub>C-F</sub> = 402 Hz, C-F), 127.12 (C), 126.96 (2CH), 126.50 (2CH), 115.45 (CH-7), 110.17 (1C, d, <sup>2</sup>*J*<sub>C-F</sub> = 26 Hz, CH-6), 100.99 (1C, d, <sup>2</sup>*J*<sub>C-F</sub> = 25 Hz, CH-4), 33.29 (CH), 23.56 (2CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>16</sub>H<sub>15</sub>FN<sub>2</sub> 255.12920, found 255.12889 [M+H]<sup>+</sup>.

#### 3.3.2.2.5. Characterisation of derivative 5,6-difluoro-2-(4-isopropylphenyl)-1*H*-benzo[*d*]imidazole (9g)

Compound **9g** (C<sub>16</sub>H<sub>14</sub>F<sub>2</sub>N<sub>2</sub>, MW 272.293) was obtained in total yield 67%; m.p. 235-237 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.75. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.17 (d, 2H, *J* = 8.4 Hz, H-2',6'), 7.82 (t, 2H, *J* = 8.8 Hz, H-4,7), 7.54 (d, 2H, *J* = 8 Hz, H-3',5'), 3.01 (m, 1H, *J* = 6.8 Hz, CH), 1.26 (d, 6H, *J* = 6.8 Hz, 2CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 153.24 (C), 151.67 (C), 147.94 (2C, dd, <sup>1</sup>*J*<sub>C,F</sub> = 242 Hz, <sup>2</sup>*J*<sub>C,F</sub> = 16 Hz, 2CF), 130.26

(2C), 127.57 (2CH), 127.41 (2CH), 122.94 (C), 102.52 (2CH, m, CH-4,7), 33.48 (CH), 23.45 (2CH<sub>3</sub>). ESI-MS ( $m/z$ ): calcd. for C<sub>16</sub>H<sub>14</sub>F<sub>2</sub>N<sub>2</sub> 273.11978, found 273.11948 [M+H]<sup>+</sup>.

#### 3.3.2.2.6. Characterisation of derivative 2-(4-(*tert*-butyl)phenyl)-1*H*-benzo[*d*]imidazole (5h)

Compound **5h** (C<sub>17</sub>H<sub>18</sub>N<sub>2</sub>, MW 250.338) was obtained in total yield 65%; m.p. 239-241 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.50. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.11 (2H, d, J= 8.4 Hz, H-2',6'), 7.65 (1H, d, J= 7.2 Hz, H-4), 7.57 (2H, d, J= 8.4 Hz, H-3',5'), 7.52 (1H, d, J= 6.8 Hz, H-7), 7.19 (2H, m, H-5,6), 1.34 (9H, s, 3CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 152.53 (C), 151.24 (C), 143.82 (C), 134.92 (C), 127.41 (C), 126.19 (2CH), 125.70 (2CH), 122.29 (CH), 121.50 (CH), 118.69 (CH), 111.16 (CH), 34.56 (C), 30.96 (3CH<sub>3</sub>). ESI-MS ( $m/z$ ): calcd. for C<sub>17</sub>H<sub>18</sub>N<sub>2</sub> 251.15428, found 251.15436 [M+H]<sup>+</sup>.

#### 3.3.2.2.7. Characterisation of derivative 2-(4-(*tert*-butyl)phenyl)-5-chloro-1*H*-benzo[*d*]imidazole (6h)

Compound **6h** (C<sub>17</sub>H<sub>17</sub>ClN<sub>2</sub>, MW 284.783) was obtained in total yield 83%; m.p. 247-249 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.76. <sup>1</sup>H-NMR (acetone-*d*<sub>6</sub>): δ 8.17 (2H, d, J=8.4 Hz, H-2',6'), 7.62-7.59 (4H, m, H-4,7,3',5'), 7.23 (1H, dd, <sup>1</sup>J=8.6 Hz, <sup>2</sup>J=1.6 Hz, 2 Hz, H-6), 1.38 (9H, s, 3CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, acetone-*d*<sub>6</sub>): δ 154.37 (C), 153.81 (C), 141.61 (C), 139.11 (C), 128.10 (C), 128.07 (C), 127.40 (2CH), 126.75 (2CH),

126.35 (CH), 116.83 (CH), 115.80 (CH), 35.45 (C), 31.45 (3CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>17</sub>H<sub>17</sub>ClN<sub>2</sub> 285.11530, found 285.11569 [M+H]<sup>+</sup>.

**3.3.2.2.8. Characterisation of derivative 2-(4-(*tert*-butyl)phenyl)-5,6-dichloro-1*H*-benzo[*d*]imidazole (7h)**

Compound **7h** (C<sub>17</sub>H<sub>16</sub>Cl<sub>2</sub>N<sub>2</sub>, MW 319.228) was obtained in total yield 63%; m.p. 259-261 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.84. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.20 (d, 2H, J= 8.4 Hz, H-2',6'), 7.96 (s, 2H, H-4,7), 7.68 (d, 2H, J= 8.4 Hz, H-3',5'), 1.35 (s, 9H, 3CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 154.82 (C), 152.88 (C), 136.28 (2C), 127.12 (2CH), 126.15 (2CH), 125.82 (2C), 124.09 (C), 115.89 (2CH), 34.81 (C), 30.80 (3CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>17</sub>H<sub>16</sub>Cl<sub>2</sub>N<sub>2</sub> 319.07633, found 319.07657 [M+H]<sup>+</sup>.

**3.3.2.2.9. Characterisation of derivative 2-(4-(*tert*-butyl)phenyl)-5-fluoro-1*H*-benzo[*d*]imidazole (8h)**

Compound **8h** (C<sub>17</sub>H<sub>17</sub>FN<sub>2</sub>, MW 268.329) was obtained in total yield 84%; m.p. >300 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.72. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.09 (2H, d, J= 8 Hz, H-2',6'), 7.57 (3H, d, J= 8.4 Hz, H-4,3',5'), 7.38 (1H, d, J= 8.4 Hz, H-7), 7.05 (1H, m, H-6), 1.33 (9H, s, 3xCH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 159.83 (C), 157.50 (C), 153.22 (C), 152.60 (C), 137.22 (1C, d, <sup>1</sup>J<sub>C,F</sub>= 401 Hz, C-F), 126.44 (C), 126.28 (2CH), 125.87 (2CH), 115.50 (1C, d, <sup>3</sup>J<sub>C,F</sub>= 10 Hz, CH-7), 110.38 (1C, <sup>2</sup>J<sub>C,F</sub>= 25 Hz, CH-6), 100.95 (1C, <sup>2</sup>J<sub>C,F</sub>= 26 Hz, CH-4),



34.55 (C), 30.84 (3CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>17</sub>H<sub>17</sub>FN<sub>2</sub> 269.14485, found 269.14484 [M+H]<sup>+</sup>.

**3.3.2.2.10. Characterisation of derivative 2-(4-(*tert*-butyl)phenyl)-5,6-difluoro-1*H*-benzo[*d*]imidazole (9h)**

Compound **9h** (C<sub>17</sub>H<sub>16</sub>F<sub>2</sub>N<sub>2</sub>, MW 286.319) was obtained in total yield 78%; m.p. 257-259 °C; TLC (PS/EA 7/3): R<sub>f</sub> 0.78. <sup>1</sup>H-NMR (DMSO-*d*<sub>6</sub>): δ 8.14 (d, 2H, J= 8.4 Hz, H-2',6'), 7.78 (t, 2H, J= 8.8 Hz, H-4,7), 7.66 (d, 2H, J= 8.4 Hz, H-3',5'), 1.35 (s, 9H, 3CH<sub>3</sub>). <sup>13</sup>C-NMR (jmod, DMSO-*d*<sub>6</sub>): δ 154.65 (C-2), 152.25 (C-4'), 147.56 (2C, dd, <sup>1</sup>J<sub>C,F</sub>= 241 Hz, <sup>2</sup>J<sub>C,F</sub>= 16 Hz, 2CF), 131.75 (C-3a,7a), 126.86 (C-2',6'), 126.15 (C-3',5'), 123.97 (C-1'), 102.66 (2CH, dd, <sup>2</sup>J<sub>C,F</sub>= 15 Hz, CH-4,7), 34.80 (C), 30.81 (3CH<sub>3</sub>). ESI-MS (*m/z*): calcd. for C<sub>17</sub>H<sub>16</sub>F<sub>2</sub>N<sub>2</sub> 287.13543, found 287.13599 [M+H]<sup>+</sup>.

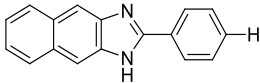
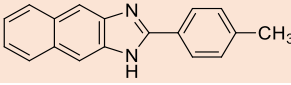
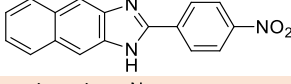
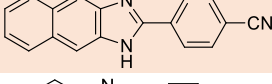
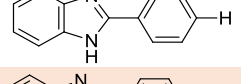
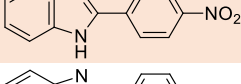
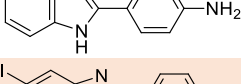
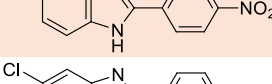
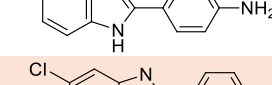
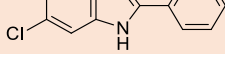
## 3.4. Results and discussion

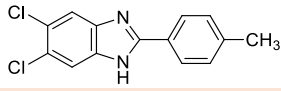
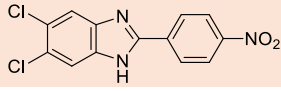
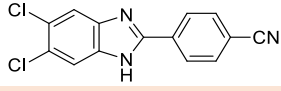
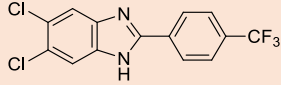
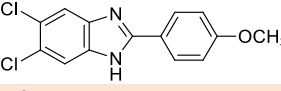
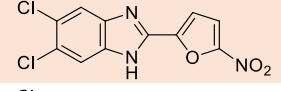
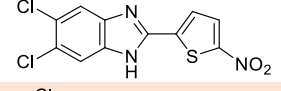
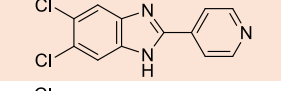
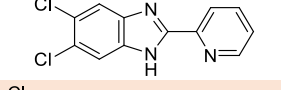
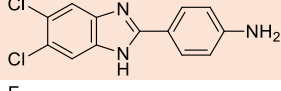
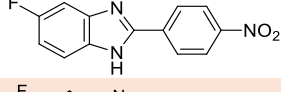
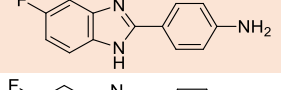
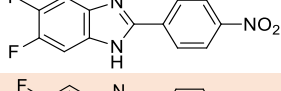
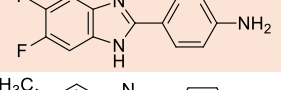
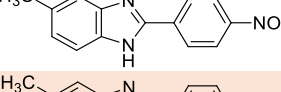
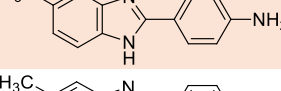
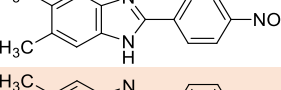
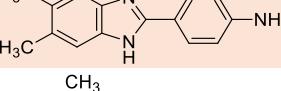
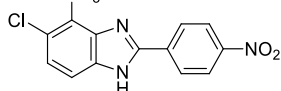
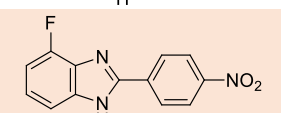
In the context of dissecting the antiviral activity of the benzimidazole moiety (as previously described) I designed and synthesised several variously substituted benzimidazole derivatives which were tested for cytotoxicity and antiviral activity in cell-based assays against several positive single-stranded RNA viruses: Bovine viral diarrhoea virus (BVDV), Yellow fever virus (YFV), Coxsackie B5 (CVB-5), and Poliovirus (Sb-1); two negative

single-stranded RNA viruses: Respiratory syncytial virus (RSV) and vesicular stomatitis virus (VSV); a double-stranded RNA virus: Reo-1; and two DNA viruses: vaccinia virus (VV) and Herpes simplex virus-1 (HSV-1).

By these assays, some derivatives have emerged as interesting inhibitors of virus infection in cell-based assay for CV-B5 and VV. A second series of derivatives were designed and synthesised introducing structural modification in order to have some preliminary structure-activity relationship (SAR) and to select the lead compound, the most effective compound for each virus.

*Table 5 Antiviral activity and cytotoxicity of benzimidazole derivatives (4a-d, 5a, 5-6c, 5-11m, 7a-f, 7i-l, 8-13c), mycophenolic acid was used as positive controls.*

Compound	R	Vero-76 <sup>a</sup> CC <sub>50</sub>	VV <sup>b</sup> EC <sub>50</sub>
4a		49	>49
4b		65	8.7±0.1
4c		>100	17.5±3.6
4d		>100	11.7±2.5
5a		>100	>100
5c		>100	14.4±3.7
5m		>100	48±4
6c		50	3.2±0.8
6m		97	>97
7a		100	53±5

7b		48	7.8±0.8
7c		>100	0.06±0.02
7d		40	0.9±0.1
7e		6.3	0.3±0.1
7f		8.9	>8.9
7i		27	>27
7j		>100	30±4
7k		>100	>100
7l		>100	>100
7m		16	15.4±2.3
8c		>100	14±0.4
8m		>100	>100
9c		>100	9.5±2.1
9m		>100	48.5±3.5
10c		>100	3.7±0.7
10m		>100	>100
11c		>100	1.0±0.2
11m		71	33±5
12c		>100	48.5±3.5
13c		>100	>100

<b>Mycophenolic acid</b>		<b>18</b>	<b>1.5</b>
<sup>a</sup> Compound concentration (μM) required to reduce the viability of mock-infected VERO-76 monolayers by 50%.			
<sup>b</sup> Compound concentration (μM) required to reduce the plaque number of VV by 50% in VERO-76 monolayers.			

Table 5. shows the antiviral activity of all synthesised compounds against VV and mycophenolic acid was used as control. To describe some SAR, these derivatives can be ranked into three groups according to their EC<sub>50</sub> as follow: ≤ 1 μM, 1-10 μM and 10-30 μM. Derivatives **7c**, **7d**, **7e**, **11c** all showed EC<sub>50</sub> values lower than 1 μM, but only **7c** and **11c** are not cytotoxic, while **7d** and **7e** showed CC<sub>50</sub> values of 40 and 6.3 μM. Derivatives most active **7c** and **11c** differ only for the substitution in positions 5 and 6 of the benzimidazole moiety, and both present a nitro-group in position 4'. In the second group of most active compounds, we can quote derivatives **6c**, **9c** and **10c**, not cytotoxic, all bringing a nitro-group in position 4', but different for substitution on benzimidazole scaffold. **6c** and **10c** are the corresponding mono-substituted derivatives of compounds **7c** and **11c** respectively, from which we can deduce that the substitution in position 5 and 6 enhanced the activity. In the third group can be counted derivatives **4c**, **4d**, **5c**, **7j** and **7m** (which also present a CC<sub>50</sub> value of 16 μM against EC<sub>50</sub> value of 15 μM).

The most active compound resulted 5,6-dichloro-2-(4-nitrophenyl)-1*H*-benzo[*d*]imidazole **7c**, whose anti-VV activity was already detected<sup>112</sup>, was selected for deeper investigation and in order to establish the mechanism

of action, the time-of-addition (TOA) assay was carried out. Time-of-addition assay<sup>113</sup> can spot the moment the compound has the maximum antiviral activity thus, knowing the virus life cycle, the virus replication step inhibited by the analysed compound can be identified. TOA results are depicted in Figure 28; the pretreatment of the cells with derivative **7c** was very subtle, but from this assay came out that derivative **7c** carried on its maximum activity during virus-entry phase.

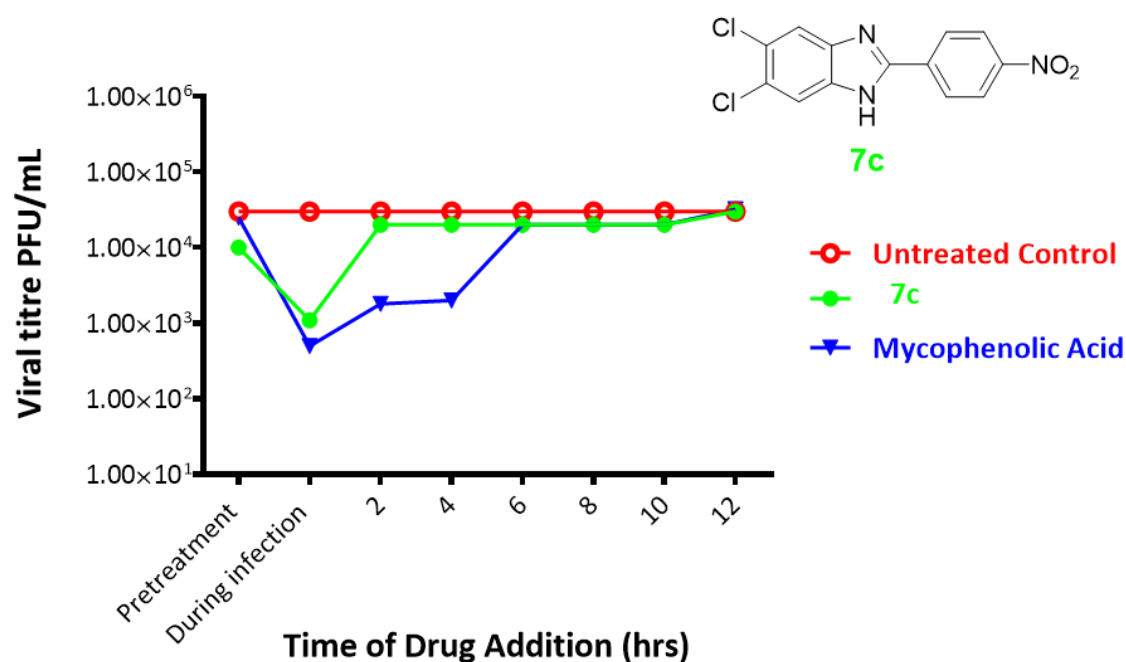
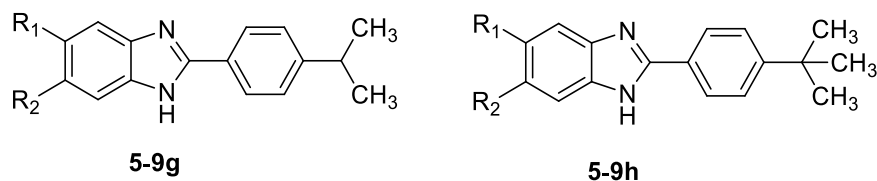


Figure 28 Effect of time of (drug)-addition on antiviral activity of derivative **7c** (green). The same test was performed using the reference compound Mycophenolic Acid (blue) for comparison. In red is reported the untreated control

Table 6 Antiviral activity and cytotoxicity of benzimidazole derivatives (**5-9g**, **5-9h**), NM 107 and pleconaril were used as positive controls



**5** R<sub>1</sub>=R<sub>2</sub>=H; **6** R<sub>1</sub>=Cl, R<sub>2</sub>=H; **7** R<sub>1</sub>=R<sub>2</sub>=Cl; **8** R<sub>1</sub>=F, R<sub>2</sub>=H; **9** R<sub>1</sub>=R<sub>2</sub>=F

Compd	Vero-76 <sup>a</sup> CC <sub>50</sub>	CVB-3 <sup>b</sup> EC <sub>50</sub>	CVB-4 <sup>c</sup> EC <sub>50</sub>	CVB-5 <sup>d</sup> EC <sub>50</sub>	EV-A71 <sup>e</sup> EC <sub>50</sub>	Hela <sup>f</sup> CC <sub>50</sub>	EV-D68 <sup>g</sup> EC <sub>50</sub>	LLC-MK2 <sup>h</sup> CC <sub>50</sub>	Echo9 <sup>i</sup> EC <sub>50</sub>
<b>5g</b>	60/24	>60	>60	>60	>24	9	>9	6	>6
<b>5h</b>	46/40	>46	>46	7	>40	9	>9	6.5	>6.5
<b>6g</b>	>100	17	>100	9.9	7	100	>100	>100	8
<b>6h</b>	>100	33	25.5	8.4	3.5	100	>100	12	4
<b>7g</b>	25.5/8	>25.5	>25.5	>25.5	>8	8	>8	7	>7
<b>7h</b>	9	>9	>9	1.5	>9	6.4	>6.4	12	>12
<b>8g</b>	>100	>100	>100	>100	>100	>100	>100	29.5	>29.5
<b>8h</b>	>100	>100	>100	>100	70	>100	>100	>100	>100
<b>9g</b>	>100	>100	>100	>100	31	>100	>100	100	>100
<b>9h</b>	>100	>100	>100	>100	6	>100	4	>100	>100
NM 107	>100	29			6				
pleconaril	>100		2	0.005		>100	0.3	>100	0.1

<sup>a,f,h</sup>Compound concentration (μM) required to reduce by 50% the viability of mock-infected Vero-76, Hela and LLC-MK2 cells, as determined by the MTT method after 3, 4 and 5 days, respectively.

<sup>b,c,d,i</sup>Compound concentration (μM) required to reduce by 50% the plaque number of CV-B3, CV-B4, CV-B5 and Echo9 in Vero-76 and LLC-MK2 cells, respectively.

<sup>e</sup>Compound concentration (μM) required to achieve 50% protection of Vero-76 cells from EV-A71-induced cytopathogenicity, as determined by the MTT method.

<sup>g</sup>Compound concentration (μM) required to achieve 50% protection of Hela cells from EV-D68-induced cytopathogenicity, as determined by the MTT method.

Some of the listed compounds (**6g**, **6h**, **7g**, **7h**) in Table 6 turned out as active against CV-B5 in the first-round screening thereby derivatives **5g-h**, **8g-h**, **9g-h** were designed and synthesised in order to dissect the role of R<sub>1</sub> and R<sub>2</sub> substituent on benzimidazole scaffold. All obtained derivatives were

subjected to a wide antivirals screening on different Enteroviruses as CVB-3, CVB-4, EV-A71, EV-D68 and Echovirus 9 (Echo9); NM 107 and pleconaril were used as reference compounds.

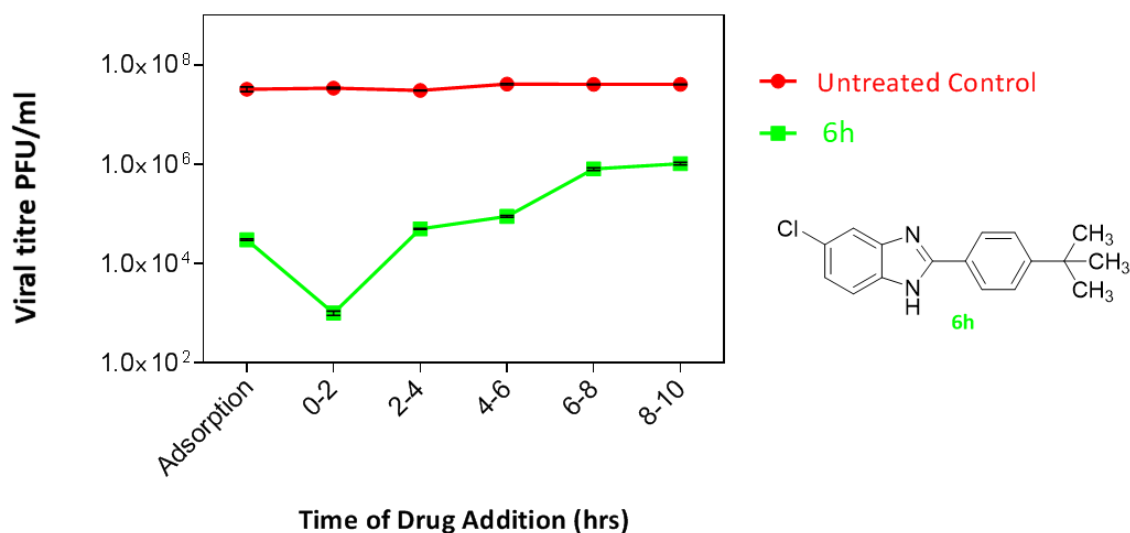


Figure 29 Time-of-addition for derivative 6h

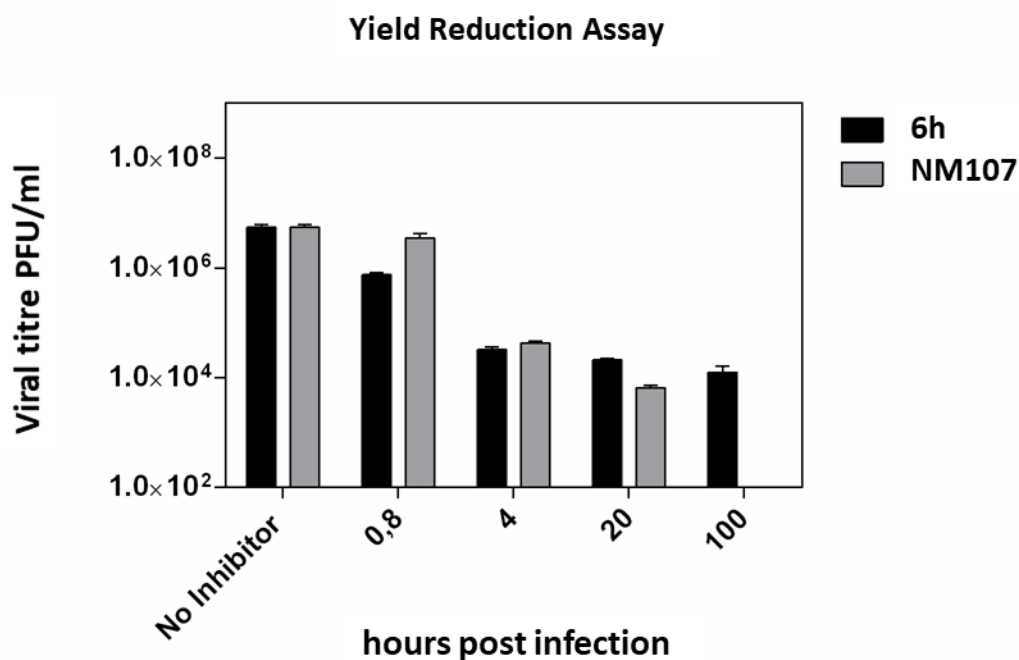
Derivative 6g was widely active on different viruses (CVB-3, CVB-5, EV-A71 and Echo9), while its tertbutyl-homologue (6h) was the most active against EV-A71 and together more selective, showing higher  $EC_{50}$  values against other EVs. Derivatives 7g and 7h turned out as the most cytotoxic, so the di-chlorine derivatives are less effective because of their high toxicity. Mono-fluorine derivatives (8g and 8h) are either not toxic or active. Among di-fluorine compounds (9g and 9h) the tertbutyl derivative 9h was found the most active against EV-D68 ( $EC_{50}$  4  $\mu$ M) and active against EV-A71 ( $EC_{50}$  6  $\mu$ M).

The most active compound 2-(4-(*tert*-butyl)phenyl)-5-chloro-1*H*-benzo[*d*]imidazole **6h** was selected for TOA assay in order to identify the virus cycle step inhibited by tested derivative.

Time of addition assay (Figure 29) was performed in EV-A71-infected Vero-76 cells (m.o.i. = 1) exposed to the compound (20  $\mu$ M) at different times of infection. Interestingly, TOA results show that **6h** compound wield its greatest effect when added with the virus. Moderate titre reduction is detectable until 2 hours post-infection. However, there was weak reduction in virus yield if compared to untreated control, even when the compound was added at the late stages of the growth cycle. This apparent inhibition indicated that **6h** could not easily be removed from cells by washing.

In order to deeply assess the mechanism of action of these class of benzimidazole derivatives, the most interesting (**6h**) was selected for deeper experiments to investigate the viral replication cycle step inhibited. 2-(4-(*tert*-butyl)phenyl)-5-chloro-1*H*-benzo[*d*]imidazole (**6h**) resulted in the most interesting since active in cell-based assay against EV-A71, a clinically relevant virus. Yield reduction assay (Figure 30) and virucidal activity assay (Figure 31) were planned and performed by collaborator Dr. G. Sanna, to measure the reduction of virus titre in presence of active compound and to assess the effect of **6h** on viral infectivity, respectively.





*Figure 30 Yield of infectious EV-A71 viruses produced in infected Vero 76 cells treated with **6h** and NM 107.*

The yield reduction assay investigated the reduction of virus titre in the presence of the active compound during a single round of viral infection. Not cytotoxic concentrations of 100, 20, 4 and 0.8  $\mu\text{M}$  were used and a dose-dependent reduction of titre was observed. Significant reduction of EV-A71 titre was detected also a low concentration (Figure 30). The potential virucidal activity of **6h** was investigated by virucidal activity assay (Figure 31) incubating EV-A71 solution ( $1 \times 10^5$  PFU/mL) at either 4 and 37  $^{\circ}\text{C}$  for 1 hour with 20  $\mu\text{M}$  compound concentration. Then the samples were titrated at high dilutions, at which **6h** was not active. No significant differences between the titre of EV-A71 treated at the two different temperatures was observed, as shown in Figure 30. Compound **6h** failed to

affect the EV-A71 infectivity, suggesting that the inhibition of EV-A71 replication observed in cell-based assays can be completely due to an interference of compound with the viral life cycle.

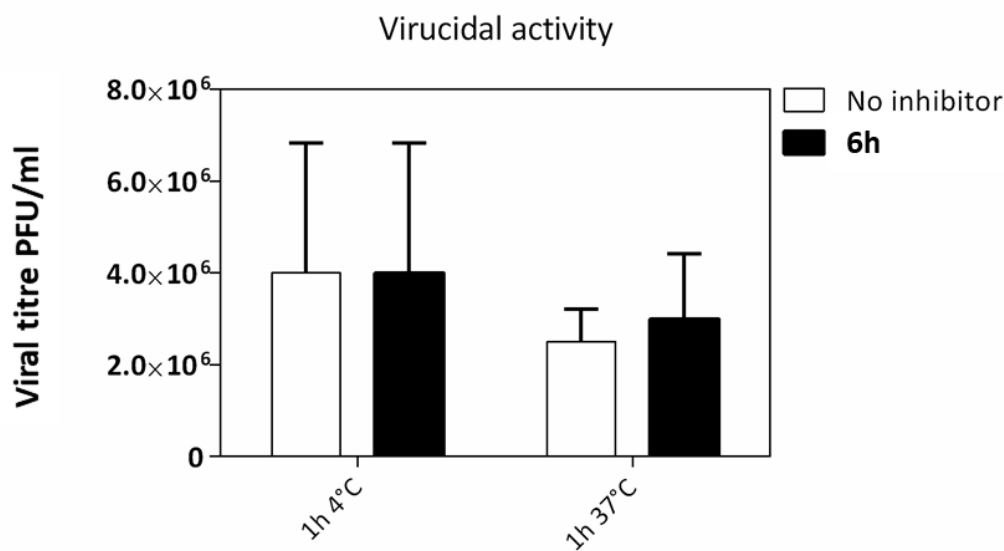


Figure 31 2. Virucidal effect (expressed as plaque-forming units (PFU)/mL) of 1509 (20  $\mu$ M) against EV-A71 virions at either 4 °C or 37 °C) for 1 hour. Dark columns, viral titer for viral and 1509 solution; Grey columns for the viral titer of not treated solution. The results presented were obtained from two independent experiments. Data are mean  $\pm$ SD.

With purpose of verifying the idea suggested by the TOA results, that compound **6h** is able to inhibit the virus entry, two more experiments were planned and carried out: adsorption assay and penetration assay. The kinetic of virus adsorption in the presence of **6h** was also evaluated. Low-temperature treatment allows binding of viruses to the cell surface receptors but prevents the internalization of virus particles into the cells. Accordingly, Vero-76 cells were incubated with EV-A71 (m.o.i. = 0.1) and **6h** for 2 hours

at 4 °C, using compound concentration of 20  $\mu$ M. The treatment with **6h** resulted in detectable reduction of the virus titre in comparison to untreated infected control. Adsorption assay has been plotted with TOA in the same diagram (Figure 32) to highlight that **6h** inhibits EV-A71 to entry the cell.

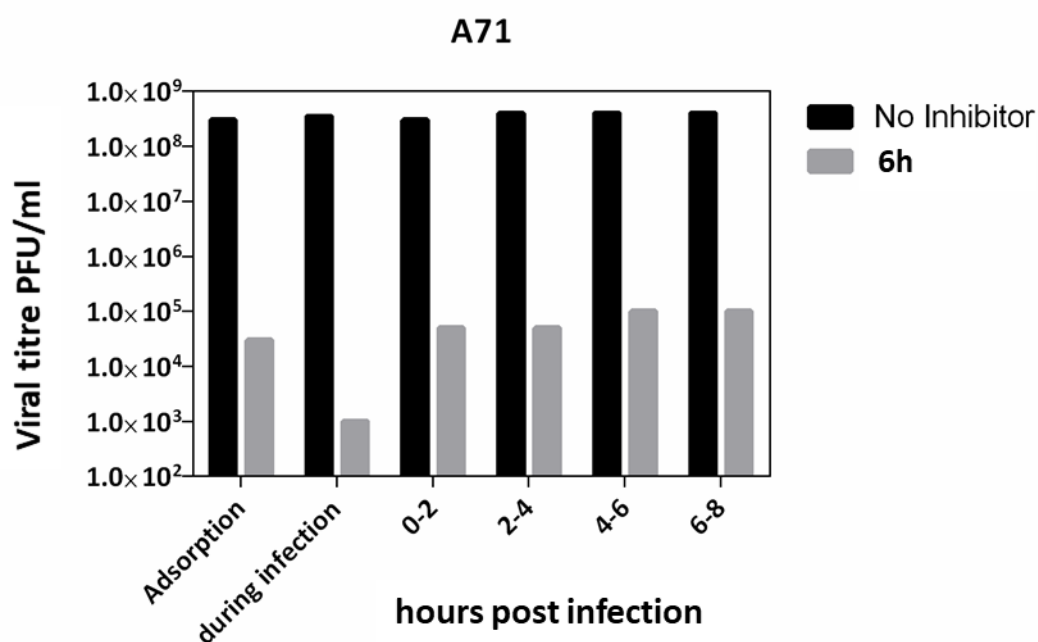


Figure 32 Adsorption Assay and Time course experiment (TOA).

To scan the mechanism of inhibition of EV-A71 entry into Vero cells by **6h** a penetration assay was carried out. Vero-76 cells were pre-chilled at 4 °C for 1 hour and then incubated with EV-A71 for 2 hours on ice. After incubation period, increasing concentrations of compound were supplemented, and to trigger the penetration of the virus *via* endocytosis, the monolayers were incubated at 37 °C under 5% CO<sub>2</sub> for 1 hour. Viral

particles were then inactivated with buffer treatment, the cells were washed and incubated at 37 °C for 4 days. Percentage of inhibition was determined. As shown in Figure 33, it emerged that **6h** could effectively inhibit the viral penetration in a dose-dependent manner.

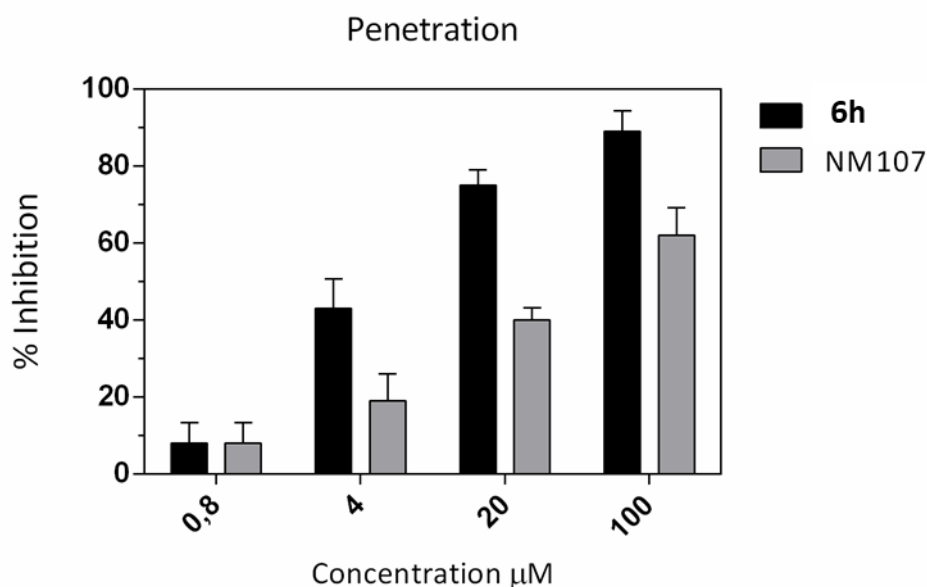
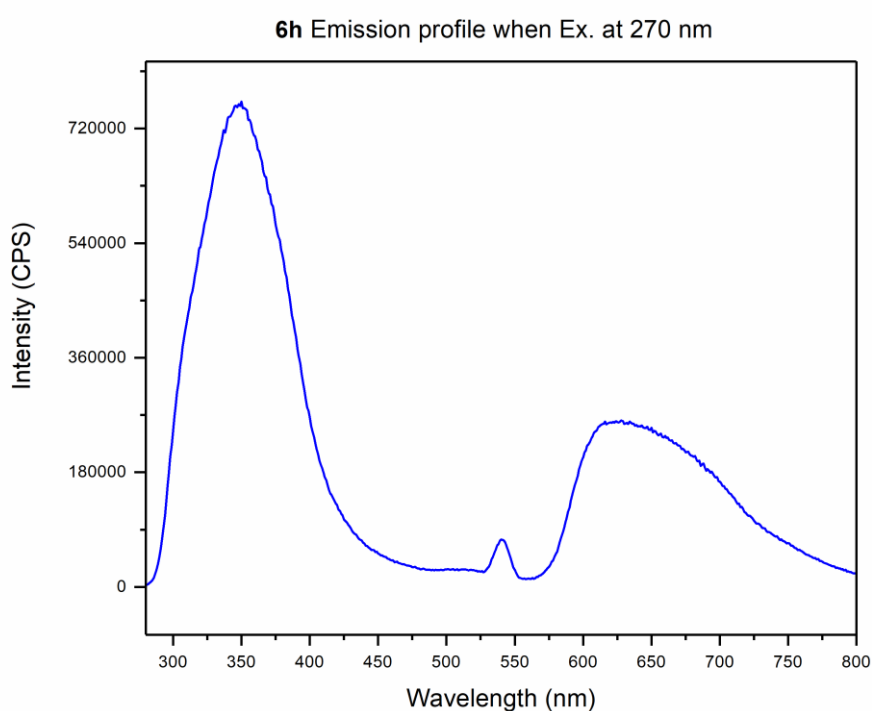


Figure 33 Penetration Assay. Dose-dependent of **6h** (Dark columns) inhibition of EV-A71 penetration. NM 107 (2'-C-Methyl-Cytidine), Grey columns was used as reference. This is a representative result (mean  $\pm$ SD) of three independent experiments.

In order to detect the target involved in the early stage of virus infection of a cell which is bound and inhibited by the synthesised benzimidazole derivative **6h**, the following experiment was planned: fluorescent detection of the compound during cell treatment with the compound at non-cytotoxic concentration.

Due to the small dimension of the molecule, binding a fluorescent tag on it could affect the activity so, considering the chemical structure of

derivative **6h**, I came up with the idea of measuring its own fluorescent properties in order to exploit its eventual fluorescence. With the help of Dr S. Mura and Prof. L. Malfatti at Department of Chemistry and Pharmacy, Laboratory of Materials Science and Nanotechnology, CR-INSTM (Via Vienna 2, 07100, Sassari, Italy) the following fluorescent measurement were carried out. It resulted that when **6h** is excited at 270 nm, emit at two different wavelengths, at about 350 nm and 625 nm.



*Figure 34 Emission profile for derivative **6h** when excited at 270 nm.*

Several linear dilutions in D-MEM (Minimum Essential Medium) of derivative **6h** were carried out, and fluorescent emission was measured. D-MEM was used as solvent in order to detect any

fluorescent emission from the cell media during the next planned experiments. As depicted in Figure 35, emissions at about 350 nm are plotted for all the analysed dilutions of compound.

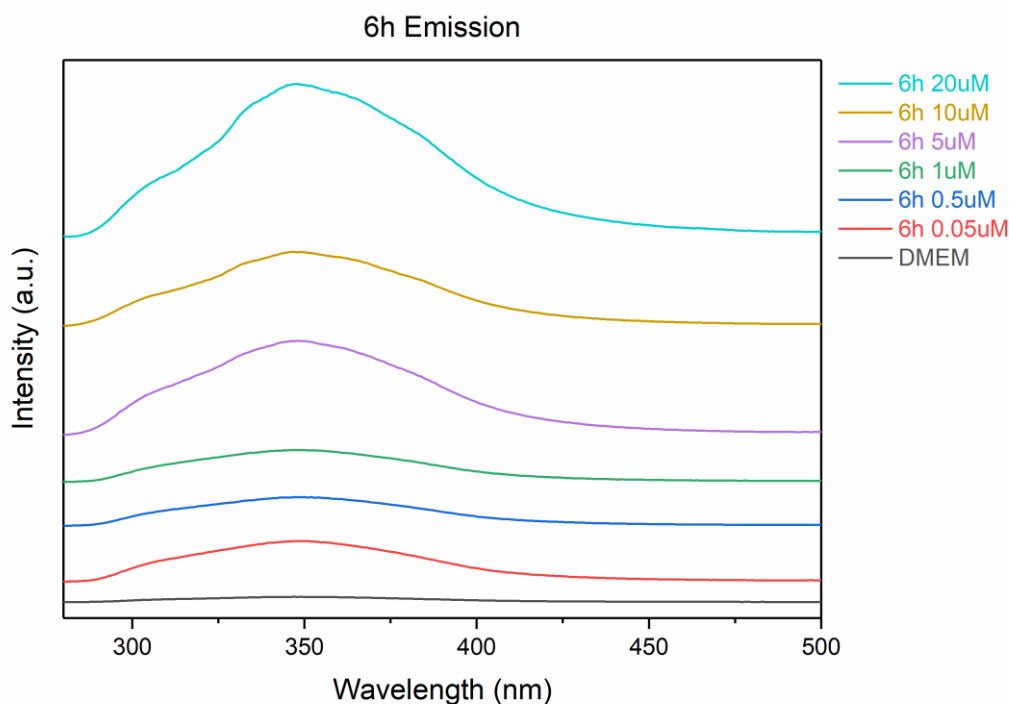


Figure 35 Emission profile (300-500 nm) for derivative **6h** in D-MEM.

With these evidences, the detection of the (eventual) binding of the lead compound on the cell surface of Vero-76 cells were planned with Prof. L. Bagella and collaborators at the Department of Biomedical Science, University of Sassari, Italy. The experiment is performed by treating the cells with concentrations of **6h** from 1 to 20  $\mu\text{M}$  and detected at selected time points with a fluorescent microscope.

### 3.5. Conclusions

From wide antiviral assays of synthesised several small benzimidazole derivatives, turned out interesting  $EC_{50}$  for some of them. SARs analysis have been made and new molecules were synthesised to obtain the most selective, active and less toxic derivative against, particularly, enteroviruses which are clinically relevant. Deeper investigations were planned and executed in order to better understand the mechanism of action and the target of the most interesting compound: **6h**. A series of specific assays, such as TOA, YRA, VAA absorption and penetration assays, suggested the idea that this molecule, inhibiting the virus entry of the cell, can bind a host cell-surface receptor, so a fluorescence investigation of the interaction between the molecule and the cell is planned and is, unfortunately, still in progress.

# Appendix

Primer	Sequence
1_F	5'-CGCTTACGACtgtTCTCTAGCCAAC-3'
1_R	5'-ACATCTGGTTCGAACTCG-3'
2_F	5'-TCGGAAGTTTtgcGTTGGTTCCC-3'
2_R	5'-TCAAATGGCAGTGTCCGC-3'
3_F	5'-TCCCAAGGATtgcTCACGTTCCC-3'
3_R	5'-TTGTGGACAATGTCCAAC-3'
4_F	5'-GTTTGGCGTTtgtTCCCGTGATG-3'
4_R	5'-TTCCGATCAAATGGCAGTG-3'
5_F	5'-TGGCGTTGGTtgcCGTGATGTGA-3'
5_R	5'-AACTTCCGATCAAATGGCAGTGTC-3'
6_F	5'-GAGTCTCTGTCCGTCAATGG-3'
6_R	5'-AGAGGGGAAAGCGGCTTT-3'
7_F	5'-ATTTCGGATCTCCCACAG-3'
7_R	5'-GTTCTTGTCTTCGGGGAAAATG-3'
<i>Ct</i> UGGT_pHLsec_F	GGTTGCGTAGCTGAAACCGGTCAAGTC GCAGCCTCTCCA
<i>Ct</i> UGGT_pHLsec_R	GATGGTGGTGCTTGGTACCCTCCCGAA CCGTCTTGAC
T7_F	TAA TAC GAC TCA CTA TAG GG
T7_R	GCT AGT TAT TGC TCA GCG G



## References

- (1) de Clercq, E. Milestones in the Discovery of Antiviral Agents: Nucleosides and Nucleotides. *Acta Pharm. Sin. B* **2012**, *2* (6), 535–548. <https://doi.org/10.1016/j.apsb.2012.10.001>.
- (2) Ellis, J. A.; West, K. H.; Cortese, V. S.; Myers, S. L.; Carman, S.; Martin, K. M.; Haines, D. M. Lesions and Distribution of Viral Antigen Following an Experimental Infection of Young Seronegative Calves with Virulent Bovine Virus Diarrhea Virus-Type II. *Can. J. Vet. Res.* **1998**, *62* (3), 161–169.
- (3) Baginski, S. G.; Pevear, D. C.; Seipel, M.; Sun, S. C.; Benetatos, C. A.; Chunduru, S. K.; Rice, C. M.; Collett, M. S. Mechanism of Action of a Pestivirus Antiviral Compound. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (14), 7981–7986. <https://doi.org/10.1073/pnas.140220397>.
- (4) Harasawa, R.; Mizusawa, H. Demonstration and Genotyping of Pestivirus RNA from Mammalian Cell Lines. *Microbiol. Immunol.* **1995**, *39* (12), 979–985. <https://doi.org/10.1111/j.1348-0421.1995.tb03301.x>.
- (5) Harasawa, R.; Sasaki, T. Sequence Analysis of the 5' Untranslated Region of Pestivirus RNA Demonstrated in Interferons for Human Use. *Biologicals* **1995**, *23* (4), 263–269. <https://doi.org/10.1006/biol.1995.0044>.

- (6) Harasawa, R.; Tomiyama, T. Evidence of Pestivirus RNA in Human Virus Vaccines. *J. Clin. Microbiol.* **1994**, *32* (6), 1604–1605.
- (7) King, R. W.; Scarnati, H. T.; Priestley, E. S.; De Lucca, I.; Bansal, A.; Williams, J. K. Selection of a Thiazole Urea-Resistant Variant of Bovine Viral Diarrhoea Virus That Maps to the RNA-Dependent RNA Polymerase. *Antivir. Chem. Chemother.* **2002**, *13* (5), 315–323.  
<https://doi.org/10.1177/095632020201300507>.
- (8) Sun, J.-H.; Lemm, J. A.; O’Boyle, D. R.; Racela, J.; Colonno, R.; Gao, M. Specific Inhibition of Bovine Viral Diarrhea Virus Replicase. *J. Virol.* **2003**, *77* (12), 6753–6760.  
<https://doi.org/10.1128/jvi.77.12.6753-6760.2003>.
- (9) Tabarrini, O.; Manfroni, G.; Fravolini, A.; Cecchetti, V.; Sabatini, S.; De Clercq, E.; Rozenski, J.; Canard, B.; Dutartre, H.; Paeshuyse, J.; et al. Synthesis and Anti-BVDV Activity of Acridones as New Potential Antiviral Agents. *J. Med. Chem.* **2006**, *49* (8), 2621–2627.  
<https://doi.org/10.1021/jm051250z>.
- (10) Bukhtiyarova, M.; Rizzo, C. J.; Kettner, C. A.; Korant, B. D.; Scarnati, H. T.; King, R. W. Inhibition of the Bovine Viral Diarrhoea Virus NS3 Serine Protease by a Boron-Modified Peptidyl Mimetic of Its Natural Substrate. *Antivir. Chem. Chemother.* **2001**, *12* (6), 367–373.  
<https://doi.org/10.1177/095632020101200607>.
- (11) Branza-Nichita, N.; Durantel, D.; Carrouée-Durantel, S.; Dwek, R. A.;

- Zitzmann, N. Antiviral Effect of N-Butyldeoxynojirimycin against Bovine Viral Diarrhea Virus Correlates with Misfolding of E2 Envelope Proteins and Impairment of Their Association into E1-E2 Heterodimers. *J. Virol.* **2001**, *75* (8), 3527–3536. <https://doi.org/10.1128/JVI.75.8.3527-3536.2001>.
- (12) Durantel, D.; Branza-Nichita, N.; Carrouée-Durantel, S.; Butters, T. D.; Dwek, R. A.; Zitzmann, N. Study of the Mechanism of Antiviral Action of Iminosugar Derivatives against Bovine Viral Diarrhea Virus. *J. Virol.* **2001**, *75* (19), 8987–8998. <https://doi.org/10.1128/JVI.75.19.8987-8998.2001>.
- (13) Zitzmann, N.; Mehta, A. S.; Carrouée, S.; Butters, T. D.; Platt, F. M.; McCauley, J.; Blumberg, B. S.; Dwek, R. A.; Block, T. M. Imino Sugars Inhibit the Formation and Secretion of Bovine Viral Diarrhea Virus, a Pestivirus Model of Hepatitis C Virus: Implications for the Development of Broad Spectrum Anti-Hepatitis Virus Agents. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (21), 11878–11882. <https://doi.org/10.1073/pnas.96.21.11878>.
- (14) Markland, W.; McQuaid, T. J.; Jain, J.; Kwong, A. D. Broad-Spectrum Antiviral Activity of the IMP Dehydrogenase Inhibitor VX-497: A Comparison with Ribavirin and Demonstration of Antiviral Additivity with Alpha Interferon. *Antimicrob. Agents Chemother.* **2000**, *44* (4), 859–866. <https://doi.org/10.1128/aac.44.4.859-866.2000>.

- (15) Vitale, G.; Corona, P.; Loriga, M.; Carta, A.; Paglietti, G.; Giliberti, G.; Sanna, G.; Farci, P.; Marongiu, M. E.; La Colla, P. 5-Acetyl-2-Arylbenzimidazoles as Antiviral Agents. Part 4. *Eur. J. Med. Chem.* **2012**, *53*, 83–97. <https://doi.org/10.1016/j.ejmech.2012.03.038>.
- (16) Carta, A.; Loriga, M.; Paglietti, G.; Ferrone, M.; Fermeglia, M.; Prich, S.; Sanna, T.; Ibba, C.; La Colla, P.; Loddo, R. Design, Synthesis, and Preliminary in Vitro and in Silico Antiviral Activity of [4,7]Phenantrolines and 1-Oxo-1,4-Dihydro-[4,7]Phenantrolines against Single-Stranded Positive-Sense RNA Genome Viruses. *Bioorganic Med. Chem.* **2007**, *15* (5), 1914–1927. <https://doi.org/10.1016/j.bmc.2007.01.005>.
- (17) Carta, A.; Briguglio, I.; Piras, S.; Corona, P.; Boatto, G.; Nieddu, M.; Giunchedi, P.; Marongiu, M. E.; Giliberti, G.; Iuliano, F.; et al. Quinoline Tricyclic Derivatives. Design, Synthesis and Evaluation of the Antiviral Activity of Three New Classes of RNA-Dependent RNA Polymerase Inhibitors. *Bioorg. Med. Chem.* **2011**, *19* (23), 7070–7084. <https://doi.org/10.1016/j.bmc.2011.10.009>.
- (18) Asthana, S.; Shukla, S.; Ruggerone, P.; Vargiu, A. V. Molecular Mechanism of Viral Resistance to a Potent Non-Nucleoside Inhibitor Unveiled by Molecular Simulations. *Biochemistry* **2014**, *53* (44), 6941–6953. <https://doi.org/10.1021/bi500490z>.
- (19) Asthana, S.; Shukla, S.; Vargiu, A. V.; Ceccarelli, M.; Ruggerone, P.;

- Paglietti, G.; Marongiu, M. E.; Blois, S.; Giliberti, G.; La Colla, P. Different Molecular Mechanisms of Inhibition of Bovine Viral Diarrhea Virus and Hepatitis C Virus RNA-Dependent RNA Polymerases by a Novel Benzimidazole. *Biochemistry* **2013**, *52* (21), 3752–3764. <https://doi.org/10.1021/bi400107h>.
- (20) Carta, A.; Briguglio, I.; Piras, S.; Corona, P.; Ibba, R.; Laurini, E.; Fermeglia, M.; Pricl, S.; Desideri, N.; Atzori, E. M.; et al. A Combined in Silico/in Vitro Approach Unveils Common Molecular Requirements for Efficient BVDV RdRp Binding of Linear Aromatic N-Polycyclic Systems. *Eur. J. Med. Chem.* **2016**, *117*, 321–334. <https://doi.org/10.1016/j.ejmech.2016.03.080>.
- (21) Ibba, R.; Corona, P.; Carta, A.; Giunchedi, P.; Loddo, R.; Sanna, G.; Delogu, I.; Piras, S. Antiviral Activities of 5-Chlorobenzotriazole Derivatives. *Monatshefte für Chemie* **2018**, *149* (7), 1247–1256. <https://doi.org/10.1007/s00706-018-2234-7>.
- (22) Chai, H.-H.; Lim, D.; Chai, H.-Y.; Jung, E. Molecular Modeling of Small Molecules as BVDV NS5B. *RdRp Inhib. Bull. Korean Chem. Soc* **2013**, *34* (3), 837. <https://doi.org/10.5012/bkcs.2013.34.3.837>.
- (23) Bahrami, K.; Khodaei, M. M.; Kavianinia, I. A Simple and Efficient One-Pot Synthesis of 2-Substituted Benzimidazoles. *Synthesis (Stuttg)*. **2007**, No. 4, 547–550. <https://doi.org/10.1055/s-2007-965878>.

- (24) Pauwels, R.; Balzarini, J.; Baba, M.; Snoeck, R.; Schols, D.; Herdewijn, P.; Desmyter, J.; De Clercq, E. Rapid and Automated Tetrazolium-Based Colorimetric Assay for the Detection of Anti-HIV Compounds. *J. Virol. Methods* **1988**, *20* (4), 309–321.
- (25) Martinez, J. P.; Sasse, F.; Brönstrup, M.; Diez, J.; Meyerhans, A. Antiviral Drug Discovery: Broad-Spectrum Drugs from Nature. *Nat. Prod. Rep.* **2015**, *32* (1), 29–48. <https://doi.org/10.1039/c4np00085d>.
- (26) Pieren, M.; Galli, C.; Denzel, A.; Molinari, M. The Use of Calnexin and Calreticulin by Cellular and Viral Glycoproteins. *J. Biol. Chem.* **2005**, *280* (31), 28265–28271. <https://doi.org/10.1074/jbc.M501020200>.
- (27) Baumann, O.; Walz, B. Endoplasmic Reticulum of Animal Cells and Its Organization into Structural and Functional Domains. *Int. Rev. Cytol.* **2001**, *205*, 149–214. [https://doi.org/10.1016/s0074-7696\(01\)05004-5](https://doi.org/10.1016/s0074-7696(01)05004-5).
- (28) Braakman, I.; Hebert, D. N. Protein Folding in the Endoplasmic Reticulum. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (5), a013201–a013201. <https://doi.org/10.1101/cshperspect.a013201>.
- (29) Fagone, P.; Jackowski, S. Membrane Phospholipid Synthesis and Endoplasmic Reticulum Function. *J. Lipid Res.* **2009**, *50* (Supplement), S311–S316. <https://doi.org/10.1194/jlr.R800049-JLR200>.

- (30) Clapham, D. E. Calcium Signaling. *Cell* **2007**, *131* (6), 1047–1058.  
<https://doi.org/10.1016/j.cell.2007.11.028>.
- (31) Friedman, J. R.; Voeltz, G. K. The ER in 3D: A Multifunctional Dynamic Membrane Network. *Trends Cell Biol.* **2011**, *21* (12), 709–717. <https://doi.org/10.1016/j.tcb.2011.07.004>.
- (32) Voeltz, G. K.; Rolls, M. M.; Rapoport, T. A. Structural Organization of the Endoplasmic Reticulum. *EMBO Rep.* **2002**, *3* (10), 944–950.  
<https://doi.org/10.1093/embo-reports/kvf202>.
- (33) Schwarz, D. S.; Blower, M. D. The Endoplasmic Reticulum: Structure, Function and Response to Cellular Signaling. *Cell. Mol. Life Sci.* **2016**, *73* (1), 79–94. <https://doi.org/10.1007/s00018-015-2052-6>.
- (34) Inoue, T.; Tsai, B. How Viruses Use the Endoplasmic Reticulum for Entry, Replication, and Assembly. *Cold Spring Harbor Perspectives in Biology*. January 2013. <https://doi.org/10.1101/cshperspect.a013250>.
- (35) L.E. Fellows, G.C. Kite, R.J. Nash, M.S.J. Simmonds, A. M. S. Distribution and Biological Activity of Alkaloidal Glycosidase Inhibitors from Plants. *Nitrogen Metab. Plants, Oxford Univ. Press* **1992**, 271–282.
- (36) Winchester, B.; Fleet, G. W. J. Amino-Sugar Glycosidase Inhibitors: Versatile Tools for Glycobiologists. *Glycobiology*. June 1992, pp 199–210. <https://doi.org/10.1093/glycob/2.3.199>.

- (37) Fischer, P. B.; Collin, M.; Karlsson, G. B.; James, W.; Butters, T. D.; Davis, S. J.; Gordon, S.; Dwek, R. A.; Platt, F. M. The Alpha-Glucosidase Inhibitor N-Butyldeoxynojirimycin Inhibits Human Immunodeficiency Virus Entry at the Level of Post-CD4 Binding. *J. Virol.* **1995**, *69* (9), 5791–5797.
- (38) Karlsson, G. B.; Butters, T. D.; Dwek, R. A.; Platt, F. M. Effects of the Imino Sugar N-Butyldeoxynojirimycin on the N-Glycosylation of Recombinant Gp120. *J. Biol. Chem.* **1993**, *268* (1), 570–576.
- (39) Wu, S.-F.; Lee, C.-J.; Liao, C.-L.; Dwek, R. A.; Zitzmann, N.; Lin, Y.-L. Antiviral Effects of an Iminosugar Derivative on Flavivirus Infections. *J. Virol.* **2002**, *76* (8), 3596–3604. <https://doi.org/10.1128/jvi.76.8.3596-3604.2002>.
- (40) Alonzi, D. S.; Scott, K. A.; Dwek, R. A.; Zitzmann, N. Iminosugar Antivirals: The Therapeutic Sweet Spot. *Biochem. Soc. Trans.* **2017**, *45* (2), 571–582. <https://doi.org/10.1042/BST20160182>.
- (41) Schallus, T.; Jaekel, C.; Fehér, K.; Palma, A. S.; Liu, Y.; Simpson, J. C.; Mackeen, M.; Stier, G.; Gibson, T. J.; Feizi, T.; et al. Malectin: A Novel Carbohydrate-Binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N-Glycosylation. *Mol. Biol. Cell* **2008**, *19* (8), 3404–3414. <https://doi.org/10.1091/mbc.E08-04-0354>.
- (42) Huang, P. N.; Jheng, J. R.; Arnold, J. J.; Wang, J. R.; Cameron, C. E.;



- Shih, S. R. UGGT1 Enhances Enterovirus 71 Pathogenicity by Promoting Viral RNA Synthesis and Viral Replication. *PLoS Pathog.* **2017**, *13* (5). <https://doi.org/10.1371/journal.ppat.1006375>.
- (43) Tax, G.; Lia, A.; Santino, A.; Roversi, P. Modulation of ERQC and ERAD: A Broad-Spectrum Spanner in the Works of Cancer Cells? *J. Oncol.* **2019**, *2019*, 1–14. <https://doi.org/10.1155/2019/8384913>.
- (44) Trombetta, S. E.; Gañan, S. A.; Parodi, A. J. The UDP-Glc:Glycoprotein Glucosyltransferase Is a Soluble Protein of the Endoplasmic Reticulum. *Glycobiology* **1991**, *1* (2), 155–161. <https://doi.org/10.1093/glycob/1.2.155>.
- (45) Roversi, P.; Marti, L.; Caputo, A. T.; Alonzi, D. S.; Hill, J. C.; Dent, K. C.; Kumar, A.; Levasseur, M. D.; Lia, A.; Waksman, T.; et al. Interdomain Conformational Flexibility Underpins the Activity of UGGT, the Eukaryotic Glycoprotein Secretion Checkpoint. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (32), 8544–8549. <https://doi.org/10.1073/pnas.1703682114>.
- (46) Guerin, M.; Parodi, A. J. The UDP-Glucose:Glycoprotein Glucosyltransferase Is Organized in at Least Two Tightly Bound Domains from Yeast to Mammals. *J. Biol. Chem.* **2003**, *278* (23), 20540–20546. <https://doi.org/10.1074/jbc.M300891200>.
- (47) Choudhury, P., Liu, Y., Bick, R. J., and Sifers, R. N. Intracellular Association between UDP-Glucose:Glycoprotein Glucosyltransferase

- and an Incompletely Folded Variant of Alpha1-Antitrypsin. *J. Biol. Chem.* **1997**, *272*, 13446–13451.
- (48) Tessier, D. C.; Dignard, D.; Zapun, A.; Radominska-Pandya, A.; Parodi, A. J.; Bergeron, J. J. M.; Thomas, D. Y. Cloning and Characterization of Mammalian UDP-Glucose Glycoprotein: Glucosyltransferase and the Development of a Specific Substrate for This Enzyme. *Glycobiology* **2000**, *10* (4), 403–412. <https://doi.org/10.1093/glycob/10.4.403>.
- (49) Izumi, M.; Makimura, Y.; Dedola, S.; Seko, A.; Kanamori, A.; Sakono, M.; Ito, Y.; Kajihara, Y. Chemical Synthesis of Intentionally Misfolded Homogeneous Glycoprotein: A Unique Approach for the Study of Glycoprotein Quality Control. *J. Am. Chem. Soc.* **2012**, *134* (17), 7238–7241. <https://doi.org/10.1021/ja3013177>.
- (50) Sakono, M.; Seko, A.; Takeda, Y.; Hachisu, M.; Ito, Y. Biophysical Properties of UDP-Glucose: Glycoprotein Glucosyltransferase, a Folding Sensor Enzyme in the ER, Delineated by Synthetic Probes. *Biochem. Biophys. Res. Commun.* **2012**, *426* (4), 504–510. <https://doi.org/10.1016/j.bbrc.2012.08.112>.
- (51) Trombetta, E. S.; Helenius, A. Glycoprotein Reglucosylation and Nucleotide Sugar Utilization in the Secretory Pathway: Identification of a Nucleoside Diphosphatase in the Endoplasmic Reticulum. *EMBO J.* **1999**, *18* (12), 3282–3292.

<https://doi.org/10.1093/emboj/18.12.3282>.

- (52) Sevvana, M.; Biadene, M.; Ma, Q.; Guo, C.; Söling, H. D.; Sheldrick, G. M.; Ferrari, D. M. Structural Elucidation of the PDI-Related Chaperone Wind with the Help of Mutants. *Acta Crystallogr. D. Biol. Crystallogr.* **2006**, *62* (Pt 6), 589–594. <https://doi.org/10.1107/S0907444906010456>.
- (53) Barak, N. N.; Neumann, P.; Sevvana, M.; Schutkowski, M.; Naumann, K.; Malešević, M.; Reichardt, H.; Fischer, G.; Stubbs, M. T.; Ferrari, D. M. Crystal Structure and Functional Analysis of the Protein Disulfide Isomerase-Related Protein ERp29. *J. Mol. Biol.* **2009**, *385* (5), 1630–1642. <https://doi.org/10.1016/j.jmb.2008.11.052>.
- (54) Vinaik, R.; Kozlov, G.; Gehring, K. Structure of the Non-Catalytic Domain of the Protein Disulfide Isomerase-Related Protein (PDIR) Reveals Function in Protein Binding. *PLoS One* **2013**, *8* (4). <https://doi.org/10.1371/journal.pone.0062021>.
- (55) Aricescu, A. R.; Lu, W.; Jones, E. Y. A Time- and Cost-Efficient System for High-Level Protein Production in Mammalian Cells. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62* (10), 1243–1250. <https://doi.org/10.1107/S0907444906029799>.
- (56) Gorrec, F. The MORPHEUS II Protein Crystallization Screen. *Acta Crystallogr. Sect. Struct. Biol. Commun.* **2015**, *71*, 831–837. <https://doi.org/10.1107/S2053230X1500967X>.

- (57) Trombetta, S. E.; Bosch, M.; Parodi, A. J. Glucosylation of Glycoproteins by Mammalian, Plant, Fungal, and Trypanosomatid Protozoa Microsomal Membranes. *Biochemistry* **1989**, *28* (20), 8108–8116. <https://doi.org/10.1021/bi00446a022>.
- (58) Caputo, A. T.; Alonzi, D. S.; Marti, L.; Reca, I. B.; Kiappes, J. L.; Struwe, W. B.; Cross, A.; Basu, S.; Lowe, E. D.; Darlot, B.; et al. Structures of Mammalian ER  $\alpha$ -Glucosidase II Capture the Binding Modes of Broad-Spectrum Iminosugar Antivirals. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (32), E4630–E4638. <https://doi.org/10.1073/pnas.1604463113>.
- (59) Newman, J.; Egan, D.; Walter, T. S.; Meged, R.; Berry, I.; Jelloul, M. Ben; Sussman, J. L.; Stuart, D. I.; Perrakis, A. Biological Crystallography Towards Rationalization of Crystallization Screening for Small-to Medium-Sized Academic Laboratories: The PACT/JCSG+ Strategy. <https://doi.org/10.1107/S09074444905024984>.
- (60) Stura E.A, Satterthwait, A.C, Calvo, J.C, Kaslow, D.C, Wilson, I. A. Reverse Screening. *Acta Cryst.* **1994**, *D50*, 448–455.
- (61) Cartron, M. L.; Mitchell, S. A.; Woodhall, M. R.; Andrews, S. C.; Watson, K. A. Preliminary X-Ray Diffraction Analysis of YcdB from *Escherichia Coli*: A Novel Haem-Containing and Tat-Secreted Periplasmic Protein with a Potential Role in Iron Transport. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2007**, *63* (1), 37–41.

<https://doi.org/10.1107/S174430910604509X>.

- (62) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; et al. Overview of the CCP4 Suite and Current Developments. *Acta Crystallographica Section D: Biological Crystallography*. April 2011, pp 235–242. <https://doi.org/10.1107/S0907444910045749>.
- (63) Davies, T. G.; Tickle, I. J. Fragment Screening Using X-Ray Crystallography; 2011; pp 33–59. [https://doi.org/10.1007/128\\_2011\\_179](https://doi.org/10.1007/128_2011_179).
- (64) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data Processing and Analysis with the AutoPROC Toolbox. *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67* (Pt 4), 293–302. <https://doi.org/10.1107/S0907444911007773>.
- (65) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
- (66) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66* (Pt 4), 486–501. <https://doi.org/10.1107/S0907444910007493>.

- (67) Blanc, E.; Roversi, P.; Vonnrhein, C.; Flensburg, C.; Lea, S. M.; Bricogne, G. Refinement of Severely Incomplete Structures with Maximum Likelihood in BUSTER-TNT. *Acta Crystallogr. D. Biol. Crystallogr.* **2004**, *60* (Pt 12 Pt 1), 2210–2221. <https://doi.org/10.1107/S0907444904016427>.
- (68) Bricogne G., B. E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonnrhein C., W. T. O. BUSTER Version 2.10.3. Global Phasing Ltd: Cambridge, United Kingdom 2016.
- (69) The PyMOL Molecular Graphics System, Version 2.0. Schrödinger, LLC.
- (70) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372. <https://doi.org/10.1038/nbt.1511>.
- (71) Lamoree, B.; Hubbard, R. E. Current Perspectives in Fragment-Based Lead Discovery (FBLD). *Essays in Biochemistry*. Portland Press Ltd November 8, 2017, pp 453–464. <https://doi.org/10.1042/EBC20170028>.
- (72) Fattori, D. Molecular Recognition: The Fragment Approach in Lead Generation. *Drug Discovery Today*. March 1, 2004, pp 229–238. [https://doi.org/10.1016/S1359-6446\(03\)03007-1](https://doi.org/10.1016/S1359-6446(03)03007-1).

- (73) Nienaber, V. L.; Richardson, P. L.; Klighofer, V.; Bouska, J. J.; Giranda, V. L.; Greer, J. Discovering Novel Ligands for Macromolecules Using X-Ray Crystallographic Screening. *Nat. Biotechnol.* **2000**, *18* (10), 1105–1108. <https://doi.org/10.1038/80319>.
- (74) Davies, D. R.; Begley, D. W.; Hartley, R. C.; Staker, B. L.; Stewart, L. J. Predicting the Success of Fragment Screening by X-Ray Crystallography. In *Methods in Enzymology*; Academic Press Inc., 2011; Vol. 493, pp 91–114. <https://doi.org/10.1016/B978-0-12-381274-2.00004-2>.
- (75) Gossert, A. D.; Jahnke, W. NMR in Drug Discovery: A Practical Guide to Identification and Validation of Ligands Interacting with Biological Macromolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy*. Elsevier B.V. November 1, 2016, pp 82–125. <https://doi.org/10.1016/j.pnmrs.2016.09.001>.
- (76) Giannetti, A. M. From Experimental Design to Validated Hits: A Comprehensive Walk-through of Fragment Lead Identification Using Surface Plasmon Resonance. In *Methods in Enzymology*; Academic Press Inc., 2011; Vol. 493, pp 169–218. <https://doi.org/10.1016/B978-0-12-381274-2.00008-X>.
- (77) Fischer, M.; Leech, A. P.; Hubbard, R. E. Comparative Assessment of Different Histidine-Tags for Immobilization of Protein onto Surface Plasmon Resonance Sensorchips. *Anal. Chem.* **2011**, *83* (5), 1800–

1807. <https://doi.org/10.1021/ac103168q>.

- (78) Linke, P.; Amaning, K.; Maschberger, M.; Vallee, F.; Steier, V.; Baaske, P.; Duhr, S.; Breitsprecher, D.; Rak, A. An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery. *J. Biomol. Screen.* **2016**, *21* (4), 414–421. <https://doi.org/10.1177/1087057115618347>.
- (79) Godemann, R.; Madden, J.; Krämer, J.; Smith, M.; Fritz, U.; Hesterkamp, T.; Barker, J.; Höppner, S.; Hallett, D.; Cesura, A.; et al. Fragment-Based Discovery of BACE1 Inhibitors Using Functional Assays. *Biochemistry* **2009**, *48* (45), 10743–10751. <https://doi.org/10.1021/bi901061a>.
- (80) Schulz, M. N.; Landström, J.; Hubbard, R. E. MTSA - A Matlab Program to Fit Thermal Shift Data. *Anal. Biochem.* **2013**, *433* (1), 43–47. <https://doi.org/10.1016/j.ab.2012.10.020>.
- (81) Vivat Hannah, V.; Atmanene, C.; Zeyer, D.; Van Dorsselaer, A.; Sanglier-Cianférani, S. Native MS: An ESI Way to Support Structure- and Fragment-Based Drug Discovery. *Future Medicinal Chemistry*. January 2010, pp 35–50. <https://doi.org/10.4155/fmc.09.141>.
- (82) Chen, X.; Qin, S.; Chen, S.; Li, J.; Li, L.; Wang, Z.; Wang, Q.; Lin, J.; Yang, C.; Shui, W. A Ligand-Observed Mass Spectrometry Approach Integrated into the Fragment Based Lead Discovery Pipeline. *Sci. Rep.* **2015**, *5*. <https://doi.org/10.1038/srep08361>.



- (83) Hubbard, R. E.; Murray, J. B. Experiences in Fragment-Based Lead Discovery. In *Methods in Enzymology*; Academic Press Inc., 2011; Vol. 493, pp 509–531. <https://doi.org/10.1016/B978-0-12-381274-2.00020-0>.
- (84) Ruiz-Carmona, S.; Schmidtke, P.; Luque, F. J.; Baker, L.; Matassova, N.; Davis, B.; Roughley, S.; Murray, J.; Hubbard, R.; Barril, X. Dynamic Undocking and the Quasi-Bound State as Tools for Drug Discovery. *Nat. Chem.* **2017**, *9* (3), 201–206. <https://doi.org/10.1038/nchem.2660>.
- (85) Collins, P. M.; Ng, J. T.; Talon, R.; Nekrosiute, K.; Krojer, T.; Douangamath, A.; Brandao-Neto, J.; Wright, N.; Pearce, N. M.; Von Delft, F. Gentle, Fast and Effective Crystal Soaking by Acoustic Dispensing. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73*, 246–255. <https://doi.org/10.1107/S205979831700331X>.
- (86) Womack, TO, Smart OS, Sharff A, Flensburg C, Keller P, Paciorek W, V. C. and B. G. RHOFIT, Version 1.2.7. Global Phasing Ltd: Cambridge, United Kingdom 2019.
- (87) Modrow, S.; Falke, D.; Truyen, U.; Schätzl, H. Historical Overview. In *Molecular Virology*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 3–15. [https://doi.org/10.1007/978-3-642-20718-1\\_1](https://doi.org/10.1007/978-3-642-20718-1_1).
- (88) Chou, W.; Ngo, T.; Gershon, P. D. An Overview of the Vaccinia Virus

- Infectome: A Survey of the Proteins of the Poxvirus-Infected Cell. *J. Virol.* **2012**, *86* (3), 1487–1499. <https://doi.org/10.1128/JVI.06084-11>.
- (89) Moss, B. Poxvirus Entry and Membrane Fusion. *Virology* **2006**, *344* (1), 48–54. <https://doi.org/10.1016/j.virol.2005.09.037>.
- (90) Moss, B. Poxvirus DNA Replication. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (9), a010199–a010199. <https://doi.org/10.1101/cshperspect.a010199>.
- (91) Townsley, A. C.; Moss, B. Two Distinct Low-PH Steps Promote Entry of Vaccinia Virus. *J. Virol.* **2007**, *81* (16), 8613–8620. <https://doi.org/10.1128/JVI.00606-07>.
- (92) Townsley, A. C.; Weisberg, A. S.; Wagenaar, T. R.; Moss, B. Vaccinia Virus Entry into Cells via a Low-PH-Dependent Endosomal Pathway. *J. Virol.* **2006**, *80* (18), 8899–8908. <https://doi.org/10.1128/JVI.01053-06>.
- (93) Schmidt, F. I.; Bleck, C. K. E.; Helenius, A.; Mercer, J. Vaccinia Extracellular Virions Enter Cells by Macropinocytosis and Acid-Activated Membrane Rupture. *EMBO J.* **2011**, *30* (17), 3647–3661. <https://doi.org/10.1038/emboj.2011.245>.
- (94) Lin, C. L.; Chung, C. S.; Heine, H. G.; Chang, W. Vaccinia Virus Envelope H3L Protein Binds to Cell Surface Heparan Sulfate and Is Important for Intracellular Mature Virion Morphogenesis and Virus

- Infection in Vitro and in Vivo. *J. Virol.* **2000**, *74* (7), 3353–3365.  
<https://doi.org/10.1128/jvi.74.7.3353-3365.2000>.
- (95) MacLeod, D. T.; Nakatsuji, T.; Wang, Z.; di Nardo, A.; Gallo, R. L. Vaccinia Virus Binds to the Scavenger Receptor MARCO on the Surface of Keratinocytes. *J. Invest. Dermatol.* **2015**, *135* (1), 142–150.  
<https://doi.org/10.1038/jid.2014.330>.
- (96) Chiu, W.-L.; Lin, C.-L.; Yang, M.-H.; Tzou, D.-L. M.; Chang, W. Vaccinia Virus 4c (A26L) Protein on Intracellular Mature Virus Binds to the Extracellular Cellular Matrix Laminin. *J. Virol.* **2007**, *81* (5), 2149–2157. <https://doi.org/10.1128/JVI.02302-06>.
- (97) Fulginiti, V. A.; Papier, A.; Lane, J. M.; Neff, J. M.; Henderson, D. A.; Henderson, D. A.; Inglesby, T. V.; O'Toole, T. Smallpox Vaccination: A Review, Part II. Adverse Events. *Clin. Infect. Dis.* **2003**, *37* (2), 251–271. <https://doi.org/10.1086/375825>.
- (98) Pons-Salort, M.; Parker, E. P. K.; Grassly, N. C. The Epidemiology of Non-Polio Enteroviruses: Recent Advances and Outstanding Questions. *Current Opinion in Infectious Diseases*. Lippincott Williams and Wilkins January 1, 2015, pp 479–487.  
<https://doi.org/10.1097/QCO.0000000000000187>.
- (99) Hober, D.; Sauter, P. Pathogenesis of Type 1 Diabetes Mellitus: Interplay between Enterovirus and Host. *Nature Reviews Endocrinology*. May 2010, pp 279–289.

<https://doi.org/10.1038/nrendo.2010.27>.

- (100) Jiang, P.; Liu, Y.; Ma, H.-C.; Paul, A. V; Wimmer, E. Picornavirus Morphogenesis. *Microbiol. Mol. Biol. Rev.* **2014**, *78* (3), 418–437. <https://doi.org/10.1128/MMBR.00012-14>.
- (101) Xing, W.; Liao, Q.; Viboud, C.; Zhang, J.; Sun, J.; Wu, J. T.; Chang, Z.; Liu, F.; Fang, V. J.; Zheng, Y.; et al. Hand, Foot, and Mouth Disease in China, 2008-12: An Epidemiological Study. *Lancet Infect. Dis.* **2014**, *14* (4), 308–318. [https://doi.org/10.1016/S1473-3099\(13\)70342-6](https://doi.org/10.1016/S1473-3099(13)70342-6).
- (102) Bian, L.; Wang, Y.; Yao, X.; Mao, Q.; Xu, M.; Liang, Z. Coxsackievirus A6: A New Emerging Pathogen Causing Hand, Foot and Mouth Disease Outbreaks Worldwide. *Expert Review of Anti-Infective Therapy*. Taylor and Francis Ltd September 2, 2015, pp 1061–1071. <https://doi.org/10.1586/14787210.2015.1058156>.
- (103) Gonzalez, G.; Carr, M. J.; Kobayashi, M.; Hanaoka, N.; Fujimoto, T. Enterovirus-Associated Hand-Foot and Mouth Disease and Neurological Complications in Japan and the Rest of the World. *International journal of molecular sciences*. NLM (Medline) October 2, 2019. <https://doi.org/10.3390/ijms20205201>.
- (104) Ooi, M. H.; Wong, S. C.; Lewthwaite, P.; Cardoso, M. J.; Solomon, T. Clinical Features, Diagnosis, and Management of Enterovirus 71. *The Lancet Neurology*. November 2010, pp 1097–1105.

[https://doi.org/10.1016/S1474-4422\(10\)70209-X](https://doi.org/10.1016/S1474-4422(10)70209-X).

- (105) Yi, E. J.; Shin, Y. J.; Kim, J. H.; Kim, T. G.; Chang, S. Y. Enterovirus 71 Infection and Vaccines. *Clinical and Experimental Vaccine Research*. Korean Vaccine Society 2017, pp 4–14. <https://doi.org/10.7774/cevr.2017.6.1.4>.
- (106) Fu, X.; Wan, Z.; Li, Y.; Hu, Y.; Jin, X.; Zhang, C. National Epidemiology and Evolutionary History of Four Hand, Foot and Mouth Disease-Related Enteroviruses in China from 2008 to 2016. *Virol. Sin.* **2019**. <https://doi.org/10.1007/s12250-019-00169-2>.
- (107) Messacar, K.; Schreiner, T. L.; Maloney, J. A.; Wallace, A.; Ludke, J.; Oberste, M. S.; Nix, W. A.; Robinson, C. C.; Glodé, M. P.; Abzug, M. J.; et al. A Cluster of Acute Flaccid Paralysis and Cranial Nerve Dysfunction Temporally Associated with an Outbreak of Enterovirus D68 in Children in Colorado, USA. *Lancet* **2015**, *385* (9978), 1662–1671. [https://doi.org/10.1016/S0140-6736\(14\)62457-0](https://doi.org/10.1016/S0140-6736(14)62457-0).
- (108) Holm-Hansen, C. C.; Midgley, S. E.; Fischer, T. K. Global Emergence of Enterovirus D68: A Systematic Review. *The Lancet Infectious Diseases*. Lancet Publishing Group May 1, 2016, pp e64–e75. [https://doi.org/10.1016/S1473-3099\(15\)00543-5](https://doi.org/10.1016/S1473-3099(15)00543-5).
- (109) Rosenfeld, A. B.; Warren, A. L.; Racaniello, V. R. Neurotropism of Enterovirus D68 Isolates Is Independent of Sialic Acid and Is Not a Recently Acquired Phenotype. *MBio* **2019**, *10* (5).

<https://doi.org/10.1128/mBio.02370-19>.

- (110) Cohen, C. J.; Shieh, J. T. C.; Pickles, R. J.; Okegawa, T.; Hsieh, J. T.; Bergelson, J. M. The Coxsackievirus and Adenovirus Receptor Is a Transmembrane Component of the Tight Junction. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (26), 15191–15196. <https://doi.org/10.1073/pnas.261452898>.
- (111) Chang, C.-W.; Leu, Y.-L.; Horng, J.-T. Daphne Genkwa Sieb. et Zucc. Water-Soluble Extracts Act on Enterovirus 71 by Inhibiting Viral Entry. *Viruses* **2012**, *4* (4), 539–556. <https://doi.org/10.3390/v4040539>.
- (112) Tonelli, M.; Simone, M.; Tasso, B.; Novelli, F.; Boido, V.; Sparatore, F.; Paglietti, G.; Pricl, S.; Giliberti, G.; Blois, S.; et al. Antiviral Activity of Benzimidazole Derivatives. II. Antiviral Activity of 2-Phenylbenzimidazole Derivatives. *Bioorg. Med. Chem.* **2010**, *18* (8), 2937–2953. <https://doi.org/10.1016/j.bmc.2010.02.037>.
- (113) Pauwels, R.; Andries, K.; Desmyter, J.; Schols, D.; Kukla, M. J.; Breslin, H. J.; Raeymaeckers, A.; Gelder, J. Van; Woestenborghs, R.; Heykants, J.; et al. Potent and Selective Inhibition of HIV-1 Replication in Vitro by a Novel Series of TIBO Derivatives. *Nature* **1990**, *343* (6257), 470–474. <https://doi.org/10.1038/343470a0>.

