Universitat Autònoma de Barcelona

Synthesis and Computational Studies of Nucleoside Analogues as Antiviral Agents

Rosa Miralles Llumà

Ph.D. Thesis

Ph.D. in Chemistry

Supervisors:

Dr. Ramon Alibés Arqués

Dr. Félix Busqué Sánchez

Dr. Jean-Didier Maréchal

Departament de Química

Facultat de Ciències

2013

Memòria presentada per aspirar al Grau
de Doctor per Rosa Miralles Llumà

Rosa Miralles Llumà

Vist i plau,

Dr. Ramon Alibés Arqués Dr. Félix Busqué Sánchez Dr. Jean-Didier Maréchal

Bellaterra, 2 de maig de 2013

Acknowledgments

Miro enrere i m'adono que aquells quatre anys que jo pensava que se'm farien eterns han passat volant. Em venen a la memòria moltíssims records de situacions que he compartit amb tota la gent que m'ha acompanyat en aquesta aventura. I per molt que és innegable que hi ha hagut tant moments bons com moments durs, tornaria a començar. Sincerament, moltíssimes gràcies a tots els que heu estat aquí i m'heu fet costat aquests anys.

Al Dr. Ramon Alibés li vull agrair, en primer lloc, l'haver-me acollit al grup i el recolzament que sempre m'ha demostrat. Moltes gràcies pel bon humor que et caracteritza, per haver confiat amb mi, per ser un xafarder de primera i, perquè no dir-ho, per la paciència que has tingut tantes vegades. L'agraïment per la paciència el faig extensiu al Dr. Félix Busqué, a qui també li agraeixo l'optimisme "químic" que m'ha encomanat quan jo ja ho donava tot per perdut. Moltes gràcies pels consells i per entendre'm sovint tan bé. Juntament amb el Ramon, us agraeixo moltíssim el suport tant científic com personal durant tot aquest temps.

A la Dra. Marta Figueredo i al Dr. Pau Bayón els vull agrair l'amabilitat i els seus valuosos consells químics. No puc deixar de mencionar també al Dr. Jordi Hernando, que amb el seu optimisme i sarcasme (en el bon sentit de la paraula) sempre aconsegueix fer-me riure.

Al Dr. Jean-Didier Maréchal li he d'agrair moltes coses. Primer de tot, l'haver confiat en mi i haver acceptat que una orgànica 100% baixés a fer càlculs, amb els riscos que això comportava. Recordo perfectament els meus primers dies a Transmet i la paciència que vas tenir per ensenyar-m'ho tot des de zero. Tampoc puc oblidar el suport i els consells que m'has donat a l'hora de pensar en el meu futur, així com l'entusiasme per la recerca que m'has encomanat. Moltes gràcies.

Vull fer extensiu aquest agraïment al Dr. Agustí Lledós i al Dr. Gregori Ujaque, per haver-me fet sentir com a casa des del primer dia que vaig arribar. Gràcies per la bona acollida i per les *penúltimes*.

I would also like to thank Prof. Chris Meier for his support during my stay at the *Universität Hamburg*. Many thanks to all the "AK Meiers", who made me spent a great time there, and especially to my lab colleagues Miri, Johanna and Thiago for their patience and kindness. *Dankeschön*.

Als Font, perquè el dia a dia al vostre costat ha estat simplement genial. Les conyes, la música, les birres i els sopars... m'és molt difícil resumir aquests anys en poques línies. Gràcies als veterans (Gladis, Javi, David, Toni, Èric, Guillaume) per tot el que m'heu ensenyat i pel bon rotllo. Especialment als del 457, valoro molt la vostra paciència amb la combinació ABBA-Queen que sovint se'm repetia! Amb els que he compartit pràcticament tota la tesi (Núria, Marta, Miguel), per tots els cops de mà i les vegades en les que ens hem fet costat mútuament. Als que han vingut darrere (Bea, Marc, Cris i Mireia), pel bon humor i per les ganes d'aprendre. I als encara més nous (Sílvia, Fran), ànims i molta sort!

També agraeixo a la resta de companys de planta la bona convivència, les festes i els congressos. Vull donar les gràcies molt especialment al Guille, l'Eva, el Josep i el Julen. Malgrat que les circumstàncies ens tenen físicament separats, sempre m'alegro de veure com via e-mail mantenim la complicitat que ens uneix. Gràcies per tots els cafès, pel safareig i, en definitiva, per ser uns bons amics.

Als Transmet: sí, sí, ho sé... havien de ser tres mesos i al final hem compartit més de dos anys... però deixeu-me dir-vos que en part és "culpa" vostra. Si no m'hagués sentit tan a gust amb vosaltres hauria tornat ràpid cap a la quarta planta! Qui m'ho havia de dir que arribaria fins aquí amb els qui jo pensava que serien una colla de *freaks*... Gràcies a tots i cada un de vosaltres. Al Víctor, per la espontaneïtat que et caracteritza i que tantes vegades ens fa riure (no la perdis mai, si us plau!), i per tots els cops de mà que m'has donat. Al Max, pel bon humor que aportava al despatx i per ser un "exaltador" excepcional. Al Sergi, perquè amb el seu aire despistat les caça totes i sap donar-te el consell adequat en el moment just. Gràcies per les converses a la cuineta (i pels croissanets!). A la Laia, pel seu somriure encomanadís. A l'Eli, per Gelsinki i per totes les bogeries compartides. I a la Bea, 50% Font-50% Transmet, pel recolzament d'aquest últim temps i per riure tan sanament. *Last but not least*, al Manu. Què dir-te que no sàpigues? No em cansaré d'agrair-te tot el que has fet per mi. I no només parlo del suport científic, que també, sinó del suport emocional. Gràcies per escoltar-me, per aconsellar-me i per donar-me un punt de vista optimista sobre tot el que m'ha preocupat, especialment aquests últims mesos. Has estat un pilar importantíssim, mil gràcies.

Seguint amb els de la primera planta, no em puc descuidar de mencionar l'inestimable paciència i amabilitat del Carles, el nostre tècnic, especialment en

“l’afer Cindy”. I del Solans, que amb l’excusa dels dolços del Sergi sempre ve de visita, m’ofereix les seves classes (un vot de confiança) i ens acaba fent riure a tots.

Als de la carrera (Amàlia, Maria, Víctor i Cris), perquè els anys passen i seguim com sempre: sopars, novetats, bon humor i alguna festeta. I als de Terrassa, especialment al Pol, a la Júlia, a la Mar i a l’Alba, perquè tot i que el temps ha passat i la situació és diferent, heu estat un puntal molt valuós durant moltíssims anys. Gràcies per tot el que hem compartit.

Als de l’alegria, Uri, Helena i Anna, perquè fem el que fem és divertidíssim si hi sou vosaltres. Uri, no perdis mai el punt de vista especial de les coses que tens i que et fa ser tan únic (m’encanta!). Helen, gràcies per la *dolce vita* que vam compartir a Roma, va ser una època genial i en molta part gràcies a la teva companyia. Anna, quants anys al teu costat! Que lluny queden aquelles classes a la masia freixa i quant que ha plogut... malgrat com hem canviat, no has deixat de ser la primera persona a qui recorro en els moments difícils amb qui primer comparteixo les bones notícies. Gràcies de tot cor per ser-hi sempre.

Fran, gran part d’aquesta tesi et pertany. Has tingut una paciència infinita amb mi i m’has cuidat com ningú. M’has donat bons consells químics. M’has fet créixer com a persona. I el que és més important, hem compartit moments que sempre que els recordo em fan somriure. Gràcies, gràcies per tot el que has fet per mi.

Als meus cosins (Núria, Bernat, Marc i Roc) els vull agrair que, cada un a la seva manera, hagin estat al meu costat aquests anys. A la Laia, perquè la vida sense ella no seria el mateix! Ja saps que hi ha coses que només les puc fer amb tu, i això és fantàstic. Gràcies per ser com ets. I a les meves àvies, per tot el que sempre m’han ensenyat. Però sobretot perquè anar a dinar a casa seva és la millor teràpia contra qualsevol maldecap. Gràcies pel vostre optimisme, per fer-me costat, i per demostrar-me que la vida és una lluita, sí, però que es pot guanyar.

I, finalment, vull agrair als meus pares la seva confiança i el recolzament que m’han permès arribar fins aquí. Gràcies per totes les vegades que m’heu animat i m’heu donat l’empenta que em faltava per fer el que realment volia, però que a vegades se’m feia massa difícil. De veritat que sense aquesta força, paciència i optimisme el camí hauria estat molt més complicat.

A tots vosaltres, moltes gràcies!

Financial and technical support

I would like to thank the *Ministerio de Educación y Ciencia* for the predoctoral FPU grant and for the travel scholarship that made possible a three months' stay in the research group of Prof. Chris Meier at the *Universität Hamburg*.

Technical support from *Servei d'Anàlisi Química* and *Servei de Ressonància Magnètica Nuclear* of the *Universitat Autònoma de Barcelona* are also acknowledged.

Happiness is only real when shared

(Into the Wild)

Als meus pares, les meves àvies i la Laia

Als meus amics

Part of the results of this thesis are described in the following scientific publications:

- "Synthesis, anti-HIV activity studies, and in silico rationalization of cyclobutane-fused nucleosides." Figueras, A.; Miralles-Llumà, R.; Flores, R.; Rustullet, A.; Busqué, F.; Figueredo, M.; Font, J.; Alibés, R.; Maréchal, J.-D. *ChemMedChem* **2012**, *7*, 1044–1056.

- "Synthesis and antiviral evaluation of cyclobutane and cyclobutene L-nucleoside analogues." Miralles-Llumà, R.; Figueras, A.; Álvarez-Larena, A.; Balzarini, J.; Figueredo, M.; Font, J.; Maréchal, J.-D.; Alibés, R.; Busqué, F. (submitted)

Contents

Abbreviations

Ac	acetyl
AcCl	acetyl chloride
ACN	acetonitrile
Ac_2O	acetic anhydride
ACV	acyclovir, 9-[(2-hydroxyethoxy)methyl]guanine
ACVDP	acyclovir diphosphate
ACVMP	acyclovir monophosphate
ACVTP	acyclovir triphosphate
ADME	drug absorption, distribution, metabolism and excretion
ADP	adenosine-5'-diphosphate
AIBN	azobis(isobutyronitrile)
AIDS	acquired immunodeficiency syndrome
ATP	adenosine-5'-triphosphate
ATR	Attenuated Total Reflectance
AZT	zidovudine, 3'-azido-3'-deoxythymidine
AZTDP	3'-azido-3'-deoxythymidine-5'-diphosphate
AZTMP	3'-azido-3'-deoxythymidine-5'-monophosphate
Bn	benzyl
Boc	*tert*-butoxycarbonyl
b.p.	boiling point
Bz	benzoyl
18-C-6	18-crown-6
CF	chemical formula
CMV	cytomegalovirus
conc.	concentrated
CSA	camphorsulfonic acid
d4A	2',3'-didehydro-2',3'-dideoxyadenosine
d4T	stavudine, 2',3'-didehydro-2',3'-dideoxythymidine
d4TDP	2',3'-didehydro-2',3'-dideoxythymidine-5'-diphosphate
d4TMP	2',3'-didehydro-2',3'-dideoxythymidine-5'-monophosphate
d4TTP	2',3'-didehydro-2',3'-dideoxythymidine-5'-triphosphate
d4U	2',3'-didehydro-2',3'-dideoxyuridine
DBAD	di-*tert*-butyl azodicarboxylate
DCG	D-Cyclohexenyl-G
ddU	2',3'-dideoxyuridine
de	diastereomeric excess
DEAD	diethyl azodicarboxylate
DIAD	diisopropyl azodicarboxylate
DIBAL-H	diisobutylaluminium hydride
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dT	deoxythymidine
dTDP	deoxythymidine-5'-diphosphate
dTMP	deoxythymidine-5'-monophosphate
dTTP	deoxythymidine-5'-triphosphate
EC	Enzyme Commission
ESI	Electrospray Ionization
Et_3N	triethylamine
EtOAc	ethyl acetate
EtOH	ethanol
FDA	US Food and Drug Administration
FF	Force Field
GA	Genetic Algorithm
GC	Gas Chromatography

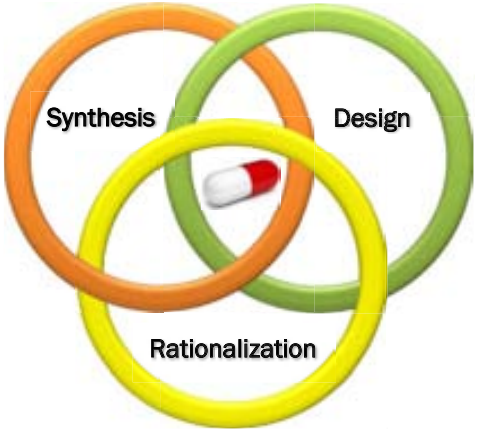
GDP	guanosine-5'-diphosphate
GMP	guanosine-5'-monophosphate
GMPK	guanylate kinase
GTP	guanosine-5'-triphosphate
HBV	hepatitis B virus
HCMV	human cytomegalovirus
HIV	human immunodeficiency virus
HRMS	High-Resolution Mass Spectra
HSV	herpes simplex virus
IC	Incremental Construction
IDU	idoxuridine, 5-iodo-2'-deoxyuridine
^{i}Pr	isopropyl
LCG	L-Cyclohexenyl-G
LDA	lithium diisopropylamide
$LiAlH_4$	lithium aluminium hydride
$LiBH_4$	lithium borohydride
M	molar
MC	Monte Carlo
MCT	methanocarbathymidine
MD	Molecular Dynamics
Me	methyl
$MeNH_2$	methylamine
MeOH	methanol
mGMPK	mouse guanylate kinase
MM	Molecular Mechanics
Mnt	menthyloxy
MOM	methoxymethyl
m.p.	melting point
mRNA	messenger ribonucleic acid
Ms	mesyl
MsCl	mesyl chloride
MW	Molecular Weight
MW	microwave heating
NDP	nucleoside diphosphate
NDPK	nucleoside diphosphate kinase
NMO	*N*-methylmorpholine-*N*-oxide
NMP	nucleoside monophosphate
NMR	Nuclear Magnetic Resonance
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitors
NRTI	Nucleoside Reverse Transcriptase Inhibitors
NTP	nucleoside triphosphate
P	octanol-water partition ratio
PCV	penciclovir, 9-[4-hydroxy-3-(hydroxymethyl)but-1-yl]guanine
PDB	Protein Data Bank
Ph	phenyl
Piv	pivaloyl
PivCl	pivaloyl chloride
p-TsOH	*p*-toluenesulfonic acid
py	pyridine
QM	quantum mechanics
RNA	ribonucleic acid
RT	Reverse Transcriptase
rt	room temperature
(S)-MCT	3'-*exo*-methanocarbathymidine
SO	Swarm Optimization
TBAF	tetrabutylammonium fluoride
TBDPS	*tert*-butyldiphenylsilyl
TBDPSCl	*tert*-butyldiphenylsilyl chloride
TBS	*tert*-butyldimethylsilyl
TBSCl	*tert*-butyldimethylsilyl chloride
^{t}Bu	*tert*-butyl
THF	tetrahydrofuran
TK	thymidine kinase
TLC	Thin Layer Chromatography
TMK	thymidylate kinase
TMSOTf	trimethylsilyl trifluoromethanesulfonate
TsCl	tosyl chloride
VSV	vesicular stomatitis virus
VV	vaccinia virus
VZV	varicella-zoster virus

Preface

Viruses cause a broad variety of diseases that strongly influence worldwide health and economy. Despite the undeniable progress achieved on antiviral therapy by the discovery of some compounds featuring activity, many efforts are still dedicated to the search for novel antiviral drugs. In this context, synthetic organic chemistry and molecular modelling are two of the major areas of research. The present thesis is located on the intersection of these two disciplines.

The work herein presented is focused on the study of nucleoside analogues as antiviral agents. Thus, the first chapter of this manuscript consists in a general introduction on viruses and antiviral drugs, with special emphasis on the application of nucleoside analogues as antiviral agents. The objectives of the present work are also included at the end of the chapter.

The following chapters are dedicated to the results and divided in three parts: synthesis, rationalization and design. The first one is focused on the synthesis of novel carbocyclic nucleoside analogues (Chapter II). After a brief overview on molecular modelling in drug design (Chapter III), the next chapters are devoted to the application of a novel protein–ligand docking approach. This approach is aimed at studying crucial phenomena that define the antiviral activity of nucleoside analogues. The procedure is used to rationalize the activity of a series of nucleosides against HIV (Chapter IV) and HSV (Chapter V), and finally to design novel nucleoside analogues as anti-HSV agents (Chapter VI).



CHAPTER I

General introduction and objectives

1. Introduction

What does the word “virus” mean? Nowadays nearly everybody can give an answer to this question. The wide range of diseases caused by viruses is definitely one of the most significant health issues in today’s society. However, nobody could have answered this question at the end of the nineteenth century. Despite the fact that infections caused by virus have been known and feared since the dawn of history, it was not until the latter part of the nineteenth century that scientists came up with the idea that living agents, smaller than any known bacteria, could cause diseases in plants and animals.[1] Nowadays it is generally accepted that viruses are capable of parasitizing all groups of living organisms.

In the last decades, the great expansion of molecular techniques and technologies for studying microorganisms has allowed to get a deeper insight into the world of viruses. Now it is known how viruses are and how they replicate, which has facilitated to face the challenge of fighting the diseases they cause. However, as we will see later in this first part of the present thesis, there is still a long way to go on this issue.

This chapter will be devoted to provide basic notions on viruses and antiviral compounds. With this aim, the first part will focus on the main characteristics of viruses as well as their replication process, whereas the second part will deal with the diseases they cause and the related antiviral compounds.

1.1. Viruses

1.1.1. What are viruses?

Although it is not easy to provide a concise definition of viruses, it can be easily pointed out why viruses are different from the rest of microorganisms.

Viruses have the following characteristics:

- They are small, ranging from about 20 to 200 nm in diameter.[a]
- They are totally dependent on living cells, either eukaryotic or prokaryotic, for replication and existence. Indeed, viruses lack the

[a] It is noteworthy that most of the bacteria are larger than 250 nm.

genetic information that encodes the tools necessary for the generation of metabolic energy or for protein synthesis.

- Their genetic material may be either DNA or RNA, and in both cases it may be double (ds) or single (ss) stranded.

Accordingly, viruses have sometimes been defined as submicroscopic, obligate intracellular parasites.[2] This dependence on the cell machinery to replicate implies that viruses can be found both inside and outside the cells. In the extracellular environment the virus particle is metabolically inert, whereas in the intracellular one its replication takes place.[3]

An extracellular virus particle, the so-called virion, is formed by its viral nucleic acid protected by a protein coat termed capsid (Figure 1). This capsid is made up of numerous capsomeres encoded by the viral genome and facilitates the entry into the host cell. The complex of the capsid and the virus genetic material is called nucleocapsid. Various enzymes needed for the replication can also be found in some virions. Moreover, some viruses possess an outer lipid envelope, derived from the plasma membrane of the infected cell when the virus particle is released. Viral envelopes also facilitate viruses' entrance into the host cells.

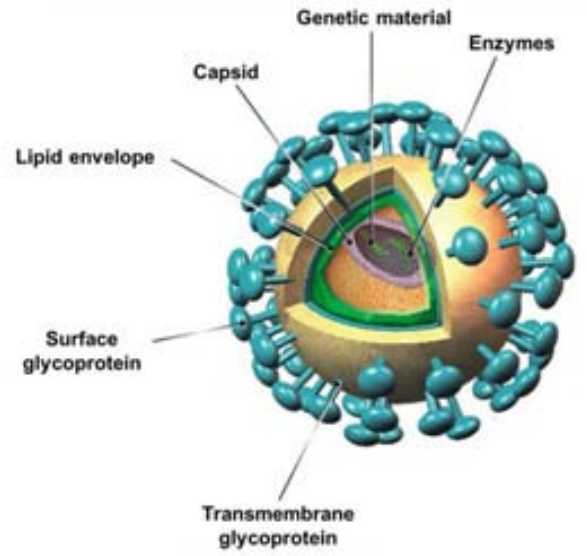


Figure 1. General structure of viruses. (Reproduced by permission of COREVIH Bourgogne – Centre Hospitalier Universitaire de Dijon)

1.1.2. How do viruses replicate?

As previously stated, viruses must rely on host cell biochemical machinery to replicate and spread. Remarkably, different strategies of replication are known, depending on the coding capability of the virus genome.

For the sake of clarity, this section will only provide a schematic overview of the major events that occur during replication. Thus, six sequential steps can be defined for the replication process (Figure 2):

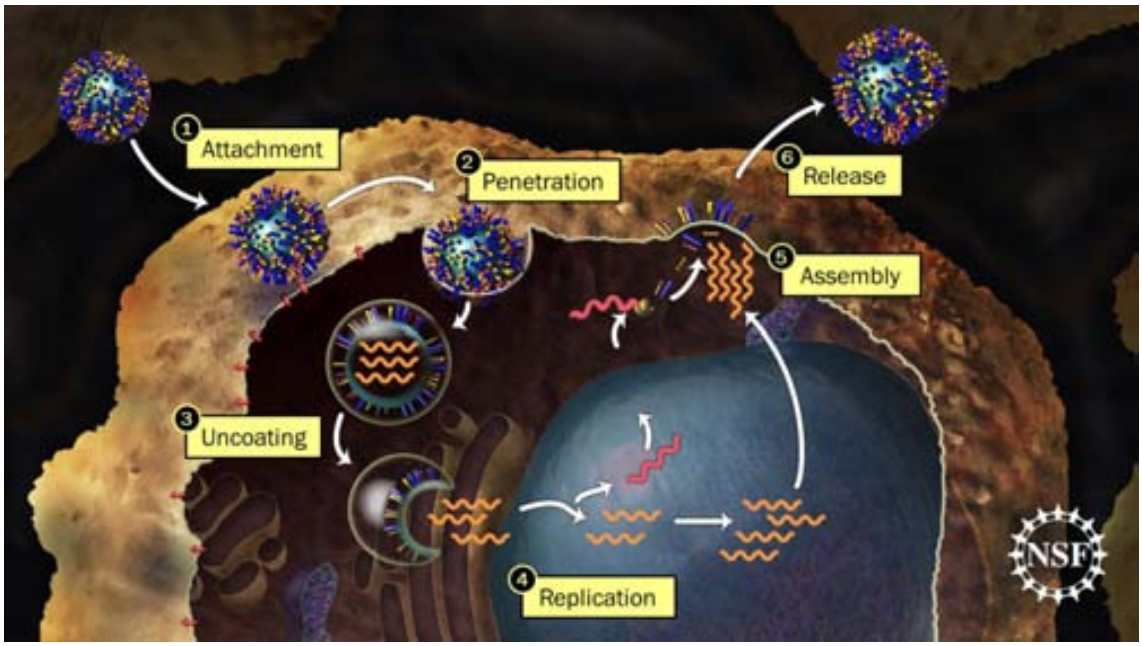


Figure 2. Viral life cycle. (Reproduced by permission of Nicole R. Fuller, National Science Foundation)

- *Attachment*. It consists of the binding of a virus surface protein (also called antireceptor) to a specific receptor on the host cell membrane. The specificity of the interaction between the cell receptor and the virus binding protein restricts the virus hosts and even the particular type of cells or tissues that will be infected.
- *Penetration* of the virion into the target cell. This process may occur by endocytosis or by fusion of the virus and host cell lipid membranes.
- *Uncoating*. The virus particle sheds its protective protein coat (i.e. the capsid) to release the viral nucleic acid.
- *Replication*. This process covers the three main events that take place when viruses infect cells: the formation of viral mRNAs, the translation into the viral proteins and the replication of the virus genome. How the

viral mRNAs are formed depends on the nature of the virus genetic material, which can be RNA or DNA. RNA viruses can be divided into three groups: positive-sense RNA viruses, which RNA can be directly translated into proteins; negative-sense RNA viruses, carrying an RNA polymerase responsible for making mirror-image copies of the original RNA (complementary to the mRNA); and retroviruses such as the human immunodeficiency virus (HIV), which use a virus-associated reverse transcriptase enzyme (RT), carried by the virus itself, to perform three sequential transformations. First, to convert the viral RNA into a DNA-RNA hybrid; second, to digest the RNA strand and, finally, to replace it by a DNA copy to give a double stranded DNA that is integrated into the host cell DNA by an enzyme called integrase. At this point, the host cell transcribes both its own DNA and the viral DNA, producing thus the corresponding viral mRNAs. Regarding DNA viruses, a host cell enzyme (DNA-dependent RNA polymerase II) is responsible for transcribing the mRNAs from the viral DNA. Once the mRNAs are formed, the virus messages are translated and viral proteins are synthesized by the host cell ribosomes. Concerning the viral genome replication, it is performed by an RNA-dependent RNA polymerase (the so-called replicase) in positive- and negative-sense RNA viruses, retroviruses use the above mentioned reverse transcriptase whereas a DNA polymerase is responsible for the replication of DNA viruses.

- *Assembly* of the newly formed RNA or DNA with the viral structural polypeptides to form virions. This process may take place either in the cytoplasm or in the nucleus.
- *Release*. The new virions are released from the cell by lysis or emerging from the cell surface and acquiring their outer lipid envelope from the host plasma membrane (budding).

The new virions formed are then free to infect neighbouring cells and start the process again. Some viruses, such as HIV, do not emerge from the cell, but they may spread to nearby cells via connecting pores or by inducing fusion of their membranes. Interestingly, the complete viral life cycle generally takes between 6 and 8 hours, and as many as 10,000 virions may be released from an infected cell.[1]

Undeniably, knowing the virus replication process has been one of the key points in the search for effective antiviral drugs.

1.1.3. Why are viruses so important?

Once the main concepts at microscopic level on viruses and their replication processes have been introduced, this section is devoted to provide a general picture of how viruses influence human life.

One of the most common misconceptions is that virus infection inevitably results in disease but, actually, only a reduced number of virus infections give rise to disease symptoms. However, nobody questions that viruses cause a large-scale impact on worldwide health and economy.

HIV, the causative agent of acquired immunodeficiency syndrome (AIDS),[4,5] is a good example. Although the number of new infections has dropped in the last years, the level of new infections is still high. According to the 2011 annual report on the global AIDS epidemic published by the World Health Organization (WHO), United Nations Children's Fund (UNICEF) and Joint United Nations Programme on HIV/AIDS (UNAIDS),[6] the number of people newly infected by HIV in 2010 was 2.7 million (to get an idea, this would represent a 0.4% of the European population newly infected by HIV in one year). Moreover, 1.8 million people died in 2010 because of this pathogen, and the estimated number of people living with HIV worldwide was 34 million in December 2010. Despite the fact that antiretroviral therapy has changed AIDS from a rapidly lethal disease into a chronic manageable illness, an effective vaccine against this virus is yet to be discovered.[7,8]

On the other hand, some viruses such as herpes simplex (HSV) cause endemic diseases, i.e. diseases that are constantly present at a significant level within a community. HSV-1 and HSV-2 infections are widespread and can be life-threatening to immunocompromised patients.[9] HSV-1 mainly causes orofacial lesions, whereas genital herpes is most frequently caused by HSV-2. Moreover, genital herpes is associated with an increased risk of HIV acquisition.[10,11] It is worth recalling that in 2008 the estimated number of people infected by HSV-2 worldwide was 535.5 million,[12] which represents 3/4 of the European population.

These are only two examples of the very long list of viruses and related diseases affecting humans. Furthermore, every day there are more evidences that

viruses and cancer are in some cases closely related.[1] It is worth mentioning that HIV and HSV are also the two target viruses of the new antiviral agents developed in this thesis.

Undoubtedly, prevention of virus infection is far better than cure, especially considering that most of the damage to cells occurs very early after the infection, often before the first clinical symptoms of disease appear. Many efforts have been made on the development of novel vaccines against viruses, some of them proving successful (e.g. hepatitis B, smallpox).[13] However, certain viruses still remain uncontrolled by immunization. These are certainly the target viruses for antiviral chemotherapy.

1.2. Antiviral drugs

The success of an antiviral drug lies mainly in its specificity, since it must be more toxic to the virus than to the host cells. Almost any stage on the replication process of viruses is susceptible to be a target for antiviral compounds.

Nowadays around 60 antiviral drugs have been approved by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA).[14] Among them, compounds targeting proteases[a] as well as viral nucleic acid replication and transcription are the largest group. On the other hand, it is noteworthy that almost half of the approved drugs are against HIV.[7,15,16] The common categories used to classify the anti-HIV drugs are the following:[7,14]

- Nucleoside Reverse Transcriptase Inhibitors (NRTIs)
- Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)
- Protease Inhibitors (PIs)
- Viral entry inhibitors (including coreceptor inhibitors -CRIs- and fusion inhibitors -FIs)
- Integrase Inhibitors (INIs)

Nowadays, HIV therapy is based on a combination of these agents,[17–19] but the first used NRTI compound, Zidovudine (AZT),[20] is still a cornerstone of HIV treatment.[8]

[a] Proteases are virus-encoded enzymes responsible of performing some structural changes to certain virus particles after the replication process, to make them infectious.

Moreover, treatment of many other widespread diseases caused by viruses is also mainly based on nucleoside analogues.[21,22] Indeed, nucleosides are nowadays the active ingredient of one third of the antiviral drugs approved by the FDA, becoming thus of great importance among the compounds with antiviral activity.[21,23–25] Additionally, a number of new nucleoside analogues are in various stages of clinical development to be approved as antiviral drugs.[7,26–28]

1.2.1. Nucleoside analogues

Nucleotides are the structural units of nucleic acids. Each nucleotide consists of three subunits: one or more phosphate groups; a sugar moiety, which can be either ribose (RNA) or deoxyribose (DNA); and a pyrimidine or purine base (Figure 3). Nucleosides differ from nucleotides in that they lack the phosphate group(s).

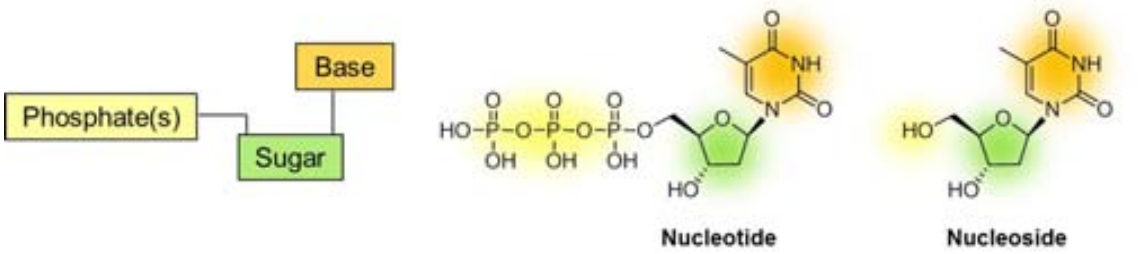


Figure 3. General structure of a nucleotide and structures of a natural pyrimidine nucleotide and nucleoside.

In order to modulate nucleosides and nucleotides activity, each of the three previously mentioned subunits may be modified. This strategy has been widely used since the synthesis of Idoxuridine (IDU) (Figure 4) in 1959,[29] the first clinically effective antiviral nucleoside analogue, which was found to be active against some herpes viruses.[23] Nowadays, around 20 of the antiviral agents approved by the FDA are based on nucleoside or nucleotide analogues. Two of the major breakthroughs in the field of nucleoside analogues were undoubtedly the synthesis of Acyclovir (ACV) (Figure 4), reported to exhibit antiviral activity against herpes simplex virus (HSV-1, HSV-2) and varicella-zoster virus (VZV) in 1977,[30] as well as the discovery in 1985 of the anti-HIV agent Zidovudine (AZT) (Figure 5).[8,20] Those findings encouraged investigators to search for novel nucleoside analogues with potent antiviral activity.[21,24,31–33]

In the case of the treatment of herpes viruses, the discovery of ACV quickly prompted the synthesis of other acyclic nucleoside analogues such as Ganciclovir[34]

(GCV) and Penciclovir[35] (PCV), as well as of their prodrugs Valaciclovir[36] (Val-ACV), Valganciclovir[37] (Val-GCV) and Famciclovir[38] (FCV) (Figure 4). Furthermore, the synthesis of Trifluridine[39] (TFT) and Brivudine[40] (BVDU) was envisaged after describing the antiviral activity of the previously mentioned IDU. It is worth noting that the nucleotide analogue Cidofovir[41] ((S)-HMPC) features a methylphosphonate group instead of a phosphate in order to increase its stability towards hydrolysis (Figure 4).

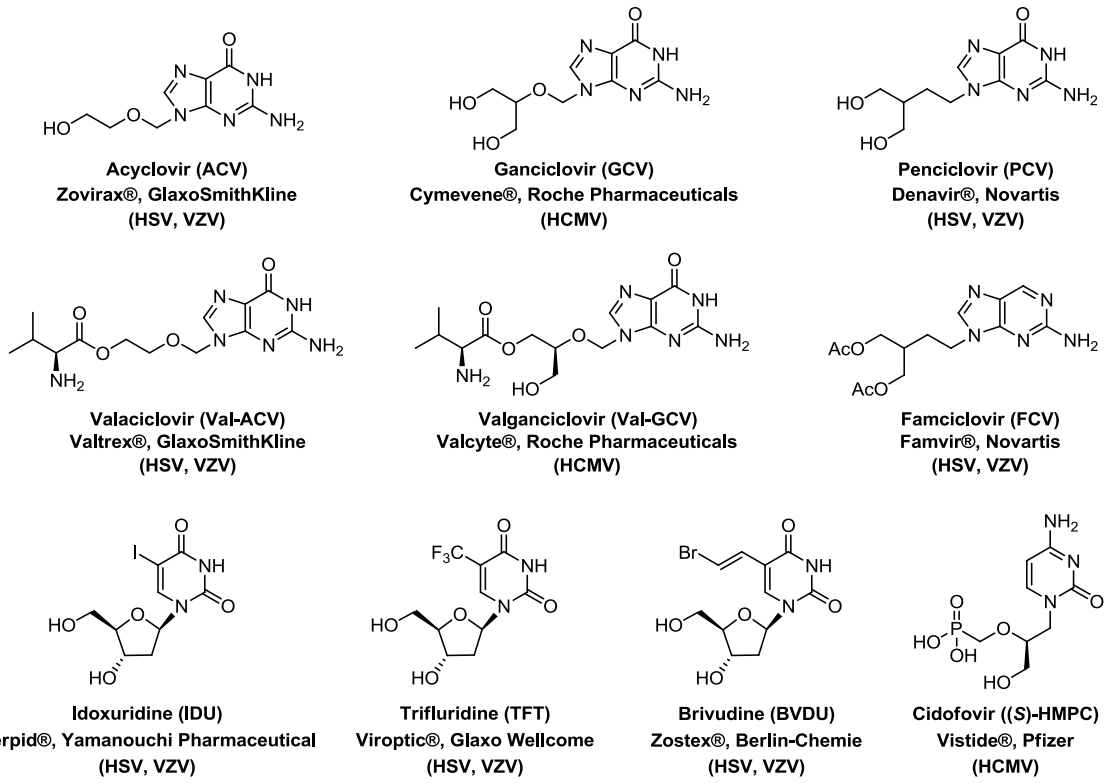


Figure 4. Antiviral drugs based on nucleoside and nucleotide analogues currently used for the treatment of herpes viruses: active principal ingredient, trade name and manufacturer. The specific herpes virus inhibited by each drug is indicated in parenthesis.

Regarding the HIV treatment, a wide range of 2',3'-dideoxynucleosides were synthesized after the discovery of AZT, some of them being active against the HIV, such as Zalcitabine[42] (ddC), Stavudine[43] (d4T) and Didanosine[44] (ddI) (Figure 5). Moreover, further structural modifications led to the discovery of the antiviral activity of some carbocyclic nucleosides, which present a carbocycle instead of the furanose ring, and L-nucleosides, enantiomers of natural or modified nucleosides.[45] Thus, the L-nucleosides Lamivudine[46] (3TC) and Emtricitabine[47] (FTC), the carbocyclic nucleoside Abacavir[48] (ABC), as well as the acyclic nucleoside

phosphonate Tenofovir disoproxil fumarate[49] (TDF) proved active against HIV (Figure 5).

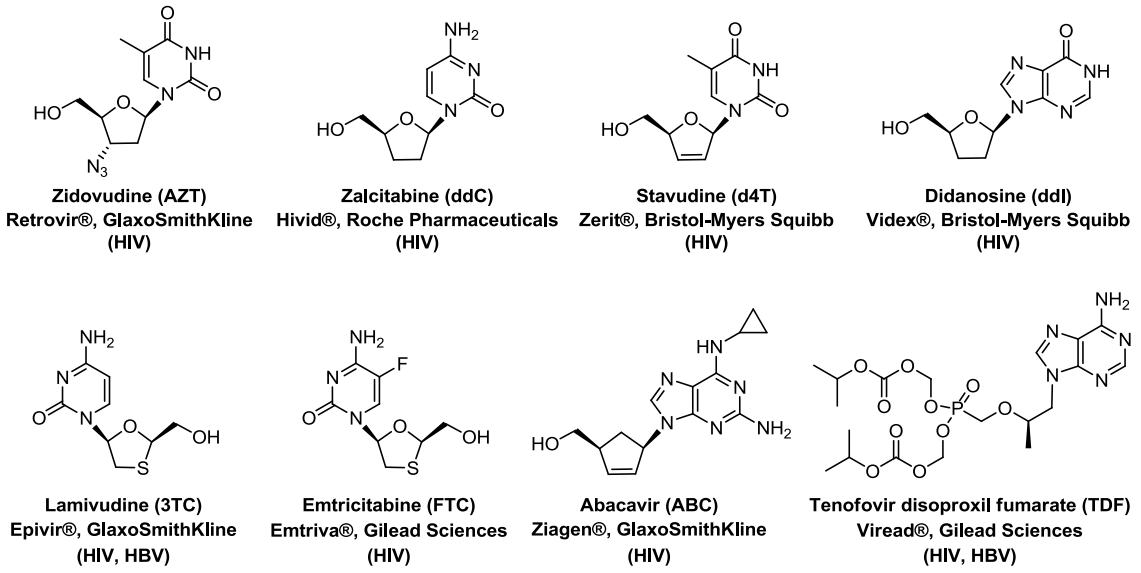


Figure 5. Antiviral drugs based on nucleoside and nucleotide analogues currently used for the treatment of HIV: active principal ingredient, trade name and manufacturer. The drugs also approved for the treatment of hepatitis B virus (HBV) are indicated in parenthesis.

Although the discovery of each one of these compounds represented a milestone on the journey towards better controlling virus infections, the development of new antiviral agents is still necessary. Major efforts are being made in the search for nucleoside analogues that overcome the main deficiencies of the current drugs, such as their side effects,[25,50–52] high toxicity[52–55] and, more importantly, the emergence of virus drug resistance.[7,27,55–62]

The patterns of virus resistance are quite complex.[27,55–57,59,60,62–72] However, it is known that the resistance is accelerated by the increasing use of antiviral drugs.[60,70,73] Therefore, an extensive knowledge on the mechanism of action of nucleoside analogues is mandatory to face the resistance issue.

The main target of nucleoside analogues are the polymerases responsible for carrying out the viral nucleic acid replication (reverse transcriptase, DNA polymerase or RNA polymerase), which substrates are the triphosphorylated nucleosides. Hence, nucleoside analogues are the prodrug forms of the active compounds, and must be converted into their triphosphorylated derivatives to be active. This process consists of three successive steps catalysed by kinases, which may be present in the host cell or encoded by the virus itself. The kinases involved in the activation of nucleosides are the following: nucleoside kinase,

monophosphate nucleoside kinase, and diphosphate nucleoside kinase, which catalyse the addition of the first, second, and third phosphoryl groups at 5' position of nucleoside analogues, respectively (Figure 6).[74,75]

Since nucleotides are the building blocks of DNA, they may now interact with the corresponding viral polymerase, competing with the endogenous nucleoside triphosphates to interfere the viral nucleic acid replication. Depending on the nature of the nucleosides, this process may be performed in two different ways. If the nucleoside analogue lacks the 3'-hydroxyl group, which is essential to add further nucleotides into the DNA growing chain, its incorporation into the primer DNA strand will definitely stop the elongation process (these compounds are therefore known as chain terminators).[55,72,76] Conversely, nucleoside analogues that contain the 3'-hydroxyl group may also interrupt the replication of the viral nucleic acid by different mechanisms, such as causing steric hindrance when some of them have been added into the DNA growing chain. Therefore, it is essentially through the inhibition of viral nucleic acid synthesis by one of the above mentioned processes that nucleosides exhibit their antiviral activity (Figure 6).

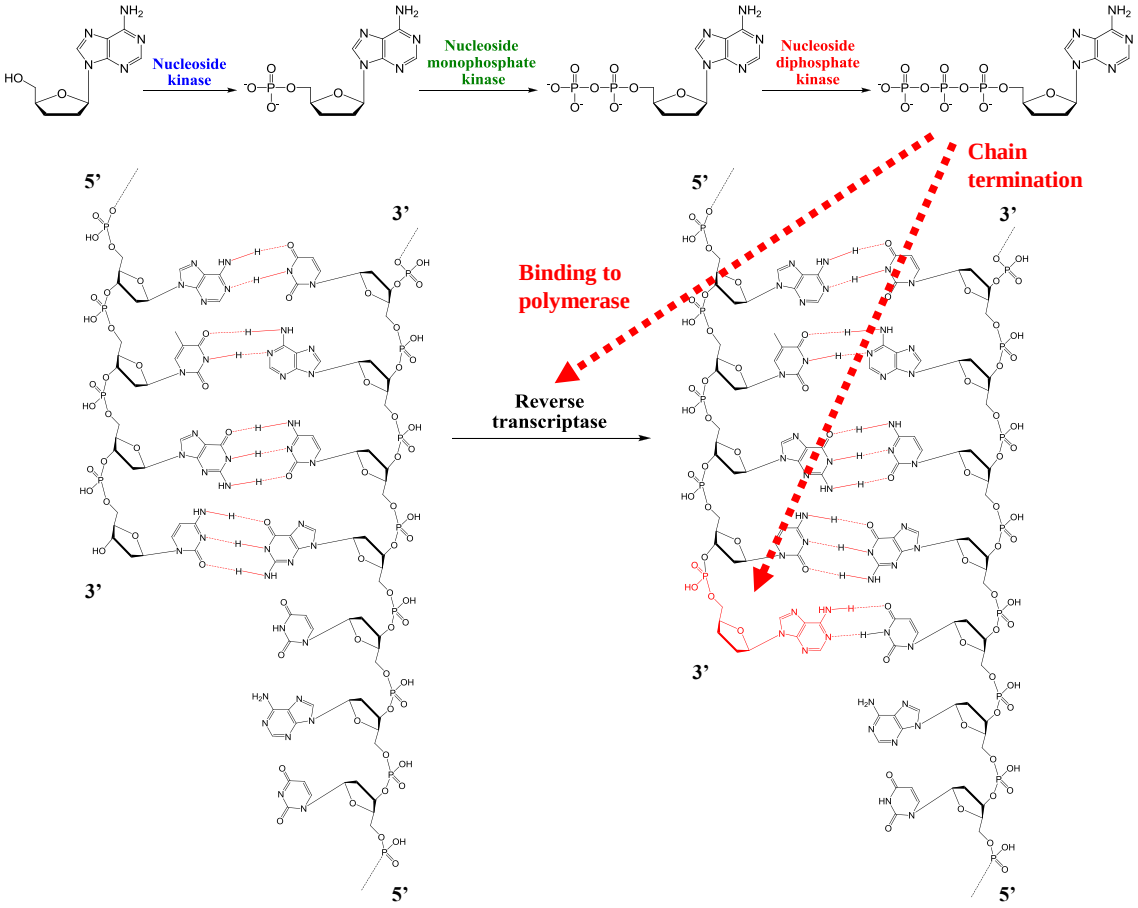


Figure 6. Mechanism of inhibition of viral replication by 2',3'-dideoxynucleosides.

As it has been mentioned, modifications on the structure of nucleoside analogues are aimed at decreasing their toxicity to the host cells and increasing their specificity towards the viruses. This is mainly achieved when the nucleoside analogue affinity for the virus-associated polymerase is higher than for the human DNA polymerase.[30,74] Remarkably, the low toxicity of some antiviral compounds, such as ACV, is also related to the narrow substrate acceptance of human kinases compared to the viral counterparts.[77,78] Thus, ACV is a substrate of the HSV kinase that catalyses its first phosphoryl transfer, but not of the human counterpart.[23,30,76] As a result, the non-infected cells will not phosphorylate ACV, preventing its activation and subsequent interaction with the human DNA polymerase, which would eventually have led to a cytotoxic compound.

The accumulative data at molecular level of all these processes, together with the extensive knowledge on viral and cellular replication events, has allowed the rational design of novel nucleoside analogues with improved selectivity and reduced toxicity. In this field, the computational tools available nowadays are also of considerable importance. Moreover, these recent significant advances have notably contributed to establish some mechanisms of antiviral resistance. One of the major mechanisms for drug resistance is based on mutations in the viral genes that encode for the viral DNA polymerase, reverse transcriptase or viral kinases. These mutations result in structural changes in certain viral enzymes that affect the proper binding of the antiviral compounds, leading to drug-resistant virus strains.[26,57,67]

The search for novel nucleoside analogues featuring higher selectivity and different resistance profiles than the currently used is still a major cornerstone of today's antiviral research. In this context, efforts have been initially focused on transforming the sugar moiety of nucleosides. Two of the main nucleoside analogues, carbocyclic nucleosides and L-nucleosides, will be presented in the following sections.

1.2.2. Carbocyclic nucleoside analogues

Besides the previously mentioned limitations on the clinical application of nucleoside analogues as antiviral agents, another important drawback of some of these compounds is their enzymatic degradation. Carbocyclic nucleosides (also called *carba*-nucleosides), where the furanose ring of the regular nucleosides is

replaced by a carbocycle, have received extraordinary attention over the last three decades.[79] The lack of the *N*-glycosidic linkage in these derivatives results in an increased metabolic stability towards phosphorylase enzymes.[80,81] Furthermore, the higher lipophilicity of carbocyclic nucleosides compared to the natural counterparts could be beneficial for their cell wall penetration.[a,82,83]

Interestingly, two *carba*-nucleosides, Aristeromycin[84] and Neplanocin A,[85] were isolated from natural sources and described to exhibit potent biological activity in 1968 and 1981, respectively (Figure 7).

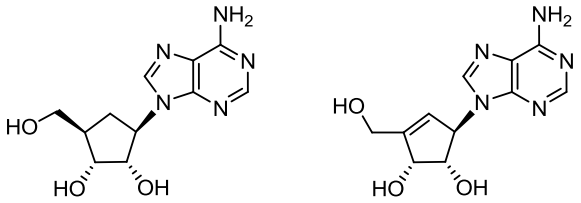


Figure 7. Carbocyclic nucleosides isolated from natural sources.

The biological properties of these compounds triggered the synthesis of a broad range of carbocyclic analogues featuring different ring sizes (Figure 8 and Figure 9).[22,33,79,86–89] The most widely studied analogues have undoubtedly been the five-membered ring.[88–90] Carbovir was prepared for the first time in 1988 and was shown to have an activity similar to AZT, but it was limited by its pharmacokinetic and toxicological deficiencies.[45,79,91–93] Better results were found for Abacavir, which was approved by the FDA for the HIV treatment in 1998.[48] Some other cyclopentane and cyclopentene analogues featuring a broad spectrum of antiviral activity are based on the natural compounds Aristeromycin and Neplanocin A (e.g. 5'-Noraristeromycin),[94] or on existing antiviral drugs (e.g. C-BVDU),[22] among others (Figure 8).

[a] Cell wall penetration is a requirement for the oral bioavailability of drugs, which is defined as the fraction of the oral dose that reaches systemic circulation. For more details on how to predict oral bioavailability of drugs, see section *2. Drug-likeness* of Chapter III.

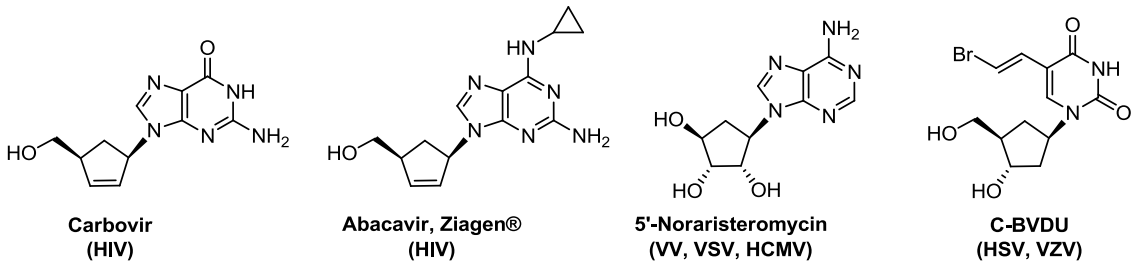


Figure 8. Five-membered carbocyclic nucleoside analogues that exhibit antiviral activity. The specific viruses inhibited by each compound are indicated in parenthesis.

Carbocyclic nucleosides bearing a cyclopropane unit have also been explored, some of them showing antiviral activity (e.g. the cyclopropane analogue A-5021 is nearly 20 times more potent than ACV) (Figure 9).[95] Among analogues presenting six-membered rings, both enantiomer nucleosides D- and L-Cyclohexenyl-G display potent anti-herpesvirus activity (HSV-1, HSV-2, VZV, CMV) (Figure 9).[96] Regarding four-membered carbocyclic analogues, cyclobutane nucleosides Cyclobut-A and Cyclobut-G inspired by the natural compound Oxetanocin-A,[97] which exhibits anti-HIV activity, also display good antiviral potency (Figure 9).[98–100]

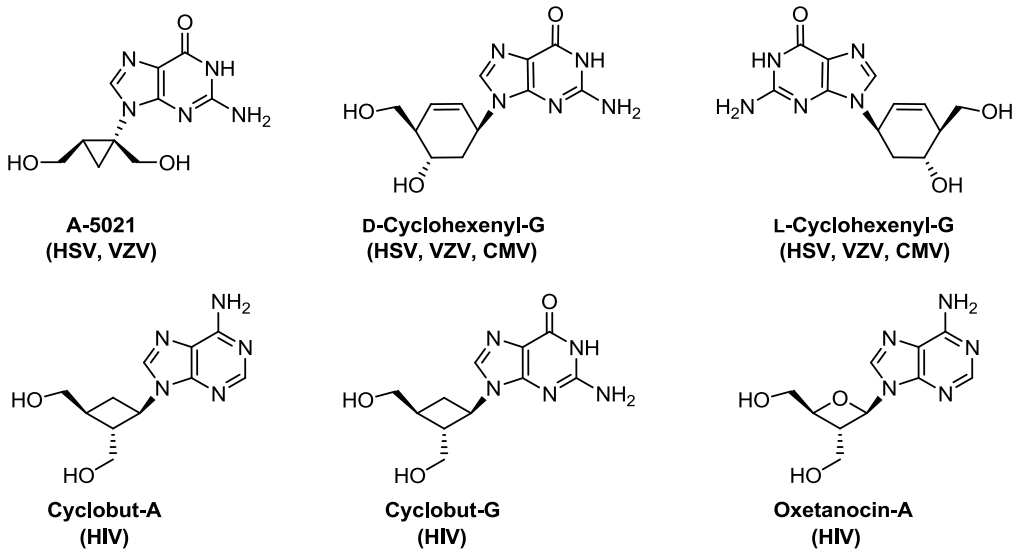


Figure 9. Three-, six- and four-membered carbocyclic nucleoside analogues that exhibit antiviral activity, and the naturally occurring nucleoside analogue Oxetanocin-A. The specific viruses inhibited by each compound are indicated in parenthesis.

Although cyclobutane analogues of Oxetanocin-A have been quite investigated,[86] only six cyclobutene nucleosides (envisaged as analogues of the active compounds Carbovir and Abacavir) have been prepared to date[a] (Figure

[a] For further details on the synthetic strategies to synthesize these analogues, see section *1.1. Cyclobutene and polyhydroxylated cyclobutane nucleoside analogues. Precedents* of Chapter II.

10).[101–104] Therefore, the synthesis of new cyclobutene nucleoside analogues may still provide useful insights on the biological properties of these compounds.

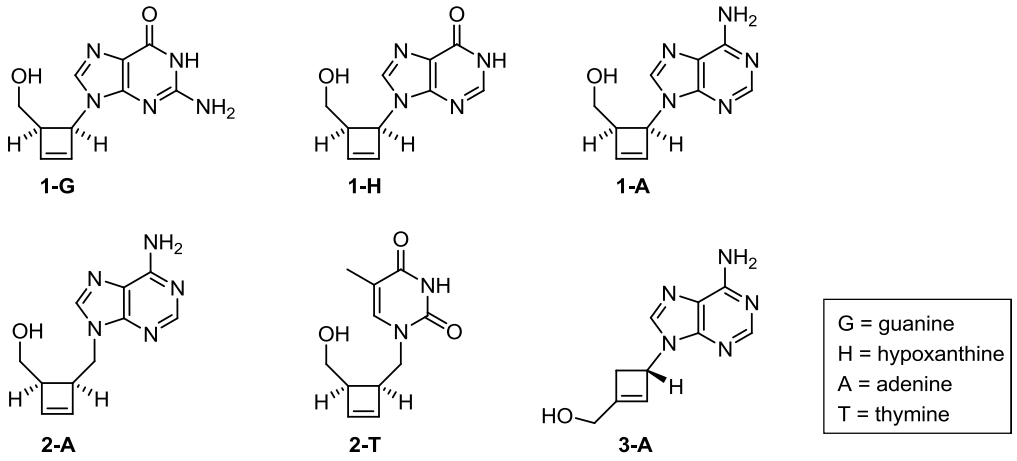


Figure 10. Cyclobutene nucleoside analogues synthesized to date.

1.2.3. L-nucleoside analogues

As it has been mentioned earlier, L-nucleosides are enantiomers of the naturally occurring D-nucleosides (Figure 11).

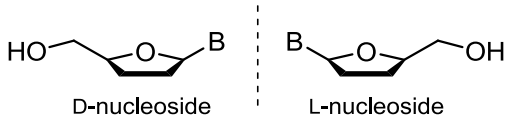


Figure 11. D- and L-nucleosides.

For a long period it was assumed that only nucleoside analogues having the natural D-configuration could exhibit antiviral activity.[105] Indeed, the first synthesis of an L-nucleoside dates back to 1964,[106] but it was not until the discovery of the antiviral activity of 3TC (Figure 12), in 1992, that this assumption was reevaluated. Since then, a plethora of analogues with an unnatural (L-) configuration have been synthesized.[107–110] Interestingly, many of them display lower toxicity, higher metabolic stability and an antiviral activity comparable and sometimes greater than the natural D- counterparts.[46,47,107,111,112]

The different behaviour of L- and D-nucleosides was first investigated for a racemic mixture containing 3TC and its D- counterpart. The resolution of that mixture showed that 3TC was more active against HIV than its D- counterpart, whereas the toxicity of the racemic was mostly caused by the latter.[46,112] Further studies proved that the higher activity of 3TC is mainly due to its major affinity for

the kinase responsible for catalysing the first phosphoryl transfer in the activation process.[113] However, both enantiomers are converted into their triphosphorylated derivatives, which may interact with the viral DNA polymerase as chain terminators.[114] Remarkably, the D-isomer also inhibits cellular DNA polymerases, resulting in its toxic effects.[115]

Moreover, some L-nucleosides also exhibit resistance toward various catabolizing enzymes, resulting in higher stability of these compounds compared to their natural counterparts.[116,117] This property may also account for their enhanced activity against some viruses.[46,107]

Hence, 3TC as well as Emtricitabine[47] (FTC, the 5-fluorinated analogue of 3TC) and Telbivudine[118] (LdT, the enantiomer of the natural nucleoside deoxythymidine, dT), have been approved by the FDA for the treatment of HIV and HBV (Figure 12).

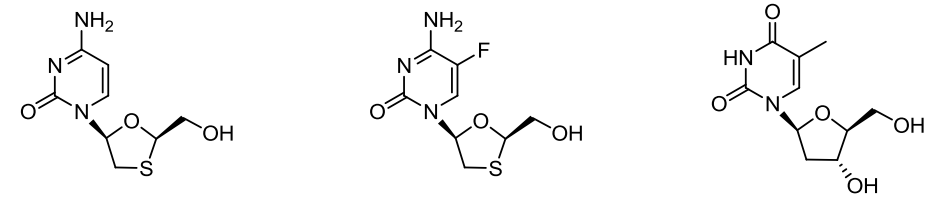


Figure 12. L-nucleoside analogues approved for the treatment of HIV and HBV: active principal ingredient, trade name and manufacturer. The specific virus inhibited by each drug is indicated in parenthesis.

Furthermore, several other L-nucleosides such as Maribavir,[119] Clevudine,[120] Elvucitabine[121] and Pentacept,[122] among others, have shown potent antiviral activity (Figure 13),[109] and some of them are currently in clinical trials.[27]

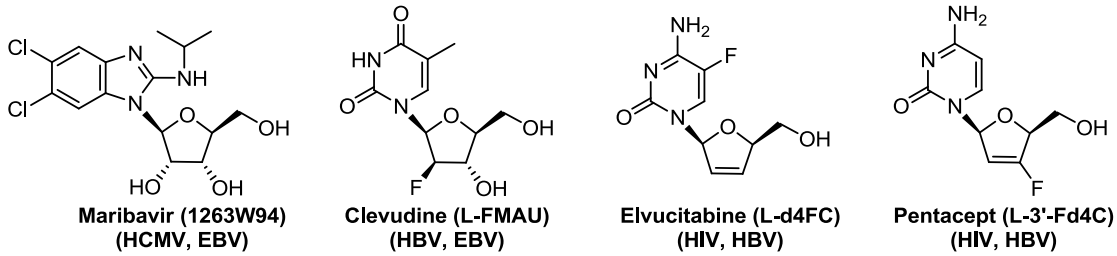


Figure 13. L-nucleoside analogues that exhibit antiviral activity. The specific virus inhibited by each compound is indicated in parenthesis.

It is worth highlighting that the chemotherapeutic potential of nucleoside analogues with an unnatural configuration is not limited to the antiviral field, since

some of these compounds have also proved to be active against other diseases, such as malaria or cancer.[109,123]

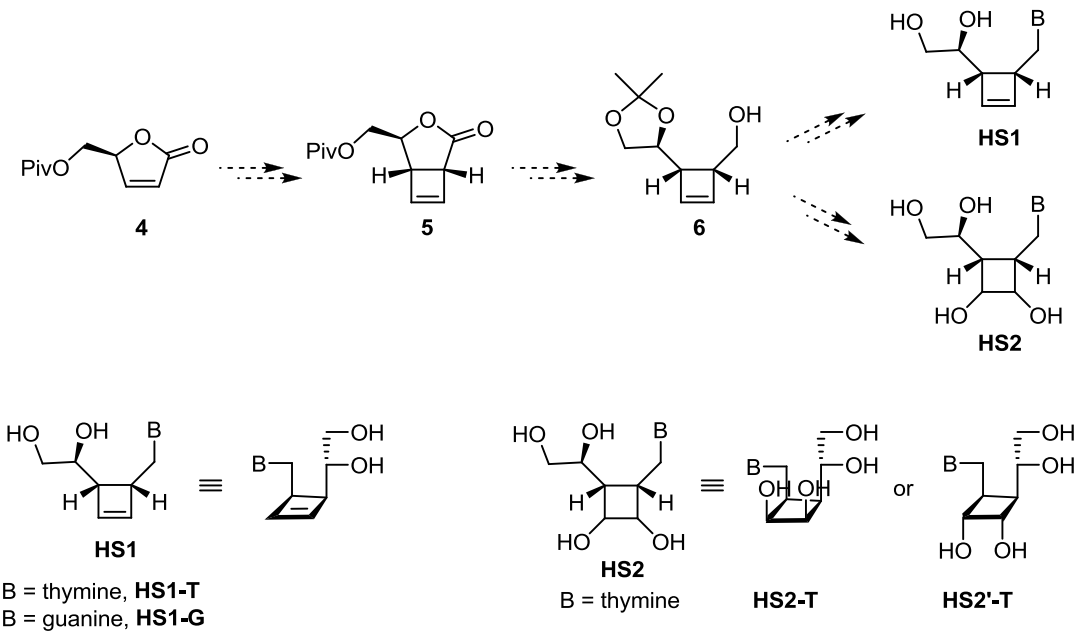
2. Objectives

The present thesis is focused on the development of new nucleoside analogues as antiviral agents, both from the synthetic and theoretical points of view. The three main objectives of this work as well as a short summary of the reasons that have encouraged us to undertake the investigations herein presented are described below.

- OBJECTIVE 1: synthesis of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues.

  Despite the important number of investigations dealing with the synthesis of carbocyclic nucleoside analogues, scarce efforts have been devoted to prepare cyclobutene nucleoside analogues. This fact, together with the antiviral potency showed by some L-nucleoside analogues, prompted us to work on the preparation of new cyclobutene L-nucleoside analogues, **HS1**, and the corresponding polyhydroxylated cyclobutane derivatives, **HS2** (Scheme 1). Note that in the nomenclature of the final nucleosides letters "HS" stand for "herpes simplex", since their mechanism of action as anti-HSV agents would be then studied by means of molecular modelling.

  The synthesis of all these compounds would start with a [2+2] photocycloaddition of the conveniently protected 2(5*H*)-furanone **4** to acetylene or a synthetic equivalent to provide bicyclic lactone **5**. Subsequent modifications of **5** would furnish alcohol **6**, which is the key intermediate to introduce the nucleobases. Removal of the protecting groups or functionalization of the cyclobutene ring would lead to the target nucleosides **HS1** and **HS2**, respectively.



Scheme 1. Synthetic pathway foreseen to prepare cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues.

Finally, these newly synthesized cyclobutene and polyhydroxylated cyclobutane nucleosides would be screened for antiviral activity against different viruses, such as HSV.

- OBJECTIVE 2: study of the mechanism of action of nucleoside analogues as antiviral agents by means of protein–ligand docking calculations to rationalize their antiviral activity results.

  In the last decade, several theoretical studies devoted to rationalize the antiviral activity of nucleoside analogues have been reported. However, most of these investigations are only focused on their first activation step or on the interaction with the target protein.

  This lack of studies on the whole activation process, along with the antiviral activity results obtained for nucleoside analogues previously prepared, encouraged us to investigate the molecular variables that are involved in their mechanism of action. With this purpose, we would apply a novel approach that is aimed at:

  i. Evaluating nucleoside analogues drug-likeness.
  ii. Analysing the activation process by which nucleosides are converted into the triphosphorylated derivatives.

iii. Investigating the interaction of the activated nucleosides with the target protein.

Two families of nucleoside analogues were planned to be investigated, leading to the following objectives:

- OBJECTIVE 2.1: rationalization of the antiviral activity against HIV of nucleoside analogues **HI1**–**HI3** previously prepared in our research group (Figure 14).

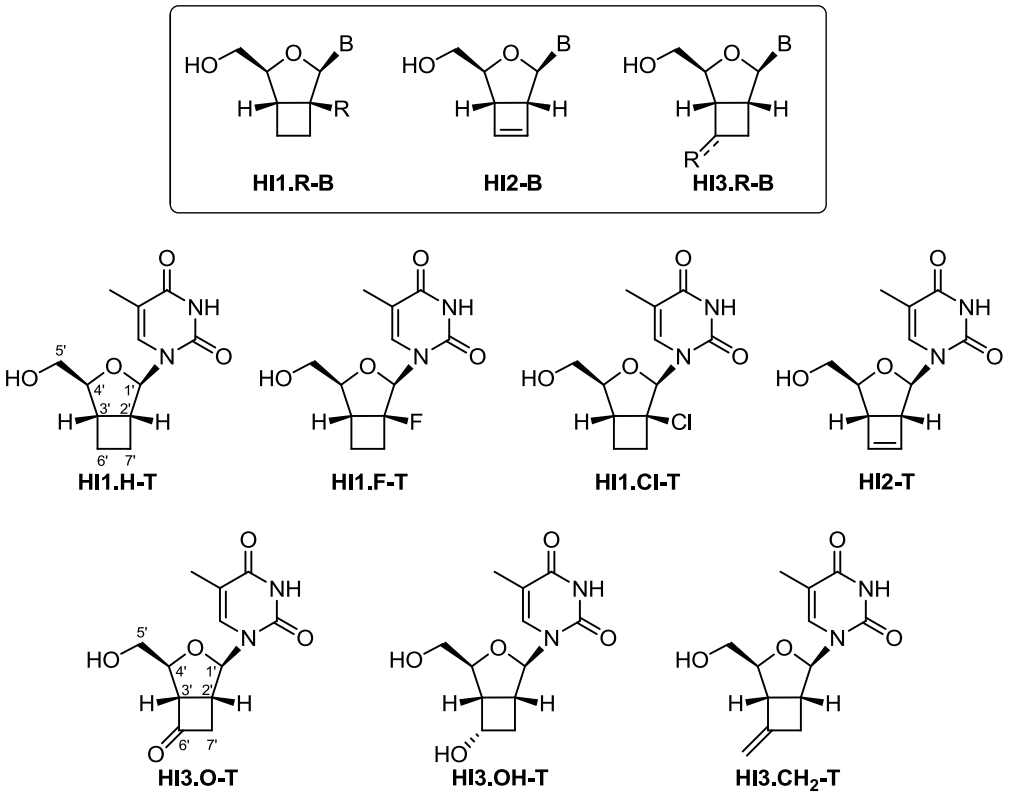


Figure 14. Nucleoside analogues to be studied against HIV.

Regarding the nomenclature, letters "HI" stand for "human immunodeficiency". The following letter(s) after the full stop denote the substituents at 2' position of nucleosides **HI1** and at 6' position of **HI3**. Finally, the last letter of the nomenclature indicates the base. This nomenclature will be used all along the manuscript.

- OBJECTIVE 2.2: rationalization of the antiviral activity against HSV of nucleoside analogues **HS1** and **HS2** synthesized in objective 1 (Figure 15).

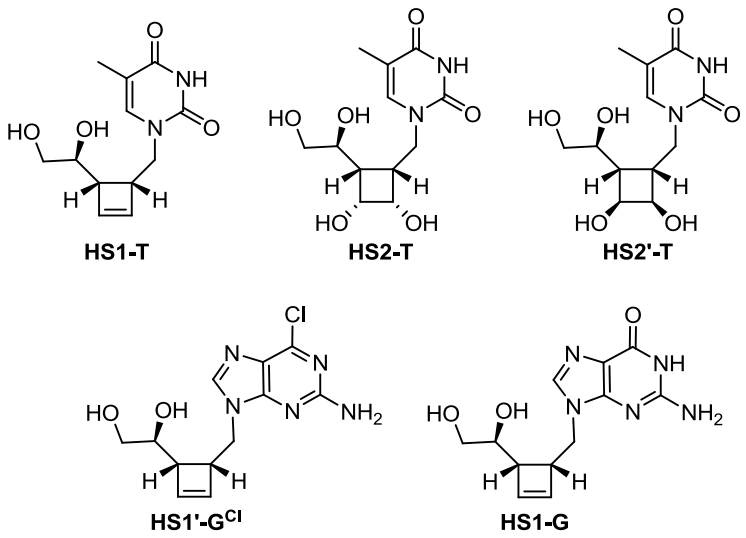


Figure 15. Nucleoside analogues to be studied against HSV.

- OBJECTIVE 3: rational design of novel carbocyclic nucleoside analogues.

  Structures of small compounds in complex with key enzymes, obtained by X-ray crystallography or derived from theoretical calculations, have been widely used for the rational design of biologically active compounds.

  Thus, taking advantage of all the information gathered in objective 2.2, we envisaged a rational design of carbocyclic nucleosides as anti-HSV agents based on these precedents. With this purpose, we aimed at:

  i. Designing a set of nucleoside analogues on the basis of a nucleoside-enzyme structure obtained by means of theoretical calculations.
  ii. Determining which nucleosides are more prone to be converted into the required triphosphorylated derivatives.

CHAPTER II

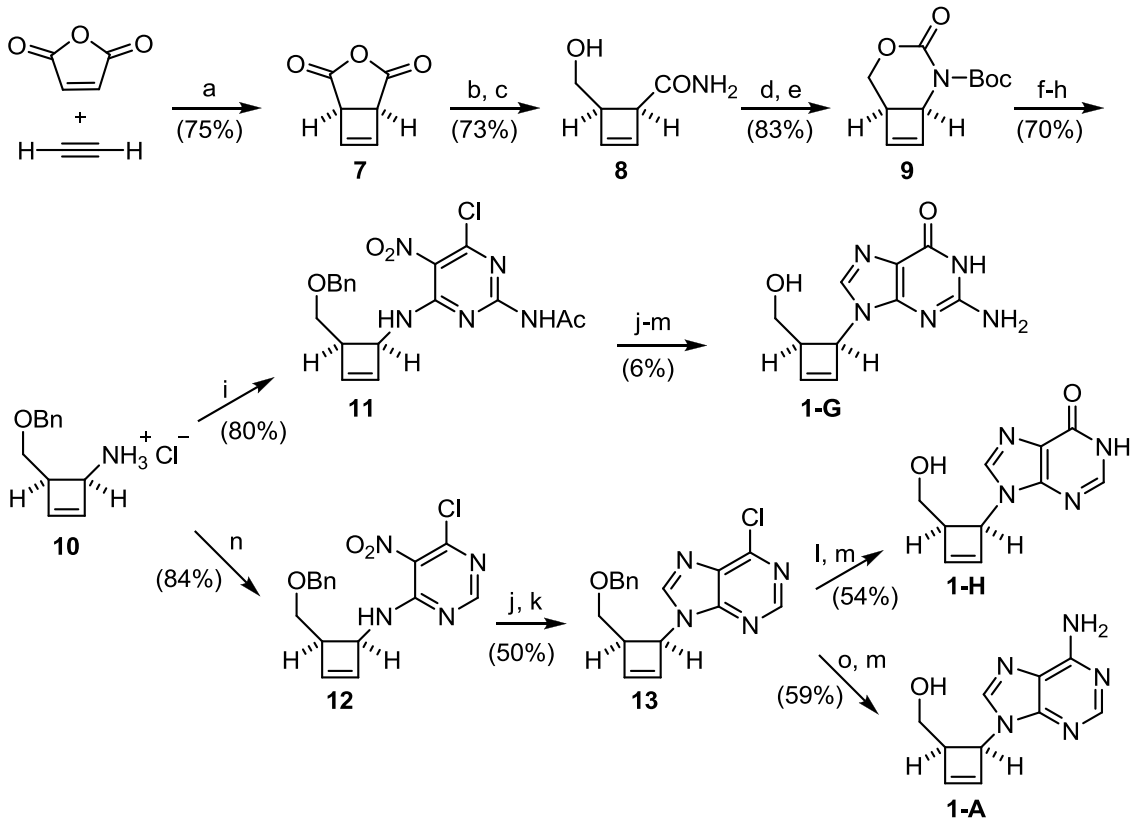
Synthesis of novel cyclobutene and cyclobutane L-nucleoside analogues

1. Introduction

1.1. Cyclobutene and polyhydroxylated cyclobutane nucleoside analogues. Precedents

As previously mentioned in the first chapter, the synthesis of cyclobutene nucleoside analogues has been scarcely studied. In fact, cyclobutene compounds with two substituents at the allylic position have been more used as intermediates to acyclic[124–126] and cyclic[127,128] dienes rather than prepared as target molecules.

The main drawback of cyclobutene compounds is the easy thermal electrocyclic ring opening, which leads to the corresponding dienes.[101,129–137] Despite the accurate conditions needed to synthesize cyclobutene compounds to avoid the thermal ring opening, Huet and co-workers published in 1997 and 1999 the synthesis of the racemic cyclobutene nucleosides **1-H**, **1-A** and **1-G**, envisaged as Carbovir analogues (Scheme 2).[101,102] The synthesis started with the [2+2] photocycloaddition of maleic anhydride to acetylene, which furnished the 3-cyclobutene-1,2-dicarboxylic anhydride, **7**. Its reduction to the corresponding lactone and the subsequent reaction with ammonia provided the cyclobutene derivative **8**. Then, Hofmann rearrangement followed by protection led to the cyclic carbamate **9**. Ring opening of its heterocyclic moiety and successive benzylation and treatment with hydrochloric acid yielded the key intermediate **10**, which led to the target nucleosides **1** after constructing the corresponding bases. Hence, cyclobutene analogues **1-G**, **1-H** and **1-A** were prepared in 13 steps and 1.5%, 7% and 8% overall yield, respectively. Disappointingly, the synthesized compounds did not show any significant activity against HIV.

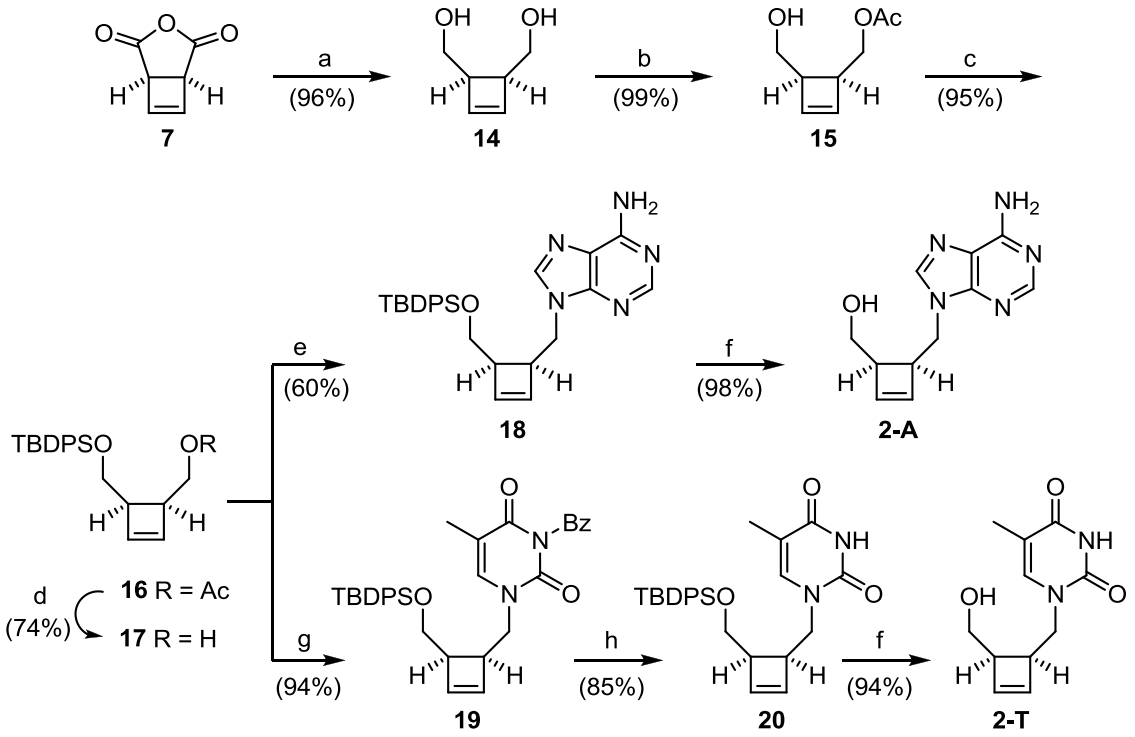


Reagents and conditions: (a) hν, EtOAc, acetophenone; (b) $NaBH_4$, THF; (c) NH_3, MeOH; (d) $PhI(OAc)_2$, MeOH, KOH, -5 °C to rt; (e) Boc_2O, Et_3N, 4-DMAP, THF; (f) LiOH, MeOH-H_2O 1:1, -10 °C; (g) BnBr, nBu_4NI, NaH, THF, -15 °C to 5 °C; (h) 3M HCl-MeOH, -5 °C to 15 °C; (i) *N*-acetyl-4-dichloro-5-nitropyrimidine, Et_3N, CH_2Cl_2; (j) $SnCl_2{\cdot}2H_2O$, EtOH, 60 °C; (k) $HC(OEt)_3$, HCl; (l) CF_3COOH-H_2O 3:1; (m) BCl_3, CH_2Cl_2, -78 °C; (n) 4,6-dichloro-5-nitropyrimidine, Et_3N, CH_2Cl_2; (o) NH_3, MeOH, 10-11 bars, 40 °C.

Scheme 2. Synthesis of cyclobutene nucleosides **1**, Huet and co-workers (1997, 1999).[101,102]

Later, the same authors suggested that the lack of activity of the prepared analogues **1** could be due to the shorter distance between the 5'-OH and the base on these nucleosides compared to the natural ones.[103] In order to evaluate this hypothesis, the enantiomerically enriched cyclobutene derivatives **2**, featuring a methylene spacer between the carbocycle and the base, were synthesized (Scheme 3). The new route started from the previously prepared cyclobutene derivative **7**, but after its reduction with $LiAlH_4$, the resulting *meso* diol **14** was desymmetrized by acetylation in the presence of *Pseudomonas fluorescens* lipase, which led to the enantiomerically enriched monoacetate **15**.[138] After protection of the alcohol of **15** as a silyl derivative and treatment with ammonia in methanol to remove the acetyl group, the alcohol **17** was obtained. Coupling of **17** with the corresponding bases via a Mitsunobu reaction and subsequent removal of the protecting groups provided the target cyclobutene nucleosides **2-A** and **2-T**. Thus, the purine analogue **2-A** was

prepared in 6 steps and 39% overall yield (from compound **7**), whereas the pyrimidine compound **2-T** was obtained in 7 steps and 50% overall yield (from **7**). None of the synthesized nucleosides displayed significant anti-HIV activity.

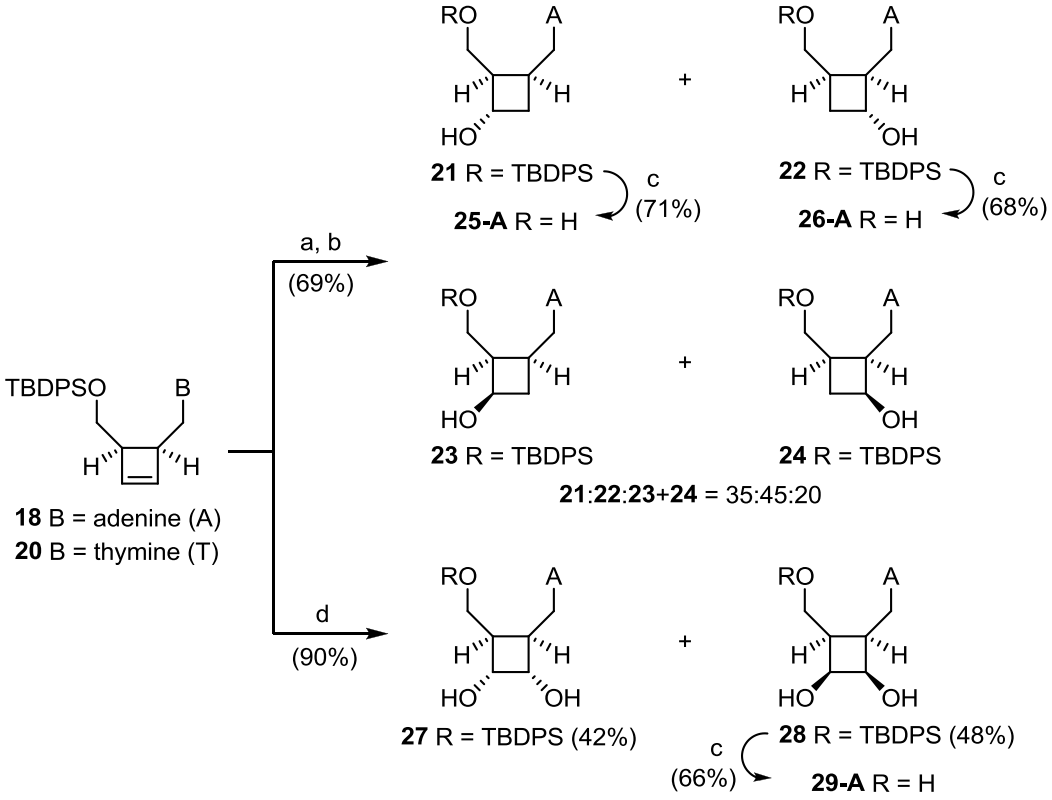


Reagents and conditions: (a) $LiAlH_4$, THF; (b) *Pseudomonas fluorescens* lipase (PFL)/vinyl acetate, -18 °C; (c) TBDPSCl, imidazole, DMF; (d) NH_3/MeOH; (e) PPh_3, DEAD, adenine, THF; (f) TBAF, THF; (g) PPh_3, DEAD, *N*3-benzoylthymine, THF; (h) NaOH.

Scheme 3. Synthesis of cyclobutene nucleosides **2**, Huet and co-workers (2002).[103]

More recently, the same research group envisioned the synthesis of polyhydroxylated derivatives from the previously described cyclobutene compounds **18** and **20** by taking advantage of the cyclobutene carbon-carbon double bond. Hence, in 2005 they reported the preparation of polyhydroxylated cyclobutane nucleoside analogues **25**, **26** and **29**, as shown in Scheme 4 for the adenine derivatives.[139] Thus, hydroboration of compound **18** provided a mixture of the four isomers **21**-**24**. The primary alcohol of the major products **21** and **22** was deprotected to deliver the dihydroxylated nucleosides **25-A** and **26-A**. On the other hand, dihydroxylation of **18** yielded a mixture of products **27** and **28**, which were separated by column chromatography. Cleavage of the silyl group of compound **28** afforded the trihydroxylated nucleoside **29-A**, whereas the deprotection of **27** failed and the corresponding nucleoside could not be obtained. The analogous thymine derivatives were synthesized following a parallel approach starting from the

intermediate **20**. Compound **25-T**, featuring a thymine base, proved to be effective against HSV-1 and HSV-2, whereas the rest of the synthesized analogues did not show significant activity against HSV-1 and HIV.



Reagents and conditions: (a) BH_3 ·THF, THF; (b) H_2O_2, NaOH; (c) HF-py, CH_2Cl_2; (d) OsO_4, NMO, THF/H_2O.

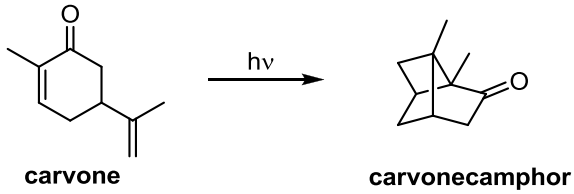
Scheme 4. Synthesis of polyhydroxylated cyclobutane nucleoside analogues **25-A**, **26-A** and **29-A**, Huet and co-workers (2005).[139]

Hence, in all these syntheses reported by Huet and co-workers the cyclobutene moiety has been constructed by the [2+2] photocycloaddition of maleic anhydride to acetylene. Indeed, [2+2] photochemical reactions have been extensively studied as a valuable synthetic tool to obtain a plethora of cyclobutane and cyclobutene compounds. In the next section, the discovery and applications of the [2+2] photocycloaddition reaction, its mechanism and some examples of [2+2] photochemical reactions of α,β-unsaturated lactones to unsaturated substrates will be briefly introduced.

1.2. [2+2] photocycloaddition

In 1908, Ciamician and Silber used the sun as energy source to induce modifications on a range of organic compounds such as carvone.[140] Although they

could not establish the structures of the resulting photoproducts unmistakably, this experiment is considered to be the discovery of the [2+2] photochemical reaction. Fifty years later, Büchi and Goldman identified the [2+2] cycloadduct obtained by irradiation of carvone with sunlight, which was called carvonecamphor (Scheme 5).[141] Ever since, the synthetic potential of the photochemical reactions started to be explored by Corey,[142,143] Eaton[144,145] and de Mayo,[146–148] which triggered a growing interest on this reaction from theoretical, mechanistic and synthetic points of view.



Scheme 5. Intramolecular [2+2] photocycloaddition of carvone.

Thus, [2+2] photocycloaddition of cyclic enones to unsaturated substrates has proved to be an effective tool to obtain a wide range of intermediates and compounds featuring a cyclobutane or cyclobutene ring on their structure,[149–153] such as annotinine, **30**,[154] caryophyllene, **31**,[143] fomannosin, **32**,[155] italicene, **33**,[156] sulcatine G, **34**,[157] grandisol, **35**,[158,159] and lineatin, **36**[160,161] (Figure 16).

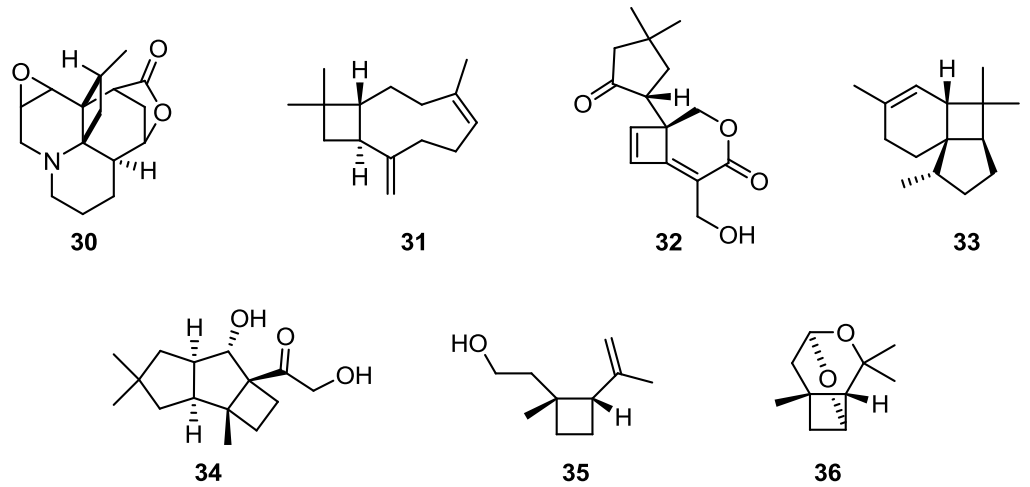
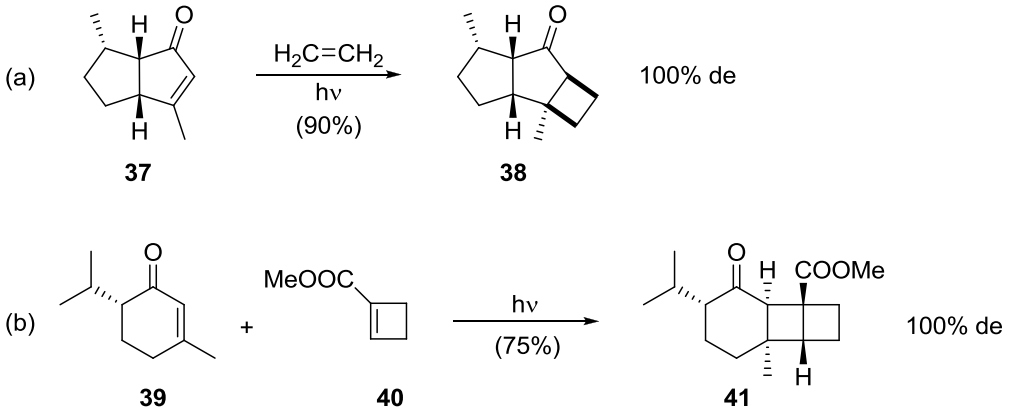


Figure 16. Natural products prepared by means of [2+2] photocycloaddition of enones to unsaturated substrates.

Moreover, the induction of stereoselectivity in this photochemical reaction has been broadly studied and applied to several stereoselective syntheses.[162–168] A stereogenic centre within the cyclic enone has been described to act as an adequate inductor of stereoselectivity, giving good facial diastereoselectivity in

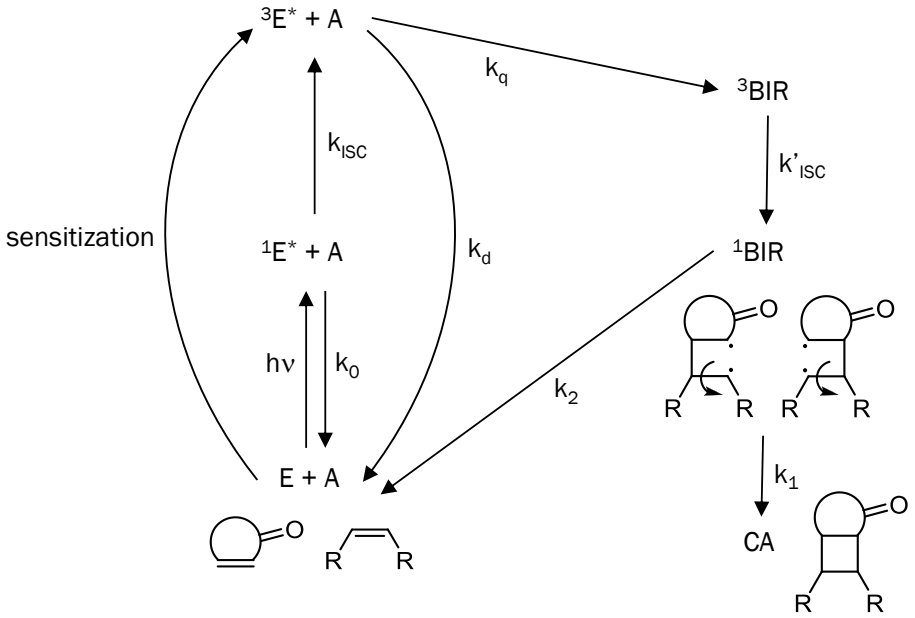
many cases, albeit its effectiveness depends on its position as well as on its substituents bulkiness. As an example, Piers and Orellana described the preparation of the tricyclic ketone **38** with a total stereoselectivity resulting from an exclusive approach of the ethylene from the less sterically hindered face of the enone **37** (Scheme 6).[169] The same behaviour was observed by Williams and Callahan when performing the [2+2] photocycloaddition of piperitone, **39**, to cyclobutene **40**, which yielded the tricyclic product **41** as a single diastereomer.[170]



Scheme 6. (a) [2+2] Photocycloaddition of the enone **37** to ethylene, Piers and Orellana (2001).[169] (b) [2+2] Photocycloaddition of the enone **39** to the cyclobutene **40**, Williams and Callahan (1980).[170]

1.2.1. Mechanism

In 1964, Corey and co-workers described [2+2] photocycloaddition reactions of cyclic enones to different substituted olefins.[142] On the basis of the experimental results, they proposed a mechanism according to which the formation of the final cycloadducts was controlled by electrostatic interactions between partial charges of the excited enone and the alkene. That mechanism was later supported by de Mayo.[147,148] However, it was not completely concordant with some experimental evidences. For this reason, an alternative hypothesis was soon postulated by Bauslaugh,[171] which was later reinforced by Schuster[172–174] and Weedon[175,176] studies leading to the nowadays accepted mechanism known as Bauslaugh-Schuster-Weedon (Scheme 7).[177]



Scheme 7. Bauslaugh-Schuster-Weedon mechanism for enone-alkene [2+2] photocycloaddition reactions. E: enone; A: alkene; BIR: 1,4-biradical; CA: cycloadducts.

According to this model, α,β-unsaturated carbonyl compounds are first excited to its singlet state ($^1E^*$), via an $n\pi^*$ or $\pi\pi^*$ excitation, which is converted into the excited triplet state ($^3E^*$) by an intersystem crossing process. It is also possible to promote an enone directly to its triplet state by using a photosensitizer, as long as the enone triplet energy is lower than the triplet sensitizer one. The excited triplet enone evolves to form the triplet 1,4-biradical (^{3}BIR), which after spin inversion leads to the corresponding singlet 1,4-biradical (^{1}BIR). This biradical may then revert to starting material or cyclize to provide the cycloadducts (CA).

It is noteworthy to mention that the relative configuration of cycloadducts resulting from [2+2] photochemical reactions involving 1,2-disubstituted olefins depends on biradical lifetimes. Thus, short lifetimes will prevent the rotation of the corresponding bond prior to the cycloadduct formation, leading to retention of the alkene configuration. By contrast, if the radical lifetime is long enough to allow free rotation, products without retention of the initial olefin configuration may be obtained.

1.2.2. [2+2] photocycloadditions of 2(5*H*)-furanones to alkenes and alkynes

In contrast to the large number of investigations dealing with [2+2] photocycloadditions of cyclic enones to unsaturated substrates published so far, the use of α,β-unsaturated lactones in such reactions has been scarcely

reported.[178–186] Despite the synthetic potential of the photocycloaddition reactions with these lactones to prepare a wide variety of four-membered ring fused γ- and δ-lactone derivatives, few studies have been devoted to their photochemical behaviour and induction of stereoselectivity in the [2+2] photocycloaddition.

Before going into the description of some of the investigations on [2+2] photocycloadditions of α,β-unsaturated lactones reported to date, it is important to establish the *anti/syn* nomenclature that will be used in the present work regarding the facial diastereoselectivity of this reaction. The *anti/syn* nomenclature describes the cycloadducts derived from the two different approaches of the unsaturated substrates to the diastereotopic faces of 5-substituted 2(5*H*)-furanones (Figure 17). Hence, an *anti* attack takes place when the unsaturated substrate approaches the furanone through its less hindered face whereas a *syn* attack is produced when approaches the contrary face, which contains the C-5 substituent.

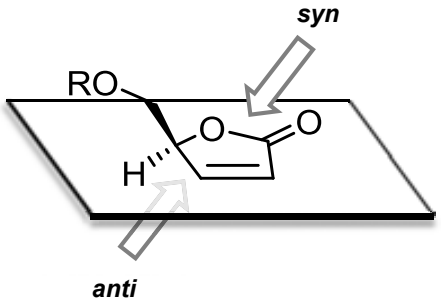
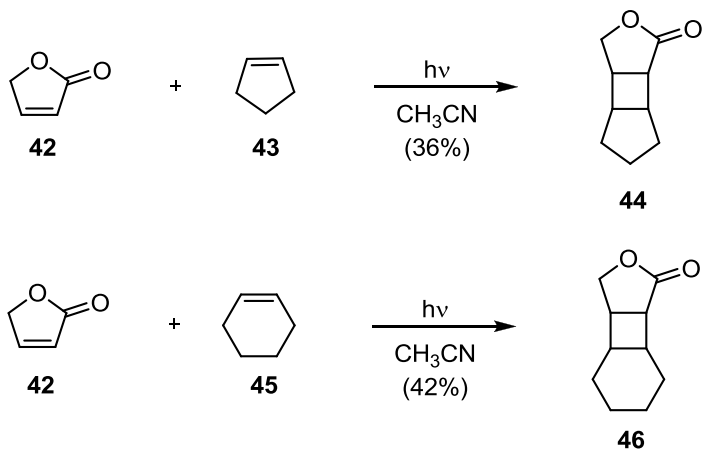


Figure 17. *Anti/syn* approaches of an unsaturated substrate to enantiopure 2(5*H*)-furanones.

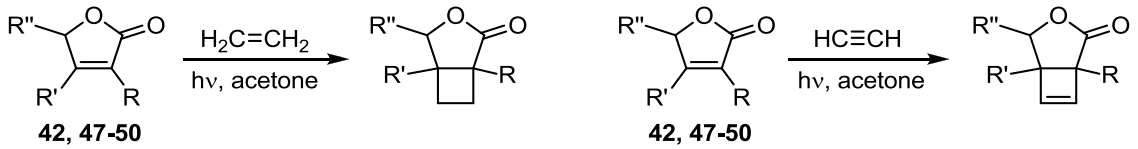
In 1972, Tada and co-workers published the first study on photochemical reactions of 2(5*H*)-furanones to olefins,[178] wherein the [2+2] photocycloaddition of crotonolactone **42** to cyclopentene and cyclohexene furnished the corresponding cycloadducts **44** and **46** in 36% and 42% yield, respectively (Scheme 8). It was also suggested that this photochemical reaction took place via the excited triplet state of the lactone.



Scheme 8. [2+2] Photocycloaddition of **42** to cyclopentene and cyclohexene, Tada and co-workers (1972).[178]

Kosugi et al. published in 1976 an article that has become a reference in the study of photochemical reactions of 2(5*H*)-furanones with unsaturated substrates.[179] The substituent effects of the 2(5*H*)-furanones **42** and **47**-**50** in the [2+2] photocycloaddition to ethylene and acetylene were investigated (Table 1). It was found that acetone, which appeared to play a role as a sensitizer, gave the best results. Interestingly, the cycloadducts derived from the photochemical reaction of 5-methyl-2(5*H*)-furanone (β-angelica lactone), **50**, with ethylene and acetylene were obtained as a 60:40 and 55:45 *anti*:*syn* diastereomeric mixture, respectively. It is worth recalling that the photocycloadditions with acetylene afforded the corresponding cyclobutene adducts from moderate to poor yields. The most significant features of the reactions were the slower reaction rates and yields compared to the reactions with ethylene, as well as the formation of a large amount of by-products at the end of the reaction periods. This behaviour is consistent with the capacity of strained cyclobutene derivatives to be excited and undergo many competitive reactions such as photoreductions or further cycloadditions.[187] Consequently, the progress of the reaction needed to be carefully monitored and stopped at appropriate reaction times.

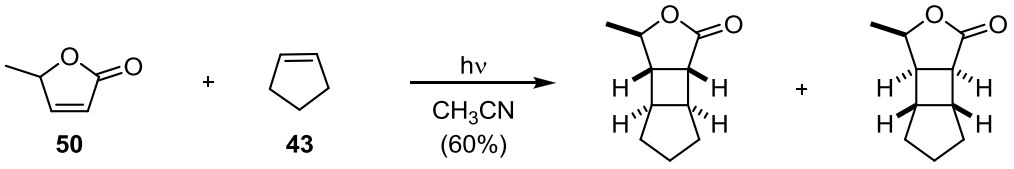
Table 1. [2+2] Photocycloaddition of 2(5*H*)-furanones to ethylene and acetylene, Kosugi and co-workers (1976).[179]



42, 47-50 **42, 47-50**

2(5*H*)-furanone	R	R'	R''	Ethylene		Acetylene	
				Yield (%)	*anti*:*syn*	Yield (%)	*anti*:*syn*
42	H	H	H	70	-	54	-
47	H	CH_3	H	44	-	29	-
48	H	Ph	H	49	-	8	-
49	Ph	H	H	52	-	23	-
50	H	H	CH_3	35	60:40	14	55:45

Simultaneously, Ohga and Matsuo reported the [2+2] photocycloaddition of the β-angelica lactone, **50**, to cyclopentene, which provided the corresponding cycloadducts in 60% yield and a 60:40 *anti*:*syn* ratio (Scheme 9).[180]

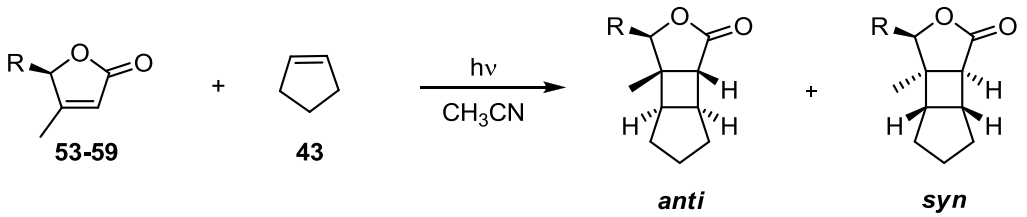


50 **43** **51** (***anti***) **52** (***syn***)

Scheme 9. [2+2] Photocycloaddition of **50** to cyclopentene, Ohga and Matsuo (1976).[180]

In the eighties, Koga and co-workers published the first [2+2] photochemical reactions using optically pure 2(5*H*)-furanones featuring different C-5 substituents to cyclopentene (Table 2).[181,182] From the experimental results, it was concluded that the efficiency in the diastereofacial differentiation correlated to the bulkiness of the substituents, obtaining the highest ratio (*anti*:*syn* 71:29) when using the pivaloyl moiety.

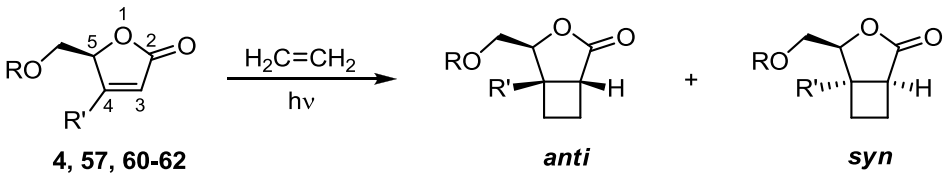
Table 2. [2+2] Photocycloaddition of 2(5*H*)-furanones to cyclopentene, Koga and co-workers (1989).[182]



2(5*H*)-furanone	R	Yield (%)	*anti*:*syn*
53	CH_2OCPh_3	29	60:40
54	CH_2OH	86	56:44
55	CH_2OMOM	80	50:50
56	$CH_2OCOCH(CH_3)_2$	62	62:38
57	CH_2OPiv	81	71:29
58	CH_2OTBS	64	68:32
59	$C(CH_3)_2OH$	57	65:35

A work published in 1991[188] was the beginning of a number of studies performed in our research group in the field of the [2+2] photocycloaddition reactions of enantiopure chiral 2(5*H*)-furanones to ethylene[189–192] and acetylene.[193,194] The main purpose of those studies was to get a deeper insight into the factors controlling the facial diastereoselectivity of these reactions. Thus, variables such as the 2(5*H*)-furanone substitution, solvent and filter were evaluated. The most representative results of the investigations on the [2+2] photocycloaddition of enantiopure chiral 2(5*H*)-furanones to ethylene are shown in Table 3.

Table 3. [2+2] Photocycloaddition of 2(5*H*)-furanones to ethylene, Font and co-workers (1991-1996).[188–192]



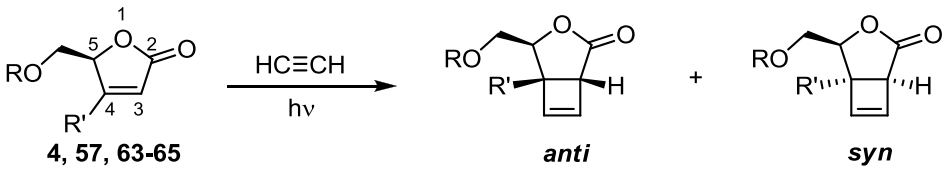
4, 57, 60-62

Entry	2(5*H*)-furanone	R	R'	Solvent	Filter	Yield (%)	*anti*:*syn*
1	**60**	H	H	acetone	pyrex	66	66:34
2	**61**	Ac	H	acetone	pyrex	82	74:26
3	**61**	Ac	H	ether	quartz	-	-
4	**4**	Piv	H	acetone	pyrex	59	78:22
5	**4**	Piv	H	ether	quartz	-	-
6	**62**	Ac	CH_3	acetone	pyrex	65	56:46
7	**57**	Piv	CH_3	acetone	pyrex	70	62:38

It was observed that the main product was always the *anti* cycloadduct, coming from the approach of the alkene through the least hindered face of the lactone. The highest *anti*:*syn* ratio was obtained when the C-5 hydroxymethyl group was protected as a pivalate (R=Piv, entry 4). Moreover, the presence of a methyl group at C-4 (R'=CH_3, entries 6 and 7) reduced the facial diastereoselectivity. On the other hand, the best results were reached by irradiation through a pyrex filter in acetone (entries 2 and 4 vs. entries 3 and 5, respectively).

As regards to [2+2] photochemical reactions with acetylene, our research group published in 2001 the first example using enantiopure 2(5*H*)-furanones as substrates (Table 4).[193]

Table 4. [2+2] Photocycloaddition of 2(5*H*)-furanones to acetylene, Font and co-workers (2001, 2003).[193,194]



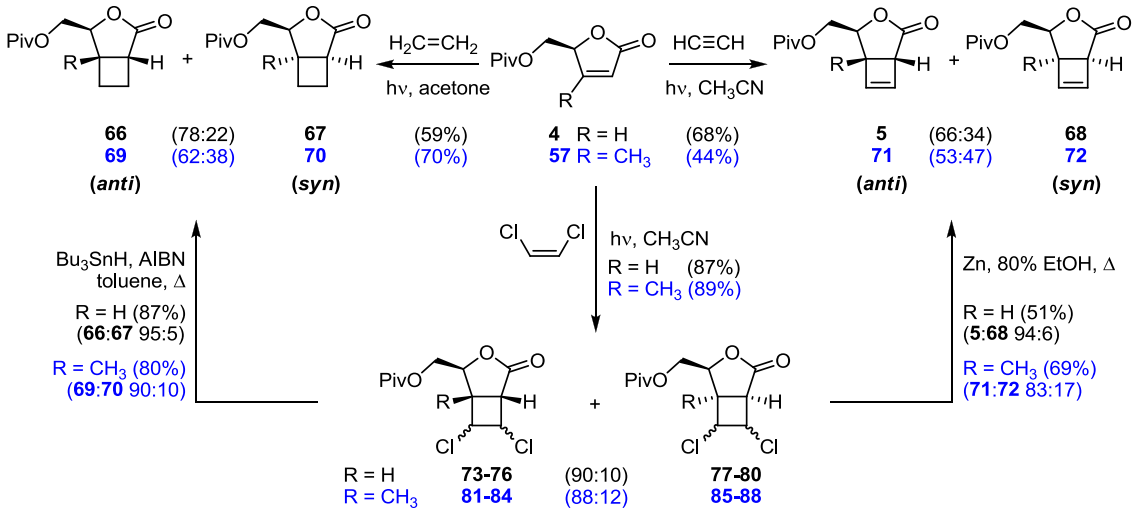
4, 57, 63-65

Entry	2(5*H*)-furanone	R	R'	Solvent	Filter	Yield (%)	*anti*:*syn*
1	**4**	Piv	H	acetone	pyrex	44	70:30
2	**4**	Piv	H	acetonitrile	quartz	68	66:34
3	**63**	CO_2Mnt	H	acetone	pyrex	42	66:34
4	**63**	CO_2Mnt	H	acetonitrile	quartz	52	59:41
5	**64**	Bz	H	acetone	pyrex	26	68:32
6	**64**	Bz	H	acetonitrile	quartz	24	66:34
7	**65**	TBDPS	H	acetone	pyrex	-	-
8	**65**	TBDPS	H	acetonitrile	quartz	-	-
9	**57**	Piv	CH_3	acetone	pyrex	32	54:46
10	**57**	Piv	CH_3	acetonitrile	quartz	44	53:47

As previously observed for the photochemical reactions with ethylene, the *anti* isomers were obtained as major products. However, the facial selectivity values and yields were, in general, lower than those obtained with ethylene (Table 4 vs. Table 3). Concerning the solvent and filter effects, better yields were achieved by irradiation through a quartz filter in acetonitrile, whereas slightly higher *anti*:*syn* ratios were obtained by using a pyrex filter and acetone (entries 1, 3, 5, 9 vs. entries 2, 4, 6, 10, respectively). Again, it was found that the 2(5*H*)-furanone **4**, bearing a pivaloyl group, gave the best facial diastereoselectivity (entry 1) and overall yield (entry 2). By contrast, the reaction proceeded less efficiently with protecting groups containing aromatic rings, such as benzoyl and *tert*-butyldiphenylsilyl (entries 5-8). The presence of a vinylic methyl group (R'=CH_3) was also detrimental for the course of the photocycloaddition, leading to the formation of the corresponding cycloadducts in lower diastereofacial selectivities and yields (entries 9 and 10 vs. entries 1 and 2, respectively).

The scarce facial selectivity and moderate yields obtained in the [2+2] photocycloadditions of enantiopure 2(5*H*)-furanones to acetylene, as well as the

hazardous character of the latter,[124,195] prompted our research group to pursue an alternative method to prepare the target cycloadducts. Hence, in 2003 it was published a highly stereoselective and efficient synthetic approach to both described cyclobutane and cyclobutene derivatives, consisting of a photochemical [2+2] cycloaddition of chiral 2(5*H*)-furanones to (*Z*)-1,2-dichloroethylene,[a] followed by either a dihydrodehalogenation or a dihaloelimination reaction (Scheme 10).[196]



Scheme 10. Methodology developed for the synthesis of cyclobutane and cyclobutene derivatives from 2(5*H*)-furanones **4** and **57**, and comparison with direct photocycloaddition to ethylene and acetylene, Font and co-workers (2003).[196]

The first step of this procedure was the [2+2] photocycloaddition of 2(5*H*)-furanones **4** and **57** to (*Z*)-1,2-dichloroethylene, in acetonitrile and irradiating through a quartz filter, that delivered a mixture of the dichlorocycloadducts in good yields (87-89%) and an excellent *anti*:*syn* ratio (90:10 for lactone **4** and 88:12 for lactone **57**). The next step was the dihydrodehalogenation reaction by treating the mixture of cycloadducts with tri-*n*-butyltin hydride and AIBN in refluxing toluene,[197] to afford the expected cyclobutanes **66**, **67** and **69**, **70** in excellent yields (87-80%). On the other hand, the reductive elimination of the vicinal chlorine atoms of the cycloadducts in the photochemical mixture was performed with activated Zn in refluxing 80% EtOH,[195,198–202] giving the corresponding cyclobutene derivatives **5**, **68** and **71**, **72** in 51% and 69% yield, respectively.

[a] Previous studies had been performed in our research group to investigate the influence of the stereochemistry of the alkene, showing that (*Z*)-1,2-dichloroethylene afforded the cycloadducts in higher yields than the (*E*)-isomer, while the *anti*:*syn* ratio was hardly affected by changing the olefin geometry.

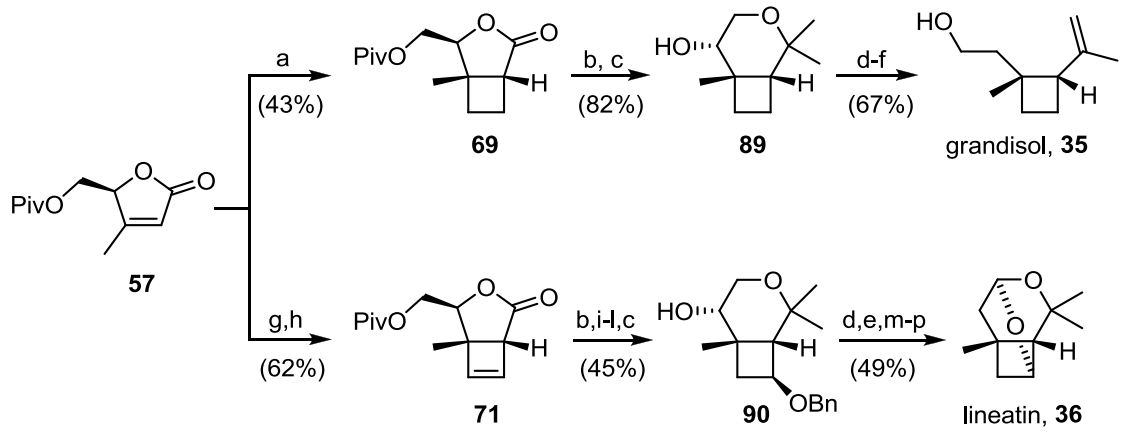
Therefore, the new approach represents a convenient and practical methodology to synthesize cyclobutane and cyclobutene derivatives. This two-step procedure significantly enhanced the facial diastereoselectivity of the photochemical reaction in comparison with the photocycloaddition to ethylene or acetylene. The easier manipulation of the safer and less cumbersome dichloroethylene instead of ethylene or acetylene allowed the scaling up of the reaction, making this new methodology more appropriate to prepare cyclobutane and cyclobutene compounds than the previous ones.

The results of all the above mentioned studies on [2+2] photocycloadditions of 2(5*H*)-furanones to unsaturated substrates will be taken into account in the development of a synthetic strategy to the target cyclobutene nucleoside analogues **HS1** and **HS2**.

1.3. Precedents in the research group using a [2+2] photocycloaddition as a key step in the synthesis of natural products and analogues

The experience acquired in our research group on the photochemical [2+2] cycloaddition of enantiopure 2(5*H*)-furanones to unsaturated substrates has been applied to synthesize naturally occurring cyclobutane pheromones, such as grandisol, **35**, and lineatin, **36**, as well as nucleoside analogues featuring a cyclobutane or cyclobutene moiety, starting thus a research program focused on these fields.

Hence, grandisol, **35**,[159] was prepared from 2(5*H*)-furanone **57**. The key step was the diastereoselective construction of the cyclobutane ring through the [2+2] photocycloaddition of **57** to ethylene, which mainly led to the *anti* cycloadduct **69**. Subsequent synthetic modifications of the bicyclic lactone **69** afforded the target compound in 6 steps and 24% overall yield (Scheme 11).



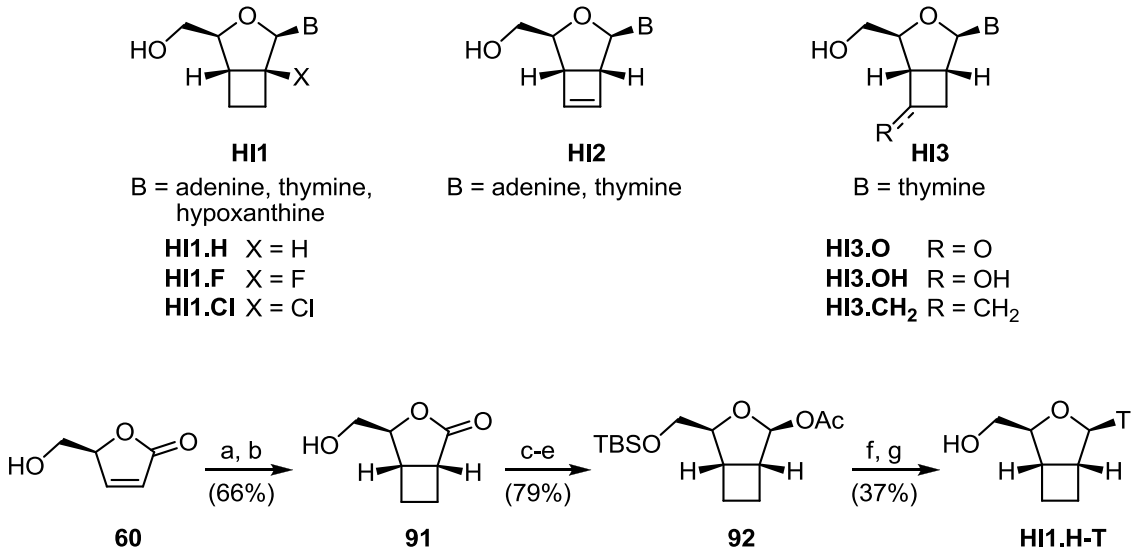
Reagents and conditions: (a) hν, ethylene, acetone; (b) MeLi, THF; (c) TsCl, DMAP, py; (d) TCDI, THF; (e) Bu_3SnH, AIBN, toluene; (f) LDA, hexane; (g) hν, (*Z*)-1,2-dichloroethylene, CH_3CN; (h) Zn, 80% EtOH, MW; (i) *p*-TsOH, acetone; (j) i. $Hg(OAc)_2$, THF-H_2O, ii. $NaBH_4$, NaOH; (k) NaH, BnBr, THF; (l) CF_3COOH, MeOH-H_2O; (m) H_2, Pd/C, EtOAc, AcOH; (n) Dess-Martin periodinane, CH_2Cl_2; (o) $RuCl_3$, $NaIO_4$, CCl_4-H_2O; (p) DIBAL-H, Et_2O, tartaric acid.

Scheme 11. Synthesis of grandisol, **35**, and lineatin, **36**, Font and co-workers (1996, 2004).[159,161]

Similarly, the synthesis of lineatin, **36**,[161] started with the photochemical [2+2] cycloaddition of furanone **57** to (*Z*)-1,2-dichloroethylene, followed by Zn promoted didehalogenation to provide the cyclobutene derivative **71**. Thus, pheromone **36** was prepared, via the intermediate **90**, in 14 steps and 14% global yield (Scheme 11).

Concerning the research program devoted to nucleoside analogues, cyclobutane-fused nucleosides as well as carbocyclic nucleosides containing a four-membered ring, such as Cyclobut-A, have been prepared.

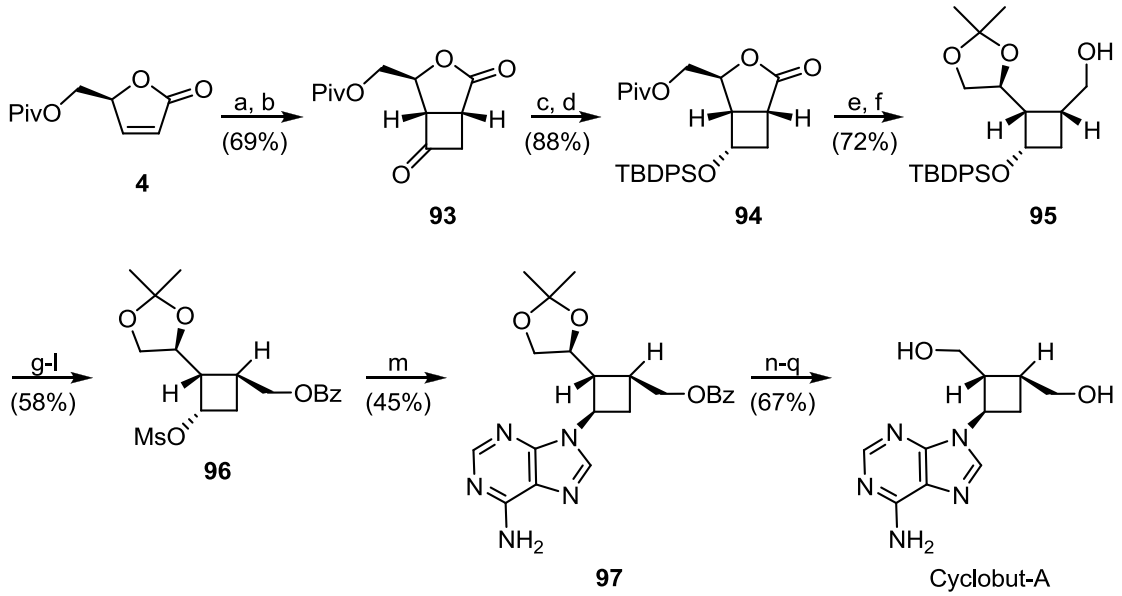
A novel family of cyclobutane- and cyclobutene-fused nucleoside analogues **HI1**–**HI3** (Scheme 12), designed as conformational mimics of the active 2',3'-didehydro-2',3'-dideoxynucleosides d4T and d4A, were synthesized.[203–205] The general synthetic pathway towards these compounds is exemplified for **HI1.H-T** nucleoside. Thus, [2+2] photocycloaddition of the 2(5*H*)-furanone **60** to (*Z*)-1,2-dichloroethylene, followed by a dihydrodehalogenation reaction with tri-*n*-butyltin hydride and AIBN in refluxing toluene, provided the bicyclic lactone **91**, which was next converted into the key intermediate **92**. Condensation of the acetate **92** with the nucleobase under Vorbrüggen glycosylation conditions and subsequent cleavage of the protecting groups yielded the target nucleoside **HI1.H-T**. The rest of the nucleosides **HI1**–**HI3** were prepared following the same photocycloaddition-acetylation-glycosylation pattern.



Reagents and conditions: (a) hν, (*Z*)-1,2-dichloroethylene, CH_3CN; (b) Bu_3SnH, AIBN, toluene, reflux; (c) TBSCl, imidazole, CH_2Cl_2; (d) DIBAL-H, toluene, -78 ºC; (e) Ac_2O, py; (f) thymine, *N*,*O*-bis(trimethylsilyl)acetamide, TMSOTf, CH_3CN; (g) TBAF, THF.

Scheme 12. Synthetic route towards cyclobutane- and cyclobutene-fused nucleoside analogues **HI1–HI3** exemplified for **HI1.H-T**, Alibés and co-workers (2006-2012).[203–205]

The synthesis of Cyclobut-A, which features potent anti-HIV activity, began with the regio- and diastereoselective [2+2] photocycloaddition of ketene diethyl acetal to the 2(5*H*)-furanone **4** (Scheme 13).[206] Subsequent hydrolysis of the acetal on the obtained adduct rendered the bicyclic derivative **93**. Further modifications led to lactone **94**, which was reduced and protected as an acetonide to afford **95**. The epimerization and protection of the primary hydroxyl group followed by the conversion of the secondary hydroxyl into a good leaving group delivered compound **96**. Next, the nucleobase was introduced by a classical nucleophilic substitution of the mesylate group, providing **97**. Removal of the protecting groups of **97**, oxidative cleavage of the vicinal diol and reduction of the resulting aldehyde led to the target nucleoside in 17 steps and 8% overall yield.



Reagents and conditions: (a) hν, ketene diethyl acetal, Et_2O; (b) *p*-TsOH, acetone, reflux; (c) L-Selectride, THF, -78 ºC; (d) TBDPSCl, imidazole, THF; (e) $LiAlH_4$, THF, 0 ºC; (f) acetone, $CuSO_4$, HCl (catalytic); (g) Dess-Martin periodinane, CH_2Cl_2; (h) Na_2CO_3, MeOH; (i) $NaBH_4$, MeOH; (j) BzCl, py, CH_2Cl_2; (k) TBAF, THF; (l) MsCl, Et_3N, CH_2Cl_2; (m) adenine, K_2CO_3, 18-C-6, DMF, 120 ºC; (n) CF_3COOH-H_2O; (o) $NaIO_4$, THF-H_2O.

Scheme 13. Synthesis of Cyclobut-A, Alibés and co-workers (2007).[206]

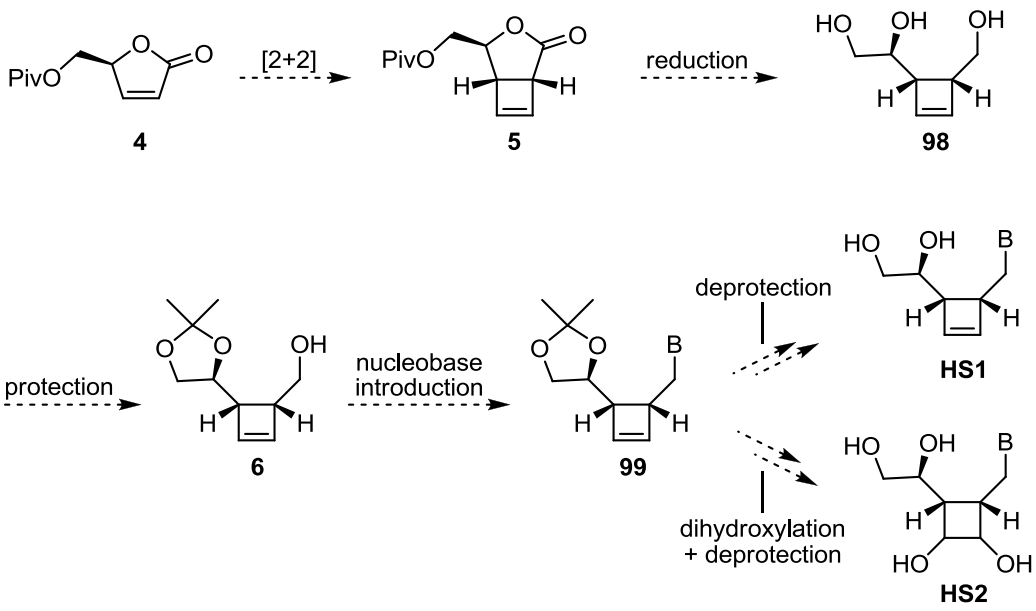
As part of this research project devoted to the synthesis of carbocyclic nucleoside analogues, the first goal of the present thesis is the preparation of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues.

2. Synthesis of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues

The interesting properties exhibited by many nucleoside analogues having an L-configuration, as well as the reduced number of reported examples dealing with cyclobutene nucleosides, encouraged us to focus on the preparation of cyclobutene L-nucleoside analogues and their polyhydroxylated cyclobutane derivatives. To the best of our knowledge, no examples of the synthesis of cyclobutene nucleosides having an L-configuration have been reported hitherto.

2.1. Synthetic strategy

The synthetic pathway towards the target L-nucleoside analogues was envisaged as depicted in Scheme 14.



Scheme 14. Synthetic pathway designed for the preparation of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues.

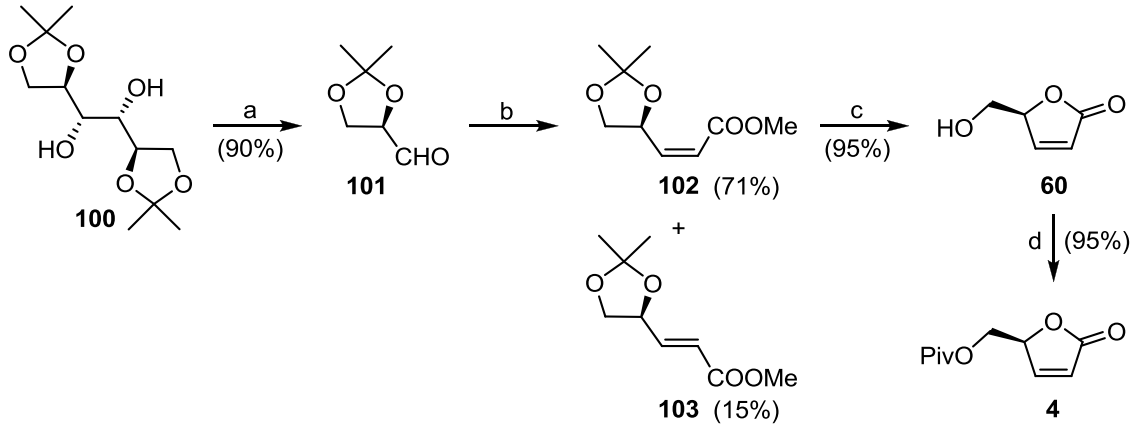
Taking advantage of the above described experience gained in the synthesis of similar compounds, the key step of this strategy would be the [2+2] photocycloaddition of the enantiopure 2(5*H*)-furanone conveniently protected **4** to (*Z*)-1,2-dichloroethylene, and subsequent Zn-promoted reductive elimination of the dichlorocycloadducts to provide bicyclic lactone **5**. At this point, our main objective is the optimization of the reaction conditions leading to lactone **5**, in order to achieve a multi-gram scale efficient synthesis of this valuable intermediate. Reductive opening of the lactone moiety of **5** followed by protection of the resulting vicinal diol should afford the key intermediate **6**, from which the nucleobases would be introduced. Finally, after removal of the protecting groups, the target nucleosides **HS1** would be obtained, whereas the functionalization of the cyclobutene ring and subsequent deprotections would provide the nucleoside analogues **HS2**.

2.2. Preparation of 2(5*H*)-furanone 4

2(5*H*)-furanone **4** was selected as a starting material to undertake the photochemical [2+2] cycloaddition as a result of the excellent diastereoselectivity induced by this compound in previous studies.[188–194,196,207]

Therefore, the first task of the present work was the preparation of the unprotected 2(5*H*)-furanone **60**, which was achieved following the synthetic procedure described by Mann and co-workers, and subsequent protection as a

pivaloyl to furnish 2(5*H*)-furanone **4** (Scheme 15).[208] This methodology allows its synthesis on a multi-gram scale and in satisfactory yield starting from the commercially available 1,2:5,6-di-*O*-isopropylidene-D-mannitol, **100**.



Reagents and conditions: (a) $NaIO_4$, THF-H_2O; (b) Ph_3P=CHCOOMe, MeOH; (c) H_2SO_4 (30%), MeOH; (d) PivCl, py, CH_2Cl_2.

Scheme 15. Synthesis of 2(5*H*)-furanone **4**.

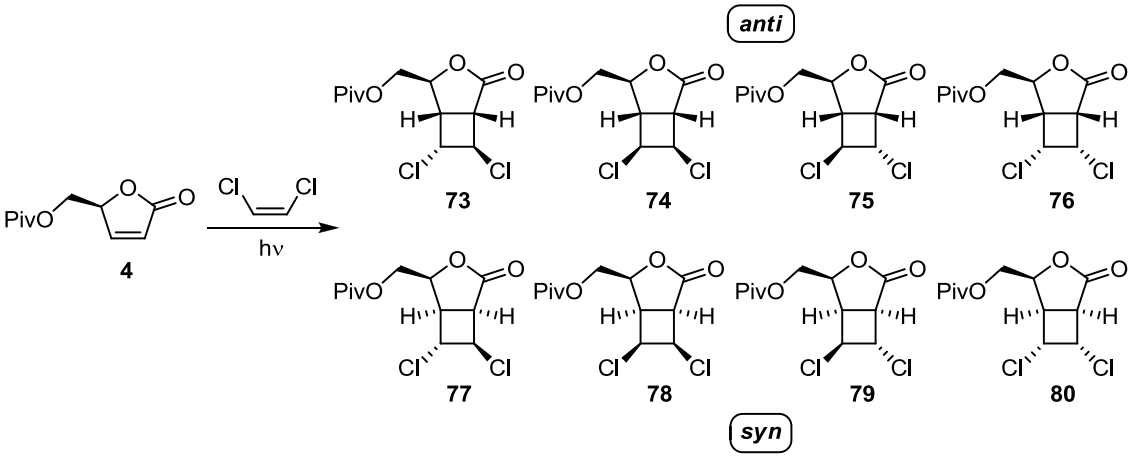
The synthetic pathway started with the oxidative cleavage of the D-mannitol diacetonide derivative **100** with sodium periodate to afford the 2,3-*O*-isopropylidene-D-glyceraldehyde, **101**. Reaction of aldehyde **101** with the stabilized phosphonium ylide Ph_3P=CHCOOMe in methanol at 0 ºC gave a 83:17 mixture of the (*Z*)- and (*E*)-alkenes **102** and **103** in 86% yield. The major isomer **102** was easily isolated by column chromatography and treated with acidic methanol to provide 2(5*H*)-furanone **60** in 95% yield, which was eventually converted into 2(5*H*)-furanone **4** by reaction with pivaloyl chloride and pyridine in CH_2Cl_2 in 95% yield (58% global yield from **100**).

2.3. Preparation of (1*R*,4*S*,5*S*)-4-pivaloyloxymethyl-3-oxabicyclo[3.2.0]hept-6-en-2-one, 5

2.3.1. [2+2] photocycloaddition of 2(5*H*)-furanone 4 to (*Z*)-1,2-dichloroethylene

As previously exemplified, facial diastereoselectivity and yields of photochemical [2+2] cycloadditions are strongly related to the solvent and filter used. Since our objective was to optimize the [2+2] photocycloaddition of 2(5*H*)-furanone **4** to (*Z*)-1,2-dichloroethylene, this reaction was carried out in the solvent previously used, acetonitrile, as well as in two other solvents, such as diethyl ether and acetone (Table 5).

Table 5. [2+2] photocycloaddition of 2(5*H*)-furanone **4** to (*Z*)-1,2-dichloroethylene.



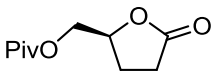
Entry	Lamp	Solvent	Filter	Time[a]	Yield (%)	73:74:75:76:*syn*	*anti*:*syn*
1[b]	125W	acetonitrile	quartz	4h 15'	87	34:19:14:23:10	90:10
2	125W	acetonitrile	quartz	6h	87	34:20:13:25:8	92:8
3	125W	diethyl ether	quartz	2h 30'	-	46:-:-:23:31	69:31
4	125W	acetone	pyrex	6h 10'	77	32:25:14:18:11	89:11
5	400W	acetonitrile	quartz	5h	88	32:18:13:23:14	86:14

[a] The reactions were monitored by GC. [b] Results previously obtained in our research group.[196]

Initially, the previously described experimental conditions[196] were evaluated (entry 1 vs. entry 2). Thus, irradiation of **4** and a 5 molar excess of (*Z*)-1,2-dichloroethylene through a quartz filter in acetonitrile afforded a mixture of the diastereomeric dichlorinated compounds **73**, **74**, **75**, **76** and ***syn*** in a 34:20:13:25:8 ratio and 87% global yield, where ***syn*** corresponds to a mixture of derivatives **77**-**80**. These results were consistent with the previous ones. Even though separation was unnecessary for synthetic purposes, the major cycloadducts had been isolated and completely characterized before.[209]

When the photochemical reaction was carried out in diethyl ether and irradiating through a quartz filter, compound **104**, resulting from the photoreduction of lactone **4**, was mainly obtained in 87% yield (Figure 18). This photoreduction reaction had been previously observed in our research group, and it is well described in the literature.[210,211] Some chlorinated cycloadducts were also formed

but in very small amounts, and it was not possible to recover them after purification of the crude by column chromatography (entry 3).



104

Figure 18. Photoreduced product obtained when performing the reaction in diethyl ether.

Irradiation of **4** and (*Z*)-1,2-dichloroethylene in acetone through a pyrex filter delivered a 32:25:14:18:11 mixture of the cycloadducts **73**:**74**:**75**:**76**:***syn*** in 77% global yield (entry 4). Hence, none of the assayed solvents improved the results previously obtained with acetonitrile and irradiating through a quartz filter.

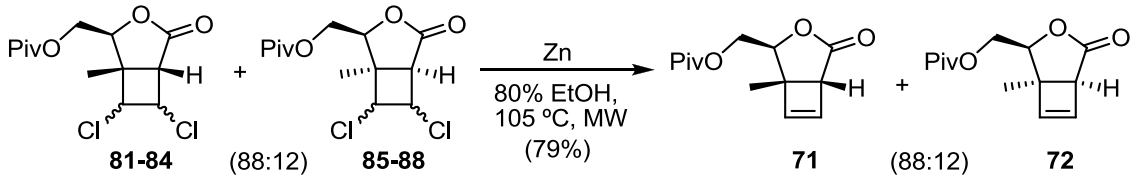
All the above mentioned experiments were performed in a scale of 100 mg of 2(5*H*)-furanone **4**. For synthetic purposes, the reaction was scaled up to 2 g of starting material, irradiating with a 400W medium pressure mercury lamp (entry 5). No major differences were observed regarding facial diastereoselectivity and yield.

Although the photochemical reaction crude could be directly used in the subsequent Zn-promoted dihaloelimination, it had been found that removal of 1,2-dichloroethylene by-products by column chromatography increased the efficiency of this reaction. However, considering that the chlorinated cycloadducts and photochemical by-products could be only detected by GC, this purification method was especially tedious and time-consuming. Therefore, it was decided to focus on finding a fast and appropriate purification process for this crude. Three different methods were assayed and their efficiency was evaluated by GC, comparing the crude chromatograms before and after its purification.

Thus, the crude was first dissolved in ethyl acetate and filtered over a pad of Celite®. The analysis by GC showed fewer by-products than before, but the crude was not clean enough to be used in the following reaction. The same results were obtained when filtering the crude through silica gel instead of Celite®. However, changing the solvent by hexane-ethyl acetate 1:1 and filtering the crude over a pad of silica gel led to a mixture of the major *anti* adducts **73**–**76** with a very small amount of by-products. It is worth noting that the minor *syn* cycloadducts **77**–**80** were lost in this purification process, so that the subsequent dihaloelimination was performed only over the *anti* dichlorinated compounds.

2.3.2. Reductive elimination of dichlorine derivatives 73–76

In our research group, Dr. Marta Racamonde studied the reductive chlorine elimination of cycloadducts **81–88**, previously performed with activated Zn in refluxing 80% EtOH (Scheme 10, page 42), under microwave irradiation.[161] Different reaction conditions were tested, obtaining the best results when the mixture of dichlorocycloadducts **81–88** was treated with Zn in 80% EtOH, in a sealed vessel at 105 ºC for 20 minutes, which furnished the target compounds in 79% yield (Scheme 16).



81-84 (88:12) **85-88** **71** (88:12) **72**

Scheme 16. Reductive elimination of chlorine cycloadducts **81–88** under microwave irradiation, Font and co-workers (2004).[161]

The outstanding decrease of the reaction time (from a 7 h reflux to 20 minutes of microwave irradiation), as well as the better yield (79%) compared to that achieved by classical thermal conditions (69%, Scheme 10), encouraged us to assay the dichloroelimination of our cycloadducts **73–76** under microwave irradiation.

Before presenting the results of the investigations on reductive chlorine elimination, the basics of the application of microwave heating in synthesis will be provided in the next subsection.

2.3.2.1. Microwave heating in organic synthesis

Using microwave irradiation in synthesis to accelerate chemical transformations has attracted much attention in recent years.[212–217] More than 2000 articles have been published in the area of microwave-assisted organic synthesis since the first reports on the use of microwave heating published in 1986.[218,219]

Although the early pioneering experiments in this field were performed in domestic microwave ovens, the current trend is to use commercial instruments equipped with good temperature control and pressure measurement, significantly enhancing the reproducibility of the experiments.

In the electromagnetic spectrum, microwave radiation region is located between infrared radiation and radio waves (Figure 19). Microwaves have wavelengths comprised between 1 mm and 1 m, corresponding to frequencies between 0.3 and 300 GHz. Importantly, the energy of the microwave photon in this frequency region is too low to break chemical bonds, preventing microwaves to induce chemical reactions, in contrast to ultraviolet and visible radiations. Accordingly, the usefulness of microwaves in organic synthesis is related to the efficient heating of materials by microwave radiations.

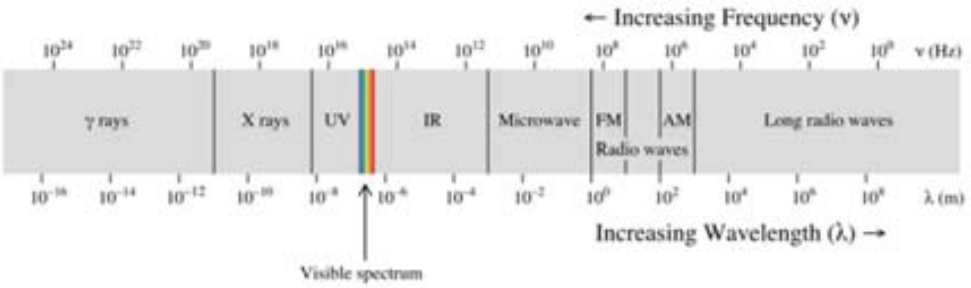


Figure 19. Electromagnetic spectrum.

The electric field component of microwave radiation is responsible for the dielectric heating of substances (solvent or reagent), exerted via two main mechanisms: dipolar polarization and ionic conduction. Both mechanisms are based in the alignment of dipoles or ions with the applied electric field when irradiating a sample at microwave frequencies. As the applied field oscillates, the dipole or ion field attempts to realign itself with the alternating electric field, causing energy to be lost, by molecular friction and collisions, in the form of heat.

Therefore, this phenomenon is dependent on the ability of a specific material to absorb microwave energy and convert it into heat. This ability is, in its turn, strongly related to the dielectric properties of each particular material, and it is determined by the so-called loss factor (tan δ). This loss factor is expressed as the quotient $\tan \delta = \varepsilon''/\varepsilon'$, where ε'' is the dielectric loss, which quantifies the efficiency with which the absorbed energy is converted into heat, and ε' is the dielectric constant, representing the ability of molecules to store electrical potential energy under the influence of an electric field. A reaction medium with a high tan δ value is thus required for efficient absorption and, consequently, for rapid heating. The loss factors, as well as boiling points, dielectric losses and dielectric constants of some common organic solvents are summarized in Table 6.[217,220]

Table 6. Boiling point, loss factor, dielectric constant and dielectric loss data for some common organic solvents (2.45 GHz, 20 °C).

Solvent	b.p. (°C)	tan δ	ε'	ε''
ethyleneglycol	197	1.35	37.0	49.950
ethanol	78	0.941	24.3	22.866
DMF	153	0.161	37.7	6.070
water	100	0.123	80.4	9.889
chlorobenzene	132	0.101	2.6	0.263
toluene	111	0.040	2.4	0.096
hexane	69	0.020	1.9	0.038

Undoubtedly, one of the main advantages of microwave irradiation over conventional heating systems is the achievement of dramatically reduced reaction times. While the external heat sources usually used in organic synthesis depend on thermal conductivity of the various materials that must be penetrated, microwave radiation passes through the walls of the reaction vessels and heats only the molecules (solvents, reagents, catalysts), producing an efficient internal heating. Thus, microwave irradiation raises the temperature of the whole volume simultaneously, producing a fast and homogeneous heating, whereas when using an oil bath the reaction mixture in contact with the vessel wall is heated first (Figure 20). Consequently, the heating profiles attained by microwave irradiation are very difficult to reproduce by standard thermal heating. Moreover, the monitoring mechanisms for temperature and pressure in modern microwave reactors allow for an excellent control of reaction parameters.

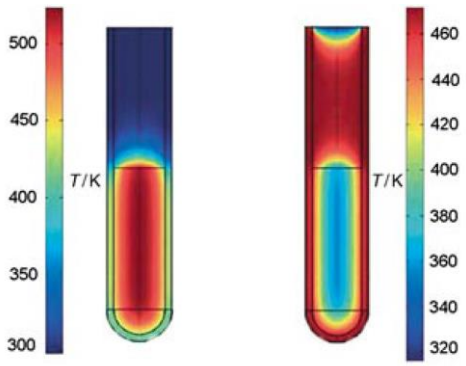


Figure 20. Difference in the temperature profiles after 1 minute of microwave irradiation (left) and treatment in an oil bath (right). (Reproduced by permission of reference 213 © 2004 JOHN WILEY AND SONS, Inc.)

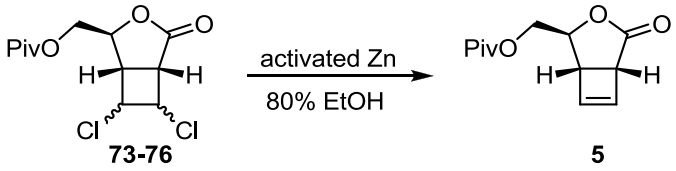
The present work was carried out with a CEM Discover® microwave instrument, which allowed us to set the following parameters: temperature, reaction time, power, maximum pressure and *PowerMAX™*.

Temperature stands for the set temperature that will be maintained during the *reaction time*. *Power* corresponds to the maximum power at which the sample will be irradiated, albeit the power may change in the course of the reaction. *PowerMAX™* is a technology that allows introducing more energy in the reaction (by increasing the power) while maintaining the temperature, by simultaneous cooling of the reaction with compressed gas on the outside of the reaction vessel. This resulted in significantly higher yields in many reactions.[221] In all the experiments performed in the present thesis, *power* was set at 250 W, maximum pressure at 250 psi and the *PowerMAX™* option was always enabled. All the reactions were performed in a sealed vessel.

2.3.2.2. Zn-promoted reductive elimination of dichlorine derivatives 73–76

With this in mind, reductive chlorine elimination of cycloadducts **73–76** with activated Zn in 80% EtOH was investigated. In order to optimize the reaction conditions, several experimental parameters, such as the amount of Zn, the stirring, the heating source, the temperature and the reaction time were evaluated (Table 7).

Table 7. Studied reaction conditions for Zn-promoted reductive elimination of cycloadducts **73–76**.



73-76 **5**

Entry	Zn eq.	73–76	Stirring	Heating	T (ºC)	Time[a]	Yield[b] (%)
1	37	100 mg	magnetic	MW	130	2x10'+50'	-
2	37	50 mg	magnetic	MW	120	10'+2x30'	-
3	37	200 mg	magnetic	MW	110	4x10'+40'	-
4	37	100 mg	magnetic	MW	95	20'+5'	47
5	37	100 mg	magnetic	MW	90	6x5'	34
6	37	70 mg	magnetic	MW	85	6x5'	24
7	25	130 mg	magnetic	oil bath	reflux	6h 30'	37
8	25	800 mg	magnetic	oil bath	reflux	7h 20'	30
9	25	1.6 g	mechanical	oil bath	reflux	7h	29
10	25	2.7 g	mechanical	oil bath	reflux	7h	22
11[c]	25	2.4 g	mechanical	oil bath	reflux	7h 30'	43

[a] All the reactions were monitored by GC. [b] Isolated yield after purification by column chromatography on silica gel. [c] Modified work-up process.

Dihaloelimination of chlorinated cycloadducts **73–76** was first attempted in a sealed vessel at 130 ºC under microwave irradiation (entry 1). Under these conditions, a small amount of cyclobutene derivative **5** was formed after the first 10 minutes run. However, attempts to drive the process to completion by further microwave irradiation resulted in the decomposition of cyclobutene **5**. The crude was purified by column chromatography, but the complexity of the resulting fractions did not allow the identification of the products. Considering that the decomposition could presumably be due to the thermal ring opening of the cyclobutene moiety, we decided to carry out the reaction at lower temperatures.

Despite the slower decomposition of cyclobutene **5** observed at 120 ºC and 110 ºC, a pure fraction of the target compound could not be obtained (entries 2 and 3). Then, the reaction was performed in a sealed vessel at 95 ºC, affording cyclobutene **5** in 44% yield after purification by column chromatography (entry 4). In

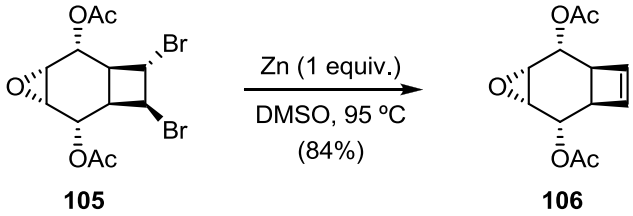
an attempt to improve the latter result, the reaction was carried out at 90 ºC and 85 ºC, but the desired product was afforded in lower yields (entries 5 and 6).

Considering that any of the conditions tested to perform the reductive elimination under microwave irradiation allowed an improvement of the results previously described in our research group (48% yield of cyclobutene **5**, Scheme 10), it was judged convenient to increase the scale of work by conventional thermal heating with an oil bath. Initially, 130 mg of chlorinated cycloadducts **73**–**76** were treated with 25 molar equivalents of Zn in refluxing 80% EtOH, obtaining the desired compound in 37% yield (entry 7), not fully reproducing the previous results. Nevertheless, we proceeded to scale up the reaction under these conditions. When performing the reductive elimination on a 800 mg scale, the formation of aggregates in the course of the reaction was observed, complicating considerably the stirring and the subsequent purification of the product, albeit cyclobutene compound **5** was finally obtained in 30% yield (entry 8).

To avoid the presence of those aggregates, magnetic stirring was changed by mechanical stirring, and the reaction was then scaled up to 1.6 g and 2.7 g of chlorinated cycloadducts (entries 9 and 10). Although the aggregates were no longer formed, lower yields were obtained when increasing the amount of starting material. At this point, it is worth mentioning that the reaction mixture was normally filtered over a pad of Celite® to remove most of the salts before purifying it by column chromatography. However, the copious amount of salts formed at this scale hampered the subsequent filtration of the crude. For this reason, and with the aim of avoiding the retention of the desired product in the salts, a different work-up process was assayed (entry 11). This consisted of an initial filtration of the reaction mixture through glass wool followed by concentration of the filtrate, which was then redissolved in dichloromethane and washed with 5% hydrochloric acid. Purification of the resulting crude by column chromatography delivered the target compound in 43% yield, improving thus all the aforementioned results accomplished with traditional heating.

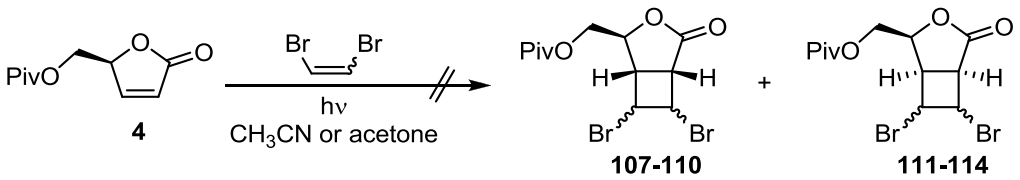
Unfortunately, there were no significant differences concerning the best yields reached with microwave and oil bath heating (entry 4 vs. entry 11). Considering that the reaction was scaled up to 2.4 g with conventional heating, this was considered the more suitable option to prepare the cyclobutene compound **5**.

On the other hand, some Zn-promoted reductive elimination reactions over brominated compounds had been described in the literature, affording the corresponding cyclobutene derivatives in moderate to high yields by using less equivalents of Zn.[222,223] For example, Balci and co-workers described the dihaloelimination of the brominated compound **105** with Zn (1 equivalent) in DMSO at 95 ºC, which provided the cyclobutene derivative **106** in 84% yield (Scheme 17).[223]



Scheme 17. Dihaloelimination of the brominated derivative **105**, Balci and co-workers (2005).[223]

This finding encouraged us to assay the [2+2] photocycloaddition of 2(5*H*)-furanone **4** to 1,2-dibromoethylene (mixture of isomers) in order to prepare the corresponding brominated cycloadducts (Scheme 18). Unfortunately, all attempts to carry out this photochemical reaction in acetonitrile and acetone met with failure, returning unaltered starting material.



Scheme 18. Attempt of [2+2] photocycloaddition of 2(5*H*)-furanone **4** to 1,2-dibromoethylene.

The *anti* relative configuration of cycloadduct **5** was confirmed by the value of its vicinal coupling constant between H-4 and H-5. According to Karplus equation, small coupling constants are expected for dihedral angles close to 90º.[224] Considering that the dihedral angle between H-4 and H-5 of the *anti* cycloadduct is close to 90º, the corresponding low value of the coupling constant ($J_{4,5}$=1.5 Hz) corroborates its *anti* relative stereochemistry (Figure 21).

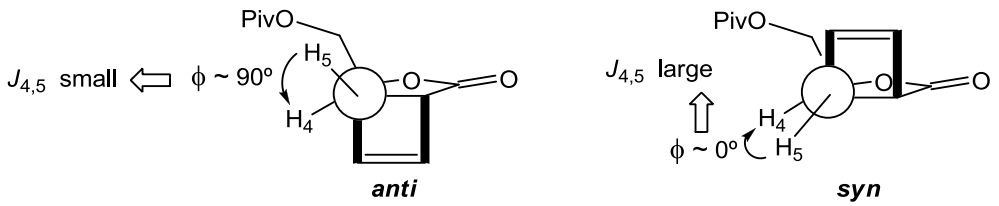
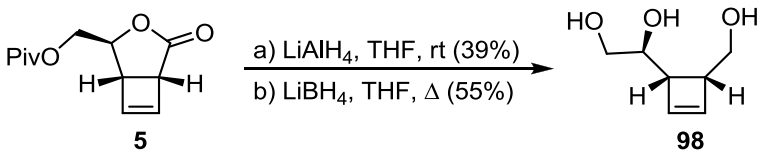


Figure 21. Dihedral angles and expected vicinal coupling constant magnitudes between H-4 and H-5 according to Karplus equation of the *anti* and *syn* isomers of compound **5**.

2.4. Preparation of {(1'*R*,4'S)-4'-[(4''S)-2'',2''-dimethyl-1'',3''-dioxolan-4''-yl]-2'-cyclobutenyl}methanol, 6

2.4.1. Reductive opening of lactone 5

Formation of triol **98** by complete reduction of bicyclic lactone **5** was initially accomplished with lithium aluminium hydride ($LiAlH_4$) in THF with yields around 39% (Scheme 19). However, the yields obtained using this methodology were barely reproducible and thus alternative reducing agents were considered. Reductive opening of **5** with lithium borohydride ($LiBH_4$) in refluxing THF was assayed, providing product **98** in 55% yield.



Scheme 19. Different reaction conditions investigated for the reductive opening of lactone **5**.

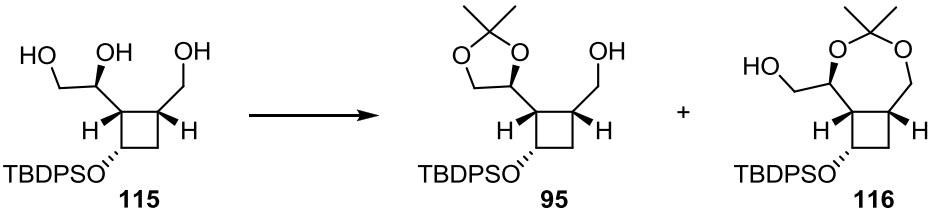
Despite achieving a slightly higher yield by using $LiBH_4$, this yield was still moderate, presumably due to the difficulties found in the purification of triol **98** by column chromatography. Therefore, to avoid the loss of product in this process, the crude resulting from the treatment of **5** with $LiBH_4$ in refluxing THF was directly used, after the reaction quenching, in the following step.

2.4.2. Protection of the vicinal diol of compound 98

Next step involved the protection of the vicinal diol of **98** as the isopropylidene acetal. A similar reaction had been previously investigated in our research group during the synthesis of Cyclobut-A.[206] In those studies it was found that in the protection of **115** to deliver **95**, variable amounts of the seven-ring acetonide **116**

were detected (Table 8). Different reaction conditions were assayed to minimize the presence of the undesired compound **116**.

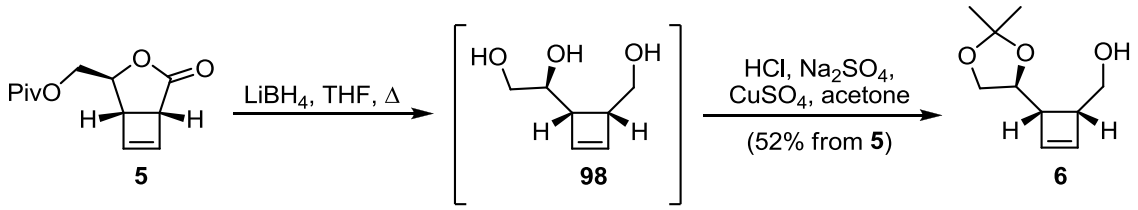
Table 8. Reaction conditions assayed to minimize the formation of the seven-ring acetonide **116**, Alibés and co-workers (2007).[206,209]



Entry	Reaction conditions	95:116
1	HCl, Na_2SO_4, acetone	71:29
2	CSA, 2,2-dimethoxypropane, acetone	91:9
3	HCl, Na_2SO_4, $CuSO_4$, acetone	97:3

The reaction was first carried out with acetone in the presence of catalytic hydrochloric acid and sodium sulfate, obtaining the desired compound **95** and the dioxepane derivative **116** in a 71:29 ratio (entry 1). When triol **115** was treated with camphorsulfonic acid (CSA), 2,2-dimethoxypropane and acetone, the ratio was increased up to 91:9 (entry 2). However, the best results were achieved by using anhydrous copper sulfate, anhydrous sodium sulfate, catalytic HCl and acetone,[225] which furnished a 97:3 mixture of the target compound **95** and the side product **116** in 85% yield (entry 3).

We have used these last reaction conditions to carry out the protection of the vicinal diol of triol **98**. Thus, treatment of **98** with acetone and catalytic HCl in the presence of anhydrous $CuSO_4$ and anhydrous Na_2SO_4 afforded, after purification by column chromatography on alumina, the target compound **6** in 52% overall yield from cycloadduct **5** (Scheme 20). The formation of the corresponding dioxepane by-product was not observed. It is worth recalling that alcohol **6** is a volatile compound, and thus great care should be taken during the evaporation of the solvents.



Scheme 20. Preparation of alcohol **6** from cycloadduct **5**.

The presence of the dioxolane in the isolated compound was confirmed by the ^{1}H-^{13}C HMBC spectrum, which showed a cross-peak between C-2'' and H-5'' (Figure 22). This assignment was also supported by the higher chemical shift value of protons H-5'' of compound **6** (4.11 and 3.71 ppm) compared to those of triol **98** (3.76 and 3.53 ppm).

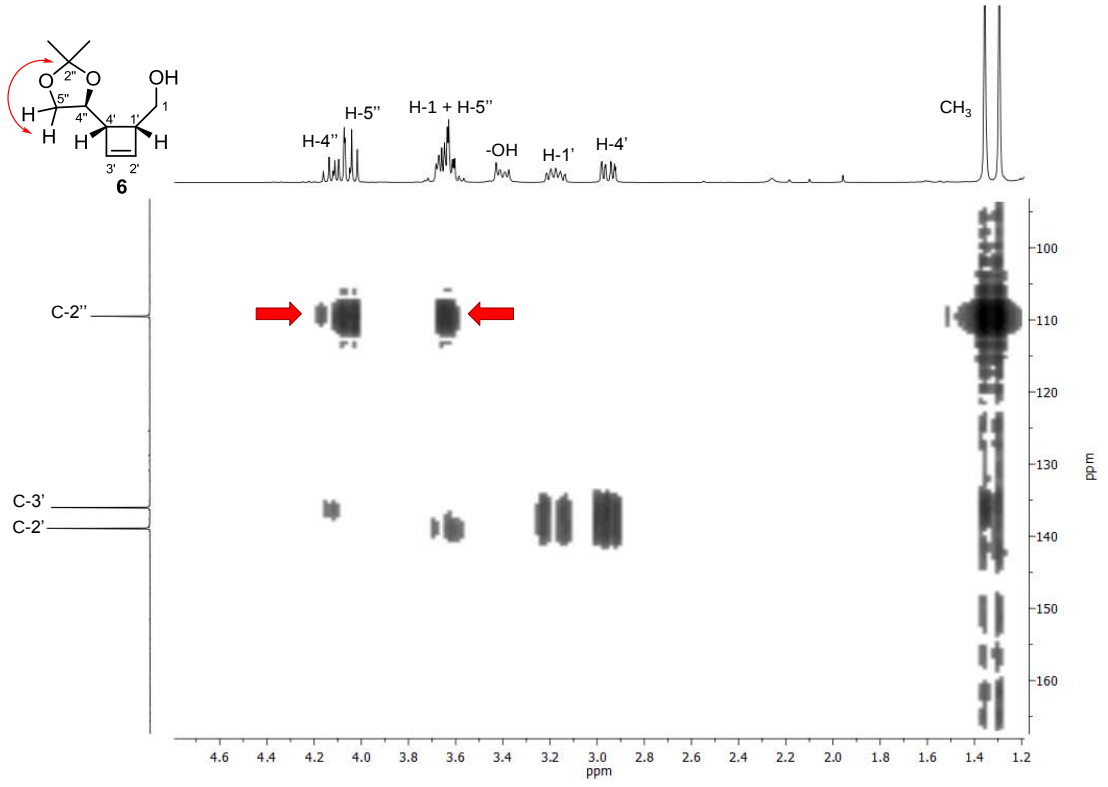
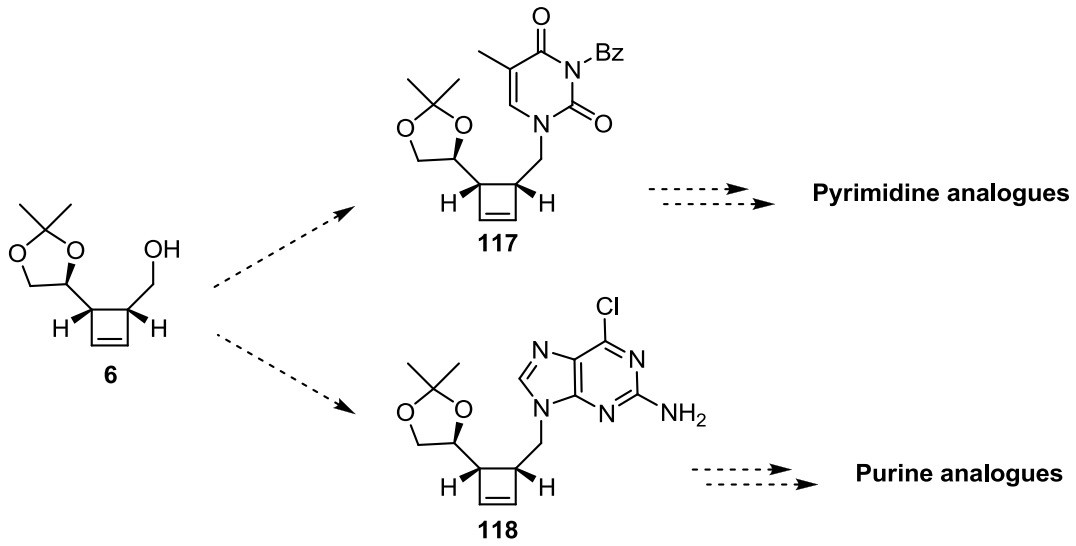


Figure 22. Detail of the HMBC experiment of **6** (250 MHz, $CDCl_3$).

This alcohol **6** was the key intermediate to introduce the nucleobases. Hence, the next section provides a brief overview on some methodologies used to introduce the bases, and the following sections will be devoted to present our experimental results.

2.5. Methodologies for the nucleobase introduction

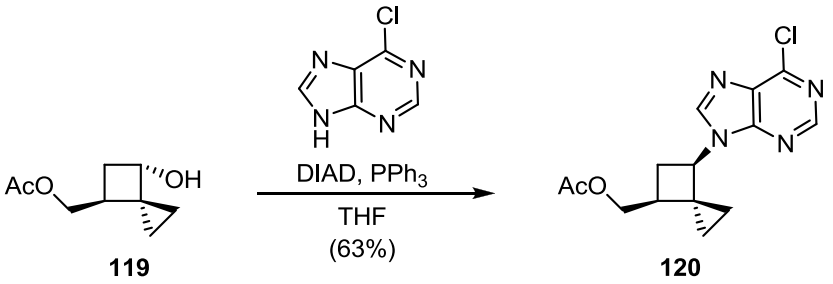
A divergent strategy from alcohol **6**, involving the introduction of a pyrimidine and a purine base, was envisaged to prepare nucleoside analogues featuring different nucleobases (Scheme 21).



Scheme 21. Foreseen divergent strategy to prepare nucleosides featuring different nucleobases.

The introduction of nucleobases in the field of carbocyclic nucleosides synthesis has been focused on two fundamental approaches: coupling of the heterocyclic base with an appropriately functionalized carbocyclic ring or construction of the heterocycle from an amine substituent present on the carbocycle.[90] Our synthetic strategy was based on the first approach, since the introduction of nucleobases was envisaged via a nucleophilic displacement of the alcohol, either directly using a Mitsunobu methodology or first converting the alcohol into a better leaving group (mesylate or tosylate) and then performing the subsequent substitution.

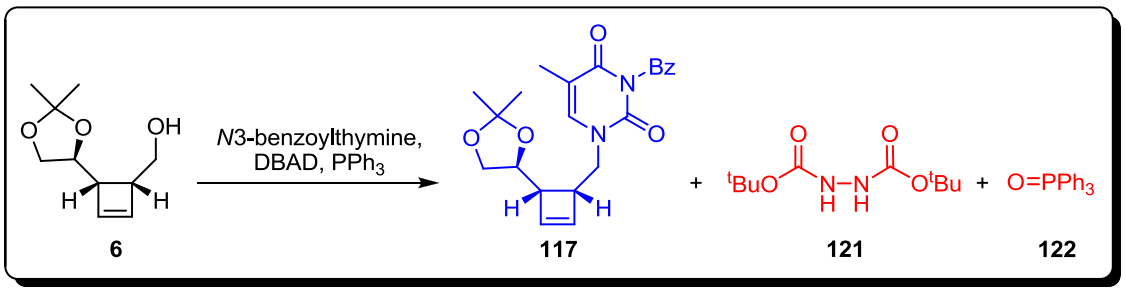
The substitution of alcohols with nucleophiles mediated by a combination of a trialkyl- or triarylphosphine and a dialkyl azodicarboxylate was described by Mitsunobu in 1981.[226–228] Nowadays, the Mitsunobu reaction has become a versatile synthetic method, being one of the most widely used methodologies for the introduction of nucleobases in the synthesis of carbocyclic nucleosides.[229,230] For instance, in the synthesis of spiro[2.3]hexane nucleosides reported by Chu and co-workers, the nucleobase was introduced via a Mitsunobu reaction with a 63% yield (Scheme 22).[231]

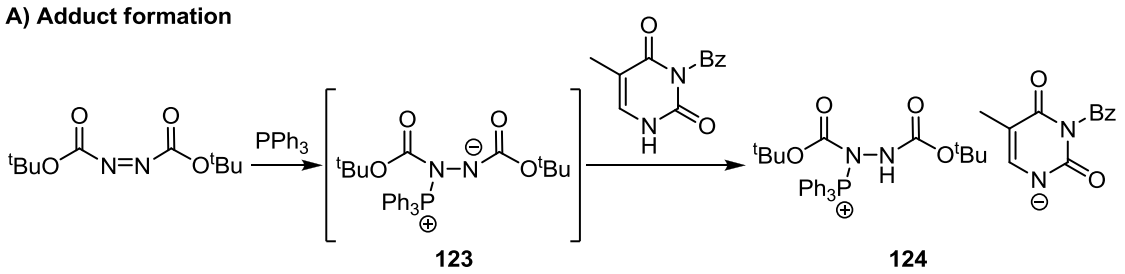


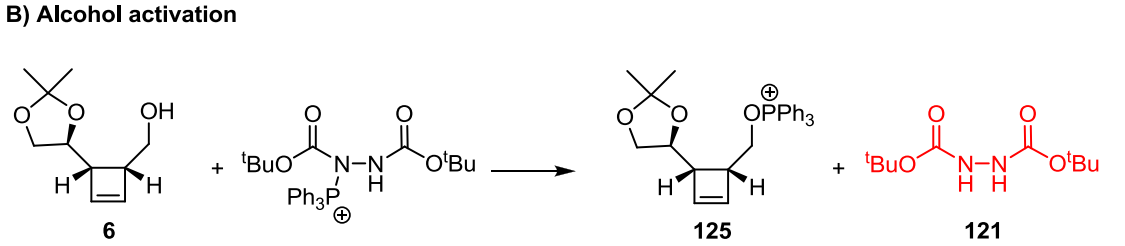
119 120

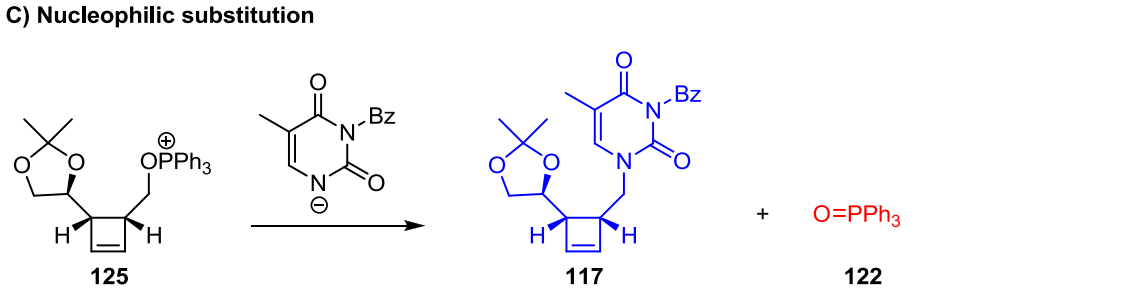
Scheme 22. Mitsunobu coupling of 6-chloropurine to compound **119**, Chu and co-workers (2004).[231]

However, it is generally accepted that one major drawback of the Mitsunobu reaction is the difficult isolation of the product from the reaction mixture.[232] This limitation is due to the fact that besides the two reactants, Mitsunobu reaction requires the use of stoichiometric quantities of two other reagents each of which produces a by-product. The reaction mechanism, which is illustrated in Scheme 23 with alcohol **6**, *N*3-benzoylthymine, di-*tert*-butyl azodicarboxylate (DBAD) and triphenylphosphine (PPh_3) as representative reagents, accounts for the formation of these side products.







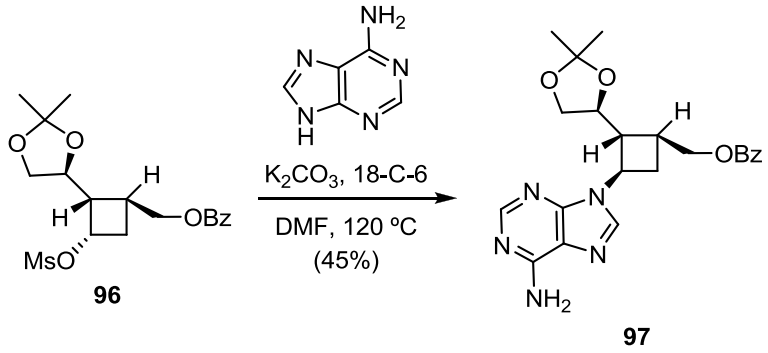


Scheme 23. Simplified Mitsunobu mechanism illustrated for the coupling of *N*3-benzoylthymine to alcohol **6**.

Thus, the first step in the Mitsunobu reaction is the nucleophilic addition of triphenylphosphine to DBAD, yielding the corresponding betaine **123** that deprotonates the nucleobase. The alcohol is then activated by the reaction with the DBAD-triphenylphosphine adduct **124**, which results in the formation of the hydrazine **121**. Nucleophilic displacement of triphenylphosphine oxide from **125** by the deprotonated base completes the reaction to provide the target compound **117** and the second side product, triphenylphosphine oxide, **122**.

To overcome the above mentioned limitation of the Mitsunobu reaction, another method commonly used for the coupling of nucleobases to carbocycles is

the nucleophilic substitution of a previously activated alcohol, for instance in the form of mesylate.[98,233–236] As an example, this methodology was used in our research group to introduce the nucleobase in the synthesis of Cyclobut-A, affording the corresponding product in moderate yield (Scheme 24).[206]



Scheme 24. Nucleophilic substitution of mesylate **96**, Alibés and co-workers (2007).[206]

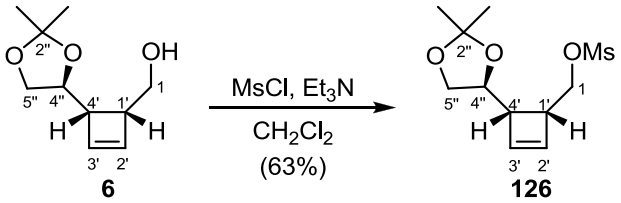
Therefore, starting from alcohol **6**, either its conversion into a good leaving group followed by a nucleophilic displacement or the Mitsunobu reaction were plausible methodologies to introduce the nucleobases.

2.6. Preparation of nucleoside analogues featuring a pyrimidine base

2.6.1. Nucleophilic substitution of the mesylate derivative of alcohol 6

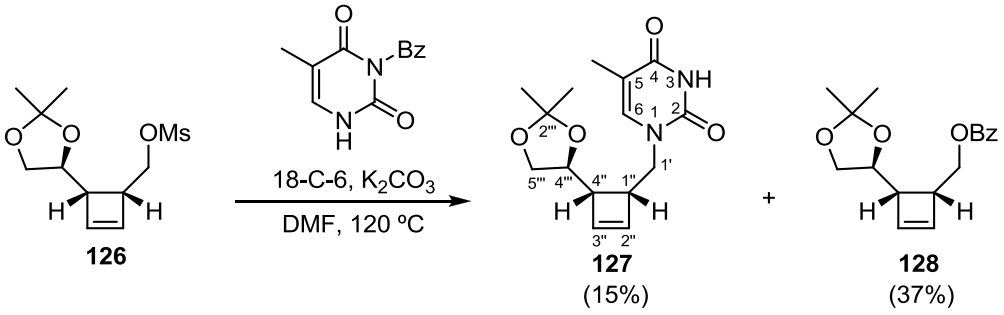
Considering the possible difficulties in the isolation of products coming from the Mitsunobu reaction, the introduction of thymine was initially attempted by the nucleophilic substitution of the corresponding mesylate.

To prepare the mesylate derivative **126**, alcohol **6** was treated with mesyl chloride and triethylamine in dichloromethane (Scheme 25). Purification of the resulting crude by column chromatography furnished, in 63% yield, a yellowish oil that was identified by its spectroscopic data as the mesylate **126**. The formation of **126** was confirmed by its characteristic methyl signal in ^{1}H-NMR and ^{13}C-NMR spectra at 3.03 and 37.5 ppm, respectively, as well as by the displacement of H-1 and C-1 signals to higher chemical shifts (H-1: from 3.71 ppm in **6** to 4.32 and 4.55 ppm in **126**; C-1: from 62.0 ppm in **6** to 70.1 ppm in **126**).



Scheme 25. Preparation of the mesylate derivative **126** from alcohol **6**.

The following step was the introduction of the nucleobase, which was initially carried out with thymine. Since both *N*1 and *N*3 nitrogen atoms of thymine may act as nucleophiles, protection of the *N*3 position should be carried out. This *N*3 protected nucleobase had been prepared following the procedure reported by Reese and co-workers, consisting of the dibenzoylation of thymine and the subsequent selective cleavage of the benzoyl group at *N*1.[237,238] Then, nucleophilic displacement of mesylate **126** was performed with *N*3-benzoylthymine employing our previously described conditions.[206] Accordingly, mesylate **126** was treated with K_2CO_3, 18-crown-6 and *N*3-benzoylthymine in DMF at 120 ºC (Scheme 26). The resulting crude was purified by column chromatography, obtaining a mixture of compound **127** (15%), which incorporated the deprotected thymine, and the by-product **128** (37%), indicating that deprotection of the base had occurred during the reaction. The benzoyl **128** likely arose from the nucleophilic displacement of the mesylate by a benzoate resulting from the deprotection of the *N*3-benzoylthymine. This could partially account for the low yield of compound **127**.



Scheme 26. Nucleophilic displacement of mesylate **126** with *N*3-benzoylthymine.

The introduction of thymine on compound **127** was observed in the ^{1}H-NMR spectrum by the disappearance of the methyl singlet peak of the mesylate along with the presence of two new peaks at 7.14 and 1.92 ppm corresponding to H-6 and C*H*$_3$-C-5 of the nucleobase, respectively.

The site of attachment of thymine through *N*1 on **127** was assigned on the basis of the corresponding ^{1}H-^{13}C HMBC experiment, which showed correlation between the aromatic proton H-6 and C-1' together with correlation between proton H-1' and C-6 (Figure 23).

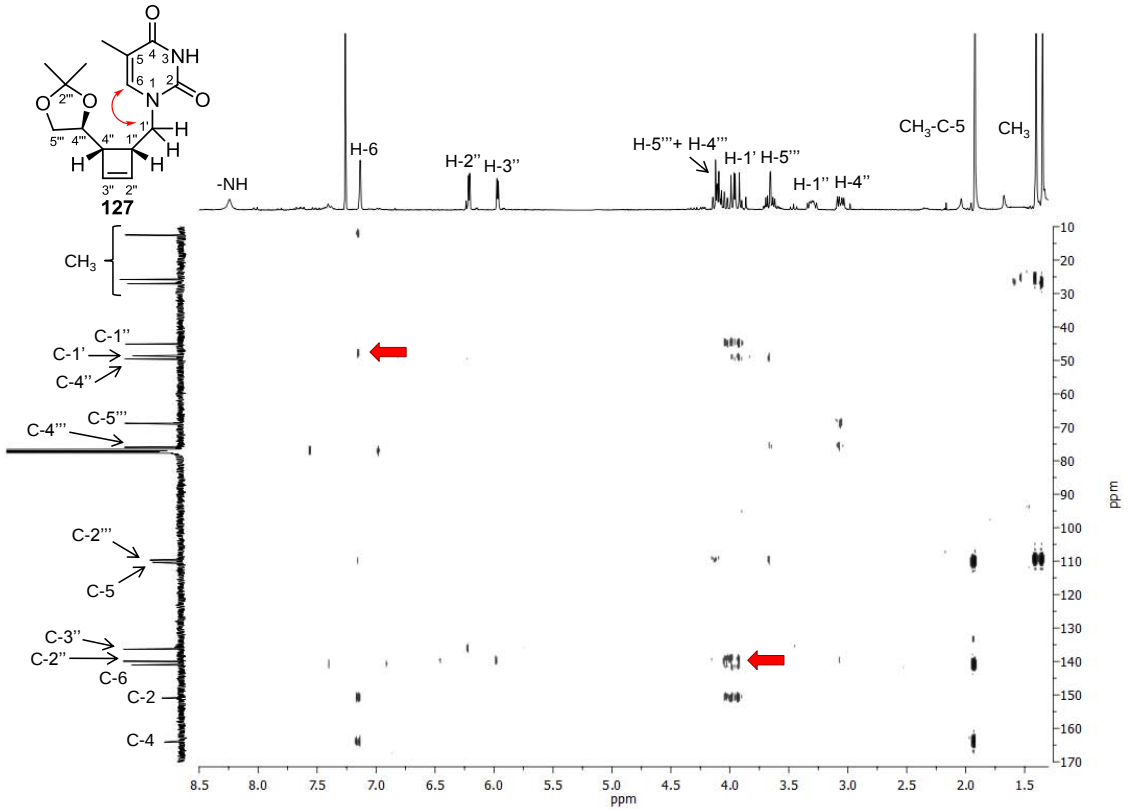
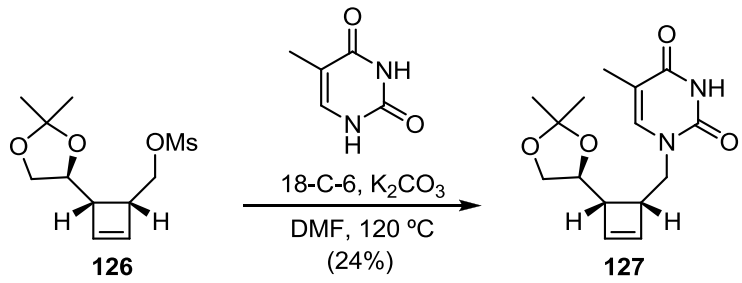


Figure 23. HMBC experiment of **127** (360 MHz, $CDCl_3$). Elucidation of *N*1 regiochemistry.

In order to avoid the formation of the side product **128**, the coupling of the nucleobase was then attempted by treating mesylate **126** with unprotected thymine, applying the previous reaction conditions. Purification of the resulting crude by column chromatography provided compound **127** albeit only in 24% yield (Scheme 27).

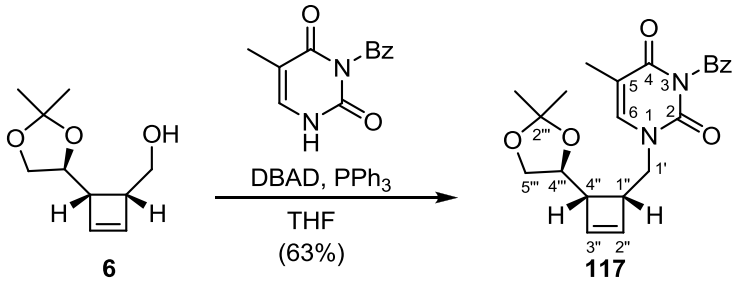


Scheme 27. Nucleophilic displacement of mesylate **126** with thymine.

Despite the poor yields obtained, these results are in accordance with previously reported works in which the yield for such reactions is around 30-40%.[236] The outcome of these reactions prompted us to investigate the Mitsunobu reaction as an alternative methodology to introduce the nucleobases.

2.6.2. Mitsunobu coupling of thymine to alcohol 6

Next, the introduction of the pyrimidine base *N*3-benzoylthymine was tested by means of a Mitsunobu reaction. Treatment of alcohol **6** with *N*3-benzoylthymine, DBAD and PPh_3 in THF provided the corresponding product **117** in 63% yield after successive column chromatography (Scheme 28).



Scheme 28. Preparation of compound **117** from **6** via a Mitsunobu reaction.

As before, the presence of the thymine base in **117** was confirmed by the two new signals in the ^{1}H-NMR spectrum at 7.26 and 1.96 ppm (corresponding to H-6 and C*H*$_3$-C-5, respectively) as well as by the three new aromatic peaks at 7.91, 7.63 and 7.48 ppm of the benzoyl group of the nucleobase. The assignment of the attachment site of the pyrimidine base on product **117** was established by the ^{1}H-^{13}C HMBC spectrum, wherein the correlation between proton H-1' and the aromatic carbon atom C-6 as well as between proton H-6 and C-1' indicated the *N*1 attachment (Figure 24).

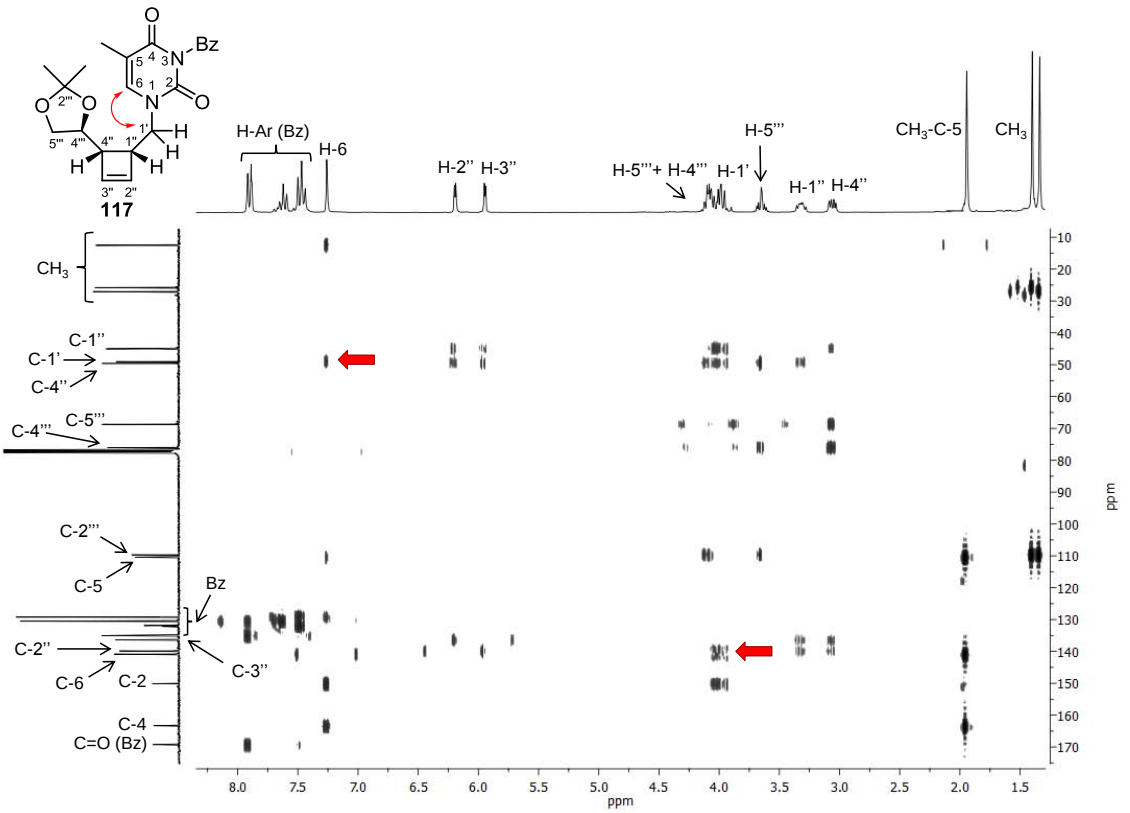


Figure 24. HMBC experiment of **117** (360 MHz, $CDCl_3$). Elucidation of *N*1 regiochemistry.

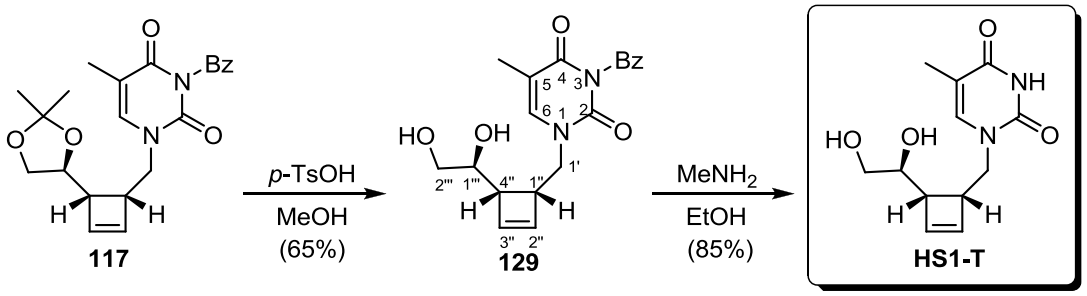
2.6.3. Synthesis of the cyclobutene nucleoside analogue HS1-T

To achieve the synthesis of the thymine nucleoside **HS1-T**, deprotection of *N*3 and the two hydroxyl groups of **117** was required.

First, cleavage of the acetonide protection of compound **117** was achieved by treatment with *p*-TsOH in MeOH at room temperature. Under these conditions, compound **129** was obtained in 65% yield after column chromatography (73% yield considering the unaltered starting material recovered) (Scheme 29). The disappearance of the signals corresponding to the acetonide in ^{1}H-NMR and ^{13}C-NMR spectra of compound **129** indicated that the removal of this group had occurred. Furthermore, displacement of H-1''', H-2''', C-1''' and C-2''' signals to lower chemical shifts was also observed (H-1''' and H-2''': from 4.09 and 3.66 ppm in **117** to 3.65 and 3.41 ppm in **129**; C-1''' and C-2''': from 76.0 and 68.7 ppm in **117** to 72.1 and 65.6 ppm in **129**).

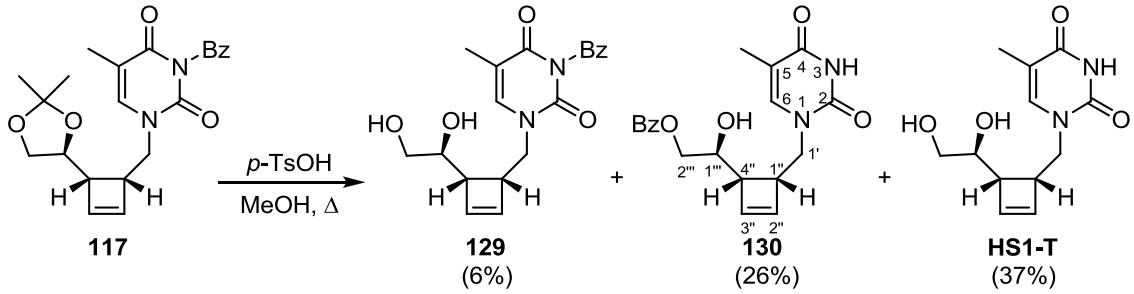
Subsequent exposure of compound **129** to a 33% solution of $MeNH_2$ in EtOH provided,[238] in 85% yield, a white solid identified as the target thymine nucleoside derivative **HS1-T** (Scheme 29). The deprotection of the nucleobase was confirmed

by the disappearance of the peaks corresponding to the benzoyl group in ^{1}H-NMR and ^{13}C-NMR spectra of the nucleoside analogue **HS1-T**.



Scheme 29. Preparation of the cyclobutene L-nucleoside **HS1-T** from **117**.

It is worth mentioning that the attempt to drive the removal of the acetonide to completion by increasing the reaction temperature to reflux resulted in the direct formation of the nucleoside derivative **HS1-T** in 37% yield, together with the intermediate **129** (6% yield) and the by-product **130** (26% yield) (Scheme 30).



Scheme 30. Deprotection of compound **117** with *p*-TsOH in refluxing MeOH.

Extensive NMR spectral data provided strong support for the structure of compound **130**. Analysis of the ^{1}H-^{13}C HMBC spectrum allowed to determine the benzoyl migration, since a cross-peak between protons H-2''' and C=O from the benzoyl group was clearly observed (Figure 25). This assignment was further supported by the higher chemical shift value of protons H-2''' of compound **130** (4.50 and 4.35 ppm) compared to those of derivative **129** (3.65 and 3.41 ppm), indicating the migration of the benzoyl group. The same behaviour was observed for C-2''' (from 65.6 ppm in compound **129** to 68.4 ppm in compound **130**).

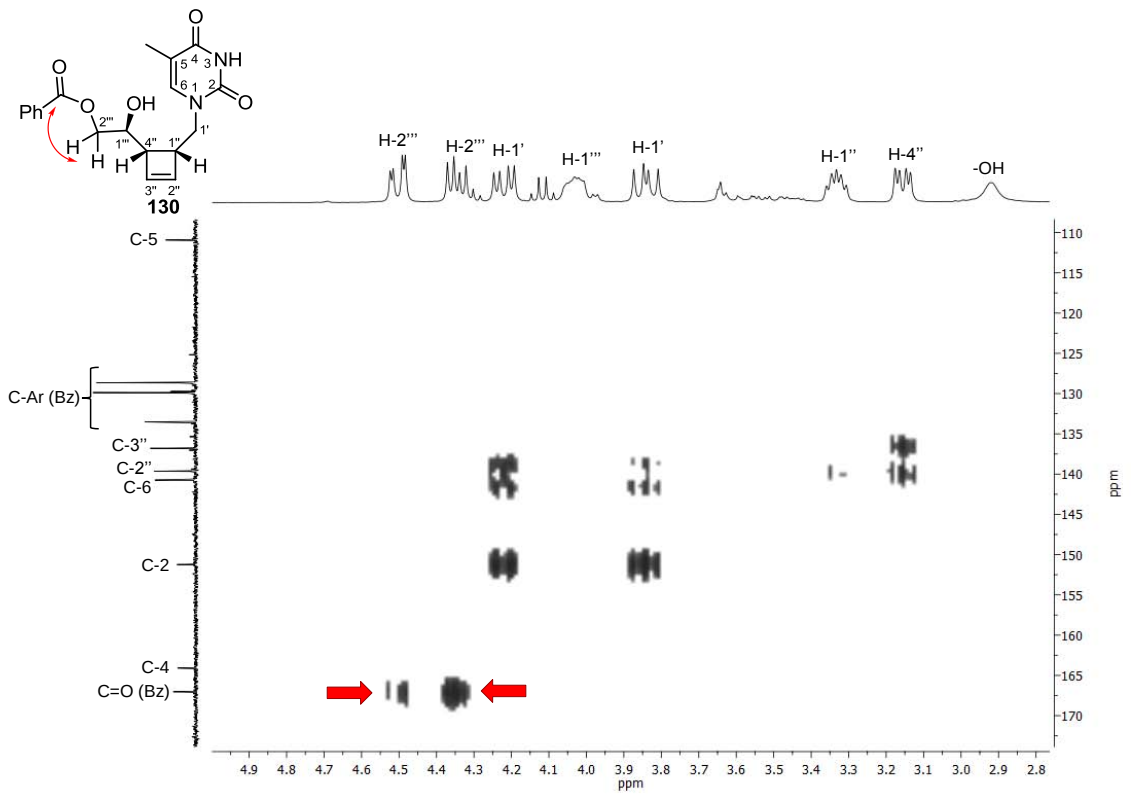
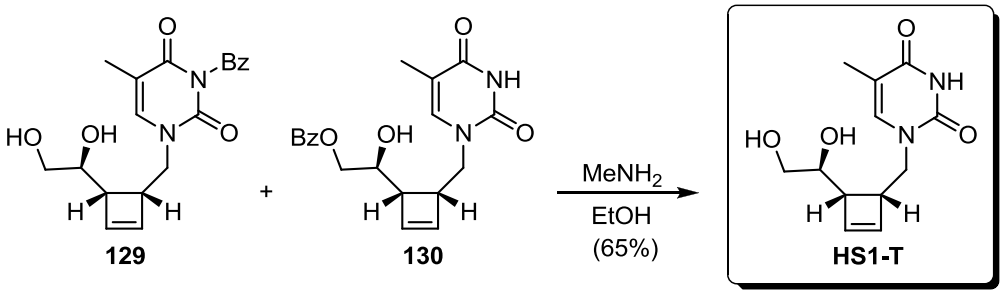


Figure 25. Detail of the HMBC experiment of **130** (360 MHz, $CDCl_3$).

A mixture of derivatives **129** and **130** were then subjected to deprotection with $MeNH_2$ in EtOH, delivering a single product identified as the nucleoside **HS1-T** in 65% yield (Scheme 31).

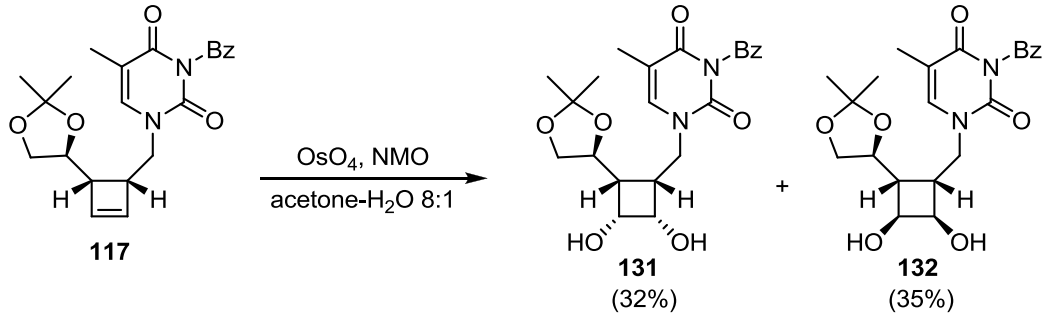


Scheme 31. Preparation of the cyclobutene L-nucleoside **HS1-T** from a mixture of **129** and **130**.

2.6.4. Synthesis of the polyhydroxylated cyclobutane nucleoside analogues HS2-T and HS2'-T

As mentioned in the introduction of the present chapter, cyclobutene compounds may be also used as precursors to prepare polyhydroxylated cyclobutane nucleosides by taking advantage of the cyclobutene carbon-carbon double bond. To this aim, dihydroxylation of the fully protected compound **117** was accomplished under mild conditions with *N*-methylmorpholine-*N*-oxide (NMO) and

catalytic OsO_4 in acetone-H_2O,[239] leading to a chromatographically separable mixture of diols **131** and **132** in 32% and 35% yield, respectively (Scheme 32).



Scheme 32. Dihydroxylation of compound **117**.

The relative configuration of diols **131** and **132** was established on the basis of selective n.O.e. experiments. Hence, in the n.O.e. study of **132**, the cyclobutane proton H-3'' signal showed an enhancement upon irradiation at the proton H-4''' frequency, indicating that both protons H-3'' and H-4''' are disposed in a *cis* relative configuration (Figure 26). Unfortunately, the chemical shifts of the ^{1}H-NMR signals of the key protons of **131** are almost identical, so unequivocal n.O.e. signal between these protons could not be determined and the relative configuration was proven in the final compound.

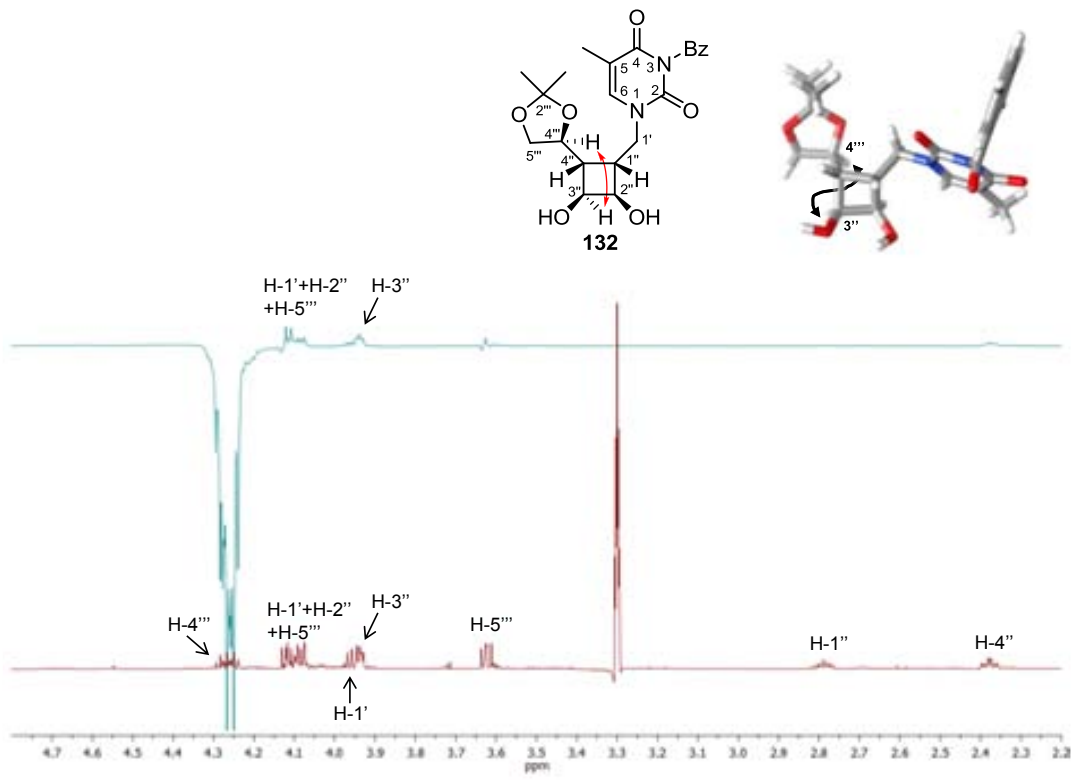
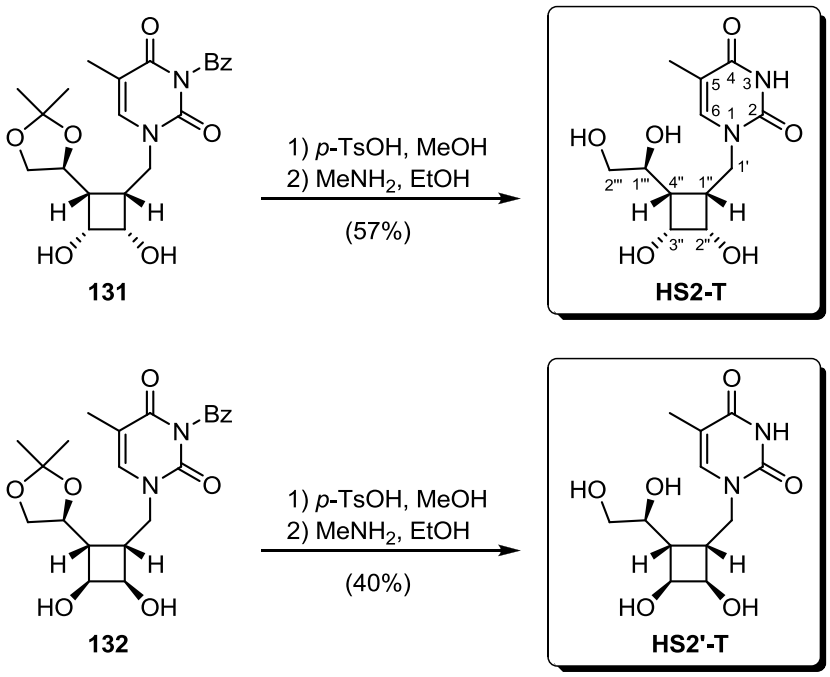


Figure 26. ^{1}H-NMR spectrum (600 MHz, CD_3OD) of **132** (red) and n.O.e. spectrum irradiated at proton H-4''' frequency (blue).

Next, diols **131** and **132** were subjected to deprotection by exposure to *p*-TsOH in MeOH and subsequent treatment with a 33% solution of $MeNH_2$ in EtOH. The tetrahydroxylated thymine nucleoside analogues **HS2-T** and **HS2'-T** were obtained in 57% and 40% overall yield from **131** and **132**, respectively (Scheme 33). As before, cleavage of the protecting groups was confirmed by ^{1}H-NMR and ^{13}C-NMR spectra of nucleosides **HS2-T** and **HS2'-T**, wherein the acetal and benzoyl signals are no longer observed and H-1''', H-2''', C-1''' and C-2''' signals are displaced to lower chemical shifts.



Scheme 33. Preparation of tetrahydroxylated cyclobutane L-nucleosides **HS2-T** and **HS2'-T** from **131** and **132**, respectively.

The relative configuration was further supported by a selective n.O.e. experiment on nucleoside **HS2-T**, wherein irradiation at proton H-4'' frequency caused n.O.e. on H-2'', confirming that this proton is on the same side than proton H-4'' (Figure 27).

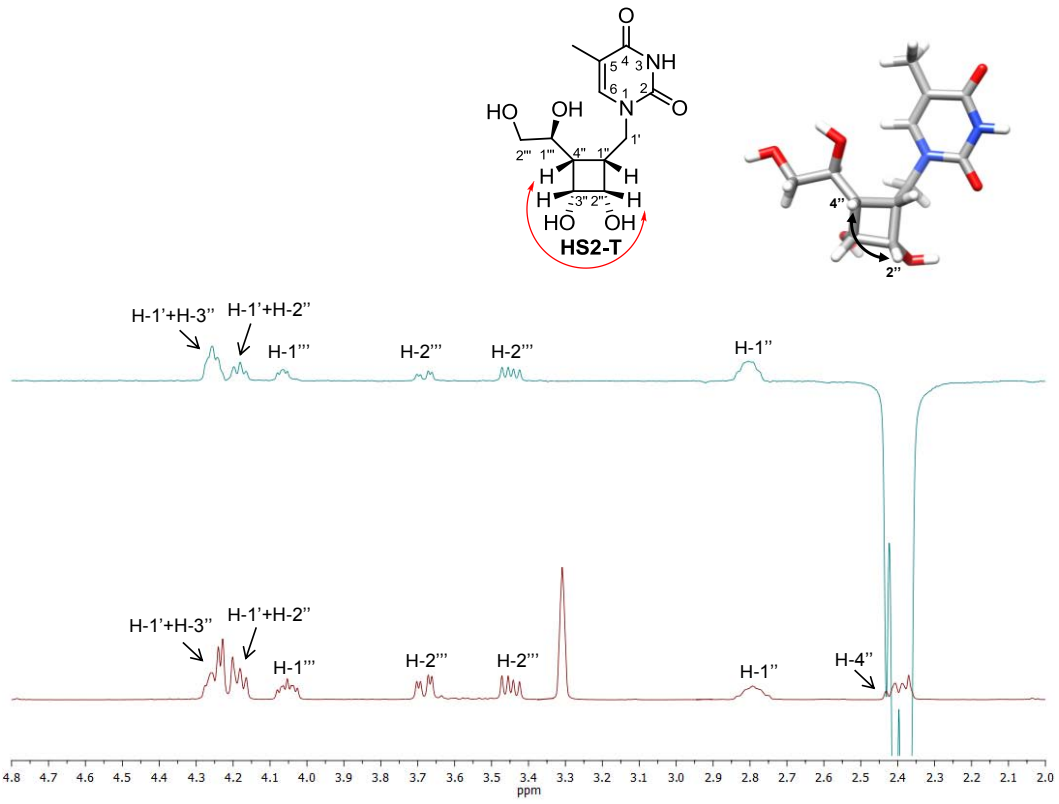
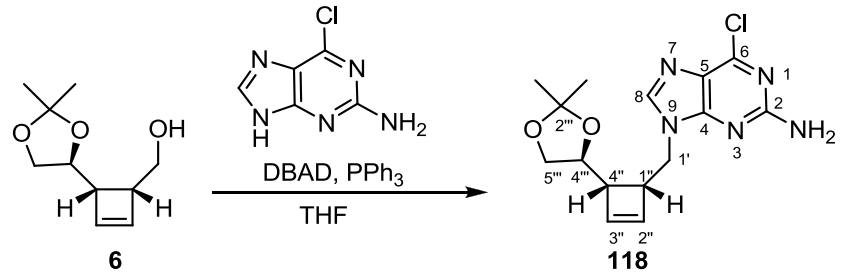


Figure 27. ^{1}H-NMR spectrum (360 MHz, CD_3OD) of **HS2-T** (red) and n.O.e. spectrum irradiated at proton H-4'' frequency (blue).

2.7. Preparation of nucleoside analogues featuring a purine base

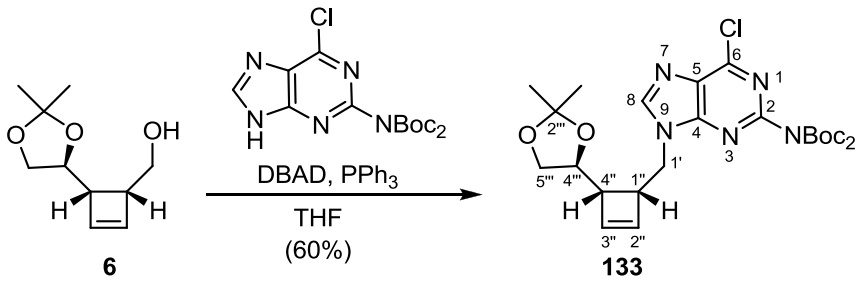
2.7.1. Mitsunobu coupling of 2-amino-6-chloropurine to alcohol 6

Afterwards, we proceeded to prepare the purine analogues from the previously obtained alcohol **6**. 2-amino-6-chloropurine was chosen as the purine nucleobase to be introduced, since it can be converted into guanine under acidic conditions.[240–248] Alcohol **6** was thus treated with 2-amino-6-chloropurine in the presence of DBAD and PPh_3 in THF. The reaction took place smoothly according to TLC controls and the ^{1}H-NMR spectrum of the resulting crude, wherein new signals corresponding to H-8 and NH_2 appeared at 7.86 and 5.35 ppm, confirmed the formation of the desired compound **118** (Scheme 34). However, it was not possible to purify the product despite successive column chromatography, so we decided to attempt the Mitsunobu coupling with the bis-Boc-protected 2-amino-6-chloropurine.



Scheme 34. Mitsunobu coupling of 2-amino-6-chloropurine to alcohol **6**.

N,*N*-bis-Boc-2-amino-6-chloropurine was easily obtained according to the procedure described by Garner and Dey, which consisted of the preparation of the tris-Boc-protected 2-amino-6-chloropurine and the subsequent deprotection of *N*9 by treatment with aqueous $NaHCO_3$.[249] Coupling of *N*,*N*-bis-Boc-2-amino-6-chloropurine to alcohol **6** was accomplished under the same conditions as before, affording compound **133** in 60% yield after repeated column chromatography (Scheme 35).



Scheme 35. Preparation of compound **133** from **6** via a Mitsunobu reaction.

The introduction of *N*,*N*-bis-Boc-2-amino-6-chloropurine was established by its distinctive signals in ^{1}H-NMR (8.27 and 1.44 ppm corresponding to H-8 and the *tert*-butyl groups of Boc, respectively) and ^{13}C-NMR spectra (five new signals in the zone comprised between 153.0 and 146.7 ppm as well as three new signals at 130.2, 83.7 and 28.0 ppm).

Analysis of the ^{1}H-^{13}C HMBC spectrum allowed to establish the *N*9 attachment of the purine base on product **133**, since a cross-peak between proton H-1' and C-4 was observed (Figure 28).

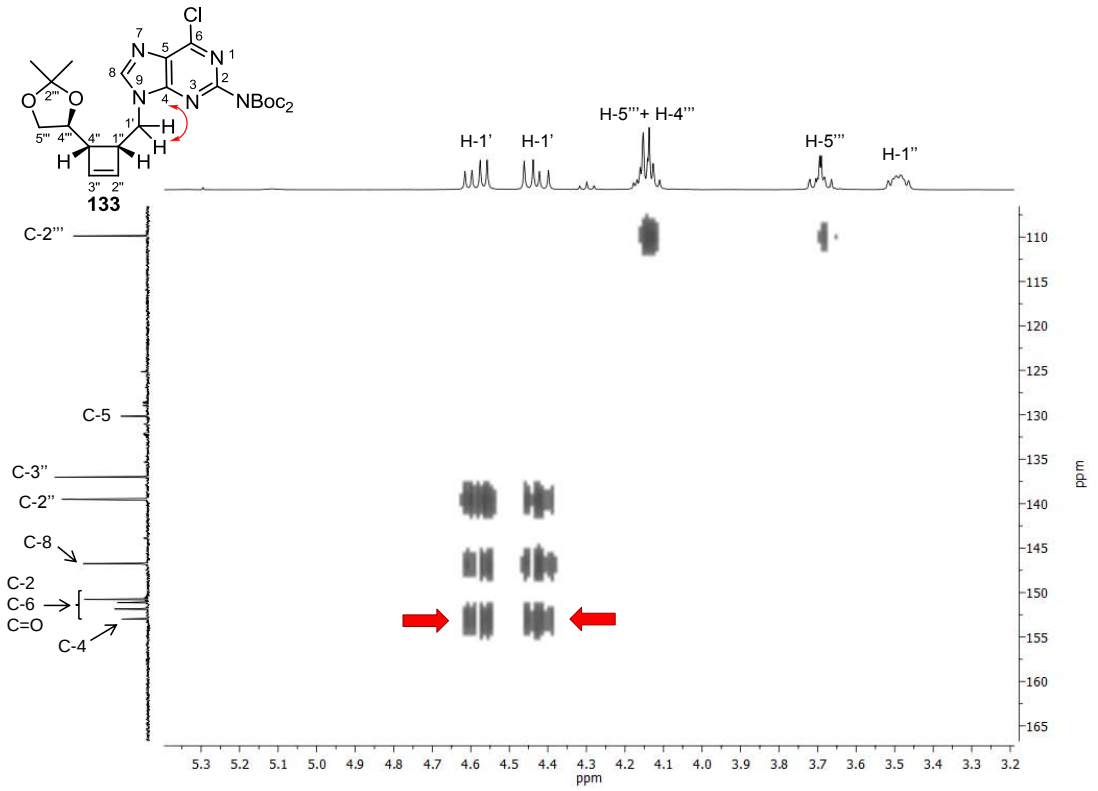
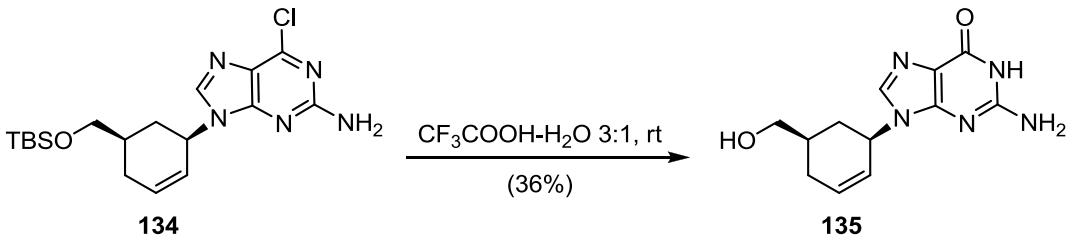


Figure 28. Detail of the HMBC experiment of **133** (360 MHz, $CDCl_3$). Elucidation of *N*9 regiochemistry.

2.7.2. Synthesis of the cyclobutene nucleoside analogues HS1-G and HS1-G^{Cl}

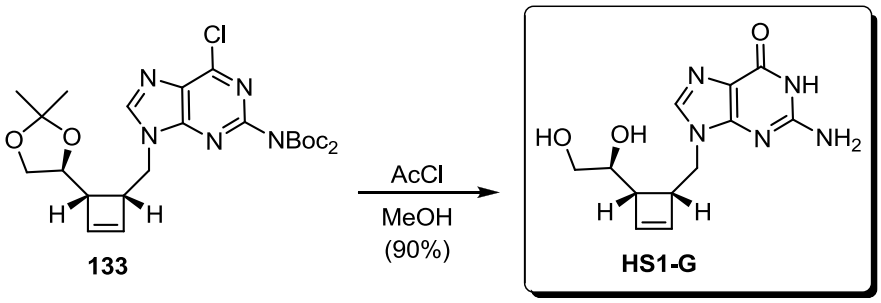
As previously mentioned, conversion of 2-amino-6-chloropurine into guanine can be accomplished by an acid treatment.[240–248] For instance, Camplo and co-workers reported the conversion of the 2-amino-6-chloropurine derivative **134** into the guanine cyclohexene nucleoside **135** by treatment with trifluoroacetic acid-H_2O 3:1 in 36% yield, simultaneously removing the TBS protecting group (Scheme 36).[242]



Scheme 36. Conversion of the 2-amino-6-chloropurine of **134** into guanine, Camplo and co-workers (2005).[242]

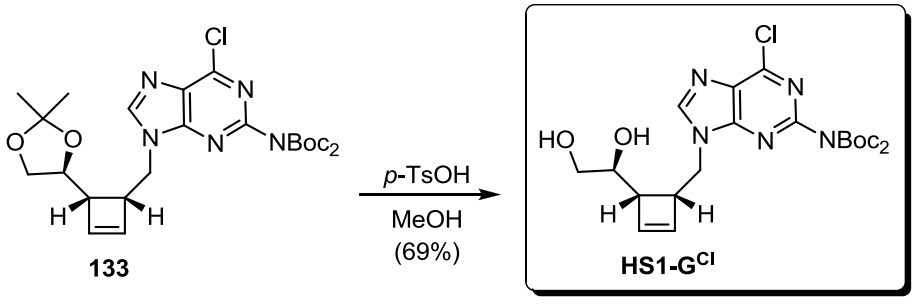
On the other hand, it is known that the Boc protecting group can also be cleaved under acidic conditions.[245,250–252] Therefore, we envisaged the

simultaneous removal of the bis-Boc and the acetonide protection, as well as the conversion of the purine base into guanine, by an acid treatment of compound **133**. To this aim, the cyclobutene derivative **133** was exposed to trifluoroacetic acid-H_2O 3:1,[240–246] leading to a complex mixture of unidentified compounds without recovering starting material. Treatment of **133** with aqueous hydrochloric acid (32%) in MeOH,[244,248,251] provided a mixture of the desired product **HS1-G** along with unidentified by-products and starting material. Eventually, it was found that the cleavage of the protecting groups and conversion of the nucleobase could be achieved by generating the hydrochloric acid *in situ* with acetyl chloride in MeOH.[252] Under these conditions, the target guanine nucleoside derivative **HS1-G** was obtained in excellent yield (Scheme 37). The removal of the acetonide and Boc protecting groups was confirmed by the disappearance of the corresponding signals in ^{1}H-NMR and ^{13}C-NMR spectra of **HS1-G**, whereas the conversion of the nucleobase into guanine was supported by the HRMS of **HS1-G**.



Scheme 37. Preparation of the cyclobutene L-nucleoside **HS1-G** from **133**.

Interestingly, selective removal of the acetonide protection was accomplished when treating compound **133** with *p*-TsOH in MeOH, furnishing a white solid in 69% yield identified by its spectroscopic data as the nucleoside analogue **HS1-G^{Cl}** (Scheme 38). Compound **HS1-G^{Cl}** was considered to be interesting by itself to evaluate its antiviral activity.



Scheme 38. Preparation of the cyclobutene L-nucleoside **HS1-G^{Cl}** from **133**.

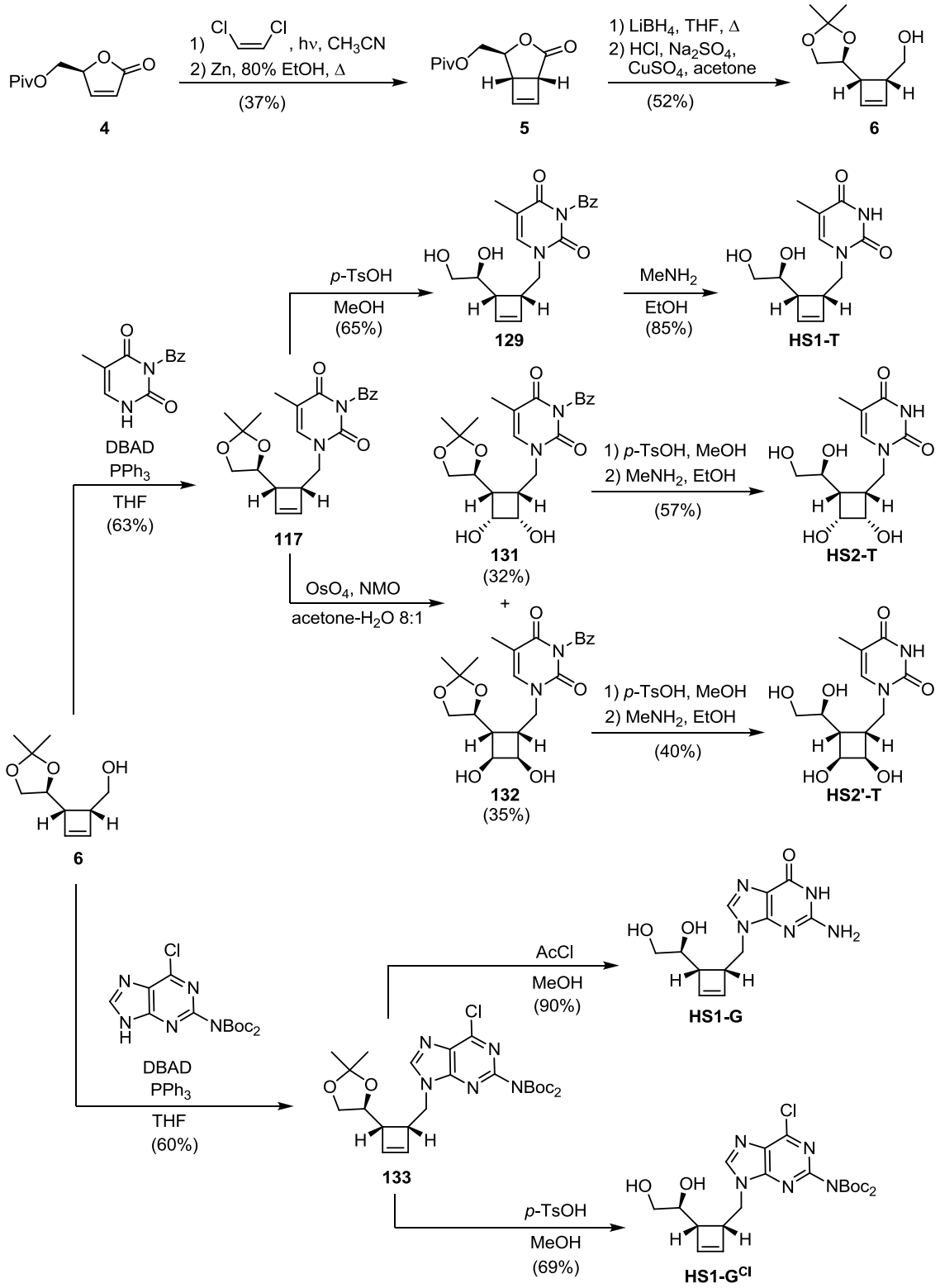
3. Antiviral activity evaluation

The antiviral activity of all the nucleoside analogues synthesized in the present work has been evaluated by the research group of Prof. Jan Balzarini (Rega Institute for Medical Research, Katholieke Universiteit Leuven, Belgium). Nucleoside analogues **HS1** and **HS2** have been tested for antiviral activity against HSV-1 (KOS), HSV-1 resistant to ACV (TK$^-$ KOS ACV^r), HSV-2 (G), vaccinia virus and vesicular stomatitis virus in human embryonic lung (HEL) cell cultures (Table ES1, Experimental Section); against vesicular stomatitis virus, coxsackie virus B4 and respiratory syncytial virus in HeLa cell cultures (Table ES2); against parainfluenza-3 virus, reovirus-1, sindbis virus, coxsackie virus B4 and punta toro virus in Vero cell cultures (Tables ES3 and ES4); against feline corona virus and feline herpes virus in Crandell-Rees feline kidney (CRFK) cell cultures (Table ES5); and against influenza viruses in Madin Darby canine kidney (MDCK) cell cultures (Table ES6). Unfortunately, nucleosides that showed weak activity against some viruses resulted also cytotoxic (i.e. **HS1-G** against vesicular stomatitis virus, coxsackie virus B4 and respiratory syncytial virus in HeLa cell cultures, **HS1-G^{Cl}** against feline corona virus and feline herpes virus in CRFK cell cultures, and **HS1-T**, **HS2-T**, **HS2'-T** and **HS1-G^{Cl}** against influenza viruses in MDCK cell cultures). On the other hand, nucleoside **HS1-G^{Cl}** showed antiviral activity against coxsackie virus B4 without showing cytotoxicity in Vero cell cultures (Tables ES3 and ES4), but these results were not reproducible in HeLa cell cultures (Table ES2). The rest of the antiviral activity evaluations revealed that our nucleosides did not exhibit activity at concentrations below 100 μg/mL.

4. Summary

Summarizing, three novel cyclobutene L-nucleoside analogues, **HS1-T**, **HS1-G^{Cl}** and **HS1-G**, as well as two polyhydroxylated cyclobutane L-nucleosides, **HS2-T** and **HS2'-T**, have been synthesized following a divergent synthetic pathway (Scheme 39). The cyclobutene moiety has been prepared by means of a [2+2] photocycloaddition of 2(5*H*)-furanone **4** to (*Z*)-1,2-dichloroethylene followed by Zn-promoted reductive elimination. Reductive opening of lactone **5** and subsequent

protection of the vicinal diol have provided the key intermediate for the introduction of the nucleobases, alcohol **6**. Pyrimidine and purine bases have been coupled via a Mitsunobu reaction, affording compounds **117** and **133** which by subsequent synthetic modifications have been converted into the target nucleosides **HS1** and **HS2**. To the best of our knowledge, these are the first examples of cyclobutene nucleoside analogues featuring the unnatural configuration L-. The antiviral activity of the synthesized nucleoside analogues has been evaluated without showing significant activity.



Scheme 39. Synthesis of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues.

CHAPTER III

Overview on molecular modelling in drug design

1. Introduction

Drugs have been defined as "compounds which interact with a biological system to produce a biological response".[253] The biological response (and consequently the effectiveness of the drug) is a function of several molecular variables that include its absorption, metabolism and interaction with its target. There are four main molecular targets, namely lipids, carbohydrates, proteins and nucleic acids. Drugs interact with all these macromolecules in a process known as binding, and the specific area where this interaction takes place is called the binding site (Figure 29). The vast majority of drugs are targeted on proteins.[253]

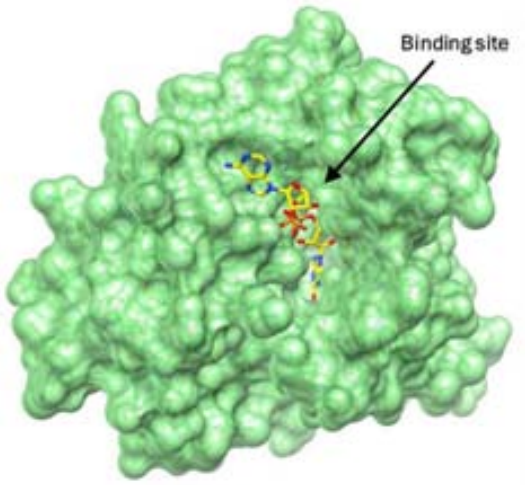


Figure 29. Representation of the surface of a protein (in green) with the crystallized ligand (in yellow; PDB code: 1LVG).

Prodrugs such as nucleoside analogues must be converted into their active form (triphosphorylated nucleosides) before reaching their target. This process is performed by the corresponding kinases,[75] which are a large family of enzymes that catalyse the phosphoryl group transfer from a donor (normally ATP) to a receptor. Afterwards, activated nucleoside analogues must interact with a protein, the viral polymerase, to cause the desired antiviral effect, as previously introduced in Chapter I. Thus, proteins are involved in the crucial processes of the mechanism of action of nucleoside analogues.

Proteins are polymers of amino acids that range in size from small to very large.[a] Using experimental techniques to investigate the properties of such macromolecules on an atomic level is very complex and time consuming. The fast advances on computer hardware and software achieved in the last decades have allowed the investigation of the above mentioned biological processes on an atomic level by means of molecular modelling. Since molecular modelling has proved to be useful in several research fields, the usage of this technique has become widespread. Nowadays molecular modelling is invariably associated with computer modelling, thus encompassing all the theoretical and computational techniques that provide insights into the behaviour of molecules and molecular systems.[254]

In general, two main approaches are applied to investigate the biological activity of a given compound: ligand-based and structure-based.[255] Ligand-based approaches, such as quantitative structure-activity relationship (QSAR) models, may be used when the structure of the biological target is unknown and a series of compounds exhibiting the desired activity have been identified. These approaches rely on the correlation between the structure of a compound and its activity. Conversely, in structure-based approaches the three-dimensional structure of the biological target is available, and the interactions of the ligand in the binding site are used to extrapolate its activity. Several molecular modelling techniques are used by these approaches, such as protein–ligand docking.

As mentioned, proteins are decisive for the antiviral activity of nucleoside analogues. The Protein Data Bank is a data collection of particular importance for this work, since all the crystallographic protein–ligand structures used in the molecular modelling studies have been retrieved from this repository. The Protein Data Bank (PDB, http://www.rcsb.org/pdb/) is a worldwide archive of structural data of biological macromolecules.[256,257] It was established in 1971 with the appearance of several crystal structures of proteins, and today over 86,000 structures are accessible via internet. Apart from containing the atomic coordinates of these structures, text included in each data entry gives pertinent information of the structure (e.g. species from which the molecule has been obtained, experimental details or related literature citations).

[a] The largest known protein, titin, presents around 27,000 amino acid residues in length.[404]

The next two sections will be devoted to the principal computational tools applied in drug design that have been used in the present thesis. The first one will deal with ligand-based approaches and be limited to "drug-likeness" and how to predict drug absorption, whereas the second one will outline a few general principles about the structure-based computational techniques aimed at modelling biological macromolecules and their interaction with small molecules.

2. Drug-likeness

Drugs have to travel through the body in order to reach their targets, so that the compound with the best binding interactions with its target is not necessarily the best drug to use in medicine. For this reason, drug absorption, distribution, metabolism and excretion (often abbreviated to ADME) should be taken into consideration when investigating the limiting steps for the biological activity of a molecule.

In this field, the term "drug-like" has been widely used in spite of the controversy on how "drug-like" is defined. Walters and Murcko define drug-like compounds as ''molecules which contain functional groups and/or have physical properties consistent with the majority of known drugs".[258] Conversely, Lipinski defines drug-like as "those compounds that have sufficiently acceptable ADME properties and sufficiently acceptable toxicity properties to survive through the completion of human Phase I clinical trials".[259] Despite these differences, and since the most common and preferred mode of administering drugs is the oral route, several methods have been developed to estimate oral bioavailability by performing retrospective analyses of known drugs.[258,260,261] Among them, the so-called "Lipinski's rule of 5" has been one of the most extensively used methods.[262–264]

Analysing 2245 compounds from the World Drug Index[a] (WDI), Lipinski et al.[262] determined that poor absorption or permeation are more likely when: there are more than 5 hydrogen bond donors (expressed as the sum of OHs and NHs); there are more than 10 hydrogen bond acceptors (expressed as the sum of Ns and Os);

[a] The World Drug Index is an authoritative index for marketed and development drugs created by Thomson Reuters.

the MW is over 500; and the log P is over 5 (where P is the octanol-water partition ratio, used to represent molecular lipophilicity). These four physico-chemical parameter ranges were associated with 90% of the studied drugs that had achieved phase II clinical status, leading to the simple mnemonic which the authors named the "rule of 5" because the cut-off points for each of the four parameters were all close to 5 or a multiple of 5. Remarkably, if a compound fails the "rule of 5" there is a high probability that oral activity problems will be encountered. However, the converse is not true: compounds passing the rule still can prove troublesome in experimental studies. Remarkably, it is well established that certain orally active therapeutic compounds (i.e. antibiotics, antifungals, vitamins and cardiac glycosides) are outside "Lipinski's rule of 5".

Another method widely used to address the oral bioavailability of compounds is the one derived from the oral bioavailability measurements in rats, for over 1100 drug candidates of the GlaxoSmithKline, reported by Veber et al.[265] From those data, the authors concluded that reduced molecular flexibility, as measured by the number of rotatable bonds,[a] and low polar surface area or total hydrogen bond count (sum of donors and acceptors) are found to be important predictors of good oral bioavailability, independently of molecular weight. Indeed, the success of the molecular weight parameter in "Lipinski's rule of 5" to predict oral bioavailability may be partially a result of the correlation of higher molecular weight with increased molecular flexibility and higher polar surface area or hydrogen bond count. Authors' observations suggested thus that compounds which meet the two criteria of (1) 10 or fewer rotatable bonds and (2) polar surface area equal to or less than 140 $Å^2$ (or 12 or fewer hydrogen bond donors and acceptors) will have a high probability of good oral bioavailability in the rat, despite the mechanistic basis for the rotatable bond filter is unclear.

[a] Rotatable bonds were defined as any single bond, not in a ring, bound to a nonterminal non-hydrogen atom. Excluded from the count were amide C-N bonds because of their high rotational energy barrier.

3. Molecular modelling of biological macromolecules

3.1. Introduction

The various operations which are carried out in molecular modelling involve the use of programs and algorithms to calculate the structure and property data of a molecule. The computational methods to estimate these parameters can be split into two categories: molecular mechanics (MM) and quantum mechanics (QM). In molecular mechanics, molecules are treated as a series of spheres (the atoms) connected by springs (the bonds), which follow the laws of classical physics. By contrast, quantum mechanics uses quantum physics to calculate the property of a molecule by considering the interactions between electrons and nuclei. Unlike molecular mechanics, atoms are not treated as solid spheres and electrons are included in the calculations. As a result, MM is less intensive on computer time than QM but it cannot treat the reactivity of the system.

One of the critical parameters in the choice of the method is the size of the system to be studied. Thus, the investigation of macromolecules formed by thousands of atoms such as proteins is generally carried out with computational methods based on MM. In this context, one of the most widely used techniques is molecular docking.[266–271]

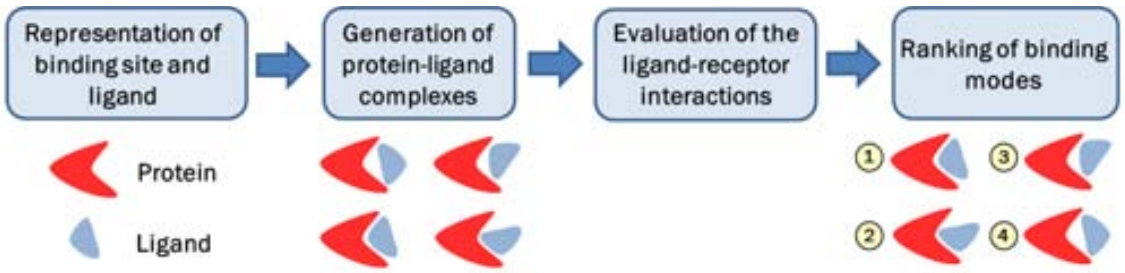
Molecular docking aims at predicting the structure of an intermolecular complex between two or more constituent molecules. In particular, protein–ligand docking deals with the molecular binding of one molecule (the ligand) to the pocket of another, usually larger, molecule (the protein, also called receptor). Pioneered during the early 1980s,[272] it remains a field of vigorous research as it has demonstrated to be a valuable tool in drug discovery programs, among others.[273–275] For instance, Vidofludimus (4SC-101), which is a novel drug candidate for the treatment of autoimmune disorders discovered by 4SC AG,[276] was developed based on a lead compound[a] that was obtained by a docking procedure.[277] It has successfully completed the Phase IIa trial for the treatment of inflammatory bowel disease.

[a] In drug discovery, a lead compound is a compound showing a desired pharmacological property which can be used to initiate a medicinal chemistry project.

In the next paragraphs, basics of docking techniques will be shortly presented, because the work in the following chapters relies substantially on the application of this technique, both to rationalize the experimental results of antiviral activity of the previously synthesized nucleoside analogues, as well as to perform a rational design of novel nucleosides as antiviral agents.

3.2. Protein–ligand docking

The objective of protein–ligand docking is the identification of the most favoured ligand conformations and orientations (commonly referred as poses, binding modes or, simply, solutions) within a given target protein.[a] The task of the docking program is twofold: first, it has to place the ligand within the active site in different orientations and second, it has to score the binding modes to identify the best ones. With this aim, docking programs follow 4 general steps: (1) identification of the physicochemical properties of the binding site and the ligand; (2) sampling of the conformational space available to the ligand in the binding site and subsequent generation of protein–ligand complexes; (3) measure of the protein–ligand binding energy; and (4) scoring of the generated protein–ligand complexes, which results in a ranking of poses (Scheme 40). Steps (2) and (3) are closely related, since most docking programs calculate the binding energies of the orientations while they are being generated, to guide the program towards the best poses.



Scheme 40. General docking procedure.

Docking programs consist of two intertwined parts: the search algorithm and the scoring function. The mathematical details and principles of these algorithms are presented in the computational methods. However, it is important to appreciate that the search algorithm is related to the sampling of the conformational space in the receptor and the generation of a number of ligand orientations whereas the

[a] From now on, we will refer to protein–ligand docking simply as "docking".

scoring function is designed to calculate the corresponding protein–ligand binding energies and accordingly rank these predicted poses to find the most favoured ones. It is then the scientist work to analyse these poses and evaluate if they are in good agreement with the known requirements of the target protein to, for example, rationalize experimental results of activity. The deep understanding of a ligand binding mode gained by means of molecular docking may also be used to further design efficient drugs by modifying the ligand.

Molecular dockings are nowadays broadly applied, being capable of predicting known ligand binding modes with average accuracies of about 1.5-2 Å and success rates in the range of 70-80%.[268] However, they present some limitations that have to be pointed out. Inaccuracies in the scoring function are one of the major limiting factors, since numerous simplifications are done to allow the computational evaluation of the ligand affinity of an extremely large number of poses. Entropy, solvent effects and electrostatic interactions are the main parameters that are completely neglected or at least not fully accounted for in most scoring schemes.[267,268,270] To address this issue, rotational entropy[278,279] and solvation[280–282] contributions have been included in some scoring functions, but these terms are only partial descriptions of the real entropic and solvation effects occurring on protein–ligand binding.

Moreover, water molecules and metals are normally disregarded when performing docking studies,[268,273] even though it is well established that they can directly participate in protein–ligand interactions. Some dockings with explicit water molecules have been performed,[283–285] and certain docking programs have optimized metal ion parameters to take them into consideration during the calculations.[286–288] However, accurately and efficiently account for the role played by metals and water in binding is still one of the major challenges in the field.

Another relevant limitation of current molecular docking programs is the flexibility of the protein receptor.[268,270,271,289–292] An ideal docking should consider many degrees of freedom: translation and rotation of one molecule relative to another, which involves 6 degrees of freedom, as well as the conformational degrees of freedom of both the ligand and the receptor. This is too costly and thus several docking methods follow the assumption that protein structures are rigid entities, and that the ligand changes its three-dimensional structure during the

binding process to find the best spatial and energetic fit to the protein binding site. This criterion is based on the historic "lock and key" model proposed by Emil Fischer in 1894 (Figure 30).[293] However, this model has long been showed obsolete and replaced with the concept that the enzyme itself usually undergoes a change in conformation when the substrate binds, induced by multiple weak interactions with the substrate. This mechanism, known as "induced fit", was postulated by Daniel Koshland in 1958.[294] Today, more modern theories that confer even more relevance to the flexibility of the receptor by itself are well established, describing the protein as an ensemble of differently populated conformational states in equilibrium rather than as a clearly dominant more stable conformation, and postulating that the ligand selects the most favoured conformation ("conformational selection").[291,295–297] Accordingly, the importance of considering protein flexibility in docking calculations is nowadays undeniable.

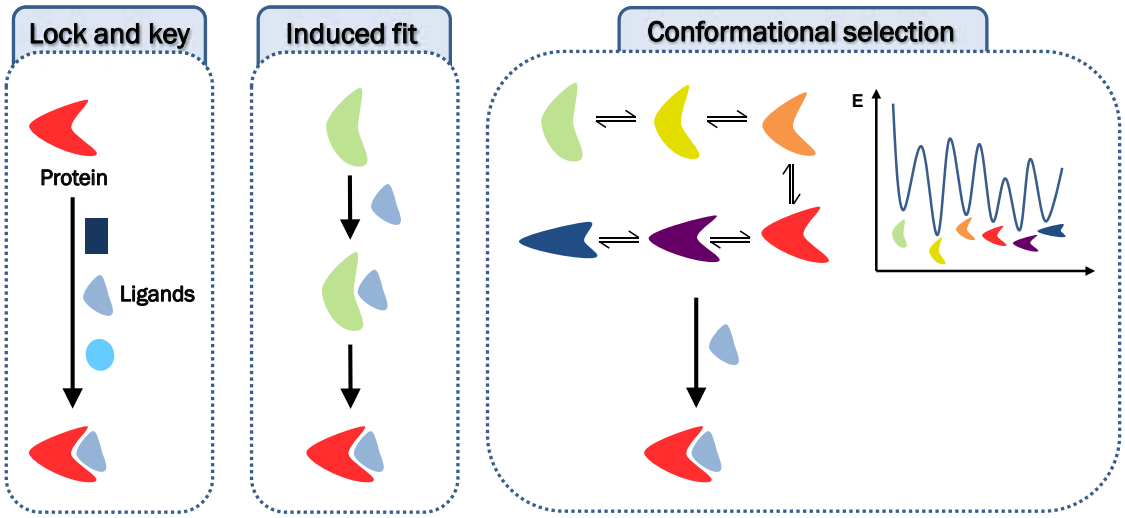


Figure 30. Simplified representation of the three main models for protein–ligand binding.

To overcome limitations of the rigid protein assumption, several approximations have been used to model protein flexibility.[271,289–291] They can be divided in two groups: models that account for the flexibility of the protein only in the binding region (local) and models which simulate the flexibility as a whole (global). Local protein flexibility was first attempted by the so-called soft-docking method, consisting in decreasing the van der Waals repulsion energy term[a] between the atoms in the binding site and those in the ligand.[298] As a result, some overlap between ligand and protein atoms is allowed when performing the conformational

[a] See section *1. Molecular mechanics* of the computational methods.

search of the flexible ligand, presupposing that low energy rearrangements of protein atoms in the binding site could eliminate these collisions while maintaining the conformation of the ligand (Figure 31). Another example of a method developed to explicitly include partial receptor flexibility made use of a library of discrete rotameric states[a] for each type of side chain.[299]

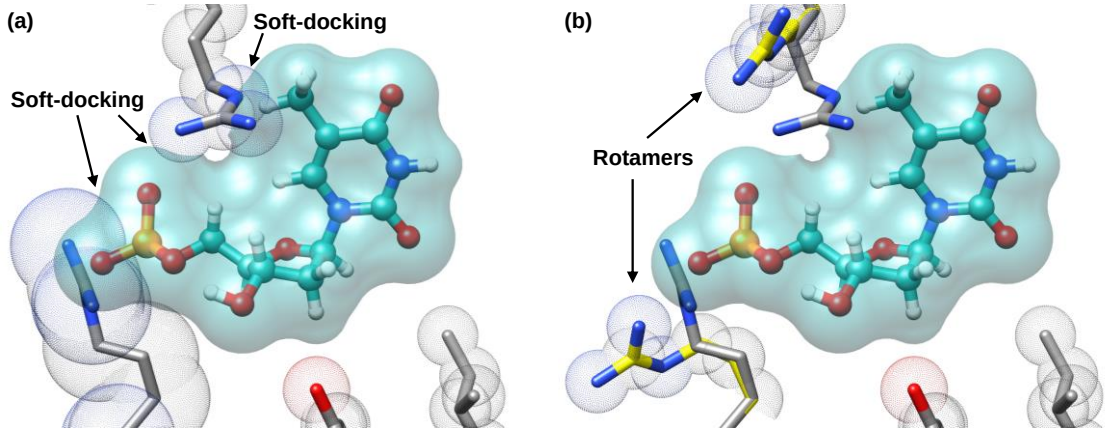


Figure 31. Methods developed to model local protein flexibility: (a) soft-docking and (b) use of rotamers (original residues shown in grey and rotamers in yellow).

Regarding full protein flexibility, one approach is to generate an ensemble of rigid protein conformations that together represent the conformational diversity of the protein.[271] These structures can be obtained from NMR studies, from multiple crystal structures or be generated by computational techniques.[290] A different representation to consider large-scale motions is to divide the protein in domains whose constituent atoms move collectively as one, while conformational changes inside domains can be taken into account using minimization.[300] In spite of the significant advances achieved by all these approaches, simulating the full process of diffusion and ligand binding without any approximation is still, disappointingly, out of our present computational capability.

[a] Rotamer libraries consist of discrete side chain conformations of low energy which are usually determined from statistical analysis of structural data derived experimentally.

CHAPTER IV

In silico study of cyclobutane-fused nucleoside analogues as anti-HIV agents

1. Introduction

Although several nucleoside analogues have been approved for the treatment of HIV, there is still a need for the development of new anti-HIV agents due to the deficiencies of the currently used antiviral drugs.[7,25,27,50–62] In this context, the study of the mechanism of action of nucleoside analogues may provide useful information for further synthetic designs. The following subsection will provide a general overview on their mechanism of action in HIV infected cells. Next, some of the recently reported investigations focused on the development of nucleosides with anti-HIV activity that include protein–ligand docking calculations will be summarized.

1.1. Mechanism of action of nucleoside analogues in HIV infected cells. Biological fundamentals

As mentioned in the introduction of the present thesis, nucleoside analogues are prodrugs and must be converted into their triphosphorylated derivatives prior to the interaction with their biological target (HIV RT in the case of anti-HIV nucleosides). Different kinases are involved in this process depending on the virus.

In cells infected by HIV, the three successive phosphoryl transfers on antiviral agents such as AZT or d4T are carried out by human kinases.[74,75] More specifically, the activation process of dT analogues is performed by the following enzymes: thymidine kinase 1 (TK1), thymidylate kinase (TMK), and nucleoside diphosphate kinase (NDPK), which mediate the formation of the mono-, di-, and triphosphorylated anabolites, respectively (Scheme 41).[75]

Scheme 41. Kinases involved in the activation of pyrimidine analogues (represented by dT) in HIV infected cells.

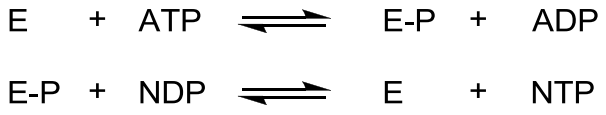
TK1 (EC number:[a] 2.7.1.21) is a cytosolic enzyme, key in human nucleotide metabolism since it catalyses the first phosphorylation step of deoxyribonucleosides in the salvage pathway. Three other deoxyribonucleoside kinases (dNKs), namely

[a] The Enzyme Commission (EC) number is a numerical classification system for enzymes based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). This system assigns to each enzyme a recommended name and number to allow it to be identified, according to the chemical reactions catalysed by the enzyme.

deoxycytidine kinase (dCK, cytosolic), thymidine kinase 2 (TK2, mitochondrial) and deoxyguanosine kinase (dGK, mitochondrial) are also responsible for catalysing this first phosphoryl transfer. In particular, pyrimidine compounds such as dT and dU are phosphorylated by TK1 and TK2, although some studies pointed out that TK2 contributes little to the cellular metabolism.[301] TK1 is well known to activate antiviral prodrugs like AZT and d4T,[74,77] despite it has a narrow substrate acceptance.

TMK (EC number: 2.7.4.9) belongs to the nucleoside monophosphate (NMP) kinase family and it is responsible for catalysing the reversible phosphoryl transfer to dTMP, using ATP as its preferred phosphoryl donor. Moreover, it is also known to participate in the activation of some anti-HIV prodrugs (e.g. AZT and d4T).[75] The NMP kinase family is completed by three other enzymes, named according to their preferred natural substrate: adenylate kinase, guanylate kinase and uridylate-cytidylate kinase. All these enzymes share a similar fold despite having a low primary structure identity.

NDPK (EC number: 2.7.4.6) catalyses the third phosphoryl transfer from a nucleoside triphosphate (usually ATP or GTP) to a diphosphate acceptor (NDP) via an unusual mechanism. While in most kinases the donor and acceptor bind at two distinct sites and normally in a non-ordered manner, in NDPK both donor and acceptor share the same binding site. Therefore, whereas normally the phosphoryl transfer is directly done from the donor to the acceptor, NDPK is known to catalyse the reaction via a ping-pong mechanism, that is, the first product is released from the enzyme before the second substrate combines (Scheme 42).[302,303] Hence, NDPK binds first the donor, then phosphorylates a catalytic histidine residue and finally transfers the phosphoryl group from the His to the diphosphorylated substrate, so that the transfer takes place through two successive steps that imply the formation of a covalent E-P (enzyme-phosphate) intermediate. This mechanism is assumed to be the same for all pyrimidine and purine NDPK substrates.[304]



Scheme 42. Successive ping-pong steps for the activation of diphosphorylated derivatives by NDPK using ATP as the phosphoryl donor. E: enzyme; NTP: nucleoside triphosphate; NDP: nucleoside diphosphate; E-P: enzyme-phosphate.

Remarkably, NDPK is abundant in most living organisms and it is assumed to be the main source of the four triphosphorylated deoxynucleosides (dNTPs) substrates of DNA polymerase. This enzyme is known to present large substrate promiscuity. Indeed, the nature of the base hardly affects the catalytic efficiency and the binding site can also accommodate non-natural bases. By contrast, the activity of NDPK drops dramatically on analogues that lack the 3'-OH as a result of the catalytic mechanism itself.[305] The 3'-OH group of the sugar is responsible for donating a hydrogen bond to the oxygen atom bridging the β- and γ-phosphates.[304,306,307] This oxygen acquires a negative charge when the γ-phosphate is transferred to the catalytic His that is stabilized by the newly formed hydrogen bond, which accelerates the transfer by at least four orders of magnitude.[305,307] Since the catalysis is assisted by the substrate, and more precisely by the 3'-OH of the furanose, substrates lacking this moiety such as AZT or d4T are phosphorylated less efficiently by NDPK, despite they bind in the same way as dT does.[308]

Once the nucleosides have been activated, they must interact with HIV-1 RT to interfere in the replication of the viral nucleic acid. HIV-1 RT is known to have two enzymatic roles, DNA polymerase and ribonuclease (RNase H) activities, but antiviral agents are only aimed at the first one. Moreover, it can employ either RNA or DNA as template.

As a DNA polymerase, HIV-1 RT binds its substrates in an ordered manner: first the DNA template:primer, and then the dNTP that has to be bound to the 3'-primer terminus. A nucleophilic attack of the 3'-OH of the primer terminus to the α-phosphorus of the incoming dNTP leads to its incorporation into the DNA growing chain, accompanied by a pyrophosphate release. Anti-HIV agents are inserted into the viral DNA via this mechanism, interrupting then the replication process.

1.2. Docking studies on nucleoside analogues as anti-HIV agents

Surprisingly, protein–ligand dockings have been scarcely applied in the development of new nucleoside analogues exhibiting anti-HIV activity. In 2009, Mehellou, Balzarini and McGuigan investigated if the first phosphorylation step was behind the poor anti-HIV activity of the nucleoside analogues 2',3'-dideoxyuridine (ddU) and 2',3'-didehydro-2',3'-dideoxyuridine (d4U) (Figure 32) by a combination of molecular modelling and experimental studies.[309] From the docking studies, the

authors suggested that these two nucleoside analogues occupy a position at the active site of TK1 which is not favourable for phosphorylation. Consequently, they found appropriate to bypass the first phosphorylation step by preparing the corresponding phosphoramidates,[a] so that the inactive ddU was turned into a moderately active anti-HIV agent whereas d4U remained inactive. Considering that their enzymatic metabolic assays revealed that d4U phosphoramidates were metabolised to release d4U monophosphate, and since d4U triphosphate is known to be a potent inhibitor of HIV reverse transcriptase, the authors concluded that the bottleneck for the eventual antiviral activity of d4U may reside in the second or third phosphorylation steps.

A combined experimental and molecular modelling study of the acyclic pyrimidine nucleotide **136** (Figure 32) featuring anti-HIV activity was also reported by Balzarini and co-workers in 2010.[310] Specifically, the authors inferred from the experimental data that this particular class of acyclic pyrimidine nucleotides, in which the pyrimidine base is linked through an oxygen atom, were incorporated into the DNA growing chain by HIV-1 RT as purine nucleotide derivatives but not as pyrimidine nucleotides. This hypothesis was demonstrated by their molecular docking studies.

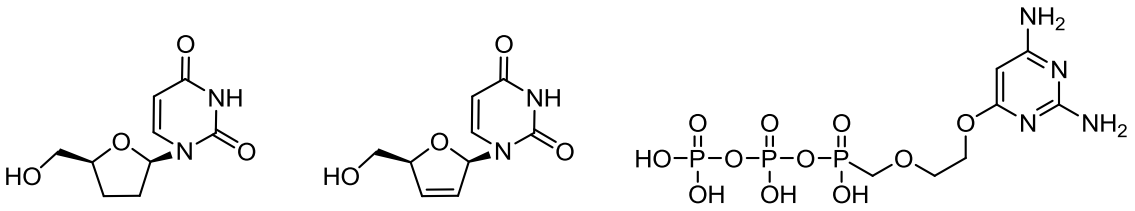


Figure 32. Structures of the nucleoside analogues investigated by means of protein–ligand dockings as anti-HIV agents.

In contrast to the reduced number of docking studies dealing with nucleoside analogues as anti-HIV agents, protein–ligand dockings have been extensively used to investigate the binding orientations of Non-Nucleoside Reverse Transcriptase Inhibitors in HIV-1 RT.[311–320]

[a] Phosphoramidates are used to mask phosphate groups, thus improving the poor membrane permeability observed in nucleosides featuring non-masked phosphate groups.

1.3. Scope

A novel class of nucleoside analogues based on d4T (Figure 33), in which the sugar moiety is conformationally restricted by an ethylene tether between the 2'- and 3'-positions,[321] had been prepared in our research group as anti-HIV agents (Scheme 12, page 45).[203] These analogues, built on a 3-oxabicyclo[3.2.0]heptane scaffold, bear different functionalities in the cyclobutane moiety, namely fluorine, chlorine, hydroxy, carbonyl group and a carbon-carbon double bond, either endo- or exocyclic (**HI1**–**HI3**, Figure 33).

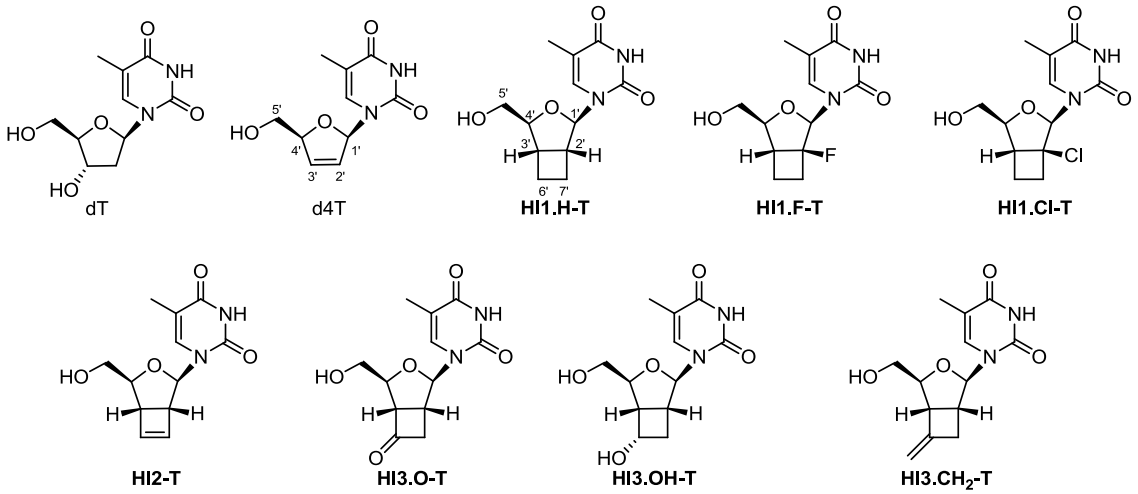


Figure 33. Nucleoside analogues **HI1–HI3** and reference compounds for their study as anti-HIV agents (dT and d4T).

Compounds **HI1–HI3** were evaluated on MT4 cells for anti-HIV-1 activity against wild-type NL4-3 strain as well as for cytotoxicity, using AZT and d4T as the positive control (Table ES7).[a] Unfortunately, none of these compounds was active at concentrations up to 25 $\mu g \cdot mL^{-1}$ despite their structural similarity to the efficient drug d4T.

In order to shed light on the molecular features responsible for this lack of activity, as well as to get a deeper insight into the activation process and mechanism of action of these nucleosides, we decided to perform a molecular modelling study on these nucleosides. To this aim, three of the key features to be met by anti-HIV nucleoside analogues were investigated: their drug-likeness, their conversion into the active form as well as their interaction with HIV-1 RT. As commented, pitfalls in drug development of nucleoside analogues have been

[a] Anti-HIV activity and cytotoxicity were assayed at the Irsicaixa Foundation, Universitat Hospital Germans Trias i Pujol, Badalona (Spain).

normally analysed by protein–ligand dockings on a unique enzyme, generally TK1.[309] Nonetheless, recent publications suggested that TMK could be also a filter into prodrug activation and that the role of NDPK may not be ruled out.[75,309] However, and to the best of our knowledge, theoretical studies that investigate the whole activation process and the subsequent interaction of nucleotides with HIV-1 RT have never been reported.

2. Computational details

Protein–ligand dockings were used to predict all the binding modes and energies in this study. Those calculations were performed with the docking program GOLD (version 5.0.1),[322] using the *ChemScore* scoring function.[278,323] Molecular graphics and visualization of docking results were performed with the *UCSF Chimera* package.[324]

The structures of the ligands were initially optimized with the *Marvin* work package[325] and the Merck molecular force field (MMFF) minimization.[326] Some modifications on the available PDB structures of the enzymes were performed with *UCSF Chimera* prior to docking calculations. Thus, all crystallographic waters, ions and crystallized ligands were removed from each enzyme. Geometrical and hydrogen bonding criteria were used to decide which rotamer of the residues was considered for the calculations. When the structures presented more than one identical subunit, duplicate parts were deleted, keeping exclusively one of the chains for the calculations (Table ES8). Hydrogen atoms were added, and the protonation state of the histidine residues was checked. Atom charges were assigned from Amber.[327] Finally, all the structures were stored in mol2 format and were ready for docking calculations.

Dockings in which the ligand is covalently bound to a DNA chain crystallized with the protein were performed on the structure of HIV-1 RT containing AZTMP (PDB code: 1N5Y). To carry out these calculations, the protein structure is initially prepared as detailed for the rest of the PDB structures used in this study. Then, AZTMP was deleted from the DNA strand and the H atom of the 3'-OH of its previous nucleoside was also deleted. Finally, the structure was stored in mol2. Covalent docking calculations were carried out by setting the remaining 3'-oxygen atom as

the protein link atom (0809), whereas one oxygen atom of the phosphate group of the monophosphorylated nucleoside was defined as the ligand link atom.

For each enzyme, the centre of the binding site in the crystallographic structure was used as the central point for the cavity (Table ES8). The radius of the cavity varies for each enzyme. Structural flexibility of the receptor was taken into account for a series of residues in the binding pocket using the free rotation scheme implemented in GOLD program (Table ES8).[322] Ligand flexibility was also considered for all the studied compounds.

Finally, each compound was docked into the corresponding enzyme and 20 predicted orientations were obtained. For each enzyme, nucleosides were docked in their reactant and product forms separately. The results were analysed both in structural and energetic terms.

3. Study of cyclobutane-fused nucleoside analogues as anti-HIV agents

Results of the present study on **HI1–HI3** analogues will be presented in three different sections, according to the main requirements of anti-HIV agents: drug-likeness, activation process and subsequent interaction with HIV-1 RT.

3.1. Drug-likeness

As commented in the previous chapter, there are many different approaches to address the drug-likeness of molecules employing retrospective analyses of known drug collections.[258,261] With the aim of studying **HI1–HI3** derivatives drug-likeness, we chose two fast and widely used approaches, the "Lipinski's rule of 5" and the Veber's criterion.

Each of these criteria uses standard descriptors to predict oral bioavailability of molecules.[a] Calculations of the standard molecular descriptors of **HI1–HI3**, as well as of the natural nucleoside dT and the antiviral drug d4T, were carried out with the *Marvin* package.[328] The full series of compounds responds positively to both criteria, sustaining their good drug-likeness (Tables ES9 and ES10). Consequently,

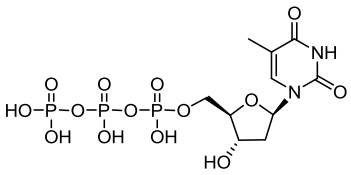
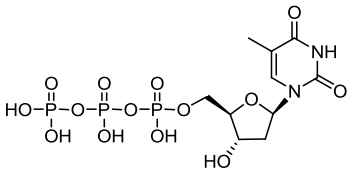
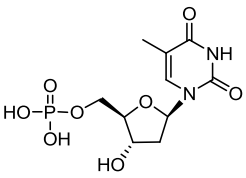
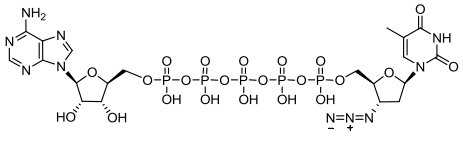
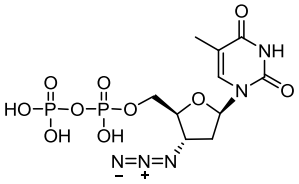
[a] For further details on these molecular descriptors, see section *2. Drug-likeness* of Chapter III.

the lack of activity of these compounds could be explained in terms of either their activation process or their interaction with the target protein, HIV-1 RT.

3.2. Activation

In order to determine the limiting steps in the activation process of the studied analogues, dockings were carried out on TK1, TMK and NDPK. The structures selected to perform the study present these enzymes complexed either with dT or with the anti-HIV drug AZT (Table 9). Moreover, a TMK structure crystallized with the inhibitor $AZTP_5A$ was also used.

Table 9. PDB structures used to study the activation process of nucleosides **HI1**–**HI3**.

Enzyme	Organism	PDB code	Crystallized ligand	Docking calculations[a]
TK1	*Homo sapiens*	1XBT	 dTTP	**HI1**–**HI3** and **HI1MP**–**HI3MP**
TK1	*Homo sapiens*	1W4R	 dTTP	**HI1**–**HI3** and **HI1MP**–**HI3MP**
TMK	*Homo sapiens*	1E2F	 dTMP (+ ADP)	**HI1MP**–**HI3MP** and **HI1DP**–**HI3DP**
TMK	*Homo sapiens*	1E9A	 $AZTP_5A$[b]	**HI1MP**–**HI3MP** and **HI1DP**–**HI3DP**
NDPK	*Dictyostelium discoideum*	1LWX	 AZTDP	**HI1DP**–**HI3DP** and **HI1TP**–**HI3TP**

[a] MP, DP and TP stand for monophosphorylated, diphosphorylated and triphosphorylated compounds, respectively. [b] $AZTP_5A$ is modelled in 1E9A X-ray structure as AZTMP and ATP.[329]

Nucleoside analogues **HI1**–**HI3** were docked separately into the active site of each kinase in their reactant and product forms. Additionally, calculations with the dT and d4T were also performed to provide with structural and energetic benchmarks. Docking results were analysed both in structural and energetic terms. In particular, our criterion was based on checking that a substantial number of low energy binding modes (from the 20 poses predicted for each compound) were consistent with a catalytic orientation, and that the corresponding binding energies were similar or even lower than those of reference.

3.2.1. 1st phosphorylation step

3.2.1.1. TK1 crystallographic structures

There is a reduced number of TK1 X-ray structures available. Among them, we selected the structures featuring dTTP (PDB code: 1XBT,[330] 1W4R[331]) in the binding site to carry out the study of the first phosphorylation step on **HI1**–**HI3** compounds.

TK1 is a homotetramer with 234 amino acids per subunit (Figure B.1, appendix B).[75] Each subunit contains two different domains, the α/β-domain and the small zinc-containing domain. The active site is buried between these two domains (Figure 34). The α/β-domain is formed by six-stranded parallel β-sheet sandwiched between a long α-helix and a flexible loop region on one side, and by three shorter helices on the other side. The Zn^{2+} ion is tetrahedrally coordinated by two cysteine pairs in its region.

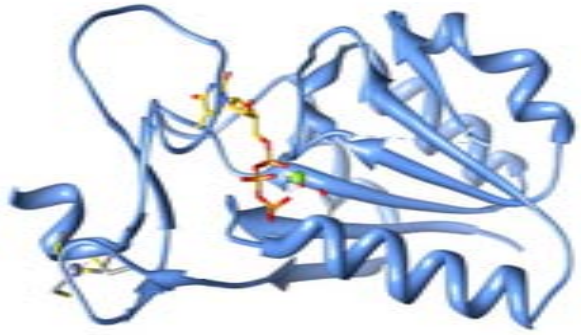


Figure 34. Subunit of TK1 crystallized with dTTP (yellow) in its binding site (PDB code: 1XBT). Mg^{2+} ion is shown in green and Zn^{2+} ion in purple.

It is widely recognized that dNTPs are inhibitors of TK1, binding as bisubstrates at the enzyme active site. Thus, the base and the deoxyribose are posed like a deoxyribonucleoside substrate, and the triphosphate is occupying both acceptor and donor phosphate sites. In particular, the α-phosphate of dTTP is at the same position of the dTMP α-phosphate, while the β-phosphate of dTTP corresponds to the γ-phosphate of ATP transferred to dT.[331] The thymine base is stacked against Arg-165 and Tyr-181 on the one side and against the perpendicularly oriented Phe-101 and Phe-133 on the other side, and it is hydrogen bonded to the main chains of Phe-128, Val-172 and Val-174 (Figure 35, Figure B.2). The 3'-OH of the sugar moiety is interacting via a hydrogen bond with the main chain amino group of Gly-176 and to the side chain of Asp-58. The 5'-oxygen atom of dTTP is close to Glu-98, which normally abstracts the proton of the 5'-OH to activate the substrate for the phosphoryl transfer. Moreover, Lys-32 is assumed to be responsible for the transfer of the γ-phosphate from ATP to the 5'-OH group and Arg-60 is properly located to stabilize the transition state of the reaction.[330,331]

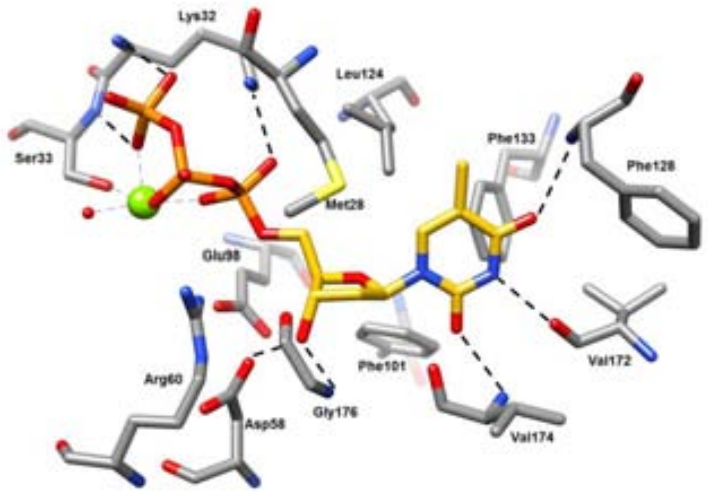


Figure 35. Representation of the main interactions of dTTP in TK1 binding site (PDB code: 1XBT). Hydrogen bonds between dTTP and residues are depicted as dotted lines. Crystallographic waters (except the one coordinated to Mg^{2+}), hydrogen atoms and amino acids Arg-165 and Tyr-181 are not shown for clarity.

3.2.1.2. Docking results

Once the TK1 binding site has been described, this section will be devoted to the docking calculations of the first phosphorylation step of **HI1–HI3** analogues.

First, the X-ray ligand dTTP was docked into the corresponding enzyme structures to validate the docking protocol. The lowest energy orientations were

perfectly overlapped to the dTTP pose in the crystallographic structure, successfully validating the performance of GOLD to be used in this study.

Afterwards, dockings of the reference compounds dT, d4T, dTMP and d4TMP were also carried out. The calculated binding modes showed the same interactions above described for dTTP. In these orientations, the thymine moiety is posed in the same way than that of the X-ray structure and the 5'-OH is pointing towards the catalytic Glu-98 for the unphosphorylated compounds. The main difference between d4T and dT predicted binding modes is the sugar pose, since the lack of the 3'-OH interaction with Gly-176 in d4T causes a slight displacement of its sugar-like moiety and the 5'-OH (Figure B.3). Concerning monophosphorylated derivatives, the predicted orientations also present both the base and the α-phosphate overlapped to the corresponding moieties in the X-ray structure. The binding energy values are -22 and -25 score units for the dT, d4T and dTMP, d4TMP, respectively (Tables ES11 and ES12). Thus, both compounds were used as benchmarks, and their described interactions will be required on **HI1–HI3** analogues to be phosphorylated by TK1.

Regarding **HI1–HI3** docking calculations, two different profiles were observed: while compounds **HI1** and **HI2** are expected to be converted into their monophosphorylated derivatives, the first phosphorylation of **HI3** is unlikely.

Most of the low energy binding modes of **HI1.H-T** in TK1 were consistent with the catalysis. There is a slight displacement of the sugar moiety due to the fused cyclobutane ring, but no clashes are observed between this additional ring and the protein residues (Figure 36).

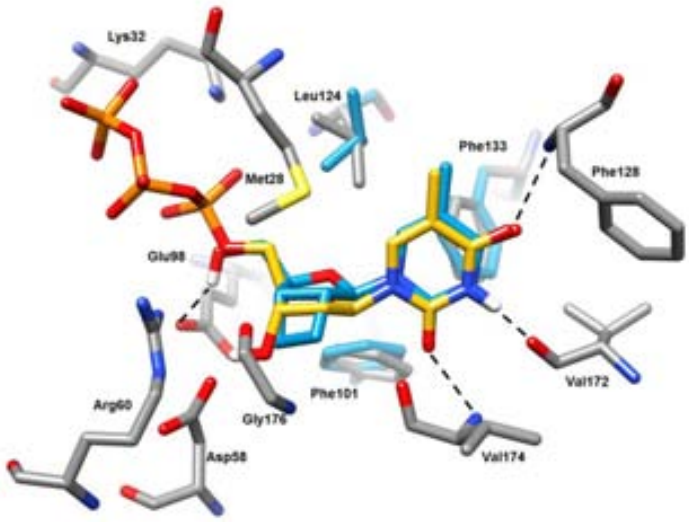


Figure 36. **HI1.H-T** (blue) superimposed to crystallographic dTTP (yellow) in TK1 binding site (PDB code: 1XBT, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between **HI1.H-T** and residues are depicted as dotted lines. (Reproduced by permission of reference 205 © 2012 JOHN WILEY AND SONS, Inc.)

Reinforcing these results, dockings on **HI1.H-TMP** led to almost all the predicted poses properly oriented for the catalysis. Moreover, the corresponding binding energy values (-26 and -27 score units for **HI1.H-T** and **HI1.H-TMP**, respectively), are lower than those of reference (Tables ES11 and ES12).

The unphosphorylated and monophosphorylated compounds featuring a 2'-halogen instead of a hydrogen atom, **HI1.F-T** and **HI1.Cl-T**, also revealed that most of their predicted binding modes were consistent with the catalytic activity of TK1. It is worth mentioning that the presence of this 2'-halogen causes a slight movement of the sugar-like moiety towards the less sterically hindered part of the binding site compared to **HI1.H-T**, but the base and the 5'-OH are still properly oriented (Figure 37). Therefore, TK1 active site has enough space to deal with a 2'-halogen. The corresponding binding energies are similar or even lower than the reference ones, being between -22 and -28 score units (Tables ES11 and ES12).

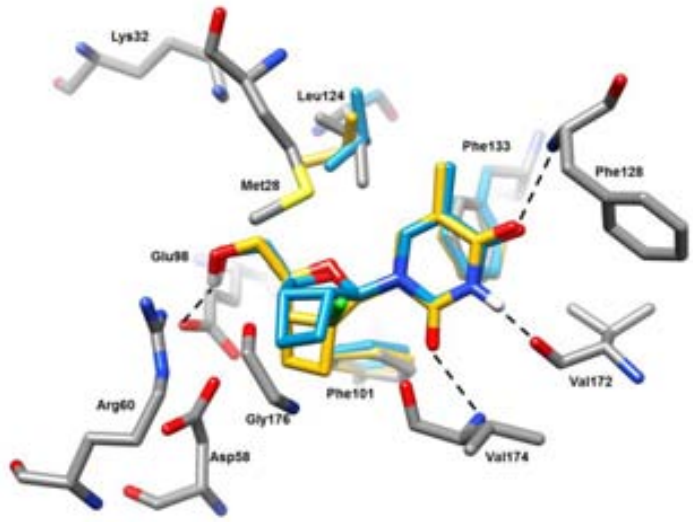


Figure 37. HI1.Cl-T (yellow) superimposed to **HI1.H-T** (blue) in TK1 binding site (PDB code: 1XBT, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between both compounds and residues are depicted as dotted lines.

The presence of an endocyclic double bond in **HI2-T** does not cause changes into the general behaviour of the compound, since most of its low energy poses calculated for both the unphosphorylated and monophosphorylated derivatives were in agreement with a first phosphoryl transfer. Additionally, the corresponding binding energies, around -25 and -28 score units for **HI2-T** and **HI2-TMP**, are lower than those of reference (Tables ES11 and ES12).

By contrast, docking calculations on compounds **HI3**, all of them substituted on the cyclobutane at C-6', led to few binding modes compatible with the catalysis, since these compounds must flip around to fit in the active site. Moreover, binding modes consistent with a catalytic reactivity of the enzyme are higher in energy than dT and d4T (Tables ES11 and ES12).

In particular, most of the low energy orientations predicted for **HI3.O-T**, **HI3.OH-T** and **HI3.CH_2-T** show the base in the 5'-OH pocket and the 5'-OH hydrogen bonded to Asp-58 and Gly-176, mimicking the 3'-OH of dT (Figure 38). This orientation is still more favoured for compound **HI3.OH-T** since there is an extra hydrogen bond between the 6'-OH and the main chain of Val-174. Structural analysis shows that the substitution at the cyclobutane ring in compounds **HI3** causes steric clashes with some amino acids of the TK1 active site when these nucleosides are properly posed for the catalysis (Figure B.4). For **HI3.O-T** and **HI3.CH_2-T**, significant bad contacts are observed between the carbonyl or methylene group bound to the

sugar-like moiety and Asp-58, whereas in ligand **HI3.OH-T** they occur between the hydroxyl group of the cyclobutane moiety and Arg-60.

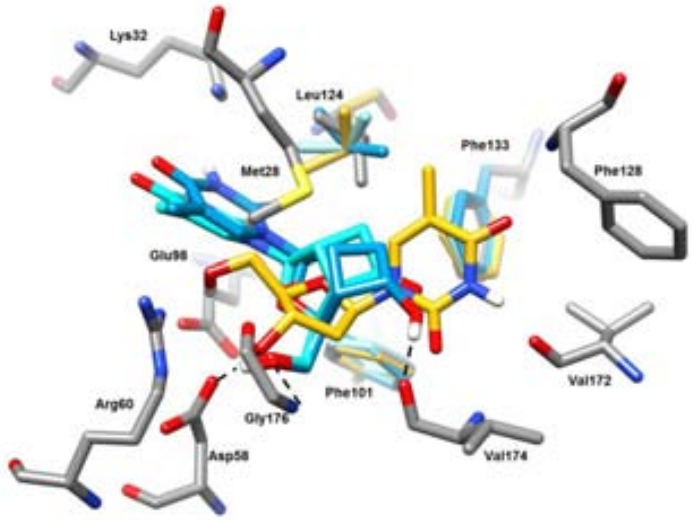


Figure 38. HI3.O-T (cyan) superimposed to **HI3.OH-T** (blue) and dT (yellow) in TK1 binding site (PDB code: 1XBT, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between compounds and residues are depicted as dotted lines. (Reproduced by permission of reference 205 © 2012 JOHN WILEY AND SONS, Inc.)

Surprisingly, a higher number of binding modes properly posed for the catalysis was obtained for monophosphorylated **HI3.O-TMP**, **HI3.OH-TMP** and **HI3.CH$_2$-TMP**. Structural analysis shows that this behaviour is consistent with the additional hydrogen bonds observed between the α-phosphate and the nearby residues, such as Phe-29, Lys-32 and Arg-60 (Figure 39). However, the corresponding binding energies are higher than the dT and d4T ones (Table ES12). These results together with the observed flip of the unphosphorylated compounds make the phosphoryl transfer not plausible to occur.

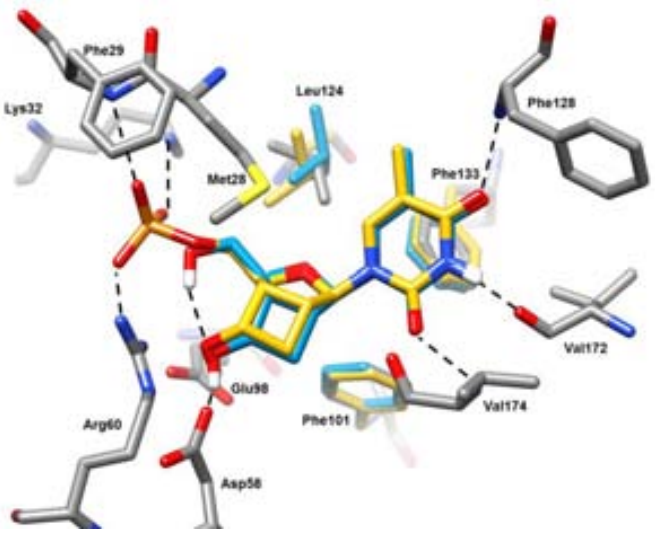
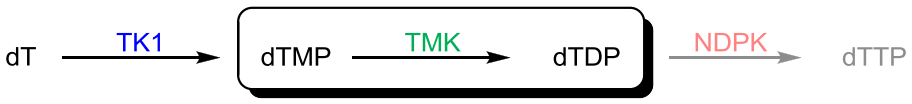


Figure 39. HI3.OH-TMP (yellow) superimposed to **HI3.OH-T** (blue) in TK1 binding site (PDB code: 1XBT, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between compounds and residues are depicted as dotted lines.

To sum up, compounds substituted at the C-6' of the cyclobutane ring, namely **HI3.O-T**, **HI3.OH-T** and **HI3.CH$_2$-T**, are not likely to be phosphorylated by TK1, whereas the rest of compounds (**HI1.H-T**, **HI1.F-T**, **HI1.Cl-T** and **HI2-T**) are expected to be converted into their monophosphorylated derivatives.

3.2.2. 2nd phosphorylation step



3.2.2.1. TMK crystallographic structures

Nucleoside monophosphate kinases are known to undergo large conformational changes upon substrate binding.[332] Thus, the X-ray structures selected to perform the study of the second phosphorylation step correspond to TMK complexed with ADP and dTMP (PDB code: 1E2F)[333] as well as to TMK crystallized with the inhibitor AZTP$_5$A (PDB code: 1E9A).[329]

Human TMK is a homodimer consisting of 212 amino acids per subunit. Each monomer is formed by a five-stranded parallel β-sheet core surrounded by nine α-helices (Figure 40).

Figure 40. TMK crystallized with dTMP (yellow) and ADP (green) in its binding site (PDB code: 1E2F). Mg^{2+} ion is shown in green.

In the X-ray structure of TMK complexed with the substrate dTMP and ADP, the thymine moiety is stacked against Phe-72 and forming a hydrogen bond with Arg-76, whereas the sugar is placed in a subpocket with enough space to deal with sugar-like moieties bulkier than deoxyribose (Figure 41, Figures B.5 and B.6). The α-phosphate is facing the ADP phosphates and hydrogen bonded to Arg-97.

Regarding the inhibitor $AZTP_5A$, which is modelled as AZTMP and ATP in the corresponding X-ray structure,[329] its orientation in the TMK binding site is mainly the same as for dTMP (Figure 41). The substrate and donor clefts are occupied by the corresponding donor and acceptor nucleosides, which are linked by a chain of five phosphate groups.

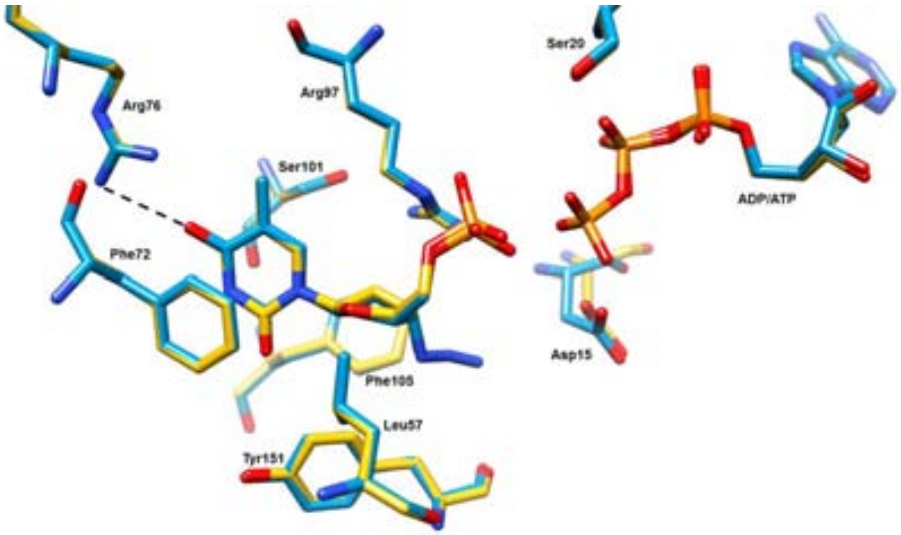


Figure 41. X-ray structure of dTMP (PDB code: 1E2F, in yellow) superimposed to AZTMP (PDB code: 1E9A, in blue) in TMK binding site. For the sake of clarity, hydrogen atoms, ions and crystallographic waters are not shown. Hydrogen bonds between compounds and residues are depicted as dotted lines.

3.2.2.2. Docking results

The docking protocol was successfully validated by performing docking calculations of the X-ray ligands into the corresponding TMK structures (Table 9, page 102), which binding modes closely matched the crystallographic structures.

Orientations consistent with the catalytic activity of the enzyme were also obtained for dT and d4T, both in their reactive and product forms (Figure 42). Surprisingly, the binding energy of the best posed dTMP orientation (-9 score units, Table ES13) was higher than that of d4TMP (-14 score units) despite its good similarity to the crystallographic structure (Figure B.7). The same behaviour was predicted for the corresponding diphosphorylated derivatives. Therefore, d4T was taken as the energetic benchmark for the investigation of this second phosphoryl transfer.

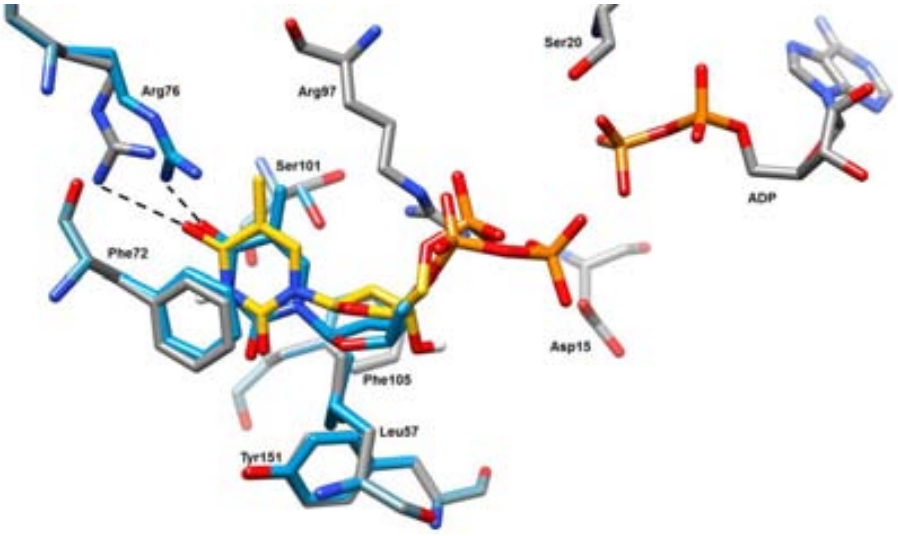


Figure 42. d4TDP (blue) superimposed to crystallographic dTMP (yellow) in TMK binding site (PDB code: 1E2F, X-ray residues shown in grey). For the sake of clarity, crystallographic waters and ions are not shown, and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between compounds and residues are depicted as dotted lines.

Regarding monophosphorylated compounds **HI1**–**HI3**, all of them are likely to be further phosphorylated by TMK. Docking calculations on these compounds led to low energy orientations properly posed for the catalysis. Moreover, the corresponding binding energies are similar or even lower than those of reference (from -14 to -18 score units, Tables ES13 and ES14).

A structural analysis of the predicted poses for **HI1MP**–**HI3MP** shows no major differences among all these compounds. When comparing these binding modes

with d4TMP, only a slight displacement of the base and the sugar-like moiety is observed. It is worth highlighting that the extra hydroxyl group on the C-6' of **HI3.OH-TMP** is hydrogen bonded to the side chain of Asp-15, which favours even more this orientation consistent with the phosphoryl transfer (Figure 43).

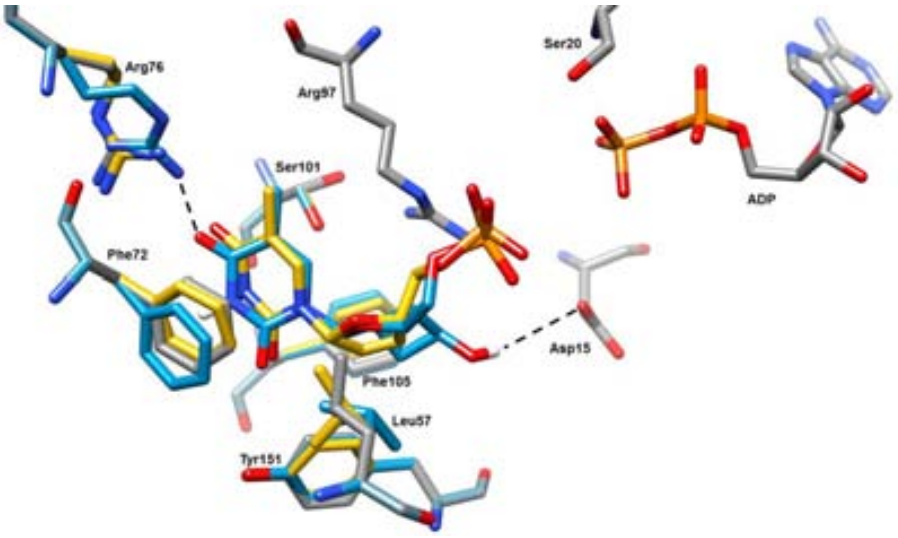


Figure 43. HI3.OH-TMP (blue) superimposed to d4TMP (yellow) in TMK binding site (PDB code: 1E2F, X-ray residues shown in grey). For the sake of clarity, crystallographic waters and ions are not shown, and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between **HI3.OH-TMP** and residues are depicted as dotted lines.

A similar behaviour is predicted for **HI1DP–HI3DP**, since all of them are posed with the thymine stacked against Phe-72 and hydrogen bonded to Arg-76, as well as the β-phosphate pointing towards ADP, such as d4TDP. Again, the hydroxyl group of **HI3.OH-TDP** is interacting with Asp-15 via a hydrogen bond.

Therefore, TMK binding site has enough room to deal with the fused cyclobutane moiety and further substitutions on the C-6', causing no remarkable differences among the mono- and diphosphorylated **HI1–HI3** calculated structures. Since these orientations are in agreement with those of d4T, our calculations showed that this enzyme does not represent a bottleneck in the activation process of nucleosides **HI1–HI3**.

3.2.3. 3rd phosphorylation step

3.2.3.1. NDPK crystallographic structure

NDPK is a homo-hexamer with short (around 150 residues) polypeptide chains. All the six subunits are identical (Figure 44).[334]

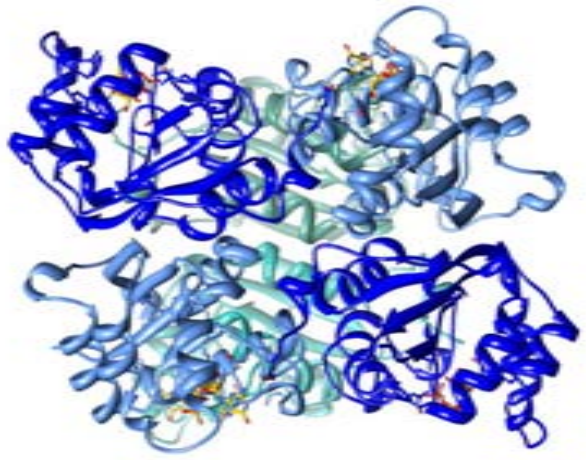


Figure 44. NDPK crystallized with GDP (yellow) in its binding site (PDB code: 1NUE).

Unfortunately, X-ray structures of human NDPK are only available with purine nucleosides, and our docking calculations of dT on these structures led to inconsistent binding modes. Therefore, dockings relative to the last phosphorylation were performed on a structure of the slime mold *Dictyostelium discoideum* NDPK complexed with AZTDP (PDB code: 1LWX).[308] This species has shown in many aspects similar behaviour with its human counterpart, showing an analogous three-dimensional structure, especially at the active site where all the residues are conserved and make the same interactions with nucleotide substrates.[334]

In the complex with AZTDP, the thymine base is pointing outside the protein whereas the phosphate groups are oriented deeper inside and towards the catalytic His-122 (Figure 45, Figure B.8). Thymine lies in between the side chains of Phe-64 and Val-116, and its polar groups make no direct interactions with the protein residues. The β-phosphate is hydrogen bonded to Arg-92 and Arg-109, and a magnesium ion is coordinated to the α- and β-phosphates of AZTDP.[308]

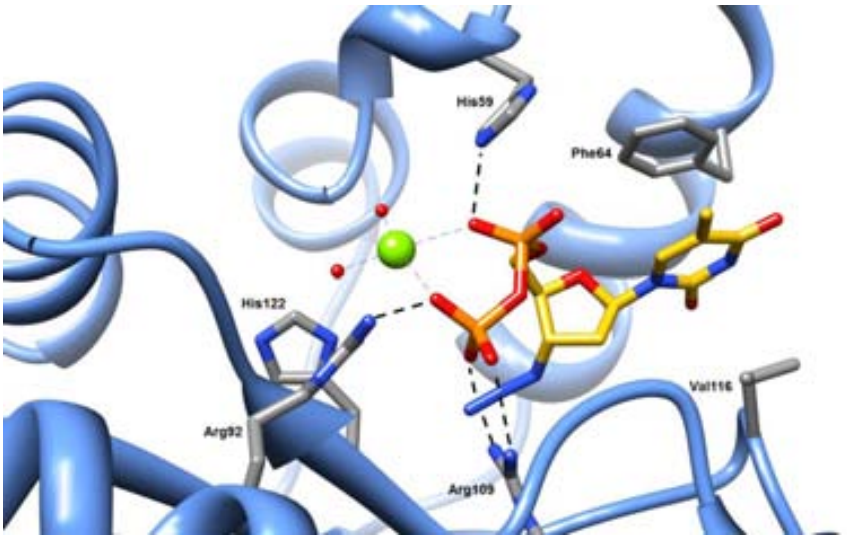


Figure 45. Representation of the main interactions of AZTDP in *Dictyostelium discoideum* NDPK binding site (PDB code: 1LWX). Hydrogen bonds between AZTDP and residues are depicted as dotted lines. The crystallographic ion Mg^{2+} is shown in green. Hydrogen atoms are not shown for clarity.

3.2.3.2. Docking results

To validate the docking protocol, AZTDP and AZTTP were docked into the X-ray structure. The first calculations were carried out considering four flexible residues, namely His-59, Leu-68, Thr-98 and Arg-109, as well as defining the binding site from a 10 Å cavity generated around Lys-16. Disappointingly, none of the predicted binding modes reproduced the crystallographic structure, and a large variability on the orientations was observed, probably because of the limitations of docking in dealing with large solvent exposed active sites.

Thus, docking calculations without considering any flexible residue were performed, but despite the fact that the corresponding calculated poses revealed less variability than before, the perfect reproduction of the X-ray structure was not achieved. Finally, different ways of defining the binding site were tested (i.e. changing the central residue, the size of the binding site or even defining the central point of the cavity from the X-ray ligand pose instead of from a residue). The best results were obtained when having as a central point the AZTDP X-ray pose and a 6 Å cavity generated around it. Carrying out the docking calculations under these conditions, some AZTDP calculated complexes were overlapped to the X-ray structure. However, the reproduction of the crystallographic structure was not as good as in the previous activation steps, so that predictions on this last phosphoryl transfer should be read with caution.

To provide with energetic and structural benchmarks, dTDP and d4TDP were docked, and most of their binding modes showed the base slightly displaced towards the solvent region and the phosphates not exactly overlapped to those of AZTDP in the X-ray structure, but still oriented towards the catalytic His-122 (Figure 46). Some predicted orientations of dTTP were also compatible with the catalysis. However, d4TTP did not lead to any favourable binding mode. Therefore, dTDP and dTTP were taken as energetic and structural references for this last step, having binding energies of -18 and -11 score units, respectively (Tables ES15 and ES16).

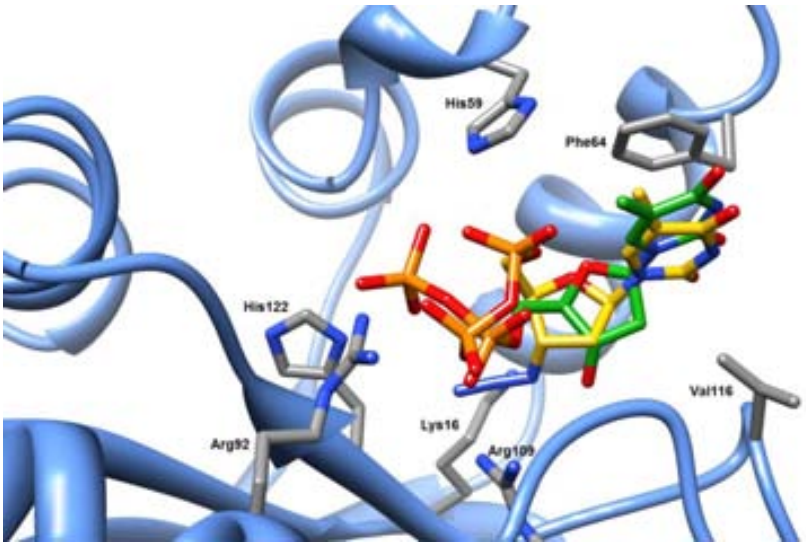


Figure 46. dTDP (green) superimposed to crystallographic AZTDP (yellow) in *Dictyostelium discoideum* NDPK binding site (PDB code: 1LWX, X-ray residues shown in grey). For the sake of clarity, hydrogen atoms, hydrogen bonds, crystallographic waters and ions are not shown.

Considering that NDPK catalysis goes through the initial phosphorylation on His-122 and the subsequent phosphoryl transfer to the substrate, docking results were structurally analysed focusing mainly in the orientation of the β-phosphate towards His-122. Following this criterion, the phosphorylation of compounds **HI3.0-T** and **HI3.0H-T** is more favoured than for the rest of the analogues.

Dockings of **HI3.0-TDP**, **HI3.0H-TDP**, **HI3.0-TTP** and **HI3.0H-TTP** led to some orientations similar to those of dTDP and dTTP (Figure 47), and with binding energies lower than those of reference (Tables ES15 and ES16). It is worth highlighting the role of the oxygen atom bound to the C-6' of the cyclobutane ring of **HI3.0-T** and **HI3.0H-T**: the extra hydrogen bond formed between this atom and the side chain of Lys-16, mimicking the interaction observed between the same residue and the 3'-OH of dTDP, seems to favour this orientation compatible with the catalysis.

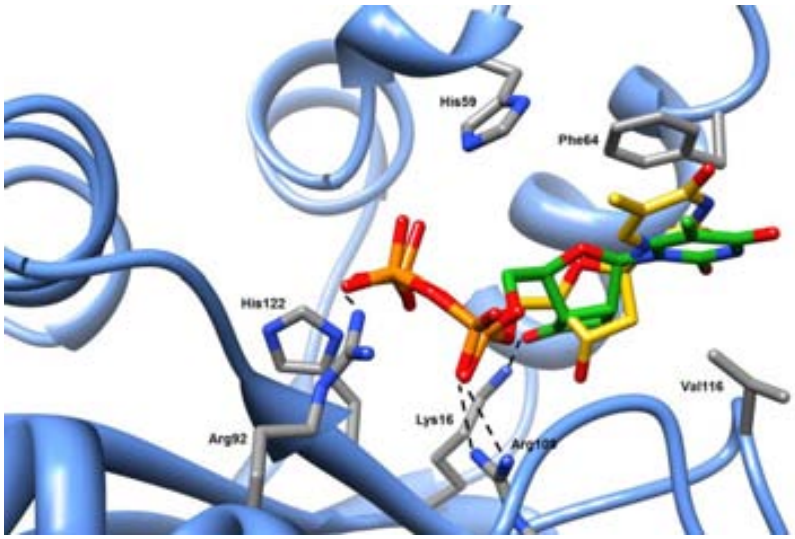


Figure 47. HI3.O-TDP (green) superimposed to dTDP (yellow) in *Dictyostelium discoideum* NDPK binding site (PDB code: 1LWX, X-ray residues shown in grey). For the sake of clarity, hydrogen atoms, crystallographic waters and ions are not shown. Hydrogen bonds between **HI3.O-TDP** and residues are depicted as dotted lines.

On the other hand, none of the binding modes of **HI1**, **HI2** and **HI3.CH$_2$-** was consistent with their last phosphoryl transfer. The predicted orientations of the diphosphorylated derivatives presented either the β-phosphate oriented towards His-122 but the base completely outside of the cleft or a 180º flip of the compounds, leaving the phosphates and the base not properly posed for the catalysis (Figure 48). Reinforcing these results, the flip of the diphosphorylated derivatives was also predicted in most of the binding modes calculated for the triphosphorylated compounds.

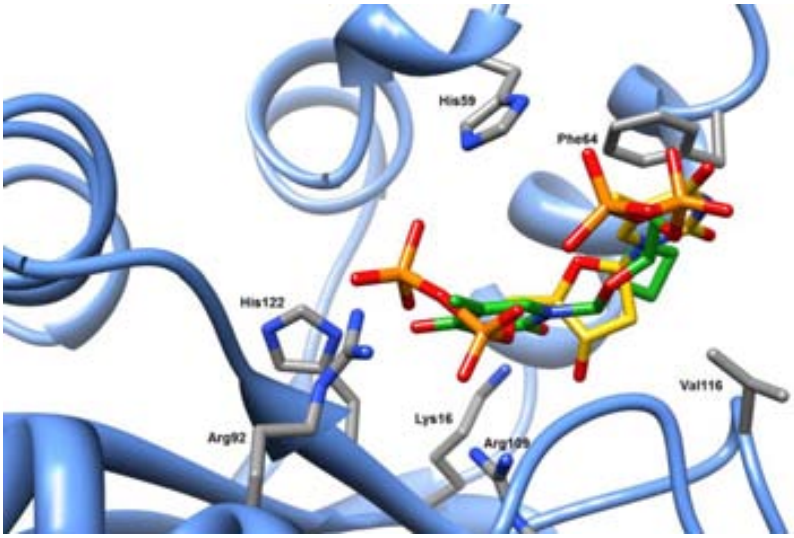


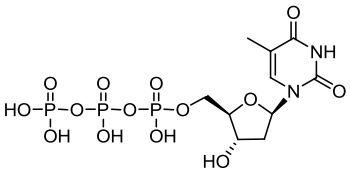
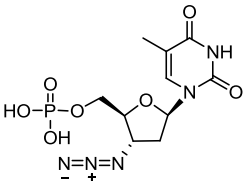
Figure 48. HI1.H-TDP (green) superimposed to dTDP (yellow) in *Dictyostelium discoideum* NDPK binding site (PDB code: 1LWX, X-ray residues shown in grey). For the sake of clarity, hydrogen atoms, hydrogen bonds, crystallographic waters and ions are not shown.

In summary, compounds bearing a cyclobutane C-6' substitution (**HI3.O-T**, **HI3.OH-T** and **HI3.CH_2-T**) are not expected to be phosphorylated in the first activation step, whereas the less favoured transfer in the activation process of the rest of the nucleosides (**HI1.H-T**, **HI1.F-T**, **HI1.Cl-T** and **HI2-T**) is the third one. Nonetheless, to overcome the uncertainties of the last kinase step, as well as to complete the full molecular profile considering the mode of action of these anti-HIV analogues, molecular insight into whether the activated nucleosides could interact with their target, HIV-1 RT, is required.

3.3. HIV-1 RT interaction

To study the interaction of nucleosides **HI1–HI3** with HIV-1 RT, docking calculations were carried out on the structures indicated in Table 10. As before, these nucleosides were docked separately into the active site of HIV-1 RT and the results were analysed in structural and energetic terms.

Table 10. PDB structures used to study the interaction of compounds **HI1**–**HI3** with HIV-1 RT.

Enzyme	Organism	PDB code	Crystallized ligand	Docking calculations[a]
RT	HIV-1	1RTD	 dTTP	**HI1TP–HI3TP**
RT	HIV-1	1N5Y	 AZTMP	**HI1MP–HI3MP**

[a] MP and TP stand for monophosphorylated and triphosphorylated compounds, respectively.

3.3.1. HIV-1 RT crystallographic structures

HIV-1 RT is a heterodimer formed by two related chains, a 66-kD subunit (p66) and a 51-kD subunit (p51). The p66 subunit is the one having the polymerase and RNase H active sites. It is folded into five separate domains, the RNase H domain

and four subdomains of the polymerase active site, namely the fingers, palm, thumb and connection (Figure 49).[335] The DNA duplex binds along a groove, about 60 Å in length, stretching from the polymerase to the RNase H active site. Remarkably, binding of the template:primer and dNTP induces significant conformational changes in parts of p66, mainly bending the fingers domain towards the palm. As a result, a number of residues are brought into contact with the incoming triphosphorylated nucleoside.[336]

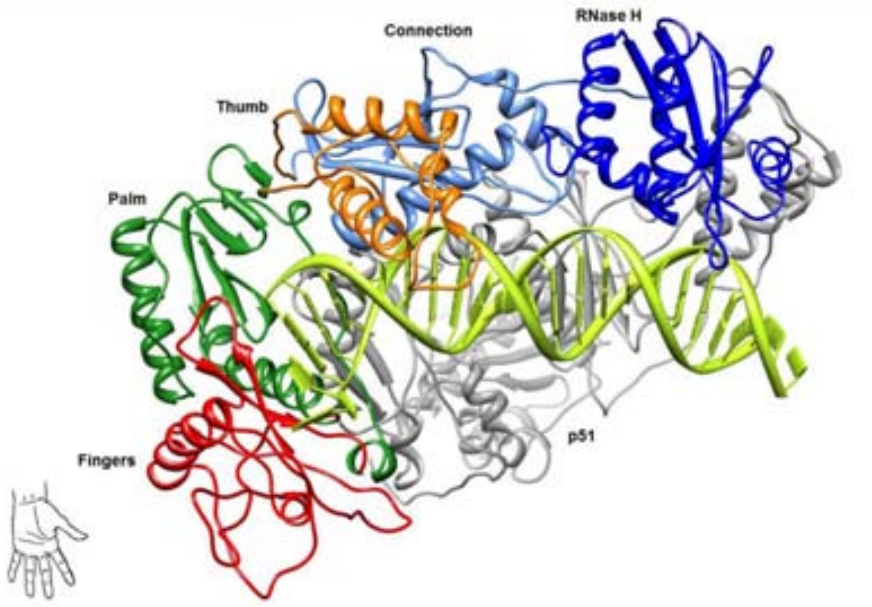


Figure 49. HIV-RT crystallized with DNA template and primer strands (green). Subunit p66 domains are shown in different colours: fingers (red), palm (green), thumb (orange), connection (cyan) and RNase H domain (blue). p51 subunit is shown in grey. Figure inspired by Kohlstaedt et al.[335] and Huang et al.[336] articles. (Reprinted with permission from AAAS)

The first X-ray structure selected to study the interaction of the synthesized analogues with the HIV-1 RT corresponds to the binding of dTTP in the cleft prior to DNA insertion (PDB code: 1RTD).[336] This structure is proposed by Huang et al. as a snapshot of the stage just after dNTP binding to HIV-1 RT but before being introduced to the DNA growing chain. The last nucleotide bound to the 3'-primer terminus in the X-ray structure is a dideoxynucleotide, which is thus unable to attack the incoming dTTP. However, the crystal of this catalytic complex contains bound dTTP precisely in the expected position for being attacked by the (missing) 3'-OH. Moreover, it accounts for the described conformational changes of HIV-1 RT that occur upon binding of the DNA template:primer and dNTP. In this structure, the incoming dTTP is pairing with the templating base (Figure 50, Figure B.9). Its thymine moiety is stacked against the base of the dideoxynucleotide bound at the

3'-terminus of the primer strand, almost as in a continuous DNA strand. Additionally, the phosphates are interacting via hydrogen bonds with the residues Lys-65, Arg-72, Asp-113 and Ala-114 and with two Mg^{2+} ions. The 3'-OH is hydrogen bonded to the main chain of Tyr-115 and to the β-phosphate.

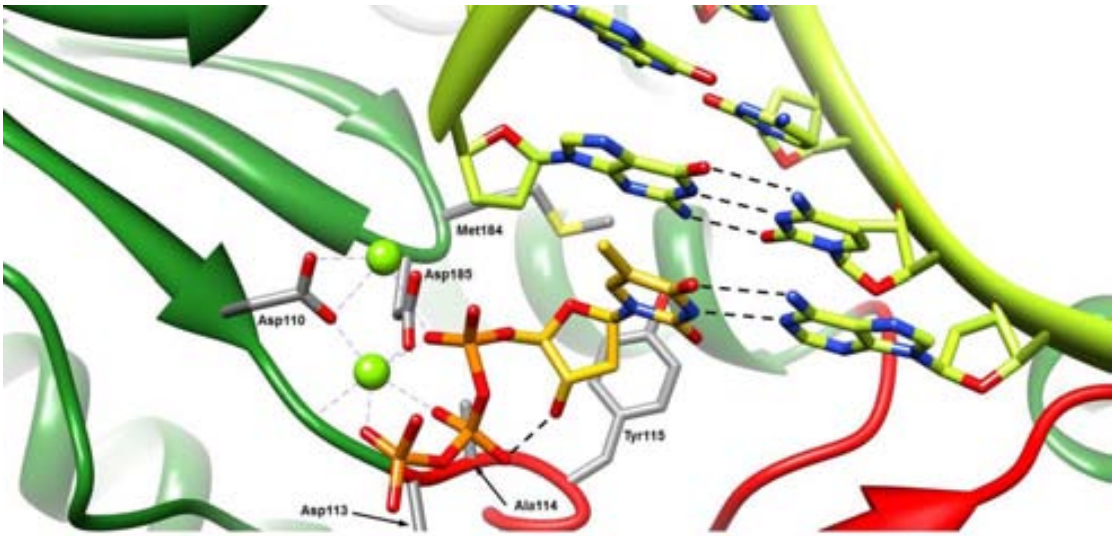


Figure 50. Representation of the main interactions of dTTP (yellow) in complex with HIV-1 RT (PDB code: 1RTD, X-ray residues shown in grey, DNA chains in green, backbone of the palm domain in forest green and backbone of the fingers domain in red). Hydrogen bonds between dTTP, residues and pairing DNA bases are depicted as dotted lines. Crystallographic waters, hydrogen atoms and amino acids Lys-65 and Arg-72 are not shown for clarity.

The second crystallographic structure used in the present study shows HIV-1 RT with AZTMP bound to a terminated DNA strand (PDB code: 1N5Y).[67] Hence, it represents a snapshot of the HIV-1 RT after the incorporation of an anti-HIV drug into the DNA growing strand. In this structure, AZTMP is paired to the corresponding nucleotide of the template and its azido group is interacting through the nitrogen atom bound to the ribose ring with Asp-185 (Figure 51).

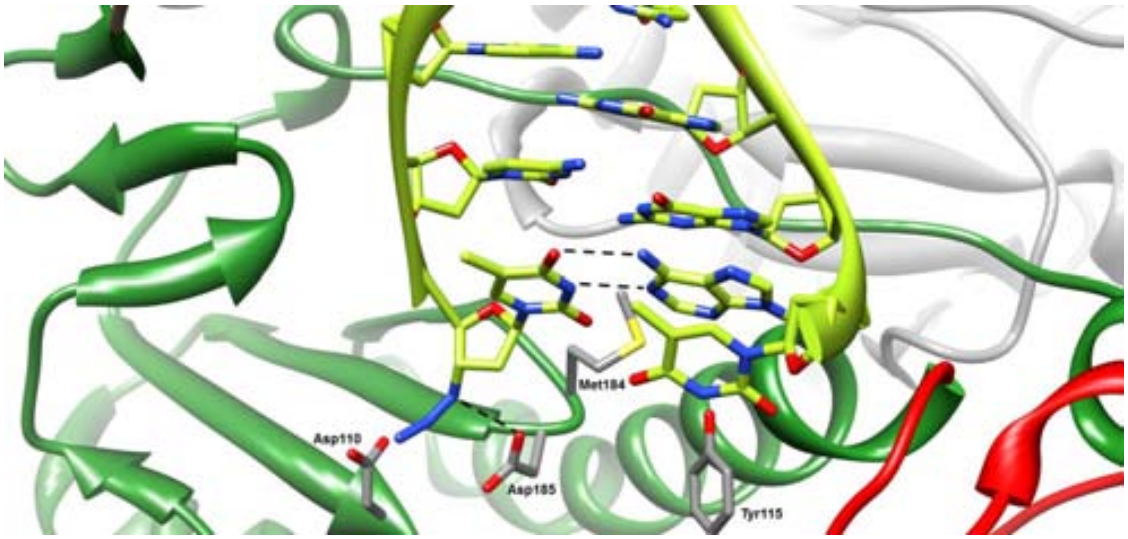
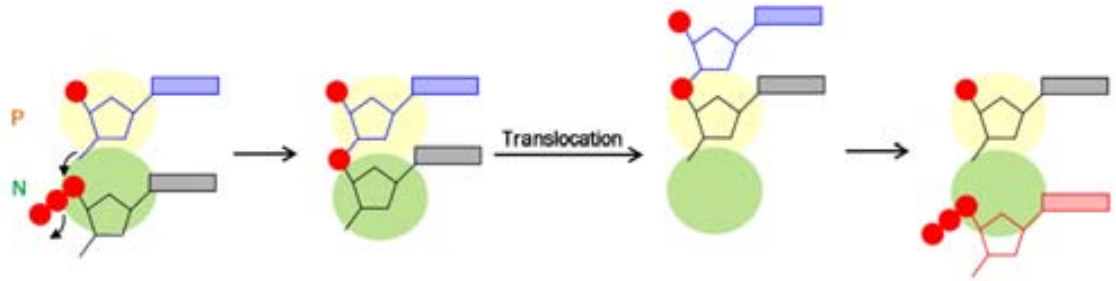


Figure 51. Representation of the main interactions of AZTMP bound to a DNA terminated chain (green) in complex with HIV-1 RT (PDB code: 1N5Y, X-ray residues shown in dark grey, backbone of the palm domain in forest green, backbone of the fingers domain in red and backbone of p51 subunit in light grey). Hydrogen bonds between AZTMP, residues and pairing DNA bases are depicted as dotted lines. Crystallographic waters and hydrogen atoms are not shown for clarity.

Regarding the location of the nucleotide that has to be incorporated into the DNA growing chain, two main sites can be defined in HIV-1 RT: the nucleotide and the priming site. The dNTP is initially located at the nucleotide site. Once it is bound to the DNA strand, the so-called translocation process takes place moving the newly incorporated dNMP to the priming site, so that the nucleotide site can accept the next dNTP (Scheme 43). According to this nomenclature, in 1RTD X-ray structure the dTTP is posed at the nucleotide site, whereas in 1N5Y the AZTMP is located at the priming site.



Scheme 43. DNA polymerization mechanism in HIV-1 RT, showing the priming site in yellow and the nucleotide site in green. Phosphate groups are depicted as red circumferences and nucleobases as different colour-filled rectangles. (Adapted by permission from Macmillan Publishers Ltd: EMBO Journal, reference 67, © 2002).

3.3.2. Docking results

As above mentioned, the structure featuring a dTTP in the nucleotide site (PDB code: 1RTD) was used to carry out the study of the triphosphorylated derivatives just before the DNA insertion, while the crystallographic complex of HIV-1 RT with the AZTMP-terminated DNA (PDB code: 1N5Y) was used to modelate the interactions of the monophosphorylated derivatives once bound to the primer DNA strand. Thus, the study of the triphosphorylated derivatives will be first presented, followed by the calculations on the monophosphorylated compounds covalently bound to the DNA chain.

The docking protocol was successfully validated by carrying out calculations of dTTP and AZTMP into the corresponding X-ray structures. Remarkably, some differences on the phosphates position of dTTP were observed, but most of the low energy complexes featured the α-phosphate properly posed for the catalysis as well as the base pairing with the corresponding adenine.

The antiviral compound d4TTP was also docked in 1RTD to provide with an additional structural and energetic benchmark, and the calculated complexes matched quite well with the X-ray data (Figure B.10). The binding energies of dTTP and d4TTP, which are -13 and -15 score units, were taken as reference for the study of the triphosphorylated derivatives (Table ES17).

Dockings of **HI1TP–HI3TP** proved that none of the compounds is likely to interact with HIV-1 RT. Most of the predicted binding modes were not consistent with the dTTP and d4TTP structures. They showed either the compounds flipped around (Figure 52) or the base displaced and the α-phosphate too far away from the previous nucleotide, preventing thus the attack of the 3'-OH.

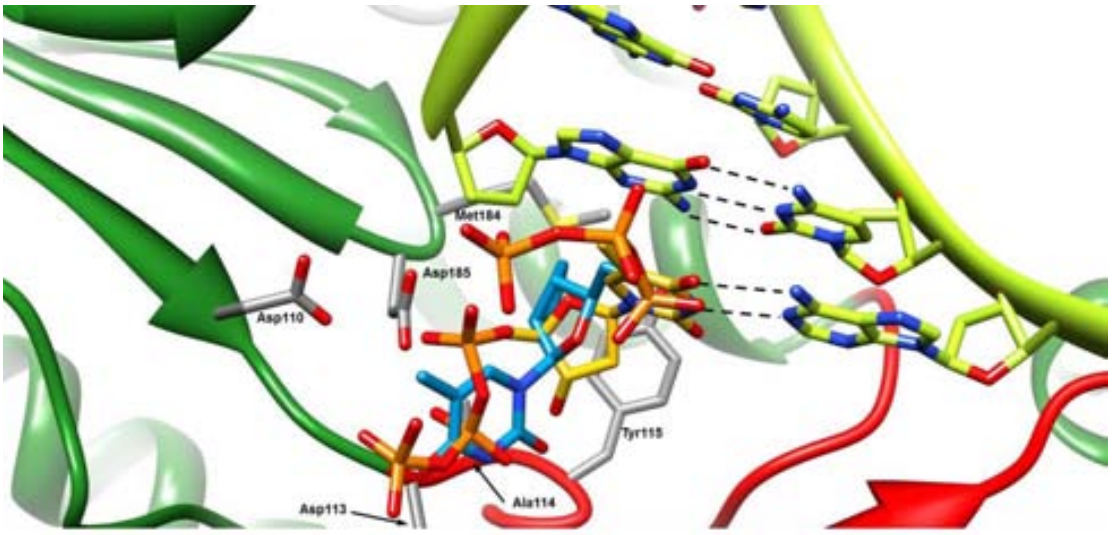


Figure 52. HI1.H-TTP (blue) superimposed to crystallographic dTTP (yellow) in HIV-1 RT nucleotide site (PDB code: 1RTD, X-ray residues shown in grey, DNA chains in green, backbone of the palm domain in forest green and backbone of the fingers domain in red). Hydrogen bonds are depicted as dotted lines. Crystallographic waters, ions, hydrogen atoms and flexible residues are not shown for clarity. (Reproduced by permission of reference 205 © 2012 JOHN WILEY AND SONS, Inc.)

Reinforcing these results, the very few orientations compatible with the binding of these analogues to the DNA growing chain afford with binding energies around -8 score units, which are higher than those of reference (Table ES17). Structural analysis shows that when these compounds are properly posed there is a steric hindrance between the cyclobutane moiety and the residues Ala-114 and Tyr-115 of the active site (Figure 53).

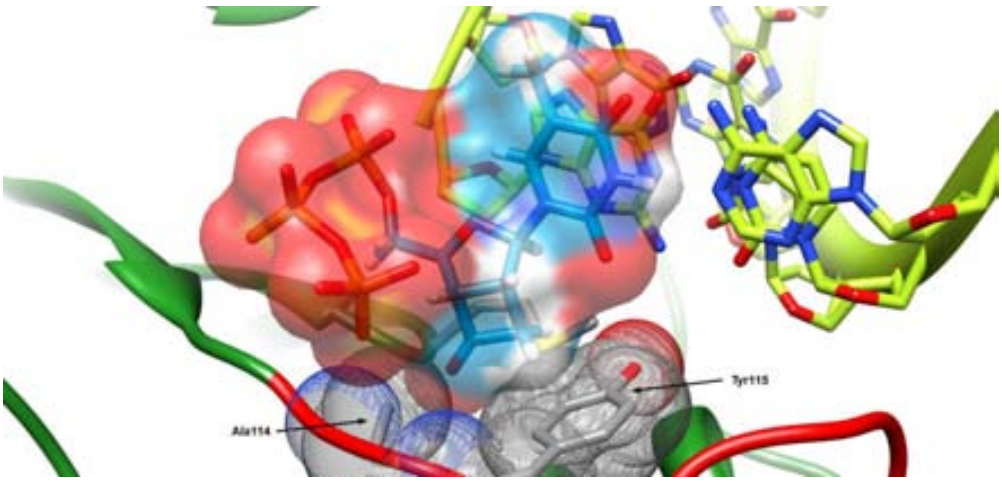


Figure 53. Van der Waals surface representation of **HI3.0-TTP** (blue) and residues (grey) featuring steric hindrance with this compound in HIV-1 RT nucleotide site (PDB code: 1RTD, DNA chains shown in green, backbone of the palm domain in forest green and backbone of the fingers domain in red). Crystallographic waters, ions and hydrogen bonds are not shown for clarity.

Regarding the study of the monophosphorylated derivatives covalently bound to the DNA chain, dTMP and d4TMP were also taken as reference since most of their predicted orientations were consistent with the crystallographic structure containing AZTMP. The corresponding binding energies are around -26 score units (Table ES18).

Dockings of **HI1MP–HI3MP** revealed that compounds **HI1** and **HI2** would be more likely to be bound to the DNA primer chain than **HI3**. A high number of complexes properly posed was obtained for compounds **HI1** and **HI2** (Figure 54). Furthermore, their binding energies were similar to the reference ones (Table ES18).

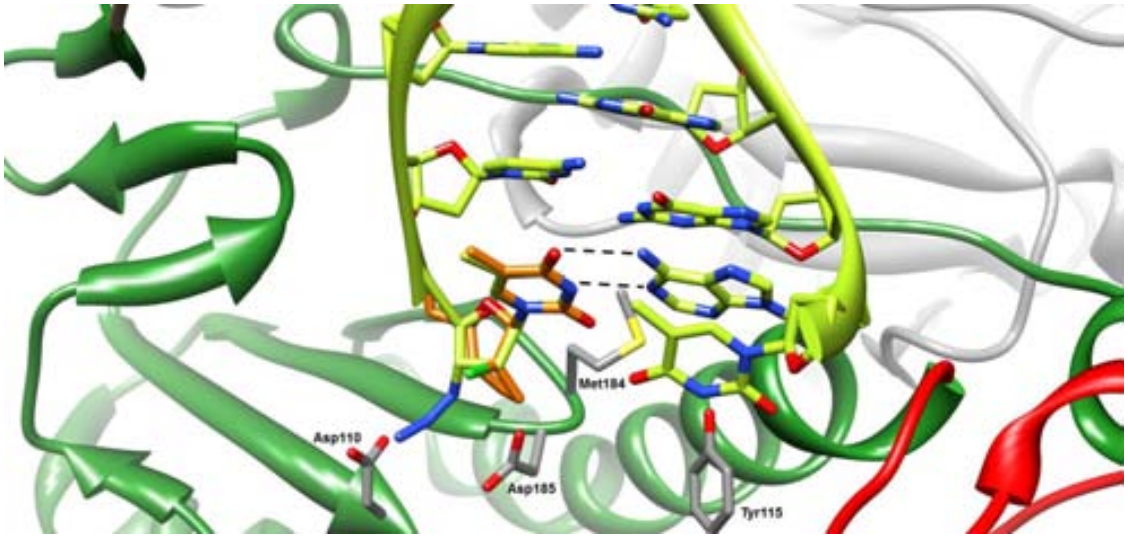


Figure 54. HI1.Cl-TMP (orange) covalently bound to the AZTMP-terminated DNA chain superimposed to crystallographic AZTMP (green) in HIV-1 RT priming site (PDB code: 1N5Y, X-ray residues shown in grey, DNA chains in green, backbone of the palm domain in forest green, backbone of the fingers domain in red and backbone of p51 subunit in light grey). Hydrogen bonds are depicted as dotted lines. Crystallographic waters and hydrogen atoms are not shown for clarity. (Reproduced by permission of reference 205 © 2012 JOHN WILEY AND SONS, Inc.)

Conversely, compounds featuring an additional substitution at C-6' (**HI3.O-TMP**, **HI3.OH-TMP** and **HI3.CH_2-TMP**) led to a lower number of predicted orientations matching the crystallographic structure. Many of their calculated binding modes show the base and the sugar-like moiety outside their pockets. Moreover, the binding energy values of the properly posed complexes are slightly higher than for the rest of compounds (Table ES18). Therefore, as observed in the first activation step, the substitution at C-6' seems to work against an appropriate binding.

4. Conclusions

The lack of anti-HIV activity of nucleosides **HI1–HI3** was investigated by means of molecular docking. To the best of our knowledge, this is the first study that explicitly considers the simulation of the entire activation process as well as of the HIV-1 RT interaction to rationalize antiviral activities.

Moreover, studies on the drug-likeness of these compounds were also performed. The evaluation of their drug-likeness confirmed that this parameter is not restrictive for the activity of these compounds.

The theoretical calculations presented so far demonstrated that the presence of a cyclobutane fused to the 2'- and 3'-positions of the sugar moiety does not work against an appropriate phosphoryl transfer by the first kinase (TK1), although further substitution on the cyclobutane at C-6' halts the prodrug activation process at the first step. Calculations also showed that TMK has low specificity, and no filtering seems to be associated with this second phosphoryl transfer. Finally, docking studies on the last activation step revealed that compounds bearing an oxygen atom bound to the C-6' of the cyclobutane ring (**HI3.O-T** and **HI3.OH-T**) would be more likely to be phosphorylated by NDPK than all the other analogues.

Regarding the interaction with HIV-1 RT, triphosphorylated nucleosides **HI1–HI3** present inadequate binding for DNA insertion due to the presence of the fused cyclobutane ring.

In summary, this study mapped out all the activation steps of **HI1–HI3**, as well as their interaction with the target protein, showing the complexity of both processes. The general conclusion that can be drawn from the theoretical calculations is that the presence of the cyclobutane fused to the 2'- and 3'-positions of the ribose moiety is restrictive for the adequate binding of the analogues to the viral DNA growing chain, causing the lack of anti-HIV activity of these compounds.

CHAPTER V

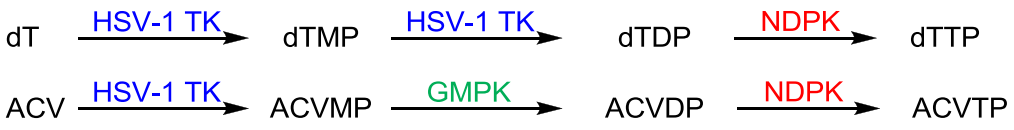
In silico study of cyclobutene and cyclobutane L-nucleoside analogues as anti-HSV agents

1. Introduction

HSV infections are still among the most frequent human diseases despite many nucleoside analogues are currently used for their treatment.[9] As for anti-HIV agents, the investigation of their mechanism of action may be useful for the development of new nucleoside analogues exhibiting anti-HSV activity. Thus, the present chapter is devoted to a molecular modelling study analogous to the previous one but focused on HSV-1 and nucleosides **HS1** and **HS2**.

1.1. Mechanism of action of nucleoside analogues in HSV-1 infected cells. Biological fundamentals

In contrast to HIV, a nucleoside kinase coded by HSV is involved in the phosphorylation process of nucleoside analogues, opening the way to the development of compounds that are selectively recognized as substrates by the viral enzyme but not by its human counterpart.[30,337] In particular, the kinases responsible for the activation of nucleosides in HSV-1 infected cells are the following: HSV-1 thymidine kinase (HSV-1 TK) for the first phosphorylation step; HSV-1 TK and human guanylate kinase (GMPK) for the second phosphorylation of pyrimidine and purine derivatives, respectively, as well as human nucleoside diphosphate kinase (NDPK) for the third phosphorylation (Scheme 44).[75] It is worth recalling that this last activation step is carried out by the same kinase (NDPK) for both anti-HSV and anti-HIV compounds.



Scheme 44. Kinases involved in the activation of pyrimidine (represented by dT) and purine analogues (represented by ACV) in HSV-1 infected cells.

HSV-1 TK (EC number: 2.7.1.21) is a key enzyme in the HSV metabolism. It is well known that HSV-1 TK accepts a broad range of substrates, in contrast to various other cellular kinases.[30,337–339] This is the basic principle for the selectivity of the established antiviral therapy with ACV and its derivatives, since HSV-1 TK accepts a broader range of substrates than its human counterpart, TK1. HSV-1 TK is able to phosphorylate both pyrimidine[340–342] and purine[30,337,343] analogues, to accept a large diversity of sugar moieties[30,337,342–345] as well as to catalyse the

introduction of the first and second phosphate to pyrimidine analogues.[75,340,346] The preferred phosphate donor of this enzyme is ATP, but it also shows high affinity for cytidine triphosphate, uridine triphosphate, and guanosine triphosphate and their deoxy analogues.[347]

GMPK (EC number: 2.7.4.8) is a monophosphate kinase that catalyses the phosphoryl transfer from ATP to GMP yielding ADP and GDP. Furthermore, it is also responsible for the second phosphorylation step of antiviral compounds such as ACV and GCV, among others.[348] This enzyme, as the TMK described in the previous chapter, belongs to the NMP kinase family.

The third activation step is carried out by NDPK (EC number: 2.7.4.6), the same enzyme that transfers the last phosphoryl on the activation of anti-HIV compounds. The main features of this enzyme have been presented in the previous chapter;[a] thus, they will not be described herein again.

Finally, triphosphorylated nucleosides have to interact with HSV-1 DNA polymerase to be clinically useful, since this enzyme is responsible for the DNA replication of HSV-1.[349] The nucleosides are then incorporated into the growing DNA strand via the attack of the 3'-OH of the primer terminus to their α-phosphorus, stopping thus the viral DNA replication.

1.2. Docking studies on nucleoside analogues as anti-HSV agents

Protein–ligand docking calculations have been more used to develop new nucleoside analogues for the treatment of HSV than for HIV. In the present section, some of the recent literature on this field is reviewed.

In 2002, Scapozza and co-workers reported a docking study of the interactions between 9-(2-hydroxypropyl)purine nucleoside analogues **137** (Figure 55) and HSV-1 TK.[350] Their results were then correlated with the *in vitro* phosphorylation and binding affinity assays data. The number of docked orientations favourable to phosphorylation from the total number of ten runs was used as the criterion to define a compound as a substrate. According to the authors, the set of criteria developed for the evaluation of docking results allows a qualitative prediction of a given ligand to be a good substrate or a potential inhibitor.

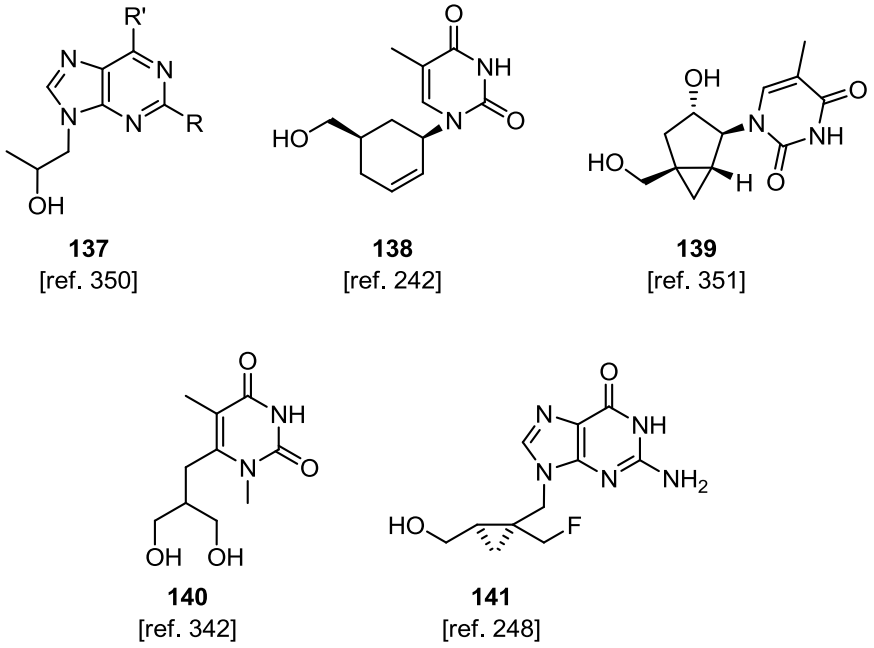
[a] See section *1.1. Mechanism of action of nucleoside analogues in HIV infected cells. Biological fundamentals* of Chapter IV.

Camplo and co-workers published in 2005 the synthesis and the antiviral evaluation of *cis*-substituted cyclohexene and cyclohexane nucleosides, including docking calculations on HSV-1 TK.[242] Some of the prepared nucleosides displayed moderate activity against HSV-1. With the aim of explaining the relatively weak antiviral activity of one of these nucleosides (compound **138**, Figure 55), the authors carried out a molecular modelling study on this derivative in the HSV-1 TK binding site. Considering that intracellular phosphorylation in the virus infected cells is required for the activation of these nucleosides, they inferred from the predicted orientations that the moderate activity of **138** could be explained by the loss of some important interactions between the ligand and certain amino acids in the binding site of HSV-1 TK.

The racemic synthesis of the new bicyclo[3.1.0]hexane thymidine analogue **139** that was successfully phosphorylated by HSV-1 TK was reported by Marquez and co-workers in 2007 (Figure 55).[351] This compound was envisaged after the encouraging results of an earlier investigation in which the authors prepared bicyclo[3.1.0]hexane nucleosides that proved to be substrates of HSV-1 TK. The crystallographic structures of these compounds in HSV-1 TK binding site along with the docking studies on compound **139** prior to its synthesis were of enormous importance for the rational design of this analogue. Moreover, docking calculations rationalized the improvement of the phosphorylation of **139** with respect to earlier series, in terms of hydrogen bond interactions and conformation of the nucleoside in HSV-1 TK binding site. Remarkably, only one enantiomer was predicted to be recognized by HSV-1 TK as a substrate according to docking experiments. Thus, this work was later extended by a publication, in 2008, of the stereoselective synthesis and biological evaluation of both enantiomers.[344] The D-enantiomer was identified as the biologically active enantiomer, as it had been predicted by molecular modelling.

The application of protein–ligand dockings to determine whether the acyclic pyrimidine nucleoside **140** (Figure 55) would be a substrate of HSV-1 TK before its synthesis was described by Scapozza and co-workers in 2011.[342] Starting from a lead compound, the authors envisaged a series of modifications that could enhance the promising properties of the lead, resulting in the proposal of the new target nucleoside **140**. Prior to performing chemical synthesis, the authors tested their hypotheses by carrying out a docking study on nucleoside **140**. The encouraging

results of the docking calculations prompted them to prepare the nucleoside analogue. Finally, the binding mode predicted by molecular modelling was confirmed by the X-ray structure of HSV-1 TK in complex with **140**. A similar approach was followed in 2012 by Sharma and co-workers,[248] resulting in the synthesis of the new purine nucleoside analogue **141** (Figure 55) that featured high activity against HeLa cells transfected with a mutant HSV-1 TK. Importantly, this property is nowadays largely exploited in some cancer treatments.[352]



137
[ref. 350]

138
[ref. 242]

139
[ref. 351]

140
[ref. 342]

141
[ref. 248]

Figure 55. Structures of the nucleoside analogues investigated by means of protein–ligand dockings as anti-HSV agents.

1.3. Scope

To the best of our knowledge, the investigation of the whole activation process of nucleosides in HSV-1 infected cells has never been reported. This fact together with the lack of activity of the cyclobutene and cyclobutane L-nucleoside analogues **HS1** and **HS2** against HSV (Table ES1) encouraged us to carry out our own molecular modelling study. In the present study calculations were carried out on **HS1** and **HS2** nucleoside analogues (Figure 56).

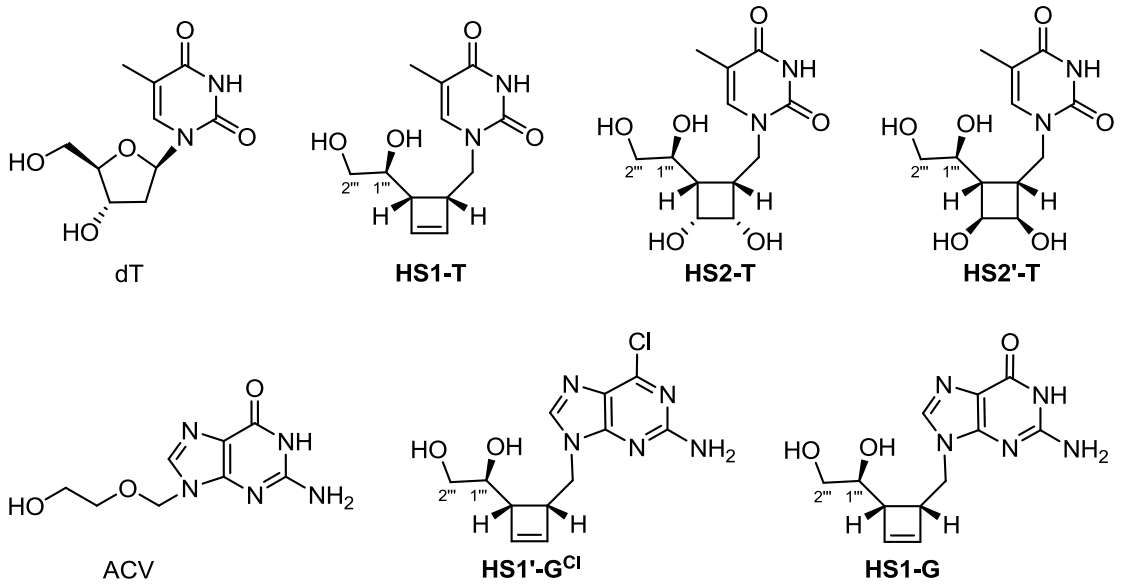
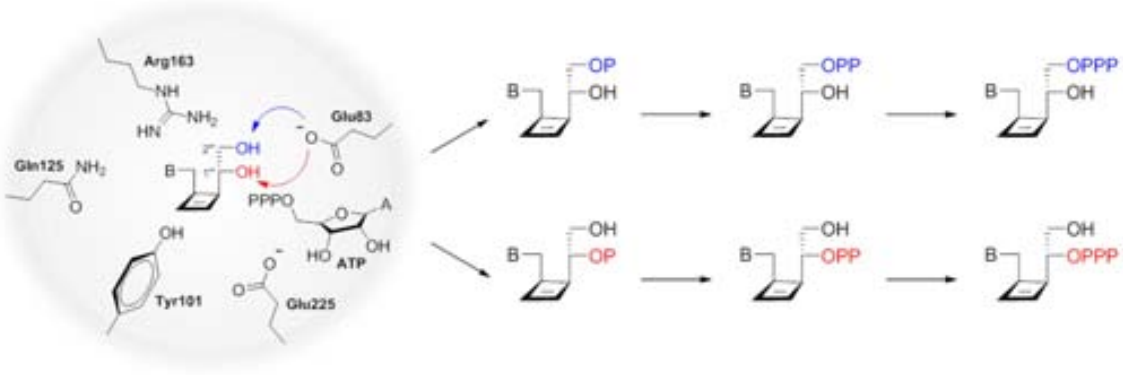


Figure 56. Nucleoside analogues **HS1** and **HS2** and reference compounds for their study as anti-HSV agents (dT and ACV). Note that the active form of compound **HS1-G^{Cl}** was assumed to be the deprotected one, **HS1'-G^{Cl}**.

As before, three are the main physiological requirements for anti-HSV nucleoside analogues to be active: their drug-likeness, their activation process and their subsequent interaction with HSV DNA polymerase. The purpose of our study was to identify if the substitution of the sugar by a cyclobutene or cyclobutane ring and the inverted configuration of our nucleosides hampered the fulfilment of one of these requirements. Disappointingly, it was not possible to study the interaction of the nucleosides with HSV-1 DNA polymerase due to the lack of structures of this enzyme crystallized with DNA. However, discerning if the failure lies in their activation or in another process is still important for further designs.

It is worth recalling that all the synthesized nucleosides contain an additional hydroxymethyl moiety compared to regular nucleosides. As a result, two different activation pathways were foreseen and studied: either successive phosphorylation at 1'''-OH or at 2'''-OH (Scheme 45).



Scheme 45. Representation of the two possible phosphorylation pathways of nucleoside analogues **HS1** and **HS2**. Upon abstraction of the corresponding proton by the residue Glu83 of HSV-1 TK, nucleosides may be phosphorylated either at their 1''' (red) or 2''' position (blue).

2. Computational details

In the present study, we adopted the same methodology than in Chapter IV. All binding modes and energies were predicted using protein–ligand dockings. Those calculations were performed with the docking program GOLD (version 5.0.1),[322] and the *ChemScore* scoring function was used.[278,323] Molecular graphics and visualization of docking results were performed with the *UCSF Chimera* package.[324]

The structures of the ligands were initially optimized using the *Marvin* work package[325] and the Merck molecular force field (MMFF) minimization.[326] Before carrying out the docking calculations, the available PDB structures of the enzymes were modified with *UCSF Chimera*. All crystallographic waters, ions and ligands were removed from each enzyme, and only one rotamer of the residues was kept for the calculations. When the structures presented more than one identical subunit, duplicate parts were deleted. Afterwards, hydrogen atoms were added and the protonation state of the histidine residues was checked. Atom charges were assigned from Amber.[327] Finally, the modified structures were saved in mol2 format.

The centre of the binding site in the X-ray structures was used as the central point of the docking cavity, the radius of which varies for each enzyme (Table ES19). A series of residues in the binding pocket were set as flexible by using the rotameric rotation scheme implemented in GOLD program (Table ES19).[322] Ligand flexibility was also considered for all the compounds.

Finally, each nucleoside was docked separately in its reactant and product form into each enzyme, and 20 predicted poses were obtained. The results were analysed both in structural and energetic terms.

3. Study of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues as anti-HSV agents

This section will be divided in two parts: the first one will focus on the evaluation of the analogues drug-likeness, and the second one will be devoted to provide molecular insights into their activation process.

3.1. Drug-likeness

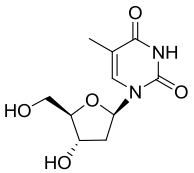
The drug-likeness evaluation of **HS1** and **HS2** nucleoside analogues was carried out according to the "Lipinski's rule of 5" and the Veber's criterion. The corresponding standard descriptors were calculated with the *Marvin* package.[328]

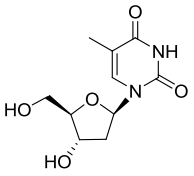
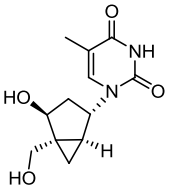
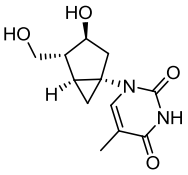
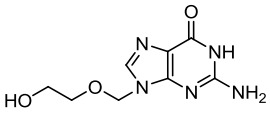
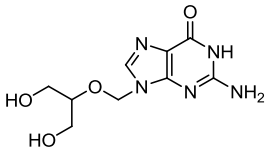
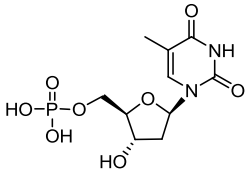
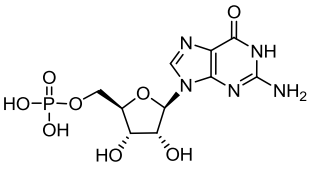
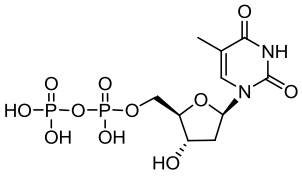
HS1 and **HS2** nucleosides positively fulfil both criteria, which supports their good drug-likeness (Tables ES20 and ES21). Therefore, these results point out that the lack of activity of our compounds should be related to more specific molecular interactions with protein receptors.

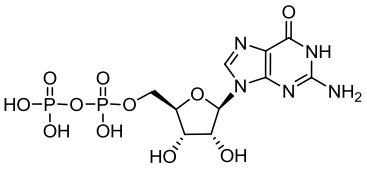
3.2. Activation

Protein–ligand dockings on HSV-1 TK, GMPK and NDPK were performed using crystallographic structures available with natural ligands, approved antiviral drugs or active carbocyclic nucleoside analogues (Table 11).

Table 11. PDB structures used to study the activation process of nucleosides **HS1** and **HS2**.

Enzyme	Organism	PDB code	Crystallized ligand	Docking calculations[a]
TK	HSV-1	1KIM	 dT	**HS1-T**, **HS2-T**, **HS2'-T**

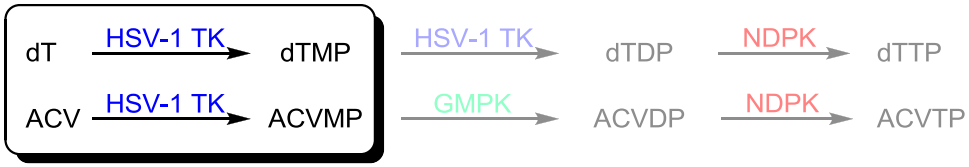
TK	HSV-1	2VTK	 dT (+ ADP)	**HS1-T**, **HS2-T**, **HS2'-T**
TK	HSV-1	1E2K	 MCT	**HS1-T**, **HS2-T**, **HS2'-T**
TK	HSV-1	1OF1	 (S)-MCT	**HS1-T**, **HS2-T**, **HS2'-T**
TK	HSV-1	2KI5	 ACV	**HS1'-G^{Cl}**, **HS1-G** and **HS1'-G^{Cl}MP**, **HS1-GMP**
TK	HSV-1	1KI3	 PCV	**HS1'-G^{Cl}**, **HS1-G** and **HS1'-G^{Cl}MP**, **HS1-GMP**
TK	HSV-1	1VTK	 dTMP (+ ADP)	**HS1-TMP**, **HS2-TMP**, **HS2'-TMP** and **HS1-TDP**, **HS2-TDP**, **HS2'-TDP**
GMPK	*Mus musculus*	1LVG	 GMP (+ ADP)	**HS1'-G^{Cl}MP**, **HS1-GMP** and **HS1'-G^{Cl}DP**, **HS1-GDP**
NDPK	*Dictyostelium discoideum*	1NDC	 dTDP	**HS1-TDP**, **HS2-TDP**, **HS2'-TDP** and **HS1-TTP**, **HS2-TTP**, **HS2'-TTP**

NDPK	*Homo sapiens*	1NUE	 GDP	**HS1'-G^{Cl}DP**, **HS1-GDP** and **HS1'-G^{Cl}TP**, **HS1-GTP**

[a] MP, DP and TP stand for monophosphorylated, diphosphorylated and triphosphorylated compounds, respectively.

HS1 and **HS2** nucleosides were docked into the active site of each kinase in their reactant and product forms. Calculations with dT and ACV were also performed to provide with structural and energetic benchmarks. The criterion to determine if a compound is expected to be phosphorylated considered the same two factors as in the HIV study: the number of binding modes favourable to phosphorylation as well as their binding energy values compared to those of reference.

3.2.1. 1st phosphorylation step



3.2.1.1. HSV-1 TK crystallographic structures

Up to date, there are more than 20 crystallographic structures of HSV-1 TK available: in its apo and holo forms as well as for wild type and mutagenic forms. We focused our attention on those containing drugs or candidates with interest for our study (Table 11). These structures represent a good approximation to take into account the required structural pre-organization of the enzyme for the catalysis.

HSV-1 TK is a homodimer with 376 residues per subunit.[338,353] Each monomer consists of 13 α-helices, two 3_{10}-helices and seven β-sheets (Figure 57). Five-stranded parallel β-sheets form part of the core of the protein, which contains the active site.[338,354] This active site is formed by an ATP and a nucleoside binding region,[355] and the first and second phosphorylation steps can occur successively with the phosphate acceptor remaining at its position.[353] In most of the kinases, the donor and the acceptor normally bind in a non-ordered manner, but HSV-1 TK binds dT before ATP.[347]

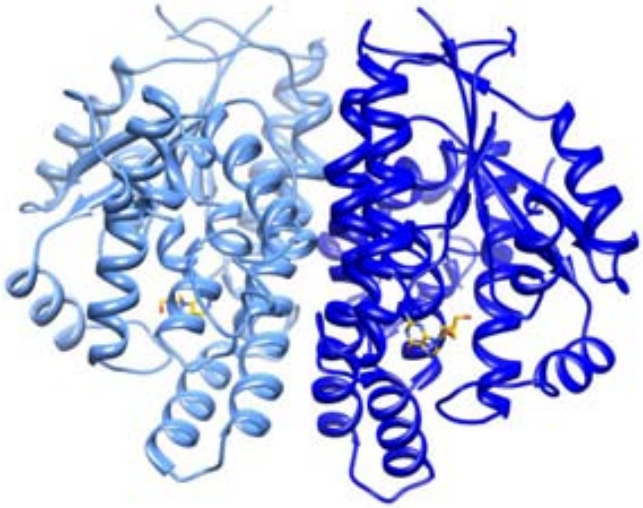


Figure 57. HSV-1 TK crystallized with PCV (orange) in its binding site (PDB code: 1KI3).

The binding mode of dT in HSV-1 TK binding site (PDB code: 1KIM)[338] is sketched in Figure 58. In this complex, the base is sandwiched between Met-128 and Tyr-172,[355] and it is further stabilized by pairwise hydrogen bond interactions with Gln-125 as well as by two water-mediated hydrogen bonds with the side chain of Arg-176. The 5-methyl group of the pyrimidine ring is placed in a hydrophobic environment formed by Tyr-132, Ala-167 and Ala-168. Additionally, the deoxyribose makes hydrogen bond interactions via its 3'-OH with Tyr-101 and Glu-225, whereas its 5'-OH is hydrogen bonded to Arg-163 and Glu-83. It is worth mentioning that Glu-83 activates the 5'-OH by deprotonating it to make it a better nucleophile.[339,353,356]

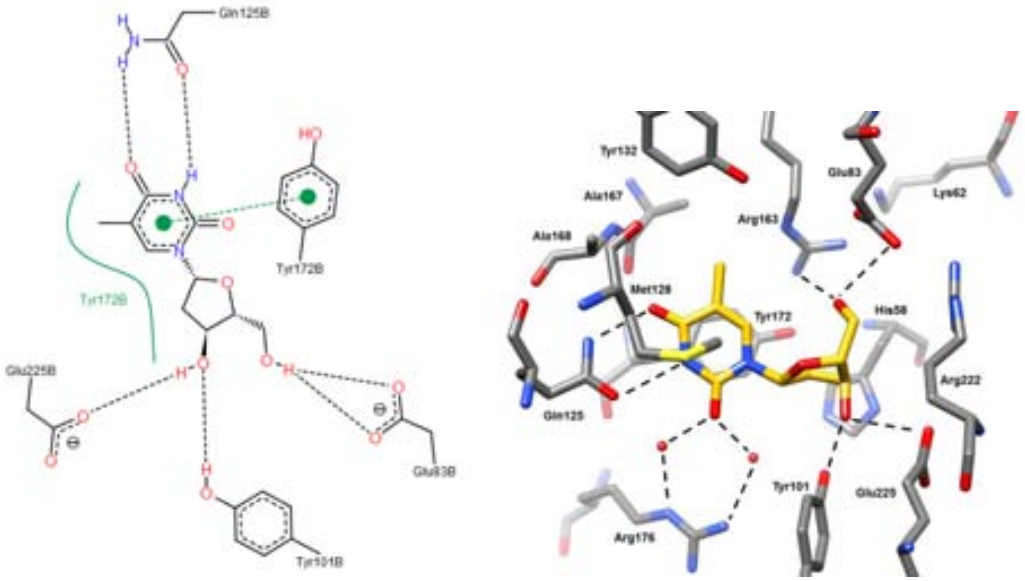


Figure 58. Representation of the main interactions of dT in HSV-1 TK binding site (PDB code: 1KIM). Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids

and the ligand (left image). Hydrogen bonds between dT and residues are depicted as dotted lines. Images generated using PoseView[357] software (left) and UCSF Chimera[324] package (right).

The binding mode of ACV (PDB code: 2KI5)[358] is quite similar to that of dT (Figure 59). Remarkably, the hydrogen bond with Gln-125 involves a 180º rotation of its amide as well as a conformational shift of the side chain. The acyclic moiety of ACV can assume two distinct orientations according to the X-ray structure, each of them matching to the previously described poses of the 3'-OH and 5'-OH of dT.[358] Only the conformation in which the OH is pointing towards Glu-83 is considered a phosphorylation orientation.

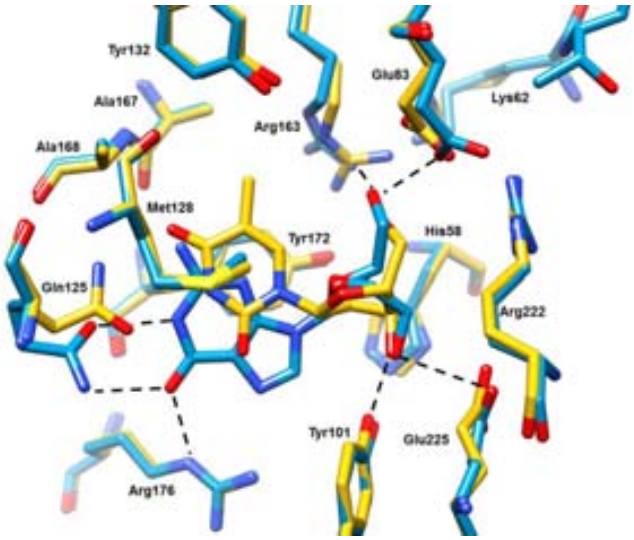


Figure 59. X-ray structure of dT (PDB code: 1KIM, in yellow) superimposed to ACV (PDB code: 2KI5, in blue) in HSV-1 TK binding site. For the sake of clarity, hydrogen atoms and crystallographic waters are not shown, and only hydrogen bonds between ACV and residues are depicted as dotted lines.

Regarding PCV (PDB code: 1KI3),[338] its guanine moiety is perfectly overlapped to the ACV one. Compared to the above described dT/HSV-1 TK complex (Figure B.11), one of the PCV hydroxyl groups parallels the binding mode of the 5'-OH of dT while the other forms an hydrogen bond interaction with Arg-222.

The binding mode of MCT (PDB code: 1E2K)[359] is very similar to the dT/HSV-1 TK complex (Figure B.12), but some differences have to be pointed out. In particular, direct hydrogen bonds with Tyr-101, Arg-163 and Glu-225 are lost. This structural difference is in agreement with the lower binding affinity of MCT to HSV-1 TK reported by Scapozza and co-workers,[359] but does not cause the absence of the phosphorylation reaction. On the other hand, compound (S)-MCT (PDB code: 1OF1)[345] superimposes very well with the active conformation of dT in HSV-1 TK binding site (Figure B.13). This behaviour is consistent with the experimental

observation that (S)-MCT analogue has a more favourable binding affinity and catalytic turnover for HSV-1 TK than MCT.[345]

All the HSV-1 TK structures described do not contain ATP in the binding site. Instead, a sulfate anion is observed in the putative position of the β-phosphate of ATP (Figure 60).[360] This sulfate binding site is located close to the 5′-OH function of dT (S-O5′ distance is 6.0 Å).

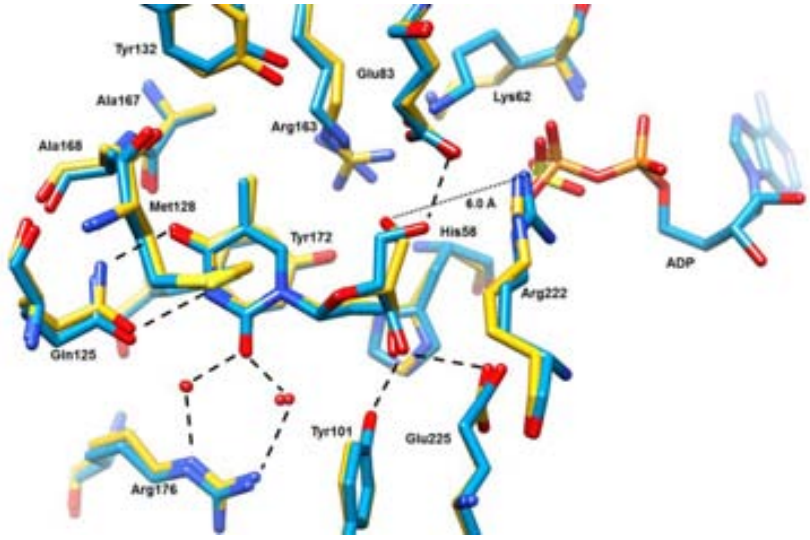


Figure 60. Superimposed X-ray structures of dT in HSV-1 TK binding site crystallized with ADP (PDB code: 2VTK, in blue) and without ADP (PDB code: 1KIM, in yellow). For the sake of clarity, hydrogen atoms are not shown, and only hydrogen bonds in 2VTK structure are depicted as dotted lines.

HSV-1 TK structures featuring ADP and dT (PDB code: 2VTK)[353] or dTMP (PDB code: 1VTK)[353] in the binding site provide a good description of the states before and after phosphorylation. Regarding dT, the main difference between the structures with and without ADP is a displacement of the 5'-OH towards the β-phosphate of ADP (Figure 60). A superposition of the structures representing the two states of the phosphorylation reaction (dT and dTMP) shows only small differences (Figure B.14). The largest displacements occur to Glu-83 and Arg-222, since the α-phosphate of dTMP pushes them outward, and to Lys-62, which is displaced towards the β-phosphate of ADP. Unfortunately, it is noteworthy that there is no crystallographic structure with a monophosphorylated purine analogue available.

Finally, it is postulated that more than 80% of the total interaction energy between the substrate and HSV-1 TK stems from the interaction with only six residues, namely Glu-83, Tyr-101, Gln-125, Met-128, Tyr-172 and Glu-

225.[339,354,361] Therefore, to consider a predicted orientation compatible with the catalysis it will be required to present most of these interactions with the substrate.

3.2.1.2. Docking results

The docking protocol was validated by carrying out docking calculations of the crystallized ligands into the corresponding HSV-1 TK X-ray structures (Table 11). The lowest energy poses showed a very good alignment with the crystallographic structures, validating thus the good performance of GOLD in predicting the binding modes of the ligands. Consequently, dT and ACV were suitable to be used as benchmarks for pyrimidine and purine analogues, respectively. The corresponding binding energies are -22 and -19 score units for dT and dTMP (Tables ES22 and ES23), and –11 and -7 score units for ACV and ACVMP (Tables ES24 and ES25). The higher binding energy values of ACV compared to dT are consistent with the lower affinity of ACV to HSV-1 TK.[338,343,358]

The study of the unphosphorylated and monophosphorylated pyrimidine analogues was not carried out on the same X-ray structure (Table 11). 1VTK was used to investigate monophosphorylated pyrimidine derivatives, since it represents the best description of HSV-1 TK state after the catalytic reaction of pyrimidine analogues.[353]

Docking calculations of the unphosphorylated and monophosphorylated **HS1** and **HS2** derivatives led to three different behaviours: for **HS1-T** and **HS2-T**, phosphorylation is more likely to occur at their 2'''-OH than at their 1'''-OH; for **HS1-G**, it could take place at 1'''-OH or 2'''-OH; and finally, for **HS2'-T** and **HS1'-G^{Cl}** the phosphorylation is unlikely.

Most of the predicted binding modes of **HS1-T** are compatible with the catalytic activity of the enzyme (Figure 61). The lack of interactions with Glu-225 and Tyr-101 as a consequence of the absence of a 3'-OH in **HS1-T** is not considered to be restrictive for the catalysis, since the same behaviour is observed for ACV.[358] Similar and even lower binding energies are calculated for **HS1-T** with respect to dT, being between -30 and -23 score units (Table ES22).

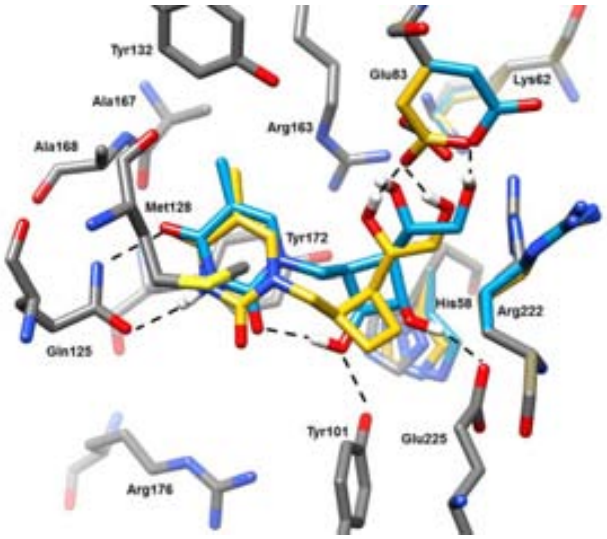


Figure 61. HS1-T (yellow) superimposed to **HS2-T** (blue) in HSV-1 TK binding site (PDB code: 1KIM, X-ray residues shown in grey). Hydrogen bonds are depicted as dotted lines. For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are only shown when bound to a heteroatom of the ligand.

Docking calculations of **HS1-TMP** led to a higher number of binding modes properly oriented for the catalysis for the 2'''-monophosphorylated (2'''-MP) derivative than for the 1'''-MP one. The predicted binding energies are in agreement with this behaviour, since the energies of the 2'''-MP derivatives are slightly lower than those of 1'''-MP (Table ES23). These results are consistent with the presence of an additional hydrogen bond between the 1'''-OH and Glu-225 for the 2'''-MP derivative (Figure B.15).

Similarly to **HS1-T**, the predicted complexes of compound **HS2-T** are in agreement with the catalytic activity of the enzyme (Figure 61), and the corresponding binding energies are similar to those of dT, being between -25 and -19 score units (Table ES22). A structural analysis of the results shows a slight displacement of the base due to the presence of an internal hydrogen bond with the 2''-OH of the sugar-like moiety. Moreover, the additional alcohol moieties at positions 2'' and 3'' are hydrogen bonded to Tyr-101 and Glu-225, respectively. Therefore, these additional interactions are expected to compensate the displacement of the nucleobase.

Calculations on **HS2-TMP** showed that phosphorylation at 2'''-OH is energetically more favoured than at 1'''-OH (Table ES23). This trend is related to the loss of one hydrogen bond with Gln-125 of 1'''-MP derivatives (Figure B.16).

Furthermore, 2'''-MP derivatives have to be less tensioned than 1'''-MP to fit in the active site.

Different poses of **HS2'-T** depending on the crystallographic structure used for the docking calculations were predicted. In 1E2K and 1OF1 structures, a flip of the cyclobutane moiety is observed in most of the binding modes, being thus not consistent with the catalysis (Figure B.17). When using 1KIM structure, two equally favoured orientations are predicted: one of them parallels the previous pose and the other is properly oriented for the phosphoryl transfer. Finally, in most of the poses predicted with 2VTK structure the base is completely displaced with respect to the crystallographic one. Therefore, a unique binding mode of **HS2'-T** in HSV-1 TK cannot be certainly determined from our docking calculations, but the high number of orientations incompatible with the catalysis indicated that the phosphoryl transfer is unlikely. In agreement with these results, few binding modes of **HS2'-TMP** derivatives are consistent with the catalytic reactivity of the enzyme. A flip of the base is observed in most of its low energy orientations (Figure 62). No clear molecular features have been identified to cause this flip, but the elevated number of binding modes featuring it suggests that the orientation with the base outside its pocket is the most plausible.

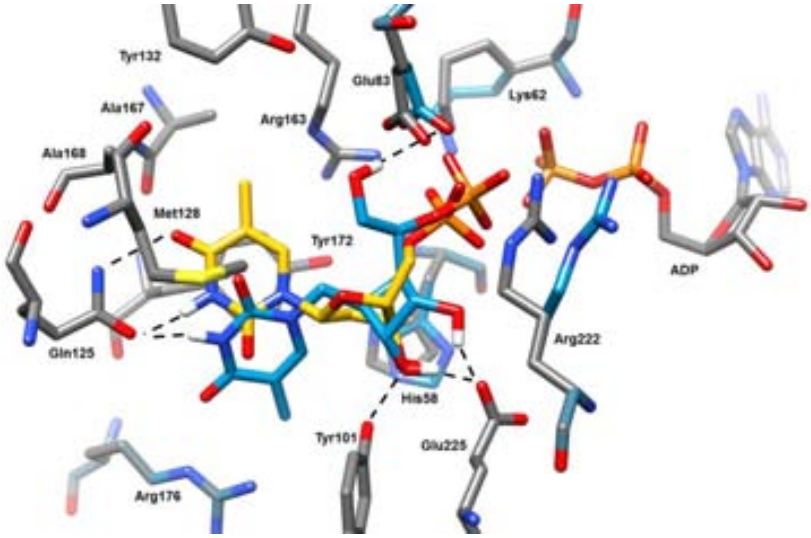


Figure 62. HS2'-T-1'''-MP (blue) superimposed to crystallographic dTMP (yellow) in HSV-1 TK binding site (PDB code: 1VTK, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between both ligands and residues are depicted as dotted lines.

Concerning purine analogues, very few binding modes of compound **HS1'-G^{Cl}** in HSV-1 TK binding site are compatible with the catalysis. Indeed, when **HS1'-G^{Cl}** is

properly oriented there are steric clashes with Arg-176, Gln-125 and His-58 (Figure 63, left). As a result, this compound flips around to minimize these clashes, leaving the base, 1'''-OH and 2'''-OH outside their pockets (Figure 63, right). Therefore, the presence of a modified purine base is restrictive for the first activation step of this cyclobutene analogue.

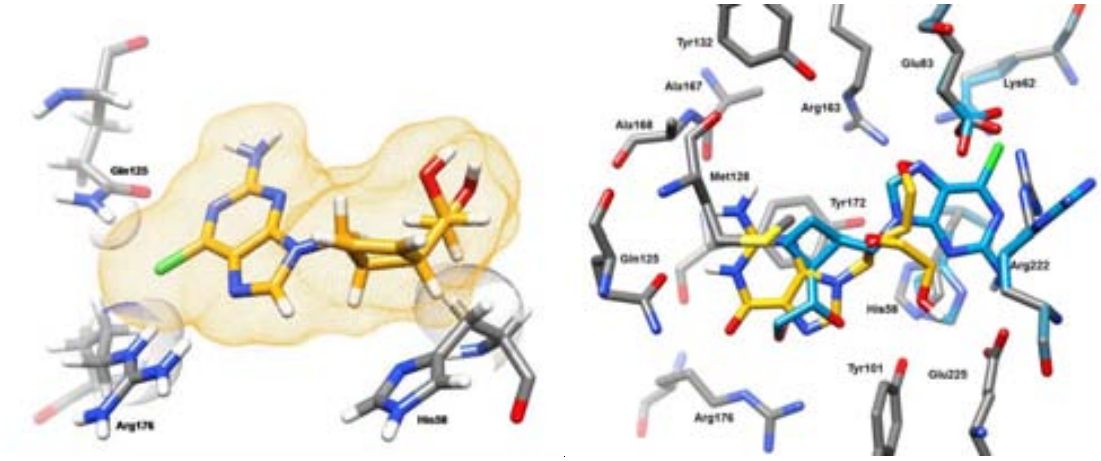


Figure 63. Van der Waals surface representation of **HS1'-G^{Cl}** (orange) and some residues (grey) in HSV-1 TK binding site (PDB code: 2KI5) (left), and **HS1'-G^{Cl}** (blue) superimposed to crystallographic ACV (yellow) in HSV-1 TK binding site (PDB code: 2KI5, X-ray residues shown in grey) (right). Crystallographic waters and hydrogen bonds are not shown for clarity.

Accordingly, calculations on **HS1'-G^{Cl}MP** led to scarce binding modes consistent with the catalytic phosphoryl transfer. Two equally favoured orientations are observed on 2KI5 structure: one of them shows the nucleobase outside its pocket, while in the other the compound is completely flipped around (Figure B.18). When carrying out the calculations with 1KI3 X-ray structure, most of the binding modes also present the base not properly posed. This behaviour is consistent with the steric clashes observed between the unphosphorylated derivative and Arg-176, Gln-125 and His-58.

Conversely, almost all the complexes of **HS1-G** are properly posed, showing the same interactions than ACV, and the corresponding binding energies are lower than those of ACV (Table ES24).

Most of the orientations of **HS1-GMP** are also compatible with the catalysis (Figure 64), and their binding energies, between -17 and -11 score units, are lower than those of ACVMP (Table ES25). Therefore, these results confirmed that **HS1-G** is expected to be converted into its monophosphorylated derivative, and that this phosphoryl transfer is feasible both at its 1''' and 2''' positions.

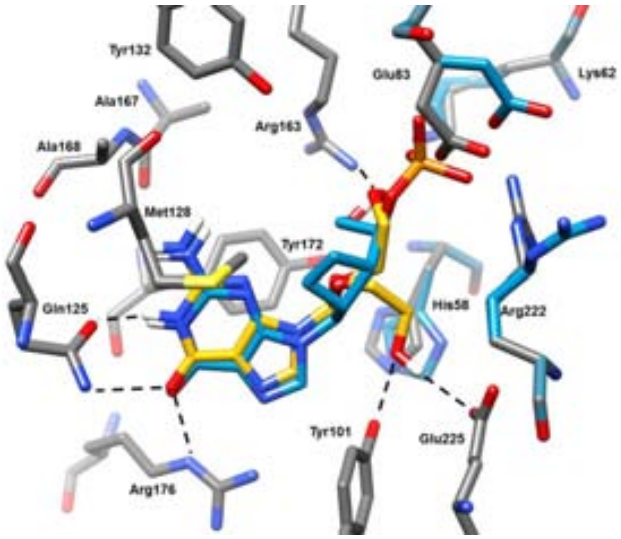
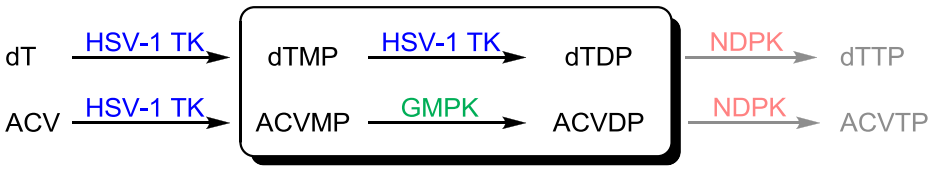


Figure 64. HS1-G-1'''-MP (blue) superimposed to crystallographic ACV (yellow) in HSV-1 TK binding site (PDB code: 2KI5, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between both ligands and residues are depicted as dotted lines.

In summary, compounds **HS1-T** and **HS2-T** are more likely to be phosphorylated at their 2''' position, whereas **HS1-G** could be phosphorylated both at its 1'''-OH or 2'''-OH. **HS2'-T** and **HS1'-G^{Cl}** are not expected to be phosphorylated by HSV-1 TK.

3.2.2. 2nd phosphorylation step



3.2.2.1. GMPK crystallographic structure

The lack of human GMPK crystallographic structures prompted us to use a mouse GMPK (mGMPK) structure to carry out the study of the second phosphorylation of purine analogues. As stated by Sekulic et al.,[348] the very high sequence similarity between human and mouse structures assures that the information of the mouse enzyme is directly transferable to the human enzyme. In particular, the structure selected contains GMP and ADP in the binding site (PDB code: 1LVG).[348]

mGMPK consists of 198 residues, and is built from eight α-helices and two β-sheets that form three structurally distinct regions, namely the core, LID and NMP-

binding region (Figure 65). This kinase is known to move these parts upon substrate binding, and so the structure presenting both the substrate and ADP in its active site is the best approximation to consider the conformational changes that occur after the binding.

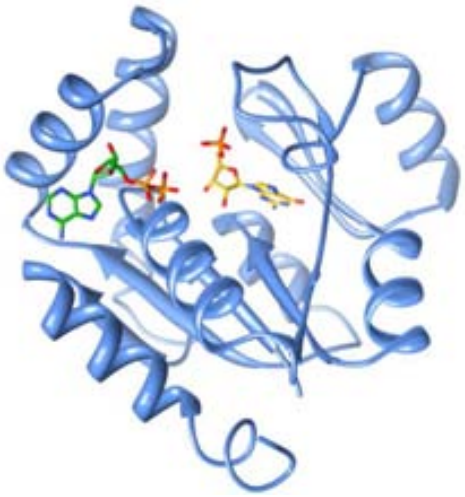


Figure 65. mGMPK crystallized with GMP (yellow) and ADP (green) in its binding site (PDB code: 1LVG).

In the complex of mGMPK with GMP and ADP, the base of GMP is hydrogen bonded to Thr-83, Ser-37, Glu-72 and Asp-103 (Figure 66, Figure B.19). The 2'-OH of the ribose moiety is stabilized by an interaction with Asp-101. Additionally, Tyr-53, Tyr-81, Arg-41, Arg-44 and Arg-148 interact with the α-phosphate of the substrate. It is worth mentioning that Arg-44 and Arg-148 are postulated to interact also with the γ-phosphate of ATP, suggesting a catalytic role of these residues in the phosphoryl transfer.[348]

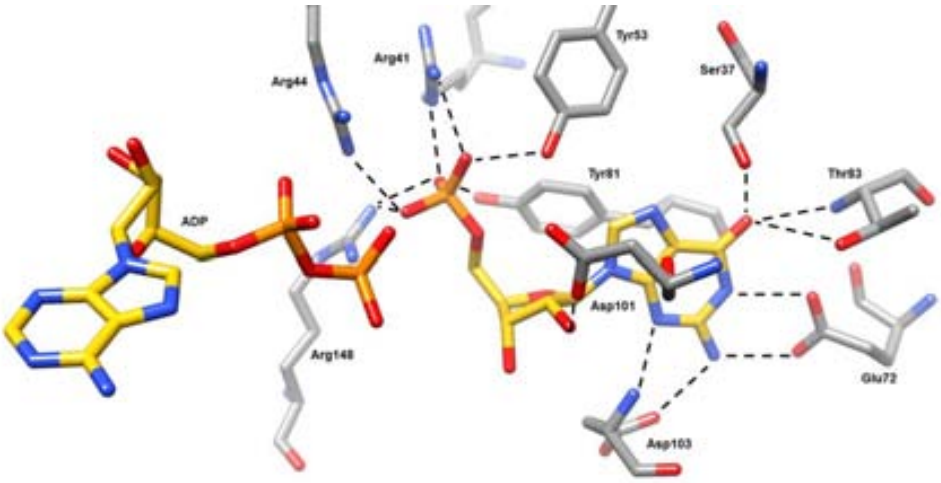


Figure 66. Representation of the main interactions of GMP in mGMPK binding site (PDB code: 1LVG). Hydrogen bonds between GMP and residues are depicted as dotted lines. Crystallographic waters and hydrogen atoms are not shown for clarity.

Regarding ADP, its base is interacting via hydrogen bonds with Asp-172 and Asn-171, as well as via a stacking interaction with Arg-133. The α- and β-phosphates of ADP are hydrogen bonded to Gly-16, Lys-17, Ser-18, Thr-19 and Arg-137 (Figure B.20).

3.2.2.2. Docking results

The second phosphorylation is carried out by HSV-1 TK for pyrimidine derivatives and GMPK for purine derivatives. Since the results of docking calculations on monophosphorylated pyrimidine derivatives with HSV-1 TK have just been presented in section 3.2.1, they will not be repeated herein. Thus, only the results on their diphosphorylated derivatives will be shown. Concerning purine nucleosides, the results of both monophosphorylated and diphosphorylated derivatives will be discussed. For this reason, the analysis of pyrimidine and purine nucleosides will be presented separately.

The study of **HS1-TDP**, **HS2-TDP** and **HS2'-TDP** was carried out with the crystallographic structure of HSV-1 TK previously used for the calculations of the monophosphorylated derivatives (PDB code: 1VTK).[353] It is noteworthy that HSV-1 TK appears to be less restrictive about the presence of a 3'-OH when catalysing the first phosphoryl transfer, considering that some compounds such as ACV (lacking the 3'-OH entirely) are efficiently phosphorylated. However, the presence and correct disposition of this alcohol becomes critical for the second phosphoryl transfer.[362] This behaviour will be taken into account in the analysis of our

calculations, considering that a compound lacking an alcohol that resembles the 3'-OH (i.e. interacting with Tyr-101 or Glu-225) will be less likely to be phosphorylated. The rest of the interactions required to consider a predicted orientation consistent with the catalysis are the same as for the monophosphorylated substrates.

Docking calculations of **HS1-TDP**, **HS2-TDP** and **HS2'-TDP** revealed that the second phosphorylation of **HS1-T** is again more favoured at its 2''' position, whereas **HS2-TMP** and **HS2'-TMP** are not expected to be converted into their diphosphorylated derivatives.

Most of the complexes of **HS1-T-1'''-DP** present the base completely outside its pocket though the rest of the interactions required are observed (Figure 67). In this orientation, the thymine ring is hydrogen bonded to Arg-176 and Tyr-101, allowing the compound to be less tensioned than when presenting the base properly oriented. Indeed, the binding energy of the only orientation compatible with the catalysis is -7 score units, which is higher than the reference one, -16.5 score units (Table ES26). By contrast, **HS1-T-2'''-DP** calculations led to several binding modes consistent with the phosphoryl transfer, featuring the base and the phosphates properly oriented and even the 1'''-OH interacting with Glu-225 (thus resembling the 3'-OH of dT). Moreover, the binding energy of **HS1-T-2'''-DP** is around -18 score units, which is lower than that of reference (Table ES26). This behaviour is in agreement with the previous results on the monophosphorylated derivative, in which phosphorylation was more likely at the 2'''-OH.

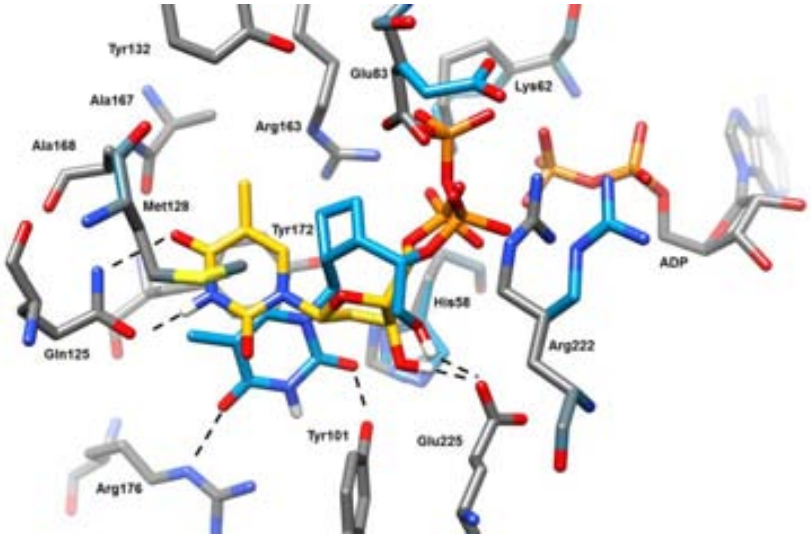


Figure 67. HS1-T-1'''-DP (blue) superimposed to crystallographic dTMP (yellow) in HSV-1 TK binding site (PDB code: 1VTK, X-ray residues shown in grey). For the sake of clarity, crystallographic waters

are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between both ligands and residues are depicted as dotted lines.

Concerning **HS2-TDP**, the binding energies calculated for all the complexes are far higher than that of reference, despite some **HS2-T-2'''-DP** poses are properly oriented for the catalysis. These high values are between -7 and +12 score units, and are mainly caused by the high internal tension of the compound to fit in the active site. Therefore, **HS2-TMP** is unlikely to be further phosphorylated.

Dockings of **HS2'-TDP** showed that the introduction of a second phosphate is neither structurally nor energetically favoured. In agreement with the results of **HS2'-TMP**, calculations of both **HS2'-T-1'''-DP** and **HS2'-T-2'''-DP** led to a reduced number of predicted orientations compatible with the catalysis which, moreover, present higher energies than the reference one (between -5 and -2 score units, Table ES26). In fact, most of the binding modes show the nucleobase outside its pocket, despite being interacting via one hydrogen bond with Gln-125 (Figure 68). As a consequence, phosphorylation of compound **HS2'-T** will fail both in the first and second activation steps.

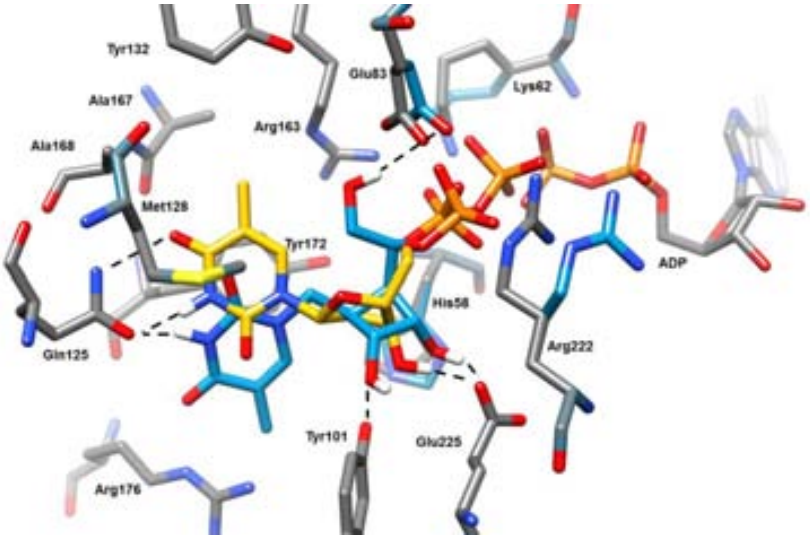


Figure 68. HS2'-T-1'''-DP (blue) superimposed to crystallographic dTMP (yellow) in HSV-1 TK binding site (PDB code: 1VTK, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between both ligands and residues are depicted as dotted lines.

Regarding purine compounds **HS1'-G^{Cl}** and **HS1-G**, the study was carried out with the X-ray structure of mGMPK previously described (PDB code: 1LVG). The docking protocol was validated by performing docking calculations of GMP with

mGMPK, which confirmed that the lowest energy orientations closely match the crystallographic structure.

Additionally, dockings of ACVMP and ACVDP on 1LVG structure were also carried out and similar poses to GMP were obtained (Figure B.21), successfully validating the use of ACV as a benchmark. The corresponding binding energy is -12 score units for both ACVMP and ACVDP (Tables ES27 and ES28).

Docking calculations on **HS1'-G^{Cl}** and **HS1-G** revealed two opposite profiles: while **HS1'-G^{Cl}MP** is not likely to be further phosphorylated, a second phosphoryl is expected to be added to **HS1-G-2'''-MP**.

None of the predicted poses of **HS1'-G^{Cl}-1'''-MP** and **HS1'-G^{Cl}-2'''-MP** is consistent with the catalysis. Indeed, a rotation of the purine base is observed in most of the complexes, minimizing thus the clash of the chlorine atom with the residue Ser-37 when the base is properly posed. These orientations are further stabilized by direct hydrogen bonds between the 2'''-OH and Tyr-53 for **HS1'-G^{Cl}-1'''-MP** (Figure 69), as well as between the 1'''-OH and Ser-13 for **HS1'-G^{Cl}-2'''-MP**.

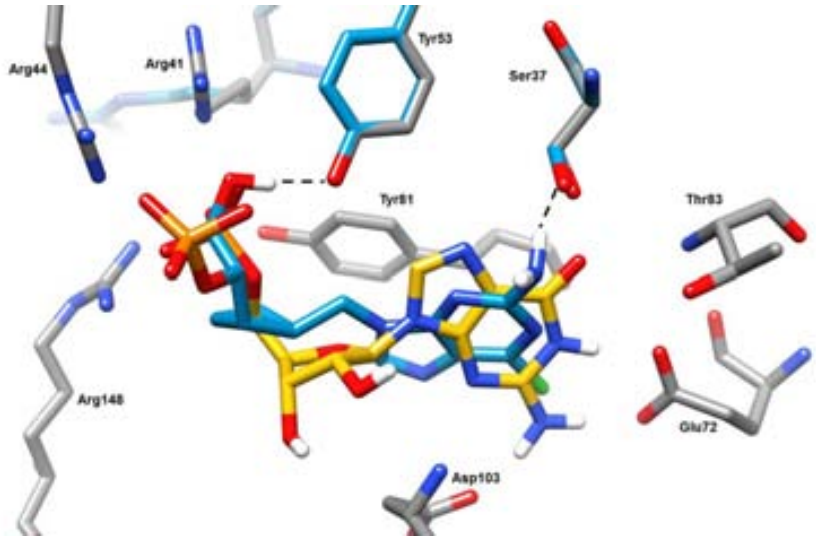


Figure 69. HS1'-G^{Cl}-1'''-MP (blue) superimposed to crystallographic GMP (yellow) in mGMPK binding site (PDB code: 1LVG, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown, hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand and only hydrogen bonds between **HS1'-G^{Cl}-1'''-MP** and residues are depicted as dotted lines.

In addition, few binding modes of **HS1'-G^{Cl}-1'''-DP** are properly posed for the catalytic activity of the enzyme, since the base is also rotated as for **HS1'-G^{Cl}MP** (Figure B.22). A large variability is observed on the predicted orientations of **HS1'-G^{Cl}-2'''-DP**, all of them being not compatible with the phosphoryl transfer.

By contrast, most of the binding modes of **HS1-G-1'''-MP** and **HS1-G-2'''-MP** are compatible with the catalysis and their binding energies, around -20 score units, are lower than the reference one (Table ES27).

Docking calculations on **HS1-GDP** showed that the phosphorylation is more favoured at its 2''' position than at its 1''' position. A higher number of orientations are compatible with the catalysis for **HS1-G-2'''-DP** than for **HS1-G-1'''-DP**, and their binding energies are lower (Table ES28). Additionally, a binding mode not consistent with the phosphoryl transfer is predicted to compete with the properly posed orientations of **HS1-G-1'''-DP**. This binding mode resembles the previously described pose of **HS1'-G^{Cl}MP**, showing the base outside its pocket and the 2'''-OH interacting via a hydrogen bond with Tyr-53 and Asp-101 (Figure 70).

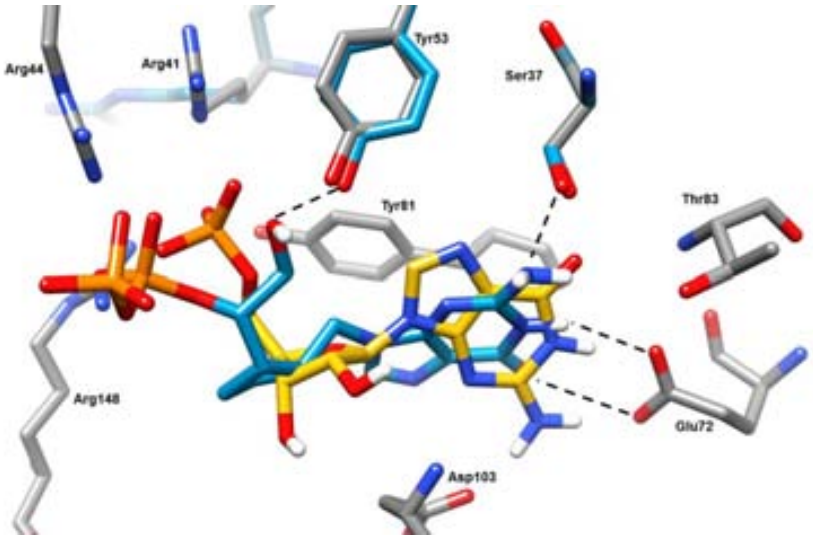
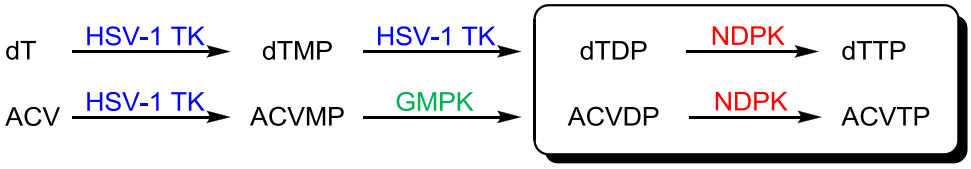


Figure 70. HS2-G-1'''-DP (blue) superimposed to crystallographic GMP (yellow) in mGMPK binding site (PDB code: 1LVG, X-ray residues shown in grey). For the sake of clarity, crystallographic waters and Asp-101 are not shown, hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand and only hydrogen bonds between **HS2-G-1'''-DP** and residues are depicted as dotted lines.

To sum up, compounds **HS2-TMP**, **HS2'-TMP** and **HS1'-G^{Cl}MP** are not likely to be further phosphorylated by the corresponding enzymes, whereas a second phosphoryl transfer is expected to occur at **HS1-T-2'''-MP** and **HS1-G-2'''-MP**.

3.2.3. 3rd phosphorylation step



3.2.3.1. NDPK crystallographic structures

Two different crystallographic structures were selected to perform the calculations of the last activation step. For pyrimidine derivatives, a complex of *Dictyostelium discoideum* NDPK with dTDP was used (PDB code: 1NDC),[304] whereas for purine derivatives dockings were carried out on a human NDPK crystallized with GDP (PDB code: 1NUE).[334]

The binding mode of GDP in human NDPK features its base filling a cleft between Phe-60 and Val-112 near the protein surface, and the β-phosphate pointing towards the catalytic His-118 (Figure 71, Figure B.23). The only polar interaction between the base and the protein links the nucleobase to Glu-152, which is the C-terminal residue of another subunit of the hexameric NDPK. The ribose moiety is totally buried in the protein, and the 2'-OH and 3'-OH are hydrogen bonded to Lys-12 and Asn-115, respectively. The 3'-OH is also interacting via an internal hydrogen bond with the β-phosphate, which is also hydrogen bonded to Thr-94, Arg-88 and His-118 via a water molecule. The α-phosphate remains accessible to the solvent and it is interacting with a magnesium ion.

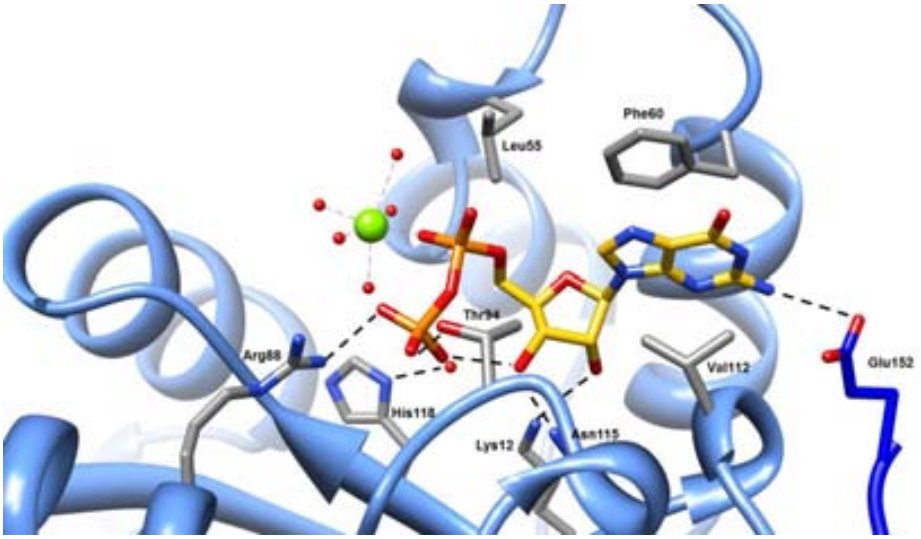


Figure 71. Representation of the main interactions of GDP in NDPK binding site (PDB code: 1NUE). Hydrogen bonds between GDP and residues are depicted as dotted lines. The crystallographic ion Mg^{2+} is shown, and Glu-152 of another subunit is shown in dark blue. Hydrogen atoms are not shown for clarity.

dTDP mode of binding in *Dictyostelium discoideum* NDPK is essentially the same as for GDP in human NDPK (Figure 72, Figure B.24). The base is also wedged between Phe-64 and Val-116, and it is interacting via two water molecules with the

carboxylate of Glu-155 from an adjacent subunit.[a] The 3'-OH is interacting with Lys-16, Asn-119 and the β-phosphate, which is also hydrogen bonded to Thr-98, Arg-92 and Arg-109 and water-mediated to the catalytic His-122. There is a magnesium ion that clearly bridges the α- and β-phosphates.

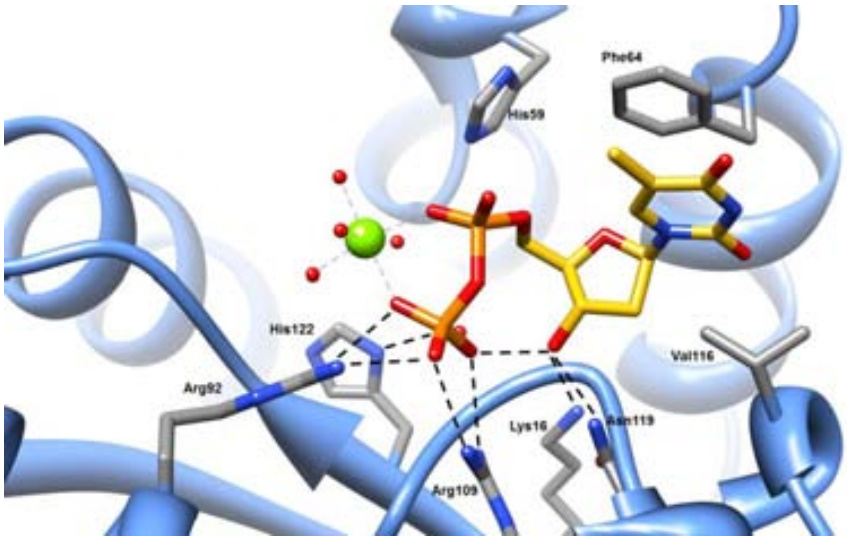


Figure 72. Representation of the main interactions of dTDP in NDPK binding site (PDB code: 1NDC). Hydrogen bonds between dTDP and residues are depicted as dotted lines. The crystallographic ion Mg^{2+} is shown, and the residue Thr-98 as well as hydrogen atoms are not shown for clarity.

3.2.3.2. Docking results

To validate the docking protocol for purine derivatives, GDP was docked into the corresponding crystallographic structure (PDB code: 1NUE). At first, only one subunit of the NDPK hexamer was used for the calculations, and residues Leu-55, Leu-64, Thr-94 and Arg-105 were set as flexible. None of the predicted binding modes matched the crystallographic structure. Hence, all the residues were kept fixed in the following docking calculations, but just a few binding modes reproduced the X-ray pose of GDP, presumably because of docking limitations in dealing with widely solvent exposed binding sites. Trying to overcome this limitation, another subunit of the hexamer that contained Glu-152 was added, and dockings were carried out again. Disappointingly, no improvement on the reproduction of the crystallographic structure was observed. The scoring function was then changed to ChemPLP, but there was any improvement neither. Finally, some different modes of defining the binding site were tested, but no better predicted orientations were

[a] The residues numbering is referring to *Dictyostelium discoideum* NDPK, but the residues are equivalent to the previously mentioned human NDPK ones.

obtained. In view of these results, the study of **HS1'-G^{Cl}** and **HS1-G** derivatives was performed on one of the NDPK subunits and without any flexible residue.

Considering that just a few orientations of GDP were properly posed, and that these complexes did not reproduce the crystallographic structure as precisely as in the previous activation steps, conclusions drawn from these calculations should be read with caution. Only a tendency on the success of this last activation step can be inferred.

Dockings of ACVDP with NDPK were carried out to provide with a structural and energetic benchmark. Again, a reduced number of binding modes were compatible with the catalysis. The lowest energy binding mode with the β-phosphate and the base nearly overlapped to the crystallographic ones (Figure 73) was taken as reference despite its high binding energy, -5 score units (Table ES31). As mentioned, this high binding energy value as well as the reduced number of results compatible with the catalysis are presumably caused by the broadly solvent exposed binding site. This is also reinforced by the lack of a hydroxyl on ACV mimicking the 3'-OH of the natural substrates, as this 3'-OH is known to stabilize the particular conformation of the nucleotides that facilitates the phosphoryl transfer.[363] The same behaviour was observed when performing the calculations of the ACVTP: not many binding modes are consistent with a catalytic phosphoryl transfer, and the corresponding binding energies are also very high (Table ES32).

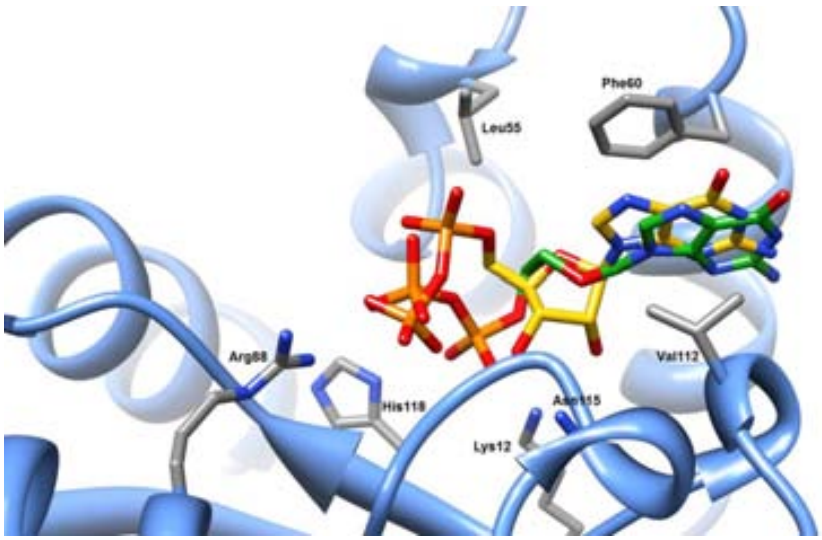


Figure 73. ACVDP (green) superimposed to crystallographic GDP (yellow) in human NDPK binding site (PDB code: 1NUE, X-ray residues shown in grey). For the sake of clarity, hydrogen atoms, hydrogen bonds, crystallographic waters and ions are not shown.

The structural criterion to analyse docking results considered that the phosphorylation process would be more likely for compounds showing the β-phosphate pointing towards the catalytic His-118. Following this criterion, the last catalytic phosphoryl transfer would be less favoured for **HS1-GDP** than for **HS1'-G^{Cl}DP**.

Some binding modes of **HS1'-G^{Cl}DP** and **HS1-GDP** were consistent with the catalysis, and the binding energies of **HS1'-G^{Cl}DP** (-16 score units, Table ES31) were lower than those of **HS1-GDP** (around -12 score units).

Concerning **HS1'-G^{Cl}TP** and **HS1-GTP**, some of their predicted poses were also properly posed (Figure 74). A higher number of orientations compatible with the catalysis as well as lower binding energies were obtained for **HS1'-G^{Cl}TP** than for **HS1-GTP** (Table ES32), indicating that the phosphoryl transfer is more likely on this nucleoside. This phosphoryl transfer is energetically more favoured at the 2''' position of **HS1'-G^{Cl}** than at its 1''' position (Table ES32).

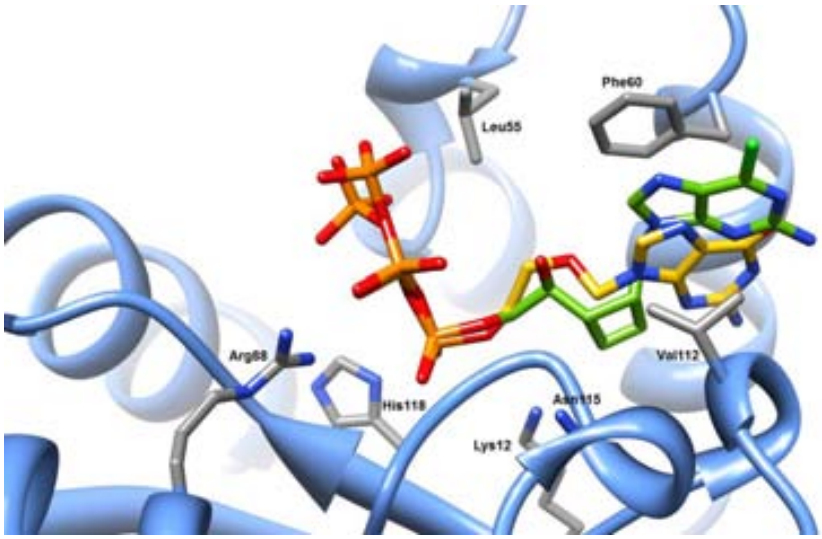


Figure 74. HS1'-G^{Cl}-2'''-TP (green) superimposed to ACVTP (yellow) in human NDPK binding site (PDB code: 1NUE, X-ray residues shown in grey). For the sake of clarity, hydrogen atoms, hydrogen bonds, crystallographic waters and ions are not shown.

Regarding calculations on pyrimidine compounds, the protocol was first validated by docking dTDP into the corresponding X-ray structure (PDB code: 1NDC). Its lowest energy binding modes featured the β-phosphate oriented towards the catalytic His-122, and the base located in the corresponding cleft near the protein surface (Figure 75). Despite the fact that the reproduction of the crystallographic structure was not as good as in the previous activation steps, dTDP was taken as a

structural and energetic benchmark. Again, docking limitations to cope with this solvent exposed binding site only allowed us to determine a tendency on the success of this last phosphoryl transfer to pyrimidine derivatives, being **HS1-T** the most favoured compound.

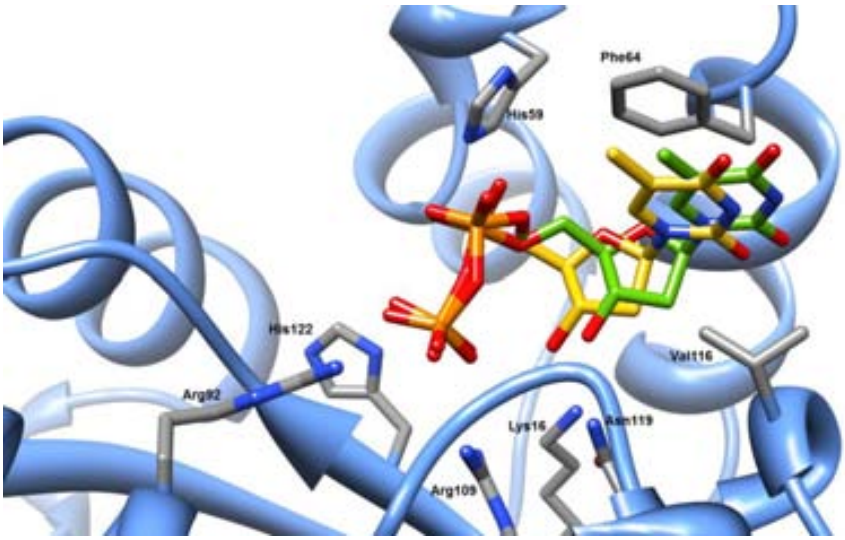


Figure 75. Docked dTDP (green) superimposed to crystallographic dTDP (yellow) in *Dictyostelium discoideum* NDPK binding site (PDB code: 1NDC, X-ray residues shown in grey). For the sake of clarity, hydrogen atoms, hydrogen bonds, crystallographic waters and ions are not shown.

Several of the binding modes predicted for **HS1-TDP**, **HS2-TDP** and **HS2'-TDP** were compatible with the catalysis (Figure 76), so all these nucleosides are likely to be phosphorylated by NDPK. As **HS1-TDP** has a higher number of orientations properly posed and lower binding energies than **HS2-TDP** and **HS2'-TDP** (Table ES29), its phosphorylation is preferred.

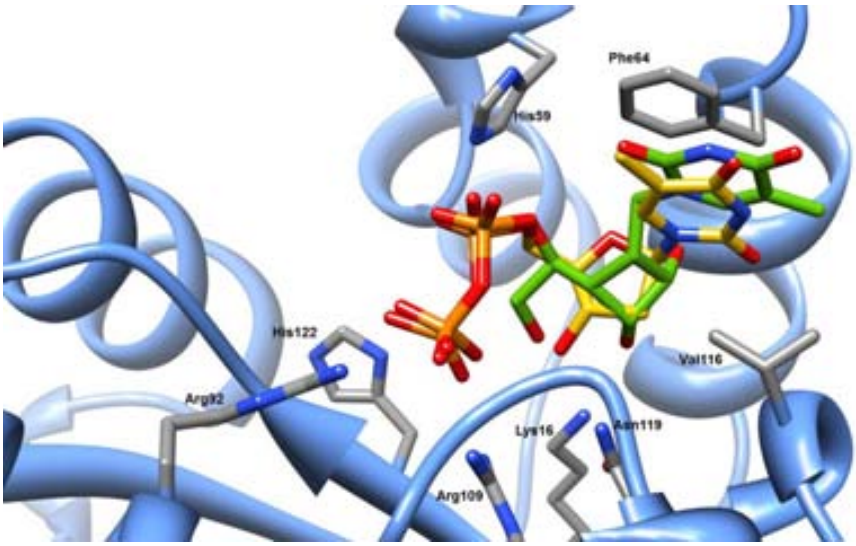


Figure 76. HS2-T-1'''-DP (green) superimposed to crystallographic dTDP (yellow) in *Dictyostelium discoideum* NDPK binding site (PDB code: 1NDC, X-ray residues shown in grey). For the sake of clarity, hydrogen atoms, hydrogen bonds, crystallographic waters and ions are not shown.

Triphosphorylated nucleosides behaved exactly as diphosphorylated derivatives: **HS1-TTP** is structurally and energetically more favoured than **HS2-TTP** and **HS2'-TTP** (Table ES30).

4. Conclusions

To rationalize the lack of activity of nucleoside analogues **HS1** and **HS2** against HSV-1, two of the main requirements for these compounds to be active were investigated: their drug-likeness and their activation process.

The evaluation of their drug-likeness showed that these nucleosides satisfy the "Lipinski's rule of 5" and the Veber's criterion, supporting their good drug-likeness.

The activation process of nucleosides **HS1** and **HS2** was then analysed by means of molecular docking. The theoretical results derived from this study revealed that nucleosides **HS2-T**, **HS2'-T** and **HS1'-G^{Cl}** are not likely to be converted into their active form, the triphosphorylated derivatives. In particular, the second phosphoryl transfer to **HS2-T** is expected to be unsuccessful, whereas **HS2'-T** and **HS1'-G^{Cl}** fail both in their first and second additions of the phosphoryl groups. Thus, the modified base of **HS1'-G^{Cl}** as well as the vicinal diol on the cyclobutane ring of **HS2-T** and **HS2'-T** are restrictive for their correct activation.

By contrast, no limitations were found in the phosphoryl transfers to **HS1-T** and **HS1-G**, indicating that the substitution of the sugar moiety by a cyclobutene and the inverted configuration of these nucleosides are not hampering their activation. Assuming that they are activated, their lack of activity against HSV-1 could be due to the failure in the incorporation into the viral DNA strand carried out by HSV-1 DNA polymerase.

Overall, this study highlighted the limiting factors in the antiviral activity of **HS1** and **HS2** nucleosides against HSV-1: whereas the activation process is restricting the activity of **HS2-T**, **HS2'-T** and **HS1'-G^{Cl}**, the absence of activity observed for **HS1-T** and **HS1-G** is neither due to their bad drug-likeness nor to their failure in becoming activated.

CHAPTER VI

Rational design of carbocyclic nucleoside analogues

1. Introduction

Protein–ligand dockings are widely used to propose potential drug candidates in drug discovery projects. In this field, dockings may be applied in two different ways. On the one hand, to provide the orientation of a given compound in the binding site of its biological target, which is subsequently used to design drug candidates based on the structure of this compound. On the other hand, to evaluate the interactions between a large pool of compounds and a biological target, which is used to select a small number of them as possible leads.[266,364–366]

In the present chapter, these two applications of docking are merged to suggest new scaffolds of nucleoside analogues as anti-HSV agents, on the basis of the molecular information gained in the preceding study on HSV (Chapter V). The approach envisaged will be introduced in the next paragraphs. However, it is worth noting that this is only a preliminary study of a complete rational design of novel carbocyclic nucleosides that is currently underway, which may eventually lead to the synthesis of the envisaged nucleoside analogue(s). Especially in this last chapter, it important to keep in mind what a pioneer in the field of computational chemistry, William L. Jorgensen, stated regarding the role of computers in drug discovery processes: "There is not going to be a *voilà* moment at the computer terminal. Instead, there is systematic use of wide-ranging computational tools to facilitate and enhance the drug discovery process".[274]

Approach

Our approach consists of two steps: (1) generation of candidates, based on previously predicted complexes between nucleoside analogues and HSV-1 TK, and (2) evaluation of these candidates, by studying the three steps of their activation process, to determine the final target nucleosides (Figure 77). The enzyme selected to generate the candidates has been HSV-1 TK since it has shown to be the most restrictive in the activation of nucleoside analogues in Chapter V. Furthermore, this enzyme has given promising results in suicide gene therapies for cancer treatment, and thus compounds that are selectively phosphorylated by HSV-1 TK may be used as anticancer agents.[345,352,367]

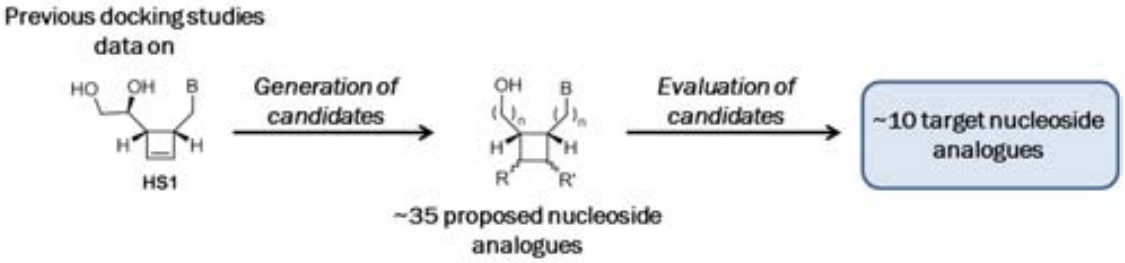


Figure 77. General strategy envisaged for the rational design of novel carbocyclic nucleosides (illustrated for cyclobutene and cyclobutane nucleoside analogues).

Generation of candidates

Three steps are followed to design and generate a set of nucleoside analogues as candidates: (1) choosing a predicted complex between a nucleoside analogue and HSV-1 TK consistent with a catalytic phosphoryl transfer; (2) identifying, in the binding site, residues side chains that could provide additional interactions with the ligand, thus favouring the binding mode compatible with the catalysis; (3) proposing novel nucleoside analogues based on the skeleton chosen in (1) but featuring the substituents needed for the protein–ligand interactions envisaged in step (2).

Evaluation of candidates

Once a series of candidates has been proposed, a computational evaluation of their activation process is performed in order to determine the compounds that are expected to be converted into their triphosphorylated derivatives. For this purpose, each nucleoside analogue is docked into the X-ray structures of each kinase involved in the activation process in HSV-1 infected cells.[a]

Some differences between the present procedure and the procedure used so far to perform the corresponding docking calculations have to be pointed out: a single PDB structure is used to investigate each activation step (instead of the multiple PDB structures employed in the previous study); compounds are only docked into the corresponding crystallographic structures in their reactant form (i.e. un-, mono- and diphosphorylated forms for the first, second and third activation steps, respectively); nucleosides that are not likely to be phosphorylated by a certain kinase will not be further studied in the following activation steps; only the lowest energy orientation from the 20 predicted binding modes of each nucleoside

[a] See section *1.1. Mechanism of action of nucleoside analogues in HSV-1 infected cells. Biological fundamentals* of Chapter V.

is considered; the results are analysed in structural terms. These simplifications reduce considerably the number of predicted complexes to analyse, allowing us to perform the study on a bigger set of compounds than in the previous studies described in Chapters IV and V in a reasonable time (Figure 78).

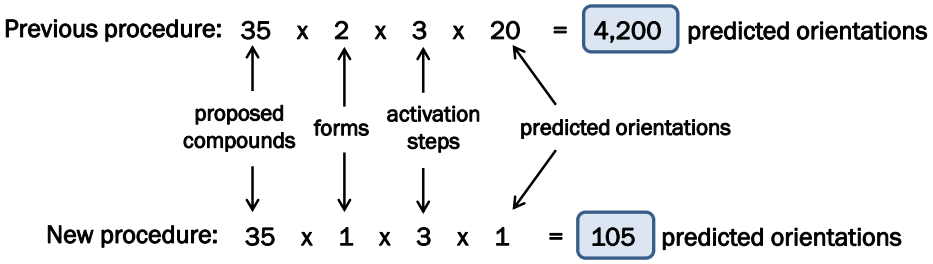


Figure 78. Comparison of the rough amount of predicted orientations to analyse for a proposed set containing 35 compounds following the procedure previously used in Chapters IV and V (top) vs. the procedure used herein (bottom). The number of PDB structures used to investigate each activation step as well as the number of nucleosides that will not be further studied when failing in a certain phosphorylation step are not considered.

Compounds that perform better in the three successive phosphorylation steps will be the target nucleosides.

2. Computational details

Protein–ligand docking calculations were performed using the same methodology and parameters than in the previous study on HSV.[a]

3. Rational design of cyclobutene and cyclobutane L-nucleoside analogues

On the basis of the molecular modelling data previously obtained in Chapter V, our approach was used to carry out a rational design of novel cyclobutene and cyclobutane L-nucleoside analogues as anti-HSV agents.

The results of the present study will be presented in two different sections in agreement with the general steps of our approach: (1) generation of candidates, followed by (2) evaluation of the proposed candidates. The conclusions of these investigations will be included at the end of the present study.

[a] See section *2. Computational details* of Chapter V.

3.1. Generation of candidates

3.1.1. Initial series of nucleosides

Following the previous investigations on cyclobutane and cyclobutene L-nucleoside analogues, two series of compounds were proposed to be investigated: nucleosides **HS3**, and nucleosides **HS8** and **HS9** (Figure 79).

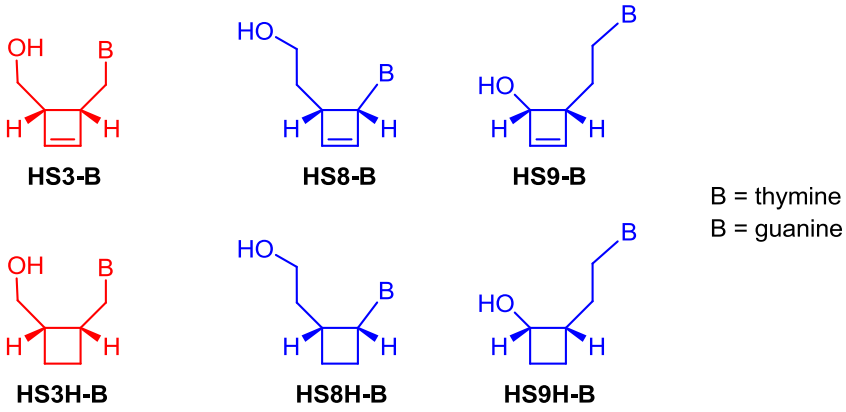


Figure 79. Nucleoside analogues **HS3** (blue), **HS8** and **HS9** (red) and their cyclobutane analogues **HS3H**, **HS8H** and **HS9H**.

We first decided to investigate cyclobutene nucleosides **HS3** that are based on the structure of **HS1** but without featuring their additional hydroxymethyl moiety. Interestingly, nucleosides **HS3** are the L- counterparts of D-cyclobutene nucleosides **2** synthesized by Huet and co-workers in 2002 (Scheme 3, page 31).[103]

Cyclobutene nucleosides **HS8** and **HS9** were proposed as candidates to check whether nucleosides featuring the nucleobase–alcohol distance of natural nucleosides but with either the base (nucleosides **HS8**) or the alcohol (nucleosides **HS9**) directly bound to the carbocycle could be activated.

The equivalent cyclobutane analogues, **HS3H**, **HS8H** and **HS9H**[a] were also proposed to be herein studied, since according to our group experience the synthesis of cyclobutane nucleosides should be easier than the preparation of cyclobutene analogues.

[a] The "H" letter is used in the nomenclature of these nucleosides to indicate that they are the cyclobutane counterparts of the corresponding cyclobutene nucleosides.

3.1.2. Proposal of substituted nucleosides

A set of substituted nucleosides based on the skeletons of **HS3** and **HS8** were designed following the above described procedure for the generation of candidates. First, nucleosides **HS3** and **HS8** were docked into HSV-1 TK binding site (step 1). Then, the orientations compatible with the catalysis were analysed to identify the residues that could provide further interactions with the ligand (step 2, Figure 80).

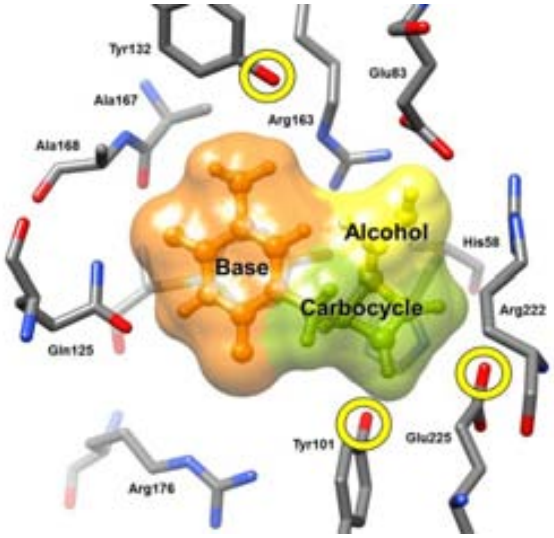
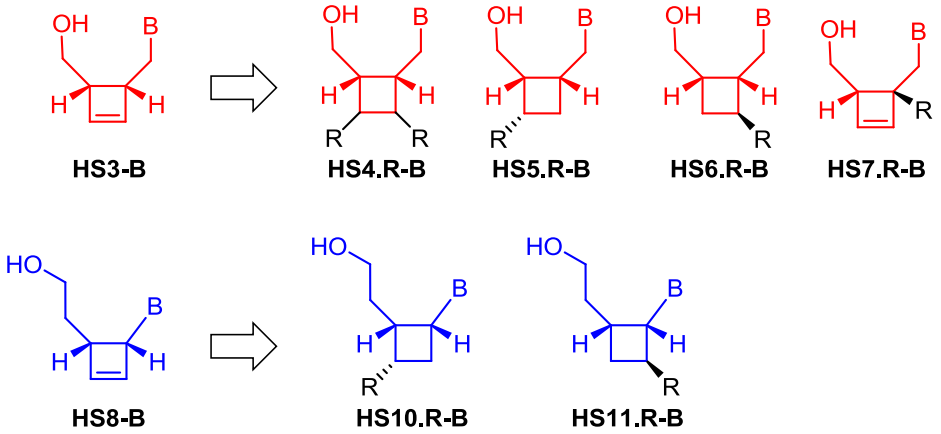
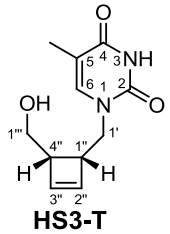
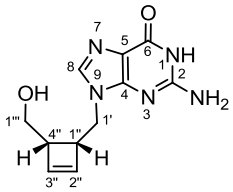
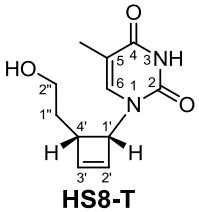


Figure 80. Location of the nucleobase (orange), carbocycle (green) and alcohol (yellow) of nucleosides **HS3** and **HS8** (exemplified for **HS3-T**) in HSV-1 TK binding site (PDB code: 1KIM, X-ray residues shown in grey). For the sake of clarity, crystallographic waters and hydrogen atoms of the residues are not shown. The residues selected to design substituted nucleosides are circled in yellow.

Finally, nucleosides **HS4**–**HS7** and **HS10**–**HS11** were proposed starting from the structures of **HS3** and **HS8** complexed with HSV-1 TK, respectively (step 3, Table 12).

Table 12. Candidates generated from **HS3** and **HS8**.



Initial compound	Possible interacting residues	Substituent (position)	Nucleoside
 HS3-T	Tyr-132	F (CH_3-C-5)	**HS3-T^F**
	Tyr-101, Glu-225	F (2'', 3'')	**HS4.F-T**[a]
	Tyr-101, Glu-225	F (2'', 3'')	**HS4'.F-T**[a]
	Glu-225	F (3'')	**HS5.F-T**
	Glu-225	CH_2OH (3'')	**HS5.CH_2OH-T**
	Tyr-101	F (2'')	**HS6.F-T**
	Tyr-101	CH_2F (2'')	**HS6.CH_2F-T**
	Tyr-101	CH_2OH (2'')	**HS6.CH_2OH-T**
 HS3-G	Tyr-101	OH (2'')	**HS6.OH-G**
	Tyr-101	CH_2F (2'')	**HS6.CH_2F-G**
	Tyr-101	CH_2OH (2'')	**HS6.CH_2OH-G**
	His-58	OH (1'')	**HS7.OH-G**
 HS8-T	Glu-225	CH_2OH (3'')	**HS10.CH_2OH-T**
	Tyr-101	CH_2OH (2'')	**HS11.CH_2OH-T**
	Tyr-101	CH_2NH_2 (2'')	**HS11.CH_2NH_2-T**
	Tyr-101	CH_2F (2'')	**HS11.CH_2F-T**

[a] **HS4.F-T** and **HS4'.F-T** differ in the absolute configuration of the atoms with the fluorine substituent (2'' and 3''): **HS4.F-T** corresponds to a (2''*R*,3''*S*) configuration and **HS4'.F-T** to (2''*S*,3''*R*).

After the evaluation of the three activation steps of nucleosides **HS3**–**HS11** (*vide infra*), some more nucleosides were proposed as candidates. In particular, the

good results of **HS8H-T** encouraged us to study nucleosides **HS12H**,[a] featuring the nucleobase directly bound to the cyclobutane ring and the alcohol bound via a methylene linker, as in natural nucleosides (Figure 81). The promising results also obtained for **HS5.CH$_2$OH-T** prompted us to propose nucleosides **HS13** as well, which present the base directly bound to the cyclobutane ring.

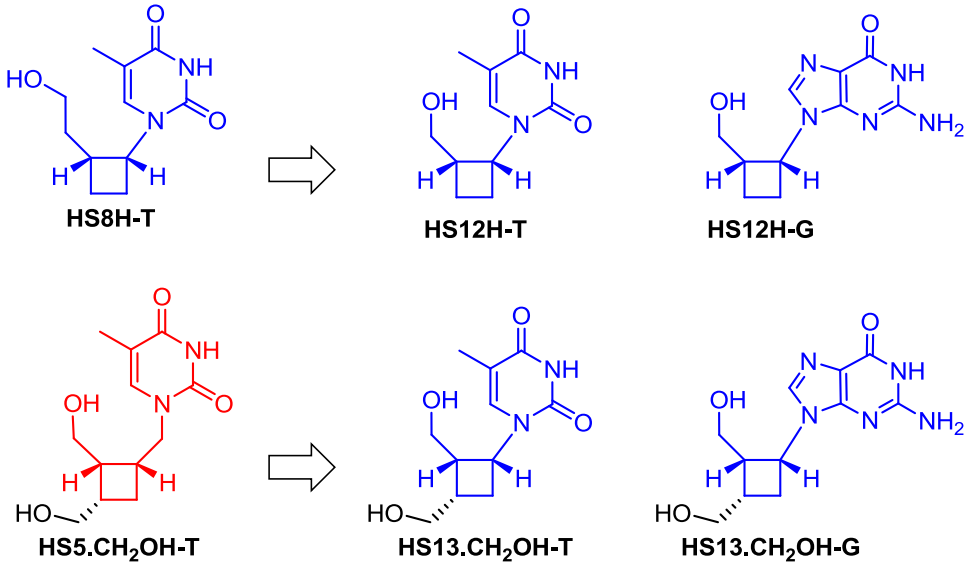


Figure 81. Nucleoside analogues **HS12H** and **HS13**.

Thus, 32 novel nucleoside analogues were proposed to be studied by means of molecular docking as anti-HSV agents. Calculations of the standard molecular descriptors to evaluate their drug-likeness were performed with the *Marvin* package[328] prior to the study of their activation process. All the investigated compounds responded positively to "Lipinski's rule of 5" and Veber's criterion, which sustained their good drug-likeness (Tables ES33 and ES34).

3.2. Evaluation of candidates

As above mentioned, the evaluation of candidates **HS3**–**HS13** is aimed at studying their conversion into the triphosphorylated derivatives as in the earlier study on **HS1** and **HS2** nucleosides. The activation process of nucleoside analogues in HSV-1 infected cells as well as the description of the kinases involved in this

[a] Despite the corresponding cyclobutene nucleosides were not investigated, the nomenclature containing the "H" letter to indicate the presence of a cyclobutane ring instead of a cyclobutene has been maintained in nucleosides **HS12** by analogy with nucleosides **HS3**, **HS8** and **HS9**.

process have been detailed in Chapter V,[a] so this information will not be presented herein again.

To establish if the proposed nucleosides could fail in becoming activated, protein–ligand docking studies on HSV-1 TK, GMPK and NDPK were performed on some of the crystallographic structures previously used (Table 13).

Table 13. PDB structures used to study the activation process of nucleosides **HS3**–**HS13**.

Enzyme	Organism	PDB code	Crystallized ligand	Docking calculations[a]
TK	HSV-1	1KIM	dT	–OH pyrimidine derivatives
TK	HSV-1	2KI5	ACV	–OH purine derivatives
TK	HSV-1	1VTK	dTMP (+ ADP)	MP pyrimidine derivatives
GMPK	*Mus musculus*	1LVG	GMP (+ ADP)	MP purine derivatives
NDPK	*Dictyostelium discoideum*	1NDC	dTDP	DP pyrimidine derivatives
NDPK	*Homo sapiens*	1NUE	GDP	DP purine derivatives

[a] –OH, MP and DP stand for unphosphorylated, monophosphorylated and diphosphorylated compounds, respectively.

Nucleoside analogues **HS3**–**HS13** were docked separately into the active site of each kinase in their reactant form, and the lowest energy binding mode predicted for each nucleoside was analysed in structural terms following the same criteria as in the preceding study. dT and ACV were also used as structural benchmarks for pyrimidine and purine compounds, respectively.

The results obtained in the protein–ligand docking studies of the three successive phosphoryl transfers are summarized in Figure 82 and discussed in the following paragraphs.

[a] See section *1.1. Mechanism of action of nucleoside analogues in HSV-1 infected cells. Biological fundamentals* of Chapter V.

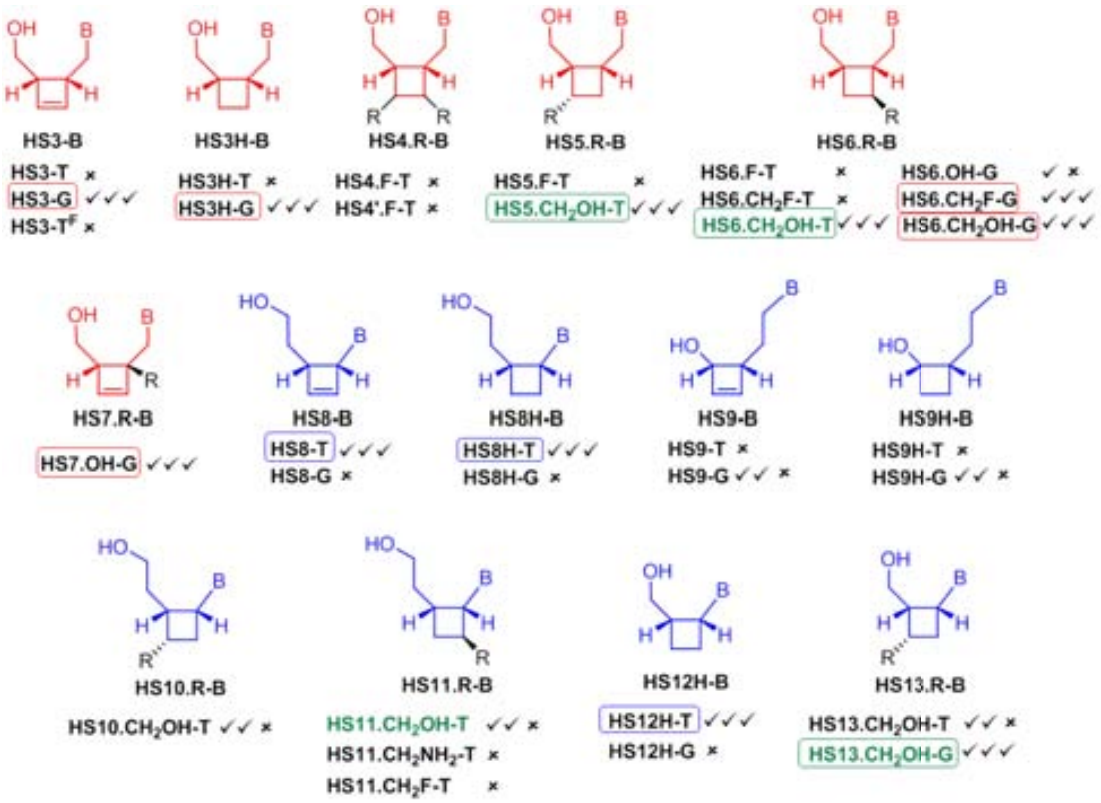


Figure 82. Protein–ligand docking results of the first, second and third activation steps of nucleosides **HS3**–**HS13**. The most favoured nucleosides based on **HS3** and **HS8**–**HS9** skeletons are indicated in red and blue boxes, respectively. Nucleosides in which the phosphoryl transfers are expected to occur on the alcohol moiety of their substituents R are indicated in green.

1st phosphorylation step

Dockings of the first activation step proved that the lack of the additional hydroxymethyl moiety in nucleosides **HS3** and **HS3H** compared to **HS1** is detrimental to the proper binding of pyrimidine nucleosides but not of purine nucleosides. Concerning the corresponding derivatives **HS4**–**HS7** substituted at different positions of the cyclobutane ring, calculations revealed that these compounds are predicted to be similarly oriented to the related **HS3** nucleosides, so that these substituents do not cause significant changes in the predicted poses.

Protein–ligand dockings also showed that reducing the distance between the nucleobase and the carbocycle of pyrimidine nucleosides is not restrictive for the catalytic activity of the enzyme (nucleosides **HS8** and **HS8H** vs. nucleosides **HS3** and **HS3H**). Conversely, this modification is unfavourable for the activation process of purine nucleosides. An opposite profile is observed with regard to nucleosides featuring the alcohol directly bound to the carbocycle (**HS9** and **HS9H**): whereas the

catalytic phosphoryl transfer to the corresponding pyrimidine compounds is unlikely, purine nucleosides are expected to be phosphorylated by HSV-1 TK.

The substituent of nucleoside **HS10** favours the binding mode compatible with the catalysis predicted for the related nucleoside **HS8-T**. By contrast, substituents of nucleosides **HS11** proved detrimental for the catalytic orientation of these compounds (except for **HS11.CH$_2$OH-T**, in which the phosphoryl transfer may occur on the alcohol moiety of the substituent, Figure 83).

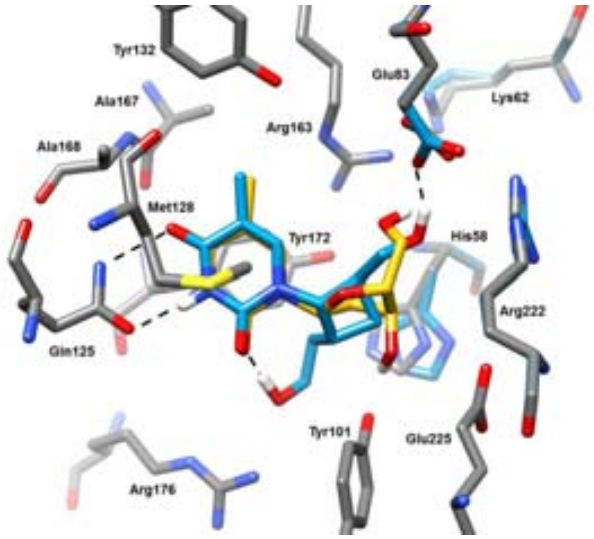


Figure 83. HS11.CH$_2$OH-T (blue) superimposed to crystallographic dT (yellow) in HSV-1 TK binding site (PDB code: 1KIM, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are only shown when bound to a heteroatom of the ligand. Hydrogen bonds between **HS11.CH$_2$OH-T** and residues are depicted as dotted lines.

Finally, maintaining the distance between the base and the alcohol moieties in pyrimidine analogues with the nucleobase directly bound to the carbocycle does not seem to be essential for the proper phosphorylation of these compounds (nucleoside **HS12H-T** vs. **HS8H-T** and **HS13.CH$_2$OH-T** vs. **HS10.CH$_2$OH-T**).

2nd phosphorylation step

The theoretical results derived from the study of the second activation step revealed that all the nucleosides featuring orientations compatible with the catalysis in the previous phosphoryl transfer are also expected to be converted into their diphosphorylated derivatives (with the exception of **HS6.OH-G**). Indeed, most of the predicted binding modes of monophosphorylated pyrimidine nucleosides closely resemble the calculated orientations of the corresponding unphosphorylated nucleosides, maintaining the additional interactions between the cyclobutane

substituents and residues in the binding site previously observed. Therefore, no filtering is associated with this phosphoryl transfer except for nucleoside **HS6.OH-G**.

3rd phosphorylation step

Protein–ligand dockings of the last activation step led to two different tendencies depending on the nature of the nucleoside base. Calculations on pyrimidine nucleosides showed that analogues substituted on the cyclobutane ring are less likely to be phosphorylated by NDPK than the related nucleosides having no substituents (compounds **HS10.CH$_2$OH-T** and **HS11.CH$_2$OH-T** vs. **HS8-T** and **HS8H-T** —Figure 84— and nucleoside **HS13.CH$_2$OH-T** vs. **HS12H-T**). Concerning purine nucleosides, the corresponding calculations revealed that the presence of substituents on the cyclobutane ring does not significantly alter the predicted binding modes of the corresponding nucleosides without substituents (nucleosides **HS6.CH$_2$F-G**, **HS6.CH$_2$OH-G** and **HS7.OH-G** vs. **HS3-G** and **HS3H-G**), and that the last phosphoryl transfer is less favoured in compounds featuring the alcohol directly bound to the carbocycle (nucleosides **HS9-G** and **HS9H-G**).

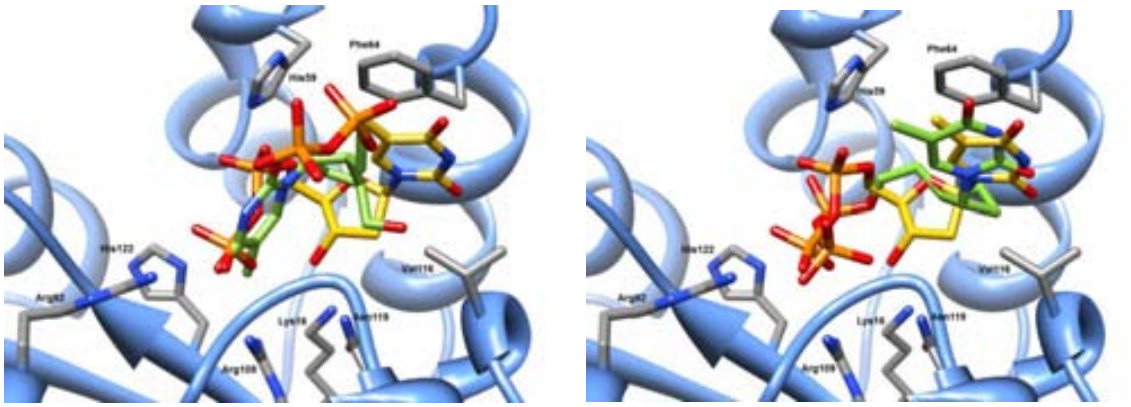


Figure 84. HS10.CH$_2$OH-TDP (green) superimposed to crystallographic dTDP (yellow) in *Dictyostelium discoideum* NDPK binding site (PDB code: 1NDC, X-ray residues shown in grey) (left), and **HS8H-TDP** (green) superimposed to crystallographic dTDP (yellow) in *Dictyostelium discoideum* NDPK binding site (PDB code: 1NDC, X-ray residues shown in grey) (right). For the sake of clarity, hydrogen atoms, hydrogen bonds, crystallographic waters and ions are not shown.

3.3. Conclusions

A rational design of cyclobutene and cyclobutane L-nucleoside analogues based on protein–ligand dockings was performed.

Calculations on each activation step revealed which skeletons and substituents favour or are detrimental to the catalytic activity of each enzyme. In

general terms, dockings showed that changing the cyclobutene by a cyclobutane ring (not substituted) does not alter the catalytic activity of the enzymes. Moreover, the conversion of nucleosides **HS3** into the triphosphorylated derivatives is more favoured for purine compounds whereas the activation process of nucleosides **HS8** and **HS12H**, both featuring the base directly bound to the carbocycle, is more likely on pyrimidine compounds. Conversely, **HS9** skeleton is restrictive for the activation of these nucleosides.

The present study demonstrated that the whole activation process is more favoured for pyrimidine nucleosides **HS5.CH_2OH-T**, **HS6.CH_2OH-T**, **HS8-T**, **HS8H-T** and **HS12H-T** as well as for purine nucleosides **HS3-G**, **HS3H-G**, **HS6.CH_2F-G**, **HS6.CH_2OH-G**, **HS7.OH-G** and **HS13.CH_2OH-G** (Figure 82). These results also highlighted the difficulty to generalize how to modify a certain nucleoside skeleton to favour its conversion into the active form, due to the complexity of this process that involves three successive phosphoryl transfers performed by different kinases. Therefore, docking calculations must be performed for each candidate on each of these activation steps. To complete the rational design, the interaction of the selected nucleosides with HSV-1 DNA polymerase should be studied.

4. Rational design of cyclohexene and bicyclo[4.1.0]heptane nucleoside analogues

Cyclohexene nucleosides are a promising class of antiviral compounds wherein a carbon-carbon double bond replaces the oxygen atom of the furanose ring, thus mimicking the conformations of natural nucleosides.[368] As mentioned in the introduction of the present thesis, it has been described that both enantiomers of the nucleoside analogues Cyclohexenyl-G (DCG and LCG, respectively) are active against some herpes viruses (HSV-1, HSV-2, VZV, CMV) (Figure 85).[96] In this field, our research group has reported the enantiodivergent synthesis of nucleosides D-**HS14-A**, D-**HS14-U** and L-**HS14-U** (Figure 85), which were evaluated as anti-HIV agents without showing antiviral activity at concentrations below 25 $\mu g \cdot mL^{-1}$.[369]

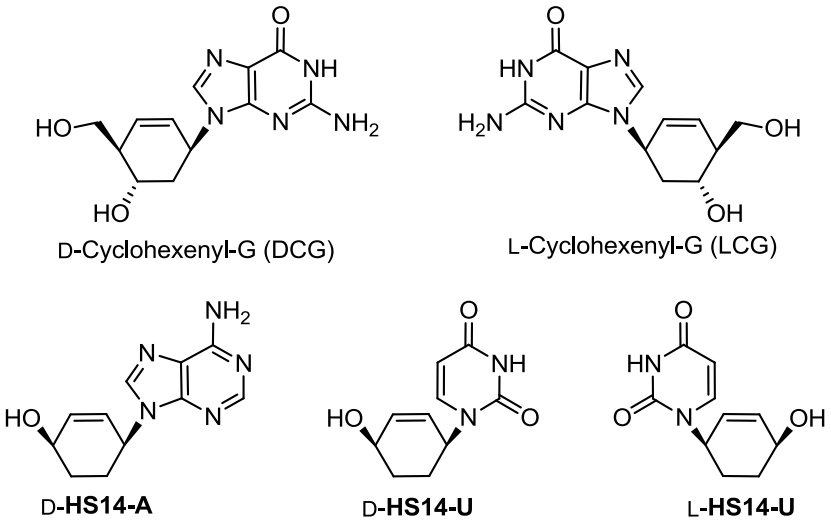


Figure 85. DCG and LCG (Herdewijn and co-workers, 2000)[96] and cyclohexene nucleosides **HS14** synthesized in our research group in the course of the research program devoted to the synthesis of carbocyclic nucleosides (Alibés and co-workers, 2009).[369]

The excellent antiviral activity exhibited by DCG and LCG as well as the experience acquired in our group to prepare chiral cyclohexene nucleosides encouraged us to perform a rational design, analogous to the previous one, of cyclohexene nucleosides as anti-HSV agents. As before, the results of this study will be presented in two different sections: (1) generation of candidates, and (2) evaluation of candidates.

4.1. Generation of candidates

4.1.1. Initial series of nucleosides

Two series of nucleoside analogues based on the skeleton of the cyclohexene moiety were envisaged: analogues **HS14** and **HS17** (Figure 86).

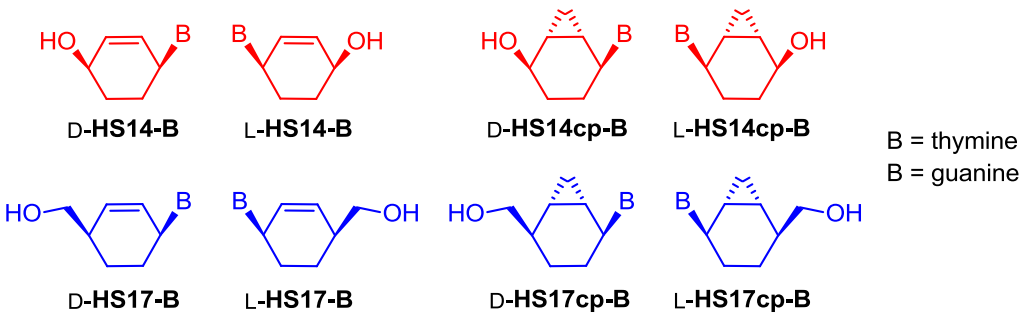


Figure 86. Nucleoside analogues **HS14** (red) and **HS17** (blue), and their cyclopropane-fused derivatives **HS14cp** and **HS17cp**.

First, nucleosides **HS14** containing different nucleobases than the previously synthesized analogues were proposed as candidates. With the aim of studying the effect of substituting the double bound of **HS14** by a fused cyclopropane ring, which could confer more flexibility to the carbocycle while still mimicking cyclohexene conformations, it was planned to investigate the corresponding cyclopropane-fused derivatives **HS14cp**. Both enantiomers of all these nucleosides were suggested to be studied considering the precedent of D- and L-Cyclohexenyl-G, which were both active.

Nucleosides **HS17**, which are analogous to **HS14** but featuring a methylene spacer between the alcohol and the carbocycle, were also candidates to test the hypothesis that the higher flexibility of the alcohol moiety favours their activation process. The corresponding cyclopropane-fused derivatives, **HS17cp**, were proposed to be studied as well.

4.1.2. Proposal of substituted nucleosides

As before, nucleosides **HS14** and **HS17** were docked into the binding site of HSV-1 TK and the poses consistent with a catalytic phosphoryl transfer were analysed to design substituted nucleosides. The residues used for this design are shown in Figure 87.

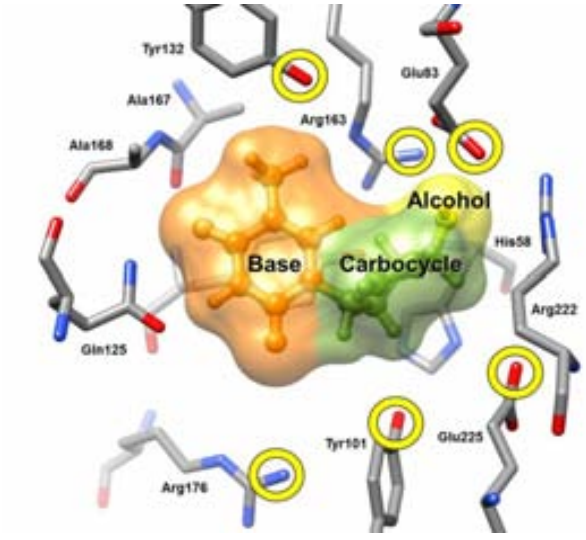
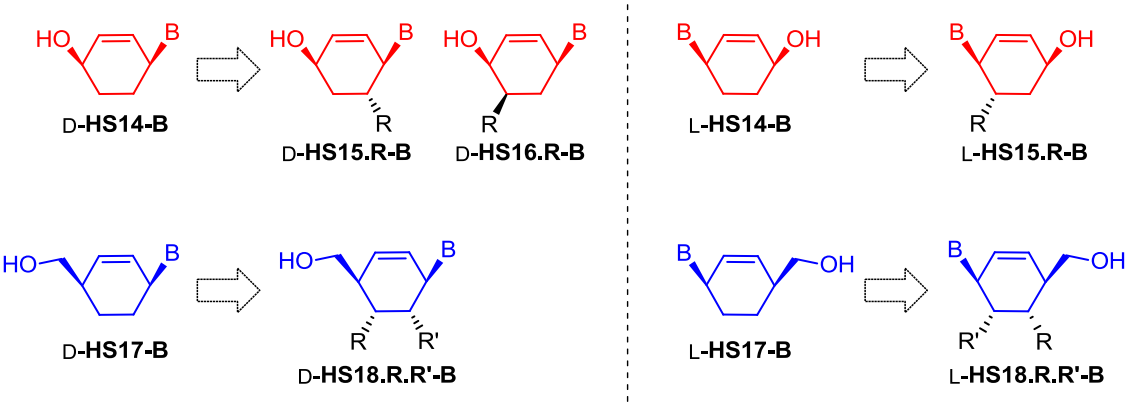
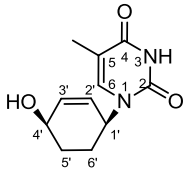
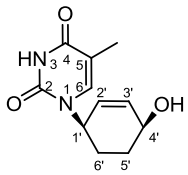
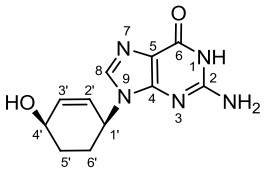
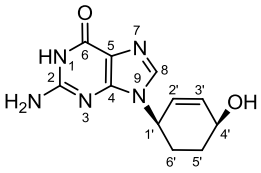
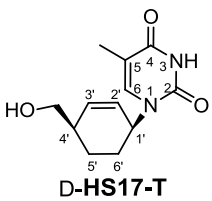


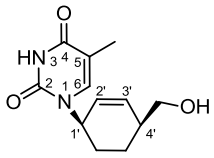
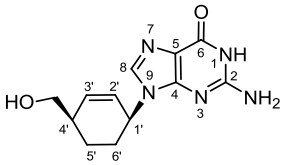
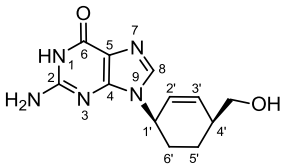
Figure 87. Location of the nucleobase (orange), carbocycle (green) and alcohol (yellow) of nucleosides **HS14** and **HS17** (exemplified for D-**HS14-T**) in HSV-1 TK binding site (PDB code: 1KIM, X-ray residues shown in grey). For the sake of clarity, crystallographic waters and hydrogen atoms of the residues are not shown. The residues selected to design substituted nucleosides are circled in yellow.

Thus, nucleosides **HS15**–**HS16** and **HS18** were designed on the basis of the binding modes of nucleosides **HS14** and **HS17**, respectively (Table 14).

Table 14. Candidates generated from **HS14** and **HS17**.



Initial compound	Possible interacting residues	Substituent (position)	Nucleoside
 D-**HS14-T**	Tyr-101	OH (6')	D-**HS15.OH-T**
	Tyr-101	CH_2OH (6')	D-**HS15.CH_2OH-T**
	Arg-163	OH (5')	D-**HS16.OH-T**
	Arg-163	CH_2OH (5')	D-**HS16.CH_2OH-T**
 L-**HS14-T**	Glu-225	CH_2CH_2OH (6')	L-**HS15.CH_2CH_2OH-T**
 D-**HS14-G**	Tyr-101	OH (6')	D-**HS15.OH-G**
	Tyr-101	CH_2OH (6')	D-**HS15.CH_2OH-G**
	Arg-163	OH (5')	D-**HS16.OH-G**
	Arg-163	CH_2OH (5')	D-**HS16.CH_2OH-G**
 L-**HS14-G**	Glu-225	CH_2CH_2OH (6')	L-**HS15.CH_2CH_2OH-G**
 D-**HS17-T**	His-58, Tyr-101	OH (5'), OH (6')	D-**HS18.OH.OH-T**
	His-58, Tyr-101	OH (5'), CH_2OH (6')	D-**HS18.OH.CH_2OH-T**

 L-**HS17-T**	Tyr-101, Arg-176	OH (5'), CH_2CH_2OH (6')	L-**HS18.OH.CH_2CH_2OH-T**
 D-**HS17-G**	His-58, Tyr-101	OH (5'), OH (6')	D-**HS18.OH.OH-G**
	His-58, Tyr-101	OH (5'), CH_2OH (6')	D-**HS18.OH.CH_2OH-G**
 L-**HS17-G**	Glu-83, Tyr-132	OH (5'), CH_2CH_2OH (6')	L-**HS18.OH.CH_2CH_2OH-G**

Hence, 32 novel nucleoside analogues were proposed to be investigated by means of molecular docking as anti-HSV agents. The standard molecular descriptors to evaluate their drug-likeness were calculated with the *Marvin* package[328] prior to the study of their activation process. All the candidates responded positively to "Lipinski's rule of 5" and Veber's criterion, with the exception of nucleosides D-**HS18.OH.OH-G**, D-**HS18.OH.CH_2OH-G** and L-**HS18.OH.CH_2CH_2OH-G** which calculated polar surface areas are above the limit fixed in Veber's studies (Tables ES35 and ES36). However, the rest of the standard descriptors are in agreement with "Lipinski's rule of 5". For this reason, we decided not to eliminate these nucleosides from the study.

4.2. Evaluation of candidates

The three successive activation steps of nucleosides **HS14**–**HS18** were studied by means of protein–ligand dockings that were carried out and analysed in the same way than in the earlier investigations on **HS3**–**HS13**. The same crystallographic structures of HSV-1 TK, GMPK and NDPK were also used (Table 13, page 166).

The results of these docking calculations are summarized in Figure 88 and discussed in the following paragraphs.

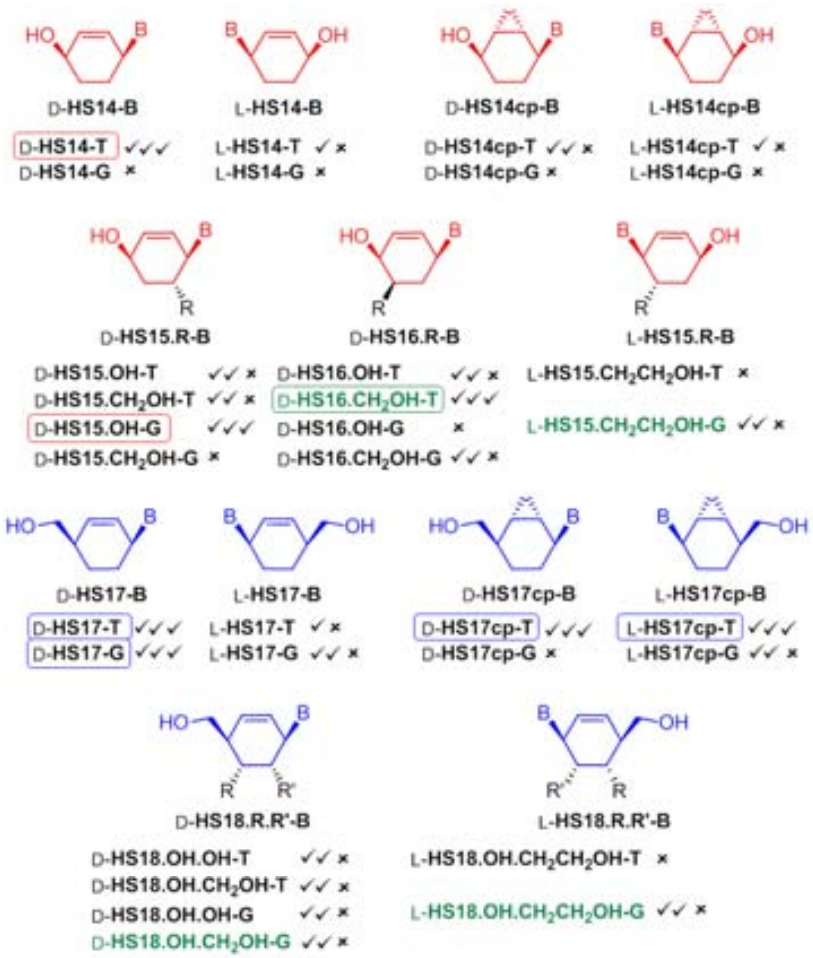


Figure 88. Protein–ligand docking results of the first, second and third activation steps of nucleosides **HS14**–**HS18**. The most favoured nucleosides based on **HS14** and **HS17** skeletons are indicated in red and blue boxes, respectively. Nucleosides in which the phosphoryl transfers are expected to occur on the alcohol moiety of their substituents R are indicated in green (nucleosides D-**HS18.OH.CH$_2$OH-G** and L-**HS18.OH.CH$_2$CH$_2$OH-G** are likely to be phosphorylated at their secondary alcohols).

1st phosphorylation step

Calculations on the first activation step of nucleosides **HS14** and **HS17** revealed that, with the exception of D-**HS17cp-G**, the presence of a fused cyclopropane ring does not alter the binding modes of the nucleosides. The study also proved that the main interactions predicted for D-nucleosides in HSV-1 TK binding site are maintained in the orientations of their L- counterparts, and so the same amino acids are involved in binding both enantiomers (Figure 89).

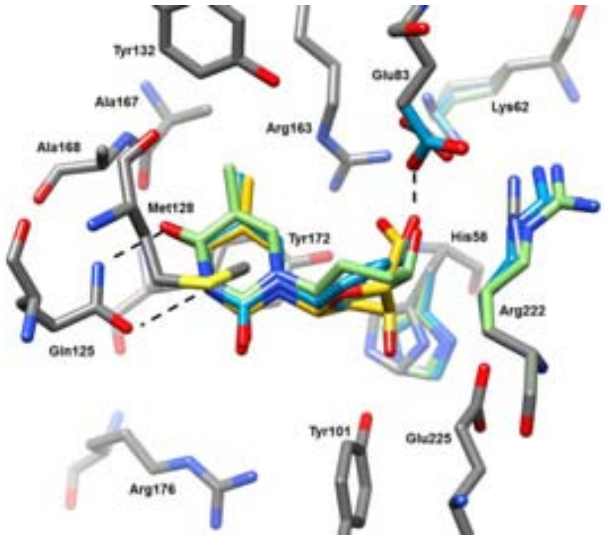


Figure 89. D-**HS14-T** (blue) superimposed to L-**HS14-T** (green) and crystallographic dT (yellow) in HSV-1 TK binding site (PDB code: 1KIM, X-ray residues shown in grey). For the sake of clarity, crystallographic waters and hydrogen atoms are not shown. Hydrogen bonds are exclusively shown for D-**HS14-T** and L-**HS14-T**. Hydrogen bonds are depicted as dotted lines.

The additional methylene spacer of nucleosides **HS17** compared to **HS14** causes different effects depending on the nature of the nucleobase. Regarding pyrimidine nucleosides, the orientations of compounds **HS14** and **HS17** are similar and compatible with the catalysis. By contrast, purine nucleosides **HS14** are not expected to be phosphorylated and the presence of the methylene spacer in the corresponding nucleosides **HS17** is beneficial to the proper binding of these compounds (except D-**HS17cp-G**).

Concerning the addition of substituents to the cyclohexene ring, docking calculations also showed different behaviours between pyrimidine and purine nucleosides. The proposed substituents on pyrimidine L-nucleosides are detrimental to the catalytic phosphoryl transfer, whereas on D-nucleosides they favour the orientations consistent with the catalysis by featuring further interactions with the expected residues in the binding site. Conversely, substituents on purine L-nucleosides are not restrictive for the catalytic activity of the enzyme, while on D-nucleosides there is no clear pattern of which positions could be substituted to enhance the phosphorylation of these compounds.

2nd phosphorylation step

The theoretical study of the second activation step confirmed that the presence of a fused cyclopropane ring instead of a carbon-carbon double bond does not affect the binding modes of the corresponding nucleosides (with the

exception of L-**HS17cp-T**). Regarding D- and L- enantiomers, no differences are observed in purine nucleosides docking results. Conversely, the L- configuration is restrictive for the second phosphoryl transfer on pyrimidine nucleosides, except L-**HS17cp-T**.

The presence of substituents on the cyclohexene ring is not detrimental to the orientations compatible with the catalysis. Indeed, the additional interactions predicted between substituents and residues in the binding site are expected to favour even more these properly posed orientations (Figure 90).

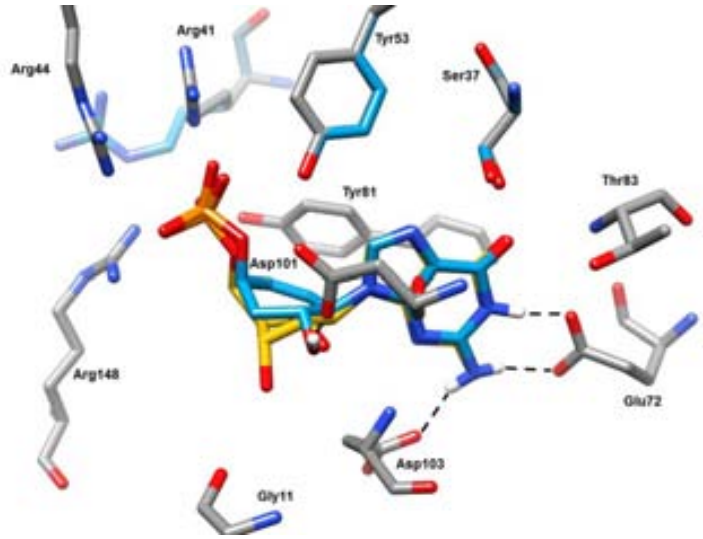


Figure 90. D-**HS15.OH-GMP** (blue) superimposed to crystallographic GMP (yellow) in mGMPK binding site (PDB code: 1LVG, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown. Hydrogen atoms bound to a heteroatom of the ligand and hydrogen bonds are exclusively shown for D-**HS15.OH-GMP**. Hydrogen bonds are depicted as dotted lines.

Remarkably, this second activation step does not represent a bottleneck in the activation process of any purine nucleoside previously predicted to be phosphorylated by HSV-1 TK.

3rd phosphorylation step

The results derived from the study of the last activation step showed that the third phosphoryl transfer is less favoured on nucleosides substituted on the cyclohexene ring (with the exception of D-**HS15.OH-G** and D-**HS16.CH_2OH-T**) than on the corresponding analogues featuring no substituents. Moreover, the substitution of the double bond by a cyclopropane is detrimental to the catalysis on pyrimidine nucleosides **HS14** but not on **HS17**.

4.3. Conclusions

A rational design of cyclohexene and bicyclo[4.1.0]heptane D- and L-nucleoside analogues was carried out by initially proposing a set of 32 novel nucleoside analogues and subsequently evaluating their conversion into the triphosphorylated derivatives by means of molecular docking.

Calculations revealed which skeletons and substituents are more favoured in each activation step. In general terms, dockings demonstrated that the substitution of the carbon-carbon double bond by a fused cyclopropane ring does not improve the activation process of the corresponding nucleosides, except for L-**HS17cp-T**. In fact, the presence of the cyclopropane ring proved even detrimental to the phosphoryl transfer on some nucleosides (e.g. D-**HS17cp-G**). By contrast, the additional methylene spacer between the alcohol and the carbocycle of nucleosides **HS17** compared to **HS14** should not be restrictive for the catalytic activity of the enzymes. Instead, this moiety is beneficial to the activation process of certain nucleosides, namely D-**HS17-G**, D-**HS17cp-T** and L-**HS17cp-T**.

Among the studied compounds, L-nucleosides are less likely to be converted into their triphosphorylated derivatives then their D- counterparts, with the exception of L-**HS17cp-T**.

In summary, the present study indicated that the whole activation process is more favoured for pyrimidine nucleosides D-**HS14-T**, D-**HS16.CH$_2$OH-T**, D-**HS17-T**, D-**HS17cp-T** and L-**HS17cp-T**, as well as for purine nucleosides D-**HS15.OH-G** and D-**HS17-G** (Figure 88). The interaction between these nucleosides and HSV-1 DNA polymerase should be studied to complete this rational design.

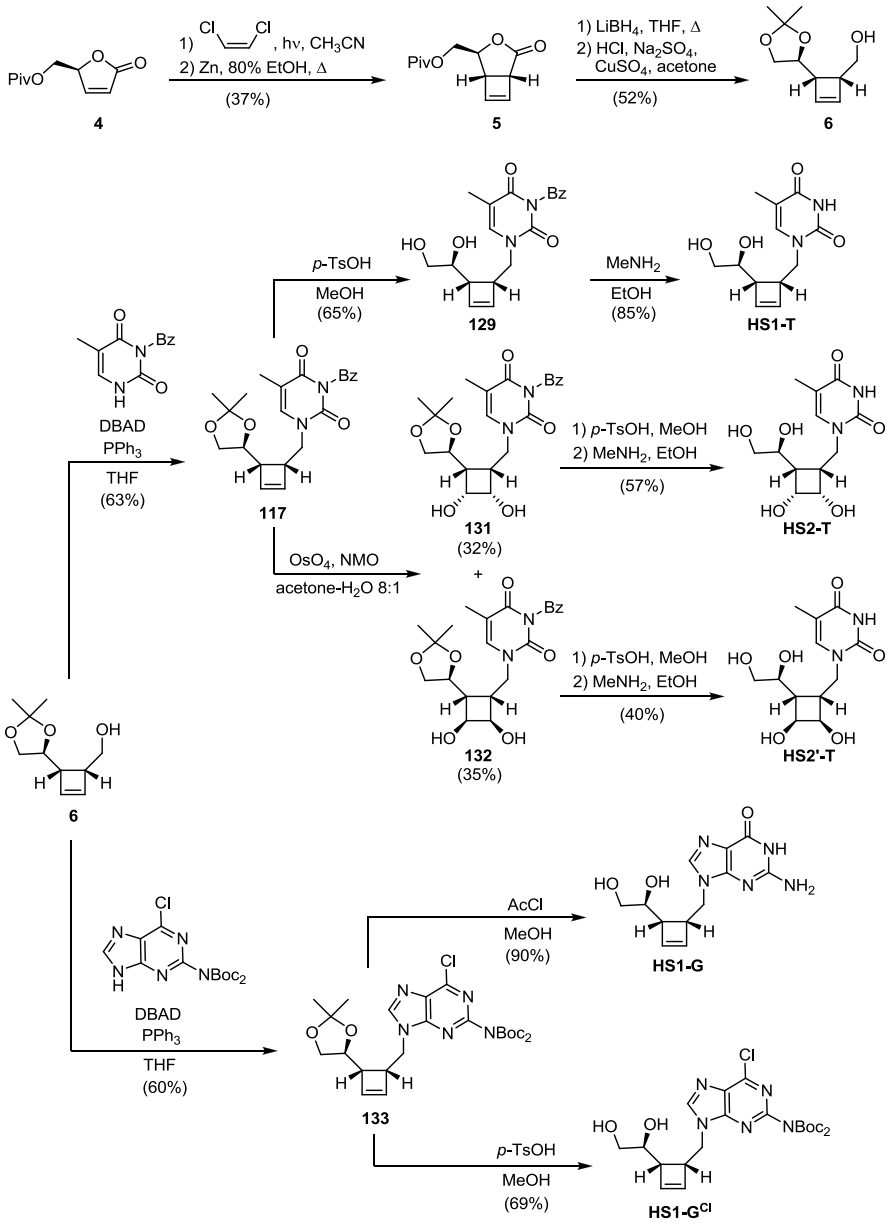
CHAPTER VII

General conclusions

General conclusions

The work presented in this thesis has been focused on the study of nucleoside analogues as antiviral agents, both from the synthetic (Chapter II) and theoretical (Chapters III–VI) points of view.

A new family of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues has been synthesized following a divergent synthetic pathway (Scheme 46). The antiviral activity of these nucleosides has been evaluated without showing significant activity (Chapter II).



Scheme 46. Synthesis of cyclobutene and polyhydroxylated cyclobutane L-nucleoside analogues.

The lack of activity against HIV of cyclobutane-fused nucleoside analogues **HI1**–**HI3** (Figure 91), previously synthesized in our research group, has been investigated by means of molecular modelling (Chapter IV). The theoretical studies on the activation process as well as on the interaction with HIV-1 RT demonstrated that the presence of the cyclobutane fused to the 2'- and 3'-positions of the ribose moiety is detrimental to the adequate binding of these analogues to the viral DNA chain, which accounts for their lack of activity.

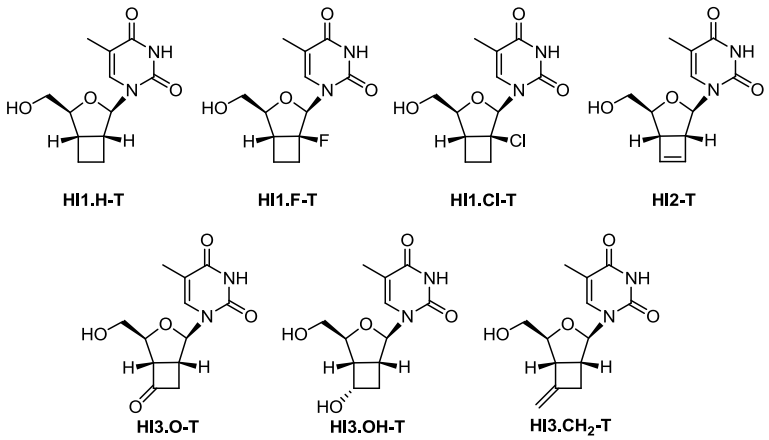


Figure 91. Nucleoside analogues **HI1**–**HI3** synthesized in our research group as anti-HIV agents.

A similar study has been carried out to rationalize the absence of activity against HSV of the newly prepared nucleosides **HS1** and **HS2** (Chapter V). The results derived from our investigations revealed that nucleosides **HS1-G^{Cl}**, **HS2-T** and **HS2'-T** cannot be activated to the required triphosphorylated form. Conversely, the antiviral activity of nucleosides **HS1-T** and **HS1-G** is neither restricted by their drug-likeness nor by their activation process, indicating that a possible limiting factor could be their incorporation into the viral DNA strand.

Finally, a preliminary study for the rational design of carbocyclic nucleoside analogues as anti-HSV agents has been performed (Chapter VI). In particular, the theoretical calculations on the activation process of a set of 32 cyclobutene and cyclobutane L-nucleosides and 32 cyclohexene and bicyclo[4.1.0]heptane nucleosides proved that analogues depicted in Figure 92 are the most prone to be converted into their triphosphorylated derivatives.

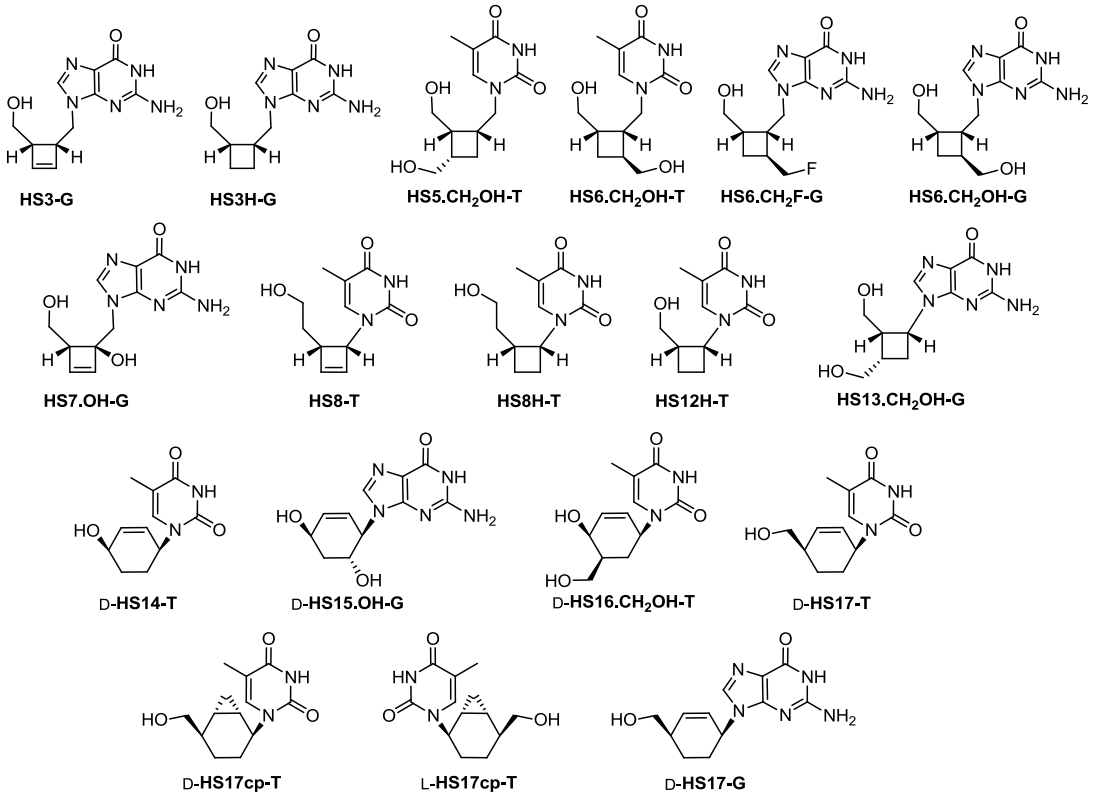


Figure 92. Carbocyclic nucleoside analogues designed as anti-HSV agents.

Overall, the results presented in this thesis highlighted the importance of both organic synthesis and molecular modelling as two intimately intertwined parts in a project aimed at the search for novel compounds with antiviral activity.

EXPERIMENTAL SECTION

1. General remarks

Commercially available reagents were used as received. Solvents were dried by distillation over the appropriate drying reagents. When required, the reactions were performed avoiding moisture by standard procedures and under nitrogen or argon atmosphere. All the reactions were monitored by analytical thin-layer chromatography (TLC) or gas chromatography (GC).

Thin-Layer Chromatography (TLC) was performed using 0.20 mm plates (Alugram Sil G/UV_{254}).

Gas Chromatography (GC) was performed with a *Hewlett Packard 6890* instrument coupled to a *Hewlett Packard 3390A* integrator and with a capillary column coated with a cross-linked dimethyl silicon phase (12 m × 0.2 mm × 0.22 μm). Instrument method used to monitor the reactions was: $T_{injector}$ = 240 ºC, $T_{detector}$ = 300 ºC, T_1 = 160 ºC, t_1 = 2 min, rate = 10 ºC/min, T_2 = 260 ºC.

Flash Column Chromatography was performed under nitrogen pressure using silica gel (230 - 400 mesh). For the purification of alcohol **6**, neutral alumina gel was used (ca. 150 mesh).

Nuclear Magnetic Resonance (NMR) was recorded at the *Servei de Ressonància Magnètica Nuclear* of the *Universitat Autònoma de Barcelona*. ^{1}H-NMR, ^{13}C-NMR, ^{1}H-^{1}H COSY, ^{1}H-^{1}H NOESY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, DEPT135 and selective n.O.e. spectra were recorded using Bruker spectrometers (DRX-250, DPX-360, ARX-400 and AVANCE-II 600 MHz). Chemical shifts (δ) are given in ppm using the residual non-deuterated solvent as internal reference. The abbreviations used to describe signal multiplicities are: s (singlet), d (doublet), t (triplet), q (quadruplet), dd (double doublet), ddd (double double doublet), dddd (double double double doublet), td (triple doublet), tt (triple triplet), m (multiplet), br s (broad singlet) and *J* to indicate the coupling constants (Hz).

Infra-red spectroscopy (IR) spectra were recorded with a Bruker Tensor 27 spectrometer using a Golden Gate ATR module with a diamond window.

High-resolution mass spectra (HRMS) were obtained by direct injection of the sample with electrospray techniques in a Bruker *microTOF-Q* instrument. These

analyses have been performed by the *Servei d'Anàlisi Química (SAQ)* of the *Universitat Autònoma de Barcelona.*

Melting points were determined using a Koffler-Reichert apparatus and were not corrected.

Polarimetry. Specific rotation ($[\alpha]_D$) values were measured at 22 ± 2 ºC and 589.6 nm using a JASCO J-175 polarimeter of SAQ in UAB.

Photochemical reactions were conducted with a 125W high pressure mercury lamp (Cathodeon, HPK 125W) or with a 400W medium pressure mercury lamp (photochemical reactors LTD, 400W MP lamp model 3040) (Figure ES1). The photochemical reactor was equipped with a Quartz or Pyrex refrigeration jacket (Figure ES2).

Figure ES1. Cathodeon HPK 125W lamp.

Figure ES2. Photochemical reactor, refrigeration jacket and photochemical reaction.

Microwave assisted reactions were performed using a *CEM Discover®* Microwave instrument, which operates between 0 and 300 W (Figure ES3). Reactions were conducted in 10 mL glass vessels sealed with a septum. Temperature was measured with an infrared sensor placed under the reaction vessel. All experiments were performed using a stirring option whereby the contents of the vessel were stirred by means of a rotating magnetic plate located below the floor of the microwave cavity and a Telfon-coated magnetic stir bar in the vessel. During the experiments compressed nitrogen was used to cool the reactor when necessary (option PowerMAX™ enabled).

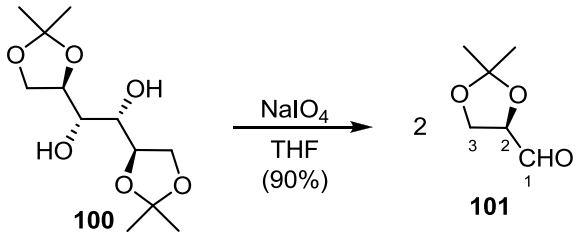
Figure ES3. CEM Discover® microwave instrument.

In this work some of the compounds prepared were already described in the literature. Therefore, only the physical and/or spectroscopic data necessary for their identification are presented.

2. Synthesis of novel cyclobutene and cyclobutane L-nucleoside analogues

2.1. Preparation of (*S*)-5-pivaloyloxymethyl-2(5*H*)-furanone, 4

2.1.1. Preparation of 2,3-*O*-isopropylidene-D-gliceraldehyde, 101

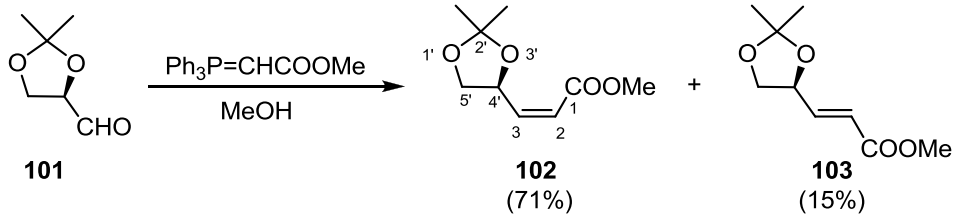


To a solution of 1,2:5,6-di-*O*-isopropylidene-D-mannitol, **100**, (12.00 g, 45.7 mmol) in THF (100 mL), a suspension of sodium periodate (10.80 g, 50.5 mmol) in a mixture of THF (37 mL) and H_2O (17 mL) was slowly added. The resulting white suspension was stirred at room temperature for 2 h. Then, diethyl ether (170 mL) was added and the mixture was stirred for 15 min prior to filtration of a white solid. The solvent was carefully removed under reduced pressure and extracted with CH_2Cl_2 (3 x 25 mL). The organic layer was dried over anhydrous Na_2SO_4 and the solvent was carefully removed under reduced pressure to avoid the loss of aldehyde. Thus, a colourless oil (10.80 g, 83.0 mmol, 90% yield) was obtained and

used in the next reaction without further purification. Variable amount of hydrated aldehyde was observed by NMR and IR.

101: **CF** $C_6H_{10}O_3$; **MW** 130.1 g/mol; **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 9.70 (d, $J_{1,2}$=1.8 Hz, 1H, C*H*O), 4.37 (ddd, $J_{2,3}$=7.4 Hz, $J_{2,3}$=4.8 Hz, $J_{2,1}$=1.8 Hz, 1H, H-2), 4.15 (dd, J_{gem}=8.9 Hz, $J_{3,2}$=7.4 Hz, 1H, H-3), 4.08 (dd, J_{gem}=8.9 Hz, $J_{3,2}$=4.8 Hz, 1H, H-3), 1.46 (s, 3H, CH_3), 1.39 (s, 3H, CH_3); **IR** (film) ν 3400, 2990, 2845, 1740, 1380, 1225, 1220 cm^{-1}.

2.1.2. Preparation of methyl (2*Z*)-3-[(4*S*)-(2,2-dimethyl-1,3-dioxolan-4-yl)]-2-propenoate, 102, and methyl (2*E*)-3-[(4*S*)-(2,2-dimethyl-1,3-dioxolan-4-yl)]-2-propenoate, 103



101 **102** (71%) **103** (15%)

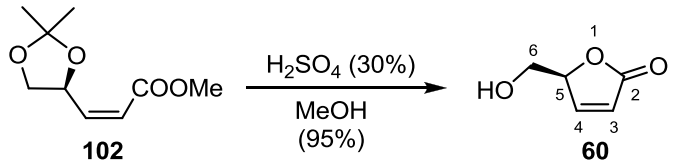
To an ice-cooled solution of aldehyde **101** (10.60 g, 81.4 mmol) in MeOH (75 mL), methoxycarbonylmethylene(triphenyl)phosphorane (27.21 g, 81.4 mmol) was carefully added in small portions. The mixture was allowed to warm to room temperature and stirred for 2 h. Then, the solvent was removed under reduced pressure and the resulting white solid was extracted with hot hexane, filtering the insoluble triphenyl phosphine oxide before hexane cooled down. Evaporation of the filtrate to dryness and purification by column chromatography (hexane-diethyl ether 3:1) afforded the (*Z*)-isomer **102** (10.72 g, 57.6 mmol, 71% yield) as an oil and the (*E*)-isomer **103** (2.23 g, 12.0 mmol, 15% yield) as an oil.

102: **CF** $C_9H_{14}O_4$; **MW** 186.2 g/mol; **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 6.35 (dd, $J_{3,2}$=11.7 Hz, $J_{3,4'}$=6.9 Hz, 1H, H-3), 5.84 (dd, $J_{2,3}$=11.7 Hz, $J_{2,4'}$=1.5 Hz, 1H, H-2), 5.48 (dddd, $J_{4',5'}$=6.9 Hz, $J_{4',5'}$=6.9 Hz, $J_{4',3}$=6.9 Hz, $J_{4',2}$=1.5 Hz, 1H, H-4'), 4.36 (dd, J_{gem}=8.4 Hz, $J_{5',4'}$=6.9 Hz, 1H, H-5'), 3.70 (s, 3H, OCH_3), 3.60 (dd, J_{gem}=8.4 Hz, $J_{5',4'}$=6.9 Hz, 1H, H-5'), 1.44 (s, 3H, CH_3), 1.39 (s, 3H, CH_3); **IR** (film) ν 2989, 2952, 2875, 1723, 1646, 1440, 1402, 1376, 1208 cm^{-1}.

103: **CF** $C_9H_{14}O_4$; **MW** 186.2 g/mol; **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 6.87 (dd, $J_{3,2}$=15.9 Hz, $J_{3,4'}$=5.3 Hz, 1H, H-3), 6.09 (dd, $J_{2,3}$=15.9 Hz, $J_{2,4'}$=1.5 Hz, 1H, H-2),

4.65 (dddd, $J_{4',5'}$=7.0 Hz, $J_{4',5'}$=6.6 Hz, $J_{4',3}$=5.3 Hz, $J_{4',2}$=1.5 Hz, 1H, H-4'), 4.16 (dd, J_{gem}=8.5 Hz, $J_{5',4'}$=6.6 Hz, 1H, H-5'), 3.73 (s, 3H, OCH_3), 3.66 (dd, J_{gem}=8.5 Hz, $J_{5',4'}$=7.0 Hz, 1H, H-5'), 1.43 (s, 3H, CH_3), 1.40 (s, 3H, CH_3); **IR** (film) ν 2989, 2950, 2880, 1727, 1663, 1438, 1264, 1217, 1166, 1125, 1062 cm^{-1}.

2.1.3. Preparation of (*S*)-5-hydroxymethyl-2(5*H*)-furanone, 60

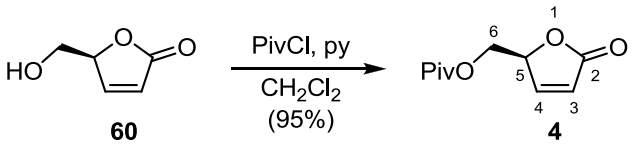


102 **60**

To a solution of **102** (10.72 g, 57.6 mmol) in MeOH (30 mL) was added a 30% aqueous solution of H_2SO_4 (260 µL). The reaction mixture was stirred for 3 h at room temperature and the solvent was removed under reduced pressure. Purification of the crude by column chromatography (EtOAc) gave lactone **60** (6.23 g, 54.6 mmol, 95% yield) as a white solid.

60: **CF** $C_5H_6O_3$; **MW** 114.1 g/mol; **m.p.** 40-41 ºC (pentane-EtOAc); **[α]$_D$** = -151.9 (*c* 2.4, H_2O); **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 7.46 (dd, $J_{4,3}$=5.8 Hz, $J_{4,5}$=1.6 Hz, 1H, H-4), 6.19 (dd, $J_{3,4}$=5.8 Hz, $J_{3,5}$=2.0 Hz, 1H, H-3), 5.13 (dddd, $J_{5,6}$=5.1 Hz, $J_{5,6}$=3.6 Hz, $J_{5,3}$=2.0 Hz, $J_{5,4}$=1.6 Hz, 1H, H-5), 3.98 (ddd, J_{gem}=12.4 Hz, $J_{6,OH}$=6.9 Hz, $J_{6,5}$=3.6 Hz, 1H, H-6), 3.77 (ddd, J_{gem}=12.4 Hz, $J_{6,OH}$=6.9 Hz, $J_{6,5}$=5.1 Hz, 1H, H-6), 2.13 (t, $J_{OH,6}$=6.9 Hz, 1H, -OH); **^{13}C-NMR** (62.5 MHz, $CDCl_3$) δ 173.0 (C=O, C-2), 153.5 (CH, C-4), 123.0 (CH, C-3), 84.0 (CH, C-5), 62.4 (CH_2, C-6); **IR** (KBr) ν 3680-3200, 3107, 2930, 2880, 1743, 1602, 1379, 1332, 1170, 1112 cm^{-1}.

2.1.4. Preparation of (*S*)-5-pivaloyloxymethyl-2(5*H*)-furanone, 4



60 **4**

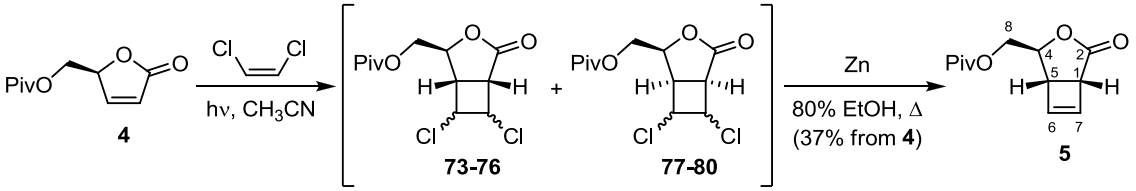
To a stirred solution of lactone **60** (1.72 g, 15.08 mmol) in CH_2Cl_2 (32 mL), pyridine (2.4 mL, 32.34 mmol) was added and the reaction mixture was cooled to 0 ºC. Then, pivaloyl chloride (3.8 mL, 30.27 mmol) was added dropwise and the mixture was stirred overnight, allowing it to warm to room temperature. At this time, water (6 mL) was added and the organic layer was washed successively with 5%

hydrochloric acid (2 x 5 mL), saturated aqueous sodium bicarbonate solution (2 x 5 mL) and brine (1 x 5 mL). The organic layer was dried (Na_2SO_4), concentrated under reduced pressure and purified by column chromatography (hexane-EtOAc 2:1) to furnish **4** (2.85 g, 14.38 mmol, 95% yield) as an oil that solidifies on standing at room temperature.

4: **CF** $C_{10}H_{14}O_4$; **MW** 198.2 g/mol; **m.p.** 32-34 ºC (EtOAc-hexane); **[α]$_D$** = -140.0 (*c* 1.3, $CHCl_3$); **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 7.40 (dd, $J_{4,3}$=5.7 Hz, $J_{4,5}$=1.6 Hz, 1H, H-4), 6.18 (dd, $J_{3,4}$=5.7 Hz, $J_{3,5}$=2.0 Hz, 1H, H-3), 5.20 (m, 1H, H-5), 4.36 (d, $J_{6,5}$=4.1 Hz, 2H, H-6), 1.13 (s, 9H, (CH_3)$_3$C); **^{13}C-NMR** (62.5 MHz, $CDCl_3$) δ 178.0 (C=O, C-2), 171.2 (C=O, Piv), 152.4 (CH, C-4), 123.1 (CH, C-3), 81.0 (CH, C-5), 62.0 (CH_2, C-6), 38.1 (C, (CH_3)$_3$*C*), 27.0 (CH_3, (*C*H$_3$)$_3$C); **IR** (KBr) ν 3088, 3065, 2976, 2928, 2864, 1758, 1730, 1616, 1484, 1330, 1284, 1172, 1157 cm^{-1}.

2.2. Preparation of (1*R*,4*S*,5*S*)-4-pivaloyloxymethyl-3-oxabicyclo[3.2.0]hept-6-en-2-one, 5

Method A (Zn-promoted dihaloelimination with conventional heating)

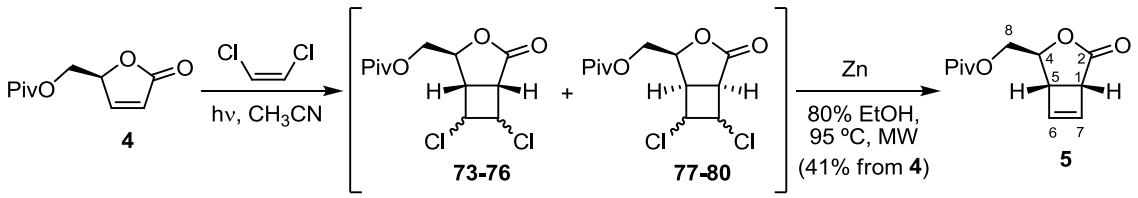


A solution of lactone **4** (1.83 g, 9.25 mmol) and (*Z*)-1,2-dichloroethylene (3.5 mL, 46.39 mmol) in acetonitrile (937 mL) was placed in a photochemical reactor (two-necked 1 L vessel fitted with a Quartz immersion type cooling jacket). The reaction mixture was initially degassed by passage of oxygen-free nitrogen through the solution for 20 min. The reactor was immersed in a cooling bath at -25 ºC and a stream of MeOH at -15 ºC was circulated throughout the refrigeration jacket. The mixture was irradiated using a medium pressure 400W mercury lamp for 5 h. The progress of the reaction was monitored by GC. Using this methodology, four major peaks were observed at retention times of 8.40, 8.55, 8.91 and 10.96 min, along with three minor peaks at 9.01, 9.54 and 11.78 min in a 32:18:13:23:7:5:2 ratio corresponding to **73**, **74**, **75**, **76**, **77**, **79** and **80**, respectively. The crude was then concentrated under reduced pressure, dissolved in hexane-EtOAc 1:1 and passed

through a silica gel pad. The pad was thoroughly washed with hexane-EtOAc 1:1 and the solvent was evaporated under reduced pressure.

The resulting crude was dissolved in 80% aqueous EtOH (45 mL) and activated Zn dust[370] (13.11 g, 200.5 mmol) was added. The mixture was heated up to 105 ºC and it was mechanically stirred at this temperature for 7 h 30 min, monitoring the reaction by GC. Using this methodology one major peak was observed at retention time of 5.76 min, corresponding to compound **5**. After cooling, the reaction mixture was filtered through glass wool and the filtrate was concentrated under reduced pressure. The resulting crude was dissolved in CH_2Cl_2 (75 mL) and washed with 5% hydrochloric acid (75 mL). The aqueous layer was extracted with CH_2Cl_2 (2 x 50 mL). The combined organic extracts were dried (Na_2SO_4), concentrated under reduced pressure and purified by column chromatography (hexane-EtOAc 6:1) to provide **5** (774 mg, 3.45 mmol, 37% yield from **4**) as a colourless oil.

Method B (Zn-promoted dihaloelimination with microwave heating)



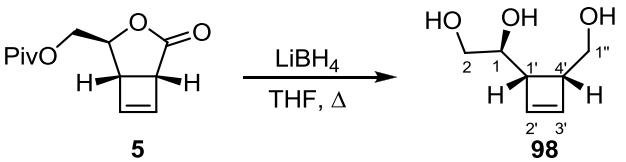
A solution of lactone **4** (700 mg, 3.53 mmol) and (*Z*)-1,2-dichloroethylene (1.4 mL, 17.55 mmol) in acetonitrile (310 mL) was placed in a photochemical reactor (two-necked 330 mL vessel fitted with a Quartz immersion type cooling jacket). The reaction mixture was initially degassed by passage of oxygen-free nitrogen through the solution for 20 min. The reactor was immersed in a cooling bath at -25 ºC and a stream of MeOH at -15 ºC was circulated throughout the refrigeration jacket. The mixture was irradiated using a high pressure 125W mercury lamp for 14 h. The progress of the reaction was monitored by GC. Using this methodology, four major peaks were observed at retention times of 8.40, 8.55, 8.91 and 10.96 min, along with three minor peaks at 9.01, 9.54 and 11.78 min in a 34:20:13:25:4:3:1 ratio corresponding to **73**, **74**, **75**, **76**, **77**, **79** and **80**, respectively. The crude was then concentrated under reduced pressure, dissolved in hexane-EtOAc 1:1 and passed through a silica gel pad. The pad was thoroughly washed with hexane-EtOAc 1:1 and the solvent was removed.

The resulting crude (900 mg) was divided in fractions of 100 mg, following the same procedure for each fraction: a solution of 100 mg of crude in 80% aqueous EtOH (2 mL) was placed in a microwave vessel. Activated Zn dust[370] (883 mg, 13.5 mmol) was added, and the system was sealed and irradiated at 95 ºC for one cycle of 20 minutes and another cycle of 5 minutes, monitoring the reaction by GC. Using this methodology one major peak was observed at retention time of 5.76 min, corresponding to compound **5**. The reaction mixture was then filtered over a pad of Celite® (EtOAc). The crudes resulting from each irradiation were combined and purified by column chromatography (hexane-EtOAc from 10:1 to 3:1) to provide **5** (330 mg, 1.47 mmol, 41% yield from **4**) as a colourless oil.

5: **CF** $C_{12}H_{16}O_4$; **MW** 224.3 g/mol; **[α]$_D$** = -211.9 (*c* 1.1, $CHCl_3$); **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 6.35 (ddd, $J_{6,7}$=2.7 Hz, $J_{6,5}$=0.7 Hz, $J_{6,4}$=0.5 Hz, 1H, H-6), 6.30 (dd, $J_{7,6}$=2.7 Hz, $J_{7,1}$=0.8 Hz, 1H, H-7), 4.60 (dddd, $J_{4,8}$=3.0 Hz, $J_{4,8}$=2.8 Hz, $J_{4,5}$=1.5 Hz, $J_{4,6}$=0.5 Hz, 1H, H-4), 4.26 (dd, J_{gem}=12.0 Hz, $J_{8,4}$=2.8 Hz, 1H, H-8), 4.12 (dd, J_{gem}=12.0 Hz, $J_{8,4}$=3.0 Hz, 1H, H-8), 3.70 (dd, $J_{1,5}$=3.5 Hz, $J_{1,7}$=0.8 Hz, 1H, H-1), 3.45 (ddd, $J_{5,1}$=3.5 Hz, $J_{5,4}$=1.5 Hz, $J_{5,6}$=0.7 Hz, 1H, H-5), 1.20 (s, 9H, $(CH_3)_3C$); **^{13}C-RMN** (62.5 MHz, $CDCl_3$) δ 178.1 (C=O, C-2), 174.9 (C=O, Piv), 140.6 (CH, C-6/C-7), 139.3 (CH, C-6/C-7), 76.2 (CH, C-4), 66.0 (CH_2, C-8), 47.7 (CH, C-1), 44.3 (CH, C-5), 39.0 (C, $(CH_3)_3C$), 27.3 (CH_3, $(CH_3)_3C$); **IR** (film) ν 2975, 2874, 1768, 1734, 1484, 1284 cm^{-1}.

2.3. Preparation of {(1*R*,4*S*)-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-2-en-1-yl}methanol, 6

2.3.1. Preparation of (1*S*)-1-[(1*S*,4*R*)-4-(hydroxymethyl)cyclobut-2-en-1-yl]ethane-1,2-diol, 98



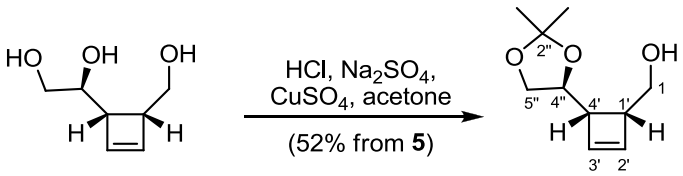
5 **98**

To a solution of **5** (462 mg, 2.06 mmol) in dry THF (34 mL), a 2.0 M solution of $LiBH_4$ in THF (5 mL, 10.00 mmol) was added dropwise under argon atmosphere. The mixture was heated up to 90 ºC and stirred at this temperature for 4 h. Then, the reaction was allowed to cool to room temperature and quenched with

$Na_2SO_4{\cdot}10H_2O$ until no bubbling was observed. The suspension was stirred overnight and then filtered through a Celite pad, thoroughly washing the pad with ethyl acetate. The filtrate was concentrated under reduced pressure and used without further purification for the following step. To characterize the product, a fraction of the crude was purified by column chromatography (from hexane-EtOAc 1:1 to CH_2Cl_2-MeOH 9:1) to afford pure **98**.

98: **CF** $C_7H_{12}O_3$; **MW** 144.2 g/mol; **[α]D** = +22.7 (*c* 1.5, $CHCl_3$); **^{1}H-NMR** (360 MHz, $CDCl_3$) δ 6.04 (dd, $J_{2',3'}$=2.9 Hz, $J_{2',1'}$=0.8 Hz, 1H, H-2'), 6.02 (dd, $J_{3',2'}$=2.9 Hz, $J_{3',4'}$=0.8 Hz, 1H, H-3'), 4.65 (br s, 1H, -OH), 4.05 (br s, 1H, -OH), 3.80 (m, 4H, 2xH-1'', H-1, H-2), 3.53 (dd, J_{gem}=11.3 Hz, $J_{2,1}$=6.9 Hz, 1H, H-2), 3.25 (dddd, $J_{4',1''}$=11.5 Hz, $J_{4',1''}$=4.0 Hz, $J_{4',1'}$=4.0 Hz, $J_{4',3'}$=0.8 Hz, 1H, H-4'), 3.02 (ddd, $J_{1',1}$=10.6 Hz, $J_{1',4'}$=4.0 Hz, $J_{1',2'}$=0.8 Hz, 1H, H-1'), 2.18 (br s, 1H, -OH); **^{13}C-NMR** (90 MHz, $CDCl_3$) δ 137.8 (CH, C-2'/C-3'), 137.2 (CH, C-2'/C-3'), 72.3 (CH, C-1), 65.1 (CH_2, C-2), 62.4 (CH_2, C-1''), 48.6 (CH, C-1'/C-4'), 48.1 (CH, C-1'/C-4'); **IR** (ATR) ν 3293 (br), 2885, 1648, 1427, 1288, 1152, 1087, 1062, 1015, 870, 726 cm^{-1}; **HRMS** (ESI+) calculated for $[C_7H_{12}O_3+Na]^+$ 167.0679, found 167.0676.

2.3.2. Preparation of {(1*R*,4*S*)-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-2-en-1-yl}methanol, 6



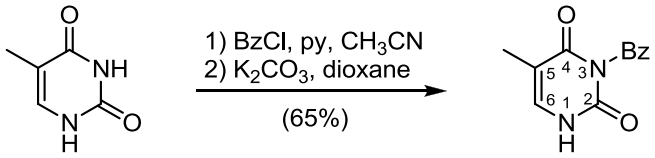
98 **6**

To a stirred solution of the previously obtained crude in acetone (36 mL), anhydrous sodium sulfate (6.70 g, 47.20 mmol), anhydrous copper sulfate (2.50 g, 15.68 mmol) and one drop of conc. HCl were added under argon atmosphere. After being stirred for 4 days at room temperature, the reaction was quenched with 30% NH_3 and filtered through a Celite pad. The solvent was thoroughly evaporated under reduced pressure and without heating (due to the volatility of alcohol **6**), and the crude material was purified by column chromatography over alumina (from hexane-diethyl ether 3:1 to diethyl ether) to furnish **6** (199 mg, 1.08 mmol, 52% overall yield from **5**) as a colourless oil.

6: **CF** $C_{10}H_{16}O_3$; **MW** 184.2 g/mol; $[\alpha]_D$ = +16.3 (*c* 0.98, $CHCl_3$); **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 6.09 (dd, $J_{2',3'}$=2.9 Hz, $J_{2',1'}$=1.1 Hz, 1H, H-2'), 5.93 (ddd, $J_{3',2'}$=2.9 Hz, $J_{3',4'}$=0.9 Hz, $J_{3',1'}$=0.4 Hz, 1H, H-3'), 4.18 (ddd, $J_{4'',4'}$=10.4 Hz, $J_{4'',5''}$=6.0 Hz, $J_{4'',5''}$=6.0 Hz, 1H, H-4''), 4.11 (dd, J_{gem}=7.9 Hz, $J_{5'',4''}$=6.0 Hz, 1H, H-5''), 3.71 (m, 3H, 2xH-1, H-5''), 3.45 (dd, $J_{OH,1}$=9.1 Hz, $J_{OH,1}$=4.1 Hz, 1H, -OH), 3.26 (m, 1H, H-1'), 3.02 (ddd, $J_{4',4''}$=10.4 Hz, $J_{4',1'}$=4.1 Hz, $J_{4',3'}$=0.9 Hz, 1H, H-4'), 1.43 (s, 3H, CH_3-C-2''), 1.37 (s, 3H, CH_3-C-2''); **^{13}C-NMR** (62.5 MHz, $CDCl_3$) δ 138.9 (CH, C-2'), 136.0 (CH, C-3'), 109.5 (C, C-2''), 75.9 (CH, C-4''), 68.5 (CH_2, C-5''), 62.0 (CH_2, C-1), 49.0 (CH, C-1'/C-4'), 48.7 (CH, C-1'/C-4'), 26.8 (CH_3, *C*H_3-C-2''), 25.6 (CH_3, *C*H_3-C-2''); **IR** (ATR) ν 3452 (br), 1370, 1212, 1150, 1058, 1019, 846, 745 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{10}H_{16}O_3+Na]^+$ 207.0992, found 207.0995. DEPT135, HMBC and COSY were recorded for **6**.

2.4. Preparation of nucleoside analogues featuring a pyrimidine base

2.4.1. Preparation of *N*3-benzoylthymine

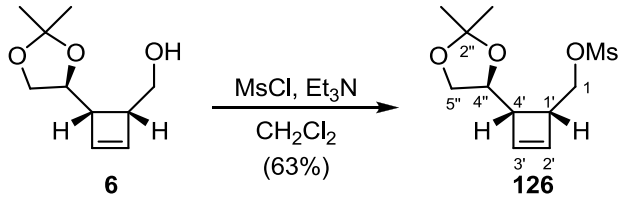


To a solution of thymine (3.00 g, 23.79 mmol) in dry acetonitrile (24 mL), anhydrous pyridine (9.5 mL) was added and the mixture was cooled to 0 ºC in an ice bath. Then, benzoyl chloride (6.2 mL, 52.57 mmol) was added dropwise. The solution was allowed to stir overnight at room temperature. Then, the solvent was removed under reduced pressure and the residue was partitioned between CH_2Cl_2 (120 mL) and H_2O (120 mL). The organic layer was separated and evaporated under reduced pressure. The residue was dissolved in a mixture of dioxane (50 mL) and a 0.5 M solution of K_2CO_3 (25 mL) and stirred for 30 min at room temperature. Then, the suspension was acidified to pH 5 by the addition of glacial acetic acid. The products were concentrated under reduced pressure and the residue was stirred with a saturated $NaHCO_3$ solution. After 1 h, the residue was filtered and washed with cold H_2O (3 x 12 mL). Finally, the product obtained was crystallized from aqueous CH_3CN to give ***N*3-benzoylthymine** (1.95 g, 8.47 mmol, 65% yield) as a white solid.

*N*3-benzoylthymine: **CF** $C_{12}H_{10}N_2O_3$; **MW** 230.2 g/mol; **^{1}H-NMR** (400 MHz, $CDCl_3$) δ 8.01 (br s, 1H, -NH), 7.75 (dd, *J*=8.0 Hz, *J*=1.2 Hz, 2H, H-Bz), 7.68 (q, $J_{6,CH3}$=1.3 Hz, 1H, H-6), 7.63 (m, 1H, H-Bz), 7.48 (dd, *J*=8.0 Hz, *J*=8.0 Hz, 2H, H-Bz), 2.03 (d, $J_{CH3,6}$=1.3 Hz, 3H, C*H*$_3$-C-5).

2.4.2. Nucleophilic substitution of the mesylate derivative of alcohol 6

2.4.2.1. Preparation of {(1*R*,4*S*)-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-2-en-1-yl}methyl methanesulfonate, 126



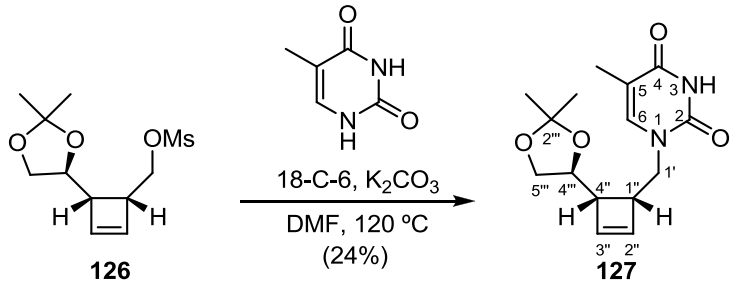
6 **126**

To a solution of **6** (39 mg, 0.21 mmol) in dry CH_2Cl_2 (2.4 mL), anhydrous triethylamine (59 µL, 0.42 mmol) and mesyl chloride (28 µL, 0.36 mmol) were added under argon atmosphere. After being stirred for 3 h at room temperature, the volatiles were removed under vacuum and the resulting oil was purified by column chromatography (from hexane-diethyl ether 3:1 to diethyl ether) to provide **126** (35 mg, 0.13 mmol, 63% yield) as a yellow oil.

126: **CF** $C_{11}H_{18}O_5S$; **MW** 262.3 g/mol; **$[\alpha]_D$** = -115.5 (*c* 1.75, $CHCl_3$); **^{1}H-NMR** (360 MHz, $CDCl_3$) δ 6.27 (dd, $J_{2',3'}$=2.8 Hz, $J_{2',1'}$=0.8 Hz, 1H, H-2'), 6.00 (d, $J_{3',2'}$=2.8 Hz, 1H, H-3'), 4.55 (dd, J_{gem}=9.7 Hz, $J_{1,1'}$=5.7 Hz, 1H, H-1), 4.32 (dd, J_{gem}=9.7 Hz, $J_{1,1'}$=8.6 Hz, 1H, H-1), 4.07 (m, 2H, H-5'', H-4''), 3.64 (m, 1H, H-5''), 3.34 (m, 1H, H-1'), 3.09 (m, 1H, H-4'), 3.03 (s, 3H, C*H*$_3$-SO_2), 1.41 (s, 3H, C*H*$_3$-C-2''), 1.33 (s, 3H, C*H*$_3$-C-2''); **^{13}C-NMR** (90 MHz, $CDCl_3$) δ 138.9 (CH, C-2'), 137.1 (CH, C-3'), 109.5 (C, C-2''), 75.7 (CH, C-4''), 70.1 (CH_2, C-1), 68.6 (CH_2, C-5''), 49.2 (CH, C-4'), 44.3 (CH, C-1'), 37.5 (CH_3, *C*H_3-SO_2), 27.1 (CH_3, *C*H_3-C-2''), 25.7 (CH_3, *C*H_3-C-2''); **IR** (ATR) ν 2985, 2936, 1352, 1172, 1060, 974, 945, 845, 810, 748 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{11}H_{18}O_5S+Na]^+$ 285.0767, found 285.0768. DEPT135, HSQC, HMBC and COSY were recorded for **126**.

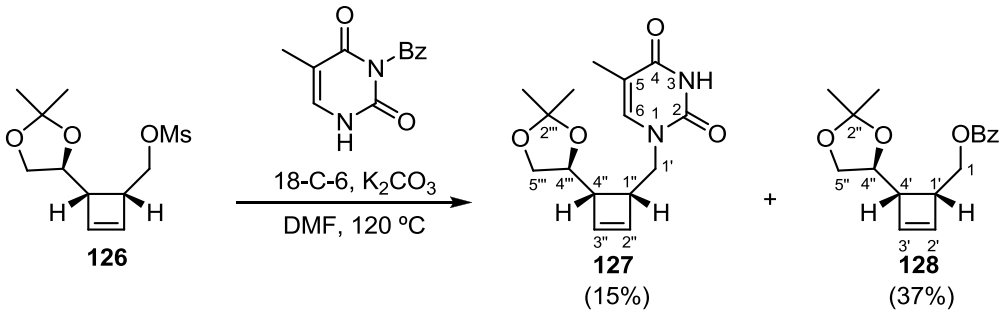
2.4.2.2. Preparation of 1-({(1*R*,4*S*)-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-2-en-1-yl}methyl)thymine, 127

Method A



A solution of the mesylate **126** (33 mg, 0.12 mmol), K_2CO_3 (52 mg, 0.38 mmol), 18-crown-6 (66 mg, 0.25 mmol), and thymine (35 mg, 0.28 mmol) in dry DMF (2.6 mL) was stirred overnight at 120 ºC. The mixture was then cooled to room temperature and concentrated under reduced pressure. The resulting crude was purified by column chromatography (from hexane-EtOAc 1:1 to CH_2Cl_2-MeOH 15:1), leading to **127** (9 mg, 0.03 mmol, 24% yield) as a colourless oil.

Method B



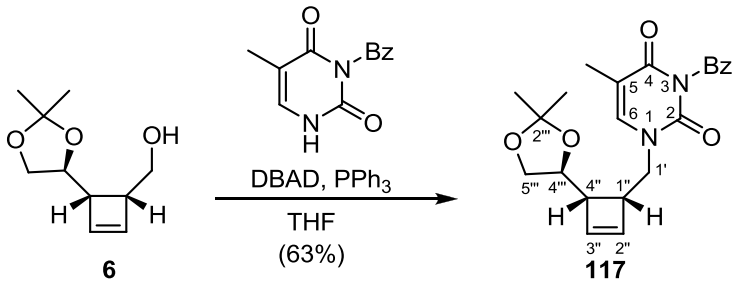
A solution of the mesylate **126** (33 mg, 0.12 mmol), K_2CO_3 (53 mg, 0.38 mmol), 18-crown-6 (70 mg, 0.26 mmol), and *N3*-benzoylthymine (59 mg, 0.26 mmol) in dry DMF (2.5 mL) was stirred for 4 h at 120 ºC. The mixture was then cooled to room temperature and concentrated under reduced pressure. The resulting crude was purified by column chromatography (from hexane-EtOAc 2:1 to EtOAc), furnishing **127** (6 mg, 0.02 mmol, 15% yield) and **128** (14 mg, 0.05 mmol, 37% yield) as colourless oils.

127: **CF** $C_{15}H_{20}N_2O_4$; **MW** 292.3 g/mol; **^{1}H-NMR** (360 MHz, $CDCl_3$) δ 8.32 (1H, -NH), 7.14 (q, $J_{6,CH3}$=0.9 Hz, 1H, H-6), 6.21 (d, $J_{2'',3''}$=2.8 Hz, 1H, H-2''), 5.96 (d, $J_{3'',2''}$=2.8 Hz, 1H, H-3''), 4.11 (m, 2H, H-4''', H-5'''), 3.97 (dd, J_{gem}=14.0 Hz,

$J_{1',1''}$=6.2 Hz, 1H, H-1'), 3.91 (dd, J_{gem}=14.0 Hz, $J_{1',1''}$=9.0 Hz, 1H, H-1'), 3.66 (m, 1H, H-5'''), 3.30 (m, 1H, H-1''), 3.07 (dd, *J*=9.7 Hz, *J*=4.3 Hz, 1H, H-4''), 1.92 (d, $J_{CH3,6}$=0.9 Hz, 3H, C*H*$_3$-C-5), 1.40 (s, 3H, C*H*$_3$-C-2''), 1.35 (s, 3H, C*H*$_3$-C-2''); **^{13}C-NMR** (90 MHz, $CDCl_3$) δ 164.0 (C=O, C-4), 150.9 (C=O, C-2), 141.1 (CH, C-6), 139.9 (CH, C-2''), 136.4 (CH, C-3''), 110.4 (C, C-5), 109.7 (C, C-2'''), 76.0 (CH, C-4'''), 68.7 (CH_2, C-5'''), 49.6 (CH, C-4''), 48.6 (CH_2, C-1'), 45.1 (CH, C-1''), 27.1 (CH_3, *C*H$_3$-C-2'''), 25.8 (CH_3, *C*H$_3$-C-2'''), 12.5 (CH_3, *C*H$_3$-C-5); **IR** (ATR) ν 3179, 1673, 1467, 1369, 1243, 1217, 1061, 848, 731 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{15}H_{20}N_2O_4+Na]^+$ 315.1315, found 315.1314. DEPT135, HSQC, HMBC and COSY were recorded for **127**.

128: **CF** $C_{17}H_{20}O_4$; **MW** 288.3 g/mol; **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 8.07 (dd, *J*=8.0 Hz, *J*=1.4 Hz, 2H, H-Bz), 7.56 (m, 1H, H-Bz), 7.44 (dd, *J*=8.0 Hz, *J*=8.0 Hz, 2H, H-Bz), 6.26 (d, $J_{2',3'}$=2.8 Hz, 1H, H-2'), 6.01 (d, $J_{3',2'}$=2.8 Hz, 1H, H-3'), 4.64 (dd, J_{gem}=11.3 Hz, $J_{1,1'}$=7.0 Hz, 1H, H-1), 4.46 (dd, J_{gem}=11.3 Hz, $J_{1,1'}$=7.0 Hz, 1H, H-1), 4.21 (ddd, $J_{4'',4'}$=9.3 Hz, $J_{4'',5''}$=7.4 Hz, $J_{4'',5''}$=6.1 Hz, 1H, H-4''), 4.07 (dd, J_{gem}=7.4 Hz, $J_{5'',4''}$=6.1 Hz, 1H, H-5''), 3.64 (dd, J_{gem}=7.4 Hz, $J_{5'',4''}$=7.4 Hz, 1H, H-5''), 3.43 (td, $J_{1',1}$=7.0 Hz, $J_{1',4'}$=4.3 Hz, 1H, H-1'), 3.14 (dd, $J_{4',4''}$=9.3 Hz, $J_{4',1'}$=4.3 Hz, 1H, H-4'), 1.55 (s, 3H, C*H*$_3$-C-2''), 1.37 (s, 3H, C*H*$_3$-C-2''); **^{13}C-NMR** (62.5 MHz, $CDCl_3$) δ 165.9 (C=O, Bz), 139.4 (CH, C-2'), 136.6 (CH, C-3'), 133.0 (CH, Bz), 130.6 (C, Bz), 129.8 (CH, Bz), 128.4 (CH, Bz), 109.3 (C, C-2''), 75.7 (CH, C-4''), 68.7 (CH_2, C-5''/C-1), 64.5 (CH_2, C-5''/C-1), 49.3 (CH, C-4'), 44.8 (CH, C-1'), 27.0 (CH_3, *C*H$_3$-C-2''), 25.6 (CH_3, *C*H$_3$-C-2''). DEPT135 was recorded for **128**.

2.4.3. Mitsunobu coupling of *N*3-benzoylthymine to alcohol 6. Preparation of 3-benzoyl-1-({(1*R*,4*S*)-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-2-en-1-yl}methyl)thymine, 117



A solution of DBAD (368 mg, 1.60 mmol) in anhydrous THF (4.5 mL) was added dropwise to a 0 ºC stirred suspension of alcohol **6** (147 mg, 0.80 mmol), *N3*-

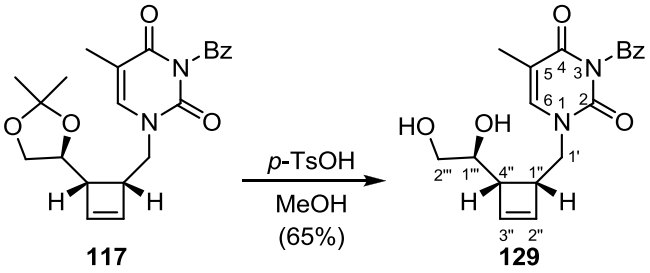
benzoylthymine (367 mg, 1.59 mmol) and triphenylphosphine (421 mg, 1.60 mmol) in anhydrous THF (5.7 mL) under argon atmosphere. The mixture was allowed to warm to room temperature and stirred overnight. The organic solvent was removed under vacuum and the resulting oil was purified by repeated column chromatography (hexane-EtOAc from 10:1 to 1:1) to give **117** (201 mg, 0.51 mmol, 63% yield) as a colourless oil.

117: **CF** $C_{22}H_{24}N_2O_5$; **MW** 396.4 g/mol; **[α]$_D$** = +3.8 (*c* 1.0, $CHCl_3$); **^{1}H-NMR** (360 MHz, $CDCl_3$) δ 7.91 (dd, *J*=8.2 Hz, *J*=1.2 Hz, 2H, H-Bz), 7.63 (m, 1H, H-Bz), 7.48 (dd, *J*=8.2 Hz, *J*=8.2 Hz, 2H, H-Bz), 7.26 (q, $J_{6,CH3}$=1.1 Hz, 1H, H-6), 6.20 (dd, $J_{2'',3''}$=2.9 Hz, $J_{2'',1''}$=0.9 Hz, 1H, H-2''), 5.97 (d, $J_{3'',2''}$=2.9 Hz, 1H, H-3''), 4.09 (m, 2H, H-5''', H-4'''), 3.99 (m, 2H, 2xH-1'), 3.66 (m, 1H, H-5'''), 3.33 (m, 1H, H-1''), 3.07 (ddd, *J*=9.8 Hz, *J*=4.2 Hz, *J*=0.8 Hz, 1H, H-4''), 1.96 (d, $J_{CH3,6}$=1.1 Hz, 3H, C*H*$_3$-C-5), 1.41 (s, 3H, C*H*$_3$-C-2'''), 1.35 (s, 3H, C*H*$_3$-C-2'''); **^{13}C-NMR** (90 MHz, $CDCl_3$) δ 169.2 (C=O, Bz), 163.3 (C=O, C-4), 150.1 (C=O, C-2), 140.8 (CH, C-6), 139.9 (CH, C-2''), 136.4 (CH, C-3''), 135.0 (CH, Bz), 131.9 (C, Bz), 130.5 (CH, Bz), 129.2 (CH, Bz), 110.4 (C, C-5), 109.7 (C, C-2'''), 76.0 (CH, C-4'''), 68.7 (CH_2, C-5'''), 49.6 (CH, C-4''), 49.0 (CH_2, C-1'), 45.0 (CH, C-1''), 27.1 (CH_3, *C*H$_3$-C-2'''), 25.8 (CH_3, *C*H$_3$-C-2'''), 12.5 (CH_3, *C*H$_3$-C-5); **IR** (ATR) ν 2984, 2930, 1745, 1696, 1648, 1436, 1238, 1062, 747 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{22}H_{24}N_2O_5+Na]^+$ 419.1577, found 419.1585. DEPT135, HSQC, HMBC and COSY were recorded for **117**.

2.4.4. Synthesis of the cyclobutene nucleoside analogue HS1-T

2.4.4.1. Preparation of 3-benzoyl-1-({(1*R*,4*S*)-4-[(1*S*)-1,2-dihydroxyethyl]cyclobut-2-en-1-yl}methyl)thymine, 129

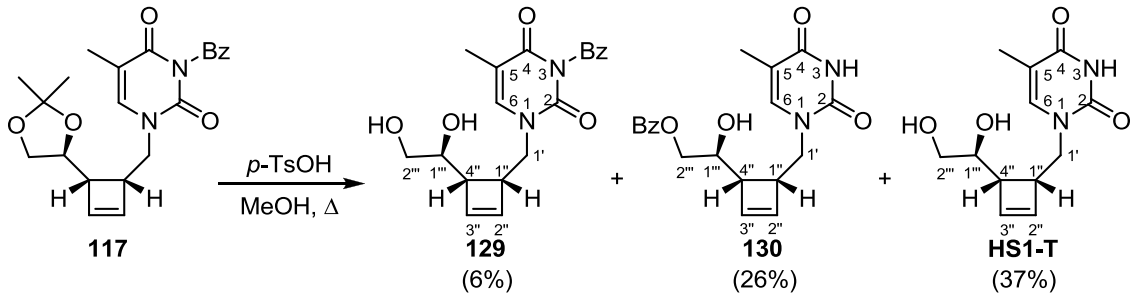
Method A (room temperature)



To a solution of **117** (52 mg, 0.13 mmol) in MeOH (5.8 mL), *p*-toluenesulfonic acid monohydrate (38 mg, 0.20 mmol) was added and the resulting mixture was

stirred for 28 h at room temperature. After removal of the solvent, the residue was purified by column chromatography (from hexane-EtOAc 3:1 to EtOAc) to furnish **129** (30 mg, 0.08 mmol, 65% yield) as a colourless oil.

Method B (refluxing MeOH)



To a solution of **117** (34 mg, 0.09 mmol) in MeOH (3.8 mL), *p*-toluenesulfonic acid monohydrate (25 mg, 0.13 mmol) was added and the resulting mixture was refluxed for 7 h and then stirred overnight at room temperature. After removal of the solvent, the residue was purified by column chromatography (from hexane-EtOAc 3:1 to EtOAc-MeOH 20:1) to afford compound **129** as a colourless oil (2 mg, 0.006 mmol, 6% yield), **130** as a colourless oil (8 mg, 0.02 mmol, 26% yield) and **HS1-T** as a white solid (8 mg, 0.03 mmol, 37% yield).

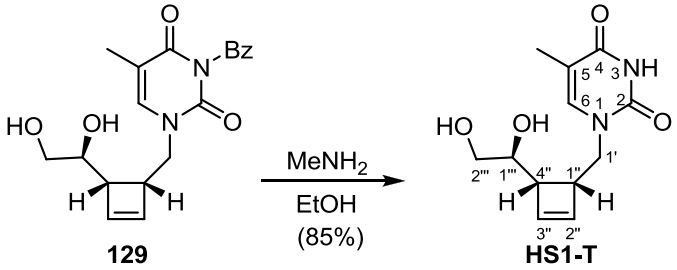
129: **CF** $C_{19}H_{20}N_2O_5$; **MW** 356.4 g/mol; $[\alpha]_D$ = -28.0 (*c* 1.1, $CHCl_3$); **^{1}H-NMR** (400 MHz, $CDCl_3$) δ 7.90 (dd, *J*=8.3 Hz, *J*=1.0 Hz, 2H, H-Bz), 7.63 (m, 1H, H-Bz), 7.48 (dd, *J*=8.3 Hz, *J*=8.3 Hz, 2H, H-Bz), 7.20 (q, $J_{6,CH3}$=0.8 Hz, 1H, H-6), 6.14 (d, $J_{2'',3''}$=2.7 Hz, 1H, H-2''), 5.99 (d, $J_{3'',2''}$=2.7 Hz, 1H, H-3''), 4.21 (dd, J_{gem}=14.0 Hz, $J_{1',1''}$=5.9 Hz, 1H, H-1'), 3.80 (dd, J_{gem}=14.0 Hz, $J_{1',1''}$=9.0 Hz, 1H, H-1'), 3.65 (m, 2H, H-2''', H-1'''), 3.41 (dd, J_{gem}=11.2 Hz, $J_{2''',1'''}$=7.2 Hz, 1H, H-2'''), 3.28 (m, 1H, H-1''), 2.98 (dd, *J*=10.3 Hz, *J*=4.1 Hz, 1H, H-4''), 1.93 (d, $J_{CH3,6}$=0.8 Hz, 3H, C*H*$_3$-C-5); **^{13}C-NMR** (100 MHz, $CDCl_3$) δ 169.4 (C=O, Bz), 163.3 (C=O, C-4), 150.4 (C=O, C-2), 140.7 (CH, C-6), 138.9 (CH, C-2''), 137.2 (CH, C-3''), 135.2 (CH, Bz), 131.7 (C, Bz), 130.5 (CH, Bz), 129.3 (CH, Bz), 110.9 (C, C-5), 72.1 (CH, C-1'''), 65.6 (CH_2, C-2'''), 49.1 (CH_2, C-1'), 48.5 (CH, C-4''), 45.0 (CH, C-1''), 12.5 (CH_3, *C*H_3-C-5); **IR** (ATR) ν 3389 (br), 2923, 2853, 1741, 1691, 1639, 1259, 1021, 798 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{19}H_{20}N_2O_5+Na]^+$ 379.1264, found 379.1264. DEPT135, HSQC, HMBC and COSY were recorded for **129**.

130: **CF** $C_{19}H_{20}N_2O_5$; **MW** 356.4 g/mol; **^{1}H-NMR** (360 MHz, $CDCl_3$) δ 8.62 (br s, 1H, -NH), 8.06 (d, *J*=7.4 Hz, 2H, H-Bz), 7.59 (t, *J*=7.4 Hz, 1H, H-Bz), 7.46 (dd, *J*=7.4

Hz, *J*=7.4 Hz, 2H, H-Bz), 7.08 (q, $J_{6,CH3}$=0.9 Hz, 1H, H-6), 6.22 (d, $J_{2'',3''}$=2.6 Hz, 1H, H-2''), 6.13 (d, $J_{3'',2''}$=2.6 Hz, 1H, H-3''), 4.50 (dd, J_{gem}=11.6 Hz, $J_{2''',1'''}$=2.8 Hz, 1H, H-2'''), 4.35 (dd, J_{gem}=11.6 Hz, $J_{2''',1'''}$=6.2 Hz, 1H, H-2'''), 4.22 (dd, J_{gem}=14.0 Hz, $J_{1',1''}$=5.6 Hz, 1H, H-1'), 4.03 (m, 1H, H-1'''), 3.84 (dd, J_{gem}=14.0 Hz, $J_{1',1''}$=9.4 Hz, 1H, H-1'), 3.34 (m, 1H, H-1''), 3.16 (dd, *J*=10.5 Hz, *J*=4.1 Hz, 1H, H-4''), 2.91 (br s, 1H, -OH), 1.92 (d, $J_{CH3,6}$=0.9 Hz, 3H, C*H*$_3$-C-5); **^{13}C-NMR** (90 MHz, $CDCl_3$) δ 167.0 (C=O, Bz), 164.1 (C=O, C-4), 151.2 (C=O, C-2), 140.7 (CH, C-6), 139.6 (CH, C-2''), 136.8 (CH, C-3''), 133.5 (CH, Bz), 129.9 (CH, Bz), 129.7 (C, Bz), 128.7 (CH, Bz), 110.9 (C, C-5), 70.7 (CH, C-1'''), 68.4 (CH_2, C-2'''), 48.8 (CH_2, C-1'), 48.8 (CH, C-4''), 45.2 (CH, C-1''), 12.5 (CH_3, *C*H_3-C-5). DEPT135, HSQC, HMBC and COSY were recorded for **130**.

HS1-T: **CF** $C_{12}H_{16}N_2O_4$; **MW** 252.3 g/mol; **m.p.** 180-185 ºC (MeOH); **[α]$_D$** = +249.9 (*c* 0.6, MeOH); **^{1}H-NMR** (360 MHz, CD_3OD) δ 7.46 (q, $J_{6,CH3}$=1.1 Hz, 1H, H-6), 6.17 (d, $J_{2'',3''}$=2.7 Hz, 1H, H-2''), 6.12 (dd, $J_{3'',2''}$=2.7 Hz, $J_{3'',4''}$=0.5 Hz, 1H, H-3''), 4.19 (dd, J_{gem}=13.7 Hz, $J_{1',1''}$=5.1 Hz, 1H, H-1'), 3.84 (dd, J_{gem}=13.7 Hz, $J_{1',1''}$=10.6 Hz, 1H, H-1'), 3.65 (m, 2H, H-1''', H-2'''), 3.49 (dd, J_{gem}=11.8 Hz, $J_{2''',1'''}$=6.6 Hz, 1H, H-2'''), 3.31 (m, 1H, H-1''), 3.03 (ddd, *J*=9.7 Hz, *J*=4.2 Hz, *J*=0.5 Hz, 1H, H-4''), 1.88 (d, $J_{CH3,6}$=1.1 Hz, 3H, C*H*$_3$-C-5); **^{13}C-NMR** (90 MHz, CD_3OD) δ 166.9 (C=O, C-4), 153.1 (C=O, C-2), 143.5 (CH, C-6), 140.0 (CH, C-2''), 138.4 (CH, C-3''), 111.0 (C, C-5), 73.4 (CH, C-1'''), 66.7 (CH_2, C-2'''), 50.3 (CH_2, C-1'), 49.9 (CH, C-4''), 46.2 (CH, C-1''), 12.2 (CH_3, *C*H_3-C-5); **IR** (ATR) ν 3362 (br), 2961, 1660, 1260, 1092, 1021, 780 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{12}H_{16}N_2O_4+Na]^+$ 275.1002, found 275.0999. Selective TOCSY, DEPT135, HMBC and COSY were recorded for **HS1-T**.

2.4.4.2. Preparation of 1-({(1*R*,4*S*)-4-[(1*S*)-1,2-dihydroxyethyl]cyclobut-2-en-1-yl}methyl)thymine, HS1-T



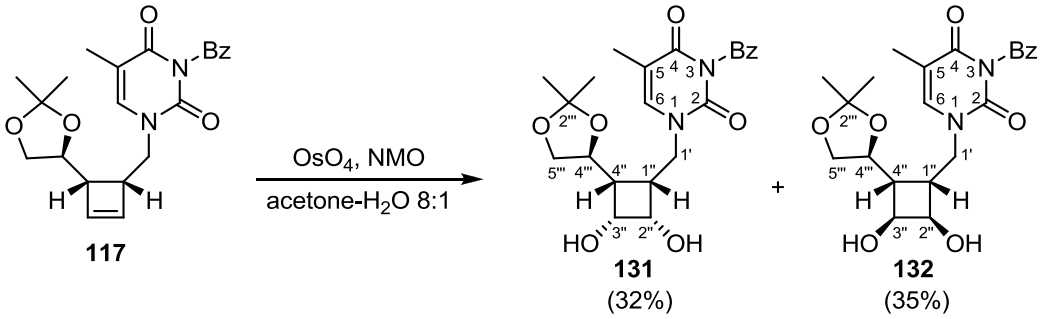
129 **HS1-T**

A solution of **129** (30 mg, 0.08 mmol) in a 33% solution of $MeNH_2$ in EtOH (0.84 mL) was stirred overnight at room temperature. The volatiles were removed

under vacuum, and the remaining crude was purified by column chromatography (CH_2Cl_2-MeOH from 50:1 to 20:1, containing 0.5% of NEt_3) to provide thymine nucleoside analogue **HS1-T** (18 mg, 0.07 mmol, 85% yield) as a white solid. Spectral and analytical data of compound **HS1-T** have been described in the previous section 2.4.4.1 of the present chapter.

2.4.5. Synthesis of the polyhydroxylated cyclobutane nucleoside analogues HS2

2.4.5.1. Preparation of 3-benzoyl-1-({(1*R*,2*S*,3*R*,4*R*)-2,3-dihydroxy-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-1-yl}methyl)thymine, 131, and 3-benzoyl-1-({(1*R*,2*R*,3*S*,4*R*)-2,3-dihydroxy-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-1-yl}methyl)thymine, 132



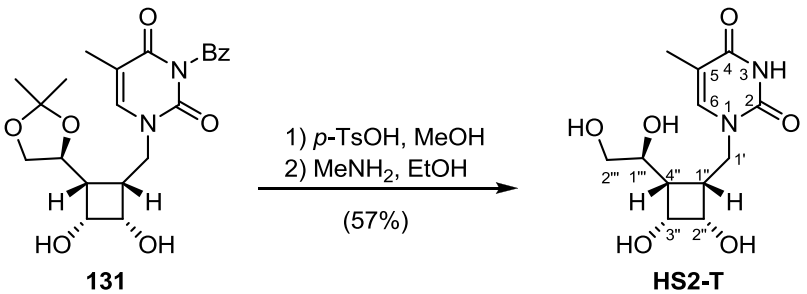
To a stirred solution of **117** (34 mg, 0.09 mmol) in acetone-water 8:1 (0.9 mL), NMO (25 mg, 0.21 mmol) and OsO_4 (2.5 wt% in *t*-BuOH solution, 54 µL, $4.31x10^{-3}$ mmol) were added. After being stirred for 6 h at room temperature, the reaction was quenched by the addition of 10% $NaHSO_3$ (0.9 mL) and the mixture was allowed to stir for 30 min. Then, the aqueous phase was extracted with EtOAc (4 x 2 mL) and the combined organic extracts were dried ($MgSO_4$) and concentrated under reduced pressure. Purification of the crude material by column chromatography (hexane-EtOAc from 1:1 to 1:6) yielded diol **131** (12 mg, 0.03 mmol, 32% yield) and the corresponding diastereomer **132** (13 mg, 0.03 mmol, 35% yield), both as white foams.

131: **CF** $C_{22}H_{26}N_2O_7$; **MW** 430.5 g/mol; **[α]$_D$** = -33.7 (*c* 0.9, MeOH); **^{1}H-NMR** (400 MHz, CD_3OD) δ 7.94 (d, *J*=8.1 Hz, 2H, H-Bz), 7.75 (q, $J_{6,CH3}$=1.0 Hz, 1H, H-6), 7.71 (m, 1H, H-Bz), 7.56 (dd, *J*=8.1 Hz, *J*=8.1 Hz, 2H, H-Bz), 4.69 (ddd, *J*=13.0 Hz, *J*=6.7 Hz, *J*=6.7 Hz, 1H, H-4'''), 4.34 (ddd, *J*=5.5 Hz, *J*=5.5 Hz, *J*=2.5 Hz, 1H, H-2''), 4.20 (m, 3H, H-1', H-3'', H-5'''), 4.07 (dd, J_{gem}=14.1 Hz, $J_{1',1''}$=9.5 Hz, 1H, H-1'), 3.52 (dd, J_{gem}=8.2 Hz, $J_{5''',4'''}$=6.7 Hz, 1H, H-5'''), 2.68 (m, 1H, H-1''), 2.57 (m, 1H, H-4''),

1.92 (d, $J_{CH3,6}$=1.0 Hz, 3H, C*H*$_3$-C-5), 1.35 (s, 3H, C*H*$_3$-C-2'''), 1.33 (s, 3H, C*H*$_3$-C-2'''); **^{13}C-NMR** (100 MHz, CD_3OD) δ 170.5 (C=O, Bz), 165.3 (C=O, C-4), 151.7 (C=O, C-2), 144.5 (CH, C-6), 136.2 (CH, Bz), 133.1 (C, Bz), 131.4 (CH, Bz), 130.4 (CH, Bz), 110.4 (C, C-5), 109.3 (C, C-2'''), 75.7 (CH, C-4'''), 71.3 (CH, C-2''), 70.6 (CH_2, C-5'''), 67.2 (CH, C-3''), 48.0 (CH, C-4''), 46.9 (CH_2, C-1'), 37.3 (CH, C-1''), 27.3 (CH_3, *C*H$_3$-C-2'''), 25.9 (CH_3, *C*H$_3$-C-2'''), 12.3 (CH_3, *C*H$_3$-C-5); **IR** (ATR) ν 3426 (br), 2923, 1745, 1692, 1642, 1442, 1258, 1056, 785 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{22}H_{26}N_2O_7+Na]^+$ 453.1632, found 453.1641. DEPT135, HSQC, HMBC and COSY were recorded for **131**.

132: **CF** $C_{22}H_{26}N_2O_7$; **MW** 430.5 g/mol; **[α]$_D$** = +1.5 (*c* 1.0, MeOH); **^{1}H-NMR** (600 MHz, CD_3OD) δ 7.96 (dd, *J*=8.4 Hz, *J*=1.2 Hz, 2H, H-Bz), 7.72 (tt, *J*=7.3 Hz, *J*=1.2 Hz, 1H, H-Bz), 7.69 (q, $J_{6,CH3}$=1.1 Hz, 1H, H-6), 7.56 (dd, *J*=8.4 Hz, *J*=7.3 Hz, 2H, H-Bz), 4.28 (dddd, *J*=10.0 Hz, *J*=10.0 Hz, *J*=6.6 Hz, *J*=6.6 Hz, 1H, H-4'''), 4.11 (m, 3H, H-1', H-2'', H-5'''), 3.96 (m, 1H, H-1'), 3.94 (m, 1H, H-3''), 3.63 (dd, *J*=8.1 Hz, *J*=6.6 Hz, 1H, H-5'''), 2.80 (dddd, *J*=10.0 Hz, *J*=10.0 Hz, *J*=6.1 Hz, *J*=6.1 Hz, 1H, H-1''), 2.39 (dddd, *J*=10.0 Hz, *J*=10.0 Hz, *J*=4.3 Hz, *J*=0.8 Hz, 1H, H-4''), 1.93 (d, $J_{CH3,6}$=1.1 Hz, 3H, C*H*$_3$-C-5), 1.39 (s, 3H, C*H*$_3$-C-2'''), 1.33 (s, 3H, C*H*$_3$-C-2'''); **^{13}C-NMR** (150 MHz, CD_3OD) δ 170.4 (C=O, Bz), 165.2 (C=O, C-4), 151.7 (C=O, C-2), 143.7 (CH, C-6), 136.2 (CH, Bz), 133.0 (C, Bz), 131.5 (CH, Bz), 130.4 (CH, Bz), 110.9 (C, C-5), 110.5 (C, C-2'''), 75.9 (CH, C-4'''), 71.1 (CH, C-2''), 70.1 (CH, C-3''), 69.4 (CH_2, C-5'''), 48.8 (CH_2, C-1'), 47.1 (CH, C-4''), 43.2 (CH, C-1''), 27.3 (CH_3, *C*H$_3$-C-2'''), 25.8 (CH_3, *C*H$_3$-C-2'''), 12.3 (CH_3, *C*H$_3$-C-5); **IR** (ATR) ν 3392 (br), 2917, 2849, 1742, 1693, 1648, 1461, 1259, 1017, 798 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{22}H_{26}N_2O_7+Na]^+$ 453.1632, found 453.1641. DEPT135, HSQC, HMBC, COSY and selective n.O.e. were recorded for **132**.

2.4.5.2. Preparation of 1-({(1*R*,2*S*,3*R*,4*S*)-2,3-dihydroxy-4-[(1*S*)-1,2-dihydroxyethyl]cyclobut-1-yl}methyl)thymine, HS2-T



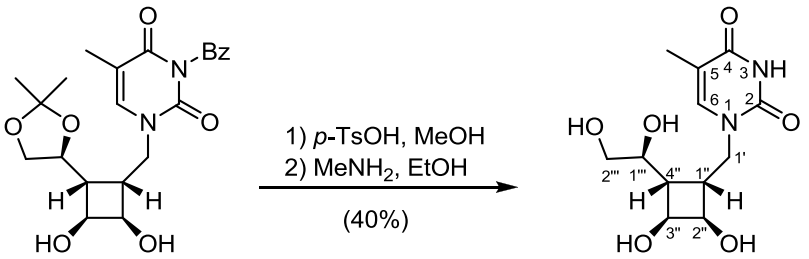
131 **HS2-T**

To a solution of diol **131** (25 mg, 0.06 mmol) in MeOH (2.8 mL), *p*-toluenesulfonic acid monohydrate (12 mg, 0.06 mmol) was added and the resulting mixture was stirred for 6 h at room temperature. Then, the reaction mixture was filtered through a basic anion exchange resin (Dowex 1X8 chloride form, 20-50 mesh). The organic layer was concentrated under reduced pressure and used for the following step without further purification.

The resulting crude was dissolved in a 33% solution of $MeNH_2$ in EtOH (0.5 mL) and the mixture was stirred at room temperature for 1 h. The volatiles were removed under reduced pressure, and the crude was dissolved in Milli-Q water (2 mL) and extracted with CH_2Cl_2 (3 x 2.5 mL). The aqueous layer was concentrated under vacuum, and the resulting crude was then dissolved in MeOH and filtered through an acidic cation exchange resin (Dowex 50WX8 hydrogen form, 200-400 mesh). The organic solvent was evaporated under reduced pressure to afford the thymine nucleoside derivative **HS2-T** (9.5 mg, 0.03 mmol, 57% yield) as a white foam.

HS2-T: **CF** $C_{12}H_{18}N_2O_6$; **MW** 286.3 g/mol; **$[\alpha]_D$** = -61.6 (*c* 1.0, MeOH); **^{1}H-NMR** (400 MHz, CD_3OD) δ 7.57 (q, $J_{6,CH3}$=1.1 Hz, 1H, H-6), 4.25 (m, 2H, H-1', H-3''), 4.19 (m, 2H, H-1', H-2''), 4.06 (ddd, $J_{1''',4''}$=9.8 Hz, $J_{1''',2'''}$=6.1 Hz, $J_{1''',2'''}$=3.7 Hz, 1H, H-1'''), 3.68 (dd, J_{gem}=11.2 Hz, $J_{2''',1'''}$=3.7 Hz, 1H, H-2'''), 3.45 (dd, J_{gem}=11.2 Hz, $J_{2''',1'''}$=6.1 Hz, 1H, H-2'''), 2.80 (m, 1H, H-1''), 2.40 (m, 1H, H-4''), 1.86 (d, $J_{CH3,6}$=1.1 Hz, 3H, C*H*$_3$-C-5); **^{13}C-NMR** (100 MHz, CD_3OD) δ 166.9 (C=O, C-4), 153.3 (C=O, C-2), 144.3 (CH, C-6), 110.4 (C, C-5), 70.7 (CH, C-3''), 69.9 (CH, C-1'''), 68.5 (CH, C-2''), 66.8 (CH_2, C-2'''), 46.8 (CH_2, C-1'), 41.8 (CH, C-4''), 41.1 (CH, C-1''), 12.2 (CH_3, CH_3-C-5); **IR** (ATR) ν 3370 (br), 1672, 1476, 1219, 1126 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{12}H_{18}N_2O_6+Na]^+$ 309.1057, found 309.1061. DEPT135, HSQC, HMBC, COSY and selective n.O.e. were recorded for **HS2-T**.

2.4.5.3. Preparation of 1-({(1*R*,2*R*,3*S*,4*S*)-2,3-dihydroxy-4-[(1*S*)-1,2-dihydroxyethyl]cyclobut-1-yl}methyl)thymine, HS2'-T



132 **HS2'-T**

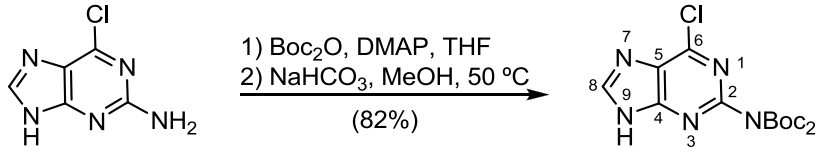
To a solution of diol **132** (22 mg, 0.05 mmol) in MeOH (2.5 mL), *p*-toluenesulfonic acid monohydrate (15 mg, 0.08 mmol) was added and the resulting mixture was stirred for 4 h at room temperature. Then, the reaction mixture was filtered through a basic anion exchange resin. The organic layer was concentrated under reduced pressure and used for the following step without further purification.

The resulting crude was dissolved in a 33% solution of $MeNH_2$ in EtOH (0.5 mL) and the mixture was stirred at room temperature for 2 h. The volatiles were removed under reduced pressure, and the crude was dissolved in Milli-Q water (2 mL) and extracted with CH_2Cl_2 (3 x 2.5 mL). The aqueous layer was concentrated under vacuum, and the resulting crude was then dissolved in MeOH and filtered through an acidic cation exchange resin. The organic solvent was evaporated under reduced pressure to obtain the thymine nucleoside derivative **HS2'-T** (5.8 mg, 0.02 mmol, 40% yield) as a white foam.

HS2'-T: **CF** $C_{12}H_{18}N_2O_6$; **MW** 286.3 g/mol; **$[\alpha]_D$** = -17.7 (*c* 0.9, MeOH); **^{1}H-NMR** (400 MHz, CD_3OD) δ 7.48 (q, $J_{6,CH3}$=1.2 Hz, 1H, H-6), 4.07 (m, 3H, H-1', H-2'', H-3''), 3.95 (dd, J_{gem}=14.0 Hz, $J_{1',1''}$=5.1 Hz, 1H, H-1'), 3.75 (ddd, $J_{1''',4'''}$=9.5 Hz, $J_{1''',2'''}$=6.9 Hz, $J_{1''',2'''}$=3.5 Hz, 1H, H-1'''), 3.64 (dd, J_{gem}=11.5 Hz, $J_{2''',1'''}$=3.5 Hz, 1H, H-2'''), 3.44 (dd, J_{gem}=11.5 Hz, $J_{2''',1'''}$=6.9 Hz, 1H, H-2'''), 2.56 (m, 1H, H-1''), 2.48 (m, 1H, H-4''), 1.87 (d, $J_{CH3,6}$=1.2 Hz, 3H, C*H*$_3$-C-5); **^{13}C-NMR** (100 MHz, CD_3OD) δ 166.9 (C=O, C-4), 153.3 (C=O, C-2), 143.2 (CH, C-6), 111.1 (C, C-5), 72.6 (CH, C-1'''), 71.1 (CH, C-2''), 69.7 (CH, C-3''), 66.4 (CH_2, C-2'''), 48.1 (CH_2, C-1'), 47.1 (CH, C-4''), 41.9 (CH, C-1''), 12.2 (CH_3, *C*H_3-C-5); **IR** (ATR) ν 3364 (br), 1677, 1475, 1260, 1091 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{12}H_{18}N_2O_6+Na]^+$ 309.1057, found 309.1061. DEPT135, HMBC and COSY were recorded for **HS2'-T**.

2.5. Preparation of nucleoside analogues featuring a purine base

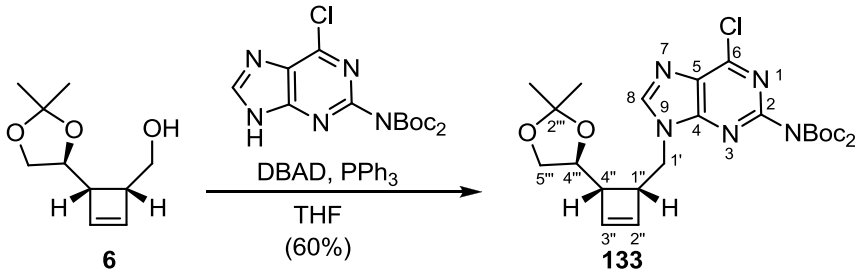
2.5.1. Preparation of *N*,*N*-bis(*tert*-butoxycarbonyl)-2-amino-6-chloropurine



To a suspension of 2-amino-6-chloropurine (1.1 g, 6.48 mmol) and DMAP (79 mg, 0.65 mmol) in dry THF (32 mL), Boc_2O (6 mL, 26.09 mmol) was added under argon atmosphere. The mixture was allowed to stir overnight at room temperature. The organic solvent was then removed under vacuum and the resulting oil was dissolved in 78 mL of MeOH and saturated aqueous sodium bicarbonate solution (35 mL) was added. The reaction mixture was stirred at 50 °C for 1 h and then at room temperature overnight. The bulk of MeOH was evaporated under reduced pressure, and water (78 mL) was added to the suspension. The aqueous layer was extracted with $CHCl_3$ (2 x 230 mL), and the combined organic extracts were dried ($MgSO_4$) and concentrated under reduced pressure. Purification of the resulting oil by column chromatography (from EtOAc to EtOAc-MeOH 95:5) afforded ***N*,*N*-bis(*tert*-butoxycarbonyl)-2-amino-6-chloropurine** (1.96 g, 5.30 mmol, 82% yield) as a white solid.

***N*,*N*-bis(*tert*-butoxycarbonyl)-2-amino-6-chloropurine**: **CF** $C_{15}H_{20}ClN_5O_4$; **MW** 369.8 g/mol; **^{1}H-NMR** (250 MHz, $CDCl_3$) δ 8.36 (s, 1H, H-8), 1.52 (s, 18H, 2x(CH_3)$_3$CO).

2.5.2. Preparation of *N*,*N*-bis(*tert*-butoxycarbonyl)-2-amino-6-chloro-9-({(1*R*,4*S*)-4-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]cyclobut-2-en-1-yl}methyl)purine, 133



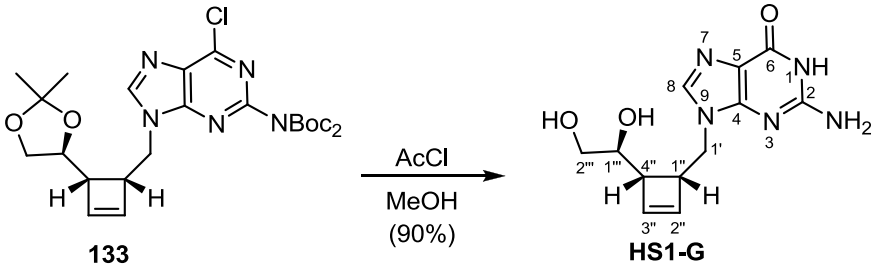
A solution of DBAD (148 mg, 0.64 mmol) in dry THF (1.8 mL) was added dropwise to a 0 ºC stirred suspension of alcohol **6** (59 mg, 0.32 mmol), *N*,*N*-bis(*tert*-

butoxycarbonyl)-2-amino-6-chloropurine (239 mg, 0.64 mmol) and triphenylphosphine (168 mg, 0.64 mmol) in anhydrous THF (2.3 mL) under argon atmosphere. The mixture was allowed to warm to room temperature and stirred overnight. The organic solvent was removed under vacuum and the resulting oil was purified by repeated column chromatography (hexane-EtOAc from 10:1 to 1:1) to afford **133** (103 mg, 0.19 mmol, 60% yield) as a pale yellow solid.

133: **CF** $C_{25}H_{34}ClN_5O_6$; **MW** 536.0 g/mol; **m.p.** 44-47 ºC ($CHCl_3$); **[α]D** = +11.1 (*c* 1.1, $CHCl_3$); **^{1}H-NMR** (360 MHz, $CDCl_3$) δ 8.27 (s, 1H, H-8), 6.16 (d, $J_{2'',3''}$=2.9 Hz, 1H, H-2''), 6.00 (d, $J_{3'',2''}$=2.9 Hz, 1H, H-3''), 4.59 (dd, J_{gem}=14.2 Hz, $J_{1',1''}$=6.5 Hz, 1H, H-1'), 4.43 (dd, J_{gem}=14.2 Hz, $J_{1',1''}$=8.4 Hz, 1H, H-1'), 4.15 (m, 2H, H-5''', H-4'''), 3.69 (ddd, J_{gem}=9.2 Hz, $J_{5''',4'''}$=9.2 Hz, $J_{5''',4'''}$=6.5 Hz, 1H, H-5'''), 3.49 (ddd, $J_{1'',1'}$=8.4 Hz, $J_{1'',1'}$=6.5 Hz, $J_{1'',4''}$=4.2 Hz, 1H, H-1''), 3.10 (dd, $J_{4'',5'''}$=9.2 Hz, $J_{4'',1''}$=4.2 Hz, 1H, H-4''), 1.44 (s, 18H, 2x(C*H*$_3$)$_3$CO), 1.40 (s, 3H, C*H*$_3$-C-2'''), 1.35 (s, 3H, C*H*$_3$-C-2'''); **^{13}C-NMR** (90 MHz, $CDCl_3$) δ 153.0 (C, C-4), 151.8 (C, C-2/C-6/C=O Boc), 151.1 (C, C-2/C-6/C=O Boc), 150.8 (C, C-2/C-6/C=O Boc), 146.7 (CH, C-8), 139.5 (CH, C-2''), 137.0 (CH, C-3''), 130.2 (C, C-5), 109.9 (C, C-2'''), 83.7 (2xC, (CH$_3$)$_3$*C*O), 76.00 (CH, C-4'''), 68.7 (CH_2, C-5'''), 49.7 (CH, C-4''), 45.6 (CH, C-1''), 44.8 (CH_2, C-1'), 28.0 (6xCH_3, (*C*H$_3$)$_3$CO), 27.0 (CH_3, *C*H$_3$-C-2'''), 25.8 (CH_3, *C*H$_3$-C-2'''); **IR** (ATR) ν 2961, 2921, 1723, 1369, 1260, 1094, 1020, 798 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{25}H_{34}ClN_5O_6+Na]^+$ 558.2090, found 558.2087. DEPT135, HSQC, HMBC and COSY were recorded for **133**.

2.5.3. Synthesis of the cyclobutene nucleoside analogues HS1-G and HS1-G^{Cl}

2.5.3.1. Preparation of 9-({(1*R*,4*S*)-4-[(1*S*)-1,2-dihydroxyethyl]cyclobut-2-en-1-yl}methyl)guanine, HS1-G

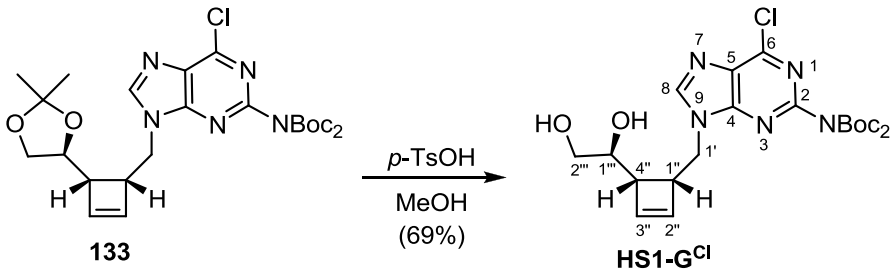


To a stirred solution of **133** (32 mg, 0.06 mmol) in MeOH (1.4 mL), acetyl chloride (86 µL, 1.19 mmol) was added. After being stirred at room temperature for

70 h, the reaction mixture was added dropwise over diethyl ether (18 mL) and it was allowed to precipitate at 5 ºC. The resulting solid was separated from the organic solvent by decantation and dried under vacuum to afford the guanine nucleoside derivative **HS1-G** (15 mg, 0.05 mmol, 90% yield) as a pale yellow solid.

HS1-G: **CF** $C_{12}H_{15}N_5O_3$; **MW** 277.3 g/mol; **m.p.** >280 ºC (MeOH); **[α]$_D$** = -126.4 (*c* 1.1, MeOH); **^{1}H-NMR** (250 MHz, CD_3OD) δ 9.08 (s, 1H, H-8), 6.18 (m, 2H, H-2'', H-3''), 4.61 (dd, J_{gem}=13.9 Hz, $J_{1',1''}$=7.4 Hz, 1H, H-1'), 4.37 (dd, J_{gem}=13.9 Hz, $J_{1',1''}$=8.1 Hz, 1H, H-1'), 3.65 (m, 3H, H-1''', H-2''', H-1''), 3.51 (dd, J_{gem}=13.7 Hz, $J_{2''',1'''}$=6.8 Hz, 1H, H-2'''), 3.08 (dd, *J*=10.0 Hz, *J*=4.2 Hz, 1H, H-4''); **^{13}C-NMR** (62.5 MHz, CD_3OD) δ 157.2 (C, C-2/C-6), 155.1 (C, C-2/C-6), 151.8 (C, C-4), 139.5 (CH, C-2''/C-3''), 138.8 (CH, C-2''/C-3''), 138.5 (CH, C-8), 108.6 (C, C-5), 73.2 (CH, C-1'''), 66.5 (CH_2, C-2'''), 49.8 (CH, C-4''), 47.1 (CH_2, C-1'), 45.8 (CH, C-1''); **IR** (ATR) ν 3315 (br), 3175 (br), 2922, 1700, 1634, 1590, 1368, 1058, 1022, 730 cm^{-1}; **HRMS** (ESI+) calculated for $[C_{12}H_{15}N_5O_3+Na]^+$ 300.1067, found 300.1072. DEPT135, HSQC, HMBC and COSY were recorded for **HS1-G**.

2.5.3.2. Preparation of *N*,*N*-bis(*tert*-butoxycarbonyl)-2-amino-6-chloro-9-({(1*R*,4*S*)-4-[(1*S*)-1,2-dihydroxyethyl]cyclobut-2-en-1-yl}methyl)purine, HS1-G^{Cl}



To a solution of **133** (28 mg, 0.05 mmol) in MeOH (2.3 mL), *p*-toluenesulfonic acid monohydrate (15 mg, 0.08 mmol) was added and the resulting mixture was stirred for 24 h at room temperature. After removal of the solvent, the residue was purified by column chromatography (from hexane-EtOAc 1:1 to EtOAc-MeOH 9:1) to furnish the *N*,*N*-bis(*tert*-butoxycarbonyl)-2-amino-6-chloropurine nucleoside derivative **HS1-G^{Cl}** (18 mg, 0.04 mmol, 69% yield) as a white solid.

HS1-G^{Cl}: **CF** $C_{22}H_{30}ClN_5O_6$; **MW** 496.0 g/mol; **m.p.** 58-62 ºC ($CHCl_3$); **[α]$_D$** = -3.2 (*c* 0.9, $CHCl_3$); **^{1}H-NMR** (400 MHz, $CDCl_3$) δ 8.25 (s, 1H, H-8), 6.15 (d, $J_{2'',3''}$=2.9 Hz, 1H, H-2''), 6.07 (dd, $J_{3'',2''}$=2.9 Hz, $J_{3'',4''}$=0.8 Hz, 1H, H-3''), 4.67 (dd, J_{gem}=14.4 Hz,

$J_{1',1''}$=7.2 Hz, 1H, H-1'), 4.45 (dd, J_{gem}=14.4 Hz, $J_{1',1''}$=6.6 Hz, 1H, H-1'), 3.74 (m, 1H, H-2'''), 3.70 (m, 1H, H-1'''), 3.54 (ddd, $J_{1'',1'}$=7.2 Hz, $J_{1'',1'}$=6.6 Hz, $J_{1'',4''}$=4.2 Hz, 1H, H-1''), 3.49 (m, 1H, H-2'''), 3.06 (ddd, $J_{4'',1'''}$=10.3 Hz, $J_{4'',1''}$=4.2 Hz, $J_{4'',3''}$=0.8 Hz, 1H, H-4''), 2.85 (br s, 1H, -OH), 2.19 (br s, 1H, -OH), 1.47 (s, 18H, 2x(CH_3)$_3$CO); **^{13}C-NMR** (100 MHz, $CDCl_3$) δ 153.2 (C, C-4), 151.7 (C, C-2/C-6/C=O Boc), 151.3 (C, C-2/C-6/C=O Boc), 151.1 (C, C-2/C-6/C=O Boc), 146.6 (CH, C-8), 138.2 (CH, C-2''), 138.0 (CH, C-3''), 130.0 (C, C-5), 84.2 (2xC, (CH_3)$_3$CO), 71.6 (CH, C-1'''), 65.8 (CH_2, C-2'''), 48.6 (CH, C-4''), 46.2 (CH, C-1''), 44.4 (CH_2, C-1'), 28.1 (6xCH_3, (CH_3)$_3$CO); **IR** (ATR) ν 2960, 2927, 1736, 1562, 1368, 1260, 1099, 1021, 796 cm^{-1}; **HRMS** (ESI+) calculated for [$C_{22}H_{30}ClN_5O_6$+Na]$^+$ 518.1777, found 518.1782. DEPT135, HSQC, HMBC and COSY were recorded for **HS1-G^{Cl}**.

3. Antiviral activity evaluation

Table ES1. Cytotoxicity and antiviral activity of nucleosides **HS1** and **HS2** in HEL cell cultures.

Compound	Minimum cytotoxic conc.[a] (μg/ml)	EC_{50}[b] (μg/ml)				
		Herpes simplex virus-1 (KOS)	Herpes simplex virus-2 (G)	Vaccinia virus	Vesicular stomatitis virus	Herpes simplex virus-1 TK$^-$ KOS ACV^r
HS1-T	>100	>100	>100	>100	>100	>100
HS2-T	>100	>100	>100	>100	>100	>100
HS2'-T	>100	>100	>100	>100	>100	>100
HS1-G^{Cl}	>100	>100	>100	>100	>100	>100
HS1-G	>100	>100	>100	>100	>100	>100
Brivudin (μM)	>250	0.01	96	1.8	>250	250
Cidofovir (μM)	>250	1.2	0.4	10	>250	2
Acyclovir (μM)	>250	0.4	0.08	>250	>250	126
Ganciclovir (μM)	>100	0.03	0.03	>100	>100	4

[a] Required to cause a microscopically detectable alteration of normal cell morphology. [b] Required to reduce virus-induced cytopathogenicity by 50%.

Table ES2. Cytotoxicity and antiviral activity of nucleosides **HS1** and **HS2** in HeLa cell cultures.

Compound	Minimum cytotoxic conc.[a] (µg/ml)	EC_{50}[b] (µg/ml)		
		Vesicular stomatitis virus	Coxsackie virus B4	Respiratory syncytial virus
HS1-T	>100	>100	>100	>100
HS2-T	>100	>100	>100	>100
HS2'-T	>100	>100	>100	>100
HS1-G^{Cl}	100	>20	>20	>20
HS1-G	100	>20	>20	>20
DS-5000	>100	100	100	1.8
(S)-DHPA (µM)	>250	>250	>250	>250
Ribavirin (µM)	>250	22	25	4.5

[a] Required to cause a microscopically detectable alteration of normal cell morphology. [b] Required to reduce virus-induced cytopathogenicity by 50%.

Table ES3. Cytotoxicity and antiviral activity of nucleosides **HS1** and **HS2** in Vero cell cultures.

Compound	Minimum cytotoxic conc.[a] (µg/ml)	EC_{50}[b] (µg/ml)				
		Para-influenza-3 virus	Reovirus-1	Sindbis virus	Coxsackie virus B4	Punta Toro virus
HS1-T	>100	>100	>100	>100	>100	>100
HS2-T	>100	>100	>100	>100	>100	>100
HS2'-T	>100	>100	>100	>100	>100	>100
HS1-G^{Cl}	>100	>100	>100	>100	9	45
HS1-G	>100	>100	>100	>100	>100	>100
DS-5000	>100	>100	>100	100	38	20
(S)-DHPA (µM)	>250	>250	>250	>250	>250	>250
Ribavirin (µM)	>250	29	112	>250	>250	29

[a] Required to cause a microscopically detectable alteration of normal cell morphology. [b] Required to reduce virus-induced cytopathogenicity by 50%.

Table ES4. Cytotoxicity and antiviral activity of nucleosides **HS1** and **HS2** in Vero cell cultures (experiment 2).

Compound	Minimum cytotoxic conc.[a] (µg/ml)	EC_{50}[b] (µg/ml)				
		Para-influenza-3 virus	Reovirus-1	Sindbis virus	Coxsackie virus B4	Punta Toro virus
HS1-G^{Cl}	>100	>100	>100	>100	9	>100
DS-5000	>100	>100	100	100	100	>100
(S)-DHPA (µM)	>250	>250	>250	>250	>250	>250
Ribavirin (µM)	>250	50	>250	>250	>250	112

[a] Required to cause a microscopically detectable alteration of normal cell morphology. [b] Required to reduce virus-induced cytopathogenicity by 50%.

Table ES5. Cytotoxicity and antiviral activity of nucleosides **HS1** and **HS2** in CRFK (Crandell-Rees Feline Kidney) cell cultures.

Compound	CC_{50}[a] (µg/ml)	EC_{50}[b] (µg/ml)	
		Feline Corona Virus (FIPV)	Feline Herpes Virus
HS1-T	>100	>100	>100
HS2-T	>100	>100	>100
HS2'-T	>100	>100	>100
HS1-G^{Cl}	46.5	>20	>20
HS1-G	>100	>100	>100
HHA	>100	12.2	18.8
UDA	52.4	0.9	1.6
Ganciclovir (µM)	>100	>100	0.7

[a] 50% Cytotoxic concentration, as determined by measuring the cell viability with the colorimetric formazan-based MTS assay. [b] 50% Effective concentration, or concentration producing 50% inhibition of virus-induced cytopathic effect, as determined by measuring the cell viability with the colorimetric formazan-based MTS assay.

Table ES6. Cytotoxicity and antiviral activity of nucleosides **HS1** and **HS2** in MDCK (Madin Darby canine kidney) cell cultures.

Compound	Concentration unit	Cytotoxicity		Antiviral EC_{50}[c] (µg/ml)					
				Influenza A H1N1 subtype		Influenza A H3N2 subtype		Influenza B	
		CC_{50}[a]	Minimum cytotoxic conc.[b]	visual CPE score	MTS	visual CPE score	MTS	visual CPE score	MTS
HS1-T	µg/ml	51.9	100	>20	>20	>20	>20	>20	>20
HS2-T	µg/ml	64.0	100	>20	>20	>20	>20	>20	>20
HS2'-T	µg/ml	58.9	100	>20	>20	>20	>20	>20	>20
HS1-G[d]	µg/ml	>100	100	>20	>20	>20	>20	>20	>20
HS1-G	µg/ml	>100	>100	>100	>100	>100	>100	>100	>100
Oseltamivir carboxylate	µM	>100	>100	9	12.6	0.8	2.0	20	32.1
Ribavirin	µM	>100	>100	9	9.6	7	2.6	7	1.9
Amantadine	µM	>200	>200	>200	>200	0.7	0.4	>200	>200
Rimantadine	µM	>200	>200	40	48.1	0.7	0.1	>200	>200

[a] 50% Cytotoxic concentration, as determined by measuring the cell viability with the colorimetric formazan-based MTS assay. [b] Minimum compound concentration that causes a microscopically detectable alteration of normal cell morphology. [c] 50% Effective concentration, or concentration producing 50% inhibition of virus-induced cytopathic effect, as determined by visual scoring of the CPE, or by measuring the cell viability with the colorimetric formazan-based MTS assay.

4. *In silico* study of cyclobutane-fused nucleoside analogues as anti-HIV agents

4.1. Anti-HIV activity and cytotoxicity

Table ES7. Anti HIV-1 activity against wild-type NL4-3 strain and cytotoxicity of nucleosides **HI1**–**HI3**.[a]

Compound	EC_{50}[b] (μg/ml)	CC_{50}[c] (μg/ml)
HI1.H-T	>25	>25
HI1.F-T	>25	>25
HI1.Cl-T	>25	>25
HI2-T	>25	>25
HI3.O-T	>25	>25
HI3.OH-T	>25	>25
HI3.CH_2-T	>25	>25
d4T	0.01	>2
AZT	0.0007	>1

[a] Data represent mean values of at least two experiments. [b] EC_{50}: effective concentration 50 or needed concentration to inhibit 50% HIV-induced cell death, evaluated with the MTT method in MT-4 cells. [c] CC_{50}: cytotoxic concentration 50 or needed concentration to induce 50% death of non-infected cells, evaluated with the MTT method in MT-4 cells.

4.2. Computational details

Table ES8. PDB structures and computational details of the docking calculations on **HI1–HI3**.

Enzyme	PDB code	Chain	Crystallized ligand	Centre of the cavity (radius)	Flexible residues
TK1	1XBT	G	dTTP	Tyr-181, CZ (20 Å)	Phe-101, Leu-124, Phe-133, Arg-165, Tyr-181
TK1	1W4R	A	dTTP	Tyr-181, CZ (20 Å)	Phe-101, Leu-124, Phe-133, Arg-165, Tyr-181
TK1	1W4R	B	dTTP	Tyr-181, CZ (20 Å)	Phe-101, Leu-124, Phe-133, Arg-165, Tyr-181
TMK	1E2F	A	dTMP (+ADP)	Phe-72, CG (12 Å)	Leu-57, Phe-72, Arg-76, Ser-101, Phe-105, Tyr-151
TMK	1E9A	A	$AZTP_5A$[a]	Phe-72, CG (12 Å)	Leu-57, Phe-72, Arg-76, Ser-101, Phe-105, Tyr-151
NDPK	1LWX	B	AZTDP	Ligand[b] (6 Å)	-
HIV-1 RT	1RTD	A, B, E, F[c]	dTTP	Tyr-115, CZ (10 Å)	Tyr-115, Met-184, Asp-185
HIV-1 RT	1N5Y	A, P, T[c]	AZTMP	Tyr-183, CZ (10 Å)	Asp-110, Tyr-115, Met-184, Asp-185

[a] AZTP5A: P1-(5'-adenosyl)P5-(5'-(3'azido-3'-deoxythymidyl))pentaphosphate. [b] The centre of the cavity was defined by the ligand position in the X-ray structure. [c] More than one chain was kept for the docking calculations on HIV-1 RT since the DNA strands were also taken into account when performing the calculations.

4.3. Drug-likeness

Table ES9. Standard molecular descriptors of compounds dT, d4T and **HI1–HI3** for "Lipinski's rule of 5".

Compound	H bond donors	H bond acceptors	MW	log P[371,372]
dT	3	5	242.23	-1.12
d4T	2	4	224.21	-0.23
HI1.H-T	2	4	252.27	-0.03
HI1.F-T	2	4	270.26	-0.06
HI1.Cl-T	2	4	286.71	0.40
HI2-T	2	4	250.25	-0.39
HI3.O-T	2	5	266.25	-0.88
HI3.OH-T	3	5	268.27	-1.37
HI3.CH_2-T	2	4	264.28	-0.09
Lipinski's criterion	≤5	≤10	≤500	≤5

Table ES10. Standard molecular descriptors of compounds dT, d4T and **HI1–HI3** for Veber's criterion.

Compound	Rotatable bonds	Polar surface area,[373] pH=7.3 (Å^2)
dT	2	99.10
d4T	2	78.87
HI1.H-T	2	78.87
HI1.F-T	2	78.87
HI1.Cl-T	2	78.87
HI2-T	2	78.87
HI3.O-T	2	95.94
HI3.OH-T	2	99.10
HI3.CH_2-T	2	78.87
Veber's criterion	≤10	≤140

4.4. Activation process

4.4.1. 1st phosphorylation step

Table ES11. Predicted binding energies (score units) of unphosphorylated dT, d4T and **HI1–HI3** in TK1.

Compound	1XBT	1W4R (chain A)	1W4R (chain B)	Average
dT	-21.7	-22.4	-21.0	-21.7
d4T	-22.1	-23.1	-21.8	-22.3
HI1.H-T	-25.7	-26.5	-25.2	-25.8
HI1.F-T	-23.0	-24.4	-20.5	-22.6
HI1.Cl-T	-24.8	-25.4	-25.3	-25.2
HI2-T	-23.3	-26.0	-25.0	-24.7
HI3.O-T	-15.4	-17.9	-17.8	-17.0
HI3.OH-T	-16.8	-19.1	-20.3	-18.8
HI3.CH_2-T	-17.0	-17.9	-17.8	-17.5

Table ES12. Predicted binding energies (score units) of monophosphorylated dT, d4T and **HI1–HI3** in TK1.

Compound	1XBT	1W4R (chain A)	1W4R (chain B)	Average
dTMP	-26.3	-27.5	-25.0	-26.3
d4TMP	-25.2	-25.3	-25.0	-25.2
HI1.H-TMP	-27.6	-27.1	-26.2	-26.9
HI1.F-TMP	-28.0	-28.2	-27.7	-28.0
HI1.Cl-TMP	-29.1	-28.3	-25.1	-27.5
HI2-TMP	-28.4	-28.1	-27.4	-28.0
HI3.O-TMP	-22.9	-20.9	-21.6	-21.8
HI3.OH-TMP	-24.2	-23.1	-22.4	-23.3
HI3.CH_2-TMP	-24.2	-21.2	-20.1	-21.8

4.4.2. 2nd phosphorylation step

Table ES13. Predicted binding energies (score units) of monophosphorylated dT, d4T and **HI1–HI3** in TMK.

Compound	1E2F	1E9A	Average
dTMP	-8.8	-9.6	-9.2
d4TMP	-13.5	-15.0	-14.3
HI1.H-TMP	-17.3	-15.6	-16.4
HI1.F-TMP	-14.1	-16.0	-15.1
HI1.Cl-TMP	-17.7	-15.9	-16.8
HI2-TMP	-16.7	-13.0	-14.5
HI3.O-TMP	-16.1	-12.9	-17.2
HI3.OH-TMP	-14.7	-14.4	-14.5
HI3.CH_2-TMP	-16.6	-17.7	-14.9

Table ES14. Predicted binding energies (score units) of diphosphorylated dT, d4T and **HI1–HI3** in TMK.

Compound	1E2F	1E9A	Average
dTDP	-12.0	-8.0	-10.0
d4TDP	-15.7	-13.4	-14.5
HI1.H-TDP	-17.7	-17.7	-17.7
HI1.F-TDP	-18.3	-17.4	-17.9
HI1.Cl-TDP	-15.0	-13.2	-14.1
HI2-TDP	-18.2	-16.7	-15.0
HI3.O-TDP	-14.3	-15.7	-16.5
HI3.OH-TDP	-13.7	-14.1	-13.9
HI3.CH_2-TDP	-16.3	-16.8	-17.5

4.4.3. 3rd phosphorylation step

Table ES15. Predicted binding energies (score units) of diphosphorylated dT, d4T and **HI1–HI3** in NDPK.

Compound	1LWX
dTDP	-17.7
d4TDP	-17.9
HI1.H-TDP	-
HI1.F-TDP	-
HI1.Cl-TDP	-
HI2-TDP	-
HI3.O-TDP	-23.4
HI3.OH-TDP	-20.9
HI3.CH_2-TDP	-

Table ES16. Predicted binding energies (score units) of triphosphorylated dT, d4T and **HI1–HI3** in NDPK.

Compound	1LWX
dTTP	-11.2
d4TTP	-
HI1.H-TTP	-
HI1.F-TTP	-14.9
HI1.Cl-TTP	-
HI2-TTP	-
HI3.O-TTP	-13.9
HI3.OH-TTP	-1
HI3.CH_2-TTP	-

4.5. HIV-1 RT interaction

Table ES17. Predicted binding energies (score units) of triphosphorylated dT, d4T and **HI1–HI3** in their interaction with pre-catalytic HIV-1 RT.

Compound	1RTD
dTTP	-13.5
d4TTP	-15.3
HI1.H-TTP	-
HI1.F-TTP	-
HI1.Cl-TTP	-
HI2-TTP	-
HI3.O-TTP	-7.5
HI3.OH-TTP	-8.7
HI3.CH_2-TTP	-

Table ES18. Predicted binding energies (score units) of monophosphorylated dT, d4T and **HI1–HI3** covalently bound to the DNA primer chain in HIV-1 RT.

Compound	1N5Y
dTMP	-26.6
d4TMP	-25.6
HI1.H-TMP	-26.5
HI1.F-TMP	-26.3
HI1.Cl-TMP	-25.5
HI2-TMP	-26.7
HI3.O-TMP	-23.3
HI3.OH-TMP	-21.5
HI3.CH_2-TMP	-24.9

5. *In silico* study of cyclobutene and cyclobutane L-nucleoside analogues as anti-HSV agents

5.1. Computational details

Table ES19. PDB structures and computational details of the docking calculations on **HS1** and **HS2**.

Enzyme	PDB code	Chain	Crystallized ligand	Centre of the cavity (radius)	Flexible residues
HSV-1 TK	1KIM	A	dT	Tyr-172, CD2 (20 Å)	His-58, Lys-62, Glu-83, Arg-222
HSV-1 TK	2VTK	A	dT (+ ADP)	Tyr-172, CD2 (20 Å)	His-58, Lys-62, Glu-83, Arg-222
HSV-1 TK	1E2K	A	MCT	Tyr-172, CD2 (20 Å)	His-58, Lys-62, Glu-83, Arg-222
HSV-1 TK	1OF1	A	(S)-MCT	Tyr-172, CD2 (20 Å)	His-58, Lys-62, Glu-83, Arg-222
HSV-1 TK	1KI3	B	PCV	Tyr-172, CD2 (20 Å)	His-58, Lys-62, Glu-83, Arg-222
HSV-1 TK	2KI5	A	ACV	Tyr-172, CD2 (20 Å)	His-58, Lys-62, Glu-83, Arg-222
HSV-1 TK	1VTK	A	dTMP (+ ADP)	Tyr-172, CD2 (20 Å)	His-58, Lys-62, Glu-83, Arg-222
Mouse GMPK	1LVG	A	GMP (+ ADP)	Asp-101, CG (10 Å)	Ser-37, Arg-41, Tyr-53
Dictyostelium discoideum NDPK	1NDC	A	dTDP	Lys-16, NZ (9 Å)	-
Human NDPK	1NUE	A	GDP	Lys-12, NZ (9 Å)	-

5.2. Drug-likeness

Table ES20. Standard molecular descriptors of compounds dT, ACV and **HS1** and **HS2** for "Lipinski's rule of 5".

Compound	H bond donors	H bond acceptors	MW	log P[371,372]
dT	3	5	242.23	-1.12
ACV	3	6	225.20	-1.55
HS1-T	3	4	252.27	-1.07
HS2-T	5	6	286.28	-3.09
HS2'-T	5	6	286.28	-3.09
HS1'-G^{Cl}	3	6	295.73	-0.24
HS1-G	4	6	277.28	-1.76
Lipinski's criterion	≤5	≤10	≤500	≤5

Table ES21. Standard molecular descriptors of compounds dT, ACV and **HS1** and **HS2** for Veber's criterion.

Compound	Rotatable bonds	Polar surface area,[373] pH=7.3 (Å^2)
dT	2	99.10
ACV	4	114.76
HS1-T	4	89.87
HS2-T	4	130.33
HS2'-T	4	130.33
HS1'-G^{Cl}	4	110.08
HS1-G	4	125.76
Veber's criterion	≤10	≤140

5.3. Activation process

5.3.1. 1st phosphorylation step

Table ES22. Predicted binding energies (score units) of unphosphorylated dT, **HS1-T**, **HS2-T** and **HS2'-T** in HSV-1 TK.

Compound	1KIM	2VTK	1E2K	1OF1	Average
dT	-20.5	-19.8	-23.9	-24.2	-22.1
HS1-T	-23.6	-25.5	-30.0	-30.0	-27.3
HS2-T	-20.9	-19.1	-25.4	-25.8	-22.8
HS2'-T	-20.0	-18.9	-	-	-19.4

Table ES23. Predicted binding energies (score units) of monophosphorylated dT, **HS1-T**, **HS2-T** and **HS2'-T** in HSV-1 TK.

Compound	1VTK
dTMP	-19.6
HS1-T-1'''-MP	-17.2
HS2-T-1'''-MP	-17.0
HS2'-T-1'''-MP	-17.7
HS1-T-2'''-MP	-19.4
HS2-T-2'''-MP	-19.6
HS2'-T-2'''-MP	-12.7

Table ES24. Predicted binding energies (score units) of unphosphorylated ACV, **HS1'-G^{Cl}** and **HS1-G** in HSV-1 TK.

Compound	1KI3	2KI5	Average
ACV	-8.6	-14.9	-11.8
HS1'-G^{Cl}	-14.3	-17.2	-15.8
HS1-G	-14.9	-27.1	-21.0

Table ES25. Predicted binding energies (score units) of monophosphorylated ACV, **HS1'-G^{Cl}** and **HS1-G** in HSV-1 TK.

Compound	1KI3	2KI5	Average
ACVMP	-4.9	-9.8	-7.4
HS1'-G^{Cl}-1'''-MP	-6.6	-12.4	-9.5
HS1-G-1'''-MP	-10.9	-17.7	-14.3
HS1'-G^{Cl}-2'''-MP	-10.1	-9.9	-10.0
HS1-G-2'''-MP	-13.3	-15.4	-14.4

5.3.2. 2nd phosphorylation step

Table ES26. Predicted binding energies (score units) of diphosphorylated dT, **HS1-T**, **HS2-T** and **HS2'-T** in HSV-1 TK.

Compound	1VTK
dTDP	-16.5
HS1-T-1'''-DP	-7.1
HS2-T-1'''-DP	-6.9
HS2'-T-1'''-DP	-4.8
HS1-T-2'''-DP	-17.9
HS2-T-2'''-DP	-7.4
HS2'-T-2'''-DP	-2.2

Table ES27. Predicted binding energies (score units) of monophosphorylated ACV, **HS1'-G^{Cl}** and **HS1-G** in mGMPK.

Compound	1LVG
ACVMP	-12.2
HS1'-G^{Cl}-1'''-MP	-
HS1-G-1'''-MP	-20.1
HS1'-G^{Cl}-2'''-MP	-
HS1-G-2'''-MP	-19.7

Table ES28. Predicted binding energies (score units) of diphosphorylated ACV, **HS1'-G^{Cl}** and **HS1-G** in mGMPK.

Compound	1LVG
ACVDP	-12.1
HS1'-G^{Cl}-1'''-DP	-5.8
HS1-G-1'''-DP	-11.2
HS1'-G^{Cl}-2'''-DP	-
HS1-G-2'''-DP	-16.7

5.3.3. 3rd phosphorylation step

Table ES29. Predicted binding energies (score units) of diphosphorylated dT and **HS1-T**, **HS2-T** and **HS2'-T** in *Dictyostelium discoideum* NDPK.

Compound	1NDC
dTDP	-16.7
HS1-T-1'''-DP	-18.0
HS2-T-1'''-DP	-11.4
HS2'-T-1'''-DP	-11.9
HS1-T-2'''-DP	-20.2
HS2-T-2'''-DP	-14.2
HS2'-T-2'''-DP	-16.3

Table ES30. Predicted binding energies (score units) of triphosphorylated dT and **HS1-T**, **HS2-T** and **HS2'-T** in *Dictyostelium discoideum* NDPK.

Compound	1NDC
dTTP	-13.1
HS1-T-1'''-TP	-20.6
HS2-T-1'''-TP	-9.2
HS2'-T-1'''-TP	-6.5
HS1-T-2'''-TP	-17.6
HS2-T-2'''-TP	-7.5
HS2'-T-2'''-TP	-14.1

Table ES31. Predicted binding energies (score units) of diphosphorylated ACV, **HS1'-G^{Cl}** and **HS1-G** in human NDPK.

Compound	1NUE
ACVDP	-5.4
HS1'-G^{Cl}-1'''-DP	-16.5
HS1-G-1'''-DP	-13.4
HS1'-G^{Cl}-2'''-DP	-16.2
HS1-G-2'''-DP	-11.7

Table ES32. Predicted binding energies (score units) of triphosphorylated ACV, **HS1'-G^{Cl}** and **HS1-G** in human NDPK.

Compound	1NUE
ACVTP	-2.5
HS1'-G^{Cl}-1'''-TP	-9.4
HS1-G-1'''-TP	-6.9
HS1'-G^{Cl}-2'''-TP	-11.7
HS1-G-2'''-TP	-5.1

6. Rational design of cyclobutene and cyclobutane L-nucleoside analogues

Table ES33. Standard molecular descriptors of compounds dT, ACV and **HS3–HS13** for "Lipinski's rule of 5".

Compound	H bond donors	H bond acceptors	MW	log P[371,372]
dT	3	5	242.23	-1.12
ACV	3	6	225.20	-1.55
HS3-T	2	3	222.24	-0.44
HS3-G	3	5	247.25	-1.13
HS3-T^{F}	2	3	240.23	-0.83
HS3H-T	2	3	224.26	-0.08
HS3H-G	3	5	249.27	-0.77
HS4.F-T	2	3	260.24	-0.68
HS4'.F-T	2	3	260.24	-0.68

HS5.F-T	2	3	242.25	-0.54
HS5.CH_2OH-T	3	4	254.28	-1.15
HS6.F-T	2	3	242.25	-0.54
HS6.CH_2F-T	2	3	256.27	-0.26
HS6.CH_2OH-T	3	4	254.28	-1.15
HS6.OH-G	4	6	265.27	-2.11
HS6.CH_2F-G	3	5	281.29	-0.95
HS6.CH_2OH-G	4	6	279.30	-1.84
HS7.OH-G	4	6	263.25	-2.01
HS8-T	2	3	222.24	-0.16
HS8-G	3	5	247.25	-0.85
HS8H-T	2	3	224.26	0.05
HS8H-G	3	5	249.27	-0.64
HS9-T	2	3	222.24	-0.16
HS9-G	3	5	247.25	-0.85
HS9H-T	2	3	224.26	0.05
HS9H-G	3	5	249.27	-0.64
HS10.CH_2OH-T	3	4	254.28	-1.02
HS11.CH_2OH-T	3	4	254.28	-0.98
HS11.CH_2NH_2-T	3	4	253.30	-1.51
HS11.CH_2F-T	2	3	256.27	-0.09
HS12H-T	2	3	210.23	-0.36
HS12H-G	3	5	235.24	-1.04
HS13.CH_2OH-T	3	4	240.26	-1.43
HS13.CH_2OH-G	4	6	265.27	-2.11
Lipinski's criterion	≤5	≤10	≤500	≤5

Table ES34. Standard molecular descriptors of compounds dT, ACV and **HS3–HS13** for Veber's criterion.

Compound	Rotatable bonds	Polar surface area,[373] pH=7.3 (Å^2)
dT	2	99.10
ACV	4	114.76
HS3-T	3	69.64
HS3-G	3	105.53

HS3-T^F	4	69.64
HS3H-T	3	69.64
HS3H-G	3	105.53
HS4.F-T	3	69.64
HS4'.F-T	3	69.64
HS5.F-T	3	69.64
HS5.CH_2OH-T	4	89.87
HS6.F-T	3	69.64
HS6.CH_2F-T	4	69.64
HS6.CH_2OH-T	4	89.87
HS6.OH-G	3	125.76
HS6.CH_2F-G	4	105.53
HS6.CH_2OH-G	4	125.76
HS7.OH-G	3	125.76
HS8-T	3	69.64
HS8-G	3	105.53
HS8H-T	3	69.64
HS8H-G	3	105.53
HS9-T	3	69.64
HS9-G	3	105.53
HS9H-T	3	69.64
HS9H-G	3	105.53
HS10.CH_2OH-T	4	89.87
HS11.CH_2OH-T	4	89.87
HS11.CH_2NH_2-T	4	97.28
HS11.CH_2F-T	4	69.64
HS12H-T	2	69.64
HS12H-G	2	105.53
HS13.CH_2OH-T	3	89.87
HS13.CH_2OH-G	3	125.76
Veber's criterion	≤10	≤140

7. Rational design of cyclohexene and bicyclo[4.1.0]heptane nucleoside analogues

Table ES35. Standard molecular descriptors of compounds dT, ACV and **HS14–HS18** for "Lipinski's rule of 5".

Compound	H bond donors	H bond acceptors	MW	log P[371,372]
dT	3	5	242.23	-1.12
ACV	3	6	225.20	-1.55
HS14-T	2	3	222.24	0.13
HS14cp-T	2	3	236.27	0.00
HS14-G	3	5	247.25	-0.56
HS14cp-G	3	5	261.28	-0.69
D-**HS15.OH-T**	3	4	238.24	-1.02
D-**HS15.CH$_2$OH-T**	3	4	252.27	-0.90
L-**HS15.CH$_2$CH$_2$OH-T**	3	4	266.29	-0.50
D-**HS15.OH-G**	4	6	263.25	-1.71
D-**HS15.CH$_2$OH-G**	4	6	277.28	-1.59
L-**HS15.CH$_2$CH$_2$OH-G**	4	6	291.31	-1.19
D-**HS16.OH-T**	3	4	238.24	-1.02
D-**HS16.CH$_2$OH-T**	3	4	252.27	-0.90
D-**HS16.OH-G**	4	6	263.25	-1.71
D-**HS16.CH$_2$OH-G**	4	6	277.28	-1.59
HS17-T	2	3	236.27	0.29
HS17cp-T	2	3	250.29	0.24
HS17-G	3	5	261.28	-0.40
HS17cp-G	3	5	275.31	-0.45
D-**HS18.OH.OH-T**	4	5	268.27	-1.82
D-**HS18.OH.CH$_2$OH-T**	4	5	282.29	-1.83
L-**HS18.OH.CH$_2$CH$_2$OH-T**	4	5	296.32	-1.65
D-**HS18.OH.OH-G**	5	7	293.28	-2.51
D-**HS18.OH.CH$_2$OH-G**	5	7	307.31	-2.52
L-**HS18.OH.CH$_2$CH$_2$OH-G**	5	7	321.33	-2.33
Lipinski's criterion	≤5	≤10	≤500	≤5

Table ES36. Standard molecular descriptors of compounds dT, ACV and **HS14–HS18** for Veber's criterion.

Compound	Rotatable bonds	Polar surface area,[373] pH=7.3 (Å^2)
dT	2	99.10
ACV	4	114.76
HS14-T	1	69.64
HS14cp-T	1	69.64
HS14-G	1	105.53
HS14cp-G	1	105.53
D-**HS15.OH-T**	1	89.87
D-**HS15.CH$_2$OH-T**	2	89.87
L-**HS15.CH$_2$CH$_2$OH-T**	3	89.87
D-**HS15.OH-G**	1	125.76
D-**HS15.CH$_2$OH-G**	2	125.76
L-**HS15.CH$_2$CH$_2$OH-G**	3	125.76
D-**HS16.OH-T**	1	89.87
D-**HS16.CH$_2$OH-T**	2	89.87
D-**HS16.OH-G**	1	125.76
D-**HS16.CH$_2$OH-G**	2	125.76
HS17-T	2	69.64
HS17cp-T	2	69.64
HS17-G	2	105.53
HS17cp-G	2	105.53
D-**HS18.OH.OH-T**	2	110.10
D-**HS18.OH.CH$_2$OH-T**	3	110.10
L-**HS18.OH.CH$_2$CH$_2$OH-T**	4	110.10
D-**HS18.OH.OH-G**	2	145.99
D-**HS18.OH.CH$_2$OH-G**	3	145.99
L-**HS18.OH.CH$_2$CH$_2$OH-G**	4	145.99
Veber's criterion	≤10	≤140

COMPUTATIONAL METHODS

The theoretical work presented in this thesis has been performed by means of molecular docking, which is grounded in molecular mechanics. The first part of the present chapter will be thus devoted to provide a brief overview of molecular mechanics, whereas in the second one the main features of protein–ligand docking techniques will be shortly outlined.

1. Molecular mechanics

Molecular mechanics techniques describe molecules as an ensemble of atoms connected by bonds. This simplification is sometimes compared to a system of balls (atoms) and springs (bonds) that follow the laws of classical physics, wherein each ball has a given radius, charge and mass.

To calculate the potential energy of a system, charges, masses, radii, bond lengths, bond angles, etc. have to be known for each atom and atom pairs. These parameters and the related equations to use them constitute the so-called *force fields*.

Despite several force fields exist (Amber, Charmm, MMFF, etc.), many of the force fields in use today consider that the potential energy V_{pot} of a system is composed by the potential from bonded (or covalent) and non-bonded (non-covalent) interactions of a system, as represented by equation (1)

$$V_{pot} = V_b + V_{nb} \tag{1}$$

where V_{pot} is the potential energy of a system, V_b the energy associated with bonded interactions and V_{nb} the energy arising from non-bonded terms. In the next paragraphs, both bonded and non-bonded terms will be explained in more detail.

Bonded terms

The energy derived from the interactions between atoms linked by a covalent bond is generally expressed as a sum of four terms.

$$V_b = V_{bond} + V_{angle} + V_{dihedral} + V_{improper} \tag{2}$$

- *Bond stretching* (V_{bond}). In molecular mechanics, a bond is represented as a string which vibrates around an equilibrium bond length ($l_{i,0}$). This vibration is

described by a harmonic function in which the system cannot escape (break) from the potential energy well.

$$V_{bond} = \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 \quad (3)$$

Equation (3) represents the energy increase as the bond length l_i deviates from the reference (equilibrium) value $l_{i,0}$, where k_i is a force constant.

- *Angle bending* (V_{angle}). Similarly to bond stretching, the angle formed between three atoms A-B-C, θ_i, vibrates around an equilibrium value $\theta_{i,0}$, and the energy associated to this deviation is again estimated by a harmonic function.

$$V_{angle} = \sum_{angles} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 \quad (4)$$

- *Dihedral motions* ($V_{dihedral}$). This term is a torsional potential that models the energy changes caused by bond rotation. Dihedral or torsion angles between four bonded atoms A-B-C-D are used to represent this bond rotation, and the corresponding energy is calculated by a periodic function, as shown in equation (5).

$$V_{dihedral} = \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) \quad (5)$$

where ω stands for the observed dihedral angle, γ for the phase angle (indicating where ω passes through a minimum), V_n for the rotation barrier and n for the periodicity.

- *Improper torsions* ($V_{improper}$). A molecule cannot be accurately described from only torsional angles. Often a function estimating the energy derived from out-of-plane motions is necessary. There are several ways in which out-of-plane bending terms can be incorporated into a force field, one of most widely used is based on improper dihedral angles. These angles are used to describe the planarity of a system of four atoms wherein a central atom is bonded to three other atoms. A torsional potential of the form expressed in equation (6) is used to maintain the improper torsion angle at 0° or 180°.

$$V_{improper} = k(1 - \cos 2\omega) \quad (6)$$

Improper dihedral angles are also used to prevent the inversion of chiral centres during a calculation.

All the bonded terms could be represented as in Figure CM1.

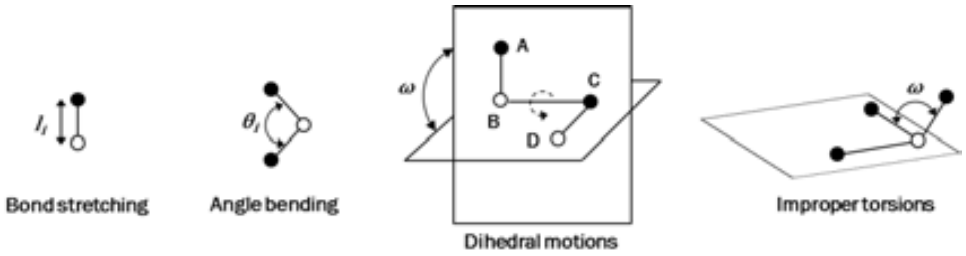


Figure CM1. Schematic representation of the four bonded terms to calculate the potential energy of a system by means of molecular mechanics.

Non-bonded terms

The non-bonded potential energy is calculated for all pairs of atoms (*i* and *j*) that are in different molecules or in the same molecule but separated by at least three bonds. This energy is expressed as a sum of the energies derived from van der Waals (V_{vdw}) and electrostatic (V_{el}) interactions.

$$V_{nb} = V_{vdw} + V_{el} \tag{7}$$

- *Van der Waals interactions* (V_{vdw}). This term describe the behaviour of two atoms approximating each other without forming a covalent bond, which is modelled as a balance between attractive and repulsive forces. Attractive forces act at long distances and represent London dispersion forces, whereas the repulsive contribution is at short distances. A Lennard-Jones potential is commonly used to represent such behaviour and thus model van der Waals interactions.

$$V_{vdw} = \sum_{i}^{N} \sum_{j}^{N} \varepsilon \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^{6} \right] \tag{8}$$

The Lennard-Jones 12-6 potential shown in equation (8) contains three adjustable parameters: the distance between atoms *i* and *j*, r_{ij}; the separation at which the energy passes through a minimum, r_m; and the potential well depth, ε.

- *Electrostatic interactions* (V_{el}). The electrostatic properties of molecules are represented in molecular mechanics by means of partial point charges assigned to each atom centre. As a result, electrostatic interactions may be estimated using Coulomb's law.

$$V_{el} = \sum_{i}^{N} \sum_{j}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (9)$$

where q_i and q_j stand for the point charges of atoms *i* and *j*, ε_0 for the dielectric constant of the medium and r_{ij} for the distance between both atoms. Dielectric constant represents the effect of the environment (solvent and protein) on the interaction between the two charges. The most exact way to represent the solvent would be to consider each solvent molecule explicitly, but such calculations tend to be particularly computer-intensive. For this reason, a common approach is to model the solvent implicitly via the dielectric constant. However, determining its value is not trivial considering that an interaction occurring in bulk solvent would not require the use of the same dielectric constant as if it took place in the surface of the protein. Indeed, the optimization of dielectric constants values of proteins has been extensively studied.[374] To account for the solvent effect on the electrostatic energy term, several other approximations have been applied such as the usage of dielectric constants dependent on the separation of the charged species. Remarkably, when performing molecular simulations with explicit solvent molecules or in gas phase, ε_0 is considered to be 1.

For each force field, a parameterization process has to be performed to determine the mathematical function to calculate the potential energy of a system as well as to optimize its parameters (i.e. constants and reference parameters, such as the ones in equations (3-9)), according to experimental data (e.g. thermodynamic data). Data obtained by accurate quantum mechanics calculations may also be used to parameterize a force field.

Importantly, the numerous degrees of freedom of protein–ligand dockings hamper the use of a function containing all these terms to calculate the potential

energy of the system, usually assuming some simplifications that will be detailed in the next section.

2. Protein–ligand docking

Protein–ligand docking programs consist of a combination of a search algorithm and a scoring function.[266–271] This section will be thus dedicated to present these two parts of docking protocols. A short overview on some of the most widely used docking programs will be also included at the end of the present section.

2.1. Search algorithms

Search algorithms generate orientations of particular ligand conformations in the protein binding site. They try to cover the search space (conformational, translational and rotational) as extensively as possible, but due to the huge number of degrees of freedom associated with both the ligand and the receptor, some approximations have to be applied in order to perform an effective sampling in a feasible time framework. The most basic level of approximation is based on treating both the ligand and the receptor as rigid, whereas a more common approach these days is modelling ligand flexibility while assuming a totally or partially rigid protein receptor. Accordingly, search algorithms are classified in two main groups, leading to the so-called rigid docking and flexible docking (with respect to the ligand structure).[a]

Rigid ligand docking is almost not currently used since only the translational and rotational degrees of freedom are considered, hence excluding any kind of flexibility. The corresponding search algorithms are based on fitting the ligand into the protein binding site by matching characteristic features of the molecules in space. Such features may be geometric (e.g. volume segments of the protein or the ligand) or chemical (e.g. complementary interactions like a hydrogen bond). A more detailed description of these search algorithms may be found in the corresponding literature.[266]

[a] Some strategies have also been developed to account for protein flexibility, which have been briefly presented in section *3.2. Protein–ligand docking* of Chapter III.

By contrast, flexible ligand docking also considers the conformational degrees of freedom of the ligand. The main algorithms devised to treat ligand flexibility are classified into three general categories: systematic methods, random or stochastic methods, and simulation methods.[266–271]

2.1.1. Systematic methods

Systematic search algorithms can be further divided into three main types: conformational or exhaustive search methods, fragmentation methods and database or conformational ensemble methods.

Exhaustive search methods

These methods are based on systematically rotating all rotatable bonds in the ligand through 360º using a fixed increment, until all possible combinations have been generated and evaluated. Their major pitfall is that the number of structures produced increases immensely with the number of rotatable bonds, and so their application is actually reduced.

Fragmentation methods

One of the most commonly used approaches to handle ligand flexibility in molecular docking is fragmentation. In this method, the ligand is divided into multiple rigid fragments. There are two approaches to grow the ligand from these fragments: either by placing one fragment in the binding site and subsequently adding the remaining fragments (*incremental construction approach*, IC), or by docking all the fragments into the active site and linking them covalently (*place-and-join approach*). FlexX[286] and DOCK[375] are examples of docking programs that use a fragmentation search method.

Conformational ensemble methods

In these methods, an ensemble of ligand conformations is initially generated with another program and each conformer is then rigidly docked into the binding site. An example of a docking program that makes use of a conformational ensemble method is FLOG.[376]

2.1.2. Random methods

Random or stochastic search algorithms sample ligand binding orientations and conformations by performing random changes in the conformational, translational or rotational space of a single ligand or a population of ligands. The alteration performed at each step will be accepted or rejected according to a predefined probability function. There are four basic types of random algorithms: Monte Carlo (MC), Tabu Search (TS), Swarm Optimization methods (SO) and Genetic Algorithms (GA).

In the following subsections, all these methods will be briefly described with a special emphasis on Genetic Algorithms, as they are implemented in the molecular docking program used in the present thesis, GOLD[322] (Genetic Optimization for Ligand Docking).

Monte Carlo methods

In Monte Carlo methods, the probability to accept a newly obtained pose is based on a Boltzmann probability function. This function depends on the difference in the energy score of the ligand before and after the random change. Some programs that use MC methods include DockVision[377] and ICM.[378]

Tabu search methods

In Tabu search methods, the new orientations generated by random changes are stored in a so-called tabu list. These methods impose restrictions that prevent the search from revisiting previously explored areas of the conformational space. Thus, an alteration will be rejected if the resulting orientation is close to one already present in the tabu list, except if its scoring is better than the best scoring in the list. PRO_LEADS[323] is the most popular docking program that uses a Tabu search algorithm.

Swarm optimization methods

Similarly to the above described random methods, swarm optimization (SO) algorithms explore the search space by generating an ensemble of orientations that are randomly located in different regions of this search space.

SO methods are bio-inspired, and the initial set of orientations represents a population of animals. Then, mimicking the movement of a swarm of birds when one of them finds food, the orientations are distorted so that they move through the search space in the direction of the fittest individuals (i.e. the best scored orientations) of this population. These movements are repeated for a given number of iterations or until the scoring function converges.

Some programs that have implemented a SO method are PLANTS[379] and a variation of the AutoDock program, PSO@AutoDock,[380] among others.

Genetic algorithm methods and GOLD

Genetic algorithms[381–383] have been widely used to solve global optimization problems, such as the search for the best orientation of a ligand in a receptor binding site, by mimicking the process of evolution. According to Darwin's theory of evolution, natural selection is one of the basic mechanisms of evolution, which involves the survival of the most adapted organisms to the environment (*the fittest*) and the propagation of their characteristics to the next generation by reproduction. Mutations are also important in evolution, since they may lead to inclusion of characteristics not present in parent(s). Thus, the diversification of the population characteristics is related to recombination of parents' genetic material during reproduction, as well as to random changes in DNA (mutations). Based on these ideas, genetic algorithms aim at sampling the conformational space in a receptor by generating an initial set of orientations (mimicking a population of individuals) that is constantly refined in a manner imitating selection and adaptation in biological evolution, thus diversifying their characteristics to generate new orientations. To this aim, some aspects related to the process of evolution have to be re-defined and modelled:

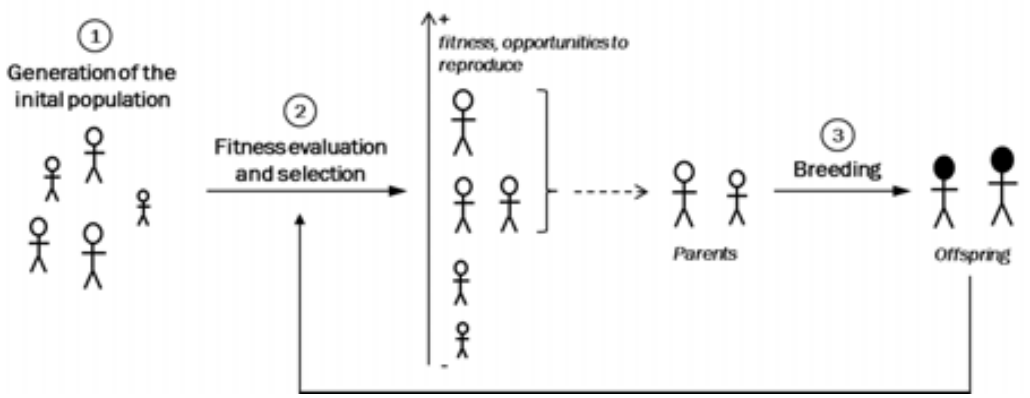
- *Individuals and genes*. Each individual is represented by a chromosome consisting of a sequence of genes. Individuals are different orientations of the ligand with respect to the protein. The variables or characteristics that define an orientation (e.g. conformation of the ligand, hydrogen bonds between the ligand and the receptor, mapping of hydrophobic points) are encoded in these genes, so that one gene corresponds to one molecular variable relevant for the problem under investigation.

- *Selection*. The fitness of each population member is related to its possibilities of survival and reproduction, mimicking natural selection. A so-called *fitness function*, which is normally a scoring function, is used to evaluate and determine this fitness, assigning a scoring to each individual.
- *Inheritance*. As in nature, individuals can combine their genes to pass them onto their offspring. The genetic operator of crossover or recombination is used to model this process, cutting two parents' chromosomes and recombining the resulting fragments to generate two children. In other words, by combining the variables of two initial orientations, two new orientations are generated.
- *Mutation*. The mutation operator simply perturbs a solution's chromosome by randomly changing the value of one of its variables, which may provide an influx of characteristics external to the population.

The general optimization scheme of genetic algorithms, which simulates the process of evolution, consists of three main steps (Scheme CM1):

- *Generation of the initial population* (step 1). An initial population of individuals is generated, wherein each individual is a different orientation of the ligand in the binding site.
- *Fitness evaluation and parent selection* (step 2). The fitness function is used to decide which individuals survive and produce offspring. Thus, each population member (orientation) is initially scored and a selection scheme is subsequently applied to make sure that fitter individuals (i.e. those with the best scoring) have more opportunities to reproduce.
- *Breeding* (step 3). Genetic operators of crossover and mutation are used to provide the first offspring generation. While the crossover operator combines parents' genes (variables), the random nature of the mutation operator may introduce some characteristics not exhibited by parents.

Steps 2 and 3 are then repeated to score children, who will replace the least-fit members of the population, and produce the next offspring generation. Iteration of these two steps is performed a certain amount of times or until the fitness function converges.



Scheme CM1. Representation of the general pattern of a genetic algorithm, wherein each individual represents an orientation of the ligand in the receptor binding site. Their sizes are proportional to their fitness score.

As mentioned, the program GOLD has been used in this thesis to perform all the protein–ligand docking studies. This program applies a genetic algorithm search strategy that follows the general optimization scheme above presented.[287,384] Some of its distinctive features are herein summarized:

- *Generation of the initial population* (step 1). An initial population of 500 individuals (orientations) is randomly created, which size will be maintained during all the optimization process. Rather than manipulate one large population of individuals, GOLD employs the so-called *island model*, which involves the creation of several subpopulations (islands) wherein independent genetic algorithm searches may be performed in parallel.[385] In particular, five subpopulations of 100 individuals each one are used. Each chromosome, representing one individual, encodes the internal conformation of the ligand and the protein binding site (by defining the torsional angle of each rotatable bond), as well as a mapping of hydrophobic points and hydrogen bond partners between protein and ligand. Remarkably, when using a library of rotameric states of certain side chains to account for protein flexibility, the specific rotamers corresponding to each orientation are also encoded in the chromosome, being defined by their torsional angle values.
- *Fitness evaluation and parent selection* (step 2). The fitness function corresponds to one of the scoring functions implemented in GOLD (namely GoldScore,[287,384] ChemScore,[278] ChemPLP,[386] and ASP[387]), according to the user selection. Depending on the chosen scoring function, certain terms are

evaluated to calculate the fitness score of each particular individual (e.g. hydrogen bonds, protein–ligand van der Waals interactions, ligand internal energy, protein–ligand clashes).[a] Parents are then randomly chosen with a probability proportional to their fitness score (i.e. the higher the fitness, the higher the probability of reproduction).

- *Breeding* (step 3). GOLD makes use of three genetic operators: crossover, mutation and migration. The two first have been previously described. The migration operator, which moves individuals from one island to another, is used to share information between subpopulations that are running independent searches. This operator requires one parent and produces one child (an exact copy of the parent), who is subsequently migrated to a different subpopulation.[385] Only one operator is applied to produce the offspring. Each operator is given a certain weight (assigned by default), and the selection of which one is used to generate the offspring is stochastically performed with a probability proportional to its weight.

The number of genetic operations (and thus, iterations of steps 2 and 3) performed on each docking run is automatically calculated by GOLD for a given ligand to make the search as efficient as possible, considering some parameters such as the ligand flexibility or the volume of the protein binding site, among others.[388] After the application of the fixed number of genetic operations, the algorithm terminates and outputs the highest scoring orientation.

It is worth recalling that one docking run corresponds to one predicted orientation. In GOLD, the number of docking runs to be performed on each ligand is 10 by default. Considering that each run may involve around 100,000 genetic operations, the number of orientations to be scored is, at least, 1,000,000. This vast amount of generated poses is common to many search algorithms. Therefore, scoring functions must contain several simplifications that should allow the evaluation of a ligand pose in a reasonable time, as it will be discussed in the following section *2.2. Scoring functions*.

The docking programs AutoDock,[279] DIVALI,[389] and DARWIN[390] also include a GA or GA-like algorithm.

[a] For a detailed discussion of scoring functions, see the following section *2.2.2. Scoring functions* of the present chapter.

2.1.3. Simulation methods

Molecular dynamics (MD) is a simulation technique that represents, for a given system at a given temperature, the movements of atoms throughout time. Newton's laws of motion together with energies derived from force fields are used to determine these atom trajectories.

One of the main drawbacks of MD is that simulations are quite time-consuming, which limits the sampling of the space for protein–ligand docking (conformational, translational and rotational) within a feasible amount of time. Besides, unrestrained MD simulations are well known to frequently get trapped in local energetic minima during this process. Some strategies have been developed to overcome this limitation, such as performing various simulations starting from different ligand positions, applying constraints between ligand and protein to force the exploration of a given part of the conformational space or mixing MD with molecular motions of high dimensionality like Normal Mode Analysis.[391] In general, MD simulations have been widely used before and after the protein–ligand docking: before docking, to optimize the protein receptor structure and account for protein flexibility; and after docking, to refine already docked complexes by incorporating solvent effects and flexibility of both the ligand and the receptor.

Remarkably, a significant advantage of random search algorithms over molecular dynamics is that these methods are more efficient in crossing high energy barriers, as their the sampling of the search space is based on the generation of an ensemble of discrete orientations all over the search space by random changes, instead of starting from an orientation and moving locally to orientations with lower energy.

2.2. Scoring functions

Scoring functions aim at evaluating and ranking the binding modes predicted on the basis of the search algorithm, so that they should enable the distinction between the true binding modes and all the alternative orientations explored. As mentioned, two important aspects of a scoring function are speed and accuracy, parameters that are inversely correlated.

Numerous scoring functions have been developed, and they can be divided into three major classes according to how they are derived: force field-based, empirical and knowledge-based.[268,271,392]

In the following sections, a short overview on each type of scoring functions will be given. The empirical scoring function used in this thesis, ChemScore,[278] will be presented in more detail in the corresponding section.

2.2.1. Force field-based scoring functions

As the name indicates, force field scoring functions are based on the decomposition of the binding energy into individual terms such as the ones previously presented for a common force field,[a] i.e. van der Waals and electrostatic energies, bond stretching, angle bending and torsional energies, among others. The set of parameters used by these scoring functions are taken from existing force fields (e.g. Amber).

A major challenge in force field-based scoring functions is how to treat the solvent in ligand binding.[268,271,392] One of the simplest methods is to use a distance-dependent dielectric constant $\varepsilon(r_{ij})$ in the electrostatic term.

Some scoring functions also include additional terms such as an explicit hydrogen-bonding term or an entropic contribution depending on the number of rotatable bonds.[279]

Examples of force field-based scoring functions include GoldScore[322] and DOCK[375] scoring functions.

2.2.2. Empirical scoring functions

In empirical scoring functions, the binding energy score of a complex is estimated on the basis of a set of weighted energy terms:

$$\Delta G = \sum_i W_i \cdot \Delta G_i \qquad (10)$$

where ΔG_i represents individual energy terms such as van der Waals energy, electrostatics, hydrogen bond, desolvation, etc. and W_i the corresponding coefficients. Experimentally determined binding affinity data of a set of protein–

[a] See section *2.1. Molecular mechanics* of the present chapter.

ligand complexes with known three-dimensional structures are used to determine the coefficients of the various terms, W_i, by means of a multivariate regression method.

Compared to force field scoring functions, empirical scoring functions are much faster in binding score calculations due to the easy computation of their energy terms. However, their general applicability depends on the experimental data set used in the parameterization process.

As mentioned, the empirical scoring function ChemScore[278] implemented in GOLD has been used in the present thesis. Some of its main features will be presented in the next paragraphs.

ChemScore function assigns general atom types to all ligand atoms and receptor atoms in contact with the ligand. The atom types are:

- *Lipophilic*. Chlorine, bromine and iodine atoms which are not ions; sulphurs which are not acceptor or polar types; carbons which are not polar type.
- *H-bond donor*. Nitrogen atoms with hydrogen attached; hydrogen atoms attached to oxygen or nitrogen.
- *H-bond donor/acceptor*. Oxygen atoms with hydrogen attached; imine nitrogen atoms.
- *H-bond acceptor*. Oxygen atoms not attached to hydrogen atoms; nitrogen atoms with no hydrogen atoms attached and one or two connections; halogens which are ions; sulphurs with only one connection (e.g. thioureas).
- *Polar (non H-bonding)*. Nitrogen atoms with no hydrogen atoms attached and more than two connections; phosphorus; sulphurs attached to one or more polar atoms (including H-bonding atoms and not including polar carbon atoms or fluorine atoms); carbons attached to two or more polar atoms (including H-bonding atoms and not including polar carbon atoms or fluorine atoms); carbons in nitriles or carbonyls; nitrogen atoms with no hydrogen atoms and four connections; fluorine atoms.
- *Metal*. Metal atoms.

Concerning the ChemScore equation, four separate terms (excluding the intercept) are included by taking into account hydrogen bonds, metal atoms, lipophilic effects of atoms and contributions from rotatable bonds that are frozen because of contact with the receptor.

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} \sum_{\mathrm{iI}} \mathrm{g}_1(\Delta r)\mathrm{g}_2(\Delta\alpha) + \Delta G_{metal} \sum_{\mathrm{aM}} \mathrm{f}(r_{aM}) + \Delta G_{lipo} \sum_{\mathrm{lL}} \mathrm{f}(r_{lL}) + \Delta G_{rot} H_{rot} \quad (11)$$

Equation (11) is the general ChemScore scoring function, where ΔG_0 is the intercept and the coefficients ΔG, which values are not shown in this general equation, were obtained by multiple linear regression. In the next paragraphs, each term will be shortly presented.

- *Hydrogen bonds* ($\sum_{il} g_1 g_2$). The hydrogen bond term is calculated for all possibilities of hydrogen bonds between ligand atoms, *i*, and protein atoms, *I*. The functions g_1 and g_2 depend, respectively, on the deviation of the hydrogen bond length ($\Delta r = r_{il} - r_0$) and D-H···A angle ($\Delta\alpha = \alpha_{il} - \alpha_0$; D: donor, A: acceptor) from their ideal values (r_0 = 1.85 Å, α_0 = 180º).
- *Metal atoms* ($\sum_{aM} f(r_{aM})$). The metal term is calculated for all acceptor and acceptor/donor atoms in the ligand, *a*, and any metal atoms in the receptor, *M*. The function $f(r_{aM})$ is a simple contact term that depends on the distance (r_{aM}) between ligand and receptor atoms of the appropriate type.
- *Lipophilic* ($\sum_{lL} f(r_{lL})$). The lipophilic term is calculated similarly to the metal term, with a function $f(r_{lL})$ depending on the distance (r_{lL}) between lipophilic ligand atoms, *l*, and lipophilic receptor atoms, *L*.
- *Rotatable bonds* (H_{rot}). The corresponding function estimates the flexibility penalty for molecules presenting frozen rotatable bonds. A rotatable bond is defined as any sp^3-sp^3 bond and any sp^2-sp^3 bond, excluding bonds to terminal CH_3, CF_3, NH_2 and NH_3 groups. Bonds are considered frozen if atoms on both sides of the rotatable bond are in contact with the receptor (i.e. the distance between any two heavy atoms is less than the sum of the van der Waals radii plus 0.5 Å).

The mathematical functions of each term may be found in the corresponding article.[278] ChemScore scoring function was calibrated using 82 protein–ligand complexes with known binding affinities and was tested against two other sets containing a further 20 and 10 complexes, respectively. The resulting function to estimate the free energy of binding, $\Delta G_{binding}$ (kJ·mol^{-1}), containing the calculated coefficients is shown in equation (12).

$$\Delta G_{binding} = -5.4800 + (-3.3400) \cdot S_{hbond} + (-6.0300) \cdot S_{metal} + (-0.1170) \cdot S_{lipo} + 2.5600 \cdot H_{rot} \quad (12)$$

where S_{hbond}, S_{metal} and S_{lipo} are the above commented scores for hydrogen bonds, metal and lipophilic atoms, respectively, and H_{rot} is the score representing the loss of conformational entropy of the ligand upon binding.

This ChemScore scoring function was then adapted for docking purposes,[322,323] and some terms deemed to be important to molecular docking were added to the regression-based part. The final form of the ChemScore function implemented in GOLD is as follows:

$$\Delta G'_{binding} = \Delta G_{binding} + E_{clash} + E_{int} + E_{cov} \quad (13)$$

- *Free energy of binding* ($\Delta G_{binding}$). The free energy of binding is calculated according to equation (12). However, to prevent unlikely hydrogen bond geometries, Baxter et al.[323] added (as a product) an extra angle term to the hydrogen bond term (S_{hbond}) that depends on the R-A···H(D) angle, β (R: atom attached to the acceptor; A: acceptor; D: donor). This term is analogous to the *r* and *α*-dependent functions above mentioned for the hydrogen bond term, that is, it depends on the difference between this angle, β, and its reference value.
- *Protein–ligand clashes* (E_{clash}). This term is calculated over all non-hydrogen protein–ligand atom pairs to penalize close contacts between these atoms. The corresponding function depends on the distance and nature (i.e. atom types) of the protein and ligand atom.
- *Internal energy of the ligand* (E_{int}). The internal energy of the ligand is a sum of a torsional and an internal clash term. The latter is calculated analogously to the protein–ligand clash energy term, but it is only evaluated for ligand atoms that are separated by at least four bonds. The torsional term depends on the deviation of the ligand torsion angles (both for rotatable and ring bonds) from their ideal values.
- *Covalent term* (E_{cov}). This term applies only to ligands bound covalently to the protein. It consists of a torsional and a bond-angle part. The torsional term is determined by the difference between torsion angles involved in the covalent bond and their reference values. Similarly, the bond-angle part depends on

the deviation of all bond angles around the covalent linkage from their ideal values.

The functional form of all these terms and reference values of the parameters used in these mathematical functions have been optimized according to a set of complexes with known binding affinities and structures, and are given in the corresponding literature.[322,323] It is noteworthy that the newly incorporated terms (protein–ligand clashes, internal energy of the ligand and covalent terms) are dimensionless penalty terms that prevent close contacts in docking and poor internal conformations. Hence, the final score ($\Delta G'_{binding}$) is also dimensionless. In the present thesis, this final score will be referred to as "binding energy" for the sake of clarity, despite it should not be considered as an accurate value of binding energy.

Examples of other empirical scoring functions are ChemPLP[386] and FlexX.[286]

2.2.3. Knowledge-based scoring functions

Knowledge-based scoring functions focus on predicting good orientations rather than binding affinities, trying to implicitly capture binding effects difficult to be explicitly modelled. These scoring functions employ energy potentials that are derived from the structural information contained in protein–ligand complexes of known structure available in the PDB.

The principle behind knowledge-based scoring functions is that the frequency of occurrence of individual contacts between protein and ligand atoms is related to their energetic contribution to binding. A certain contact that occurs more frequently than it should be expected from an average distribution indicates an attractive interaction, whereas when it occurs less frequently is indicative of a repulsion between the involved atoms.[393]

Hence, the occurrence frequency of each protein–ligand atom pair in a large dataset of complexes structures is evaluated and converted into atom pair potentials. The energy of ligand binding for a given complex is then calculated by summing the potential terms for all the protein–ligand atom pairs in the studied structure. Therefore, knowledge-based scoring functions are as fast as empirical scoring functions.

Examples of knowledge-based scoring functions include DrugScore[394,395] and GOLD/ASP.[387]

2.3. Protein–ligand docking software

More than 60 docking programs have been developed up to now featuring different combinations of search algorithms and scoring functions,[269] which reflects the lack of a search algorithm and a scoring function that generally give better results than the others.

Several studies have been devoted to evaluate the performance of the most common docking programs in terms of reproduction of X-ray structures, prediction of binding affinities, and ability to identify active compounds from a set of randomly chosen molecules.[268,269] However, it is difficult to draw conclusions of general applicability, since the results obtained for many programs are highly dependent on the protein considered. Overall, the majority of these comparative studies conclude that there is no single program that certainly outperforms the others.

Among all the existing docking programs, AutoDock is the most cited one,[a] followed by GOLD, DOCK, FlexX and ICM (Figure CM2).

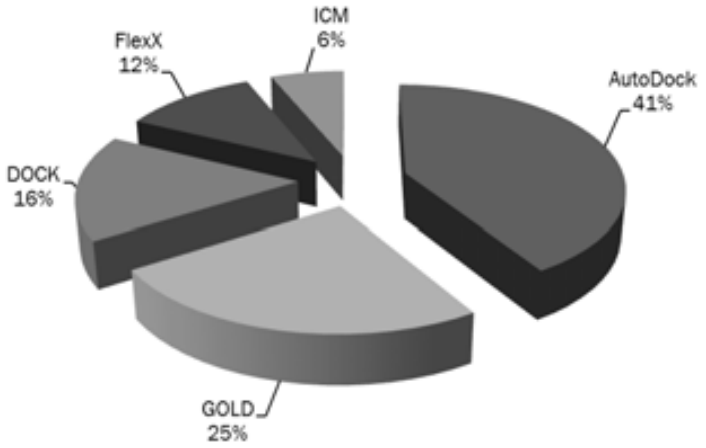


Figure CM2. Percentage of citations of the top five docking programs (from ISI Web of Knowledge, January 2013, considering any of the original references indicated in Table CM1).

The main features of the top five docking programs are summarized in Table CM1. There are no clear preferences regarding the type of search algorithm or scoring function implemented in these programs. Remarkably, AutoDock and DOCK

[a] ISI Web of Knowledge search on the articles citing any of the original references of each program indicated in Table CM1. (January 2013).

are freely available for academic institutions, converting GOLD in the most cited non-free software.

Table CM1. Most common docking programs' search algorithms, scoring functions and their distribution as free software.

Program	Search algorithm	Scoring function(s)	Free[a]
AutoDock[279,396–398] (version 4.2)	Random (GA-like)	Semi-empirical[b]	Yes
GOLD[287,322,384,399] (version 5.1)	Random (GA)	FF-based (GoldScore), empirical (ChemScore, ChemPLP), knowledge-based (ASP)	No
DOCK[272,375,400–402] (version 6)	Systematic (IC)	FF-based	Yes
FlexX[286]	Systematic (IC)	Empirical	No
ICM[378]	Random (MC)	FF-based	No

[a] Software available free of charge for academic institutions. [b] Semi-empirical scoring functions combine traditional molecular mechanics FF with empirical weights and/or empirical functional forms for some terms.[403]

BIBLIOGRAPHY

(1) Collier, L. H.; Oxford, J. S. *Human virology*; 3rd ed.; Oxford University Press, 2006.

(2) Cann, A. J. *Principles of molecular virology*; 5th ed.; Academic Press, 2011.

(3) Madigan, M. T.; Martinko, J. M.; Stahl, D. A.; Clark, D. P. *Brock Biology of Microorganisms*; 13th ed.; Pearson Education, 2011.

(4) Barré-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vézinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. *Science* **1983**, *220*, 868–871.

(5) Popovic, M.; Sarngadharan, M. G.; Read, E.; Gallo, R. C. *Science* **1984**, *224*, 497–500.

(6) UNAIDS. "Progress report 2011: global HIV/AIDS response". http://www.unaids.org (accessed Sep 27, 2012).

(7) Esté, J. A.; Cihlar, T. *Antiviral Res.* **2010**, *85*, 25–33.

(8) Broder, S. *Antiviral Res.* **2010**, *85*, 1–18.

(9) Brady, R. C.; Bernstein, D. I. *Antiviral Res.* **2004**, *61*, 73–81.

(10) Celum, C.; Levine, R.; Weaver, M.; Wald, A. *Bull. W. H. O.* **2004**, *82*, 447–453.

(11) Wald, A.; Link, K. *J. Infect. Dis.* **2002**, *185*, 45–52.

(12) Bulletin of the World Health Organization. http://www.who.int/bulletin/volumes/86/10/07-046128/en/ (accessed Sep 27, 2012).

(13) Arvin, A. M.; Greenberg, H. B. *Virology* **2006**, *344*, 240–249.

(14) De Clercq, E. *Future Virol.* **2008**, *3*, 393–405.

(15) Naeger, L. K.; Struble, K. A.; Murray, J. S.; Birnkrant, D. B. *Antiviral Res.* **2010**, *85*, 232–240.

(16) De Clercq, E. In *Advances in Antiviral Drug Design*; Elsevier Ltd, 2007; Vol. 5, pp. 1–58.

(17) Zolopa, A. R. *Antiviral Res.* **2010**, *85*, 241–244.

(18) Arribas, J. R.; Pozniak, A. L.; Gallant, J. E.; DeJesus, E.; Gazzard, B.; Campo, R. E.; Chen, S.-S.; McColl, D.; Holmes, C. B.; Enejosa, J.; Toole, J. J.; Cheng, A. K. *Acquir. Immune. Defic. Syndr.* **2008**, *47*, 74–78.

(19) Mehellou, Y.; De Clercq, E. *J. Med. Chem.* **2010**, *53*, 521–538.

(20) Mitsuya, H.; Weinhold, K. J.; Furman, P. A.; St Clair, M. H.; Lehrman, S. N.; Gallo, R. C.; Bolognesi, D.; Barry, D. W.; Broder, S. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7096–7100.

(21) De Clercq, E. *Antiviral Res.* **2010**, *85*, 19–24.

(22) Kulikowski, T. *Pharm. World Sci.* **1994**, *16*, 127–138.

(23) Périgaud, C.; Gosselin, G.; Imbach, J. L. *Nucleosides, Nucleotides Nucleic Acids* **1992**, *11*, 903–945.

(24) Martin, J. C.; Hitchcock, M. J. M.; De Clercq, E.; Prusoff, W. H. *Antiviral Res.* **2010**, *85*, 34–38.

(25) Razonable, R. R. *Mayo Clin. Proc.* **2011**, *86*, 1009–1026.

(26) Cihlar, T.; Ray, A. S. *Antiviral Res.* **2010**, *85*, 39–58.

(27) Asahchop, E. L.; Wainberg, M. A.; Sloan, R. D.; Tremblay, C. L. *Antimicrob. Agents Chemother.* **2012**, *56*, 5000–5008.

(28) Price, N. B.; Prichard, M. N. *Curr. Opin. Virol.* **2011**, *1*, 548–554.

(29) Prusoff, W. H. *Biochim. Biophys. Acta* **1959**, *32*, 295–296.

(30) Elion, G. B.; Furman, P. A.; Fyfe, J. A.; De Miranda, P.; Beauchamp, L.; Schaeffer, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5716–5720.

(31) Lin, T.; Shen, J. Z.; August, E. M.; Brankovan, V.; Yang, H.; Ghazzouli, I.; Prusoff, W. H. *J. Med. Chem.* **1989**, *32*, 1891–1895.

(32) Mansuri, M. M.; Starrett, J. E.; Ghazzouli, I.; Hitchcock, M. J. M.; Sterzycki, R. Z.; Brankovan, V.; Lin, T.-S.; August, E. M.; Prusoff, W. H.; Sommadossi, J.-P.; Martin, J. C. *J. Med. Chem.* **1989**, *32*, 461–466.

(33) Huryn, D. M.; Okabe, M. *Chem. Rev.* **1992**, *92*, 1745–1768.

(34) Smith, K. O.; Galloway, K. S.; Kennell, W. L.; Ogilvie, K. K.; Radatus, B. K. *Antimicrob. Agents Chemother.* **1982**, *22*, 55–61.

(35) Harnden, M. R.; Jarvest, R. L. *Tetrahedron Lett.* **1985**, *26*, 4265–4268.

(36) Colla, L.; De Clercq, E.; Busson, R.; Vanderhaeghe, H. *J. Med. Chem.* **1983**, *26*, 602–604.

(37) Pescovitz, M. D.; Rabkin, J.; Merion, R. M.; Paya, C. V; Pirsch, J.; Freeman, R. B.; O'Grady, J.; Robinson, C.; To, Z.; Wren, K.; Banken, L.; Buhles, W.; Brown, F. *Antimicrob. Agents Chemother.* **2000**, *44*, 2811–2815.

(38) Gill, K. S.; Wood, M. J. *Clin. Pharmacokinet.* **1996**, *31*, 1–8.

(39) Heidelberger, C.; Parsons, D.; Remy, D. C. *J. Am. Chem. Soc.* **1962**, *84*, 3597–3598.

(40) De Clercq, E.; Descamps, J.; De Somer, P.; Barr, P. J.; Jones, A. S.; Walker, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 2947–2951.

(41) Snoeck, R.; Sakuma, T.; De Clercq, E.; Rosenberg, I.; Holy, A. *Antimicrob. Agents Chemother.* **1988**, *32*, 1839–1844.

(42) Mitsuya, H.; Broder, S. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1911–1915.

(43) Lin, T. S.; Schinazi, R. F.; Prusoff, W. H. *Biochem. Pharmacol.* **1987**, *36*, 2713–2718.

(44) Yarchoan, R.; Mitsuya, H.; Thomas, R. V.; Pluda, J. M.; Hartman, N. R.; Perno, C. F.; Marczyk, K. S.; Allain, J. P.; Johns, D. G.; Broder, S. *Science* **1989**, *245*, 412–415.

(45) Tan, X.; Chu, C. K.; Boudinot, F. D. *Adv. Drug Delivery Rev.* **1999**, *39*, 117–151.

(46) Schinazi, R. F.; Chu, C. K.; Peck, A.; McMillan, A.; Mathis, R.; Cannon, D.; Jeong, L.-S.; Beach, J. W.; Choi, W.-B.; Yeola, S.; Liotta, D. C. *Antimicrob. Agents Chemother.* **1992**, *36*, 672–676.

(47) Schinazi, R. F.; McMillan, A.; Cannon, D.; Mathis, R.; Lloyd, R. M.; Peck, A.; Sommadossi, J.-P.; Clair, M. S.; Wilson, J.; Furman, P. A.; Painter, G.; Choi, W.-B.; Liotta, D. C. *Antimicrob. Agents Chemother.* **1992**, *36*, 2423–2431.

(48) Daluge, S. M.; Good, S. S.; Faletto, M. B.; Miller, W. H.; St Clair, M. H.; Boone, L. R.; Tisdale, M.; Parry, N. R.; Reardon, J. E.; Dornsife, R. E.; Averett, D. R.; Krenitsky, T. A. *Antimicrob. Agents Chemother.* **1997**, *41*, 1082–1093.

(49) Balzarini, J.; Holy, A.; Jindrich, J.; Naesens, L.; Snoeck, R.; Schols, D.; De Clercq, E. *Antimicrob. Agents Chemother.* **1993**, *37*, 332–338.

(50) Hawkins, T. *Antiviral Res.* **2010**, *85*, 201–209.

(51) Haubrich, R. H.; Riddler, S. A.; DiRienzo, A. G.; Komarow, L.; Powderly, W. G.; Klingman, K.; Garren, K. W.; Butcher, D. L.; Rooney, J. F.; Haas, D. W.; Mellors, J. W.; Havlir, D. V. *AIDS* **2009**, *23*, 1109–1118.

(52) Calmy, A.; Hirschel, B.; Cooper, D. A.; Carr, A. *Antiviral Ther.* **2009**, *14*, 165–179.

(53) Gallant, J. E.; Staszewski, S.; Pozniak, A. L.; DeJesus, E.; Suleiman, J. M. A. H.; Miller, M. D.; Coakley, D. F.; Lu, B.; Toole, J. J.; Cheng, A. K. *J. Am. Med. Assoc.* **2004**, *292*, 191–201.

(54) Chen, C.-H.; Vazquez-Padua, M.; Cheng, Y.-C. *Mol. Pharmacol.* **1991**, *39*, 625–628.

(55) El Safadi, Y.; Vivet-Boudou, V.; Marquet, R. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 723–737.

(56) Paredes, R.; Clotet, B. *Antiviral Res.* **2010**, *85*, 245–265.

(57) Menéndez-Arias, L. *Antiviral Res.* **2010**, *85*, 210–231.

(58) Little, S. J.; Holte, S.; Routy, J.-P.; Daar, E. S.; Markowitz, M.; Collier, A. C.; Koup, R. A.; Mellors, J. W.; Connick, E.; Conway, B.; Kilby, M.; Wang, L.; Withcomb, J. M.; Hellmann, N. S.; Richman, D. D. *N. Engl. J. Med.* **2002**, *347*, 385–394.

(59) Morfin, F.; Thouvenot, D. *J. Clin. Virol.* **2003**, *26*, 29–37.

(60) Collins, P. *Ann. Med.* **1993**, *25*, 441–445.

(61) Larder, B. A.; Darby, G.; Richman, D. D. *Science* **1989**, *234*, 1731–1734.

(62) Kuritzkes, D. R. *Curr. Opin. Virol.* **2011**, *1*, 582–589.

(63) Chevillotte, M.; Von Einem, J.; Meier, B. M.; Lin, F.-M.; Kestler, H. A.; Mertens, T. *Antiviral Res.* **2010**, *85*, 318–327.

(64) Boyer, P. L.; Clark, P. K.; Hughes, S. H. *J. Virol.* **2012**, *86*, 5885–5894.

(65) Kimberlin, D. W.; Coen, D. M.; Biron, K. K.; Cohen, J. I.; Lamb, R. A.; McKinlay, M.; Emini, E. A.; Whitley, R. J. *Antiviral Res.* **1995**, *26*, 369–401.

(66) Pelosi, E.; Mulamba, G. B.; Coen, D. M. *Antiviral Res.* **1998**, *37*, 17–28.

(67) Sarafianos, S. G.; Clark, A. D.; Das, K.; Tuske, S.; Birktoft, J. J.; Ilankumaran, P.; Ramesha, A. R.; Sayer, J. M.; Jerina, D. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E. *EMBO J.* **2002**, *21*, 6614–6624.

(68) Schuetz, J. D.; Connelly, M. C.; Sun, D.; Paibir, S. G.; Flynn, P. M.; Srinivas, R. V.; Kumar, A.; Fridland, A. *Nat. Med.* **1999**, *5*, 1048–1051.

(69) Erice, A. *Clin. Microbiol. Rev.* **1999**, *12*, 286–297.

(70) Delaney IV, W. E.; Locarnini, S.; Shaw, T. *Antiviral Chem. Chemother.* **2001**, *12*, 1–35.

(71) Larder, B. A.; Kemp, S. D. *Science* **1989**, *246*, 1155–1158.

(72) Arts, E. J.; Wainberg, M. A. *Antimicrob. Agents Chemother.* **1996**, *40*, 527–540.

(73) Kimberlin, D. W.; Whitley, R. J. *J. Antimicrob. Chemother.* **1996**, *37*, 403–421.

(74) Furman, P. A.; Fyfe, J. A.; St. Clair, M. H.; Weinhold, K.; Rideout, J. L.; Freeman, G. A.; Lehrman, S. N.; Bolognesi, D. P.; Broder, S.; Mitsuya, H.; Barry, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8333–8337.

(75) Deville-Bonne, D.; El Amri, C.; Meyer, P.; Chen, Y.; Agrofoglio, L. A.; Janin, J. *Antiviral Res.* **2010**, *86*, 101–120.

(76) Elion, G. B. *J. Antimicrob. Chemother.* **1983**, *12*, 9–17.

(77) Eriksson, S.; Munch-Petersen, B.; Johansson, K.; Eklund, H. *Cell. Mol. Life Sci.* **2002**, *59*, 1327–1346.

(78) Prichard, M. N.; Keith, K. A.; Johnson, M. P.; Harden, E. A.; McBrayer, A.; Luo, M.; Qiu, S.; Chattopadhyay, D.; Fan, X.; Torrence, P. F.; Kern, E. R. *Antimicrob. Agents Chemother.* **2007**, *51*, 1795–1803.

(79) Zhu, X. F. *Nucleosides, Nucleotides Nucleic Acids* **2000**, *19*, 651–690.

(80) Shealy, Y. F.; Clayton, J. D. *J. Am. Chem. Soc.* **1966**, *88*, 3885–3887.

(81) Shealy, Y. F.; Clayton, J. D. *J. Am. Chem. Soc.* **1969**, *91*, 3075–3083.

(82) Wachtmeister, J.; Classon, B.; Samuelsson, B. *Tetrahedron* **1995**, *51*, 2029–2038.

(83) Wachtmeister, J.; Classon, B.; Samuelsson, B. *Tetrahedron* **1997**, *53*, 1861–1872.

(84) Kusaka, T.; Yamamoto, H.; Shibata, M.; Muroi, M.; Kishi, T.; Mizuno, K. *J. Antibiot.* **1968**, *21*, 255–263.

(85) Yaginuma, S.; Muto, N.; Tsujino, M.; Sudate, Y.; Hayashi, M.; Otani, M. *J. Antibiot.* **1981**, *34*, 360–366.

(86) Marquez, V. E. In *Advances in Antiviral Drug Design*; Elsevier Ltd, 1996; Vol. 2, pp. 89–146.

(87) Ferrero, M.; Gotor, V. *Chem. Rev.* **2000**, *100*, 4319–4347.

(88) Jeong, L. S.; Lee, J. A. *Antiviral Chem. Chemother.* **2004**, *15*, 235–250.

(89) Wang, J.; Rawal, R. K.; Chu, C. K. In *Medicinal Chemistry of Nucleic Acids*; John Wiley & Sons, Inc., 2011; pp. 1–100.

(90) Crimmins, M. T. *Tetrahedron* **1998**, *54*, 9229–9272.

(91) Vince, R.; Hua, M.; Brownell, J.; Daluge, S.; Lee, F. C.; Shannon, W. M.; Lavelle, G. C.; Qualls, J.; Weislow, O. S.; Kiser, R.; Canonico, P. G.; Schultz, R. H.; Narayanan, V. L.; Mayo, J. G.; Shoemaker, R. H.; Boyd, M. R. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 1046–1053.

(92) Vince, R.; Brownell, J. *Biochem. Biophys. Res. Commun.* **1990**, *168*, 912–916.

(93) Carter, S. G.; Kessler, J. A.; Rankin, C. D. *Antimicrob. Agents Chemother.* **1990**, *34*, 1297–1300.

(94) Patil, S. D.; Schneller, S. W.; Hosoya, M.; Snoeck, R.; Andrei, G.; Balzarini, J.; De Clercq, E. *J. Med. Chem.* **1992**, *35*, 3372–3377.

(95) Sekiyama, T.; Hatsuya, S.; Tanaka, Y.; Uchiyama, M.; Ono, N.; Iwayama, S.; Oikawa, M.; Suzuki, K.; Okunishi, M.; Tsuji, T. *J. Med. Chem.* **1998**, *41*, 1284–1298.

(96) Wang, J.; Froeyen, M.; Hendrix, C.; Andrei, G.; Snoeck, R.; De Clercq, E.; Herdewijn, P. *J. Med. Chem.* **2000**, *43*, 736–745.

(97) Hoshino, H.; Shimizu, N.; Shimada, N.; Takita, T.; Takeuchi, T. *J. Antibiot.* **1987**, *40*, 1077–1078.

(98) Ichikawa, Y.; Narita, A.; Shiozawa, A.; Hayashi, Y.; Narasaka, K. *J. Chem. Soc., Chem. Commun.* **1989**, 1919–1921.

(99) Honjo, M.; Maruyama, T.; Sato, Y.; Horii, T. *Chem. Pharm. Bull.* **1989**, *37*, 1413–1415.

(100) Norbeck, D. W.; Kern, E.; Hayashi, S.; Rosenbrook, W.; Sham, H.; Herrin, T.; Plattner, J. J.; Erickson, J.; Clement, J.; Swanson, R.; Shipkowitz, N.; Hardy, D.; Marsh, K.; Arnett, G.; Shannon, W.; Broder, S.; Mitsuya, H. *J. Med. Chem.* **1990**, *33*, 1285–1288.

(101) Gourdel-Martin, M.-E.; Huet, F. *J. Org. Chem.* **1997**, *62*, 2166–2172.

(102) Gourdel-Martin, M.-E.; Huet, F. *Nucleosides, Nucleotides Nucleic Acids* **1999**, *18*, 645–648.

(103) Hubert, C.; Alexandre, C.; Aubertin, A.-M.; Huet, F. *Tetrahedron* **2002**, *58*, 3775–3778.

(104) Maruyama, T.; Hanai, Y.; Sato, Y. *Nucleosides, Nucleotides Nucleic Acids* **1992**, *11*, 855–864.

(105) Maury, G. *Antiviral Chem. Chemother.* **2000**, *11*, 165–190.

(106) Smejkal, J.; Sorm, F. *Collect. Czech. Chem. Commun.* **1964**, *29*, 2809–2813.

(107) Wang, P.; Hong, J. H.; Cooperwood, J. S.; Chu, C. K. *Antiviral Res.* **1998**, *40*, 19–44.

(108) Gumina, G.; Song, G. Y.; Chu, C. K. *FEMS Microbiol. Lett.* **2001**, *202*, 9–15.

(109) Mathé, C.; Gosselin, G. *Antiviral Res.* **2006**, *71*, 276–281.

(110) Graciet, J.-C. G.; Schinazi, R. F. In *Advances in Antiviral Drug Design*; Elsevier Ltd, 1999; Vol. 3, pp. 1–68.

(111) Jarvis, B.; Faulds, D. *Drugs* **1999**, *58*, 101–141.

(112) Coates, J. A. V.; Cammack, N.; Jenkinson, H. J.; Mutton, I. M.; Pearson, B. A.; Storer, R.; Cameron, J. M.; Penn, C. R. *Antimicrob. Agents Chemother.* **1992**, *36*, 202–205.

(113) Shewach, D. S.; Liotta, D. C.; Schinazi, R. F. *Biochem. Pharmacol.* **1993**, *45*, 1540–1543.

(114) Lee, K.; Chu, C. K. *Antimicrob. Agents Chemother.* **2001**, *45*, 138–144.

(115) Hart, G. J.; Orr, D. C.; Penn, C. R.; Figueiredo, H. T.; Gray, N. M.; Boehme, R. E.; Cameron, J. M. *Antimicrob. Agents Chemother.* **1992**, *36*, 1688–1694.

(116) Pélicano, H.; Pierra, C.; Eriksson, S.; Gosselin, G.; Imbach, J.-L.; Maury, G. *J. Med. Chem.* **1997**, *40*, 3969–3973.

(117) Chang, C.-N.; Doong, S.-L.; Zhou, J. H.; Beach, J. W.; Jeong, L. S.; Chu, C. K.; Tsai, C. H.; Cheng, Y. C. *J. Biol. Chem.* **1992**, *267*, 13938–13942.

(118) Kim, J. W.; Park, S. H.; Louie, S. G. *Ann. Pharmacother.* **2006**, *40*, 472–478.

(119) Biron, K. K.; Harvey, R. J.; Chamberlain, S. C.; Good, S. S.; Smith III, A. A.; Davis, M. G.; Talarico, C. L.; Miller, W. H.; Ferris, R.; Dornsife, R. E.; Stanat, S. C.; Drach, J. C.; Townsend, L. B.; Koszalka, G. W. *Antimicrob. Agents Chemother.* **2002**, *46*, 2365–2372.

(120) Chu, C. K.; Ma, T.; Shanmuganathan, K.; Wang, C.; Xiang, Y.; Pai, S. B.; Yao, G. Q.; Sommadossi, J. P.; Cheng, Y. C. *Antimicrob. Agents Chemother.* **1995**, *39*, 979–981.

(121) Chen, S.-H. *Curr. Med. Chem.* **2002**, *9*, 899–912.

(122) Asif, G.; Hurwitz, S. J.; Gumina, G.; Chu, C. K.; McClure, H. M.; Schinazi, R. F. *Antimicrob. Agents Chemother.* **2005**, *49*, 560–564.

(123) Grove, K. L.; Guo, X.; Liu, S.-H.; Gao, Z.; Chu, C. K.; Cheng, Y.-C. *Cancer Res.* **1995**, *55*, 3008–3011.

(124) Binns, F.; Hayes, R.; Ingham, S.; Saengchantara, S. T.; Turner, R. W.; Wallace, T. W. *Tetrahedron* **1992**, *48*, 515–530.

(125) Binns, F.; Hayes, R.; Hodgetts, K. J.; Saengchantara, S. T.; Wallace, T. W.; Wallis, C. J. *Tetrahedron* **1996**, *52*, 3631–3658.

(126) Gauvry, N.; Huet, F. *J. Org. Chem.* **2001**, *66*, 583–588.

(127) Trost, B. M.; McDougal, P. G. *J. Org. Chem.* **1984**, *49*, 458–468.

(128) Vos, G. J. M.; Benders, P. H.; Reinhoudt, D. N.; Egberink, R. J. M.; Harkema, S.; Van Hummel, G. J. *J. Org. Chem.* **1986**, *51*, 2004–2011.

(129) Cooper, W.; Walters, W. D. *J. Am. Chem. Soc.* **1958**, *80*, 4220–4224.

(130) Niwayama, S.; Wangt, Y.; Houk, K. N. *Tetrahedron Lett.* **1995**, *36*, 6201–6204.

(131) Niwayama, S.; Kallel, E. A.; Spellmeyer, D. C.; Sheu, C.; Houk, K. N. *J. Org. Chem.* **1996**, *61*, 2813–2825.

(132) Dolbier, W. R.; Koroniak, H.; Houk, K. N.; Sheu, C. *Acc. Chem. Res.* **1996**, *29*, 471–477.

(133) Murakami, M.; Hasegawa, M.; Igawa, H. *J. Org. Chem.* **2004**, *69*, 587–590.

(134) Sheldrake, H. M.; Wallace, T. W.; Wilson, C. P. *Org. Lett.* **2005**, *7*, 4233–4236.

(135) Gauvry, N.; Lescop, C.; Huet, F. *Eur. J. Org. Chem.* **2006**, 5207–5218.

(136) Hasegawa, M.; Murakami, M. *J. Org. Chem.* **2007**, *72*, 3764–3769.

(137) Hasegawa, M.; Usui, I.; Konno, S.; Murakami, M. *Org. Biomol. Chem.* **2010**, *8*, 4169–4175.

(138) Pichon, C.; Hubert, C.; Alexandre, C.; Huet, F. *Tetrahedron: Asymmetry* **2000**, *11*, 2429–2434.

(139) Marsac, Y.; Nourry, A.; Legoupy, S.; Pipelier, M.; Dubreuil, D.; Aubertin, A.-M.; Bourgougnon, N.; Benhida, R.; Huet, F. *Tetrahedron* **2005**, *61*, 7607–7612.

(140) Ciamician, G.; Silber, P. *Chem. Ber.* **1908**, *41*, 1928–1935.

(141) Büchi, G.; Goldman, I. M. *J. Am. Chem. Soc.* **1957**, *79*, 4741–4748.

(142) Corey, E. J.; Bass, J. D.; LeMahieu, R.; Mitra, R. B. *J. Am. Chem. Soc.* **1964**, *86*, 5570–5583.

(143) Corey, E. J.; Mitra, R. B.; Uda, H. *J. Am. Chem. Soc.* **1964**, *86*, 485–492.

(144) Eaton, P. E. *J. Am. Chem. Soc.* **1962**, *84*, 2344–2348.

(145) Eaton, P. E. *Acc. Chem. Res.* **1968**, *1*, 50–57.

(146) De Mayo, P.; Reid, S. T.; Yip, R. W. *Can. J. Chem.* **1964**, *42*, 2828–2835.

(147) De Mayo, P. *Acc. Chem. Res.* **1971**, *4*, 41–47.

(148) Loutfy, R. O.; De Mayo, P. *J. Am. Chem. Soc.* **1977**, *99*, 3559–3565.

(149) Lee-Ruff, E.; Mladenova, G. *Chem. Rev.* **2003**, *103*, 1449–1483.

(150) Namyslo, J. C.; Kaufmann, D. E. *Chem. Rev.* **2003**, *103*, 1485–1537.

(151) Iriondo-Alberdi, J.; Greaney, M. F. *Eur. J. Org. Chem.* **2007**, 4801–4815.

(152) Hoffmann, N. *Chem. Rev.* **2008**, *108*, 1052–1103.

(153) Bach, T.; Hehn, J. P. *Angew. Chem. Int. Ed.* **2011**, *50*, 1000–1045.

(154) Wiesner, K.; Poon, L.; Jirkovský, I.; Fishman, M. *Can. J. Chem.* **1969**, *47*, 433–444.

(155) Miyano, K.; Ohfune, Y.; Azuma, S.; Matsumoto, T. *Tetrahedron Lett.* **1974**, *15*, 1545–1548.

(156) Faure, S.; Piva, O. *Tetrahedron Lett.* **2001**, *42*, 255–259.

(157) Mehta, G.; Sreenivas, K. *Tetrahedron Lett.* **2002**, *43*, 3319–3321.

(158) Webster, F. X.; Silverstein, R. M. *J. Org. Chem.* **1986**, *51*, 5226–5231.

(159) Alibés, R.; Bourdelande, J. L.; Font, J.; Parella, T. *Tetrahedron* **1996**, *52*, 1279–1292.

(160) Mori, K.; Sasaki, M. *Tetrahedron* **1980**, *36*, 2197–2208.

(161) Alibés, R.; De March, P.; Figueredo, M.; Font, J.; Racamonde, M.; Parella, T. *Org. Lett.* **2004**, *6*, 1449–1452.

(162) Ogino, T.; Yamada, K.; Isogai, K. *Tetrahedron Lett.* **1977**, *18*, 2445–2448.

(163) Tolbert, L. M.; Ali, M. B. *J. Am. Chem. Soc.* **1982**, *104*, 1742–1744.

(164) Lange, G. L.; Decicco, C.; Tan, S. L.; Chamberlain, G. *Tetrahedron Lett.* **1985**, *26*, 4707–4710.

(165) Demuth, M.; Palomer, A.; Sluma, H.-D.; Dey, A. K.; Krüger, C.; Tsay, Y.-H. *Angew. Chem. Int. Ed.* **1986**, *25*, 1117–1119.

(166) Lange, G. L.; Decicco, C.; Lee, M. *Tetrahedron Lett.* **1987**, *28*, 2833–2836.

(167) Lange, G. L.; Organ, M. G. *Tetrahedron Lett.* **1993**, *34*, 1425–1428.

(168) García-Expósito, E.; Álvarez-Larena, A.; Branchadell, V.; Ortuño, R. M. *J. Org. Chem.* **2004**, *69*, 1120–1125.

(169) Piers, E.; Orellana, A. *Synthesis* **2001**, 2138–2142.

(170) Williams, J. R.; Callahan, J. F. *J. Org. Chem.* **1980**, *45*, 4479–4483.

(171) Bauslaugh, P. G. *Synthesis* **1970**, *2*, 287–300.

(172) Schuster, D. I.; Greenberg, M. M.; Nuñez, I. M.; Tucker, P. C. *J. Org. Chem.* **1983**, *48*, 2615–2619.

(173) Schuster, D. I.; Dunn, D. A.; Heibel, G. E.; Brown, P. B.; Rao, J. M.; Woning, J.; Bonneau, R. *J. Am. Chem. Soc.* **1991**, *113*, 6245–6255.

(174) Kaprinidis, N. A.; Lem, G.; Courtney, S. H.; Schuster, D. I. *J. Am. Chem. Soc.* **1993**, *115*, 3324–3325.

(175) Hastings, F. J.; Weedon, A. C. *J. Am. Chem. Soc.* **1991**, *113*, 8525–8527.

(176) Andrew, D.; Weedon, A. C. *J. Am. Chem. Soc.* **1995**, *117*, 5647–5663.

(177) Schuster, D. I.; Lem, G.; Kaprinidis, N. A. *Chem. Rev.* **1993**, *93*, 3–22.

(178) Tada, M.; Kokubo, T.; Sato, T. *Tetrahedron* **1972**, *28*, 2121–2125.

(179) Kosugi, H.; Sekiguchi, S.; Sekita, R.; Uda, H. *Bull. Chem. Soc. Jpn.* **1976**, *49*, 520–528.

(180) Ohga, K.; Matsuo, T. *Bull. Chem. Soc. Jpn.* **1976**, *49*, 1590–1594.

(181) Tomioka, K.; Tanaka, M.; Koga, K. *Tetrahedron Lett.* **1982**, *23*, 3401–3404.

(182) Tomioka, K.; Tanaka, M.; Koga, K. *Chem. Pharm. Bull.* **1989**, *37*, 1201–1207.

(183) Avetisyan, A. A.; Margaryan, A. K.; Nalbandyan, G. K.; Avetisyan, T. V. *Zh. Org. Khim.* **1989**, *25*, 530–536.

(184) Hoffmann, N.; Scharf, H.-D.; Runsink, J. *Tetrahedron Lett.* **1989**, *30*, 2637–2638.

(185) Fillol, L.; Miranda, M. A.; Morera, I. M.; Sheikh, H. *Heterocycles* **1990**, *31*, 751–782.

(186) D'Annibale, A.; D'Auria, M.; Mancini, G.; Pace, A. D.; Racioppi, R. *Eur. J. Org. Chem.* **2012**, 785–791.

(187) Houk, K. N. *Chem. Rev.* **1976**, *76*, 1–74.

(188) Alibés, R.; Bourdelande, J. L.; Font, J. *Tetrahedron: Asymmetry* **1991**, *2*, 1391–1402.

(189) Alibés, R.; Bourdelande, J. L.; Font, J. *Tetrahedron Lett.* **1993**, *34*, 7455–7458.

(190) Alibés, R.; Bourdelande, J. L.; Font, J. *Tetrahedron Lett.* **1994**, *35*, 2587–2588.

(191) Alibés, R.; Bourdelande, J. L.; Font, J.; Gregori, A. *J. Braz. Chem. Soc.* **1995**, *6*, 119–121.

(192) Alibés, R.; Bourdelande, J. L.; Font, J.; Gregori, A.; Parella, T. *Tetrahedron* **1996**, *52*, 1267–1278.

(193) Alibés, R.; De March, P.; Figueredo, M.; Font, J.; Racamonde, M. *Tetrahedron Lett.* **2001**, *42*, 6695–6697.

(194) Alibés, R.; De March, P.; Figueredo, M.; Font, J.; Fu, X.; Racamonde, M.; Álvarez-Larena, A.; Piniella, J. F. *J. Org. Chem.* **2003**, *68*, 1283–1289.

(195) Gauvry, N.; Comoy, C.; Lescop, C.; Huet, F. *Synthesis* **1999**, *4*, 574–576.

(196) Alibés, R.; De March, P.; Figueredo, M.; Font, J.; Racamonde, M.; Rustullet, A.; Álvarez-Larena, A.; Piniella, J. F.; Parella, T. *Tetrahedron Lett.* **2003**, *44*, 69–71.

(197) Neumann, W. P. *Synthesis* **1987**, 665–683.

(198) Shimo, T.; Somekawa, K.; Kuwakino, J.; Uemura, H.; Kumamoto, S.; Tsuge, O.; Kanemasa, S. *Chem. Lett.* **1984**, 1503–1504.

(199) Pitt, I. G.; Russell, R. A.; Warrener, R. N. *J. Am. Chem. Soc.* **1985**, *107*, 7176–7178.

(200) Akhtar, M.; Nigel P., B.; Mark A., C.; Gani, D. *Tetrahedron* **1987**, *43*, 5899–5908.

(201) Inoue, M.; Sato, T.; Hirama, M. *Angew. Chem. Int. Ed.* **2006**, *45*, 4843–4848.

(202) Zhang, F.; Simpkins, N. S.; Blake, A. J. *Org. Biomol. Chem.* **2009**, *7*, 1963–1979.

(203) Alibés, R.; Alvárez-Larena, A.; De March, P.; Figueredo, M.; Font, J.; Parella, T.; Rustullet, A. *Org. Lett.* **2006**, *8*, 491–494.

(204) Flores, R.; Rustullet, A.; Alibés, R.; Álvarez-Larena, A.; De March, P.; Figueredo, M.; Font, J. *J. Org. Chem.* **2011**, *76*, 5369–5383.

(205) Figueras, A.; Miralles-Llumà, R.; Flores, R.; Rustullet, A.; Busqué, F.; Figueredo, M.; Font, J.; Alibés, R.; Maréchal, J.-D. *ChemMedChem* **2012**, *7*, 1044–1056.

(206) Rustullet, A.; Alibés, R.; De March, P.; Figueredo, M.; Font, J. *Org. Lett.* **2007**, *9*, 2827–2830.

(207) Racamonde, M.; Alibés, R.; Figueredo, M.; Font, J.; De March, P. *J. Org. Chem.* **2008**, *73*, 5944–5952.

(208) Mann, J.; Partlett, N. K.; Thomas, A. *J. Chem. Res. (S)* **1987**, 369.

(209) Rustullet, A. Síntesi estereoselectiva d'anàlegs ciclobutànics de nucleòsids, UAB, 2006.

(210) Kropp, P. J. *J. Am. Chem. Soc.* **1973**, *95*, 4611–4619.

(211) Bartlett, P. D.; Roof, A. A. M.; Winter, W. J. *J. Am. Chem. Soc.* **1981**, *103*, 6520–6522.

(212) Lidström, P.; Tierney, J.; Wathey, B.; Westman, J. *Tetrahedron* **2001**, *57*, 9225–9283.

(213) Kappe, C. O. *Angew. Chem. Int. Ed.* **2004**, *43*, 6250–6284.

(214) Kappe, C. O. *Chem. Soc. Rev.* **2008**, *37*, 1127–1139.

(215) Baig, R. B. N.; Varma, R. S. *Chem. Soc. Rev.* **2012**, *41*, 1559–1584.

(216) Kappe, C. O.; Dallinger, D.; Murphree, S. *Practical microwave synthesis for organic chemists*; Wiley-VCH, 2009.

(217) Kappe, C. O.; Stadler, A.; Dallinger, D. *Microwaves in organic and medicinal chemistry*; Wiley-VCH, 2012.

(218) Gedye, R.; Smith, F.; Westaway, K.; Ali, H.; Baldisera, L.; Laberge, L.; Rousell, J. *Tetrahedron Lett.* **1986**, *27*, 279–282.

(219) Giguere, R. J.; Bray, T. L.; Duncan, S. M. *Tetrahedron Lett.* **1986**, *27*, 4945–4948.

(220) Gabriel, C.; Gabriel, S.; Grant, E. H.; Halstead, B. S. J.; Mingos, D. M. P. *Chem. Soc. Rev.* **1998**, *27*, 213–223.

(221) CEM corporation. http://cem.com (accessed Oct 24, 2012).

(222) Mascitti, V.; Corey, E. J. *J. Am. Chem. Soc.* **2004**, *126*, 15664–15665.

(223) Kelebekli, L.; Kara, Y.; Balci, M. *Carbohydr. Res.* **2005**, *340*, 1940–1948.

(224) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871.

(225) Kotecha, N. R.; Ley, S. V.; Mantegani, S. *Synlett* **1992**, 395–399.

(226) Mitsunobu, O. *Synthesis* **1981**, 1–28.

(227) Schenk, S.; Weston, J.; Anders, E. *J. Am. Chem. Soc.* **2005**, *127*, 12566–12576.

(228) But, T. Y. S.; Toy, P. H. *Chem. Asian J.* **2007**, *2*, 1340–1355.

(229) Jenny, T. F.; Previsani, N.; Benner, S. A. *Tetrahedron Lett.* **1991**, *32*, 7029–7032.

(230) Swamy, K. C. K.; Kumar, N. N. B.; Balaraman, E.; Kumar, K. V. P. P. *Chem. Rev.* **2009**, *109*, 2551–2651.

(231) Bondada, L.; Gumina, G.; Nair, R.; Ning, X. H.; Schinazi, R. F.; Chu, C. K. *Org. Lett.* **2004**, *6*, 2531–2534.

(232) Dembinski, R. *Eur. J. Org. Chem.* **2004**, 2763–2772.

(233) Mévellec, L.; Huet, F. *Tetrahedron* **1997**, *53*, 5797–5812.

(234) Nair, V.; Jeon, G.-S. *ARKIVOC* **2004**, 133–140.

(235) Kim, J. W.; Ko, O. H.; Hong, J. H. *Arch. Pharm. Res.* **2005**, *28*, 745–749.

(236) Kim, A.; Hong, J. H. *Nucleosides, Nucleotides Nucleic Acids* **2006**, *25*, 941–950.

(237) Cruickshank, K. A.; Jiricny, J.; Reese, C. B. *Tetrahedron Lett.* **1984**, *25*, 681–684.

(238) Frieden, M.; Giraud, M.; Reese, C. B.; Song, Q. *J. Chem. Soc., Perkin Trans. 1* **1998**, 2827–2832.

(239) Fresneda, M. Á.; Alibés, R.; Font, J.; Bayón, P.; Figueredo, M. *J. Org. Chem.* **2012**, *77*, 5030–5035.

(240) Pérez-Pérez, M.-J.; Rozenski, J.; Busson, R.; Herdewijn, P. *J. Org. Chem.* **1995**, *60*, 1531–1537.

(241) Wang, J.; Morral, J.; Hendrix, C.; Herdewijn, P. *J. Org. Chem.* **2001**, *66*, 8478–8482.

(242) Barral, K.; Courcambeck, J.; Pèpe, G.; Balzarini, J.; Neyts, J.; De Clercq, E.; Camplo, M. *J. Med. Chem.* **2005**, *48*, 450–456.

(243) Diederichsen, U.; Weicherding, D.; Diezemann, N. *Org. Biomol. Chem.* **2005**, *3*, 1058–1066.

(244) Jeong, L. S.; Lee, J. A.; Kim, H. O.; Tosh, D. K.; Moon, H. R.; Lee, S.-J.; Lee, K. M.; Sheen, Y. Y.; Chun, M. W. *Nucleosides, Nucleotides Nucleic Acids* **2007**, *26*, 1021–1024.

(245) Fletcher, S.; Shahani, V. M.; Gunning, P. T. *Tetrahedron Lett.* **2009**, *50*, 4258–4261.

(246) Hocková, D.; Holý, A.; Masojídková, M.; Keough, D. T.; De Jersey, J.; Guddat, L. W. *Bioorg. Med. Chem.* **2009**, *17*, 6218–6232.

(247) Liu, L. J.; Yoo, J. C.; Hong, J. H. *Nucleosides, Nucleotides Nucleic Acids* **2009**, *28*, 150–164.

(248) Sundaram, G. S. M.; Harpstrite, S. E.; Kao, J. L.-F.; Collins, S. D.; Sharma, V. *Org. Lett.* **2012**, *14*, 3568–3571.

(249) Dey, S.; Garner, P. *J. Org. Chem.* **2000**, *65*, 7697–7699.

(250) Harnden, M. R.; Parkin, A.; Parratt, M. J.; Perkins, R. M. *J. Med. Chem.* **1993**, *36*, 1343–1355.

(251) Shendage, D. M.; Fröhlich, R.; Haufe, G. *Org. Lett.* **2004**, *6*, 3675–3678.

(252) Jacobsen, M. F.; Knudsen, M. M.; Gothelf, K. V. *J. Org. Chem.* **2006**, *71*, 9183–9190.

(253) Patrick, G. L. *An introduction to medicinal chemistry*; 3rd ed.; Oxford University Press, 2005.

(254) Leach, A. R. *Molecular modelling: principles and applications*; 2nd ed.; Pearson Education, 2001.

(255) Kirchmair, J.; Distinto, S.; Liedl, K. R.; Markt, P.; Rollinger, J. M.; Schuster, D.; Spitzer, G. M.; Wolber, G. *Infect. Disord.: Drug Targets* **2011**, *11*, 64–93.

(256) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Eur. J. Biochem.* **1977**, *80*, 319–324.

(257) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(258) Walters, W. P.; Murcko, M. A. *Adv. Drug Delivery Rev.* **2002**, *54*, 255–271.

(259) Lipinski, C. A. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249.

(260) Walters, W. P.; Ajay; Murcko, M. A. *Curr. Opin. Chem. Biol.* **1999**, *3*, 384–387.

(261) Muegge, I. *Med. Res. Rev.* **2003**, *23*, 302–321.

(262) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

(263) Lipinski, C. A. *Drug Discovery Today: Technol.* **2004**, *1*, 337–341.

(264) Kerns, E. H.; Di, L. *Drug-like Properties: Concepts, Structure Design and Methods*; Elsevier, 2008.

(265) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615–2623.

(266) Muegge, I.; Rarey, M. In *Reviews in Computational Chemistry*; John Wiley & Sons, Inc., 2001; Vol. 17, pp. 1–60.

(267) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.

(268) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 15–26.

(269) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. *Br. J. Pharmacol.* **2008**, *153*, S7–S26.

(270) Corbeil, C. R.; Therrien, E.; Moitessier, N. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 241–263.

(271) Huang, S.-Y.; Zou, X. *Int. J. Mol. Sci.* **2010**, *11*, 3016–3034.

(272) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.

(273) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.

(274) Jorgensen, W. L. *Science* **2004**, *303*, 1813–1818.

(275) Clark, D. E. *Expert Opin. Drug Discov.* **2006**, *1*, 103–110.

(276) 4SC AG. http://www.4sc.de/ (accessed Jan 22, 2013).

(277) Leban, J.; Saeb, W.; Garcia, G.; Baumgartner, R.; Kramer, B. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 55–58.

(278) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

(279) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(280) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(281) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(282) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput. Chem.* **2003**, *24*, 1637–1656.

(283) Birch, L.; Murray, C. W.; Hartshorn, M. J.; Tickle, I. J.; Verdonk, M. L. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 855–869.

(284) De Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. E. *J. Med. Chem.* **2005**, *48*, 2308–2318.

(285) Roberts, B. C.; Mancera, R. L. *J. Chem. Inf. Model.* **2008**, *48*, 397–408.

(286) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470–489.

(287) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.

(288) Friesner, R. a; Banks, J. L.; Murphy, R. B.; Halgren, T. a; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(289) Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213–218.

(290) Teodoro, M. L.; Phillips Jr, G. N.; Kavraki, L. E. *IEEE International Conference on Robotics and Automation* **2001**.

(291) Carlson, H. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447–452.

(292) Tuffery, P.; Derreumaux, P. *J. R. Soc. Interface* **2012**, *9*, 20–33.

(293) Fischer, E. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2984–2993.

(294) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–104.

(295) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.

(296) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; De Groot, B. L. *Science* **2008**, *320*, 1471–1475.

(297) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789–796.

(298) Jiang, F.; Kim, S.-H. *J. Mol. Biol.* **1991**, *219*, 79–102.

(299) Leach, A. R. *J. Mol. Biol.* **1994**, *235*, 345–356.

(300) Sandak, B.; Wolfson, H. J.; Nussinov, R. *Proteins: Struct., Funct., Bioinf.* **1998**, *32*, 159–174.

(301) Pontarin, G.; Gallinaro, L.; Ferraro, P.; Reichard, P.; Bianchi, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12159–12164.

(302) Garces, E.; Cleland, W. W. *Biochemistry* **1969**, *8*, 633–640.

(303) Lascu, I.; Gonin, P. *J. Bioenerg. Biomembr.* **2000**, *32*, 237–246.

(304) Cherfils, J.; Moréra, S.; Lascu, I.; Véron, M.; Janin, J. *Biochemistry* **1994**, *33*, 9062–9069.

(305) Schaertl, S.; Konrad, M.; Geeves, M. A. *J. Biol. Chem.* **1998**, *273*, 5662–5669.

(306) Chen, Y.; Gallois-Montbrun, S.; Schneider, B.; Véron, M.; Moréra, S.; Deville-Bonne, D.; Janin, J. *J. Mol. Biol.* **2003**, *332*, 915–926.

(307) Schneider, B.; Xu, Y. W.; Sellam, O.; Sarfati, R.; Janin, J.; Veron, M.; Deville-Bonne, D. *J. Biol. Chem.* **1998**, *273*, 11491–11497.

(308) Xu, Y.; Sellam, O.; Moréra, S.; Sarfati, S.; Biondi, R.; Véron, M.; Janin, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7162–7165.

(309) Mehellou, Y.; Balzarini, J.; McGuigan, C. *Org. Biomol. Chem.* **2009**, *7*, 2548–2553.

(310) Herman, B. D.; Votruba, I.; Holý, A.; Sluis-Cremer, N.; Balzarini, J. *J. Biol. Chem.* **2010**, *285*, 12101–12108.

(311) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.

(312) Barreca, M. L.; Balzarini, J.; Chimirri, A.; De Clercq, E.; De Luca, L.; Höltje, H. D.; Höltje, M.; Monforte, A. M.; Monforte, P.; Pannecouque, C.; Rao, A.; Zappalà, M. *J. Med. Chem.* **2002**, *45*, 5410–5413.

(313) Zhou, Z.; Madura, J. D. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 493–503.

(314) Ragno, R.; Artico, M.; De Martino, G.; La Regina, G.; Coluccia, A.; Di Pasquali, A.; Silvestri, R. *J. Med. Chem.* **2005**, *48*, 213–223.

(315) Lu, X.; Chen, Y.; Guo, Y.; Liu, Z.; Shi, Y.; Xu, Y.; Wang, X.; Zhang, Z.; Liu, J. *Bioorg. Med. Chem.* **2007**, *15*, 7399–7407.

(316) Lagos, C. F.; Caballero, J.; Gonzalez-Nilo, F. D.; Pessoa-Mahana, C. D.; Perez-Acle, T. *Chem. Biol. Drug Des.* **2008**, *72*, 360–369.

(317) Pawar, V. S.; Lokwani, D. K.; Bhandari, S. V.; Bothara, K. G.; Chitre, T. S.; Devale, T. L.; Modhave, N. S.; Parikh, J. K. *Med. Chem. Res.* **2010**, *20*, 370–380.

(318) La Regina, G.; Coluccia, A.; Brancale, A.; Piscitelli, F.; Gatti, V.; Maga, G.; Samuele, A.; Pannecouque, C.; Schols, D.; Balzarini, J.; Novellino, E.; Silvestri, R. *J. Med. Chem.* **2011**, *54*, 1587–1598.

(319) Ma, X.-D.; He, Q.-Q.; Zhang, X.; Yang, S.-Q.; Yang, L.-M.; Gu, S.-X.; Zheng, Y.-T.; Chen, F.-E.; Dai, H.-F. *Eur. J. Med. Chem.* **2012**, *58*, 504–512.

(320) Li, D.; Zhan, P.; Liu, H.; Pannecouque, C.; Balzarini, J.; De Clercq, E.; Liu, X. *Bioorg. Med. Chem.* **2013**, *21*, 2128–2134.

(321) Marquez, V. E.; Ben-Kasus, T.; Barchi, J. J.; Green, K. M.; Nicklaus, M. C.; Agbaria, R. *J. Am. Chem. Soc.* **2004**, *126*, 543–549.

(322) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins: Struct., Funct., Bioinf.* **2003**, *52*, 609–623.

(323) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins: Struct., Funct., Bioinf.* **1998**, *33*, 367–382.

(324) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(325) Marvin 5.4, 2010, ChemAxon (http://www.chemaxon.com).

(326) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.

(327) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(328) JChem for Excel 5.6, 2011, ChemAxon (http://www.chemaxon.com).

(329) Ostermann, N.; Lavie, A.; Padiyar, S.; Brundiers, R.; Veit, T.; Reinstein, J.; Goody, R. S.; Konrad, M.; Schlichting, I. *J. Mol. Biol.* **2000**, *304*, 43–53.

(330) Welin, M.; Kosinska, U.; Mikkelsen, N.-E.; Carnrot, C.; Zhu, C.; Wang, L.; Eriksson, S.; Munch-Petersen, B.; Eklund, H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17970–17975.

(331) Birringer, M. S.; Claus, M. T.; Folkers, G.; Kloer, D. P.; Schulz, G. E.; Scapozza, L. *FEBS Lett.* **2005**, *579*, 1376–1382.

(332) Vonrhein, C.; Schlauderer, G. J.; Schulz, G. E. *Structure* **1995**, *3*, 483–490.

(333) Ostermann, N.; Schlichting, I.; Brundiers, R.; Konrad, M.; Reinstein, J.; Veit, T.; Goody, R. S.; Lavie, A. *Structure* **2000**, *8*, 629–642.

(334) Moréra, S.; Lacombe, M. L.; Xu, Y.; LeBras, G.; Janin, J. *Structure* **1995**, *3*, 1307–1314.

(335) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. *Science* **1992**, *256*, 1783–1790.

(336) Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. *Science* **1998**, *282*, 1669–1675.

(337) Fyfe, J. A.; Keller, P. M.; Furman, P. A.; Miller, R. L.; Elion, G. B. *J. Biol. Chem.* **1978**, *253*, 8721–8727.

(338) Champness, J. N.; Bennett, M. S.; Wien, F.; Visse, R.; Summers, W. C.; Herdewijn, P.; De Clerq, E.; Ostrowski, T.; Jarvest, R. L.; Sanderson, M. R. *Proteins: Struct., Funct., Bioinf.* **1998**, *32*, 350–361.

(339) Pilger, B. D.; Perozzo, R.; Alber, F.; Wurth, C.; Folkers, G.; Scapozza, L. *J. Biol. Chem.* **1999**, *274*, 31967–31973.

(340) Chen, M. S.; Prusoff, W. H. *J. Biol. Chem.* **1978**, *253*, 1325–1327.

(341) Bohman, C.; Balzarini, J.; Wigerinck, P.; Van Aerschot, A.; Herdewijn, P.; De Clercq, E. *J. Biol. Chem.* **1994**, *269*, 8036–8043.

(342) Martić, M.; Pernot, L.; Westermaier, Y.; Kraljević, T. G.; Krištafor, S.; Raić-Malić, S.; Scapozza, L.; Ametamey, S. *Nucleosides, Nucleotides Nucleic Acids* **2011**, *30*, 293–315.

(343) Balzarini, J.; Bohman, C.; De Clercq, E. *J. Biol. Chem.* **1993**, *268*, 6332–6337.

(344) Comin, M. J.; Vu, B. C.; Boyer, P. L.; Liao, C.; Hughes, S. H.; Marquez, V. E. *ChemMedChem* **2008**, *3*, 1129–1134.

(345) Schelling, P.; Claus, M. T.; Johner, R.; Marquez, V. E.; Schulz, G. E.; Scapozza, L. *J. Biol. Chem.* **2004**, *279*, 32832–32838.

(346) Chen, M. S.; Walker, J.; Prusoff, W. H. *J. Biol. Chem.* **1979**, *254*, 10747–10753.

(347) Perozzo, R.; Jelesarov, I.; Bosshard, H. R.; Folkers, G.; Scapozza, L. *J. Biol. Chem.* **2000**, *275*, 16139–16145.

(348) Sekulic, N.; Shuvalova, L.; Spangenberg, O.; Konrad, M.; Lavie, A. *J. Biol. Chem.* **2002**, *277*, 30236–30243.

(349) Liu, S.; Knafels, J. D.; Chang, J. S.; Waszak, G. A.; Baldwin, E. T.; Deibel, M. R.; Thomsen, D. R.; Homa, F. L.; Wells, P. A.; Tory, M. C.; Poorman, R. A.; Gao, H.; Qiu, X.; Seddon, A. P. *J. Biol. Chem.* **2006**, *281*, 18193–18200.

(350) Pospisil, P.; Pilger, B. D.; Marveggio, S.; Schelling, P.; Wurth, C.; Scapozza, L.; Folkers, G.; Pongracic, M.; Mintas, M.; Malic, S. R. *Helv. Chim. Acta* **2002**, *85*, 3237–3250.

(351) Comin, M. J.; Agbaria, R.; Ben-Kasus, T.; Huleihel, M.; Liao, C.; Sun, G.; Nicklaus, M. C.; Deschamps, J. R.; Parrish, D. A.; Marquez, V. E. *J. Am. Chem. Soc.* **2007**, *129*, 6216–6222.

(352) Dey, D.; Evans, G. R. D. In *Targets in gene therapy*; InTech, 2011; pp. 65–76.

(353) Wild, K.; Bohner, T.; Folkers, G.; Schulz, G. E. *Protein Sci.* **1997**, *6*, 2097–2106.

(354) De Winter, H.; Herdewijn, P. *J. Med. Chem.* **1996**, *39*, 4727–4737.

(355) Alber, F.; Kuonen, O.; Scapozza, L.; Folkers, G.; Carloni, P. *Proteins: Struct., Funct., Bioinf.* **1998**, *4531*, 453–459.

(356) Sulpizi, M.; Schelling, P.; Folkers, G.; Carloni, P.; Scapozza, L. *J. Biol. Chem.* **2001**, *276*, 21692–21697.

(357) Stierand, K.; Rarey, M. *ACS Med. Chem. Lett.* **2010**, *1*, 540–545.

(358) Bennett, M. S.; Wien, F.; Champness, J. N.; Batuwangala, T.; Rutherford, T.; Summers, W. C.; Sun, H.; Wright, G.; Sanderson, M. R. *FEBS Lett.* **1999**, *443*, 121–125.

(359) Prota, A.; Vogt, J.; Pilger, B.; Perozzo, R.; Wurth, C.; Marquez, V. E.; Russ, P.; Schulz, G. E.; Folkers, G.; Scapozza, L. *Biochemistry* **2000**, *39*, 9597–9603.

(360) Brown, D. G.; Visse, R.; Sandhu, G.; Davies, A.; Rizkallah, P. J.; Melitz, C.; Summers, W. C.; Sanderson, M. R. *Nat. Struct. Mol. Biol.* **1995**, *2*, 876–881.

(361) Luyten, I.; De Winter, H.; Busson, R.; Lescrinier, T.; Crenven, I.; Durant, F.; Balzarini, J.; De Clercq, E.; Herdewijn, P. *Helv. Chim. Acta* **1996**, *79*, 1462–1474.

(362) Marquez, V. E.; Choi, Y.; Comin, M. J.; Russ, P.; George, C.; Huleihel, M.; Ben-Kasus, T.; Agbaria, R. *J. Am. Chem. Soc.* **2005**, *127*, 15145–15150.

(363) Schneider, B.; Babolat, M.; Xu, Y. W.; Janin, J.; Véron, M.; Deville-Bonne, D. *Eur. J. Biochem.* **2001**, *268*, 1964–1971.

(364) Walters, W. P.; Stahl, M. T.; Murcko, M. A. *Drug Discovery Today* **1998**, *3*, 160–178.

(365) Shoichet, B. K. *Nature* **2004**, *432*, 862–865.

(366) Klebe, G. *Drug Discovery Today* **2006**, *11*, 580–594.

(367) Fillat, C.; Carrió, M.; Cascante, A.; Sangro, B. *Curr. Gene Ther.* **2003**, *3*, 13–26.

(368) Herdewijn, P.; De Clercq, E. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1591–1597.

(369) Ferrer, E.; Alibés, R.; Busqué, F.; Figueredo, M.; Font, J.; De March, P. *J. Org. Chem.* **2009**, *74*, 2425–2432.

(370) Newman, M. S.; Evans, F. J. *J. Am. Chem. Soc.* **1955**, *77*, 946–947.

(371) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.

(372) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752–781.

(373) Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714–3717.

(374) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Bioinf.* **2001**, *44*, 400–417.

(375) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. *RNA* **2009**, *15*, 1219–1230.

(376) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 153–174.

(377) Hart, T. N.; Read, R. J. *Proteins: Struct., Funct., Bioinf.* **1992**, *13*, 206–222.

(378) Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comput. Chem.* **1994**, *15*, 488–506.

(379) Korb, O.; Stützle, T.; Exner, T. E. In *Lecture Notes in Computer Science*; Springer, 2006; pp. 247–258.

(380) Namasivayam, V.; Günther, R. *Chem. Biol. Drug Des.* **2007**, *70*, 475–484.

(381) Clark, D. E.; Westhead, D. R. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 337–358.

(382) Devillers, J. In *Genetic Algorithms in Molecular Modeling*; Elsevier Ltd, 1996; pp. 1–34.

(383) Venkatasubramanian, V.; Sundaram, A. In *Encyclopedia of computational chemistry*; John Wiley & Sons, Inc., 1998; pp. 1115–1127.

(384) Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43–53.

(385) Jones, G.; Willett, P.; Glen, R. C. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532–549.

(386) Korb, O.; Stützle, T.; Exner, T. E. *J. Chem. Inf. Model.* **2009**, *49*, 84–96.

(387) Mooij, W. T. M.; Verdonk, M. L. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 272–287.

(388) Using Automatic (Ligand-Dependent) Genetic Algorithm Parameter Settings. http://www.ccdc.cam.ac.uk/support/documentation/gold/5_1/gold/gold.1.155.html (accessed Jan 10, 2013).

(389) Clark, K. P. *J. Comput. Chem.* **1995**, *16*, 1210–1226.

(390) Taylor, J. S.; Burnett, R. M. *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 173–191.

(391) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. *Med. Res. Rev.* **2006**, *26*, 531–568.

(392) Huang, S.-Y.; Grinter, S. Z.; Zou, X. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899–12908.

(393) Böhm, H.-J. In *Protein-Ligand Interactions. From Molecular Recognition to Drug Design*; Wiley-VCH, 2003; pp. 3–20.

(394) Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337–356.

(395) Velec, H. F. G.; Gohlke, H.; Klebe, G. *J. Med. Chem.* **2005**, *48*, 6296–6303.

(396) Goodsell, D. S.; Olson, A. J. *Proteins: Struct., Funct., Bioinf.* **1990**, *8*, 195–202.

(397) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293–304.

(398) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

(399) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. *J. Med. Chem.* **2005**, *48*, 6504–6515.

(400) Ewing, T. J. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175–1189.

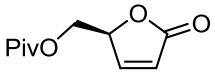
(401) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.

(402) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601–619.

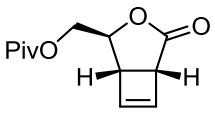
(403) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. *J. Comput. Chem.* **2007**, *28*, 1145–1152.

(404) Zhang, Y.; Featherstone, D.; Davis, W.; Rushton, E.; Broadie, K. *J. Cell Sci.* **2000**, *113*, 3103–3115.

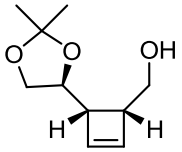
Formula Index



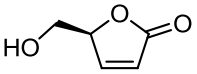
4



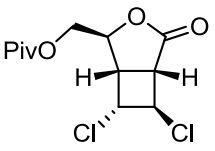
5



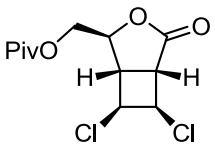
6



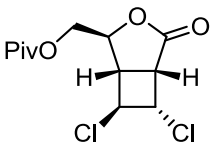
60



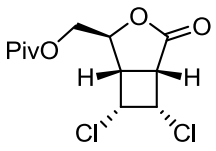
73



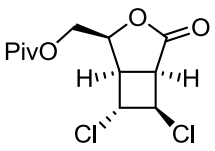
74



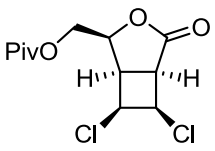
75



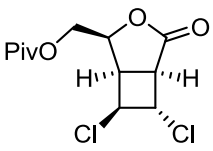
76



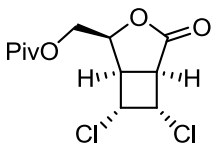
77



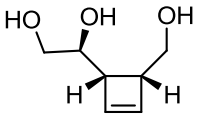
78



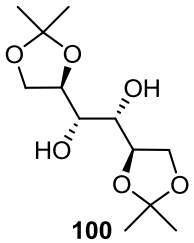
79



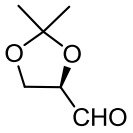
80



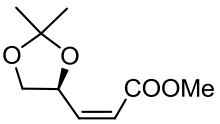
98



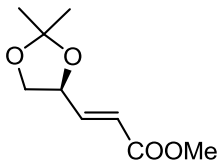
100



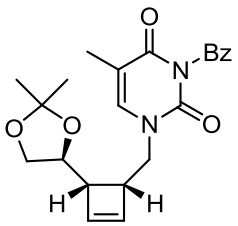
101



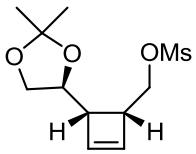
102



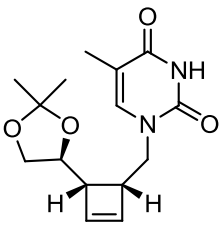
103



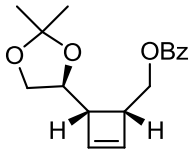
117



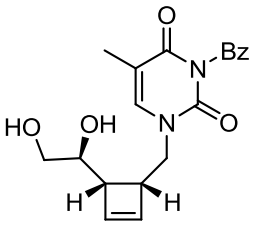
126



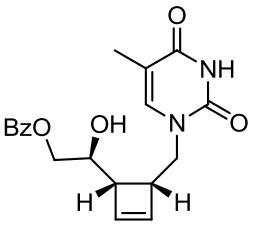
127



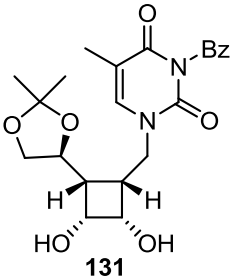
128



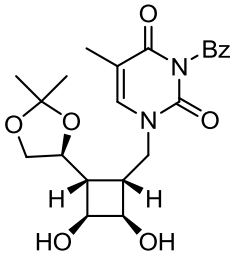
129



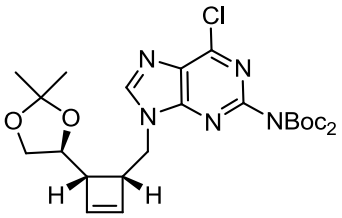
130



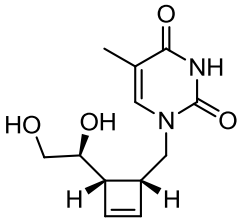
131



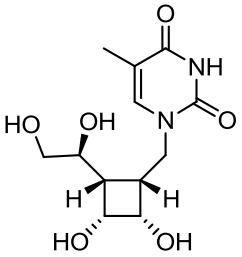
132



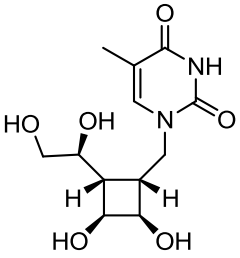
133



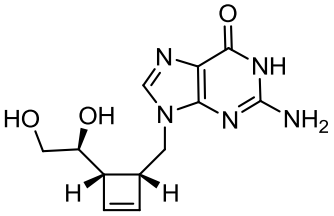
HS1-T



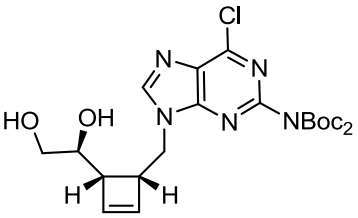
HS2-T



HS2'-T



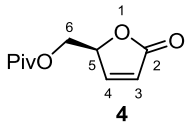
HS1-G



HS1-G^{CI}

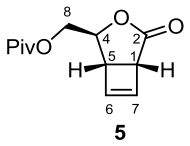
APPENDIX A

NMR spectra



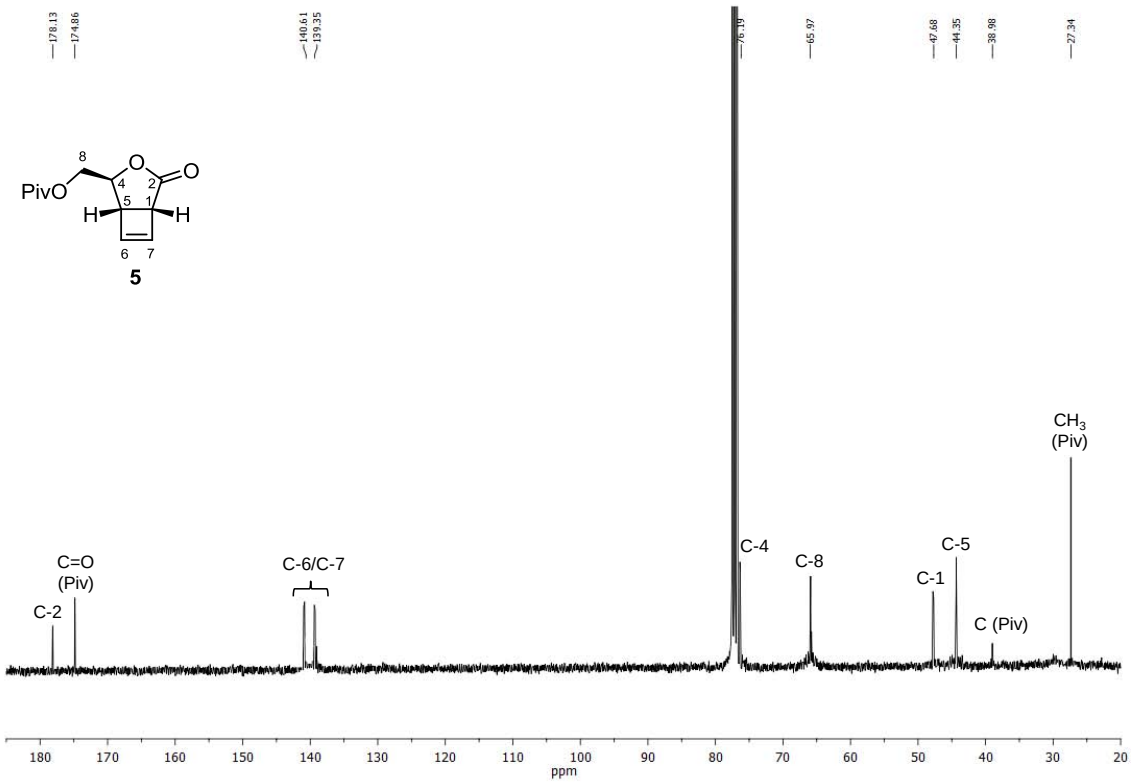
4

^{1}H-NMR (250 MHz, $CDCl_3$)

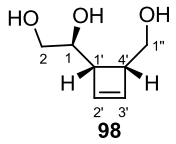


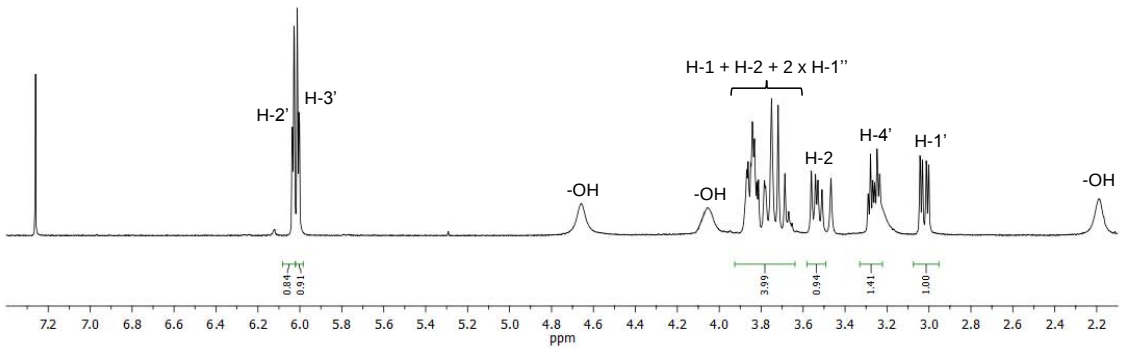
5

^{1}H-NMR (250 MHz, $CDCl_3$)

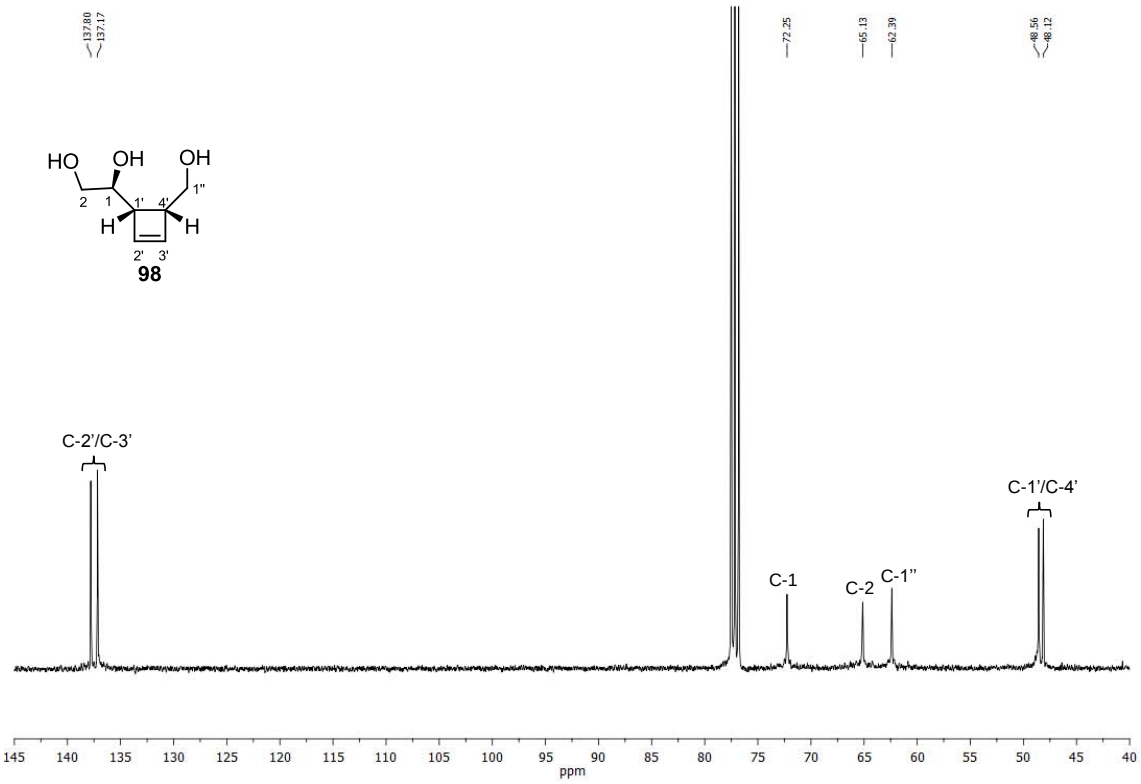


^{13}C-NMR (62.5 MHz, $CDCl_3$)

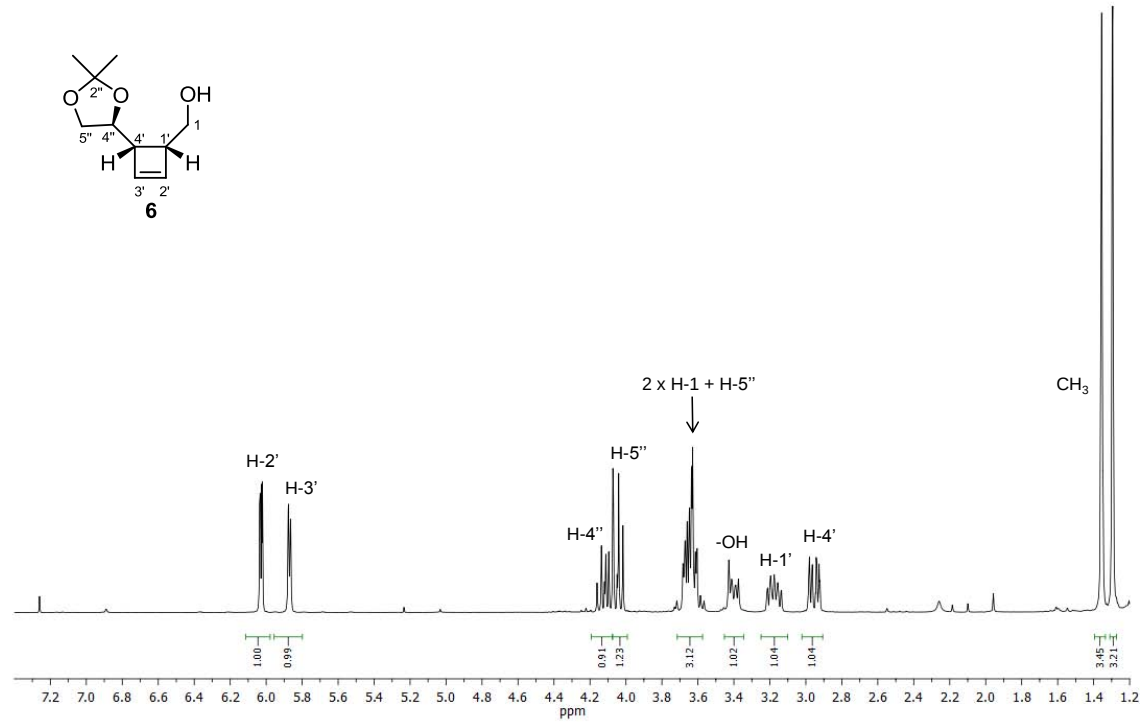




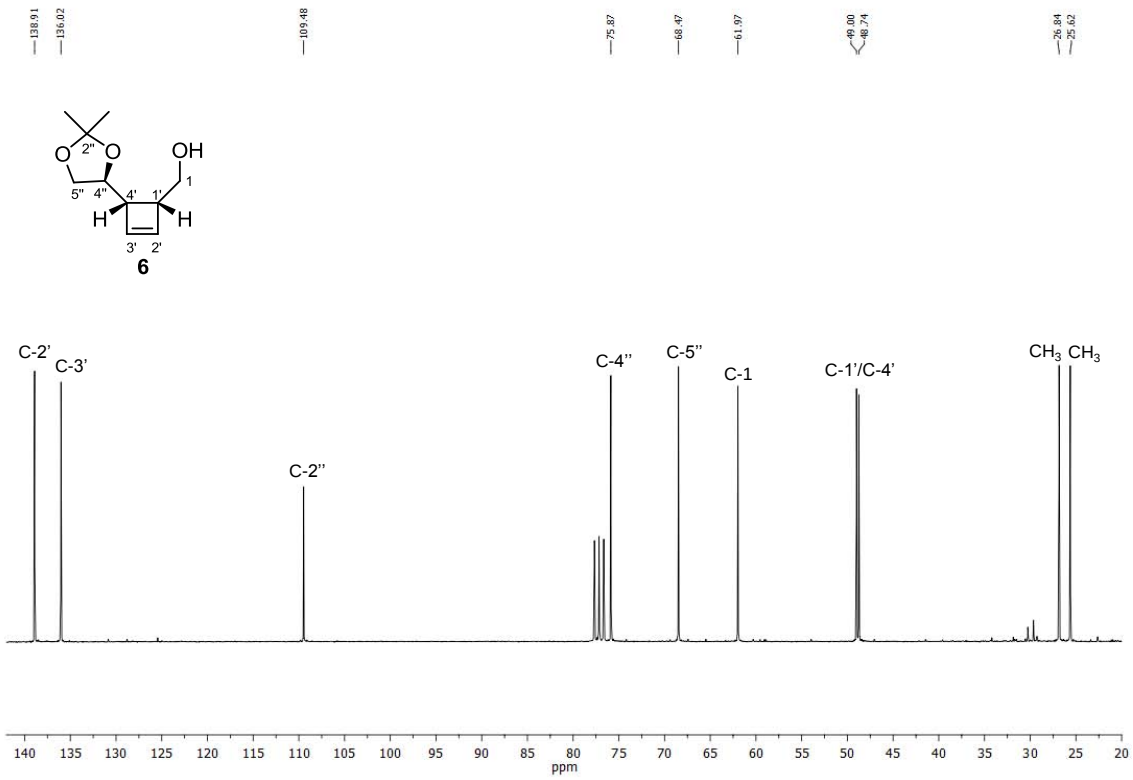
^{1}H-NMR (360 MHz, $CDCl_3$)



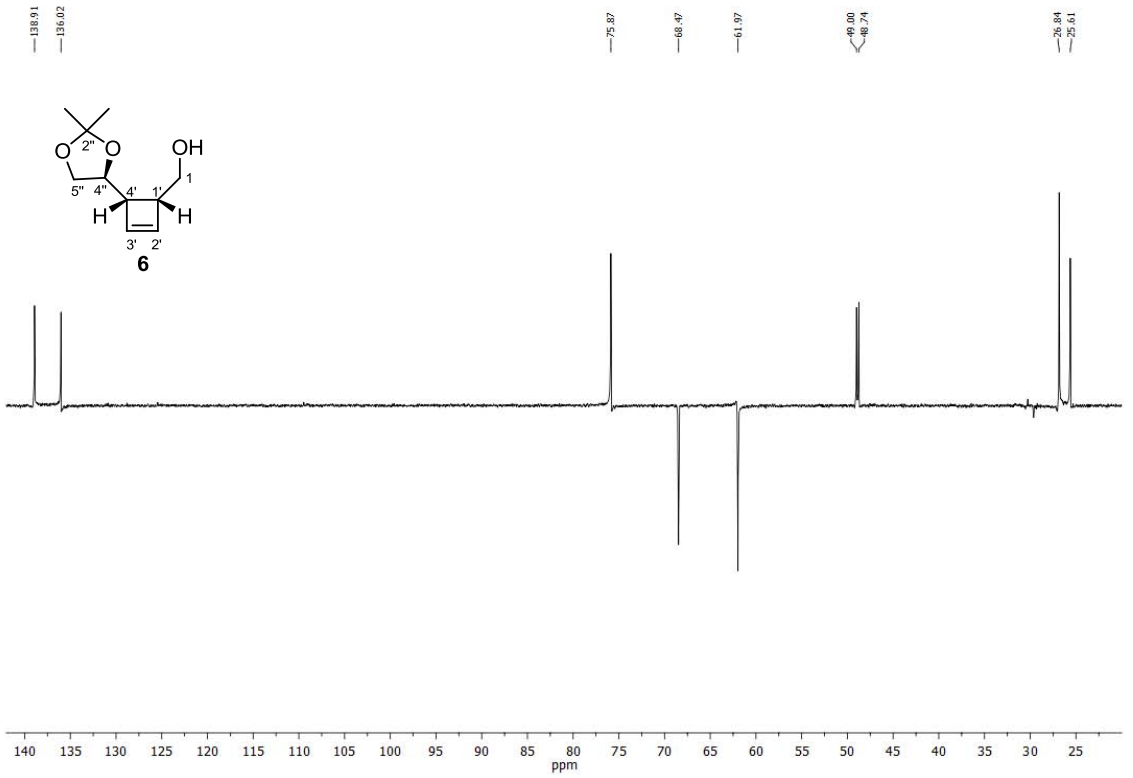
^{13}C-NMR (90 MHz, $CDCl_3$)



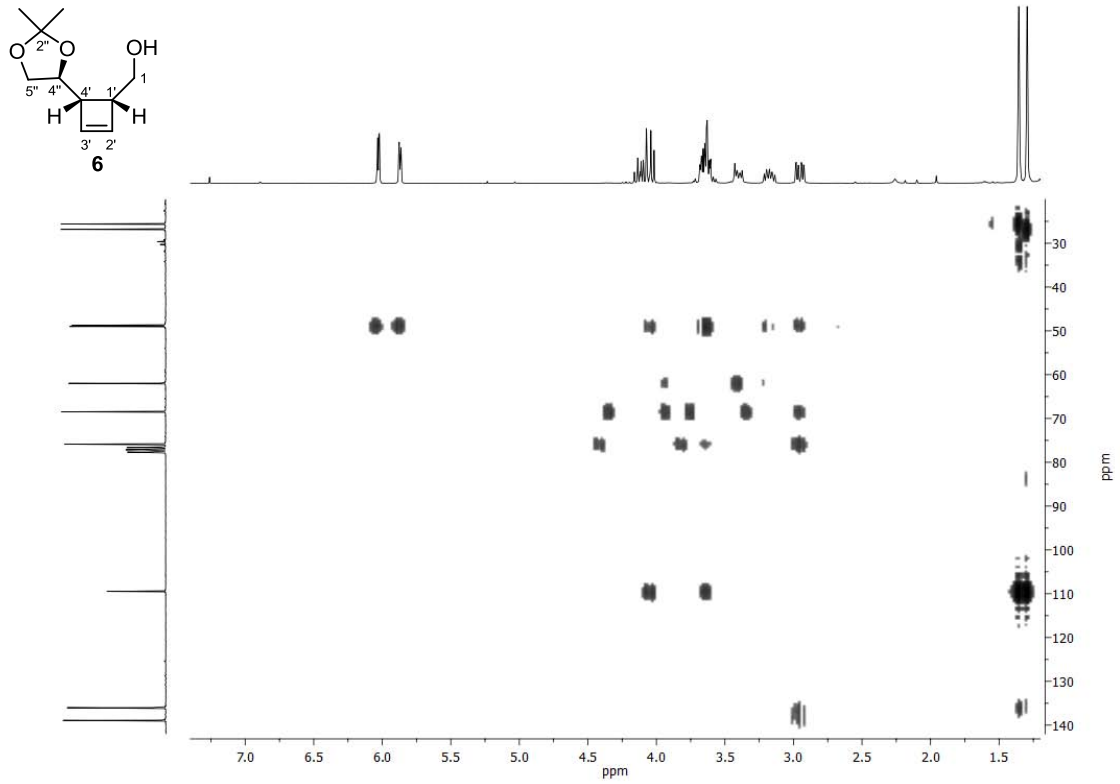
^{1}H-NMR (250 MHz, $CDCl_3$)



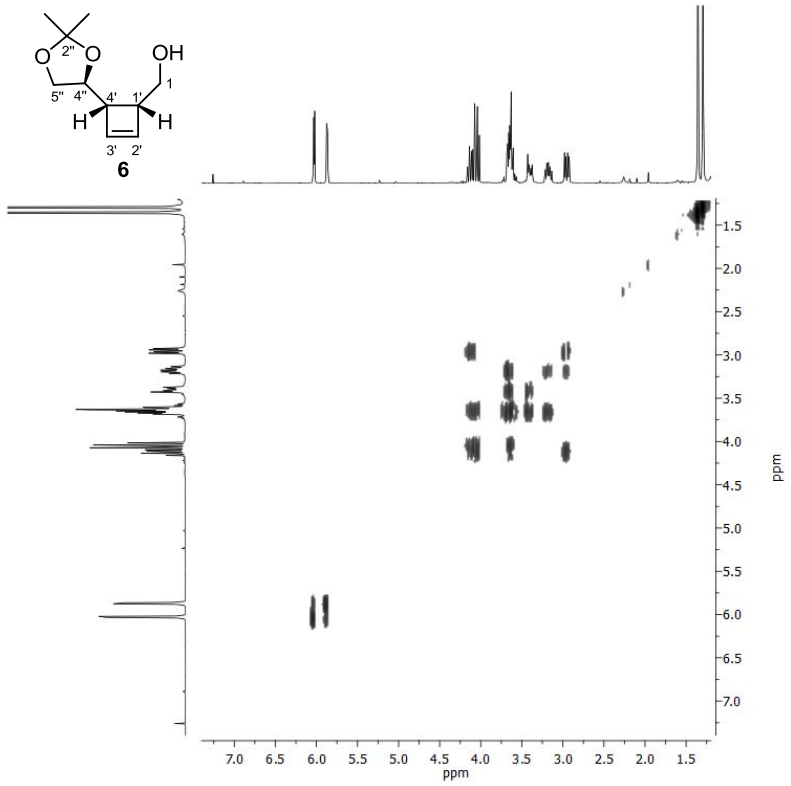
^{13}C-NMR (62.5 MHz, $CDCl_3$)



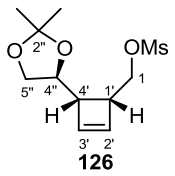
DEPT135 (250 MHz, $CDCl_3$)



HMBC (250 MHz, $CDCl_3$)

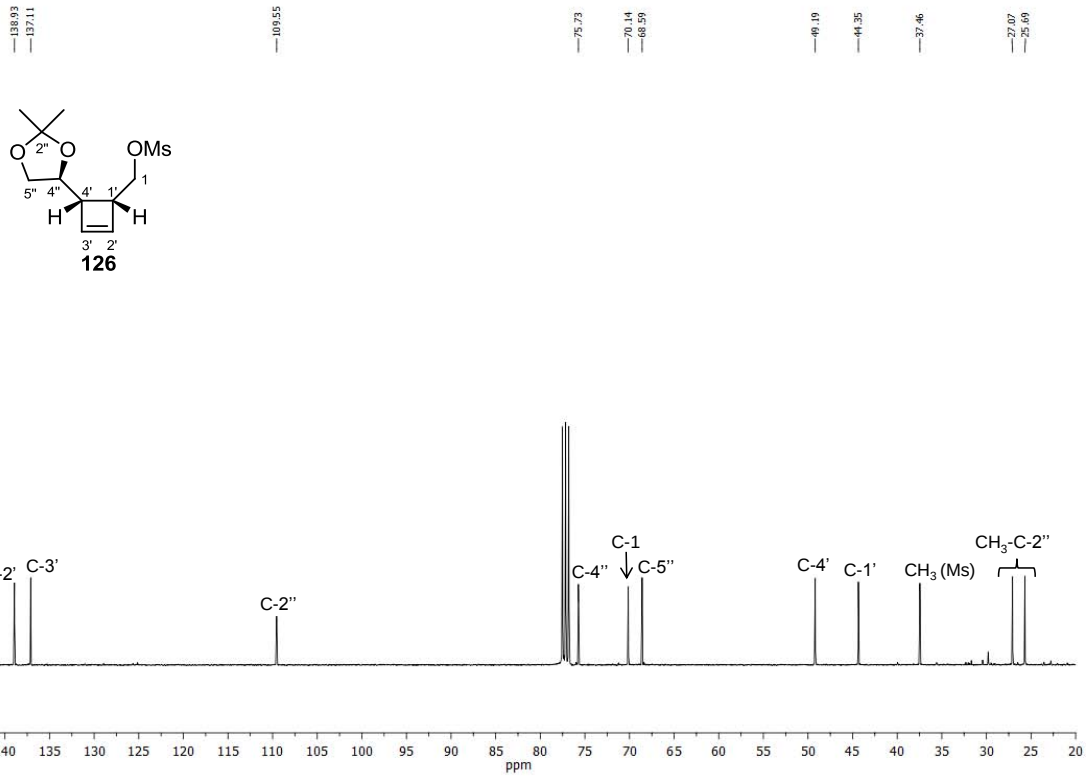


COSY (250 MHz, $CDCl_3$)



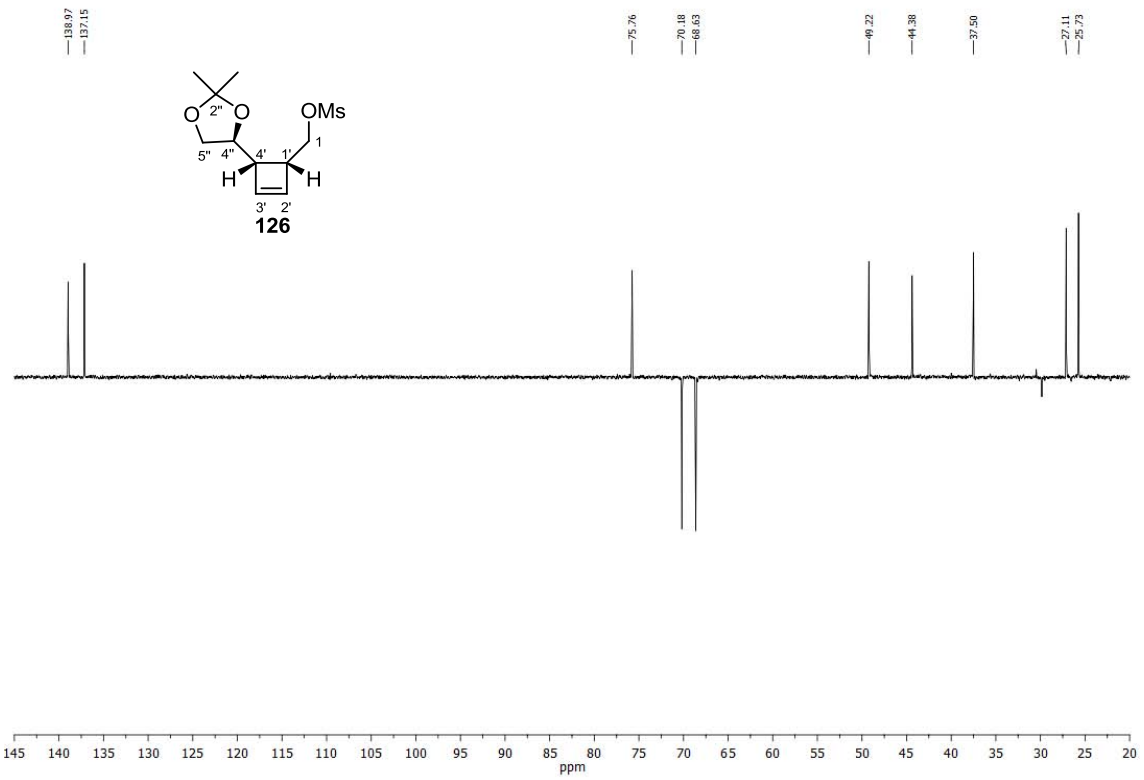
126

^{1}H-NMR (360 MHz, $CDCl_3$)

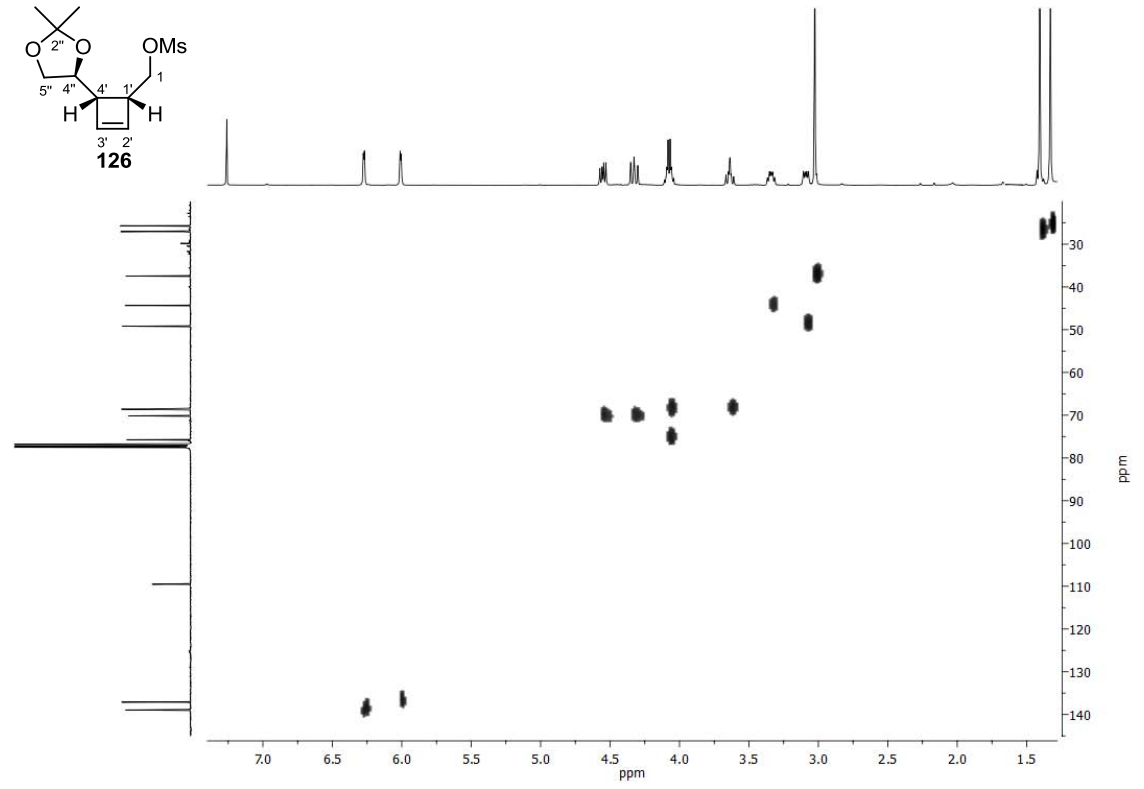


126

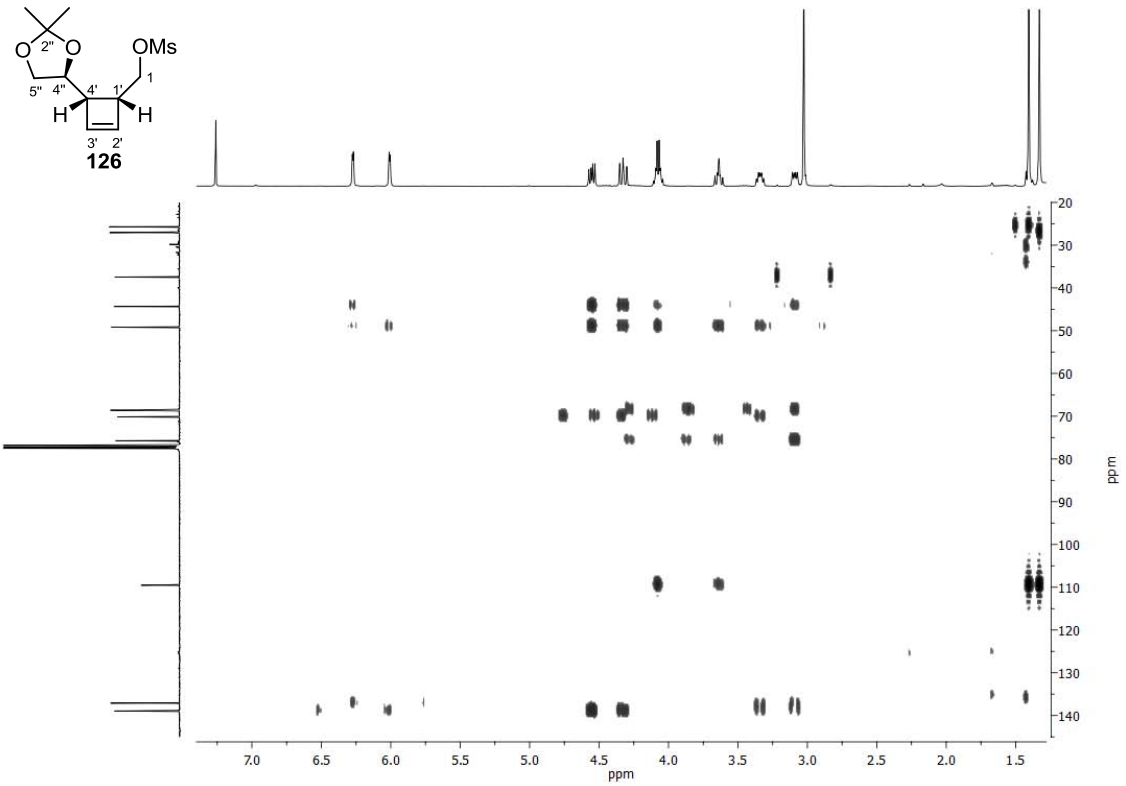
^{13}C-NMR (90 MHz, $CDCl_3$)



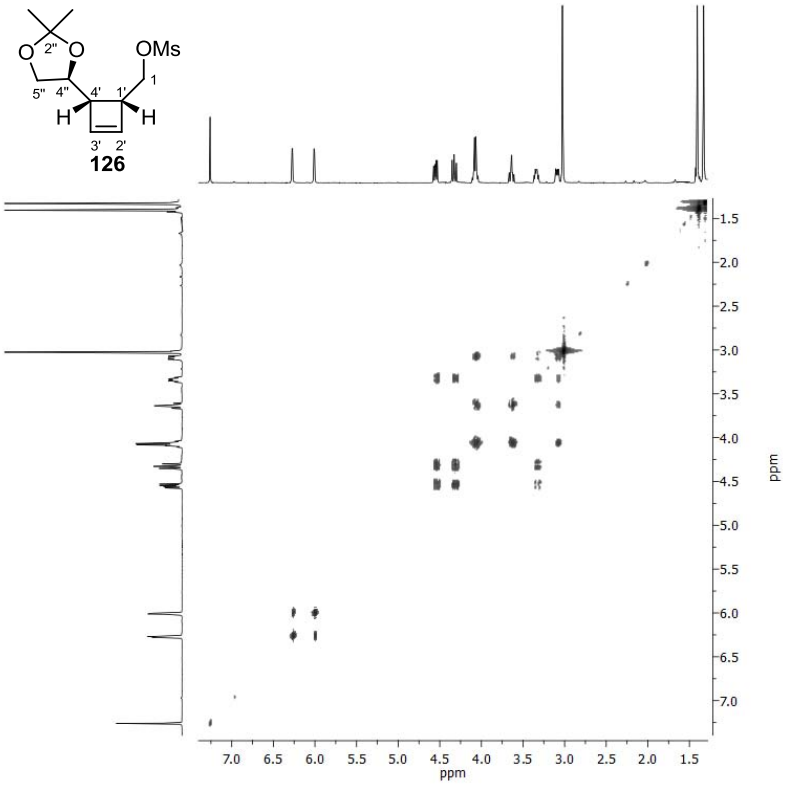
DEPT135 (360 MHz, $CDCl_3$)



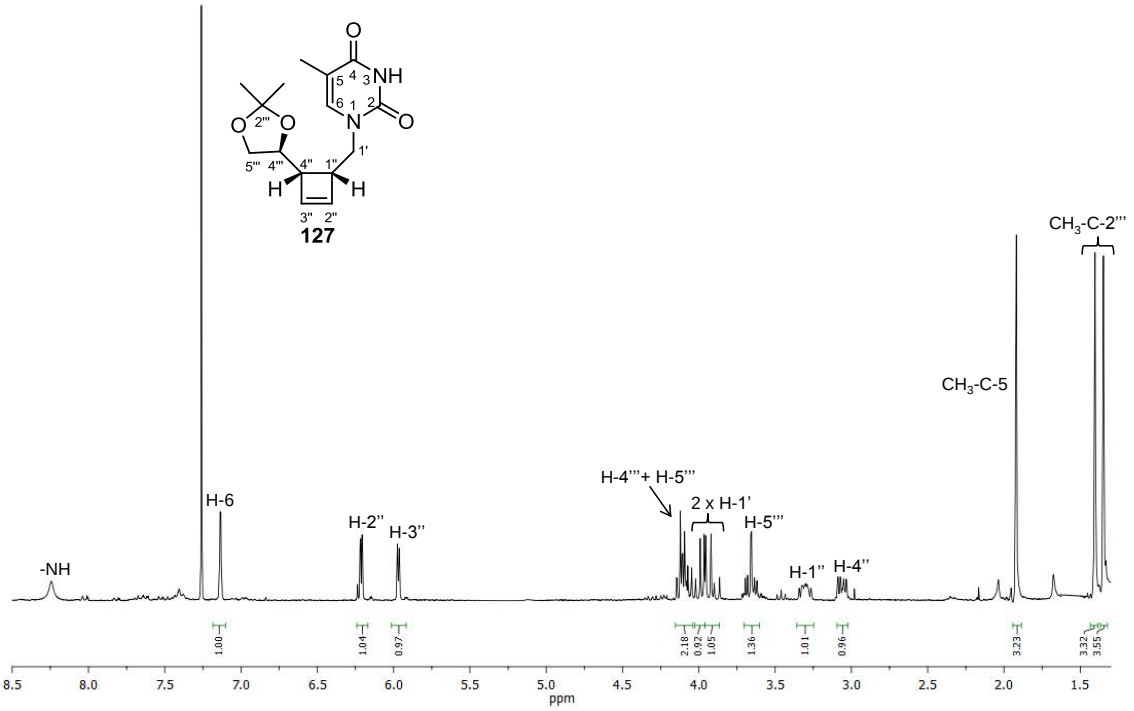
HSQC (360 MHz, $CDCl_3$)



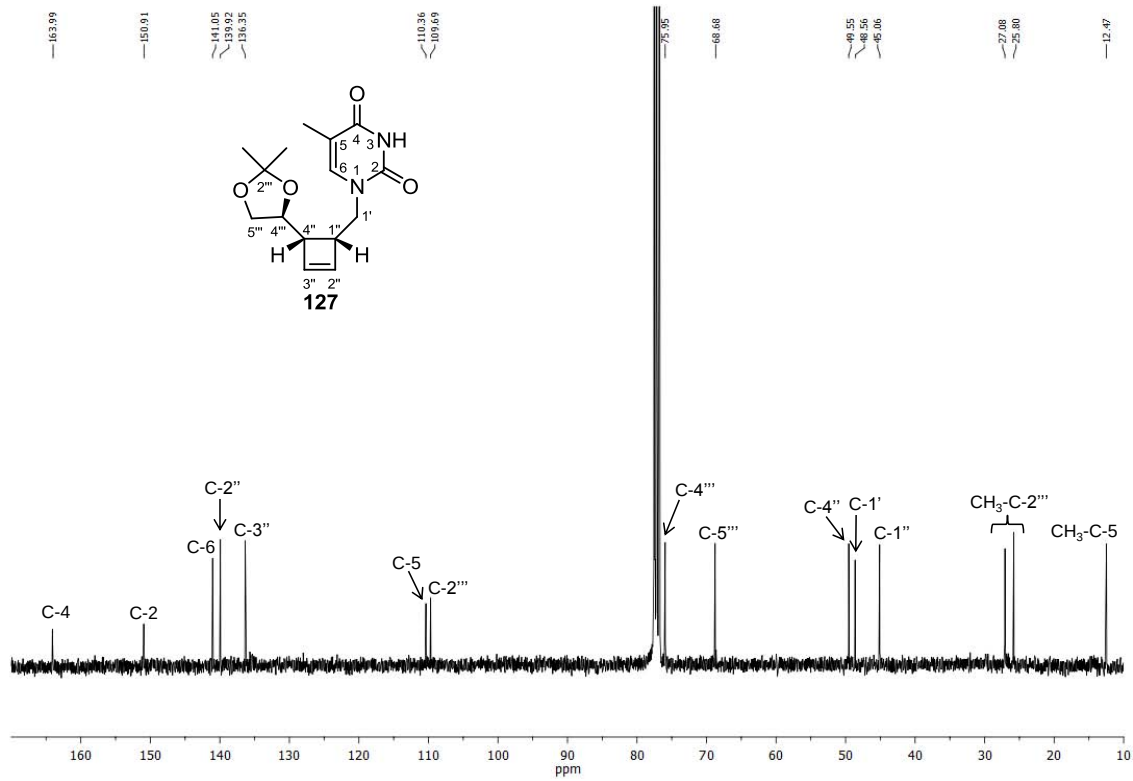
HMBC (360 MHz, $CDCl_3$)



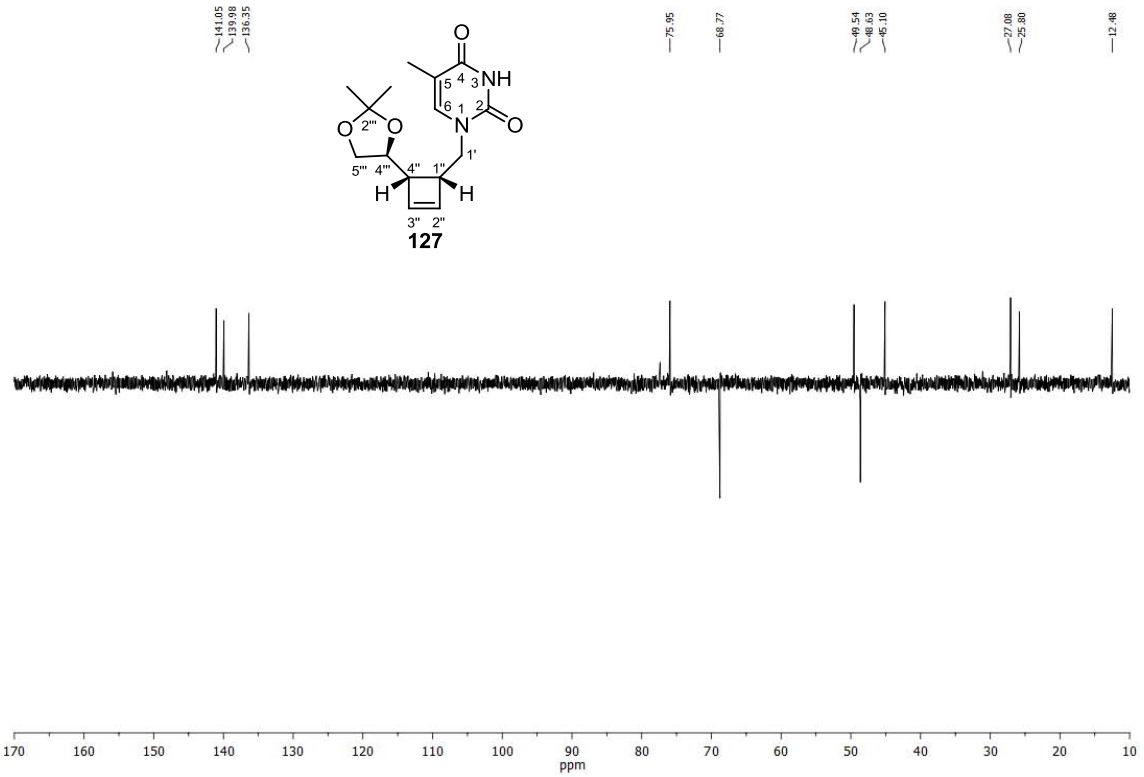
COSY (360 MHz, $CDCl_3$)



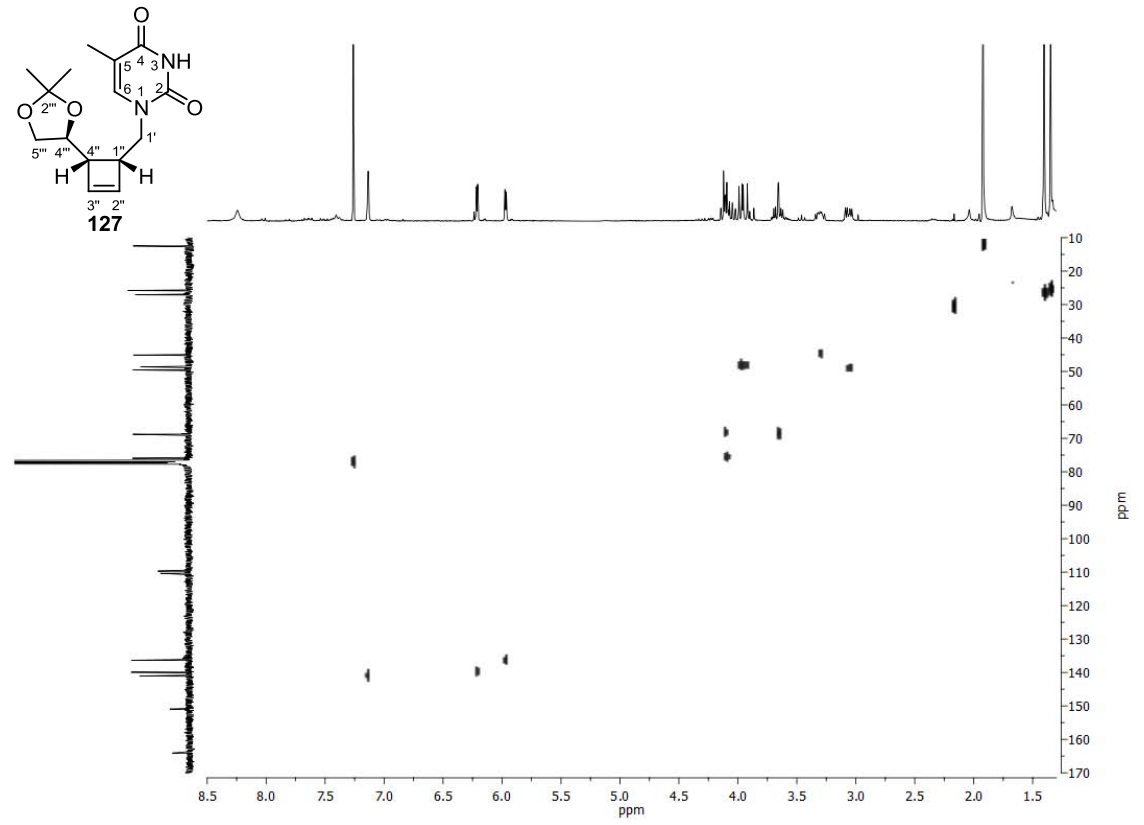
^{1}H-NMR (360 MHz, $CDCl_3$)



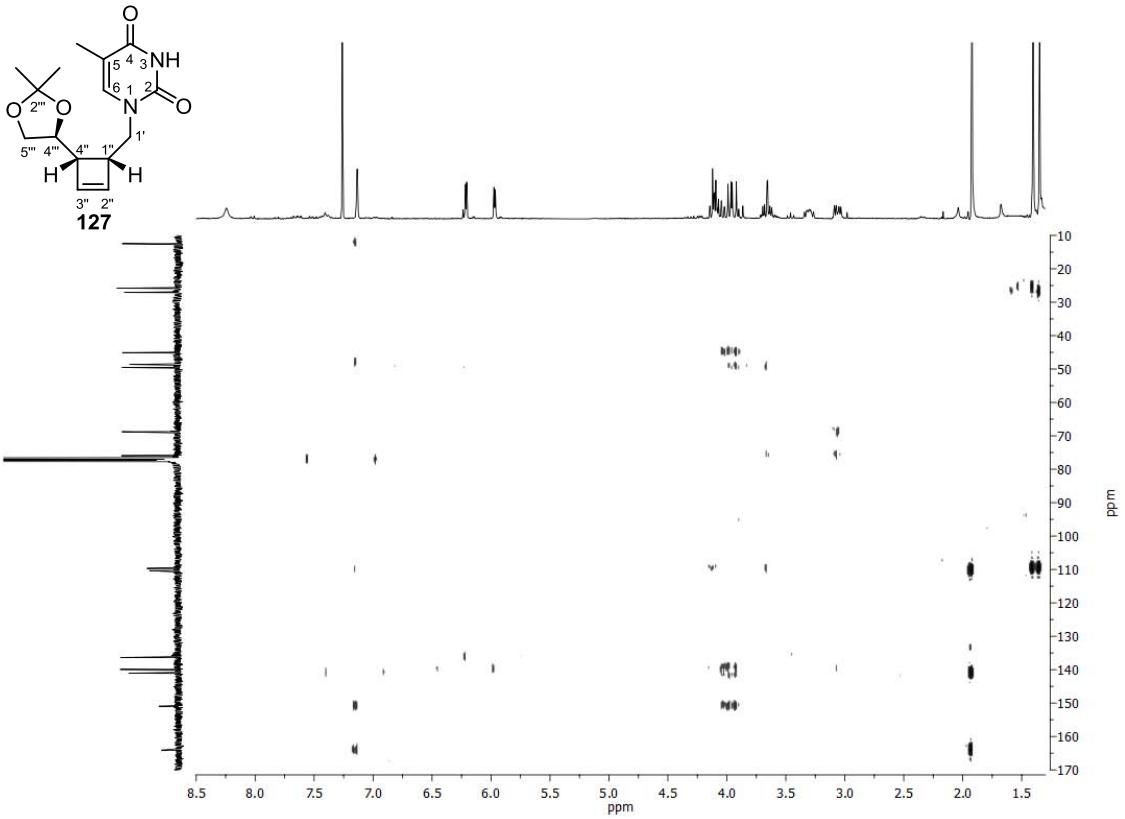
^{13}C-NMR (90 MHz, $CDCl_3$)



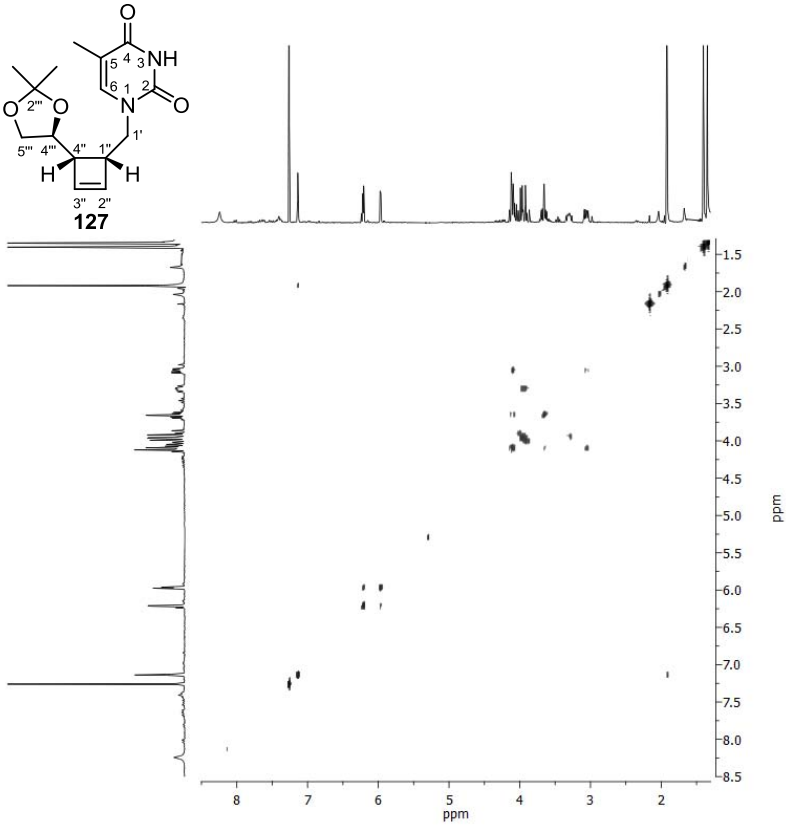
DEPT135 (360 MHz, $CDCl_3$)



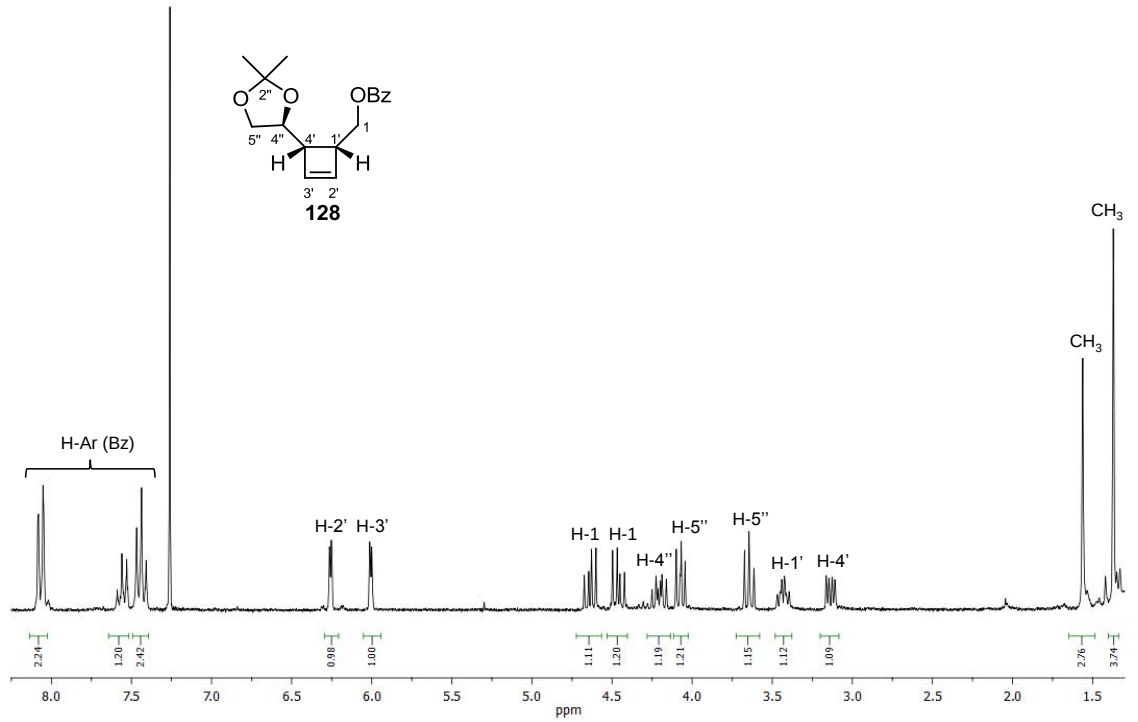
HSQC (360 MHz, $CDCl_3$)



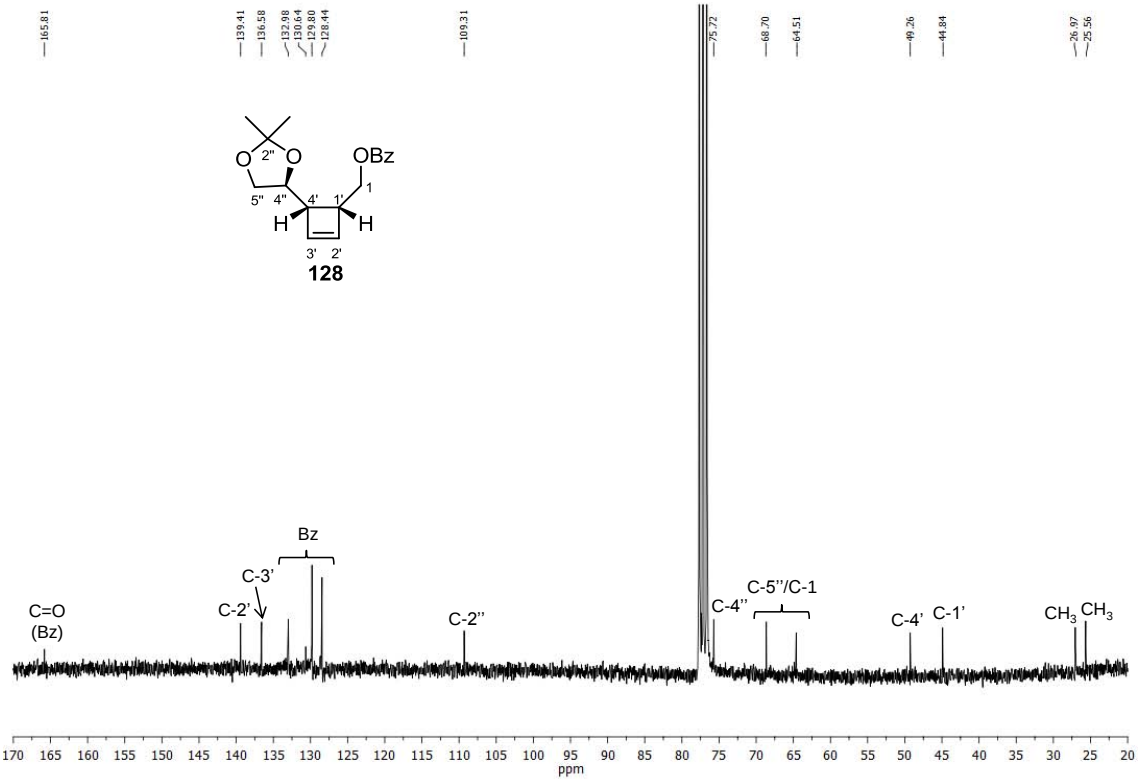
HMBC (360 MHz, $CDCl_3$)



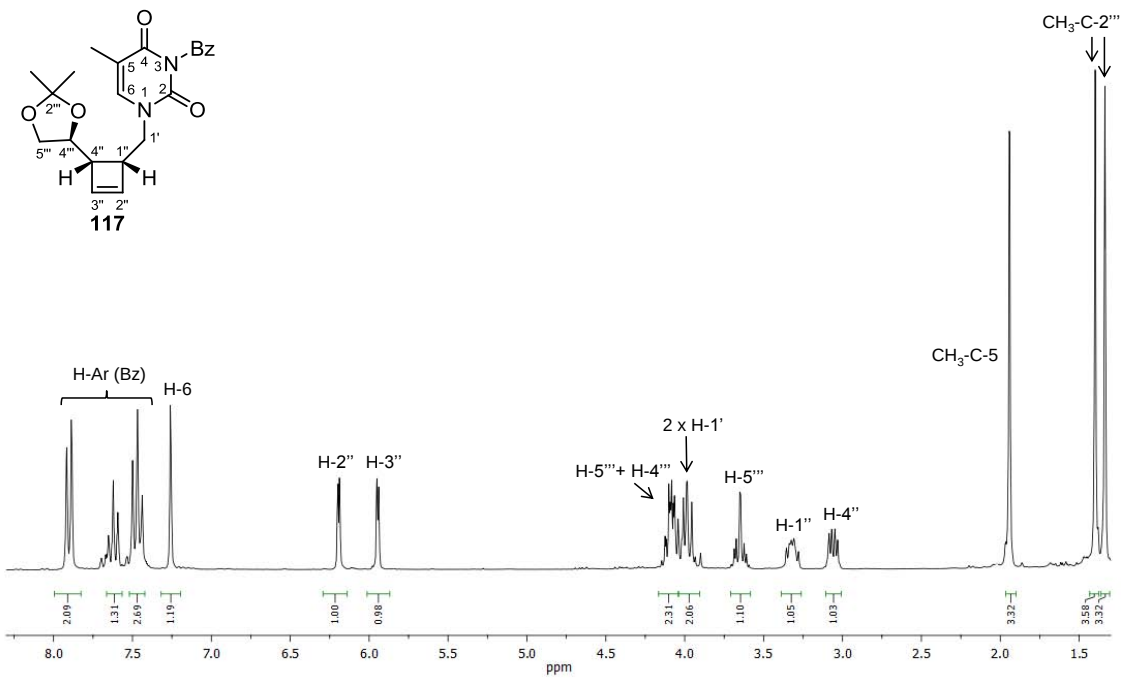
COSY (360 MHz, $CDCl_3$)



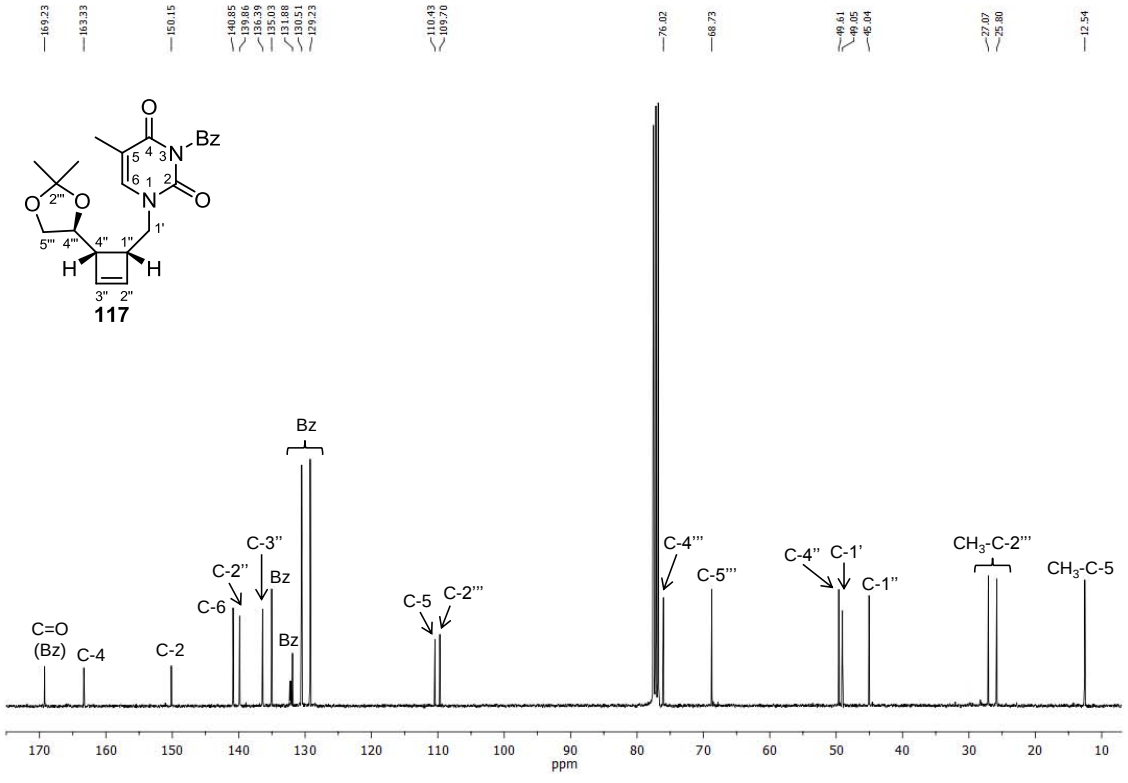
^{1}H-NMR (250 MHz, $CDCl_3$)



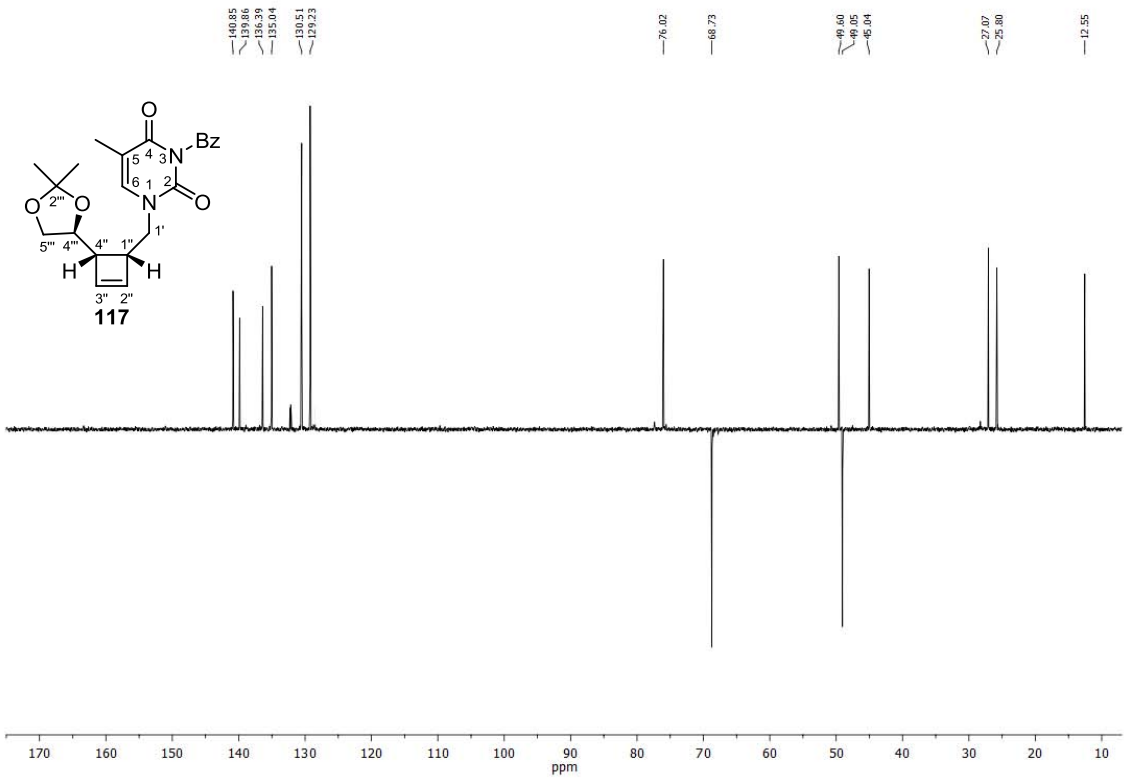
^{13}C-NMR (62.5 MHz, $CDCl_3$)



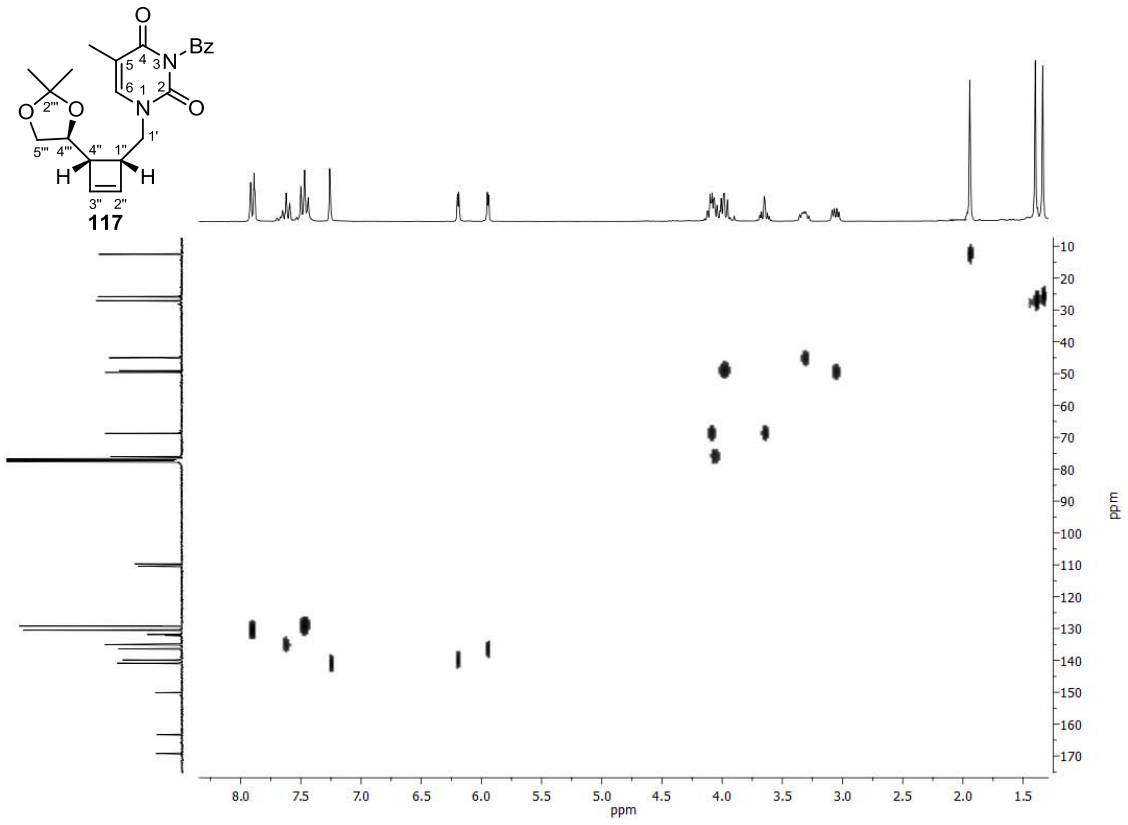
^{1}H-NMR (360 MHz, $CDCl_3$)



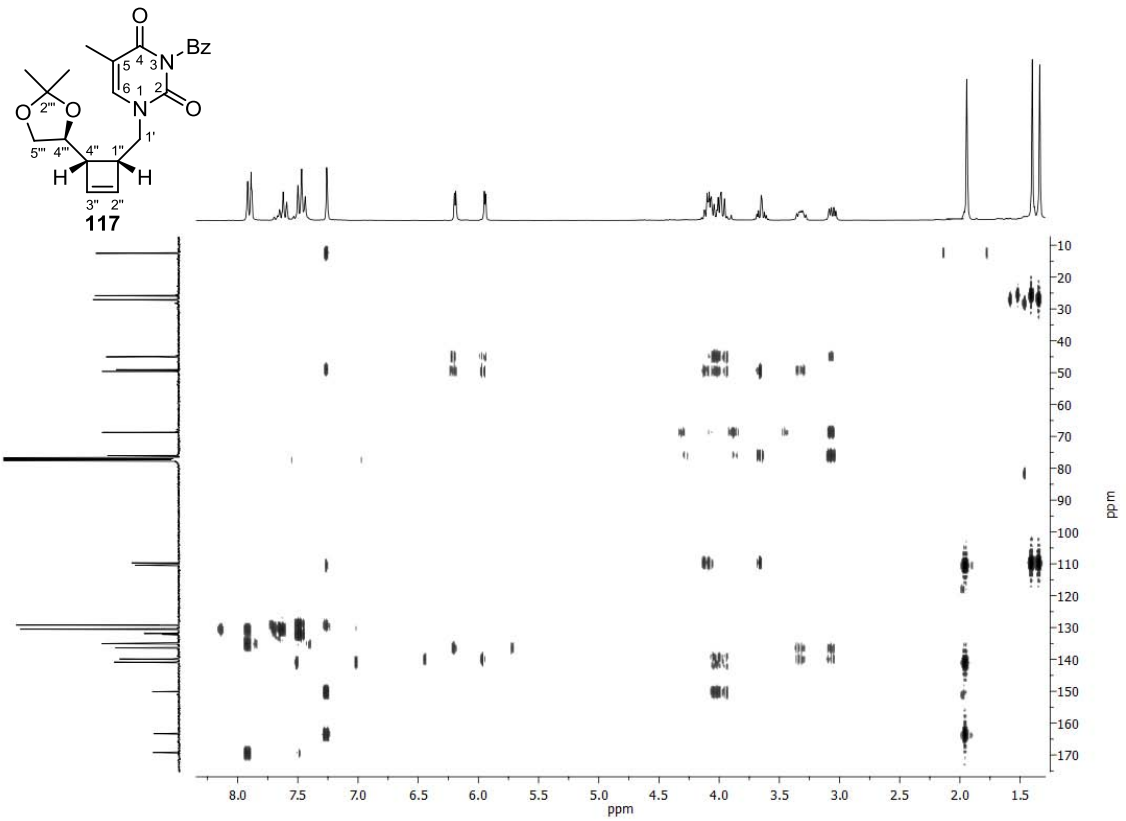
^{13}C-NMR (90 MHz, $CDCl_3$)



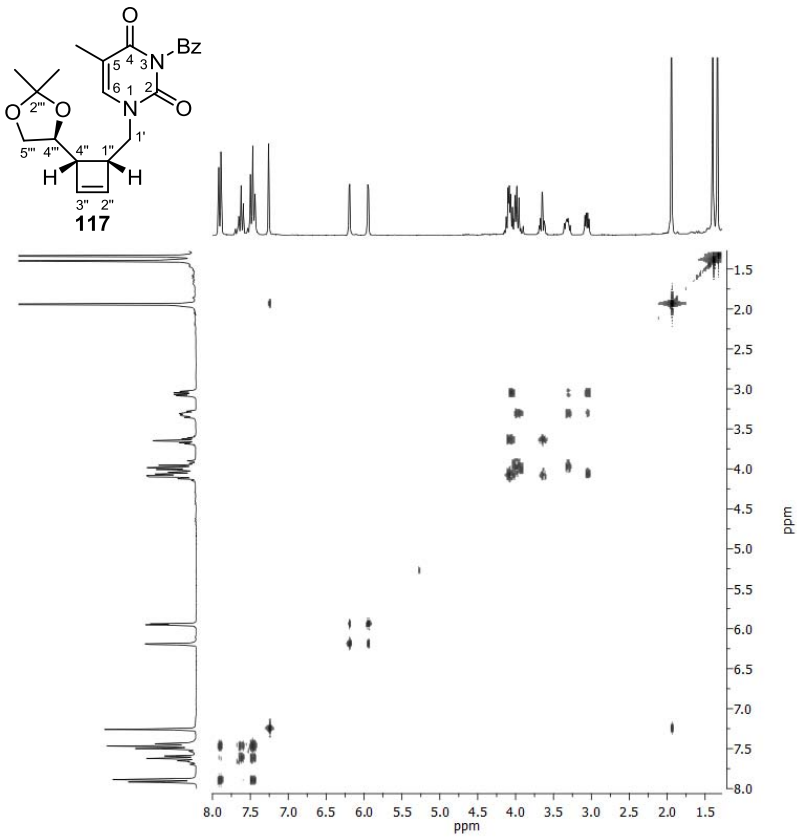
DEPT135 (360 MHz, $CDCl_3$)



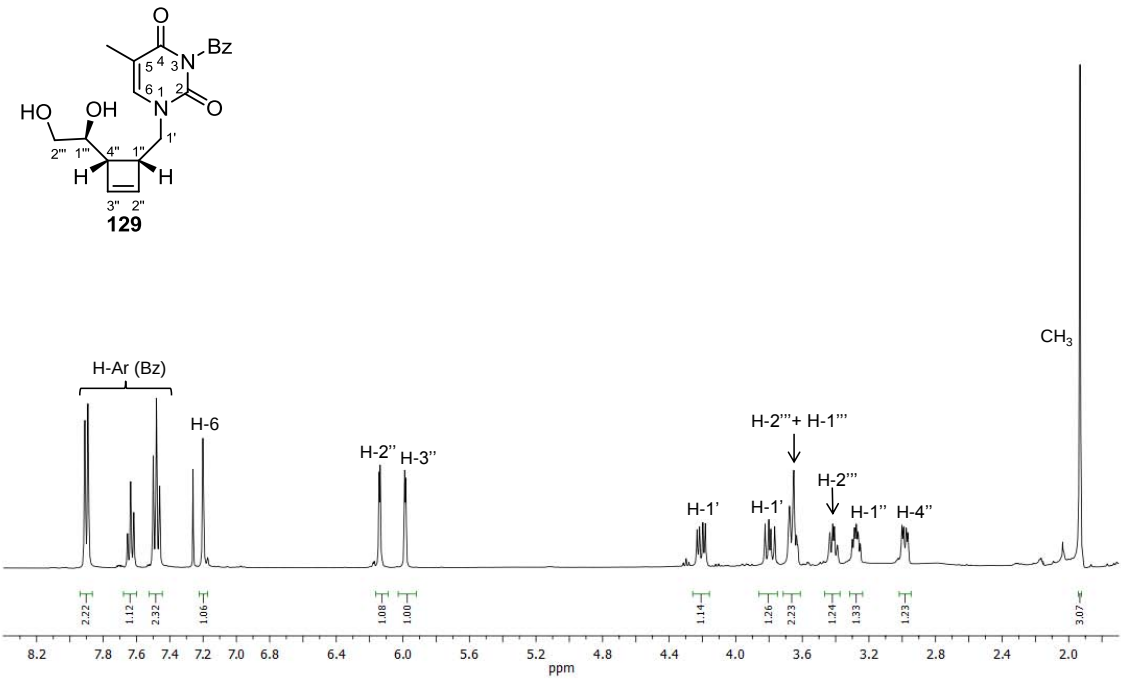
HSQC (360 MHz, $CDCl_3$)



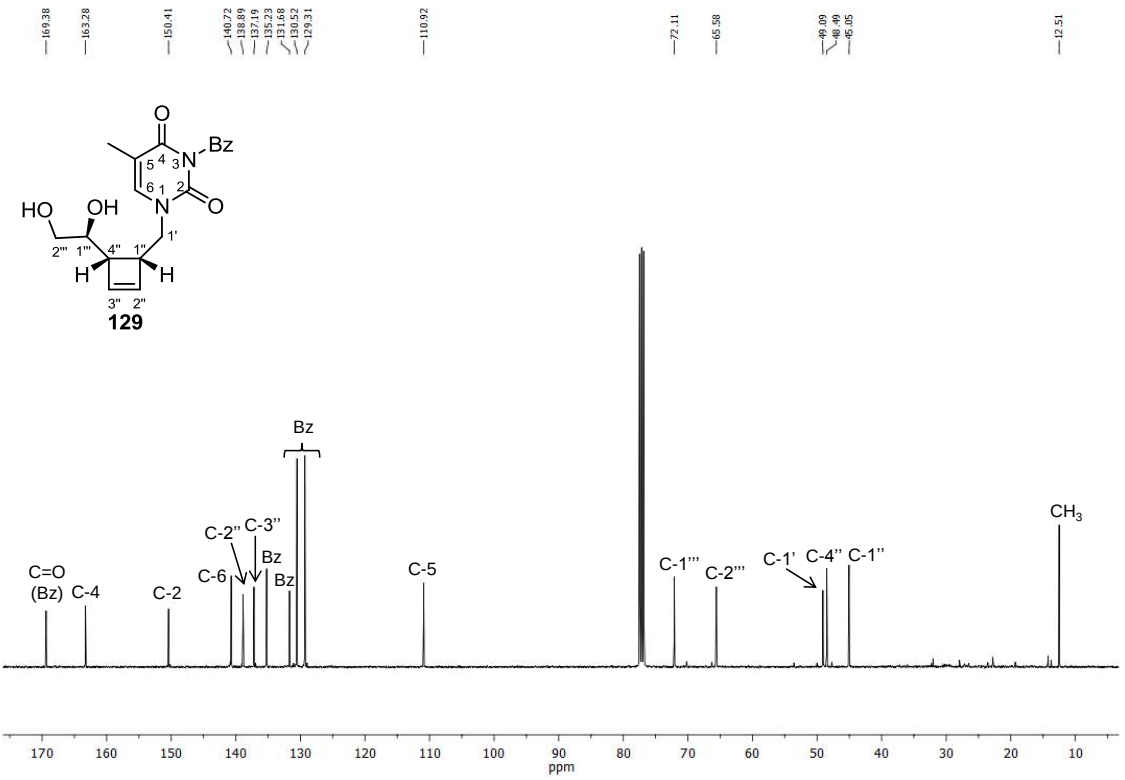
HMBC (360 MHz, $CDCl_3$)



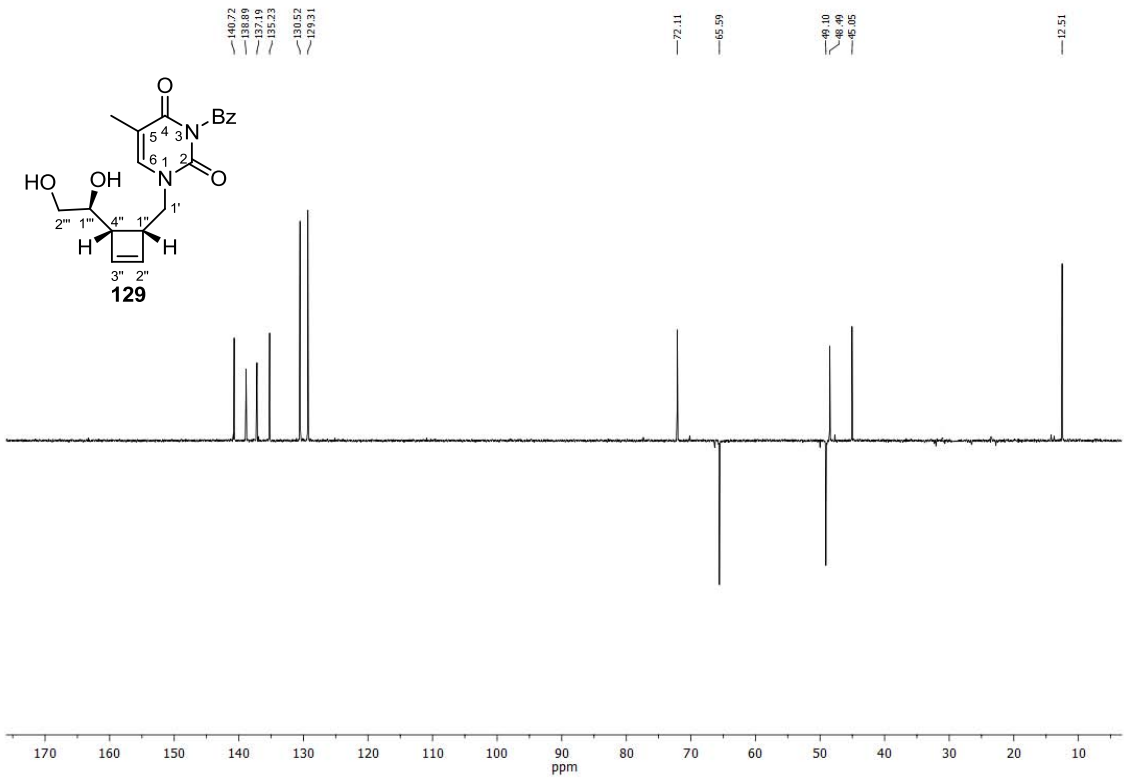
COSY (360 MHz, $CDCl_3$)



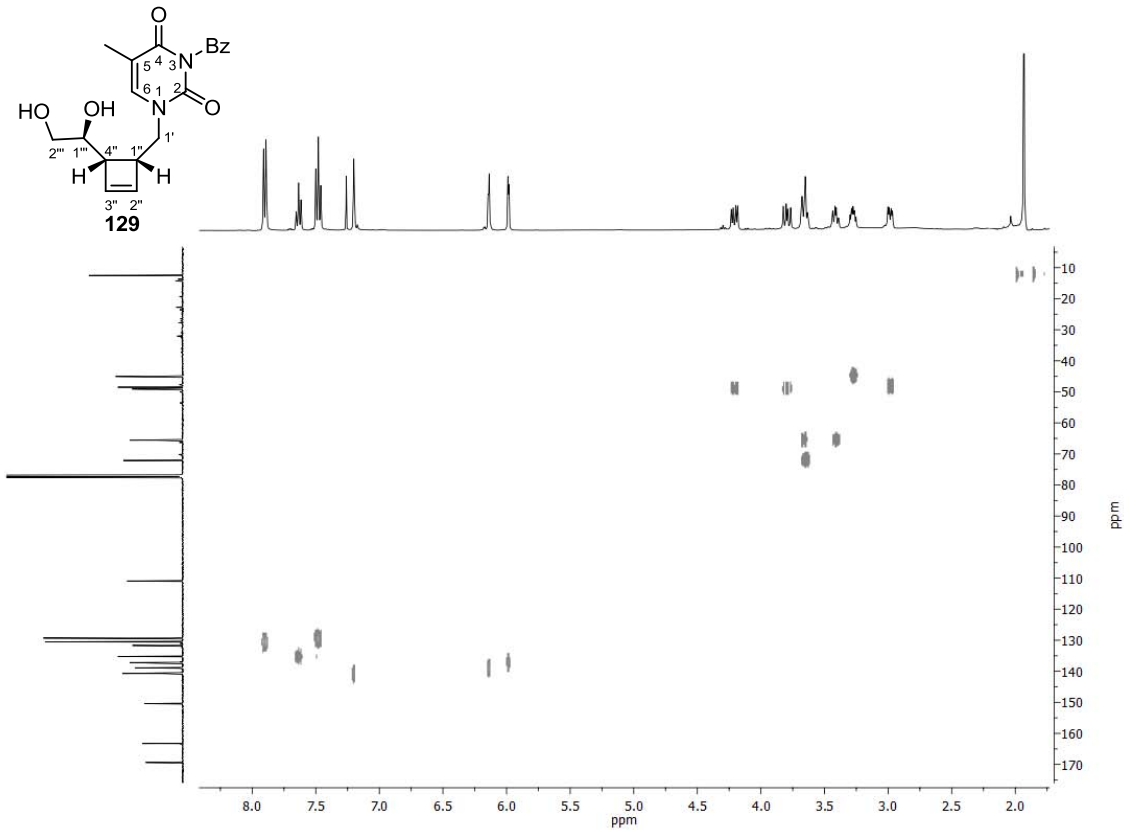
^{1}H-NMR (400 MHz, $CDCl_3$)



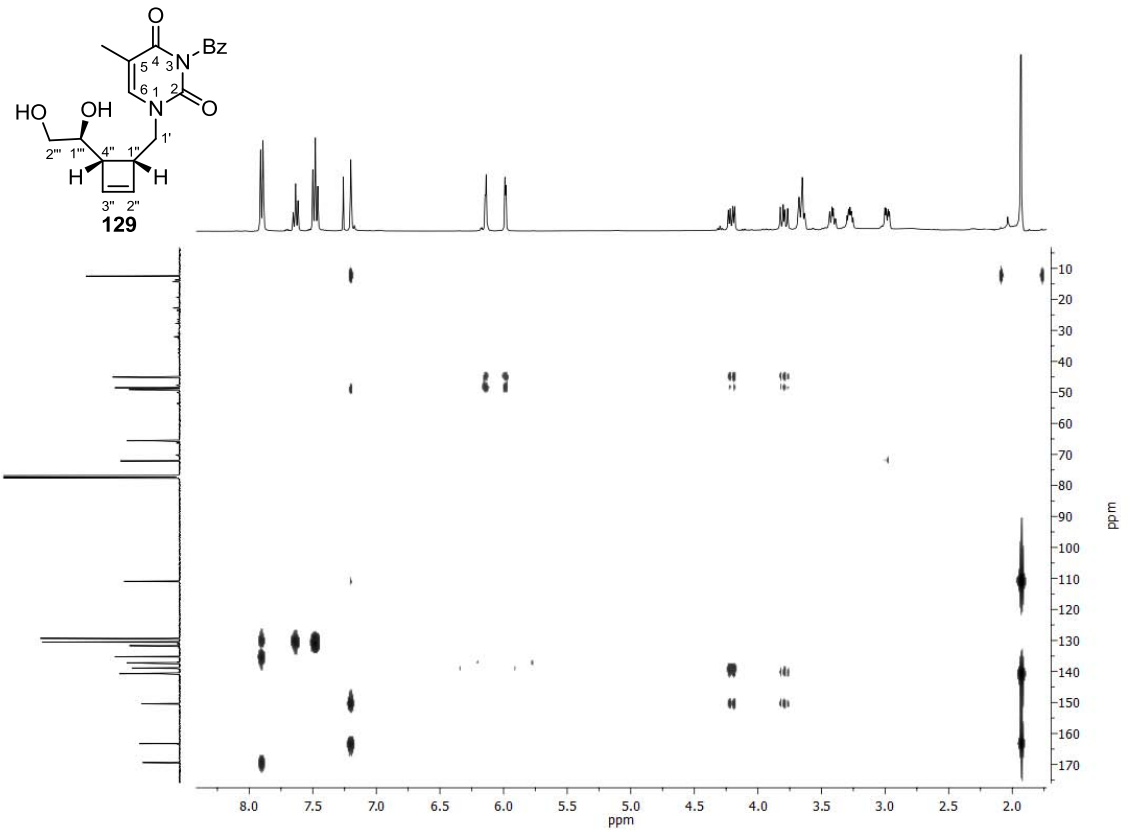
^{13}C-NMR (100 MHz, $CDCl_3$)



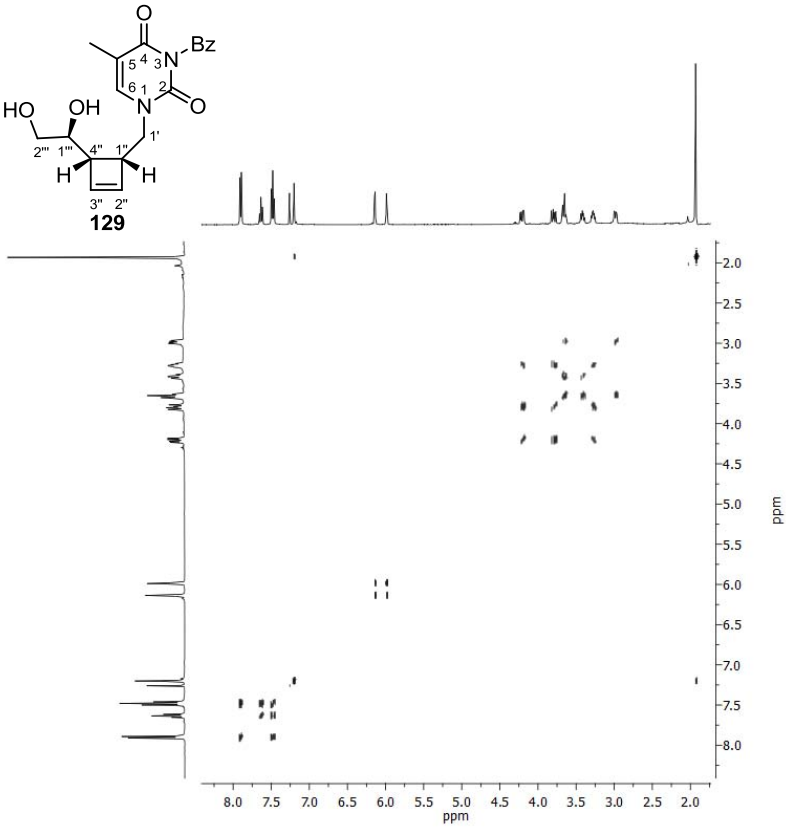
DEPT135 (400 MHz, $CDCl_3$)



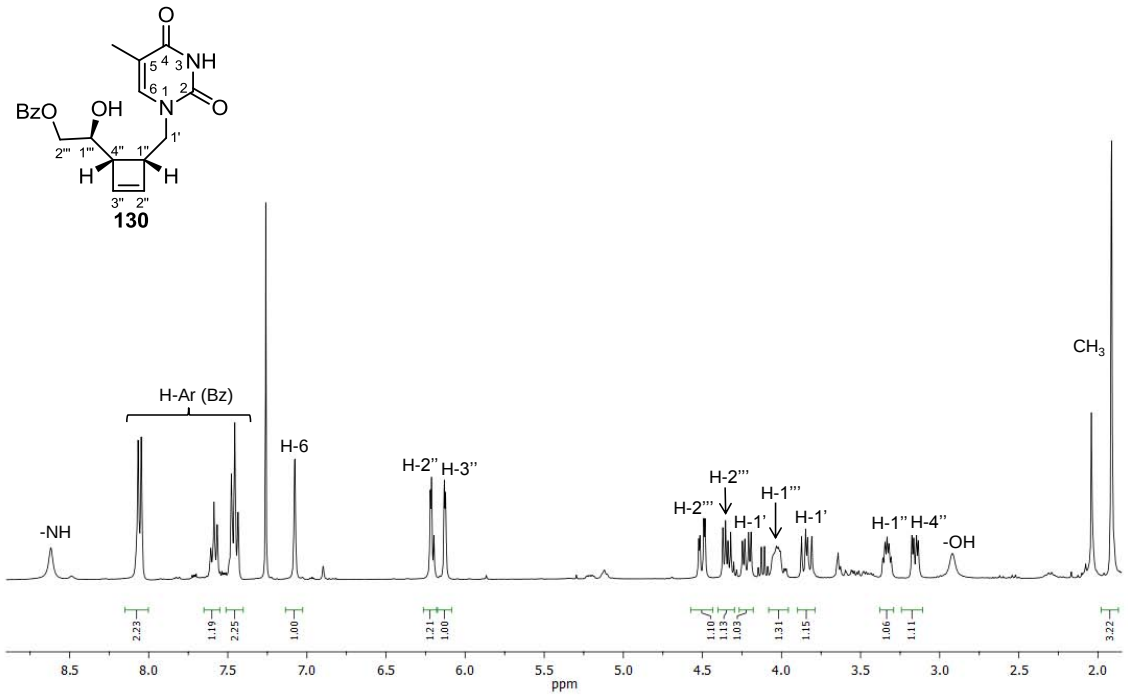
HSQC (400 MHz, $CDCl_3$)



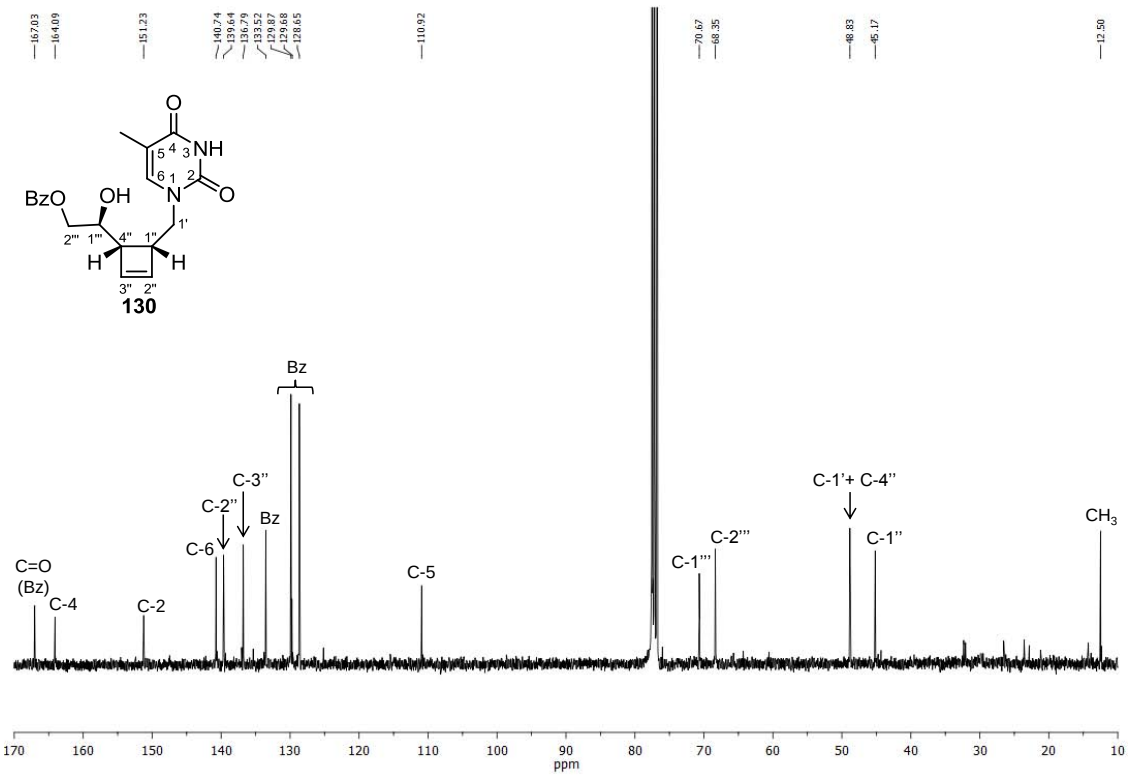
HMBC (400 MHz, $CDCl_3$)



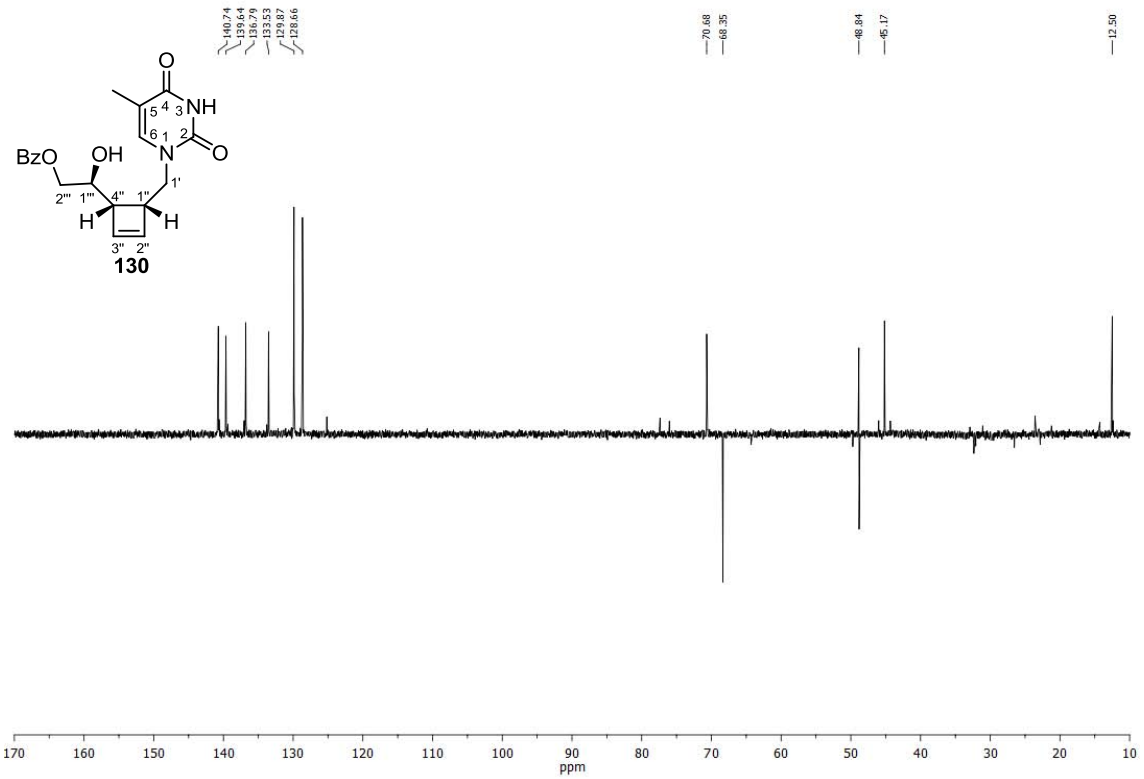
COSY (400 MHz, $CDCl_3$)



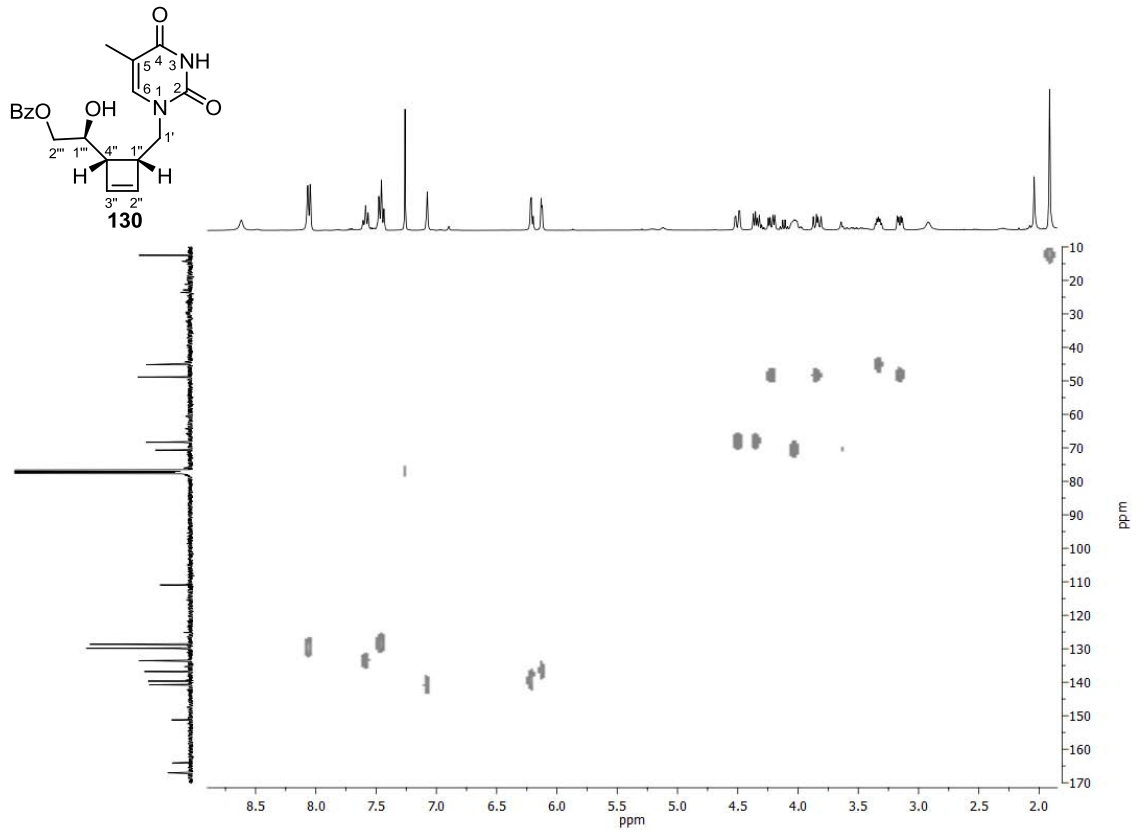
^{1}H-NMR (360 MHz, $CDCl_3$)



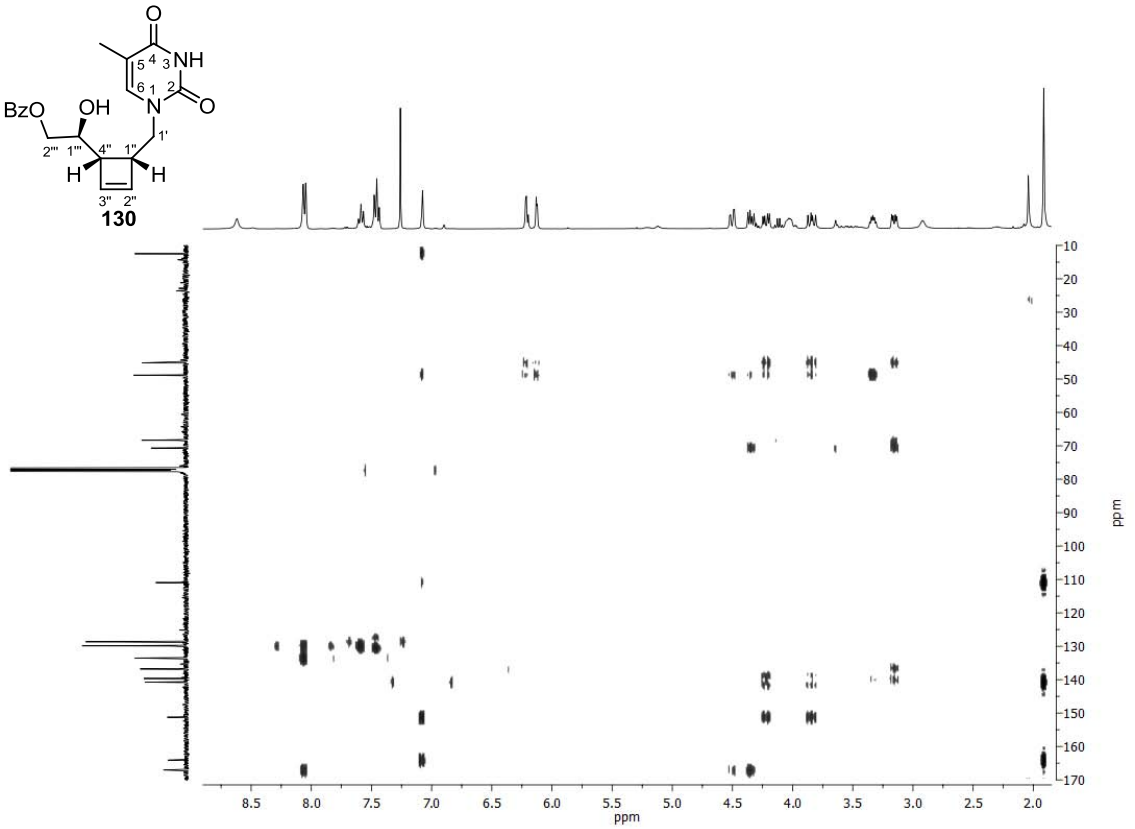
^{13}C-NMR (90 MHz, $CDCl_3$)



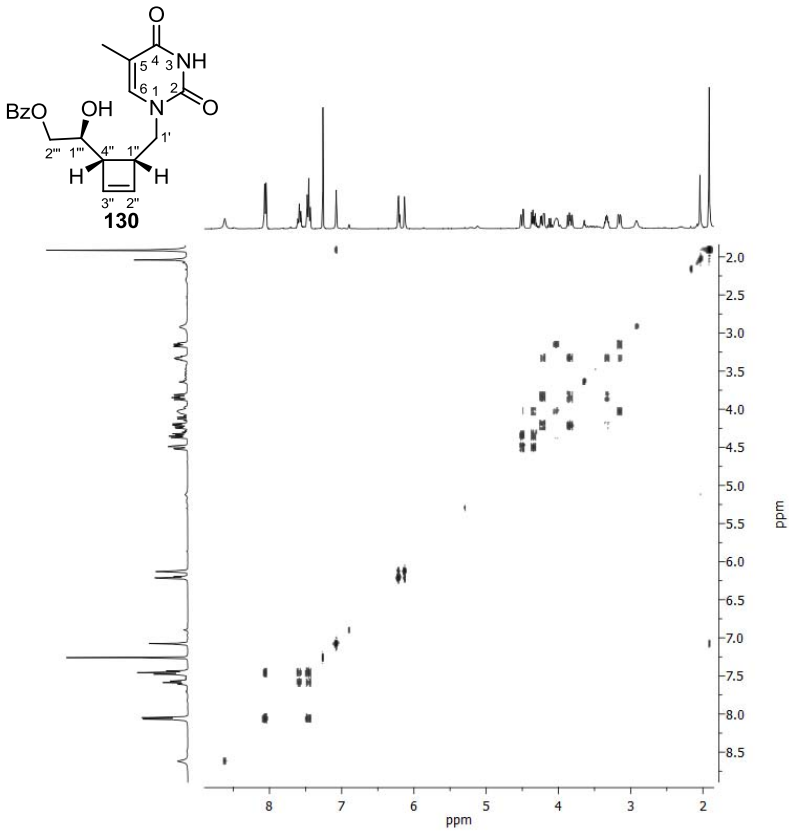
DEPT135 (360 MHz, $CDCl_3$)



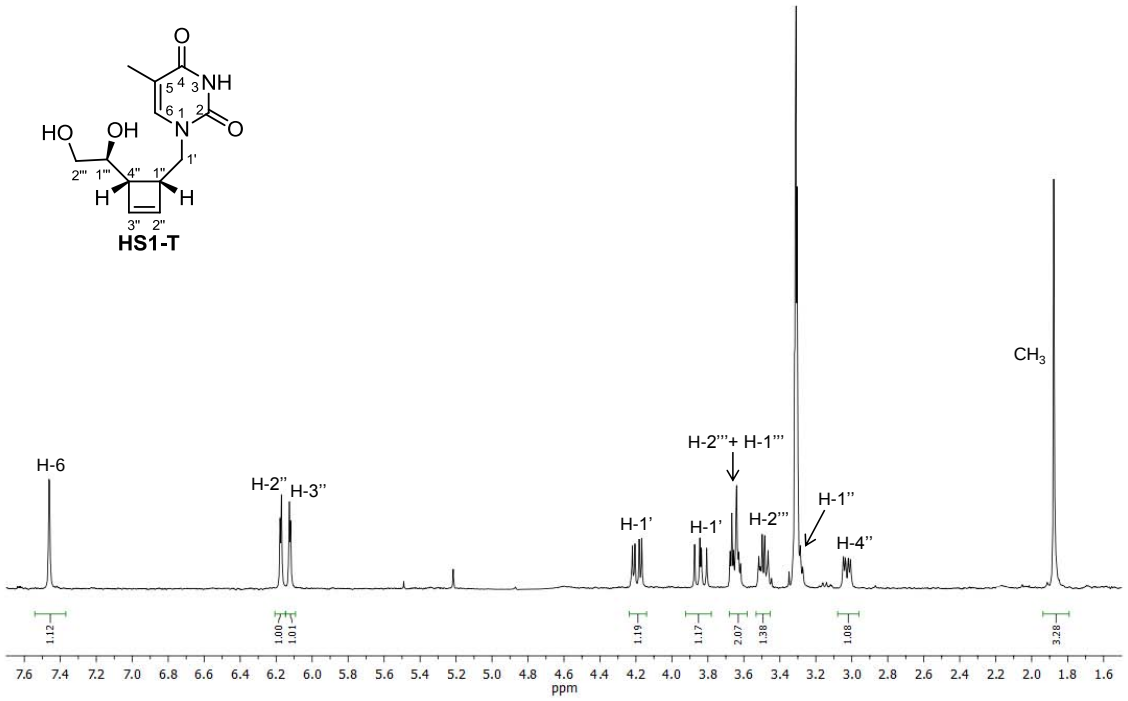
HSQC (360 MHz, $CDCl_3$)



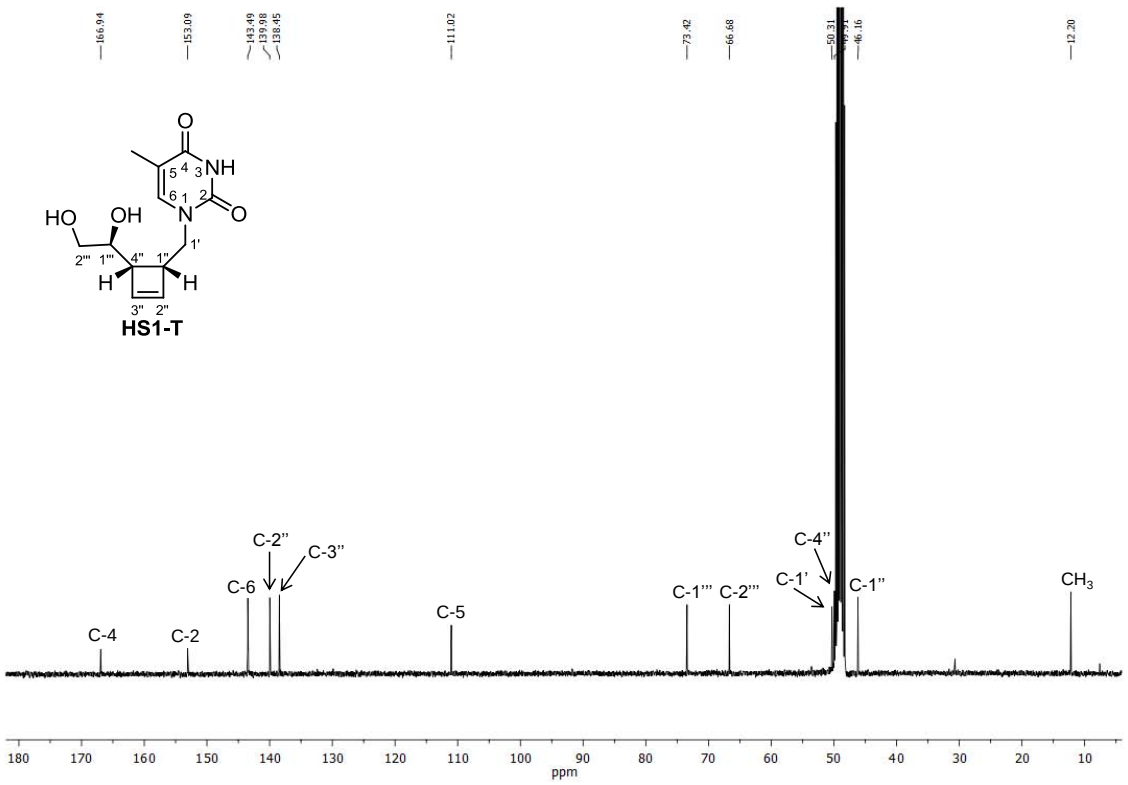
HMBC (360 MHz, $CDCl_3$)



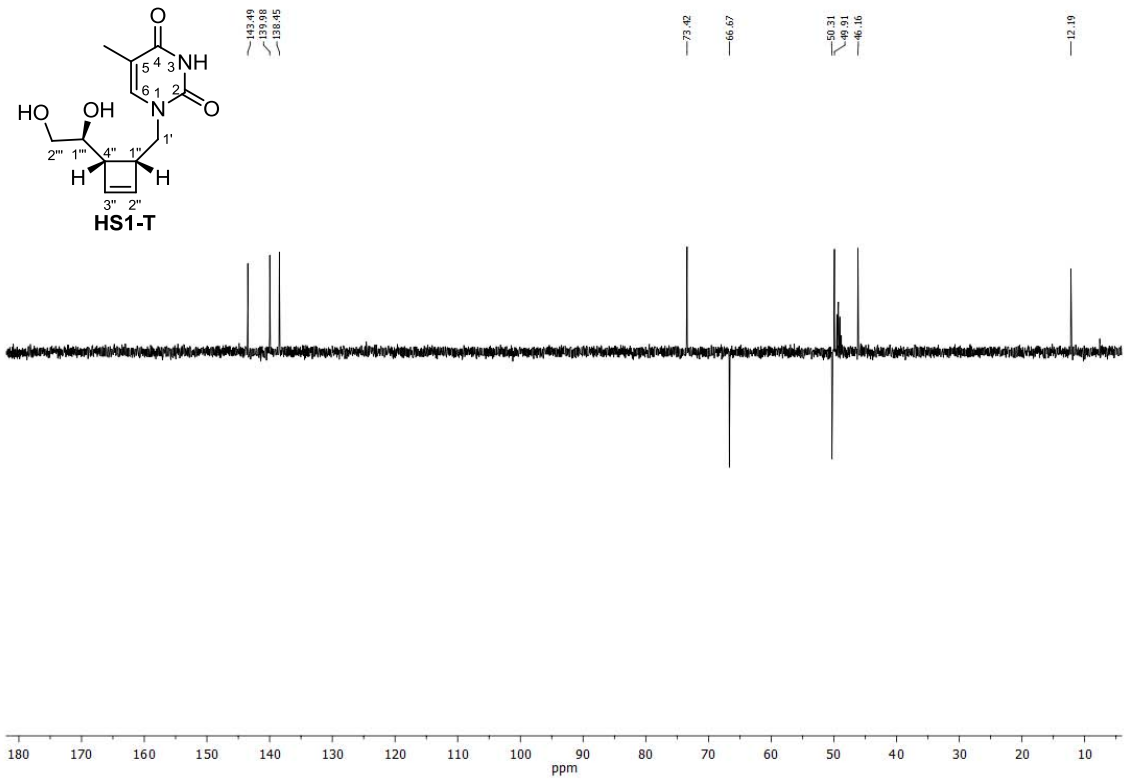
COSY (360 MHz, $CDCl_3$)



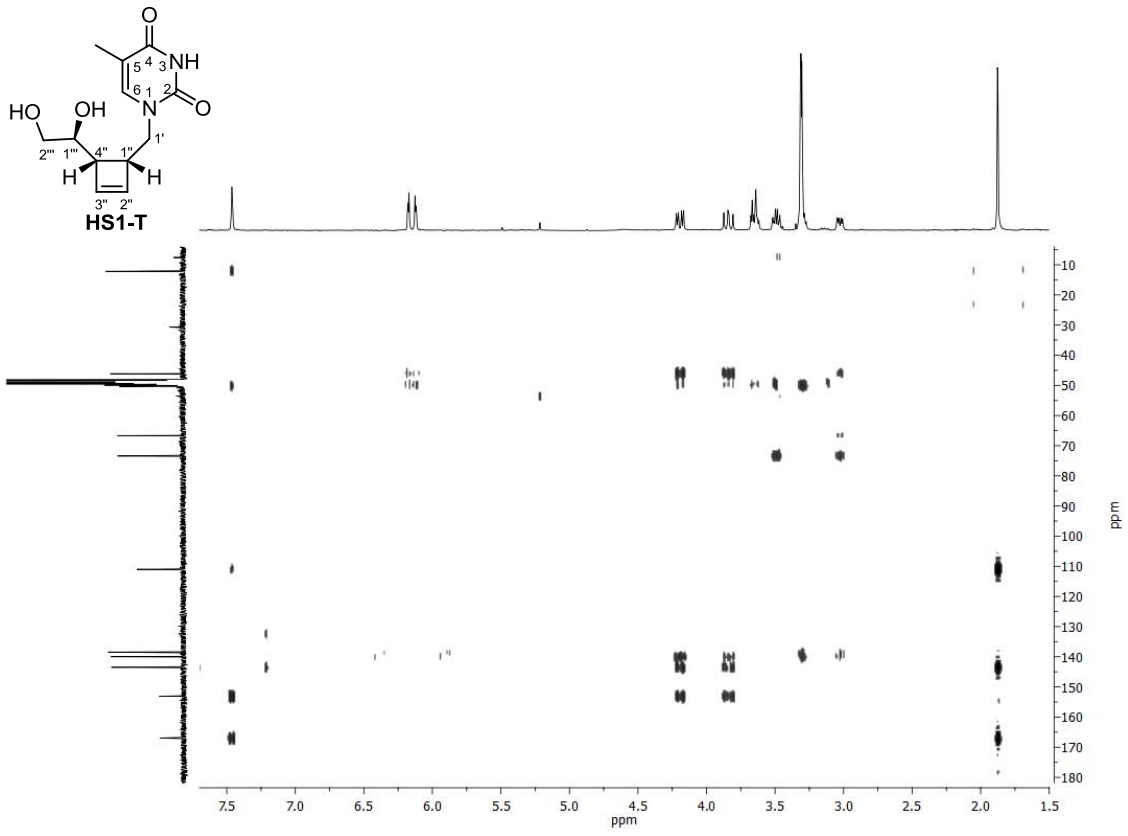
^{1}H-NMR (360 MHz, CD_3OD)



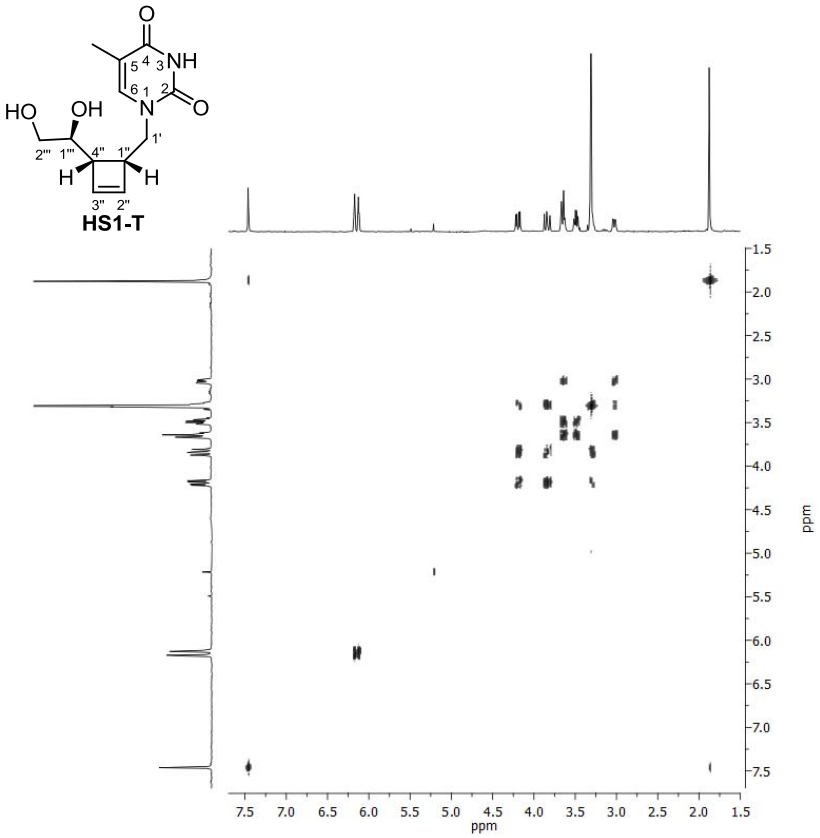
^{13}C-NMR (90 MHz, CD_3OD)



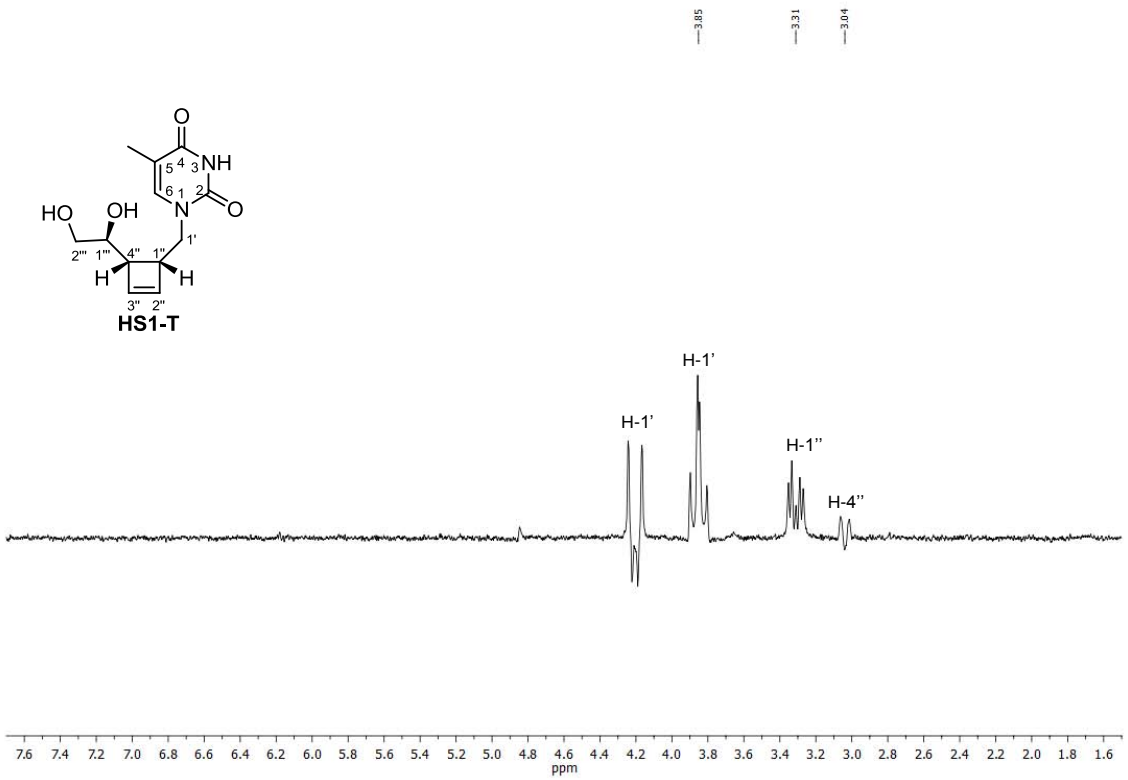
DEPT135 (360 MHz, CD_3OD)



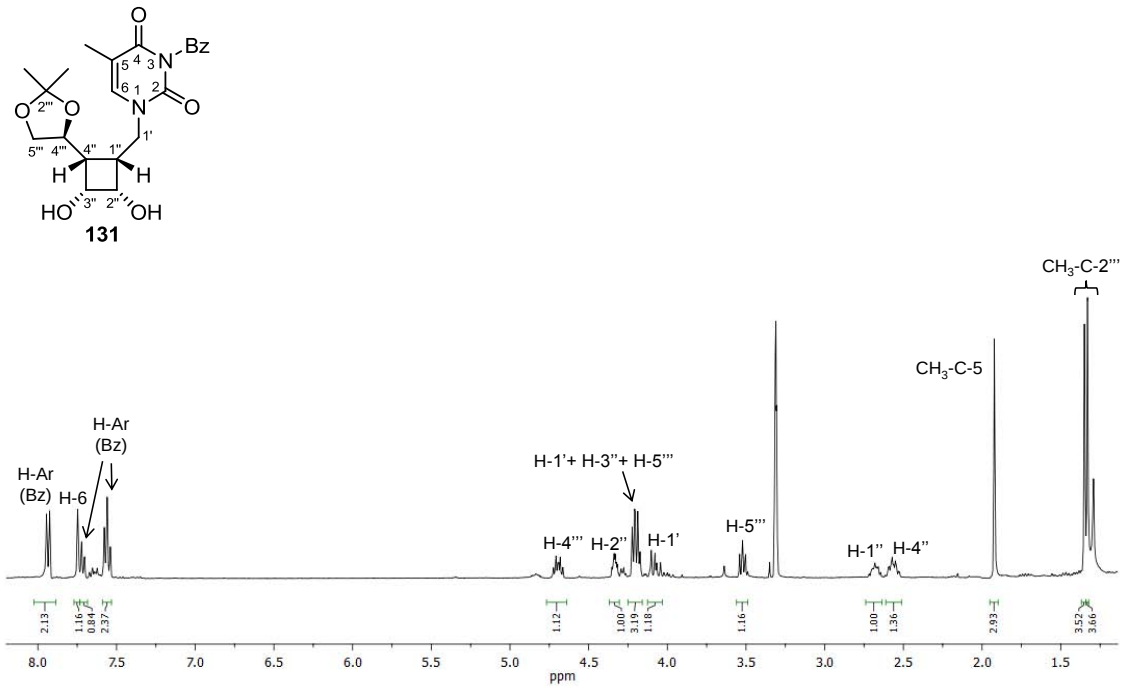
HMBC (360 MHz, CD_3OD)



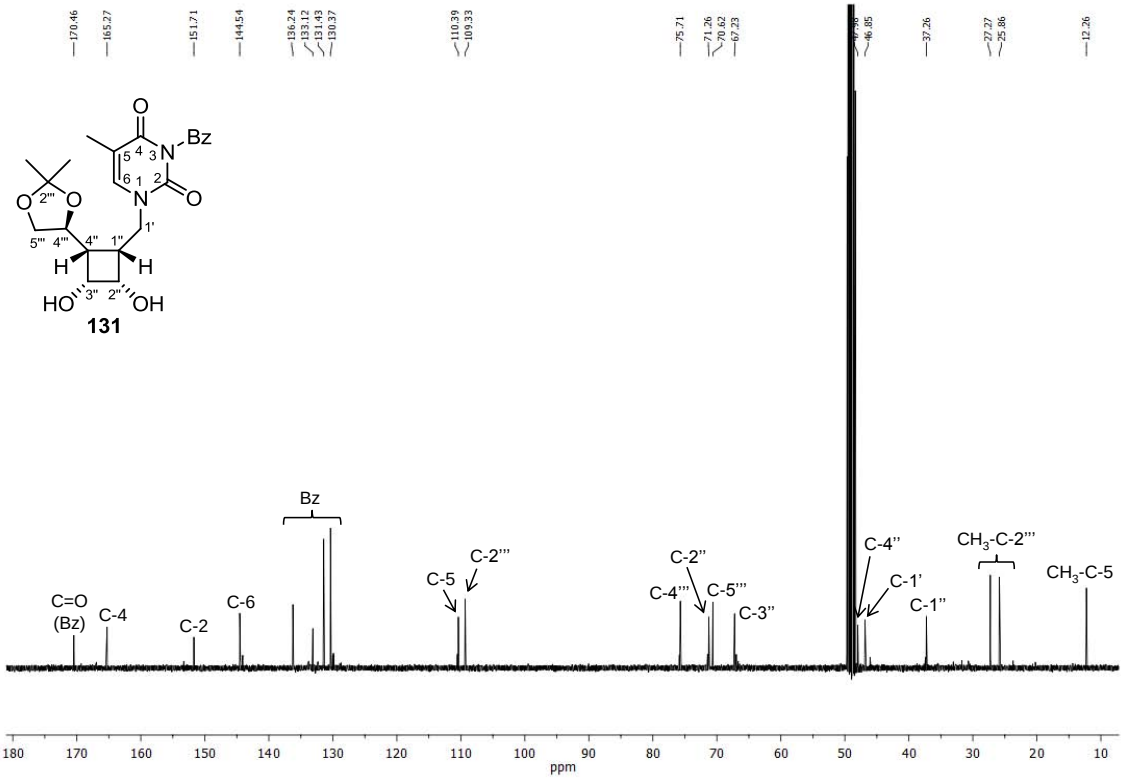
COSY (360 MHz, CD_3OD)



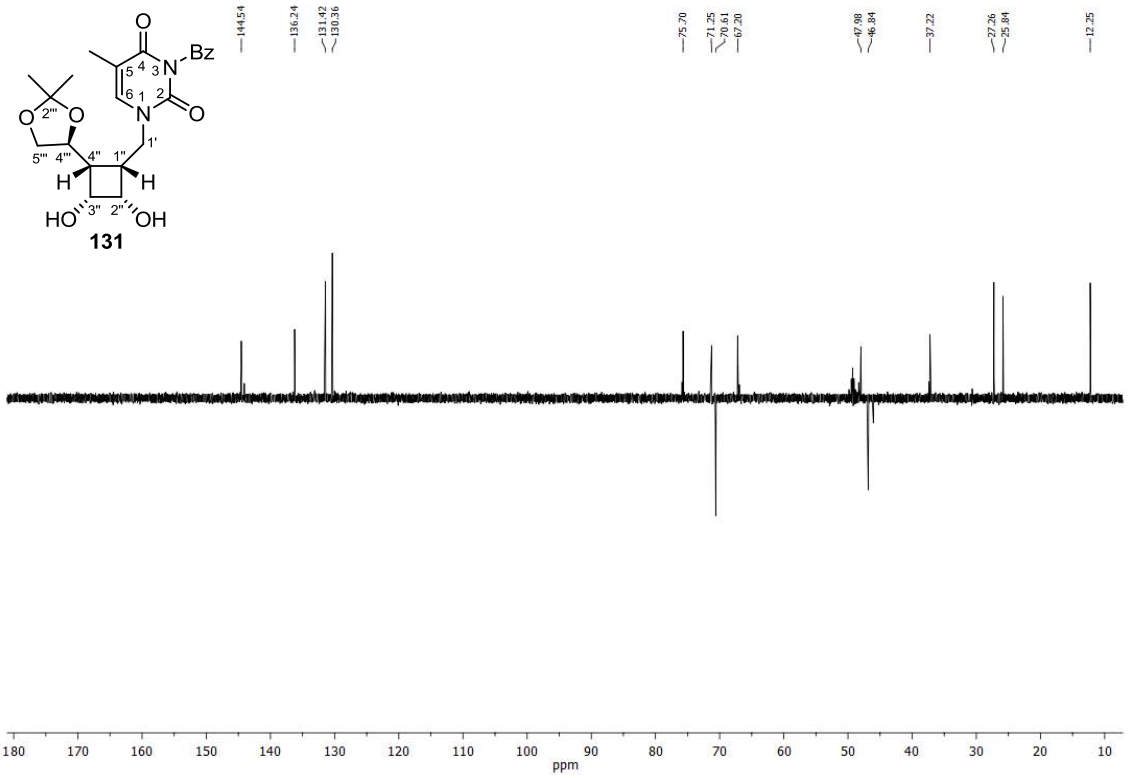
Selective TOCSY irradiated at H-1' frequency (360 MHz, CD_3OD)



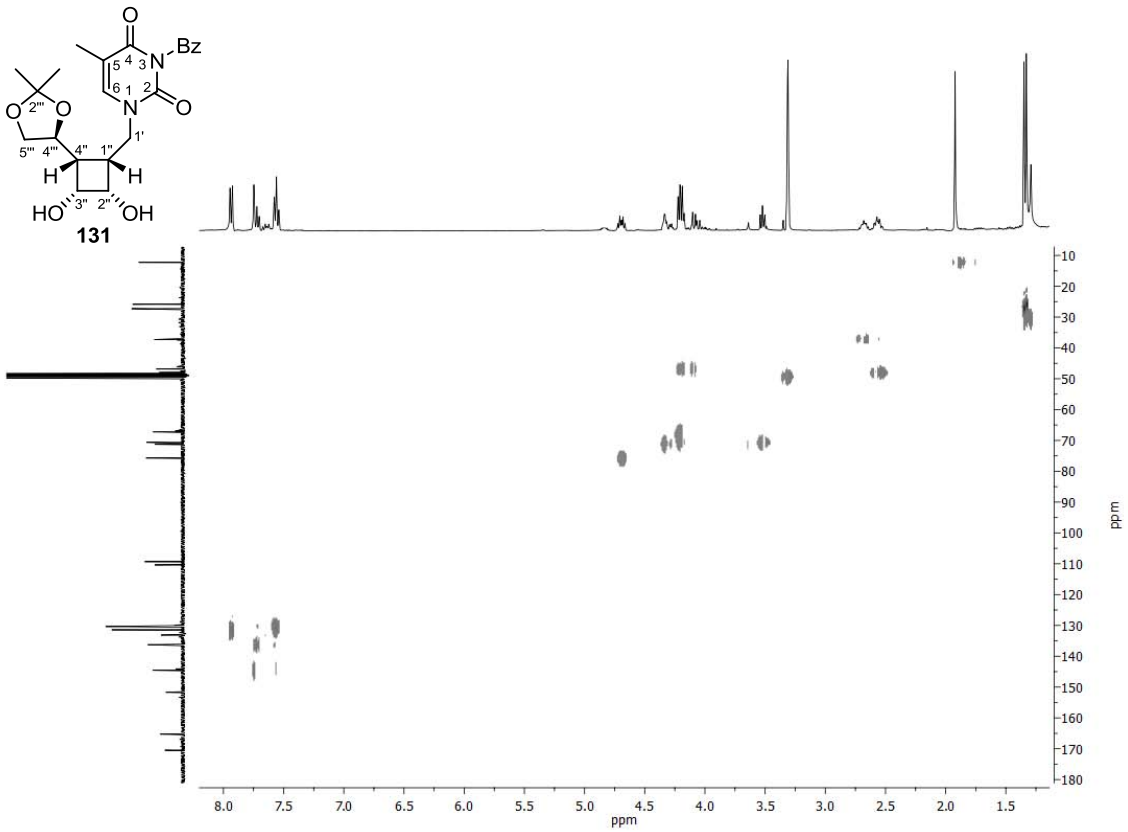
^{1}H-NMR (400 MHz, CD_3OD)



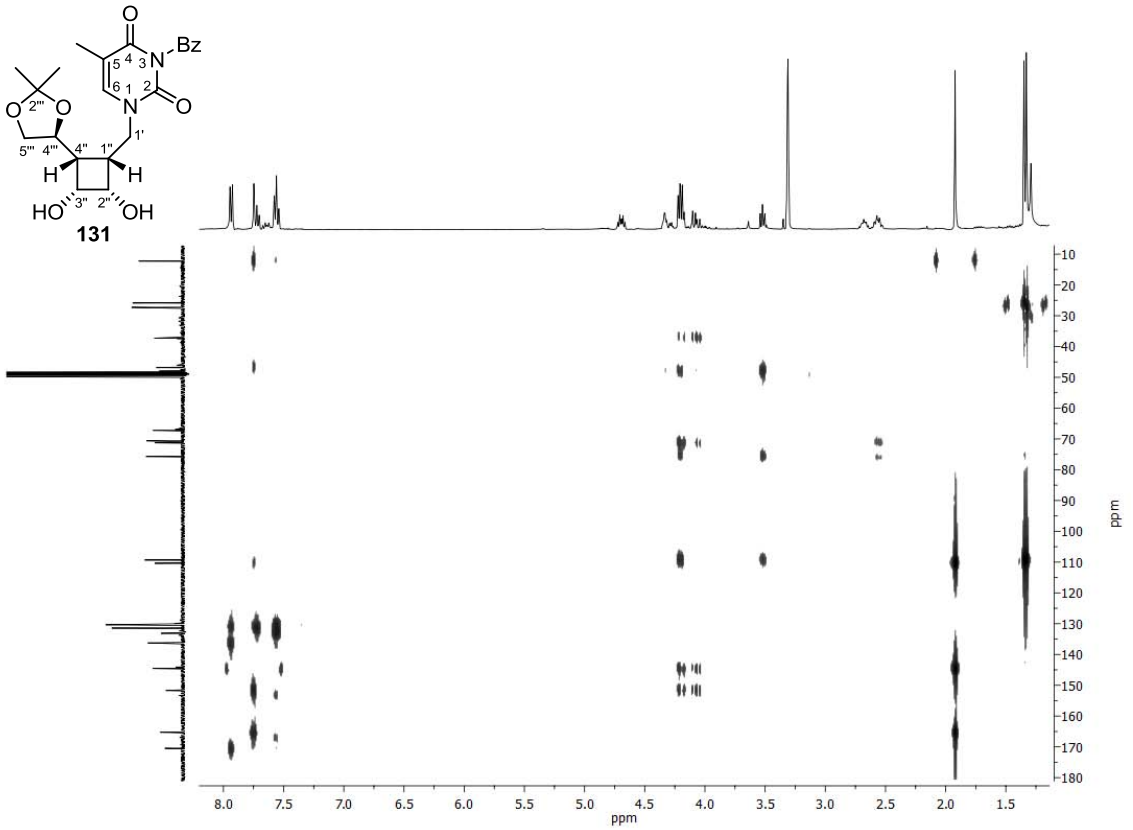
^{13}C-NMR (100 MHz, CD_3OD)



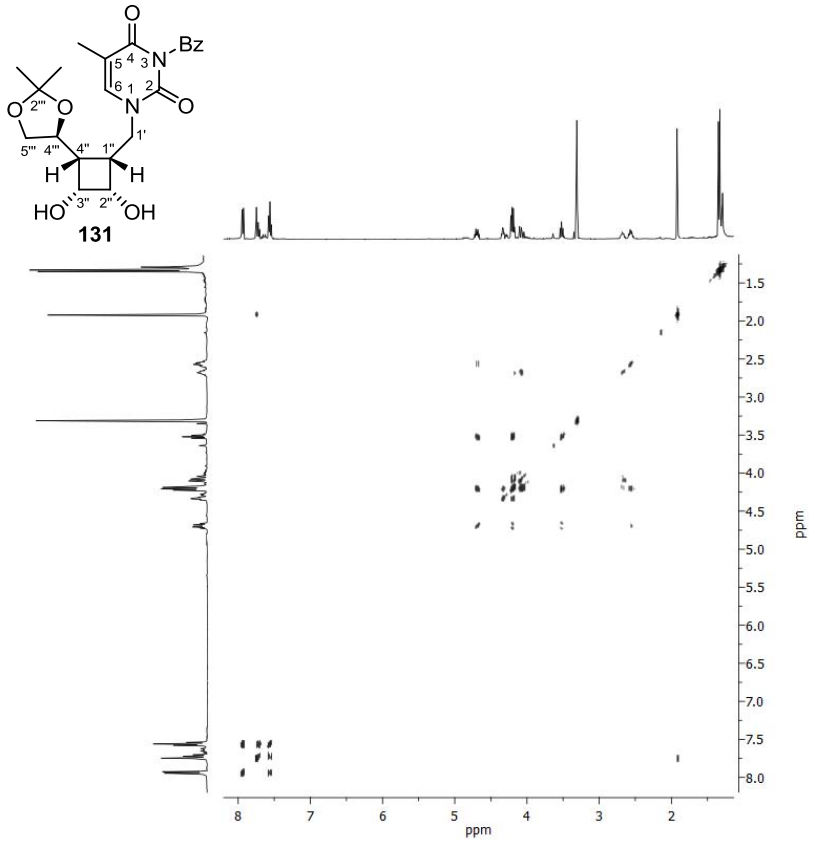
DEPT135 (400 MHz, CD_3OD)



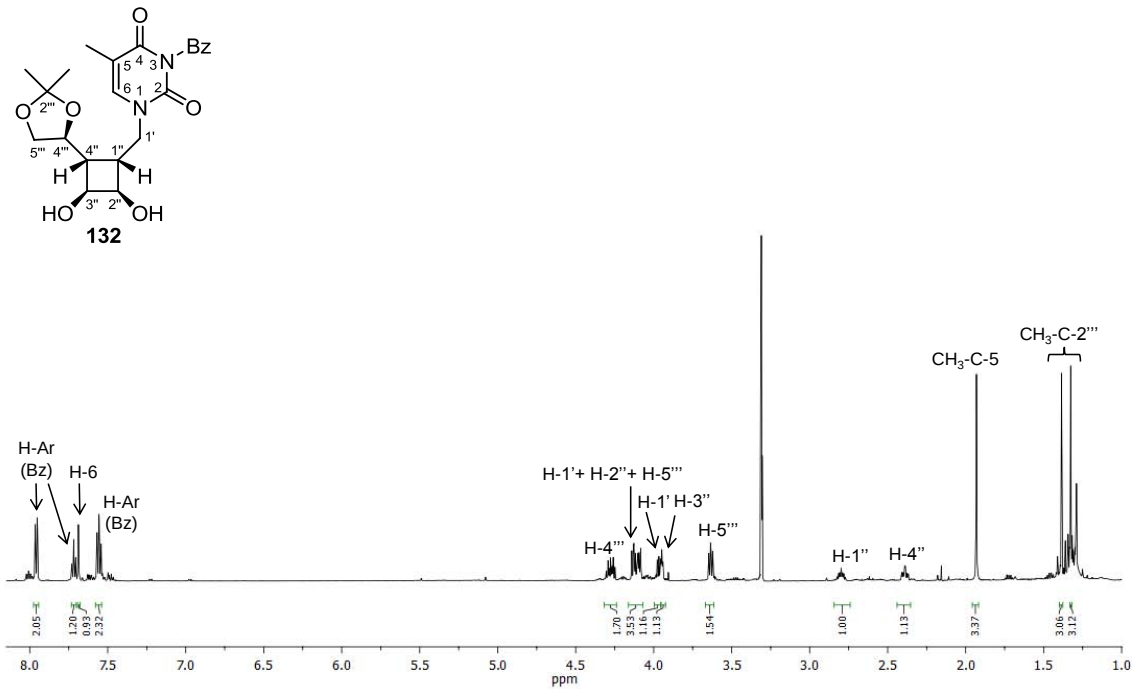
HSQC (400 MHz, CD_3OD)



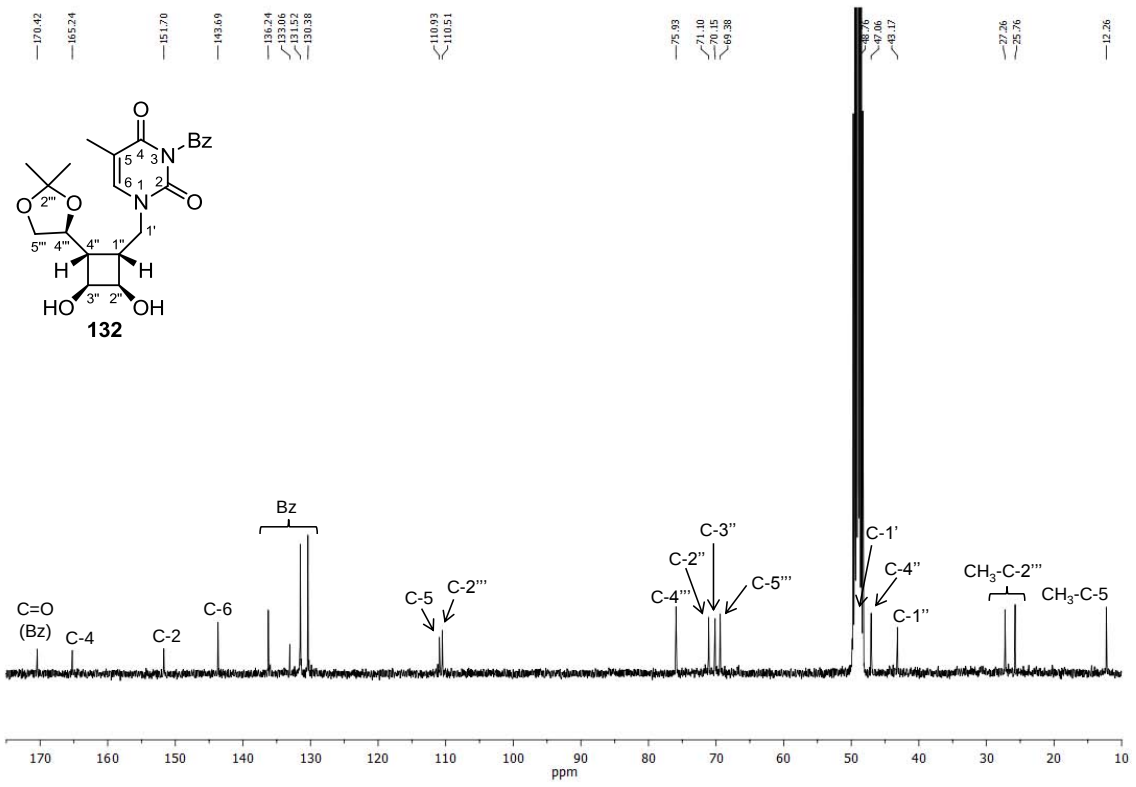
HMBC (400 MHz, CD_3OD)



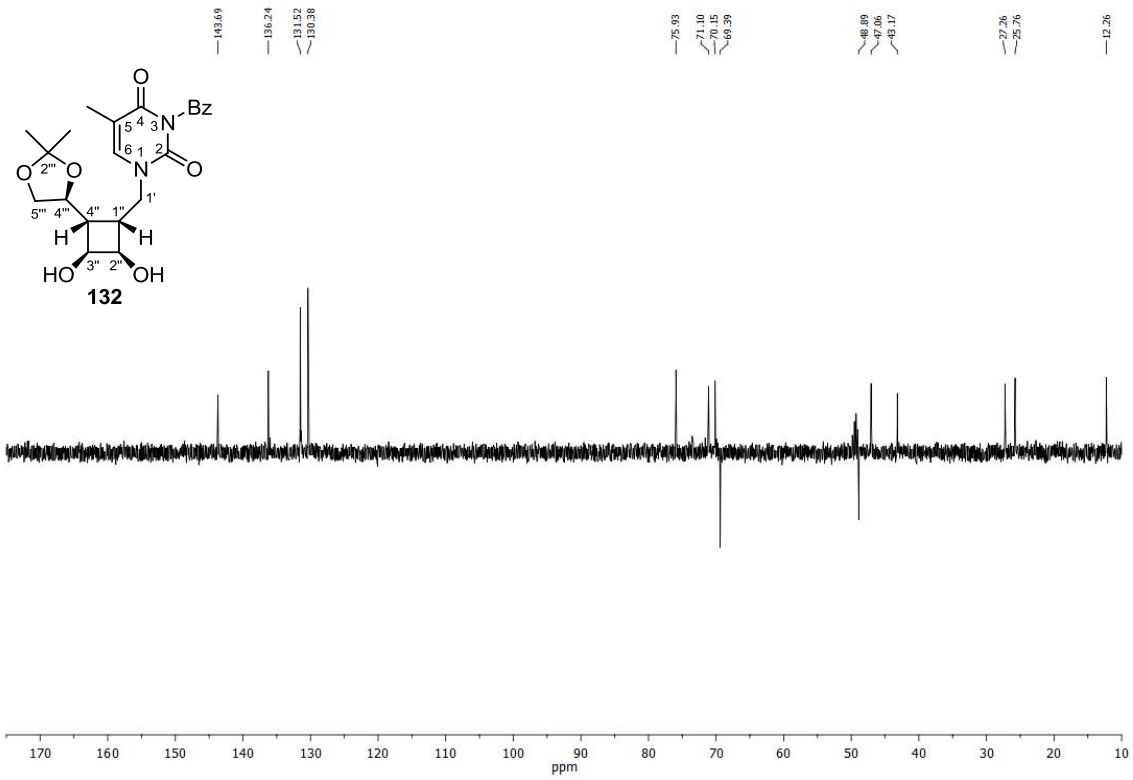
COSY (400 MHz, CD_3OD)



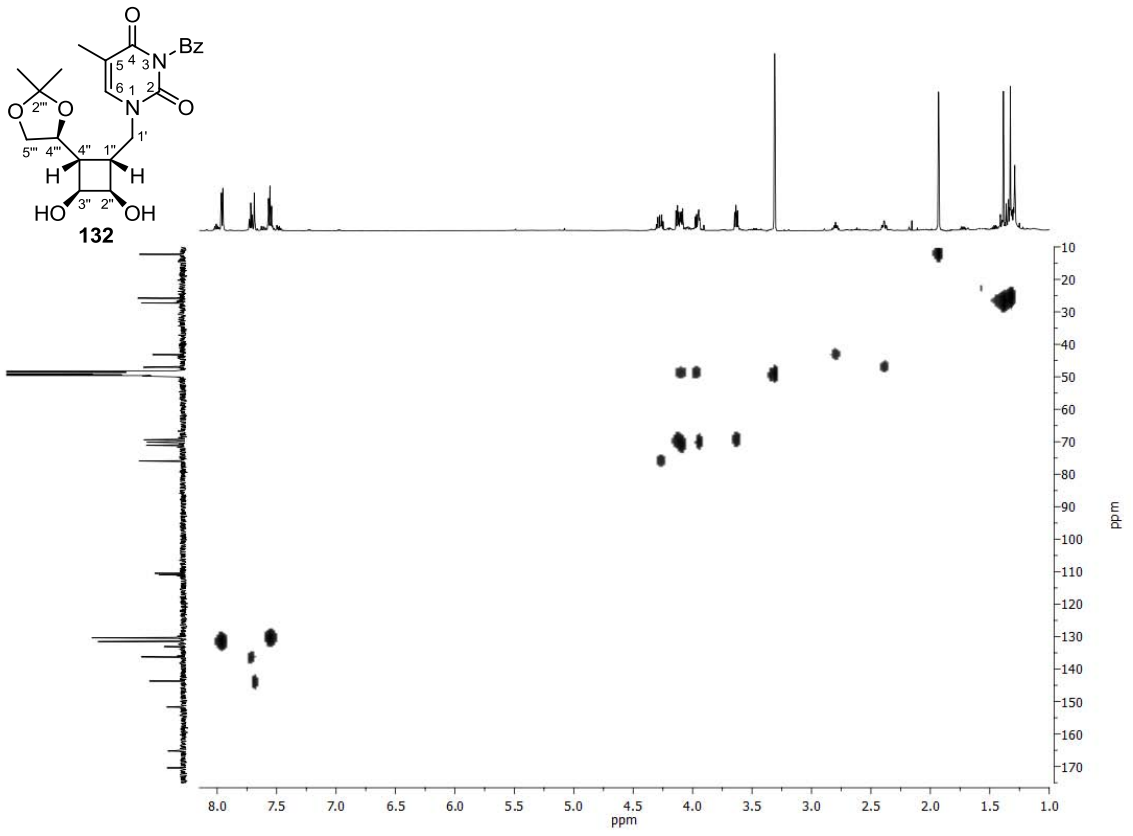
^{1}H-NMR (600 MHz, CD_3OD)



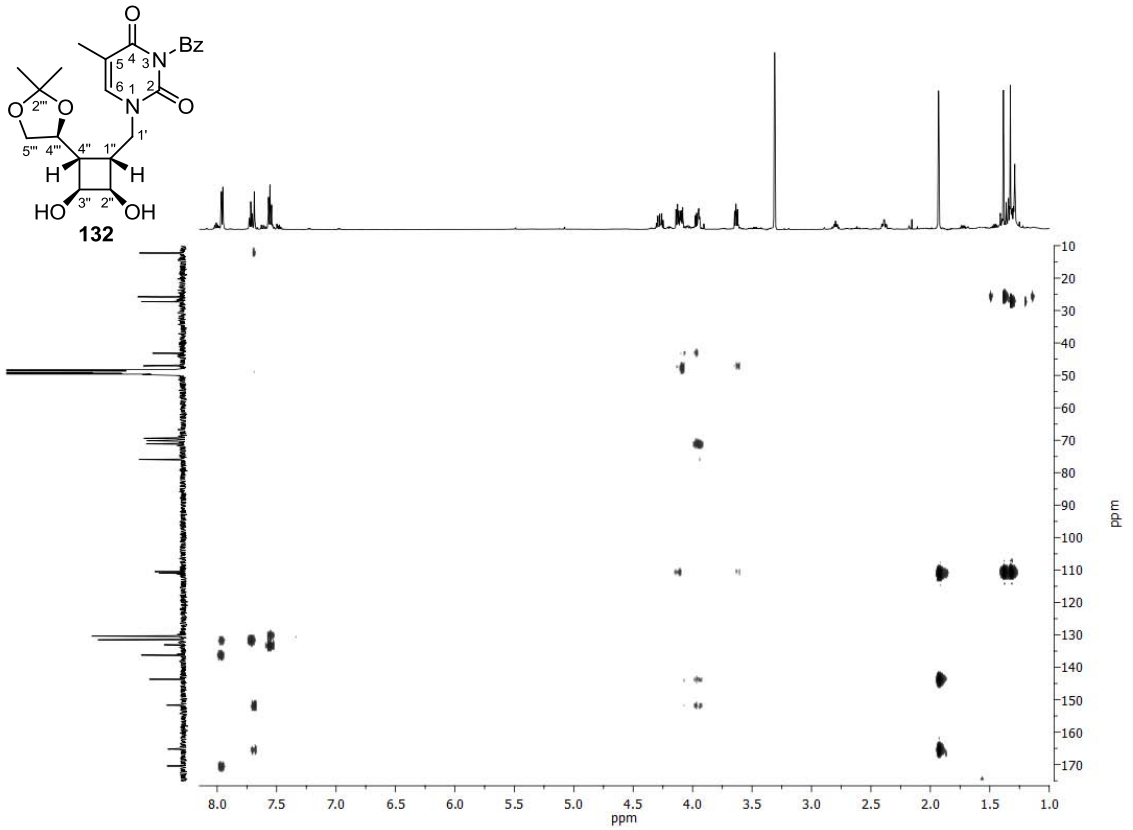
^{13}C-NMR (150 MHz, CD_3OD)



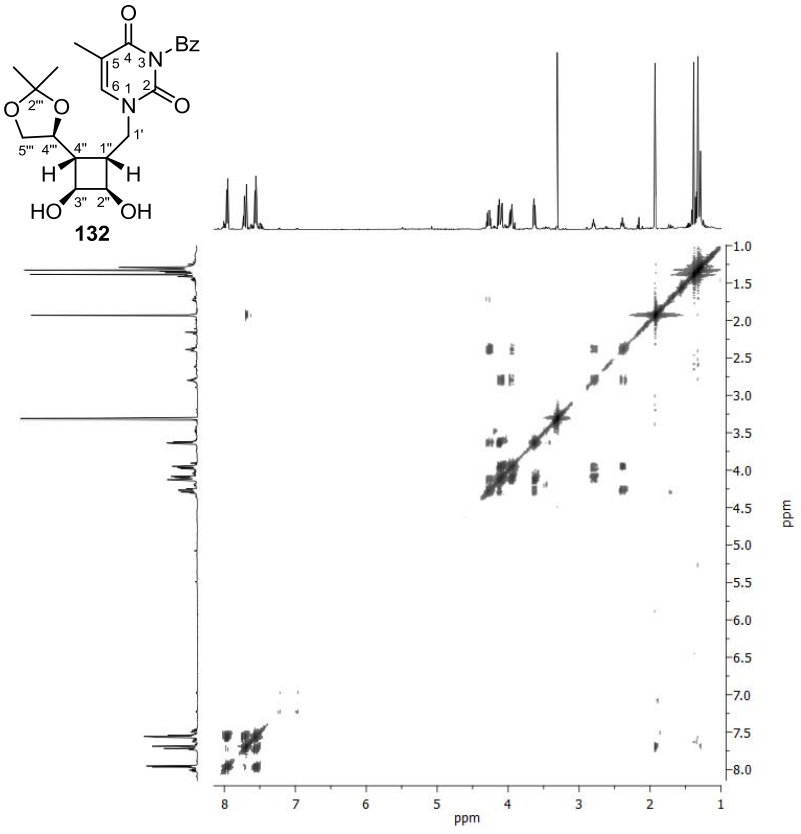
DEPT135 (600 MHz, CD_3OD)



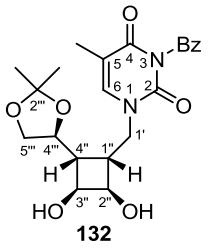
HSQC (600 MHz, CD_3OD)



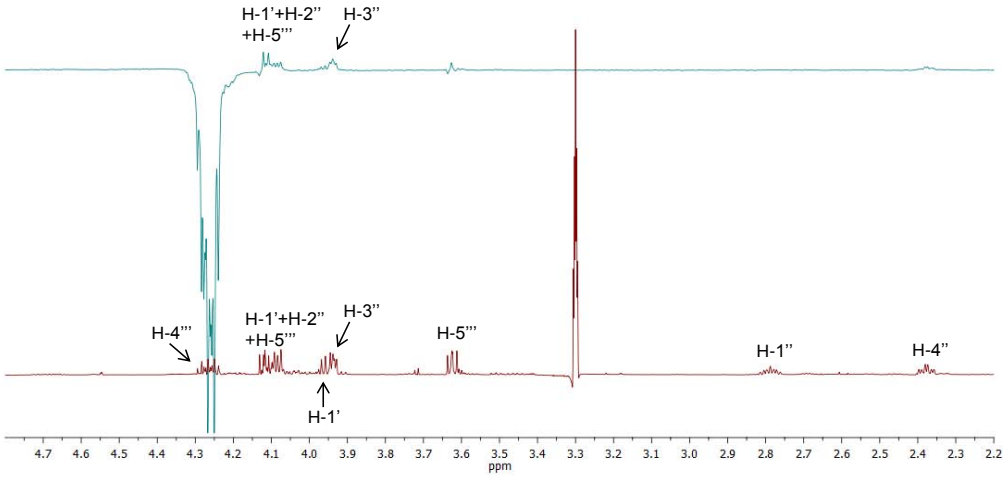
HMBC (600 MHz, CD_3OD)



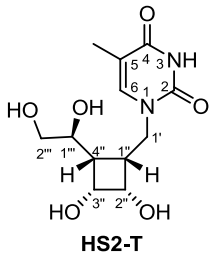
COSY (600 MHz, CD_3OD)



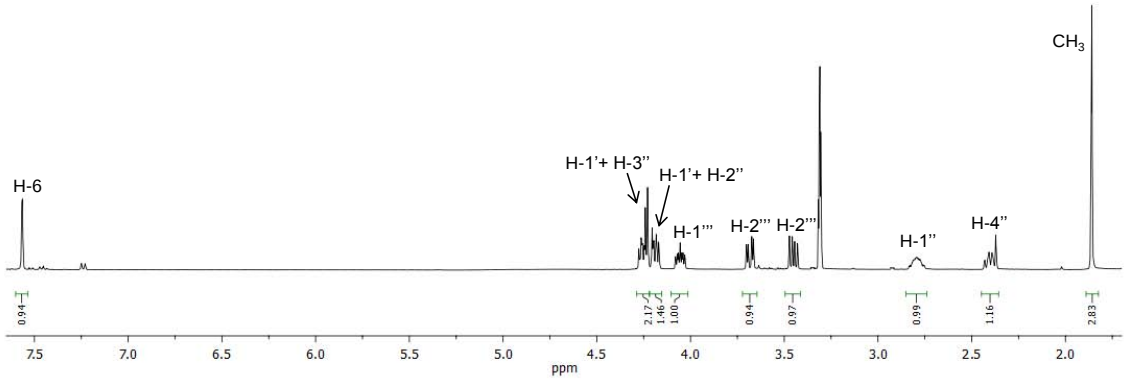
132



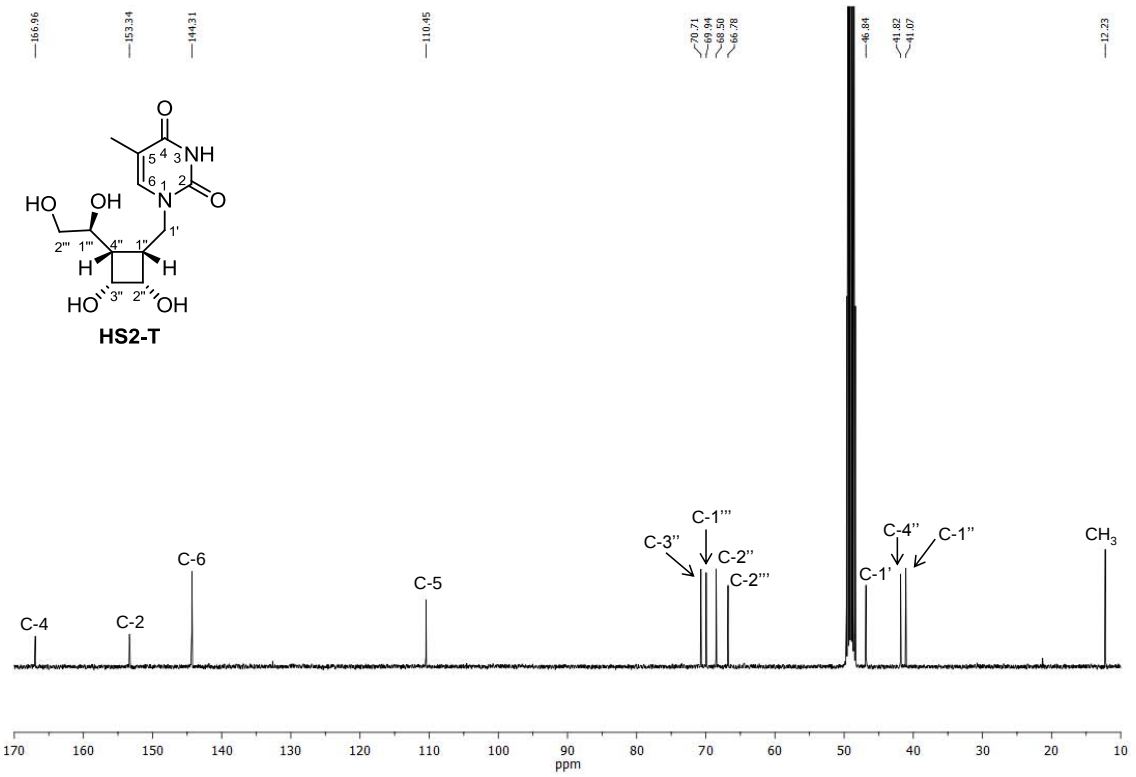
^{1}H-NMR (red) and selective n.O.e. irradiated at H-4''' frequency (blue) (600 MHz, CD_3OD)



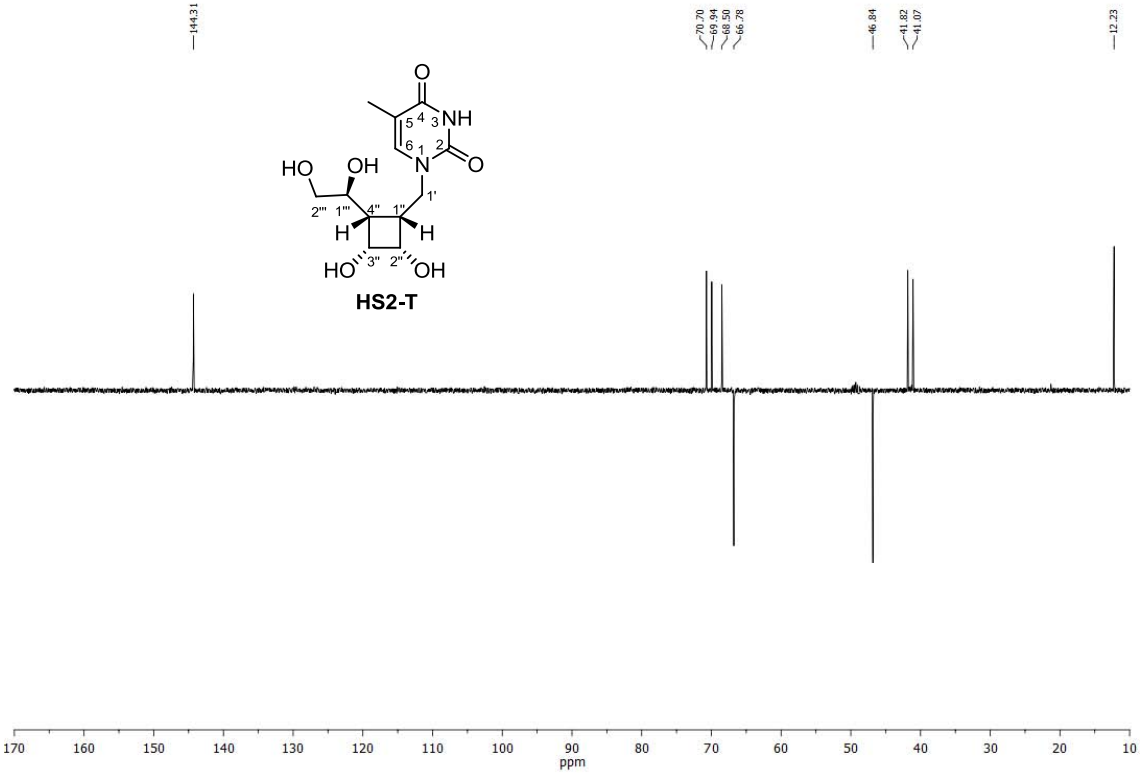
HS2-T



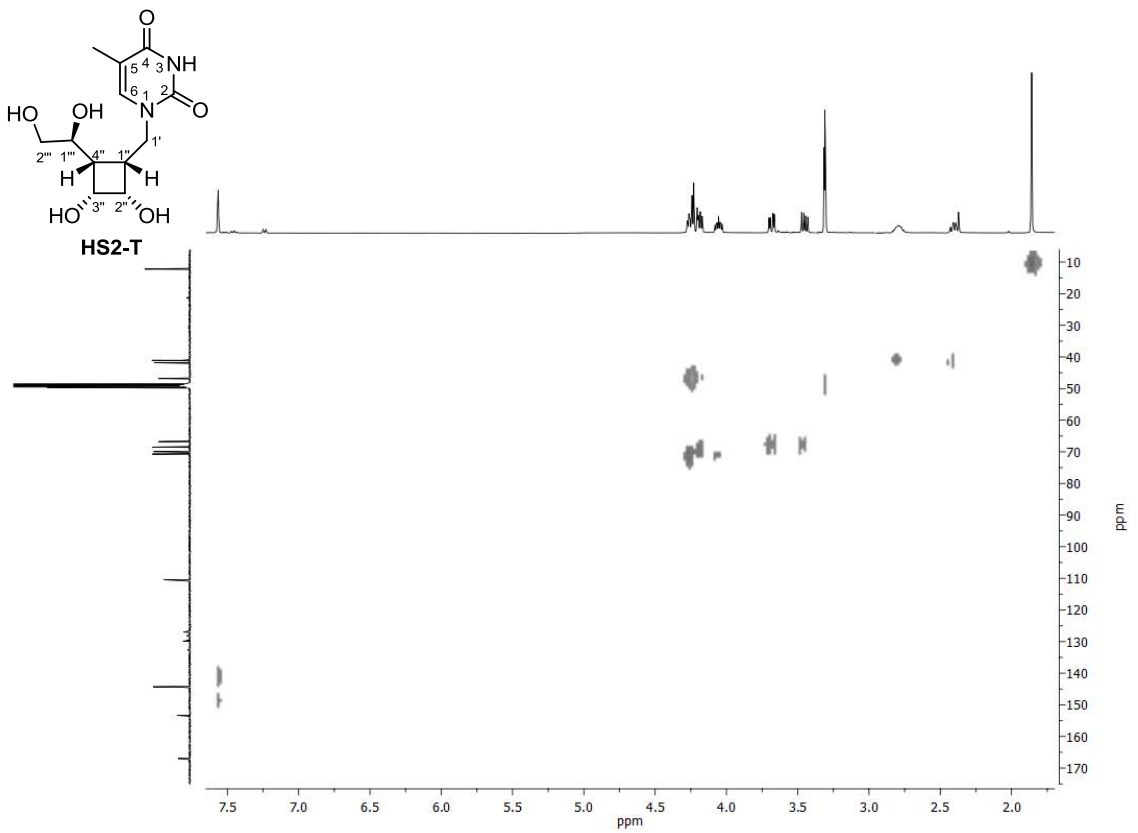
^{1}H-NMR (400 MHz, CD_3OD)



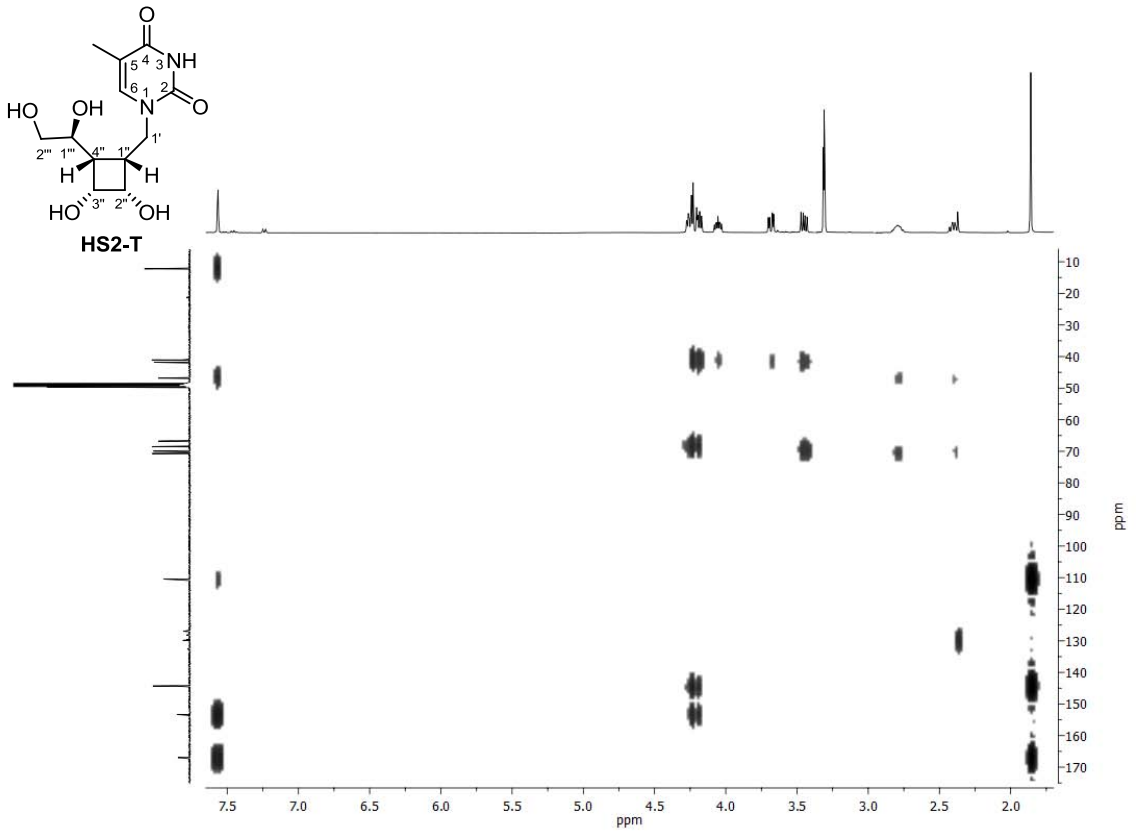
^{13}C-NMR (100 MHz, CD_3OD)



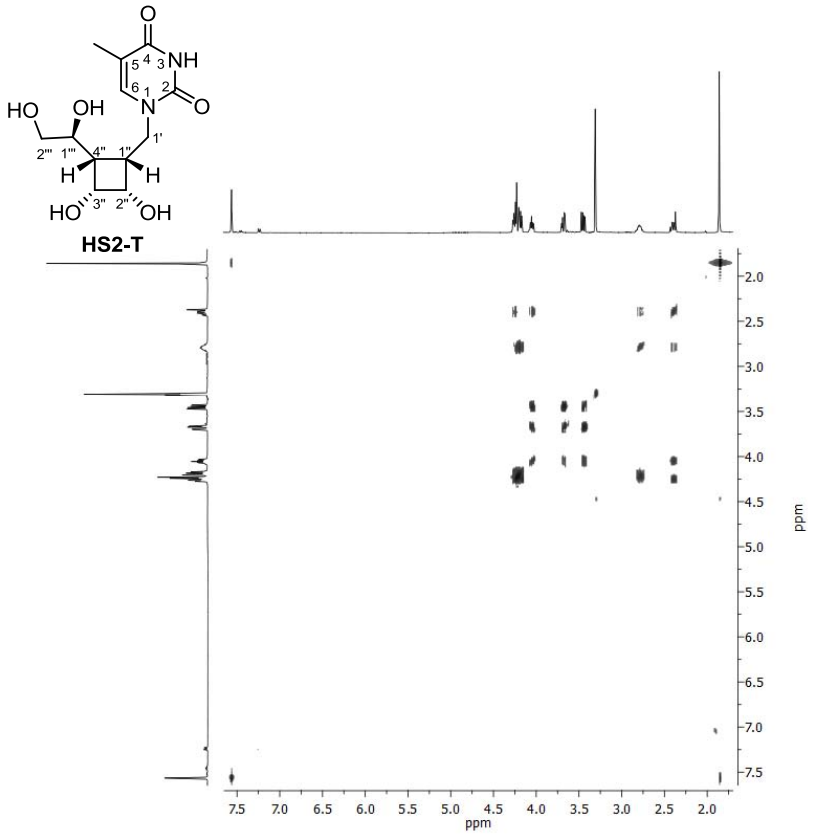
DEPT135 (400 MHz, CD_3OD)



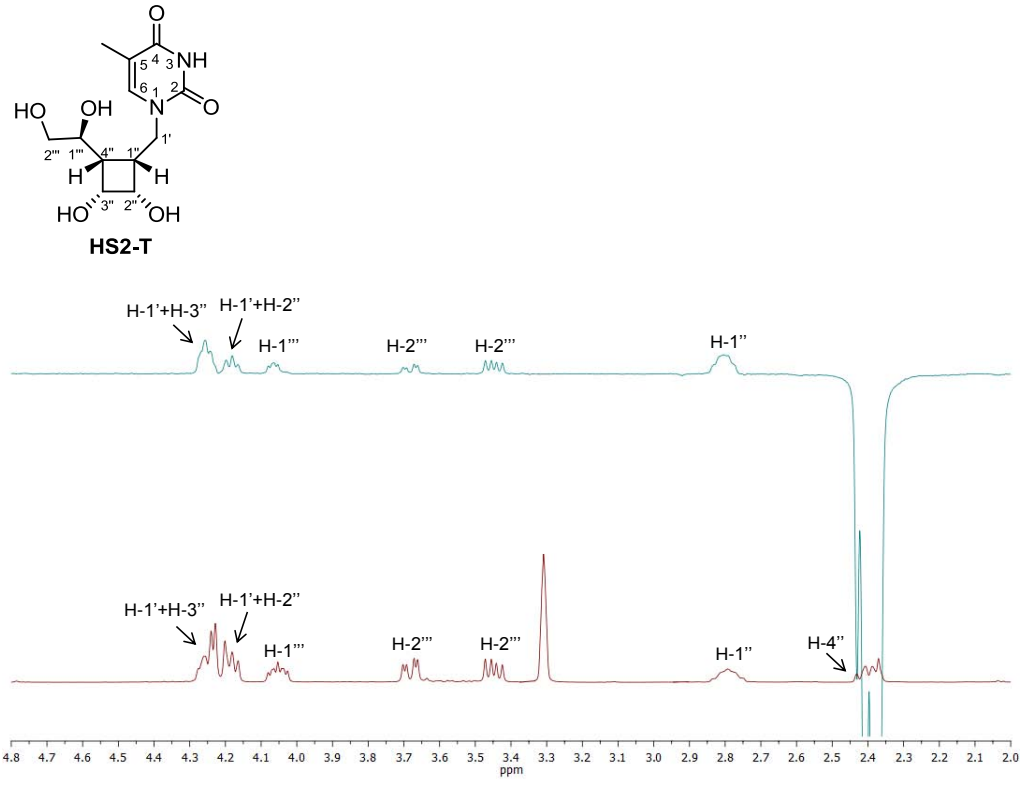
HSQC (400 MHz, CD_3OD)



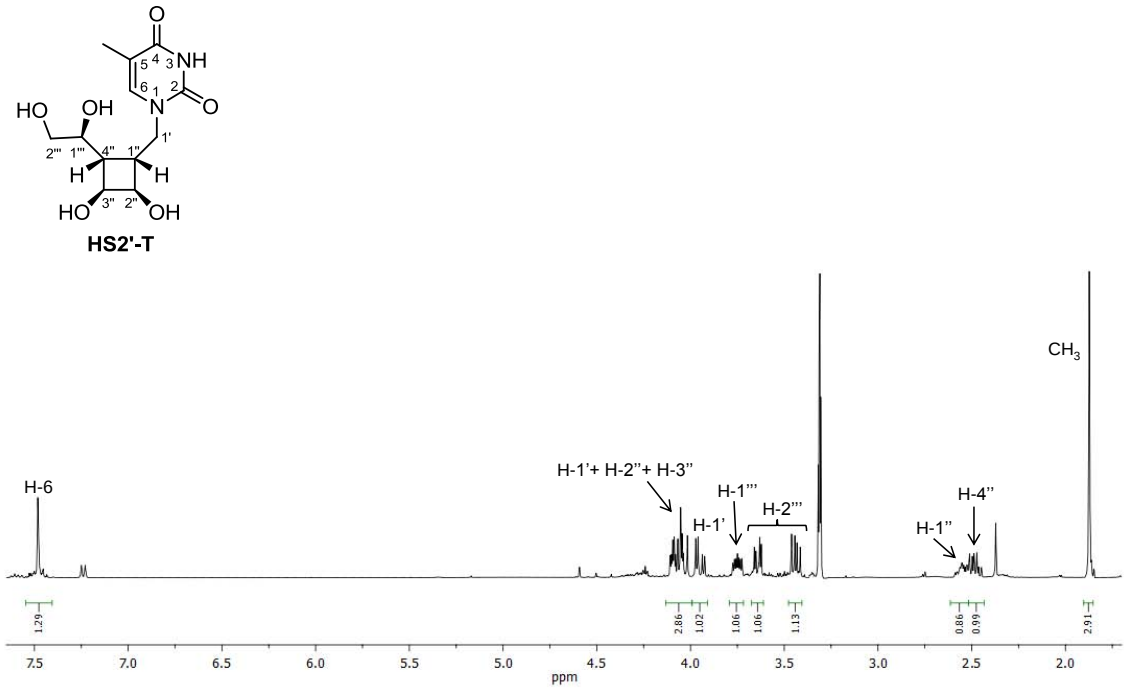
HMBC (400 MHz, CD_3OD)



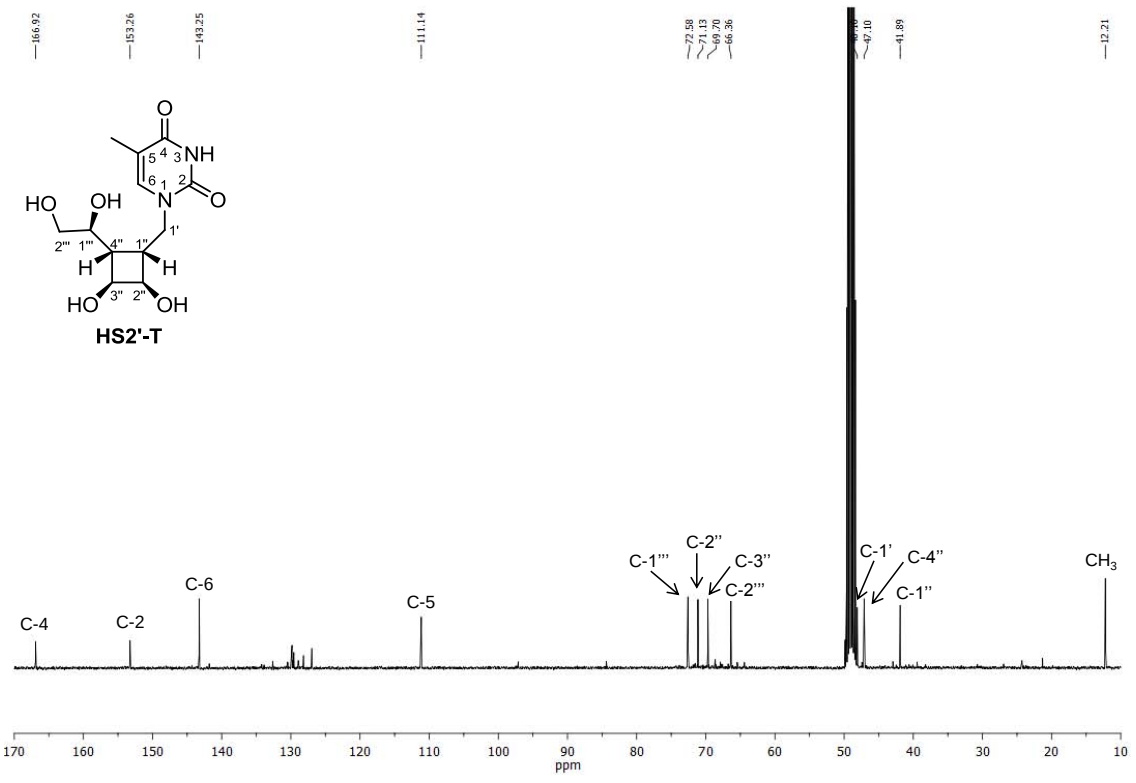
COSY (400 MHz, CD_3OD)



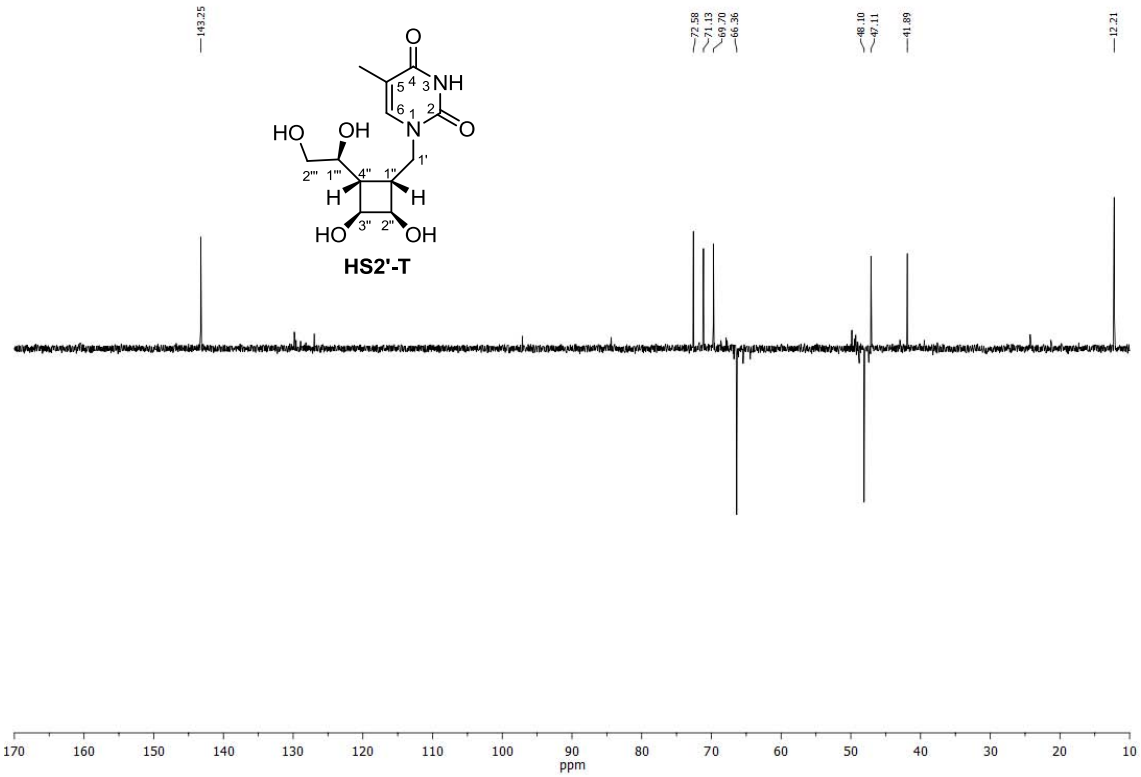
^{1}H-NMR (red) and selective n.O.e. irradiated at H-4'' frequency (blue) (360 MHz, CD_3OD)



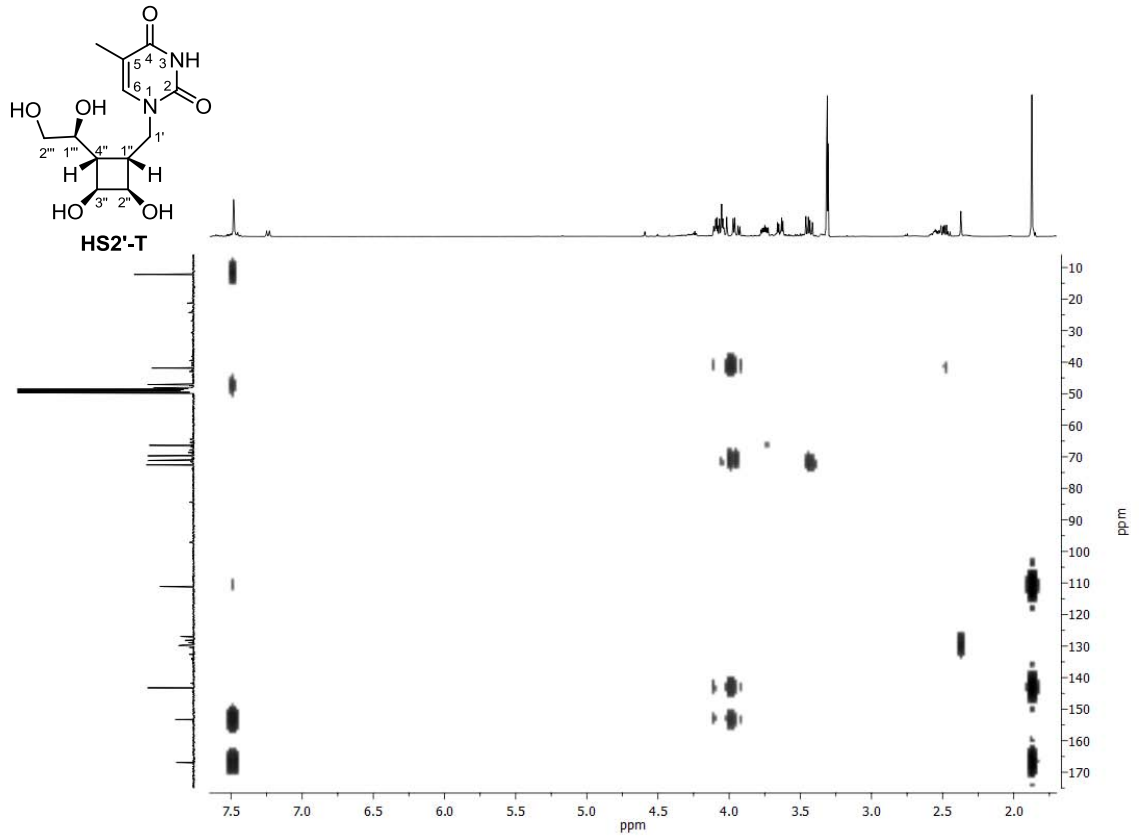
^{1}H-NMR (400 MHz, CD_3OD)



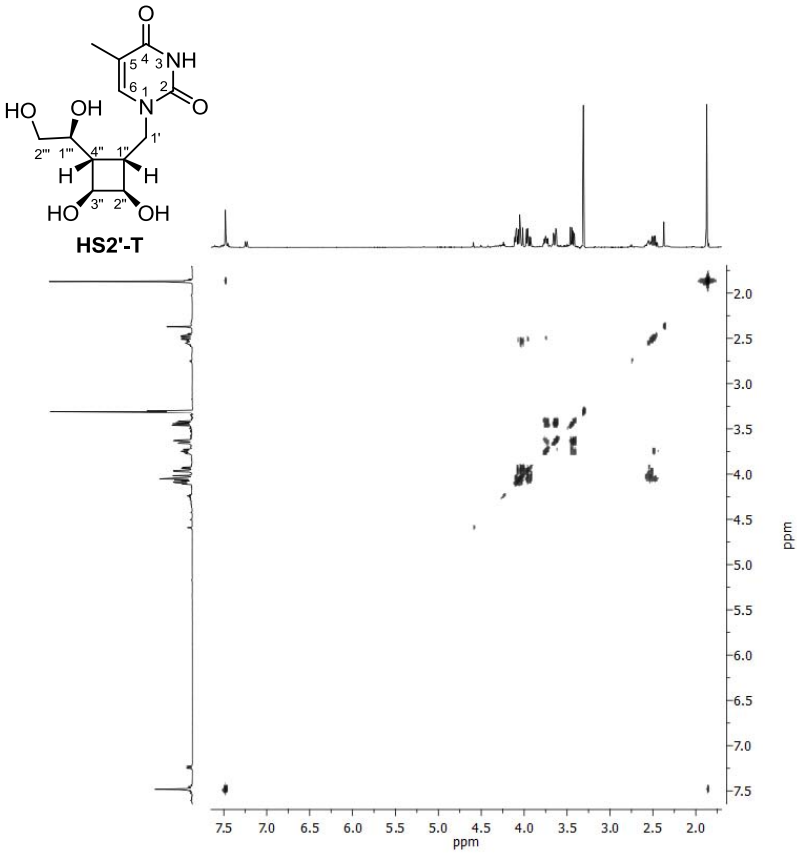
^{13}C-NMR (100 MHz, CD_3OD)



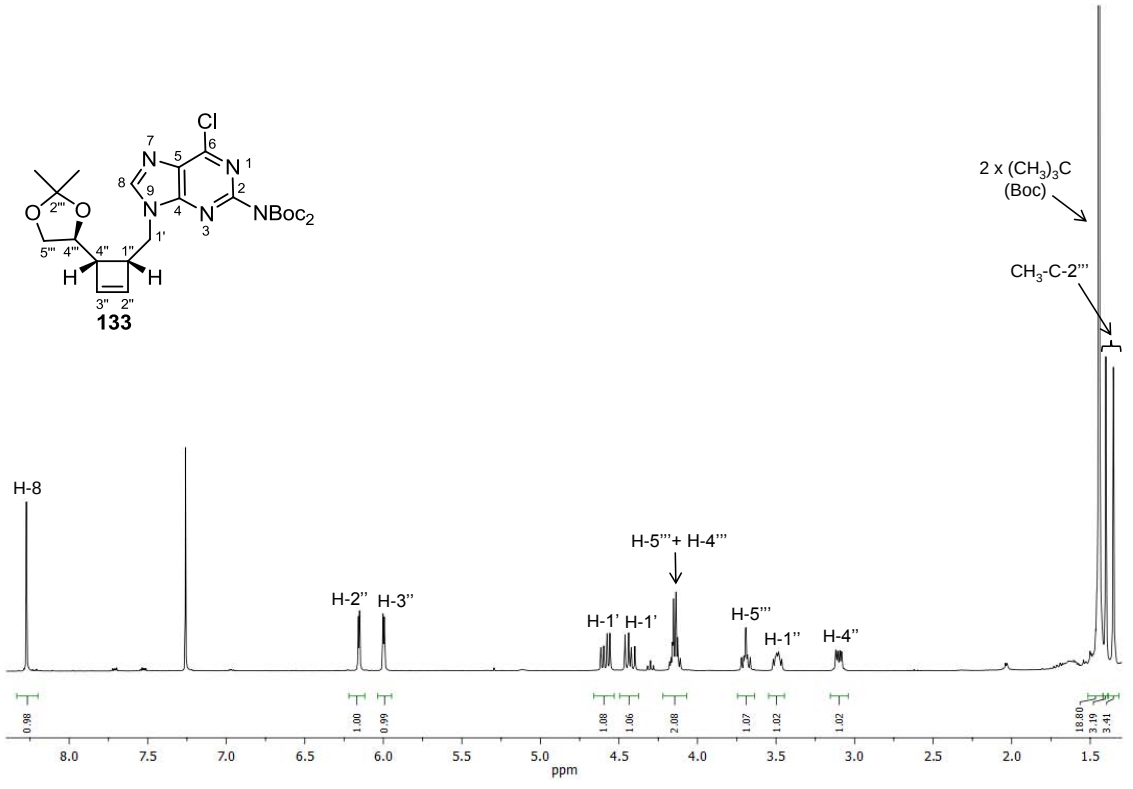
DEPT135 (400 MHz, CD_3OD)



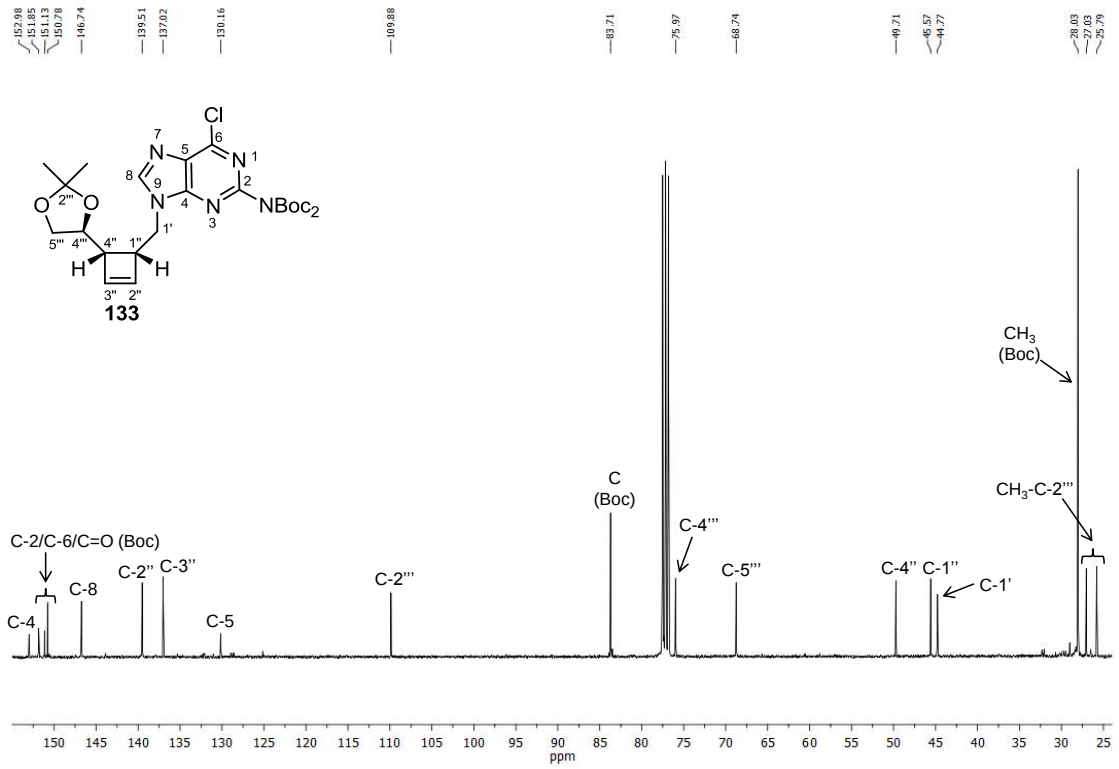
HMBC (400 MHz, CD_3OD)



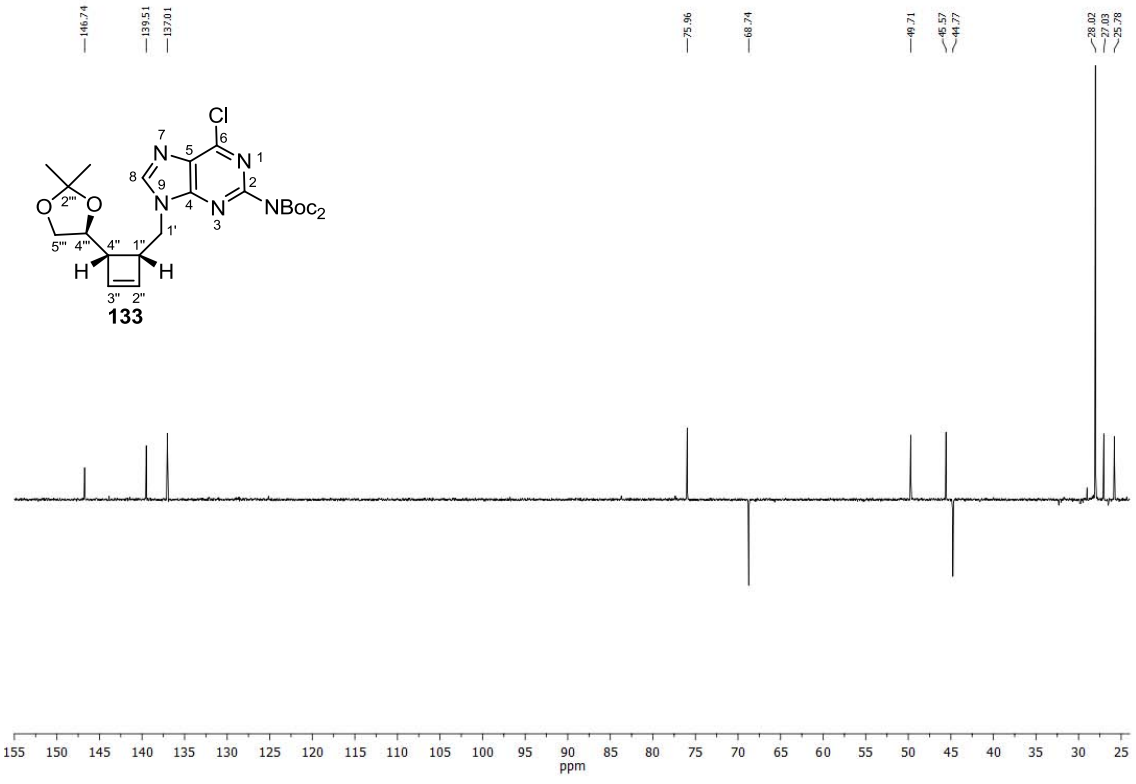
COSY (400 MHz, CD_3OD)



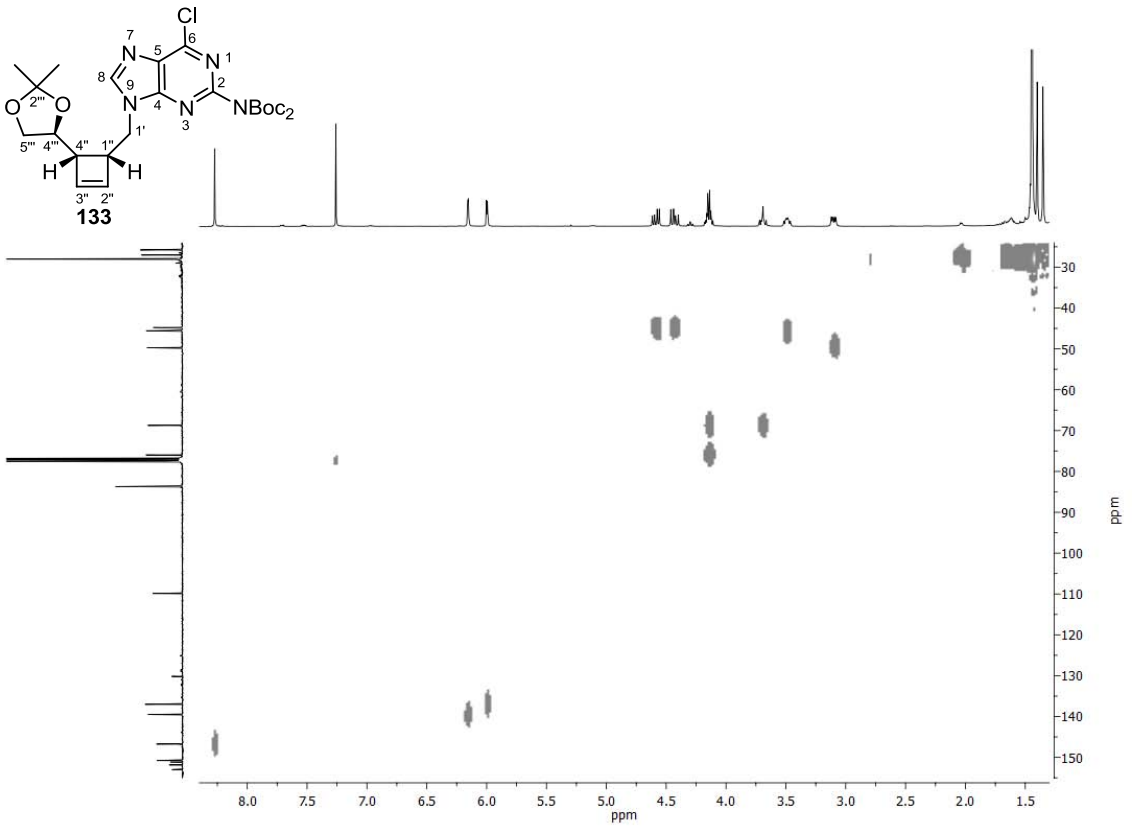
^{1}H-NMR (360 MHz, $CDCl_3$)



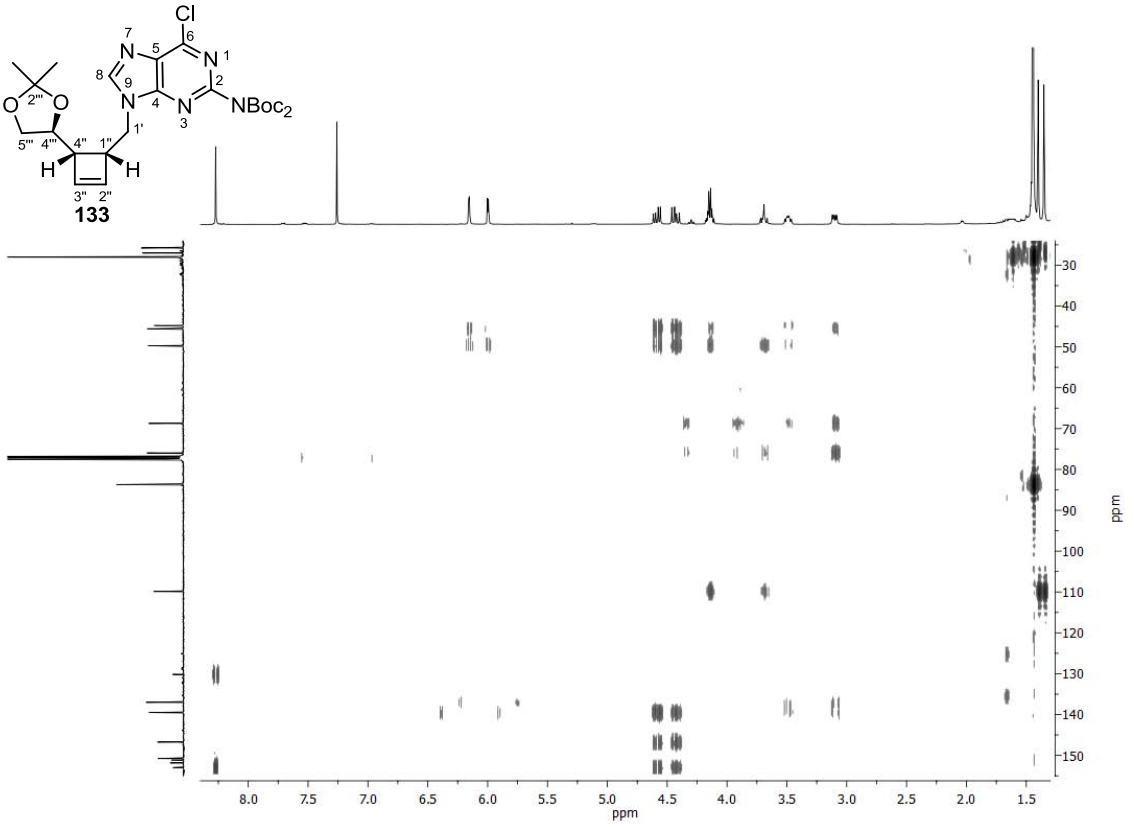
^{13}C-NMR (90 MHz, $CDCl_3$)



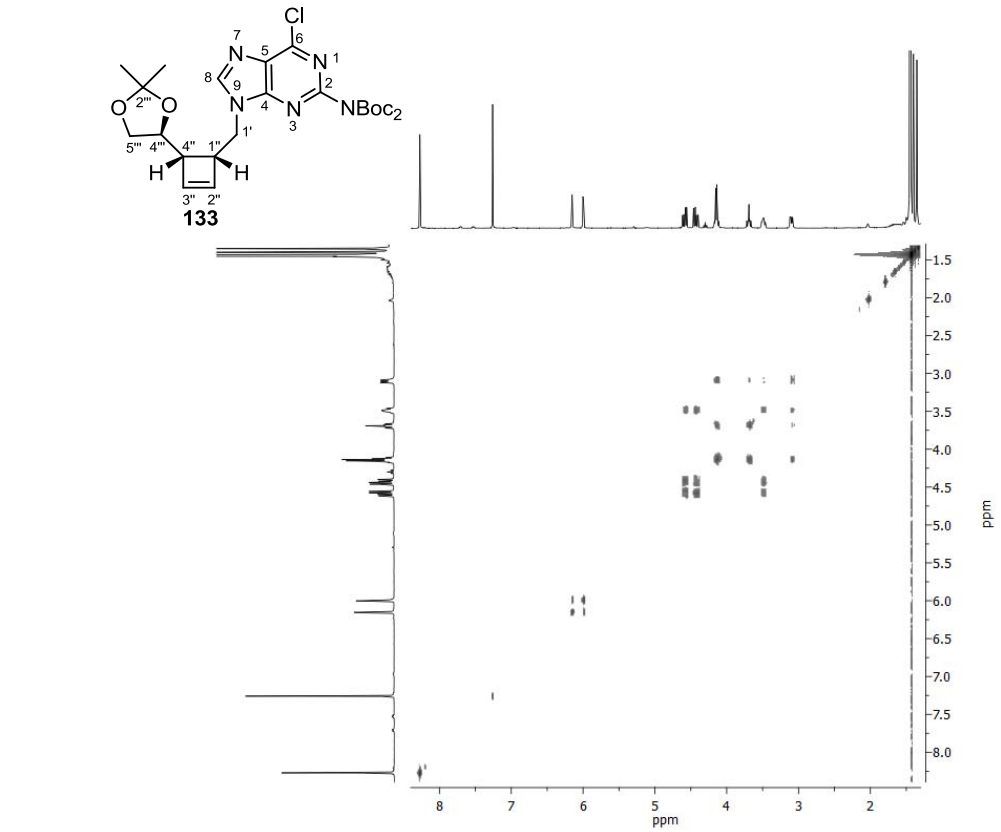
DEPT135 (360 MHz, $CDCl_3$)



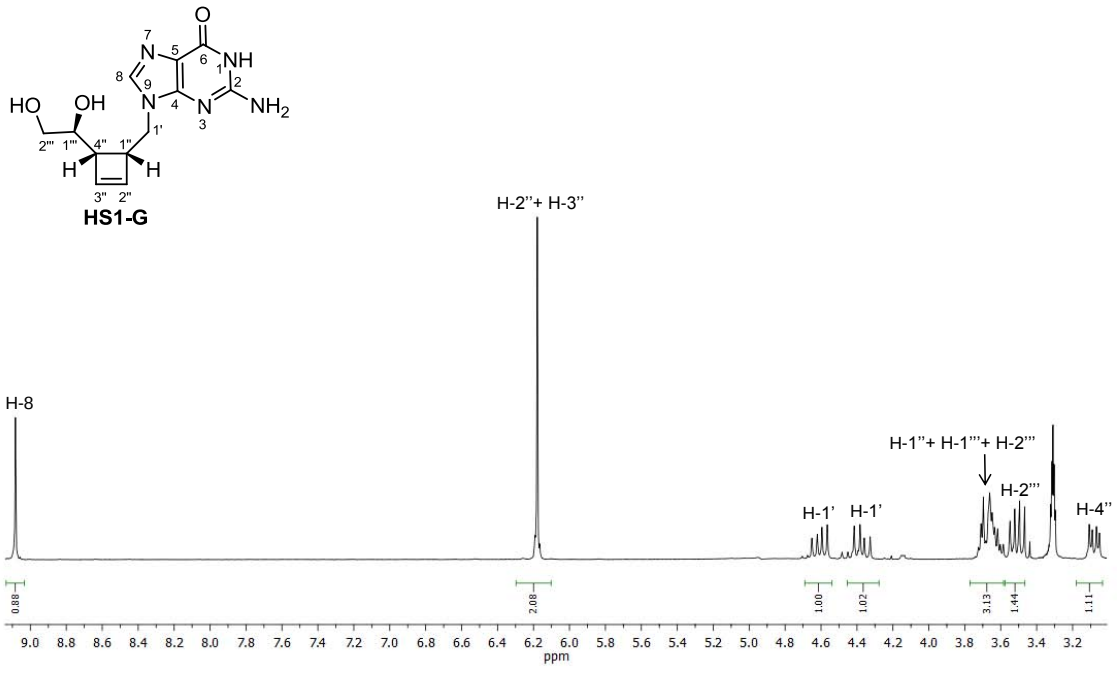
HSQC (360 MHz, $CDCl_3$)



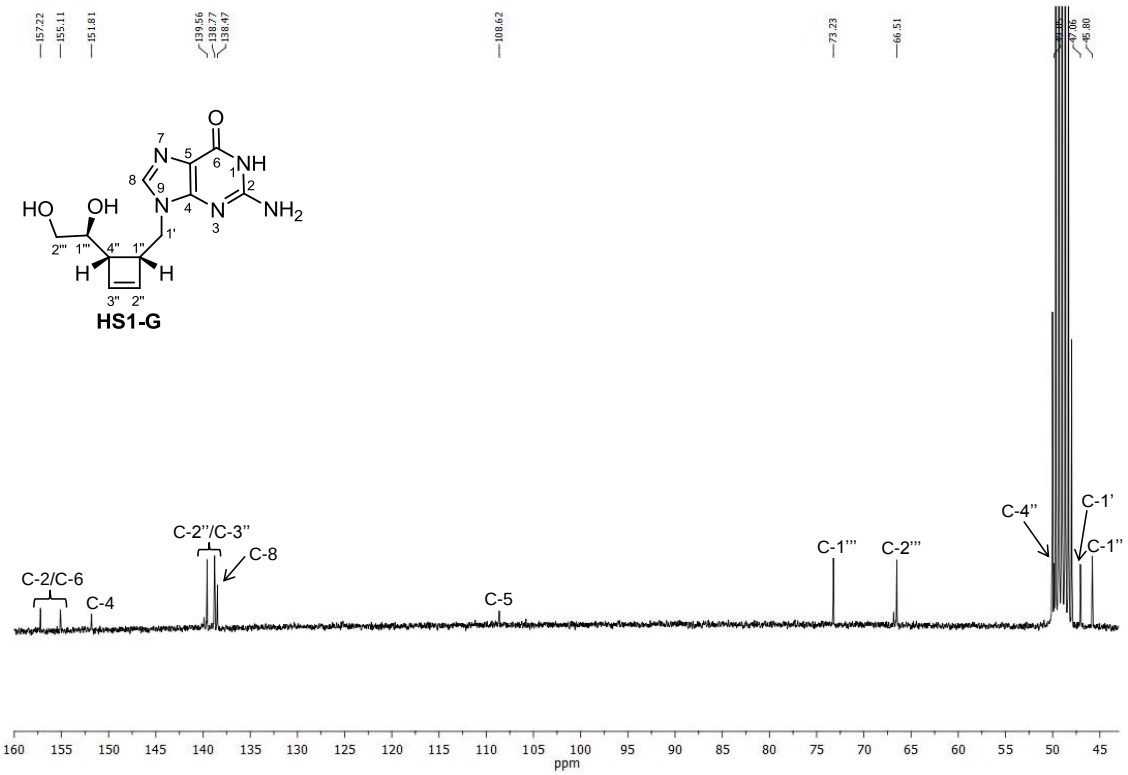
HMBC (360 MHz, $CDCl_3$)



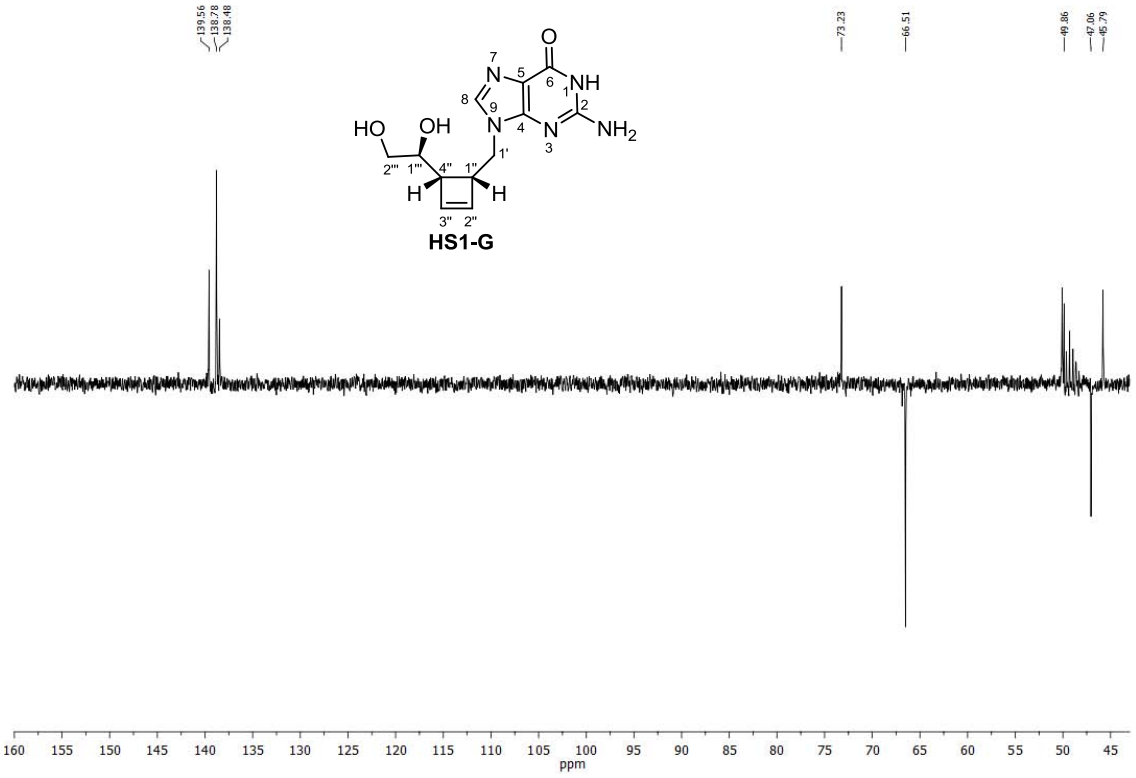
COSY (360 MHz, $CDCl_3$)



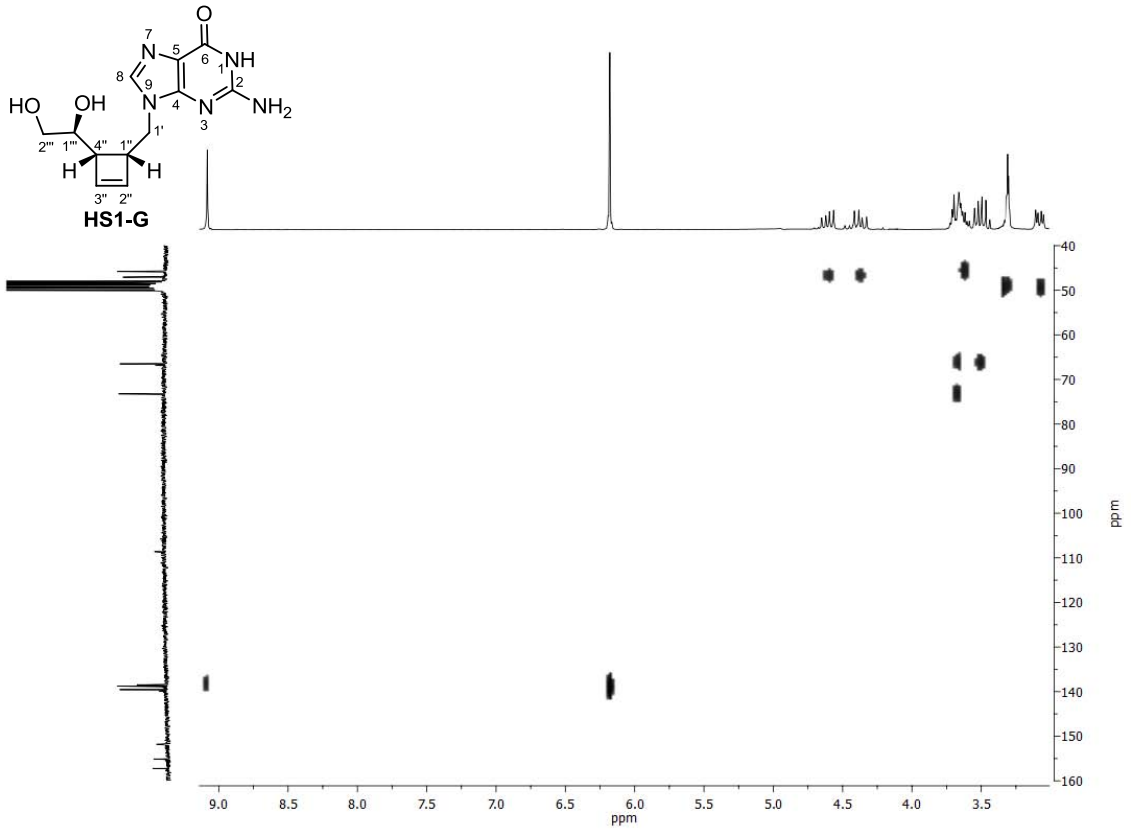
^{1}H-NMR (250 MHz, CD_3OD)



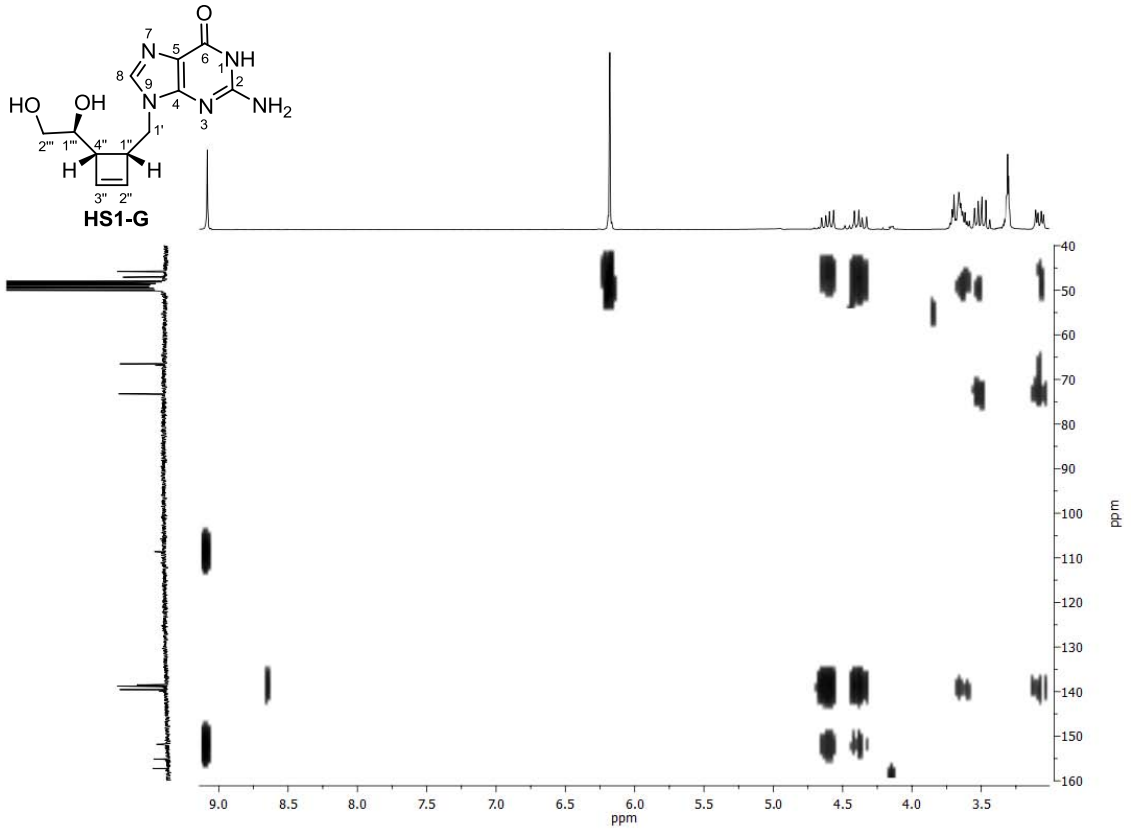
^{13}C-NMR (62.5 MHz, CD_3OD)



DEPT135 (250 MHz, CD_3OD)

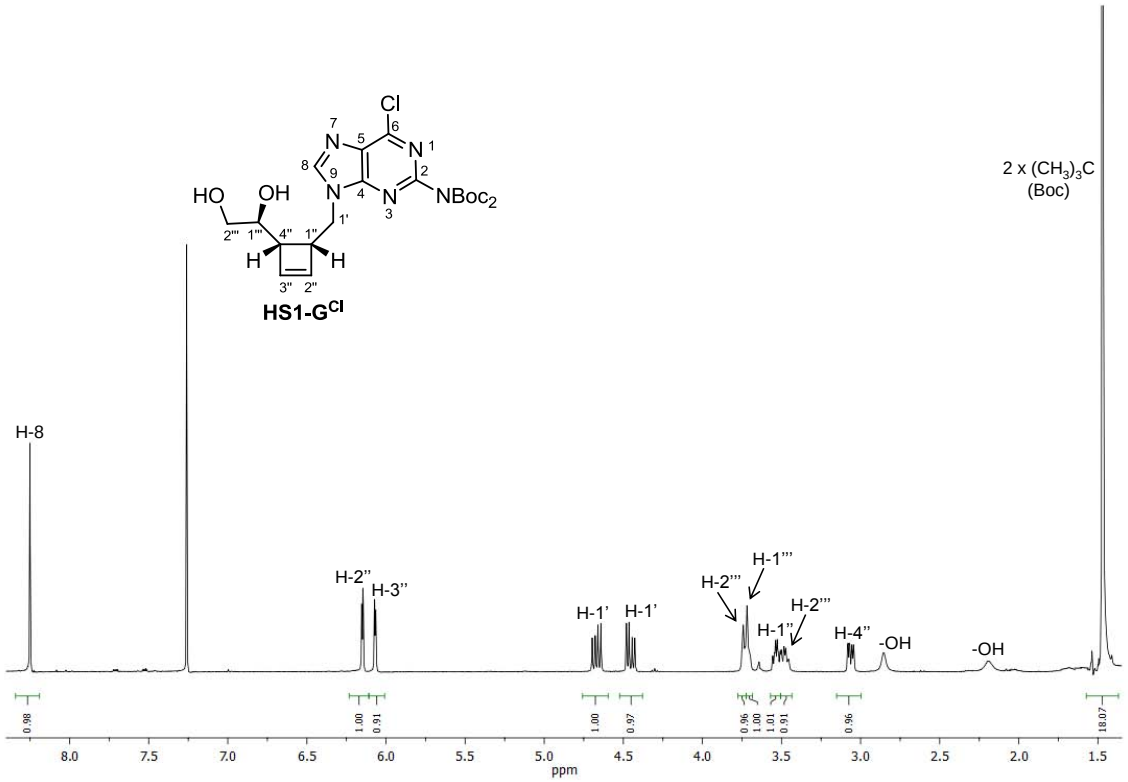


HSQC (250 MHz, CD_3OD)

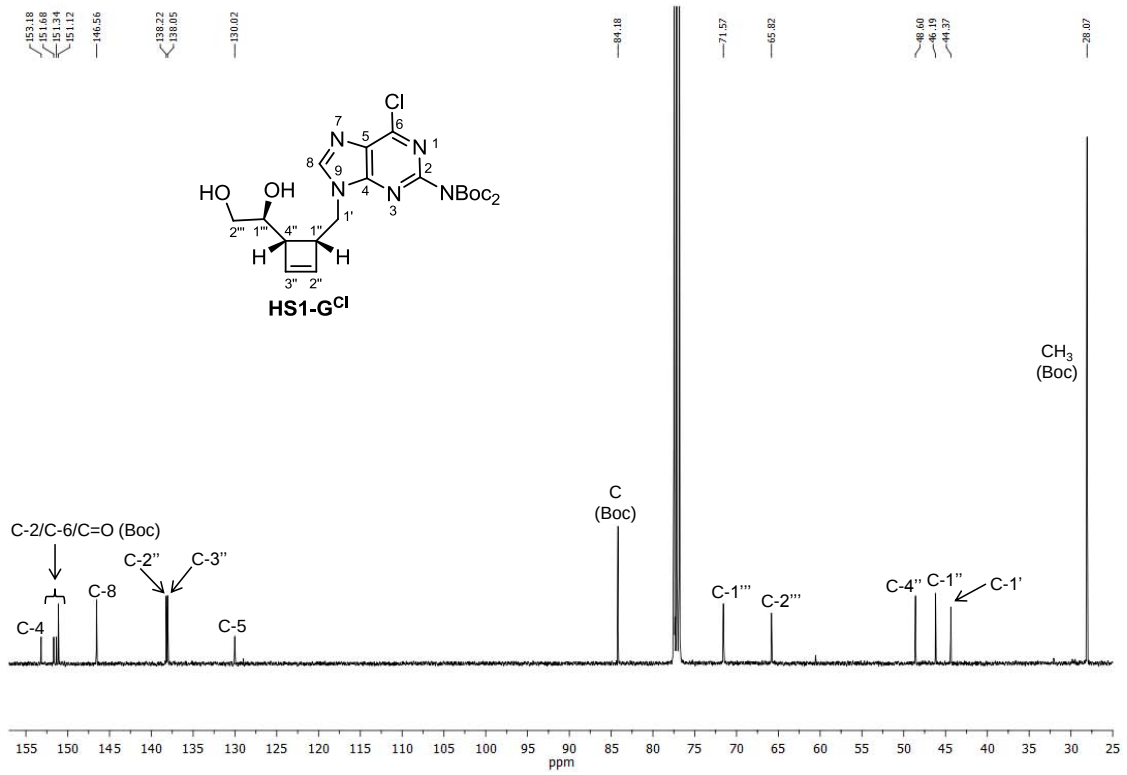


HMBC (250 MHz, CD_3OD)

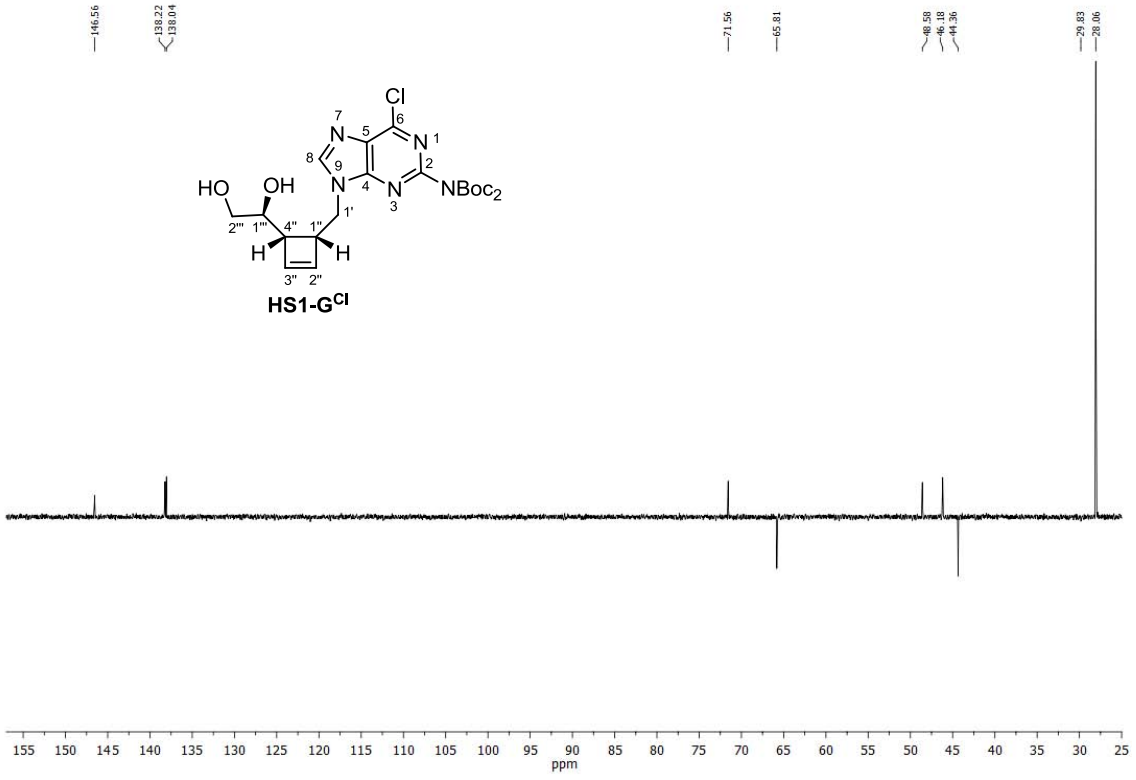
COSY (250 MHz, CD_3OD)



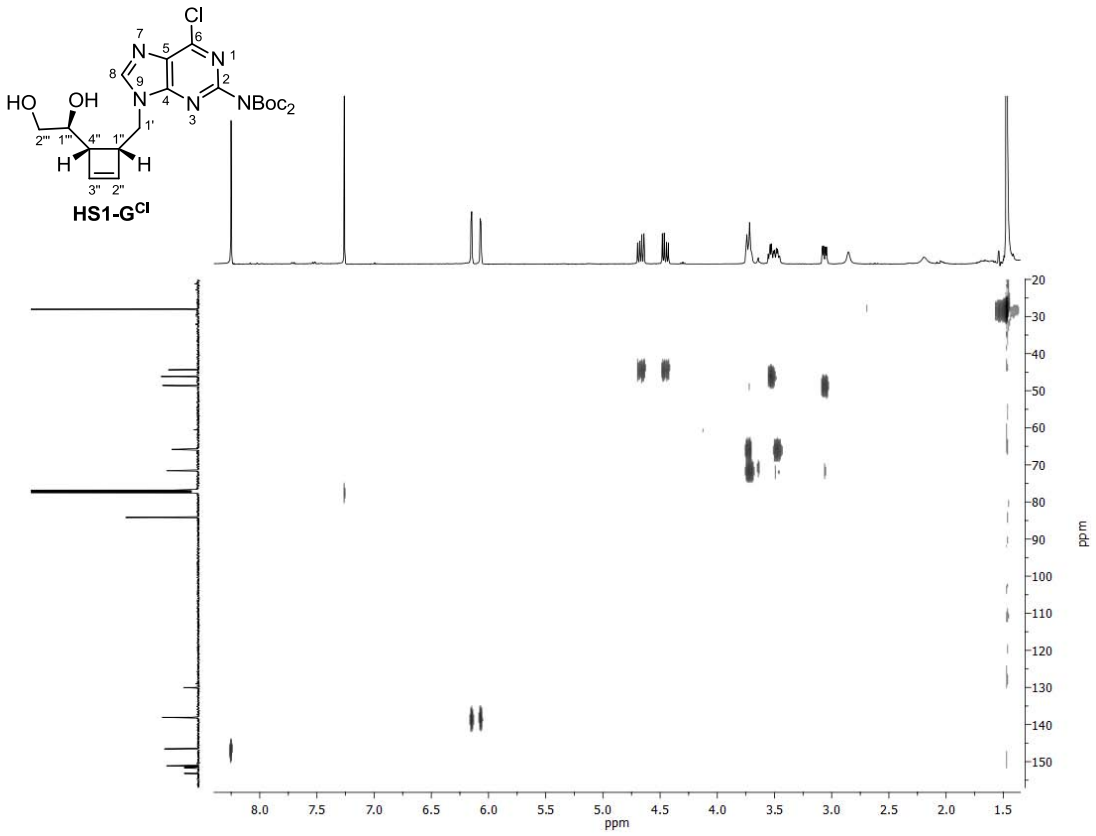
^{1}H-NMR (400 MHz, $CDCl_3$)



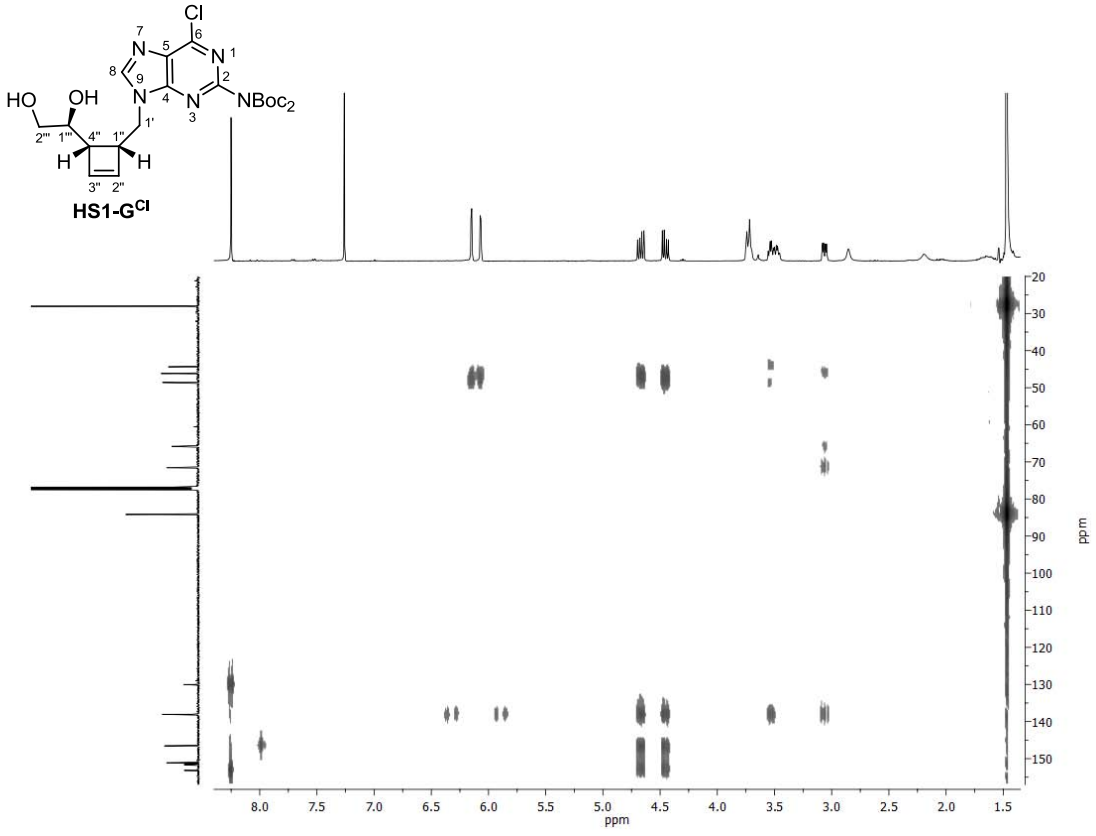
^{13}C-NMR (100 MHz, $CDCl_3$)



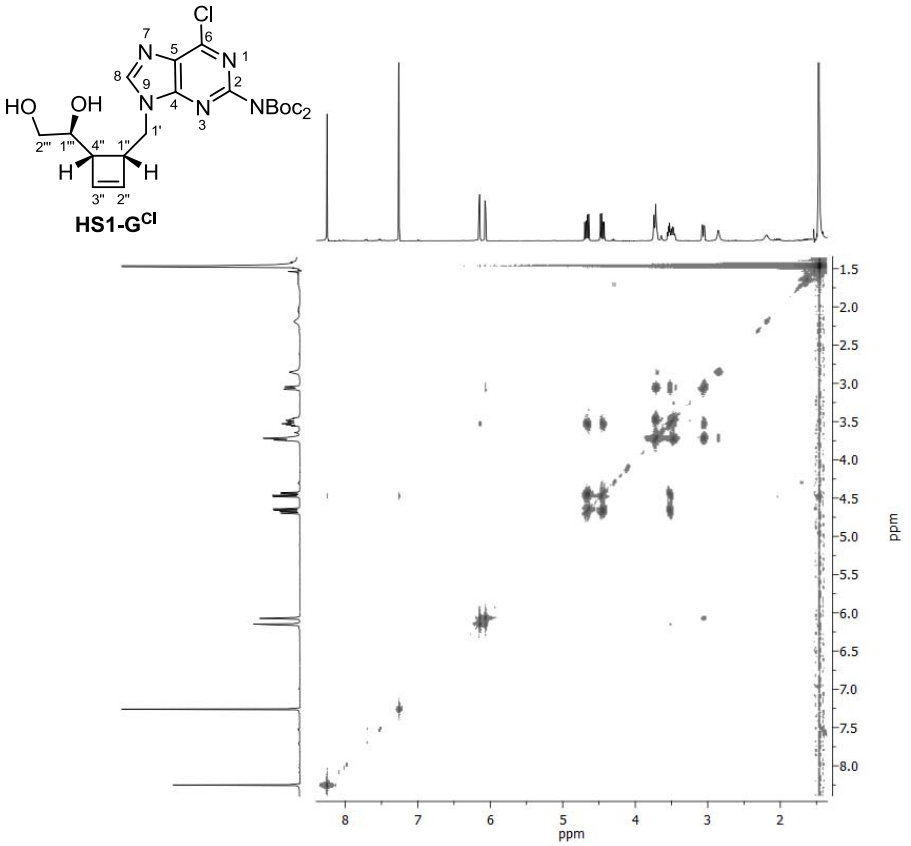
DEPT135 (400 MHz, $CDCl_3$)



HSQC (400 MHz, $CDCl_3$)



HMBC (400 MHz, $CDCl_3$)



COSY (400 MHz, $CDCl_3$)

APPENDIX B

Figures not included in the manuscript

1. *In silico* study of cyclobutane-fused nucleoside analogues as anti-HIV agents

1.1. 1st phosphorylation step

1.1.1. TK1 crystallographic structures

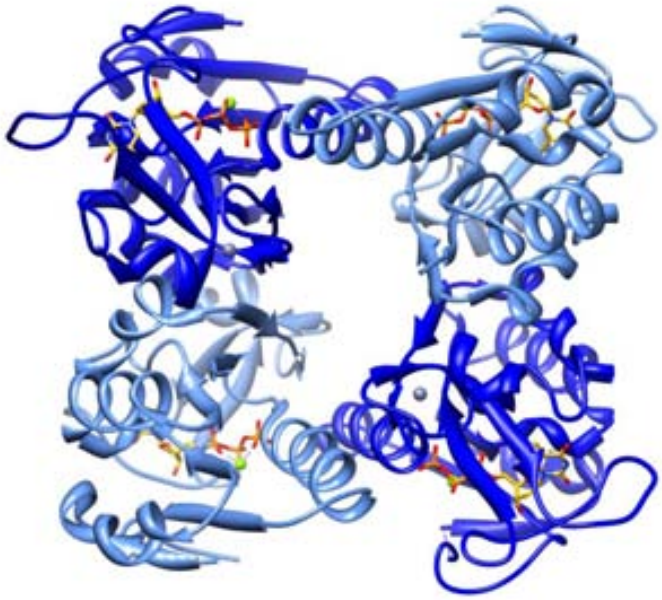


Figure B.1. TK1 crystallized with dTTP (yellow) in its binding site (PDB code: 1XBT). Mg^{2+} ions are shown in green and Zn^{2+} ions in purple.

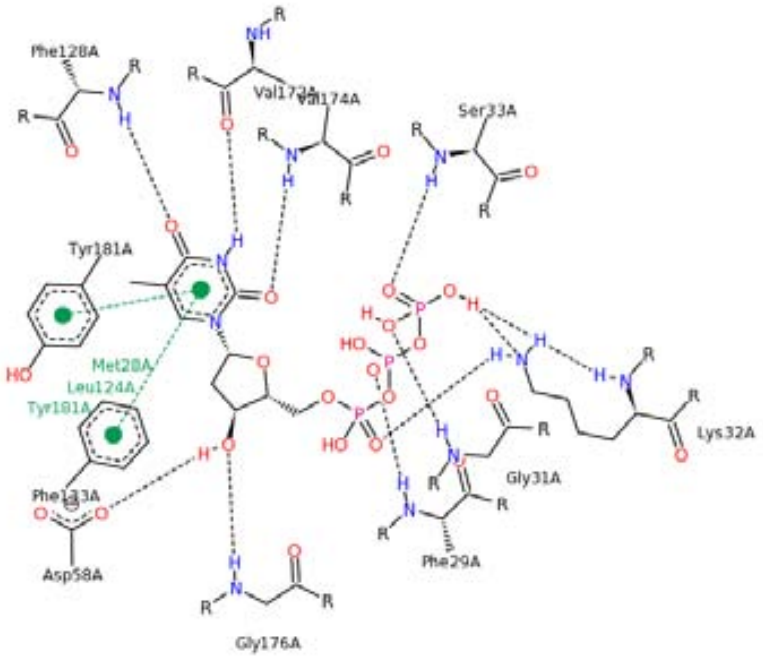


Figure B.2. Representation of the main interactions of dTTP in TK1 binding site (PDB code: 1W4R). Hydrogen bonds are depicted as dashed lines. Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids and the ligand. Image generated using PoseView software.

1.1.2. Docking results

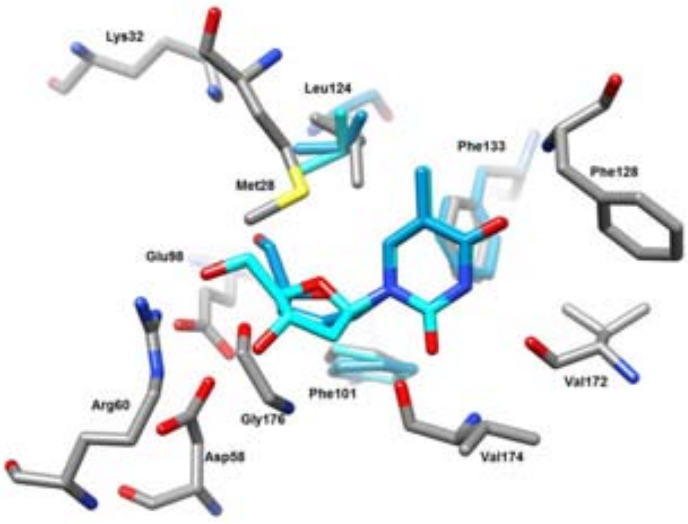


Figure B.3. Docked dT (cyan) superimposed to docked d4T (blue) in TK1 binding site (PDB code: 1XBT, X-ray residues shown in grey). Crystallographic waters, hydrogen atoms and hydrogen bonds are not shown for clarity.

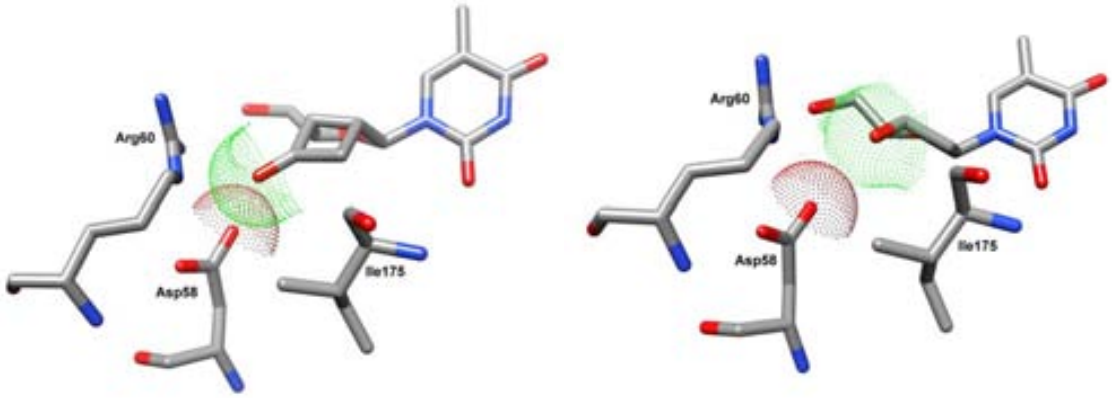


Figure B.4. Van der Waals radius representation of certain atoms to show the presence of steric clashes between **HI3.0-T** and certain residues in TK1 binding site (PDB code: 1XBT) (left), and the absence of those steric clashes between dT and the same residues in TK1 binding site (PDB code: 1XBT) (right). Crystallographic waters, hydrogen atoms and hydrogen bonds are not shown for clarity. (Reproduced by permission of reference 205 © 2012 JOHN WILEY AND SONS, Inc.)

1.2. 2nd phosphorylation step

1.2.1. TMK crystallographic structures

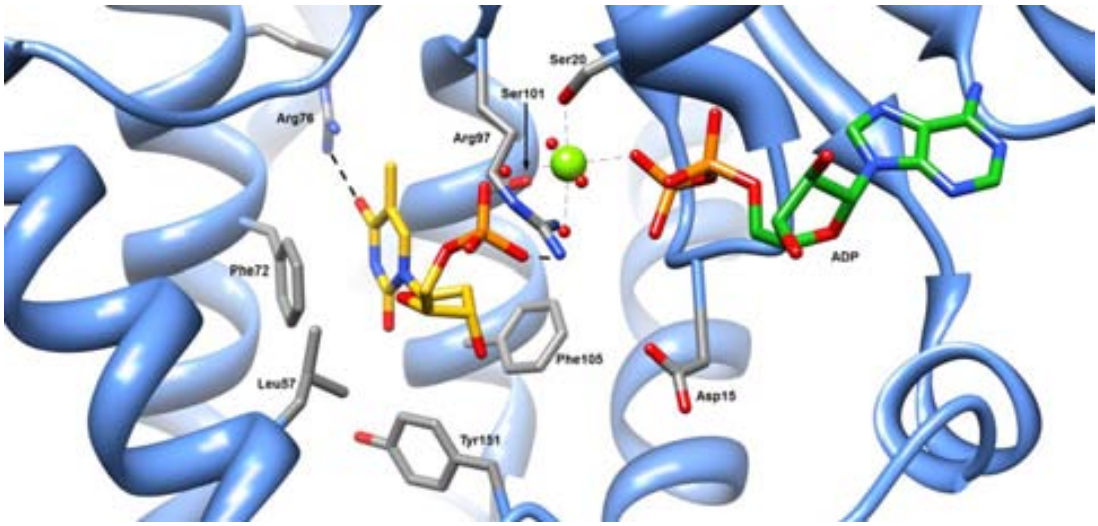


Figure B.5. Representation of the main interactions of dTMP in TMK binding site (PDB code: 1E2F). Hydrogen bonds between dTMP and residues are depicted as dotted lines. Mg^{2+} ion is shown in green. Crystallographic waters (except the ones coordinated to Mg^{2+}) and hydrogen atoms are not shown for clarity.

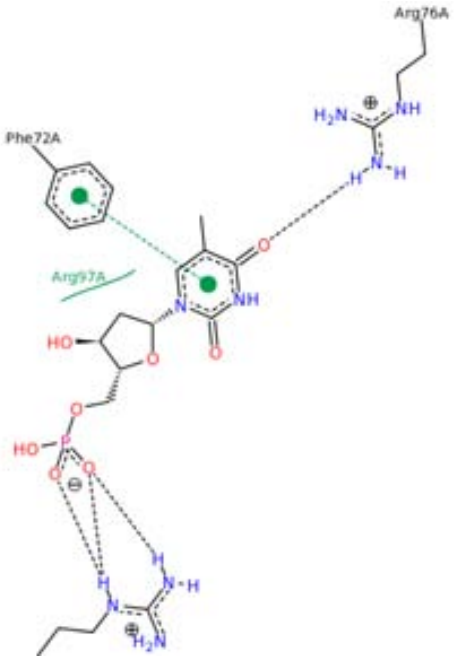


Figure B.6. Representation of the main interactions of dTMP in TMK binding site (PDB code: 1E2F). Hydrogen bonds are depicted as dashed lines. Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids and the ligand. Image generated using PoseView software.

1.2.2. Docking results

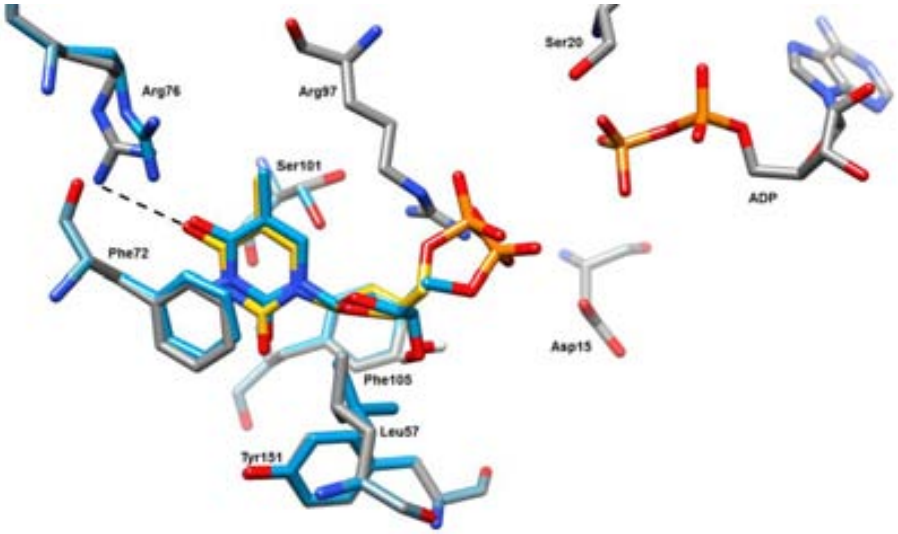


Figure B.7. Docked dTMP (blue) superimposed to crystallographic dTMP (yellow) in TMK binding site (PDB code: 1E2F, X-ray residues shown in grey). For the sake of clarity, crystallographic waters and ions are not shown, and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds between compounds and residues are depicted as dotted lines.

1.3. 3rd phosphorylation step

1.3.1. NDPK crystallographic structures

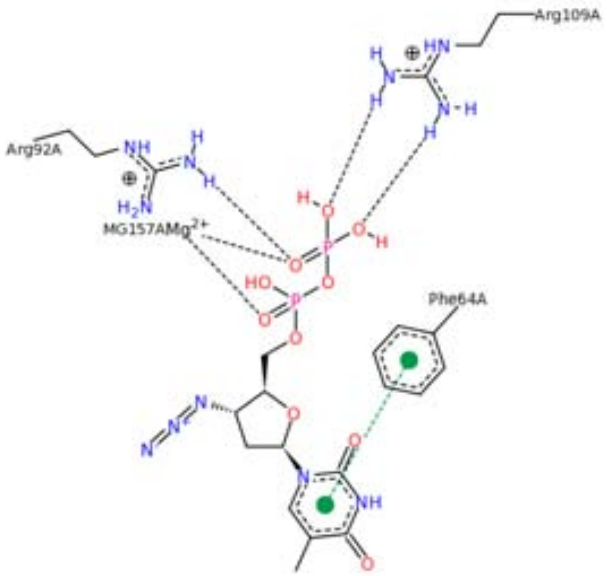


Figure B.8. Representation of the main interactions of AZTDP in *Dictyostelium discoideum* NDPK binding site (PDB code: 1LWX). Hydrogen bonds are depicted as dashed lines. Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids and the ligand. Image generated using PoseView software.

1.4. HIV-1 RT interaction

1.4.1. HIV-1 RT crystallographic structures

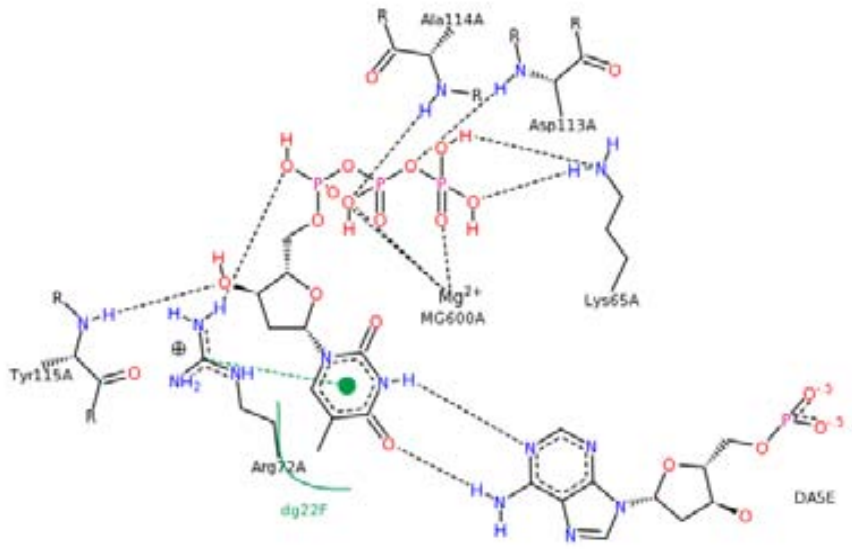


Figure B.9. Representation of the main interactions of dTTP in HIV-1 RT nucleotide site (PDB code: 1RTD). Hydrogen bonds are depicted as dashed lines. Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids and the ligand. Image generated using PoseView software.

1.4.2. Docking results

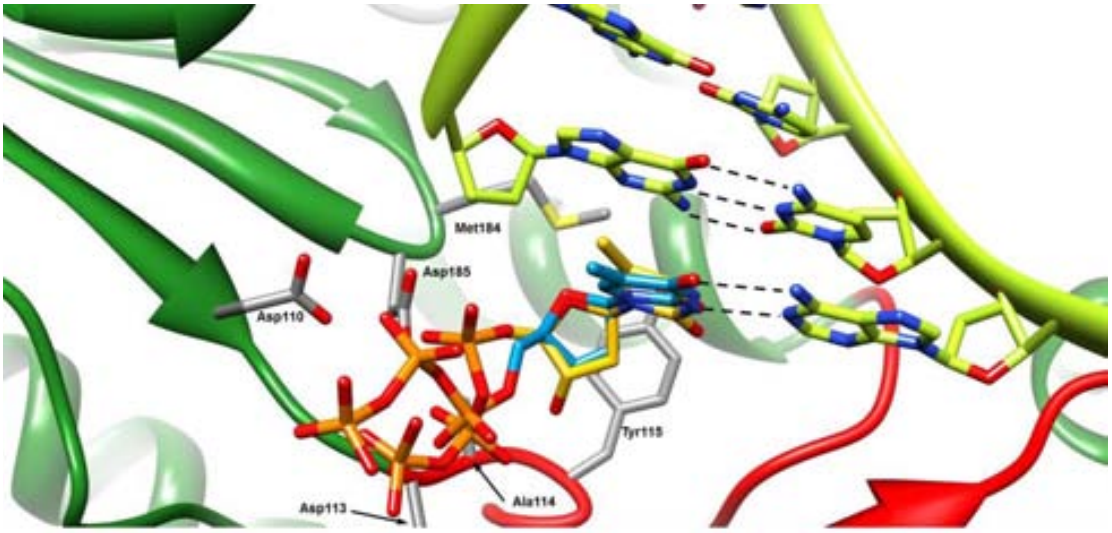


Figure B.10. Docked d4TTP (blue) superimposed to crystallographic dTTP (yellow) in HIV-1 RT nucleotide site (PDB code: 1RTD, X-ray residues shown in grey, DNA chains in green, backbone of the palm domain in forest green and backbone of the fingers domain in red). Hydrogen bonds are depicted as dotted lines. Crystallographic waters, ions, hydrogen atoms and flexible residues are not shown for clarity. (Reproduced by permission of reference 205 © 2012 JOHN WILEY AND SONS, Inc.)

2. *In silico* study of cyclobutene and cyclobutane L-nucleoside analogues as anti-HSV agents

2.1. 1st phosphorylation step

2.1.1. HSV-1 TK crystallographic structures

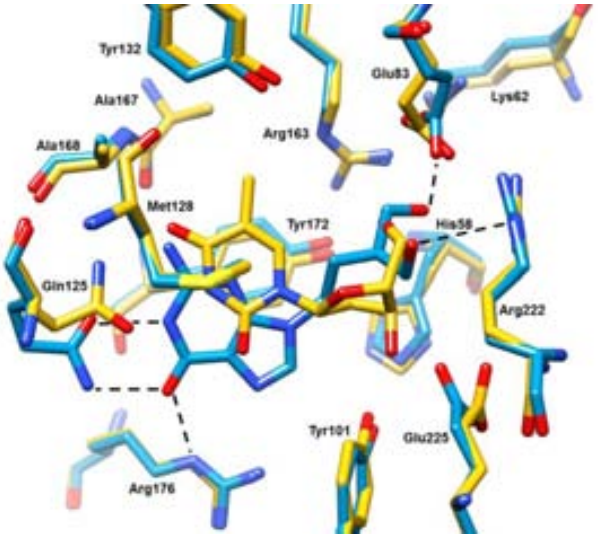


Figure B.11. Superimposed X-ray structures of dT (PDB code: 1KIM, in yellow) and PCV (PDB code: 1KI3, in blue) in HSV-1 TK binding site. For the sake of clarity, hydrogen atoms and crystallographic waters are not shown, and only hydrogen bonds between PCV and residues are depicted as dotted lines.

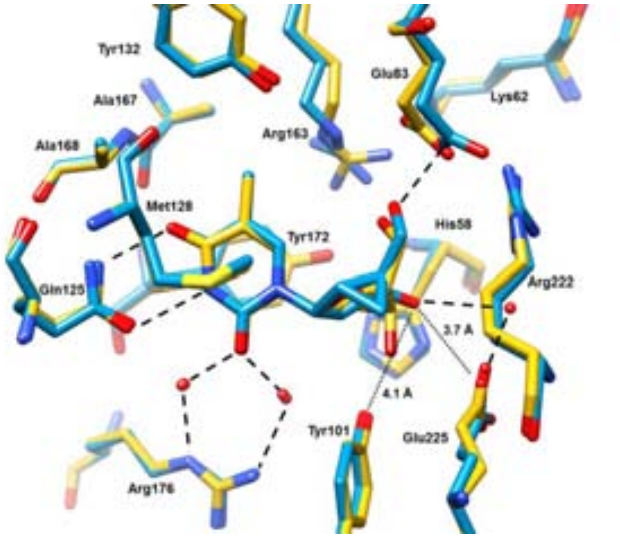


Figure B.12. Superimposed X-ray structures of dT (PDB code: 1KIM, in yellow) and MCT (PDB code: 1E2K, in blue) in HSV-1 TK binding site. For the sake of clarity, hydrogen atoms are not shown, and only hydrogen bonds between MCT and residues are depicted as dotted lines.

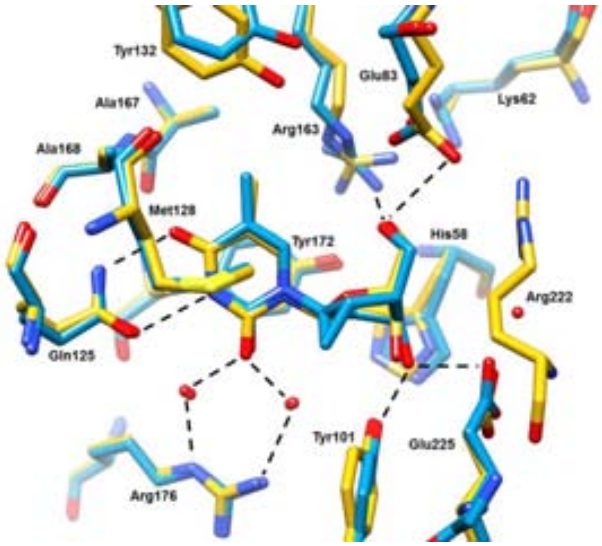


Figure B.13. Superimposed X-ray structures of dT (PDB code: 1KIM, in yellow) and (S)-MCT (PDB code: 1OF1, in blue) in HSV-1 TK binding site. Hydrogen bonds between nucleosides and residues are depicted as dotted lines. Hydrogen atoms are not shown for clarity.

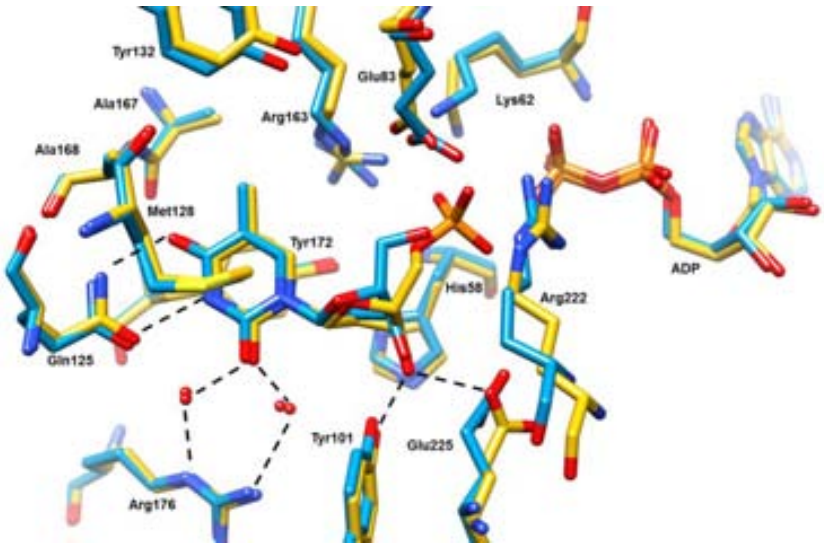


Figure B.14. Superimposed X-ray structures of dT (PDB code: 2VTK, in blue) and dTMP (PDB code: 1VTK, in yellow), both containing also ADP, in HSV-1 TK binding site. For the sake of clarity, hydrogen atoms are not shown and only hydrogen bonds between dTMP and residues are depicted as dotted lines.

2.1.2. Docking results

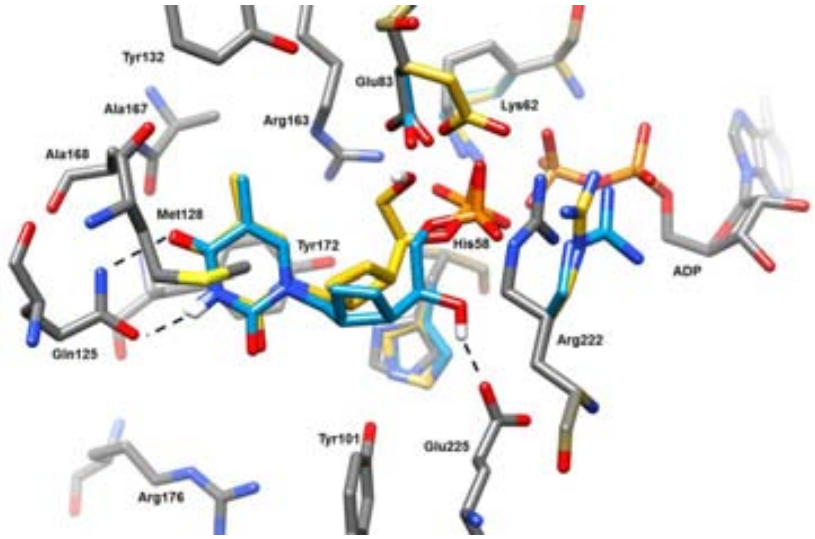


Figure B.15. HS1-T-1'''-MP (yellow) superimposed to **HS1-T-2'''-MP** (blue) in HSV-1 TK binding site (PDB code: 1VTK, X-ray residues shown in grey). Hydrogen bonds are depicted as dotted lines. For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are only shown when bound to a heteroatom of the ligand.

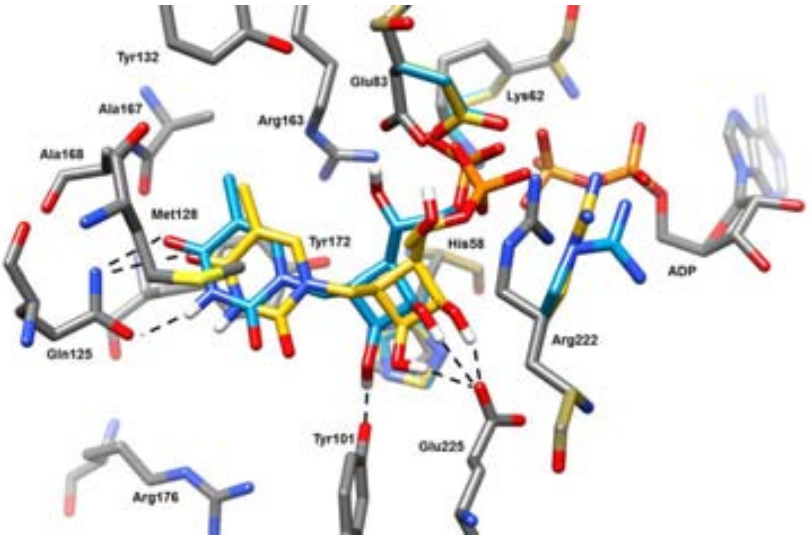


Figure B.16. HS2-T-1'''-MP (yellow) superimposed to **HS2-T-2'''-MP** (blue) in HSV-1 TK binding site (PDB code: 1VTK, X-ray residues shown in grey). Hydrogen bonds are depicted as dotted lines. For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are only shown when bound to a heteroatom of the ligand.

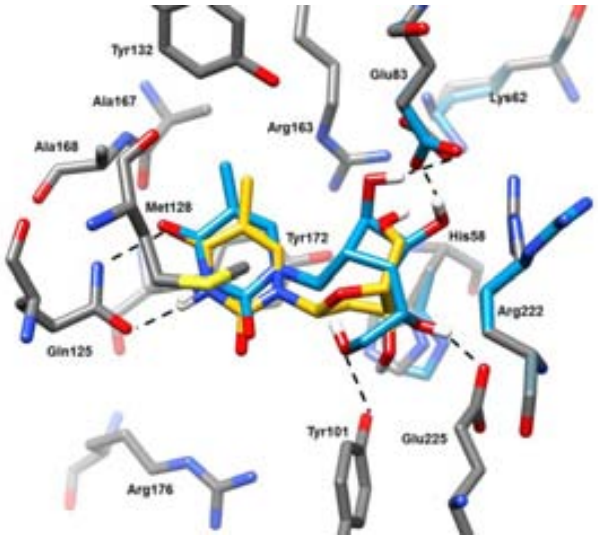


Figure B.17. HS2'-T (blue) superimposed to crystallographic dT (yellow) in HSV-1 TK binding site (PDB code: 1KIM, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown, hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand and only hydrogen bonds between **HS2'-T** and residues are depicted as dotted lines.

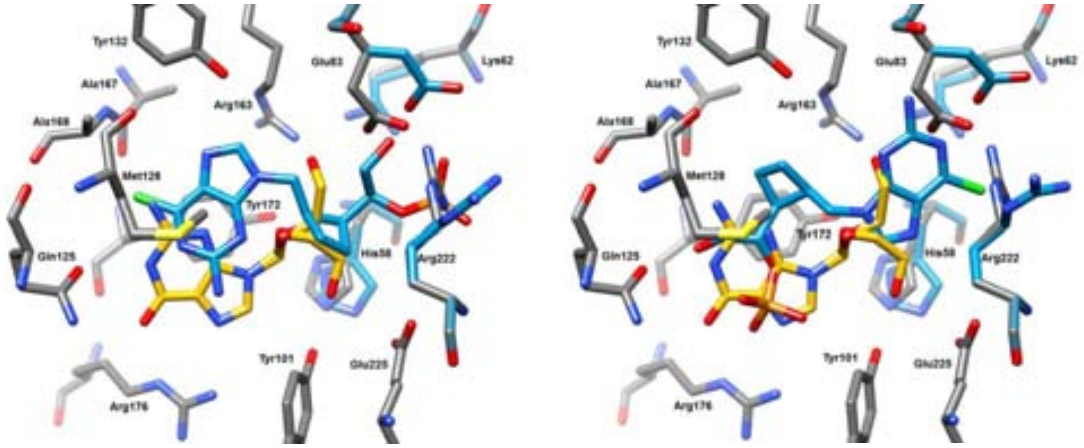


Figure B.18. Two equally favoured orientations of **HS1'-G^{CL}-1'''-MP** (blue) superimposed to crystallographic ACV (yellow) in HSV-1 TK binding site (PDB code: 2KI5, X-ray residues shown in grey). Crystallographic waters, hydrogen atoms and hydrogen bonds are not shown for clarity.

2.2. 2nd phosphorylation step

2.2.1. GMPK crystallographic structure

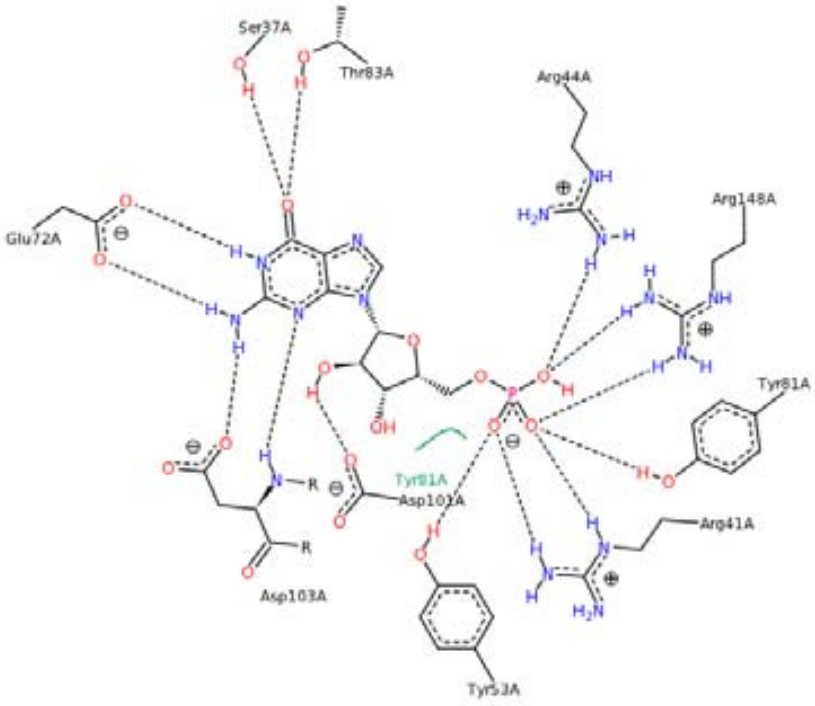


Figure B.19. Representation of the main interactions of GMP in mGMPK binding site (PDB code: 1LVG). Hydrogen bonds are depicted as dashed lines. Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids and the ligand. Image generated using PoseView software.

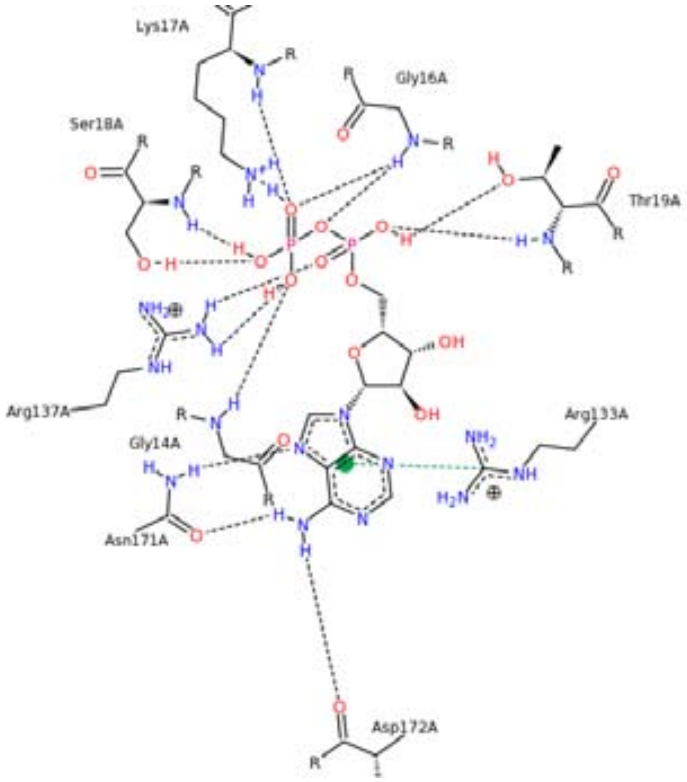


Figure B.20. Representation of the main interactions of ADP in mGMPK binding site (PDB code: 1LVG). Hydrogen bonds are depicted as dashed lines. Image generated using PoseView software.

2.2.1. Docking results

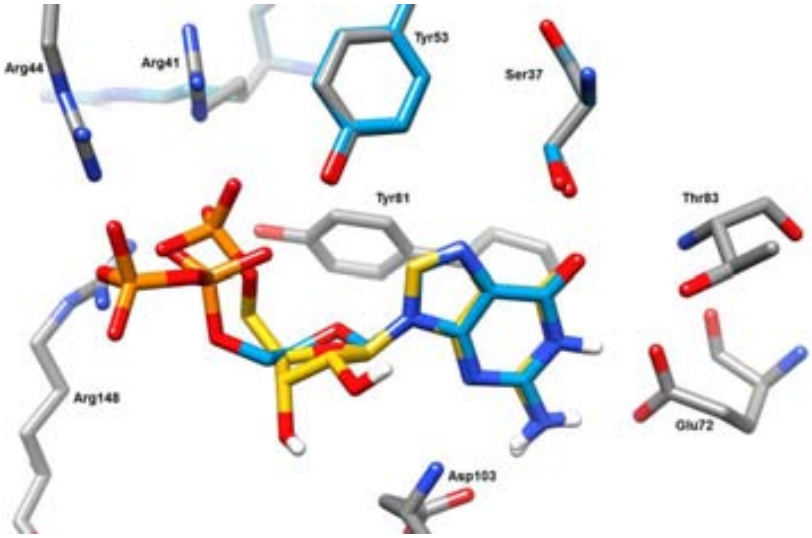


Figure B.21. Docked ACVDP (blue) superimposed to crystallographic GMP (yellow) in mGMPK binding site (PDB code: 1LVG, X-ray residues shown in grey). Crystallographic waters and hydrogen bonds are not shown for clarity. Hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand.

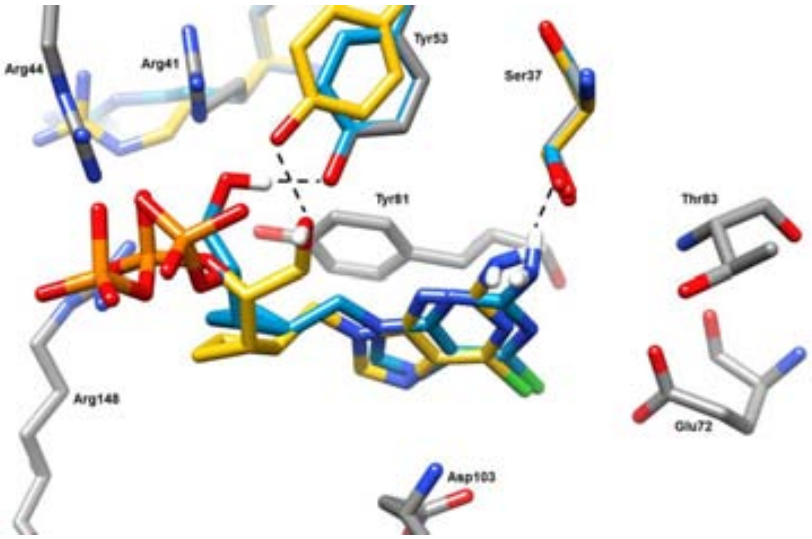


Figure B.22. HS1'-G^{CL}-1'''-MP (blue) superimposed to **HS1'-G^{CL}-1'''-DP** (yellow) in mGMPK binding site (PDB code: 1LVG, X-ray residues shown in grey). For the sake of clarity, crystallographic waters are not shown and hydrogen atoms are exclusively shown when bound to a heteroatom of the ligand. Hydrogen bonds are depicted as dotted lines.

2.3. 3rd phosphorylation step

2.3.1. NDPK crystallographic structures

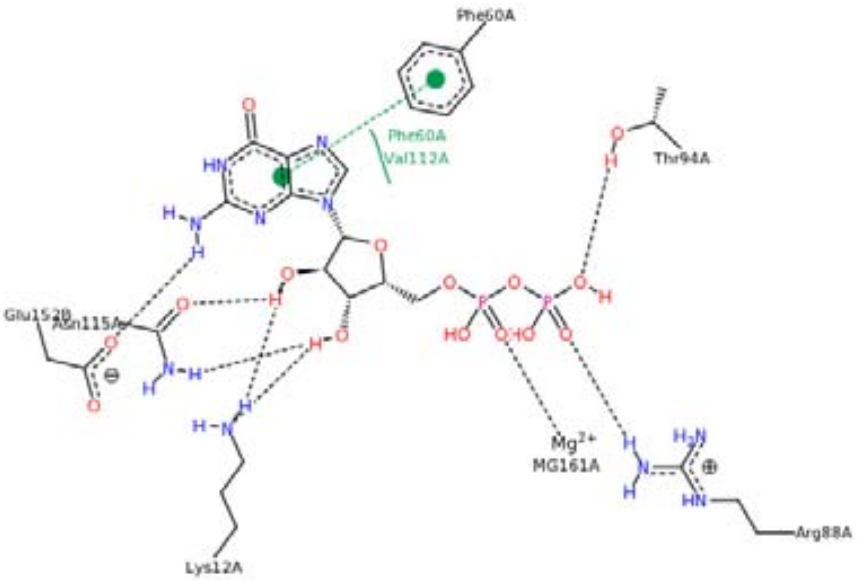


Figure B.23. Representation of the main interactions of GDP in NDPK binding site (PDB code: 1NUE). Hydrogen bonds are depicted as dashed lines. Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids and the ligand. Image generated using PoseView software.

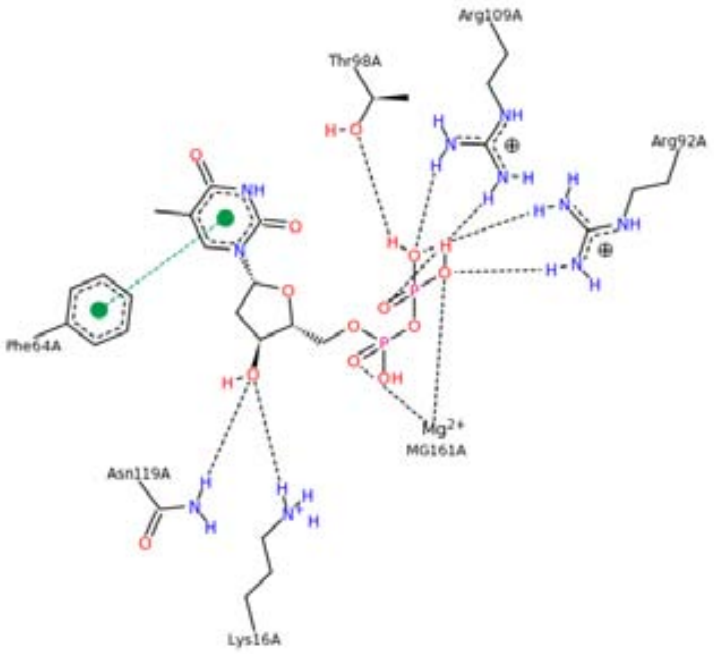


Figure B.24. Representation of the main interactions of dTDP in NDPK binding site (PDB code: 1NDC). Hydrogen bonds are depicted as dashed lines. Hydrophobic interactions are illustrated as smooth contour lines between the respective amino acids and the ligand. Image generated using PoseView software.