KU Leuven
Groep Biomedische Wetenschappen
Faculteit Geneeskunde
Departemen van Microbiologie, Immunologie en
Transplantatie
Rega Institute for Medical Research
Laboratorium van Virologie en Chemotherapie

KU LEUVEN

**DOCTORAL SCHOOL
BIOMEDICAL SCIENCES**

STRUCTUUR-GEBASEERDE ONTWIKKELING EN SYNTHESE VAN NUCLEOSIDE-TRIFOSFAAT ANALOGEN ALS EEN NIEUWE KLASSE VIRALE POLYMERASE INHIBITOREN

Weijie Gu

Promotor: Prof. Dr. Kalyan Das

Copromotor: Prof. Dr. Piet Herdewijn
　　　　　　Dr. Steven De Jonghe

Proefschrift voorgedragen
tot het behalen van de
graad van Doctor in de
biomedische wetenschap

September 2021

KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Microbiology, Immunology and
Transplantation
Rega Institute for Medical Research
Laboratory of Virology and Chemotherapy

KU LEUVEN

**DOCTORAL SCHOOL
BIOMEDICAL SCIENCES**

STRUCTURE-BASED DESIGN AND SYNTHESIS OF NUCLEOSIDE-TRIPHOSPHATE ANALOGS AS A NEW CLASS OF VIRAL POLYMERASE INHIBITORS

Weijie Gu

<u>Jury</u>:

Supervisor: Prof. Dr. Kalyan Das
Co-supervisor: Prof. Dr. Piet Herdewijn
                Dr. Steven De Jonghe
Chair: Prof. Dr. Ghislain Opdenakker
Jury members: Prof. Dr. Eddy Arnold
                Prof. Dr. Christophe Pannecouque
                Prof. Dr. Matheus Froeyen
                Dr. Tongfei Wu

Dissertation presented in
partial fulfilment of the
requirements for the
degree of Doctor in
Biomedical Science

September 2021

Acknowledgements

Weije Gu

Leuven, 2021

Summary

Nucleoside analogs play a crucial role in antiviral chemotherapy. Currently, approximately 30 nucleos(t)ide analogs received marketing approval for the treatment of various viral infections. In order to exert their therapeutic effects, nucleoside analogs must be phosphorylated stepwise via multiple cellular kinases into their biologically active nucleoside triphosphate (NTP) analogs.

However, due to high substrate specificity of kinases, the conversion process is usually inefficient, particularly at the first step which is often rate-limiting. To bypass this inefficient first phosphorylation step, nucleoside monophosphate analogs (NMPs) were introduced. Unfortunately, NMPs suffer from *in vivo* instability due to dephosphorylation by cellular enzymes. In addition, they are too polar to penetrate the cellular membrane. Thus, prodrug strategies, such as the ProTide approach, are utilized to improve cellular permeability and oral bioavailability of NMPs. As isosteres of NMPs, nucleoside phosphonates are also able to skip the first phosphorylation step, and their phosphonate moiety has the advantage of being chemically and metabolically stable.

Although several prodrugs of NMPs and nucleoside phosphonates delivered successful antiviral drugs, which can bypass the first phosphorylation step, they still need to undergo two subsequent phosphorylations in order to become pharmaceutically active. The two remaining steps also are bottlenecks for a number of nucleos(t)ide analogs. In addition, the currently used antiviral drugs face problems such as toxicity and drug resistance in the long run. Therefore, the design and synthesis of NTP analogs that can circumvent the kinase activation pathway and behave as direct viral polymerase substrates or inhibitors, and new antiviral agents with few adverse effects and high genetic barrier to resistance, are of great significance.

This doctoral thesis mainly aims at the design and synthesis of novel nucleoside analogs acting as NTP analogs/mimics for HIV-1 reverse transcriptase (RT). The HIV-1 RT inhibitory activity of these compounds was evaluated by *in vitro* enzymatic assays and the anti-HIV activity was investigated in cell-based assays. The binding mode of these compounds to HIV-1 RT was investigated by X-ray crystallography.

In the first part of this doctoral thesis (**Chapter 2**), the synthesis, biological evaluation and X-ray structural analysis of a series of amino acid conjugates of dAMP and tenofovir (TFV) is described. It was demonstrated that a number of these conjugates act as dNTP analogs, with HIV-1 RT being able to catalytically incorporate the TFV part in a primer, thereby acting as a chain terminator. X-ray structures in complex with HIV-1 RT/dsDNA showed binding of the conjugates in different modes at the polymerase active site in the presence of Mg^{2+} vs. Mn^{2+} ions. The adaptability of the compounds is seemingly essential for catalytic incorporation of TFV by RT. Among these newly synthesized

compounds, L-Glu-TFV with a carboxyl sidechain was most efficiently incorporated. L-Met-TFV showed weak incorporation and rather behaved as a dNTP-competitive RT inhibitor. This result advocates the feasibility of designing NTP/dNTP analogs by chemical substitutions to nucleotide analogs.

Because of the chemical instability and high polarity of the amino acid conjugates of TFV, a second series of TFV analogs in which the phosphonoamidate moiety was replaced by a stable and neutral amide linker was prepared in **Chapter 3**. To increase membrane permeability for cell-based assay, several ester prodrugs were also prepared. These compounds were evaluated in a biochemical HIV-1 RT inhibition assay, an anti-HIV assay, and four X-ray structures of three indivial analogs were obtained in complex with RT/dsDNA.

In the third part of this doctoral thesis (**Chapter 4**), a new series of acyclic nucleoside phosphonate analogs as NTP mimics was prepared by introduction of metal chelating moieties including phosphonate and carboxylic acid groups. The impacts of chain length, chirality and linker atoms were also taken into consideration. Several challenging synthetic procedures were developed for syntheses of these compounds. HIV-1 RT inhibition assays were performed to evaluate their inhibitory effect on DNA synthesis catalyzed by RT. X-ray structures of this series of compounds in complex with HIV-1 RT/dsDNA demonstrated that their binding modes are distinct from that of compounds in Chapter 2 and 3. The detailed structural understanding of these new compounds provides opportunities for the design of new HIV-1 RT inhibitors.

From an X-ray crystallographic fragment screening in our lab, two fragments (called fragments 048 and 166) were discovered that bind to a transient P pocket of HIV-1 RT. In the last part of this thesis (**Chapter 5**), the design by a docking study and synthesis of new compounds based on the X-ray structure of fragment 166 in complex with HIV-1 RT/dsDNA is described. Five new analogs were synthesized and evaluated in an HIV-1 RT inhibition assay and an anti-HIV assay. In addition, the synthesis of two selected compounds based on docking results was attempted.

Samenvatting

Nucleoside-analogen spelen een cruciale rol in de behandeling van virale infecties. Tot op heden werd een 30-tal nucleos(t)ide-analogen gecommercialiseerd als antiviraal geneesmiddel. Nucleoside-analogen moeten stapsgewijs gefosforyleerd worden door verschillende cellulaire kinases om uiteindelijk de biologisch actieve nucleosidetrifosfaat-analogen te genereren. De beperkte substraatspecificiteit van deze kinases heeft echter als gevolg dat dit proces inefficiënt verloopt. In vele gevallen is de omzetting van het nucleoside naar het nucleosidemonofosfaat (d.w.z. de eerste fosforylering) de snelheidsbepalende stap. Om dit te vermijden zouden in principe nucleosidemonofosfaat-analogen kunnen worden aangewend. Vermits ze onderhevig zijn aan enzymatische defosforylatie, zijn deze nucleosidemonofosfaat-analogen metabool labiel. Bovendien zijn ze zeer polair, wat hun geringe cellulaire permeabiliteit verklaart. De ontwikkeling van prodrugs heeft toegelaten om de cellulaire permeabiliteit en de orale biologische beschikbaarheid van nucleoside monofosfaat-analogen te verbeteren. De meest succesvolle prodrugs zijn de ProTides. Daarnaast werden ook nucleosidefosfonaten, isosteren van nucleosidemonofosfaten, gesynthetiseerd, die ook in staat om de eerste fosforylatie stap te omzeilen.

Verschillende prodrugs van nucleosidemonofosfaat- en nucleosidefosfonaat-analogen werden reeds op de markt gebracht als succesvolle antivirale geneesmiddelen. Deze verbindingen moeten echter nog steeds tweemaal gefosforyleerd worden om de biologisch actieve metaboliet te genereren. Voor een aantal nucleoside-analogen zijn deze bijkomende fosforyleringen problematisch. Daarnaast zijn de nucleoside-analogen die momenteel gebruikt worden in antivirale chemotherapie op lange termijn toxisch en leiden ze tot de selectie van resistente virussen. Daarom is de zoektocht naar nucleosidetrifosfaat-analogen die geen kinase-gemedieerde activering nodig hebben, en dus functioneren als directe polymerase inhibitoren, van cruciaal belang. Bovendien is er steeds nood aan nieuwe antivirale geneesmiddelen met minder neveneffecten en een verbeterd resistentieprofiel.

In deze doctoraatsthesis wordt de ontwikkeling en synthese van nieuwe nucleosidetrifosfaat-analogen beschreven. Deze verbindingen werden geëvalueerd als potentiële inhibitoren van HIV-1 reverse transcriptase (RT) in biochemische experimenten en de anti-HIV activiteit werd nagegaan in cellulaire testen. Daarenboven werd de manier waarop deze verbindingen binden aan HIV-1 RT bestudeerd via X-stralen kristallografie.

In het eerste deel van dit doctoraatswerk (**Hoofdstuk 2**) wordt de synthese, biologische evaluatie en structurele analyse van een serie aminozuurconjugaten van dAMP en tenofovir (TFV) beschreven. Er werd aangetoond dat een aantal van deze conjugaten functioneren als dNTP-analogen, waarbij HIV-1 RT in staat is om het TFV-gedeelte katalytisch te incorporeren aan een primer. Deze verbindingen fungeren aldus als 'chain terminators'. De kristalstructuren van deze verbindingen in

complex met HIV-1 RT/dsDNA toonden aan dat de interactie van deze conjugaten met de actieve plaats van het virale polymerase verschillend is naargelang de aanwezigheid van Mg^{2+} vs. Mn^{2+} ionen. L-Glu-TFV (met een carbonzure zijketen) werd het meest efficiënt geïncorporeerd. L-Met-TFV, daarentegen, werd niet ingebouwd in een groeiende DNA keten, maar functioneerde als een dNTP-competitieve RT-inhibitor. Deze resultaten tonen aan dat het mogelijk is om nieuwe NTP/dNTP-analogen te ontwikkelen door structurele modificatie van nucleotide-analogen.

Omwille van de chemische instabiliteit en de polariteit van het fosfonoamidaat gedeelte in de aminozuurconjugaten van TFV, werd een tweede serie TFV-analogen gesynthetiseerd waarbij de fosfonoamidaat groep werd vervangen door een stabiele en neutrale amidelinker. Dit onderzoek is beschreven in **Hoofdstuk 3**. Hierbij werden ook verschillende esterprodrugs bereid om de cellulaire permeabiliteit van de carbonzuren te verbeteren. De activiteit van deze verbindingen werd nagegaan in zowel een biochemische HIV-1 RT als in een cellulaire anti-HIV test. Daarenboven werden de kristallografische structuren van vier verschillende analogen, in complex met RT/dsDNA, opgehelderd.

In het derde deel van deze doctoraatsthesis (**Hoofdstuk 4**) wordt de synthese van een nieuwe serie acyclische nucleosidefosfonaten beschreven via de invoering van metaal-complexerende functionele groepen zoals fosfonaten en carbonzuren. Hierbij werd ook de ketenlengte, de stereochemie van de chirale centra en het type linkeratomen gewijzigd. Verschillende uitdagende syntheseprocedures werden hierbij op punt gesteld. Voor alle verbindingen werd nagegaan of ze de DNA- synthese, gekatalyseerd door HIV-1 RT, kunnen inhiberen. Kristallografische analyse van deze producten, in complex met HIV-1 RT/dsDNA, toonde aan dat ze op een andere manier binden aan HIV-1 RT, dan de nucleoside-analogen vermeld in Hoofdstukken 2 en 3. Een gedetailleerde analyse van de structuren opent mogelijkheden voor de ontwikkeling en synthese van nieuwe HIV-1 RT inhibitoren.

Een X-stralen kristallografische fragment screening leidde tot de ontdekking van fragmenten 048 en 166 die beiden binden aan een transiënte P pocket van HIV-1 RT. In het laatste deel van deze thesis (**Hoofdstuk 5**) wordt een moleculaire docking studie uitgevoerd voor de ontwikkeling van nieuwe HIV RT inhibitoren, gebaseerd op de structuur van fragment 166 in complex met HIV-1 RT/dsDNA. Vijf nieuwe analogen van fragment 166 werden gesynthetiseerd en geëvalueerd in zowel een biochemische HIV-1 RT test, als in een cellulair anti-HIV experiment. Daarenboven werd uitgebreid gewerkt aan de, tot op heden niet succesvolle, synthese van twee bijkomende analogen.

Abbreviations

3D	Three dimensional
3TC	Lamivudine ((-)-L-2′,3′-Dideoxy-3′-thiacytidine)
A	Adenine
ABC	Abacavir
α-CNPs	Alpha-carboxynucleoside phosphonates
AIDS	Acquired immune deficiency syndrome
ANPs	Acyclic nucleoside phosphates
AZT	3'-Azido-3'-deoxythymidine
Bn	Benzyl
Boc	*t*-Butoxycarbonyl
C	Cytosine
CbzCl	Carbobenzoxy chloride
CC_{50}	Half maximal (50 %) cytotoxic concentration
$CDCl_3$	Deuterated chloroform
CMV	Cytomegalovirus
CNPs	Cyclic nucleoside phosphonates
d4T	Stavudine (2',3'-didehydro-2',3'-dideoxythymidine)
d4U	2',3'-didehydro-2',3'-dideoxyuridine
dAMP	2'-Deoxyadenosine monophosphate
DCC	*N*, *N*'-dicyclohexylcarbodiimide
ddU	2',3'-dideoxyuridine
DENV-2	Dengue virus type 2
DIAD	Diisopropyl azodicarboxylate
DIEA	*N*, *N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	*N*, *N*-dimethylformamide
DMSO	Dimethyl sulfoxide
DCM	Dichloromethane
D_2O	Deuterium oxide (heavy water)
DNA	Deoxyribonucleic acid
DPPA	Diphenylphosphoryl azide
dsDNA	Double stranded DNA
EA	Ethyl acetate
EBOV	Ebola virus
EC_{50}	Half maximal (50 %) effective concentration
EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EFV	Efavirenz
FDA	Food and Drug Administration
FTC	Emtricitabine ((-)-2',3'-Dideoxy-5-fluoro-3'-thiacytidine)
G	Guanine
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HOBt	*N*-hydroxybenzotriazole
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSV	Herpes simplex virus
IC_{50}	Half maximal (50 %) inhibitory concentration
IDA	Iminodiacetic acid
IDP	Iminodipropionic acid

INDOPY	Indolopyridone
MHz	Megahertz
mM	Millimolar
MS	Mass spectrometry
n-BuLi	*n*-Butyl lithium
NcRTI	Nucleotide competing reverse transcriptase inhibitor
NDP	Nucleoside diphosphate
NMP	Nucleoside monophosphate
NMR	Nuclear magnetic resonance
NNRTI	Nonnucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
NRTTI	Nucleoside reverse transcriptase translocation inhibitor
NtRTI	Nucleotide reverse transcriptase inhibitor
(d)NTP	(2'-Deoxy) nucleoside triphosphate
NS5B	Non-structural 5B
ppm	Parts per million
PMBCl	4-Methoxybenzyl chloride
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic acid
RP-HPLC	Reversed-phase high-performance liquid chromatography
RT	Reverse transcriptase
RTIs	Reverse transcriptase inhibitors
r.t.	Room temperature
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
T	Thymine
T-*α*-CNP	Thymine alpha-carboxynucleoside phosphonate
TAF	Tenofovir alafenamide
TBAF	Tetrabutylammonium fluoride
TBDPSCl	*tert*-Butyl(chloro)diphenylsilane
TBSCl	*tert*-Butyl(chloro)dimethylsilane
TDF	Tenofovir disoproxil fumarate
TEA	Triethylamine
TEAB	Triethylammonium bicarbonate
TEMPO	2,2,6,6-Tetramethylpiperidinyloxyl
TFA	Trifluoroacetic acid
TFV	Tenofovir
THF	Tetrahydrofuran
TLC	Thin-layer chromatography
TMSBr	Trimethylsilyl bromide
U	Uracil
VZV	Varicella zoster virus
YFV	Yellow fever virus

Table of amino acids and their abbreviations

Amino acid	Three-letter code	One-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table of Contents

Chapter 1. General introduction

1.1 General structure of nucleoside analogs

Natural nucleosides are fundamental biomolecules, which play essential roles in numerous cellular processes such as DNA/RNA synthesis, cell signalling, enzyme regulation and metabolism.[1] Nucleotides are composed of a sugar moiety (*β*-D-ribofuranose or *β*-D-2'-deoxyribofuranose) attached to a purine (adenine and guanine) or pyrimidine (cytosine, uracil and thymine) base via a glycosidic bond, and at least one phosphate group (**Figure 1-1**).[2] The purine bases and the pyrimidine base cytosine are common for DNA and RNA, but mostly the pyrimidine base thymine (T) is in DNA, while uracil (U) is in RNA. Only occasionally does T occur in RNA or U in DNA. Nucleosides are classified as N-nucleosides or C-nucleosides, respectively, based on the type of bond between the sugar moiety and the nucleobase.
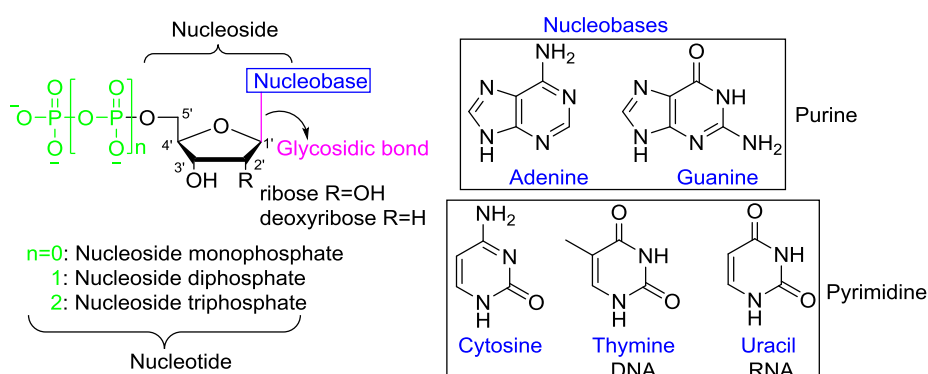
Figure 1-1. General structure of natural nucleos(t)ides and nucleobases.

Nucleoside analogs are synthetically modified natural nucleosides. They are used to disrupt cellular division or viral replication by impairing DNA/RNA synthesis or by inhibiting essential enzymes involved in nucleos(t)ide metabolism.[3] Structural modifications in the sugar moiety, the nucleobase and/or the glycosidic bond led to the discovery of novel nucleoside analogs with potent biological activity. To date, there are more than 30 nucleos(t)ide analogs approved on the market for the treatment of viral and fungal infections and cancers.[3-5]

1.2 Nucleoside analogs as antivirals

Due to their impressive potency to inhibit the replication of various viruses, such as the human immunodeficiency virus (HIV), herpesviruses, hepatitis B virus (HBV), hepatitis C virus (HCV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), nucleoside analogs play a significant role in antiviral chemotherapy.[3, 6-10] Since idoxuridine was reported by William H. Prusoff in 1959[11] and approved as the first antiviral nucleoside analog,[12] a wide variety of novel antiviral nucleos(t)ide analogs have been discovered over the past several decades. Currently, more than 20 nucleoside analogs are being used for treating viral infections, as shown in **Figure 1-2**.
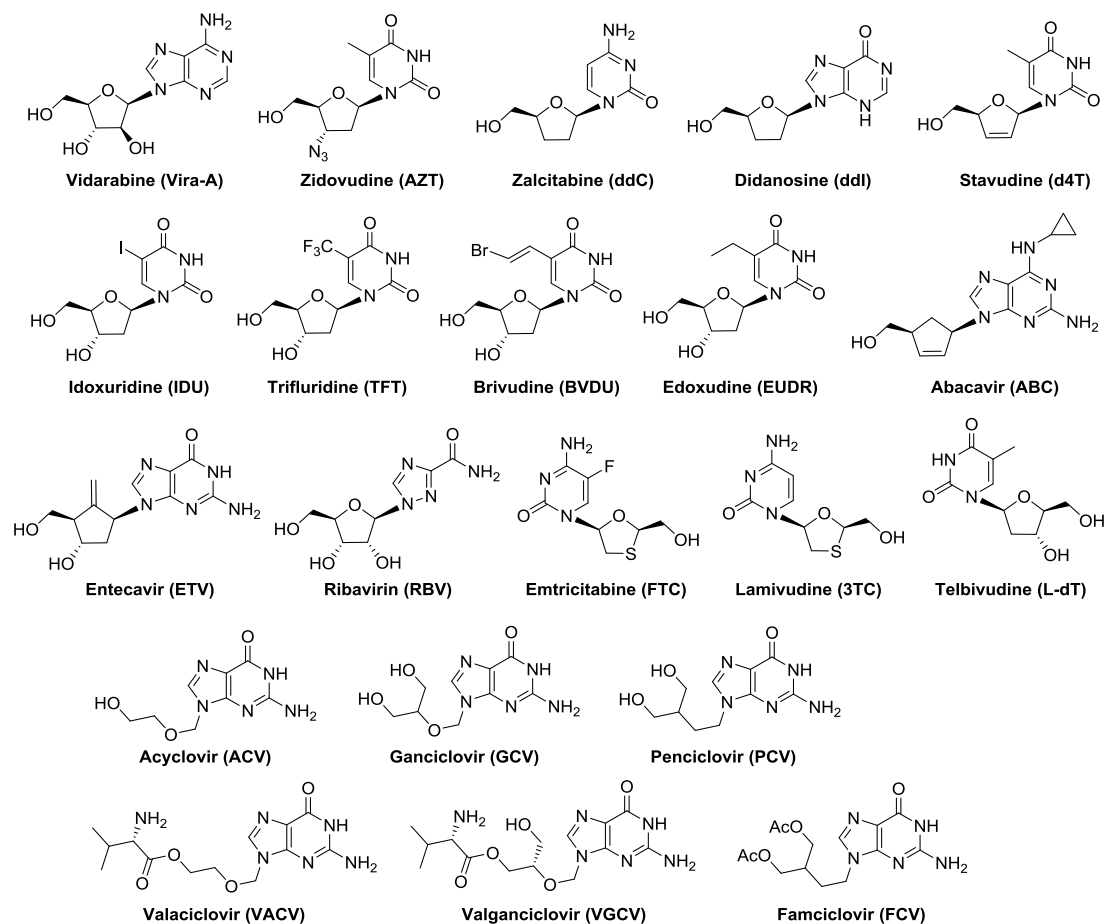
Figure 1-2. Nucleoside analogs approved as antiviral drugs.

Novel nucleoside analogs were obtained by structural modifications of the sugar or nucleobase moiety, (or a combination thereof), thereby allowing for a large structural diversity of nucleosides.

Modifications of the sugar moiety

Vidarabine (Ara-A), where the stereochemistry of the hydroxyl group at the C2′ position is inverted "up", compared to the "down" position of natural nucleosides, was first approved by the FDA in 1976 and has been used against various DNA viruses such as Herpes Simplex Virus 1 and 2 (HSV-1 and HSV-2), as well as Varicella-Zoster Virus (VZV).[13-14]

The 2',3'-dideoxynucleoside analogs such as zidovudine (AZT), zalcitabine (ddC), didanosine (ddI), stavudine (d4T), lamivudine (3TC) and emtricitabine (FTC) are incorporated in the viral genome. However, due to absence of the 3'-hydroxyl group, no further incorporation occurs, and extension of the growing DNA chain is halted. As a result, these nucleoside analogs act as chain terminators, with proven efficacy as therapeutic agents against HIV.

Replacement of the furanose ring by a cyclopentyl ring generates carbocyclic nucleoside analogs, which often display enhanced biostability because of their resistance to hydrolysis by phosphorylases.[15] Representative examples are abacavir (ABC) and entecavir (ETV), which both have been approved by the FDA for the treatment of HIV and HBV infections, respectively. ABC is

a chain terminator featuring an unsaturated sugar ring, whereas ETV is not a chain terminator. Interestingly, ETV has a unique exocyclic double bond that causes unfavorable steric hinderance, leading to delayed chain termination, despite the presence of 3′-OH group.[16]

The 2′ and 3′-carbons of the sugar ring were also removed to generate acyclic nucleoside analogs. A famous example of this class of compounds is acyclovir (ACV), an acyclic guanosine analog which has shown potent activity and specificity towards HSV-1 and HSV-2.[17-18] Ganciclovir (GCV) is also an acyclic guanosine analog, but it still keeps the 3′-hydroxyl group. Penciclovir (PCV), used for treating HSV-1 and HSV-2 infections, is an acyclic guanosine analog with a methylene group rather than an oxygen in the acyclic sugar moiety. ACV, GCV and PCV display low oral bioavailability, so a prodrug strategy has been applied, resulting in the development of valaciclovir, valganciclovir and famciclovir.[19]

The sugar moiety of natural nucleosides is D-ribofuranose or D-2'-deoxyribofuranose, so natural nucleosides are all D-nucleosides. Since the discovery of 3TC, much attention has been paid to the synthesis and biological evaluation of L-nucleoside analogs, which are enantiomers of the natural D-nucleosides. FTC and telbivudine (L-dT) are other examples of approved antiviral L-nucleoside analogs. Both 3TC and FTC have an additional sulfur atom inserted in their sugar rings. FTC also contains a modified nucleobase. Compared to the natural D-nucleosides, L-nucleoside analogs have comparable or greater antiviral activity, lower toxicity and improved metabolic stability.[20] 3TC is approved for the treatment of both HIV and HBV infections, while FTC and L-dT are used against HIV and HBV infections, respectively.[21-22]

Modifications of the nucleobase

Idoxuridine, trifluridine, brivudine and edoxudine are 2′-deoxyuridine analogs, with a substituent at the C5 position of the uracil base. This modification changes the steric and electronic environment, lipophilicity and even the hydrogen bonding/hydrophobic interactions between the enzyme binding site and the nucleoside analog[11, 23-25] Among the four compounds, idoxuridine is applied topically to treat herpes simplex keratitis, trifluridine is primarily used for ocular herpes infections, whereas brivudine is indicated for the treatment of VZV infections. Edoxudine shows effectiveness against HSV.[11, 25-26] Ribavirin is a synthetic guanosine nucleoside analog possessing a 1,2,4-triazole-3-carboxamide moiety as a modified nucleobase. It has broad-spectrum antiviral activity against various RNA viruses.[27]

1.3 Nucleoside monophosphate analogs and nucleoside phosphonates as antivirals

1.3.1 Nucleoside monophosphate analogs

In order to exert their therapeutic effects, nucleoside analogs must be phosphorylated stepwise via multiple cellular kinases into the biologically active nucleoside triphosphates (NTPs). However, due to their structural differences from natural nucleosides, the conversion process for an analog is often inefficient, particularly for the first phosphorylation step which is very often rate-limiting.[28-29] For instance, the first phosphorylation step of the anti-HIV drug d4T to d4T 5'-monophosphate, catalyzed by thymidine kinase, is the rate-limiting step in human cells.[30] An inefficient phosphorylation results in a diminished or even complete lack of biological activity of a nucleoside analog. To bypass this inefficient first phosphorylation step, nucleoside monophosphate analogs (NMPs) were introduced. Unfortunately, NMPs suffer from *in vivo* instability, due to dephosphorylation by cellular enzymes, particularly phosphatases. In addition, they are too polar to penetrate the cellular membrane. However, the phosphate moiety offers the possibility to design prodrugs, allowing to improve cellular permeability and oral bioavailability of NMPs. Various prodrug strategies mask one or two hydroxyl groups of the monophosphate moiety increasing lipophilicity and improving cellular uptake.[31-35] The ProTide strategy, pioneered by the late Chris McGuigan,[36] is the most successful example of an NMP prodrug approach.[34, 37] Sofosbuvir[38] and remdesivir[39] are both examples of ProTide prodrugs of NMPs (**Figure 1-3**).

Sofosbuvir   **Remdesivir (GS-5734)**
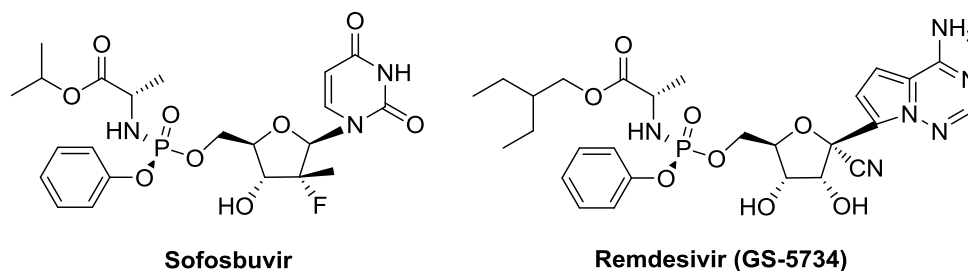
Figure 1-3. FDA-approved antiviral ProTides of NMPs.

Sofosbuvir that contains a 2′-fluoro C-methyl sugar moiety, has been approved by the FDA to treat HCV infections. It is a phosphoramidate prodrug, that after intracellular metabolism, releases the NMP metabolite, which is subsequently converted into its di- and triphosphate analog. The NTP is the biologically active species that functions as a substrate of the viral NS5B polymerase.

Remdesivir is currently the only C-nucleoside-based ProTide approved by the FDA. Structurally, it is a phenyl 2-ethylbutyl L-alanine ester phosphoramidate prodrug of 1′-cyano-4-aza-7,9-dideaza adenosine. It was reported that the 1′-cyano group is vital for its selectivity for the viral polymerase over human polymerases, since this group occupies a pocket that is only available in the NTP-binding site of viral polymerase.[39-40] Although initially developed by Gilead Sciences for the treatment of HCV infections, it displayed broad spectrum antiviral activity against various RNA viruses such as Ebola virus (EBOV), yellow fever virus (YFV), dengue-2 virus (DENV-2), influenza A virus (IAV), parainfluenza 3 (PIV), SARS-CoV and most recently SARS-CoV-2.[39, 41-42] Compared

with the anti-EBOV activity of its parent C-nucleoside analog ($EC_{50} = 0.78\,\mu M$), remdesivir displayed more pronounced inhibition of the EBOV replication ($EC_{50} = 0.06\,\mu M$).[40] This was attributed to higher intracellular concentrations of remdesivir triphosphate. In addition, as the only FDA-approved nucleoside analog for treatment of COVID-19, it is a delayed translocation inhibitor of SARS-CoV-2 replication. Structural studies revealed that remdesivir monophosphate blocked RNA translocation after incorporation of three additional nucleotides, leading to delayed chain termination.[43-44]

1.3.2 Nucleoside phosphonates

Based on the structure of sugar moiety, nucleoside phosphonates can be divided into two types, namely, acyclic nucleoside phosphonates (ANPs) and cyclic nucleoside phosphonates (CNPs). Irrespective of acyclic or cyclic, all nucleoside phosphonates have a common phosphonate (P-C-O) moiety, which unlike the monophosphate (P-O-C) moiety of the natural nucleotides, is resistant to hydrolysis by cellular enzymes, in particular phosphodiesterases and phosphatases.[45-46]

Acyclic nucleoside phosphonates

In addition to the chemical and enzymatic stability of the phosphonate moiety, ANPs have a flexible acyclic chain, which permits the molecules to adopt to the correct conformation at the active site of target enzyme, such as DNA polymerase and reverse transcriptase (RT). ANPs are effective against various DNA viruses and retroviruses and have become a key class of antivirals (**Figure 1-4**).
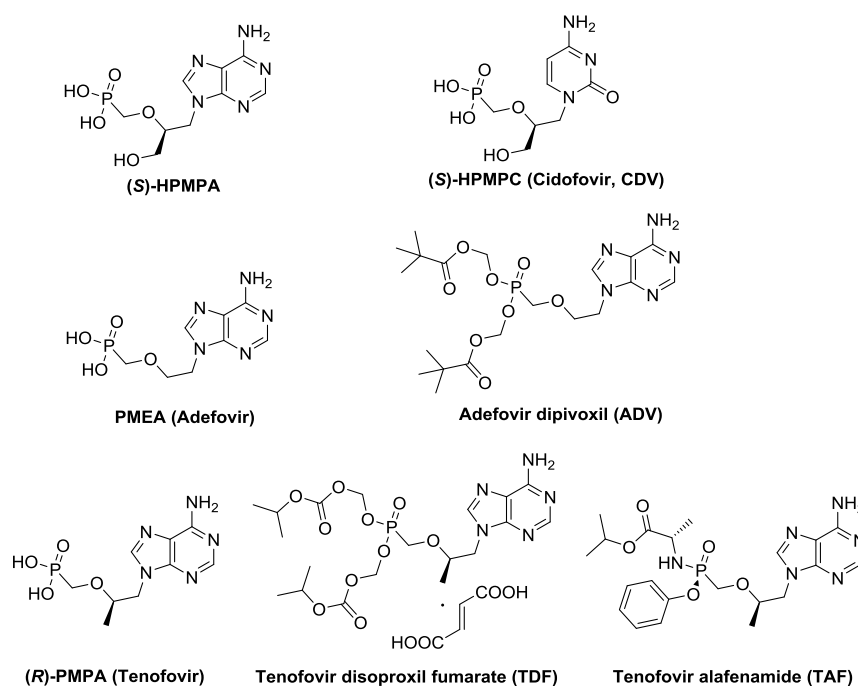
Figure 1-4. Chemical structures of therapeutic ANPs and/or their prodrugs.

Based on the structure of acyclic moiety, three subclasses of ANPs are distinguished. The first ANP, (*S*)-9-(3-hydroxy-2-phosphonylmethoxypropyl)adenine (also known as (*S*)-HPMPA), was

originally reported in 1986 and showed broad spectrum antiviral activity against various DNA viruses.[47] (*S*)-HPMPA is the prototype of the first subclass of ANPs, which are known as the 3-hydroxy-2-phosphonylmethoxypropyl (HPMP) derivatives. The stereochemistry of the carbon atom bearing the hydroxymethylene side chain is important as the (*S*)-enantiomers are the most potent. (*S*)-HPMPC (cidofovir), the cytosine counterpart of (*S*)-HPMPA, was equally active as (S)-HPMPA, but less toxic.[48] Its potent activity against cytomegalovirus (CMV) made cidofovir to be the first ANP to be licensed for clinical use in 1996 for the treatment of CMV retinitis in acquired immune deficiency syndrome (AIDS) patients.

The second subclass of ANPs features a 2-phosphonylmethoxyethyl (PME) moiety connected with the nucleobase. Unlike the first subclass, these ANPs do not possess a chiral carbon. The best studied example of this class is adefovir (9-(2-phosphonylmethoxyethyl)adenine, PMEA), which was initially developed as an anti-HIV drug, but suffered from nephrotoxicity at the required high dosage. However, at a much lower dosage, adefovir can be used for the treatment of chronic HBV infections. Because of its low oral biovailability, adefovir was marketed as its ester prodrug adefovir dipivoxil, which received marketing approval in 2002 as treatment for HBV infected patients.

The third subclass of ANPs contains a 2-phosphonylmethoxypropyl (PMP) moiety linked to the nucleobase, with the (*R*)-enantiomer being the most active. The most widely used PMP derivative is (*R*)-9-(2-(phosphonylmethoxypropyl)adenine (also known as (*R*)-PMPA, tenofovir), which displays potent activity against HIV and HBV. Unfortunately, like most ANPs, tenofovir (TFV) is endowed with a very poor bioavailability. Therefore, tenofovir disoproxil was prepared as a prodrug and formulated as the fumarate salt (tenofovir disoproxil fumarate, TDF), which was approved in 2001 and 2008, for the treatment of HIV and HBV infections, respectively. Tenofovir alafenamide (TAF) is a ProTide prodrug of tenofovir, that received FDA approval in 2015 for the treatment of HIV and HBV infections. Compared with TDF, a lower dosage of TAF can be used in patients. In addition, the emergence of viral drug resistance against TAF is seldomly observed, yielding TAF a safe and effective antiviral drug.

Cyclic nucleoside phosphonates

Unlike widespread ANPs, a limited number of CNPs have shown antiviral activity. Some examples are illustrated in **Figure 1-5**. The lack of antiviral activity of CNPs is most likely due to their poor substrate properties for cellular kinases. In order to improve their substrate properties, new compounds were designed by introducing the phosphono-alkoxy group at the 3'-position of the sugar moiety. For example, the L-2'-deoxythreosyl phosphonate nucleosides PMDTT and PMDTA were reported as anti-HIV and anti-HBV agents.[49] The synthesis of the amidate prodrugs of PMDTT and PMDTA led to a boost in antiviral potentcy.[50] Another CNP was 2',3'-dideoxy-2',3'-

didehydrothymidine phosphonate (d4TP), which is the most active phosphonomethoxy analog among a series of cyclic pyrimidine nucleotides.[51] Further research on purine analogs resulted in the discovery of phosphonomethoxy-2'-fluoro-2',3'-dideoxy-2',3'-didehydroadenosine (Fd4AP, GS-9148) that exhibited low mitochondrial toxicity and cytotoxicity, and retained activity against different drug resistant variants.[52] GS-9193, the ProTide prodrug of GS-9148, has shown a higher activity at lower doses, rendering this compound a promising HIV drug candidate.[53]
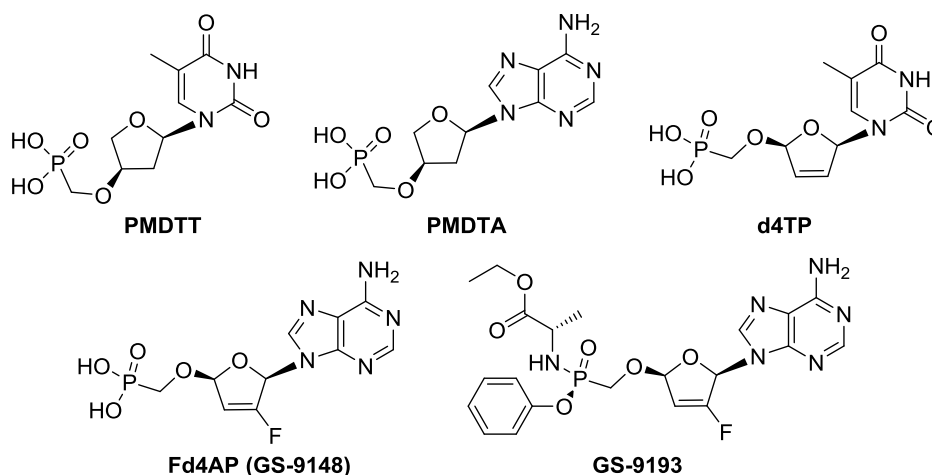
Figure 1-5. Structures of reported CNPs and/or their prodrugs.

1.4 Mechanism of action of nucleoside analogs and nucleoside phosphonates

Nucleoside analogs follow an identical metabolic pathway as the natural nucleosides (**Figure 1-6**). With the aid of specific nucleoside transporters, nucleoside analogs get into cells, which is a dynamic process involving equilibrative nucleoside transporters and concentrative nucleoside transporters.[54-55] Once inside cells, nucleoside analogs need three consecutive phosphorylation steps by different cellular kinases to become biologically active. The first phosphorylation step is catalyzed by nucleoside kinases, which result in the formation of NMPs. The second phosphorylation step is conducted by NMP kinases, generating nucleoside diphosphates (NDPs). The last phosphorylation is carried out by NDP kinases, leading to the production of NTPs that are the pharmacologically active forms of therapeutic nucleoside analogs. Triphosphorylated nucleosides are competitive inhibitors or substrates of viral polymerases, which catalytically incorporate the nucleotide part as RNA or DNA chain terminator.[3]
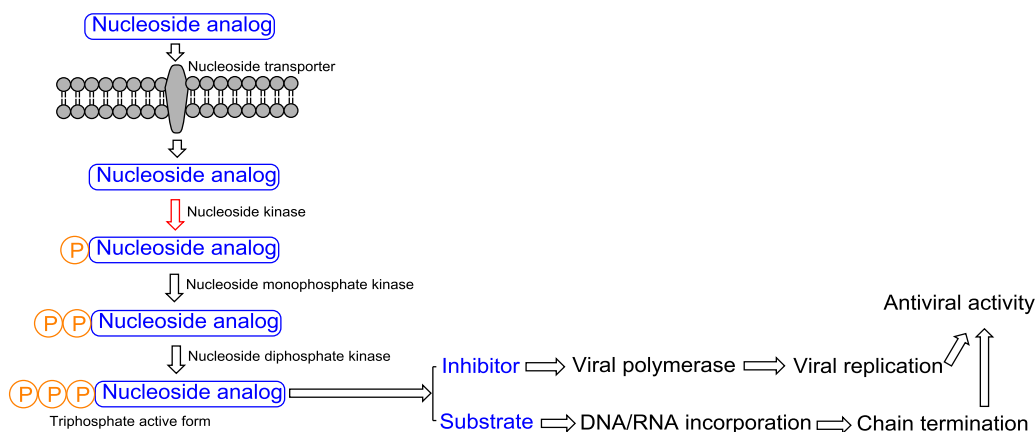
Figure 1-6. Mechanism of action of nucleoside analogs.

The phosphonate moiety of nucleoside phosphonates is isosteric to the phosphate group of NMP, with the advantage of being chemically and metabolically stable. Therefore, nucleoside phosphonates can skip the first phosphorylation step and directly undergo the next two steps. Specifically, after cellular uptake, nucleoside phosphonates are converted into their corresponding monophosphates and diphosphates, respectively, catalyzed by cellular kinases.[56-57] The nucleoside phosphonate diphosphates are the equivalent of NTPs, and hence, are the biologically active species, which can inhibit viral DNA polymerases or be incorporated into the growing DNA chain.[58-59]

1.5 Nucleoside triphosphate mimics/analogs as antivirals

Although prodrugs of NMP and nucleoside phosphonate analogs have been explored as successful antiviral drugs, which can bypass the first, and often rate-limiting, phosphorylation step, they still need to undergo two subsequent phosphorylations in order to become active. However, the two remaining steps may also be a bottleneck for a number of nucleoside analogs. For example, it has been reported that the second phosphorylation of AZT is rate-limiting.[60-61] Even the third phosphorylation of 2',3'-didehydro-2',3'-dideoxyuridine (d4U) and 2',3'-dideoxyuridine (ddU) is a hurdle, because their diphosphate analogs are poor substrates of NDP kinase.[62] Therefore, the design and synthesis of NTP analogs or mimics which can skip the kinase activation pathway and behave as direct viral polymerase substrates or inhibitors, are of great interest and significance.

1.5.1 Generally chemical synthesis of nucleoside triphosphate analogs

In principle, NTPs can be synthesized both enzymatically and chemically. However, due to the high substrate specificity, the enzymatic methods are mainly applicable to natural NTPs only, while chemical methods are usually the preferred option for the synthesis of unnatural NTPs. Numerous chemical approaches have been developed for the synthesis of NTP analogs starting from nucleoside analogs.[63-66] Among these approaches, Ludwig's "one-pot, three-step" approach[67] has become one of the most commonly used methods due to its wide applicability to various nucleoside analogs and its simplicity (protection-free for nucleosides, one-pot synthesis and short reaction time). This method

involves an unprotected nucleoside analog reacting with phosphorus oxychloride (POCl₃), with the resulting nucleoside phosphorodichloridate subsequently reacting with bis-(tri-*n*-butylammonium) pyrophosphate, and final hydrolysis of the resulting cyclic intermediate. POCl₃ and trialkylphosphate function as phosphorylating agent and solvent, respectively, which are based on Yoshikawa's procedure.[68]

This method has been applied to a large number of unprotected nucleosides and their analogs. For example, the triphosphate form of remdesivir[40] has been synthesized by such a method, as shown in **Scheme 1-1**. Reaction of the parent nucleoside of remdesivir with POCl₃ using trimethyl phosphate as the solvent, is followed by the treatment with tributylammonium pyrophosphate using tributylamine as a base and dry acetonitrile as a solvent. Hydrolysis of cyclic intermediate with triethylammonium bicarbonate (TEAB) buffer afforded remdesivir-triphosphate. The reaction scheme involves the formation of remdesivir dichlorophosphoridate, followed by the generation of cyclic intermediate and subsequent hydrolysis to give the final product. Similarly, the ribofuranosyl triphosphate form of favipiravir (T-705)[69] was also prepared by this approach, as illustrated in **Scheme 1-2**. T-705-ribonucleoside is treated with POCl₃ and N-methylimidazole in trimethylphosphate yielding a phosphorodichloridate intermediate. Then, reaction with bis(tri-*n*-butylammonium) pyrophosphate in dry DMF results in the cyclic entity, which is hydrolyzed with water to furnish the triphosphate form of ribofuranosyl T-705.
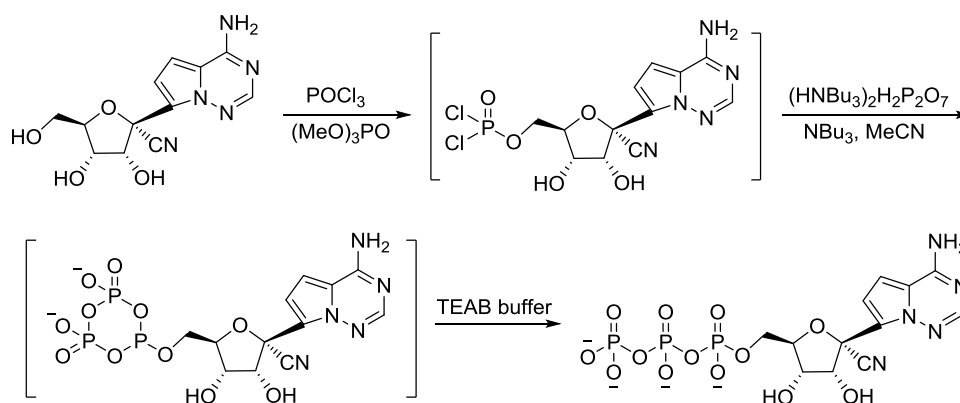
Scheme 1-1. Synthesis of triphosphate form of remdesivir using "one-pot, three-step" approach.
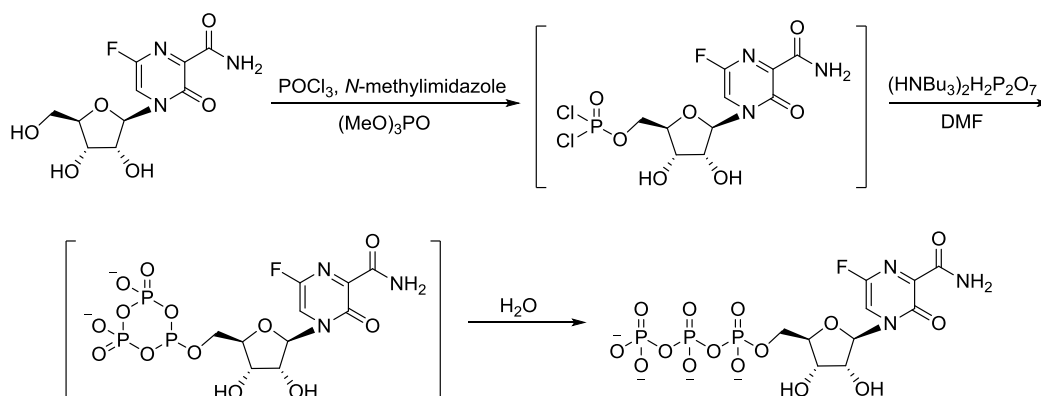
Scheme 1-2. Synthesis of the triphosphate form of ribofuranosyl T-705 using "one-pot, three-step" approach.

The Ludwig-Eckstein method[70] has also witnessed widespread use. In this one-pot triphosphorylation procedure, a 2'-3'-protected nucleoside analog is treated with salicylic chlorophosphite. The resulting activated phosphite is reacted with pyrophosphate to generate a cyclic P(III) intermediate. This intermediate is then oxidized to a cyclic P(V) intermediate using I_2/pyridine/H_2O, which is finally hydrolyzed with aqueous ammonia to obtain the triphosphate product, as shown in **Scheme 1-3**. Compared with the "one-pot, three-step" approach, this one-pot reaction requires a 2'-3'-protected nucleoside analog to avoid the formation of by-products, despite the higher selectivity for the 5'-position.
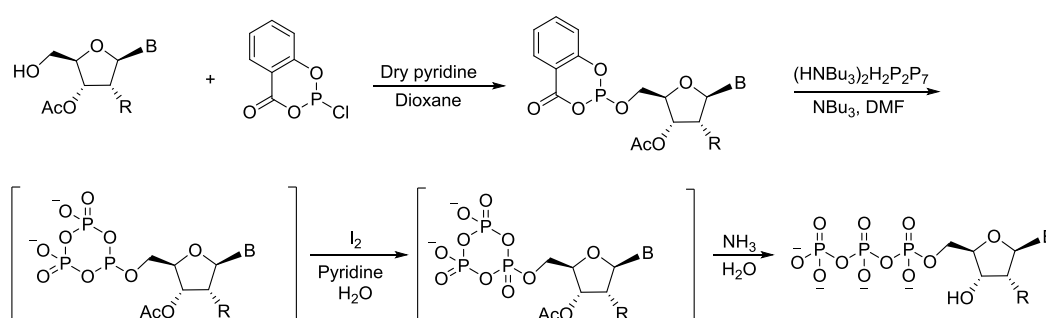
Scheme 1-3. Synthesis of base-modified NTPs via the Ludwig-Eckstein method (B = natural or modified nucleobase; R = H, OH, OAc).

1.5.2 Conjugates of nucleoside monophosphates and pyrophosphate mimic as nucleoside triphosphate analogs

Besides modifications of the nucleoside moiety of NTP, the triphosphate part can also be subjected to structural variation. Owing to the general assumption that the non-bridging oxygen atoms are vital for the metal chelation of the triphosphate moiety at the polymerase active site and for efficient enzymatic DNA polymerization, chemical modifications of the triphosphate moiety in the early days mainly concentrated on the replacement of the bridging oxygen atoms of the triphosphate moiety with S, NH, CH_2, CH_2O, CFH, and CF_2,[71-74] or the simple substitution of the P=O bond with a P=S bond.[75-77]

Further research demonstrated that some natural amino acids (e.g., L-aspartic acid, L-histidine)[78-79], unnatural amino acids (e.g., iminodiacetate (IDA),[80] iminodipropionate (IDP),[81] and 3-phosphono-L-alanine[82]) and even dipeptides[83] were able to function as pyrophosphate mimics in enzymatic DNA synthesis. The coupling of 2'-deoxyadenosine monophosphate (dAMP) with these pyrophosphate mimics yielded phosphoramidate analogs that act as NTP mimics (**Figure 1-7**), which are recognized successfully by polymerases. For example, L-Asp-dAMP and L-His-dAMP can be catalytically incorporated into the growing DNA strand by HIV-1 RT, which suggests that L-aspartic

acid and L-histidine possess the required structural and electronic features for efficient coordination with the catalytic metal ions at the polymerase active site. L-Asp-dAMP was incorporated more efficiently than L-His-dAMP by HIV-1 RT, whereas L-His-dAMP showed slightly better results for chain elongation compared with L-Asp-dAMP. Molecular modelling of these conjugates in the active site of HIV-1 RT demonstrated that they bind differently.[79] Unexpectedly, L-Glu-dAMP was not recognized by HIV-1 RT as a substrate, which implies that recognition and incorporation of these NTP mimics is a very specific process. Also, compared with D-Asp-dAMP, L-Asp-dAMP is a better substrate for RT, which indicates that the incorporation reaction is stereoselective.[79]
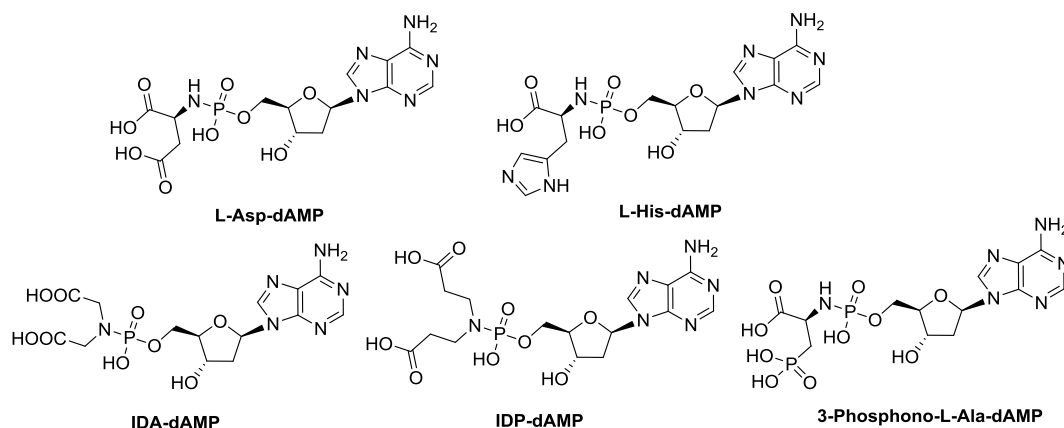
Figure 1-7. Examples of NTP mimics with amino acid as pyrophosphate mimic.

IDA-dAMP was also recognized by HIV-1 RT as a substrate for single nucleotide incorporation and shows a typical Michaelis-Menten kinetic profile. Better incorporation kinetics and chain elongation were obtained for IDA-dAMP that lacks a chiral center compared to L-Asp-dAMP and L-His-dAMP. The improved substrate properties of IDA-dAMP, when compared to L-Asp-dAMP, might result from the enhanced chelating capacity of the N-diacetate group. Furthermore, modeling of IDA-dAMP at the active site of HIV-1 RT indicates that the interaction between the triphosphate moiety, metal ions and residues of RT is improved by the involvement of six amino acid residues, rather than four for the natural substrate and five for L-Asp-dAMP.[80] IDP-dAMP, an analog of IDA-dAMP with an extended side chain in the amino acid moiety, is a better substrate than IDA-dAMP for DNA synthesis catalyzed by HIV-1 RT, with better kinetics and increased chain elongation capacity.[81]

Replacing the β-carboxylic acid of L-Asp-dAMP with a phosphonate group afforded 3-phosphono-L-Ala-dAMP, which is an improved substrate for HIV-1 RT catalyzed DNA synthesis, with a measured V_{max} that is only 1.3-fold lower than that for natural substrate dATP and a K_m which is 78-fold higher. The V_{max}/K_m value is 99 times lower than that for dATP. Compared with L-Asp-dAMP, this compound shows enhanced chain elongation, although the elongation is still less effective than that for dATP. Molecular docking of 3-phosphono-L-Ala-dAMP at the HIV-1 RT active site

suggested that the phosphoramidate and the phosphonate group mimic the *α*- and *γ*-phosphate groups of dNTPs, respectively. The carboxyl group acts as a mimic of the *β*-phosphate group of dNTPs.[84-85]

1.5.3 Alpha-carboxynucleoside phosphonates as nucleoside triphosphate mimics

From a structural standpoint of view, alpha-carboxynucleoside phosphonates (*α*-CNPs) are nucleoside phosphonate analogs with an additional carboxylic acid group at the alpha position of the phosphonate moiety and a cyclic or acyclic bridge that links the carboxy phosphonate moiety to the nucleobase (**Figure 1-8**). It is noteworthy that all *α*-CNPs were formed as mixtures of diastereomers at the carboxyphosphonate group. Different from typical nucleoside phosphonates, *α*-CNPs do not need any phosphorylation steps to exert their inhibitory activity against viral DNA polymerases. Therefore, *α*-CNPs mimic dNTPs and act directly as viral DNA polymerase inhibitors.[86] Unlike amino-acid-dNTP conjugates, *α*-CNPs do not get incorporated, but act as dNTP-competing inhibitors.
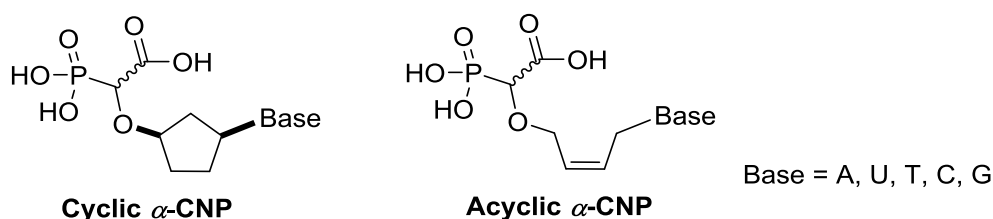
Figure 1-8. Chemical structures of cyclic and acyclic *α*-CNPs.

It was demonstrated that cyclic *α*-CNPs were able to selectively inhibit retroviral RT in the presence of a competing dNTP,[87-88] whereas acyclic *α*-CNPs showed potent inhibitory activity against herpesvirus DNA polymerases regardless of the presence of a competing dNTP.[89] Thus, it seems that cyclic *α*-CNPs bind HIV-1 RT in a way which is obviously different from the binding of acyclic *α*-CNPs to herpesvirus DNA polymerases. Specifically, cyclic *α*-CNPs competitively interact with the substrate-binding site of RT, but acyclic *α*-CNPs appear to bind to the herpetic DNA polymerase at a site that is different from the dNTP-binding site. To gain insight into the structural interactions of cyclic thymine alpha-carboxynucleoside phosphonate (T-*α*-CNP) with HIV-1 RT, crystal structures of T-*α*-CNP bound to RT-dsDNA or RT-DNA-aptamer were determined.[87, 90] The structural studies showed that the thymine of T-*α*-CNP base-paired with the adenine base and the *α*-carboxymethylphosphonate moiety mimics the coordination of triphosphate moiety of dNTP (**Figure 1-9**). Only the L-enantiomer of the T-*α*-CNP racemic mixture was active against HIV-1 RT.
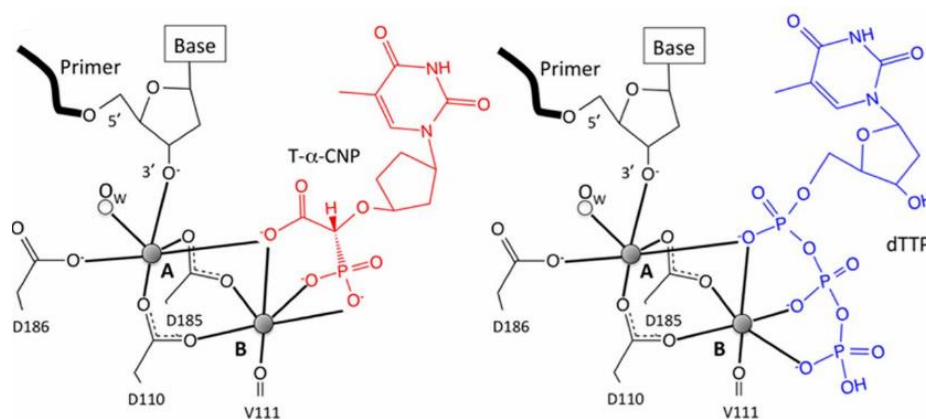
Figure 1-9. Structural interactions of cyclic T-α-CNP (Left) and the natural dTTP (Right) with HIV-1 RT and the template/primer. (Image taken from reference[87])

1.5.4 Prodrugs of nucleoside triphosphate analogs

Although many NTP analogs can bypass the activation pathway, they suffer from high polarity and poor stability, which are unfavorable for entering the host cells and thereby difficult to be used as antiviral drugs. Therefore, it is of great significance to develop and deliver appropriate prodrugs of NTP analogs which can not only skip all phosphorylation steps but also maximize the intracellular concentration of NTP analogs. However, because of their chemical instability, difficult synthesis, poor cellular uptake, and high sensitivity for enzymatic dephosphorylation, only very limited reports in the literature mentioned the design and synthesis of NTP prodrugs in the early days.[91-93] Even in the reported cases of NTP prodrugs, the yields of their chemical synthesis were very low.

When it comes to the development of NTP prodrugs, researchers always face one challenge associated with the labile phosphate anhydride bonds within the triphosphate moiety. Under physiological conditions, the reactive phosphate anhydride linkages are kinetically stable due to the presence of negative charges which stop nucleophilic attacks at the phosphate groups and turn the phosphates into bad leaving groups. Thus, a completely masked triphosphate prodrug may cause quick hydrolysis of the anhydride bonds.[94]

Recently, a Tri*PPPro* approach, a new prodrug strategy for NTPs, has been reported by Chris Meier's group.[95-100] They masked the negative charges at the γ phosphate of NTPs with two lipophilic and biodegradable moieties, which resulted in improved membrane permeability and chemical/enzymatic stability. The two masking groups could be identical or different, mainly including the ester moiety (acyloxybenzyl group) and the carbonate moiety (alkoxycarbonyloxybenzyl group), as exemplified with d4TTP prodrugs in **Figure 1-10**. Also, the lipophilicity can be modulated by utilizing different masking groups so that the NTP prodrugs pass the cell membrane efficiently.

R, R$_1$, R$_2$ = alkyl.
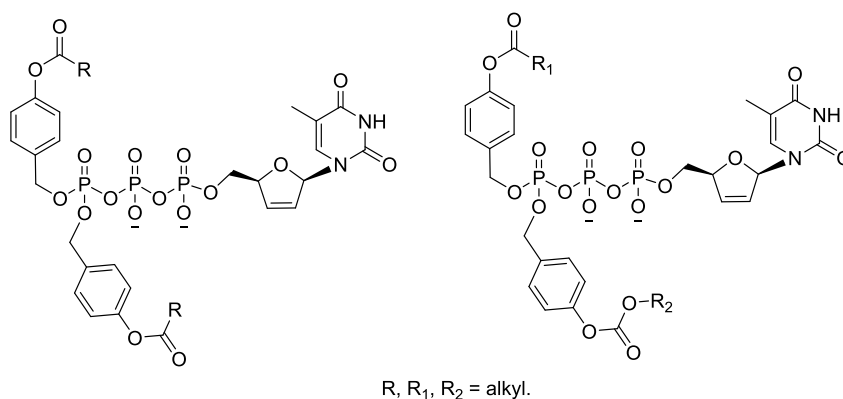
Figure 1-10. Examples of Tri*PPPro*-d4TTP prodrugs with symmetric (left) and asymmetric (right) masking moieties.

1.6 HIV-1 reverse transcriptase and its inhibitors

1.6.1 Structure of HIV-1 reverse transcriptase and DNA synthesis

The structure of HIV-1 RT was first reported nearly three decades ago.[101] Structural studies have demonstrated that HIV-1 RT is a heterodimer of two subunits, p66 and p51.[102] The two subunits share the same amino acid sequence, but p51 subunit lacks the RNase H domain. The p66 contains a polymerase domain at the N-terminus and an RNase H domain at the C-terminus. The polymerase domain resembles a right hand, possessing fingers, palm, thumb and connection subdomains. p51 also has these four subdomains, but their spatial arrangements are different from that in p66; p51 has no enzymatic activities, and it provides structural support to p66 by assembling its subdomains into a relatively rigid structure (**Figure 1-11**). The polymerase active site that is located in the palm subdomain of p66, is composed of three aspartic acid residues (D110, D185 and D186) which chelate the metal ions cofactor (Mg^{2+}) via their carboxylate groups.[103] D185 and D186 are part of the YMDD motif that is highly conserved in polymerases.[104] There are also conserved surrounding residues that take part in the formation of the dNTP binding site. For example, R72 and K65 are involved in the binding of the triphosphate moiety of the incoming dNTP. Q151 interacts with the 3'-OH of the incoming dNTP[105] and Y115 stacks with the deoxyribose ring of the incoming dNTP.[106]
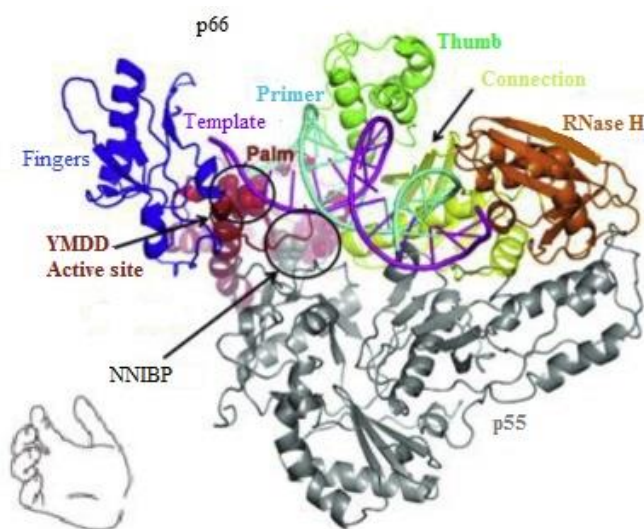
Figure 1-11. Structure of HIV-1 RT. (Image adapted from reference[107])

The DNA synthesis by HIV-1 RT starts with the binding of RT to the template/primer, which leads to a conformational change of the p66 thumb subdomain (from "close" to "open"). For a nucleotide incorporation, the primer 3'-end is positioned at the priming site (P-site) and a dNTP binds at the nucleotide binding site (N site), resulting in the formation of an HIV-1 RT/DNA/dNTP ternary complex.[108] The p66 fingers subdomain closes down to form the dNTP-binding pocket. The 3'-OH of the primer and the α-phosphate of the dNTP should be aligned precisely for catalytic incorporation of the dNMP part with the release of pyrophosphate. This is the rate-limiting step during the DNA synthesis by HIV-1 RT.[105, 109] The p66 fingers opening allows the pyrophosphate to be released and the translocation of the primer 3'-end to the P site that permits the binding and incorporation of the next nucleotide.

During DNA synthesis by HIV-1 RT, two divalent metal ions (Mg^{2+} *in vivo*) are required. This is because the metals can coordinate the oxygens of triphosphate moiety of the incoming dNTP and the side chains of the three catalytic residues (D110, D185 and D186). This kind of metal coordination at the polymerase site facilitates the attack of the 3'-OH on the α-phosphate of the incoming dNTP by activating the hydroxyl group (**Figure 1-12**). In addition, the two metal ions stabilize the negative charges of the resulting intermediates during DNA synthesis.[110]
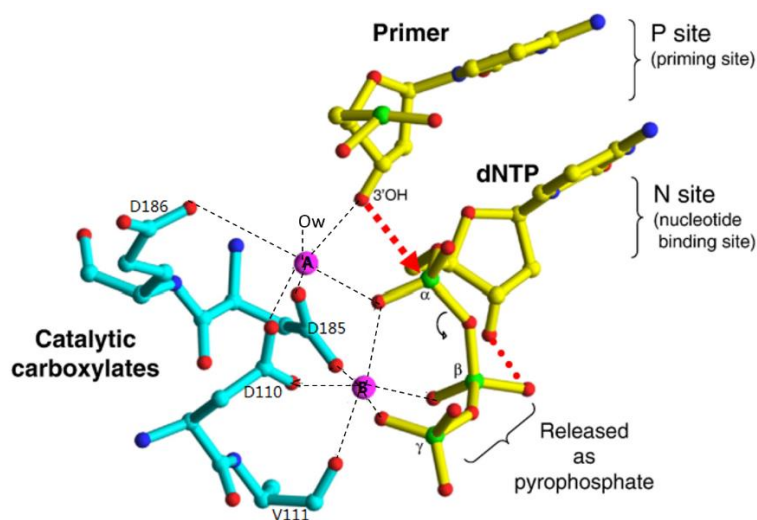
Figure 1-12. Metal chelation at the polymerase active site of HIV-1 RT. (Image modified from reference[111])

1.6.2 HIV-1 reverse transcriptase inhibitors

The approved RT inhibitors (RTIs) used for HIV/AIDS treatment are divided into two types that inhibit RT by different mechanisms of action. The first type is known as nucleoside/nucleotide RT inhibitors (NRTIs/NtRTIs) as discussed earlier, and the second group are the non-nucleoside RT inhibitors (NNRTIs). The standard three-drug regimens usually involve two NRTIs with either an NNRTI or a protease inhibitor (PI) or an integrase inhibitor (INSTI). Other combination therapies, which may involve the entry inhibitors, are used when the treatment is not successful using the standard therapies.

Drug Resistance to NRTIs/NtRTIs

Drug resistance remains a primary limitation in life-long HIV/AIDS treatment. Certain RT mutations can lead to NRTI/NtRTI resistance, which usually involves two kinds of mechanisms. The first mechanism is called as discrimination or exclusion in which mutations help discriminate an NRTI-TP or NtRTI-DP from the natural dNTP substrate. For example, M184V mutation can selectively reduce the incorporation of 3TC-TP by steric hindrance.[112-113] The second mechanism is named as excision which involves selective removal ("unblocking") of an NRTI-MP or NtRTI from the DNA primer by reversing the direction of RT-catalysis to excise the chain terminator from the primer 3'-end.[114] For instance, AZT resistance can result from a set of five mutations (D67N, K70R, K219Q, M41L, T215Y), which are also called as thymidine-analog mutations (TAMs).[115] TAMs facilitate binding of an ATP as the excision substrate and help remove the incorporated AZT-MP from DNA strand with the release of the excision product AZTppppA.[116-117]

Non-nucleoside RT Inhibitors. NNRTIs non-competitively bind to a hydrophobic pocket near the polymerase active site of RT, which induces conformational changes in some residues, makes the

p66 thumb subdomain more rigid and thus allosterically block DNA synthesis.[118] Unlike NRTIs, NNRTIs do not need cellular conversion to exert their activity. Nowadays, there are six NNRTIs approved by the FDA and they can be classified into two generations according to their binding mode and resilience of resistance mutations. The first generation NNRTIs include nevirapine, efavirenz, delavirdine, while the second generation NNRTIs contain etravirine, rilpivirine and doravirine (**Figure 1-13**).
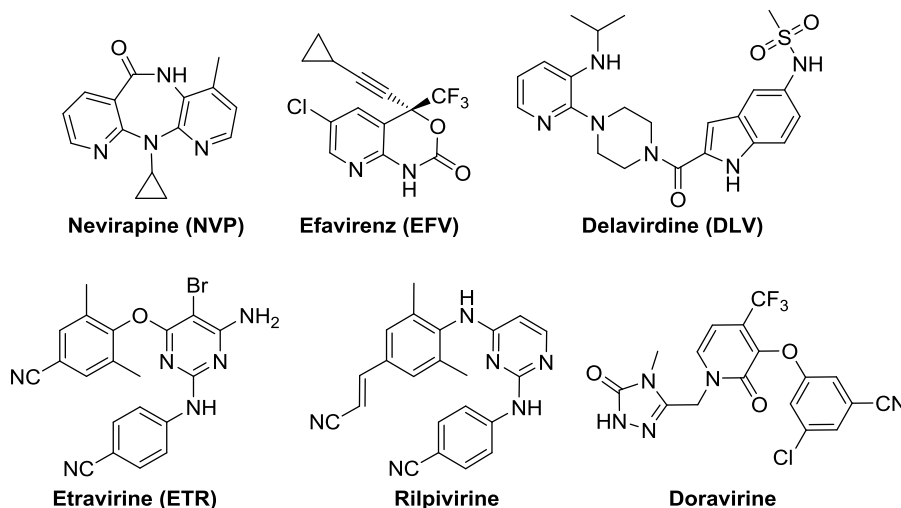
Figure 1-13. Approved NNRTIs for HIV/AIDS treatment.

RT mutations can cause resistance to all approved NNRTIs, and most of these mutations are found in or around the NNRTIs binding pocket (NNIBP). Despite their diverse chemical structures, the first-generation NNRTIs share a "butterfly-like" binding conformation.[119] They are vulnerable to several resistance mutations which reduce their inhibitory activity rapidly. For example, the mutations (Y181C, Y188L and F227L) in the hydrophobic core of NNIBP can lead to NNRTI resistance due to the loss of vital hydrophobic interactions with NNRTIs. The V106A and V179D mutations induce resistance by changing the hydrophobic interaction.[120] The mutations (L100I and G190A) in the middle region of NNIBP can also induce resistance because L100I directly changes the shape of NNIBP whereas G190A can develop steric hindrance with an NNRTI. In addition, the mutations (K103N and K101E) at the rim of NNIBP entrance result in resistance by hindering NNRTIs getting into NNIBP.[111] Compared with the first generation NNRTIs, the second generation NNRTIs show a higher genetic barrier to drug resistance, because they display conformational flexibility which helps them adapt to RT binding site and maintain their binding efficacy to RT even in the presence of some NNIBP mutations.[121] However, they still face issues of resistance mutations. For instance, mutations Y181I and Y181V cause severe resistance to rilpivirine and etravirine, whereas the K101P mutation induces resistance to rilpivirine.[122]

Other RT Inhibitors. Despite the strong potency and extensive use of marketed N(t)RTIs and NNRTIs, the emergence of drug resistance and toxicity due to long-term use of these drugs urges the

need for finding new RTIs with a different resistance profile and/or a distinct mechanisms of action. As a result, many new classes of RT inhibitors were discovered, such as nucleoside RT translocation inhibitors (NRTTIs), nucleotide competing RT Inhibitors (NcRTIs) and pyrophosphate analog inhibitors.[123-124]

NRTTIs block the translocation of HIV-RT. 4'-Ethynyl-2-fluoro-2'-deoxyadenosine (EFdA or MK-8591) (**Figure 1-14**) is a representative example, which has a 3'-hydroxyl and a 4'-ethynyl group on the sugar moiety. EFdA-TP can be recognized by HIV-1 RT, and is then incorporated into the growing DNA strand. However, because RT cannot be translocated efficiently on the DNA strand, further DNA elongation is blocked by the incorporated EFdA-MP.[125-126] EFdA is a highly potent RT inhibitor that inhibits both wild type and NRTI-resistant HIV-1 strains and is currently under clinical development.[127]
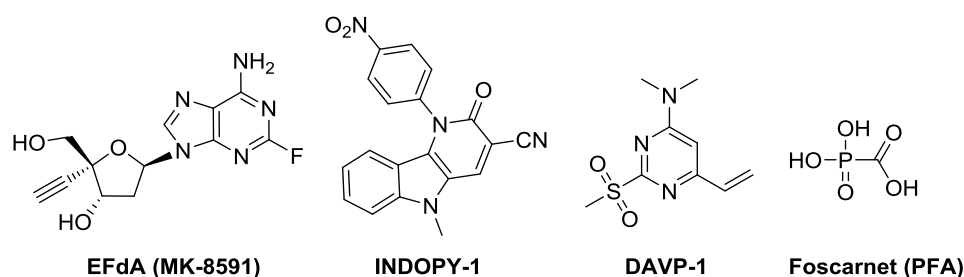
Figure 1-14. Examples of other RTIs.

NcRTIs bind near the polymerase active site of HIV-1 RT and compete with the natural nucleotide substrates. Indolopyridones (INDOPYs), such as INDOPY-1 (**Figure 1-14**), were first reported in 2006.[128-129] It was shown that INDOPY-1 was not only active against HIV-1, but also displayed antiviral activity against HIV-2, which is different from typical NNRTIs that are very specific for HIV-1.[128] Also, INDOPYs display a unique drug resistance profile. For example, INDOPY-1 retains activity toward common RTI-induced mutations (K103N, Y181C, Y188C, *etc*.) and TAMs. In contrast, M184V/Y115F double mutations reduce its activity, and K65R mutation increases its activity.[129-130] 4-Dimethylamino-6-vinylpyrimidines (DAVPs), such as DAVP-1 (**Figure 1-14**), were also discovered and classified as NcRTIs due to their competitive inhibition with the natural nucleotide substrate.[131-132] DAVP-1 maintains its potency against the L100I, Y179D or Y181I mutations, but it is severely weakened by the K103N mutation.[131] Structural analysis demonstrated that the DAVP-1 binding site is located in a region that is vital for the correct positioning of the 3'-OH of the DNA primer during DNA synthesis.[133] In addition to INDOPYs and DAVPs, cyclic α-CNPs compete with dNTPs for binding to RT. Therefore, α-CNPs are considered as NcRTIs as well.

A representative pyrophosphate analog inhibitor is foscarnet (phosphonoformate, PFA) (**Figure 1-14**), which not only inhibits DNA polymerase of herpesviruses but is also an inhibitor of RT of

retroviruses.[134] PFA is utilized to treat herpesvirus infections, particularly CMV retinitis in HIV-infected patients as part of salvage therapy.[135] It is reported that PFA competitively blocks pyrophosphorolysis and pyrophosphate exchange reactions, indicating that the binding site of PFA and pyrophosphate can overlap.[136]

1.7 Structure-based design approaches in the discovery of antivirals

Classical drug discovery methods are usually time-consuming and expensive. To overcome the limitations, rational drug design becomes more and more popular since it is more efficient and economical. The increasing knowledge of the structure and function of proteins lays a foundation for rational drug design and discovery. Advances in biomolecular structural biology techniques (e.g., X-ray crystallography, nuclear magnetic resonance and cryo-electron microscopy) have offered a huge number of (> 100,000) three-dimensional (3D) structures of various proteins and complexes.[137] The availability of the 3D structures of target macromolecules and their interaction with a ligand has paved the way for structure-based design approaches. X-ray crystallography plays a central role in visualizing and studying the binding mode of various classes of inhibitors at or near atomic resolution. Detailed structural studies of ligand-protein complexes can provide important molecular insight into the binding sites of various protein targets, and the structural information is extensively used to develop new compounds with increased potency, affinity and selectivity.[138]

Structure-based design approaches have played an important role in the discovery of antivirals, especially HIV protease inhibitors. Saquinavir was the first HIV-1 protease inhibitor on the market. In the early days, the knowledge and experience from designing renin inhibitors, and the determination of X-ray structure of HIV-1 protease resulted in rapid progress of structure-based design approaches.[139] The availability of more X-ray structures of HIV protease, inhibitor-bound HIV-1 protease, and mutant proteases greatly assisted in designing new inhibitors. Analogous approaches were also applied to many other antivirals like neuraminidase inhibitors and NNRTIs. Currently, structure-based approaches have contributed to the discovery of at least 16 new drugs approved for the treatment of viral diseases such as HIV/AIDS and influenza.

1.8 Objectives of the research

Objective 1

As mentioned above, NTP analogs hold great potential as antiviral drugs because they can skip the inefficient phosphorylation steps. Previous literature data demonstrated that selected amino acid conjugates of dAMP can function as substrates in DNA synthesis, catalyzed by HIV RT. Therefore, in the first part of this doctoral thesis, the synthesis of a series of NTP analogs with dAMP and TFV as the NMP moiety was carried out (**Figure 1-15**). These compounds were evaluated in two different

assays. A first HIV-1 RT incorporation assay was applied to investigate if these compounds were recognized as potential substrates by RT and incorporated into a growing DNA primer strand. In addition, a biochemical assay was used to investigate if these compounds act as inhibitors of HIV-1 RT. In order to study their binding mode, individual compounds were soaked into crystals of an RT/dsDNA catalytic complex, allowing to obtain X-ray structures of RT/dsDNA/inhibitor ternary complexes. The systematic study of these newly synthesized NTP analogs by enzymatic experiments and X-ray crystallography will assist in the design of new NTP analogs with different metal chelating groups as pyrophosphate mimics.

R= L-Asp, D-Asp, L-His, L-Glu, L-Met, L-Arg, L-Phe.
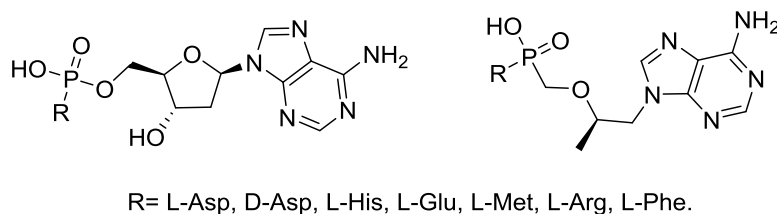
Figure 1-15. Target NTP analogs with dAMP and TFV as the NMP moiety.

Objective 2

Because of the chemical instability and high polarity of the amino acid conjugates of dAMP or TFV, a second series of TFV analogs, in which the phosphonoamidate moiety was replaced by a stable and neutral amide linker (**Figure 1-16**), was prepared in **Chapter 3**. These compounds were evaluated as potential inhibitors of HIV-1 RT in a biochemical assay. Several ester prodrugs were also prepared to improve membrane permeability. All compounds were screened in an antiviral HIV assay. Moreover, individual compounds were soaked into the crystals of RT/dsDNA catalytic complex to obtain the crystal structures of RT/dsDNA/inhibitor ternary complexes.

AA= β -Ala, D-Asp, L-Asp, etc.
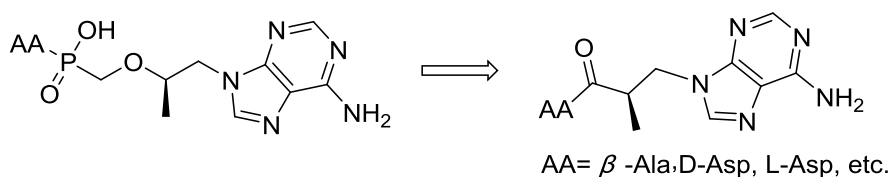
Figure 1-16. Introduction of a stable and uncharged amide linker.

Objective 3

 The lack of desirable metal chelation in the amide analogs of **Chapter 3** enforced us to prepare a novel series of ANP analogs, mostly carrying a phosphonate group and a carboxyl group on the acyclic side chain (**Figure 1-17**). Variations were made on chain length, linker atoms as well as chirality so that they could form new interactions. Seven compounds were synthesized and evaluated for HIV-1 RT inhibition. In addition, the crystal structures of RT/dsDNA/inhibitor ternary complexes were obtained to get insights into their binding mode at the polymerase active site.
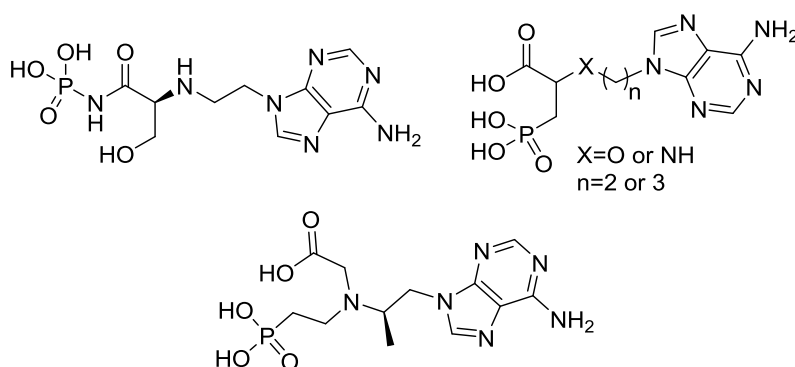
Figure 1-17. Target ANP analogs.

Objective 4

Our laboratory carried out an X-ray crystallographic screening of fragments for binding to HIV-1 RT. This experiment was conducted at the Xchem facility in the Diamond Light Source, UK. From the screening of 300 fragments, two fragments **048** and **166** (**Figure 1-18**) were discovered that bind to the P site of HIV-1 RT. This P-pocket is a novel transient site that is presumably created while RT slides over its dsDNA substrate. Based on X-ray structure of fragment hit **166** and the pocket composition, we designed a series of new ligands (**Figure 1-19**). To study their possible binding mode, molecular docking was performed using the software AutoDock Vina.[140] Five ligands were initially synthesized, and another two ligands were selected for synthesis based on results of the docking study. Structural studies and biological evaluation of the synthesized compounds were also performed.
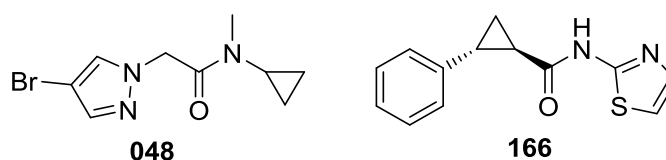
Figure 1-18. Hit from fragment screening.
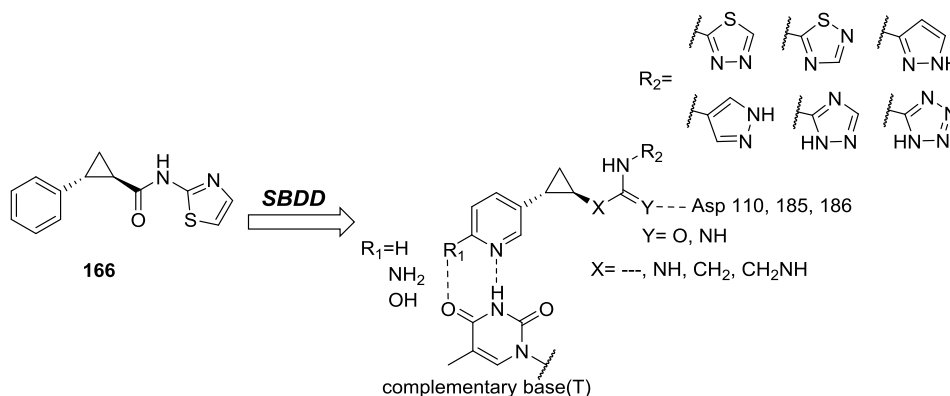
Figure 1-19. Designed compounds based on fragment **166**.

1.9 References

1. Blackburn, G. M.; Gait, M. J.; Loakes, D.; Williams, D. M. *Nucleic Acids in Chemistry and Biology: Edition 3*. Royal Society of Chemistry Publishing: **2006**.
2. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry, 6th Edition*. W.H. Freeman and Company, New York, USA: **2012**.
3. Jordheim, L. P.; Durantel, D.; Zoulim, F.; Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discov.* **2013**, 12, 447-464.
4. De Clercq, E.; Li, G. Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev* **2016,** 29, 695-747.
5. Andriole, V. T. Current and future antifungal therapy: new targets for antifungal agents. *J Antimicrob Chemoth* **1999,** 44, 151-162.
6. Cihlar, T.; Ray, A. S. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res.* **2010,** 85, 39-58.
7. Safadi, Y. E.; Vivet-Boudou, V.; Marquet, R. HIV-1 reverse transcriptase inhibitors. *Appl Microbiol Biot* **2007,** 75, 723-737.
8. Burton, J. R.; Everson, G. T. HCV NS5B Polymerase Inhibitors. *Clin Liver Dis* **2009,** 13, 453-465.
9. Deval, J. Antimicrobial Strategies. *Drugs* **2009,** 69, 151-166.
10. Sheahan, T. P.; Sims, A. C.; Zou, S.; *et al.* An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci. Transl. Med.* **2020,** 12, eabb5883.
11. Prusoff, W. H. Synthesis and biological activities of iododeoxyuridine, an analog of thymidine. *Biochim. Biophys. Acta* **1959,** 32, 295-296.
12. De Clercq, E. In Search of Selective Antiviral Chemotherapy. *Clin Microbiol Rev* **1997,** 10, 674-693.
13. Field, H. J.; De Clercq, E. Antiviral drugs - a short history of their discovery and development. *Microbiol. Today* **2004,** 31, 58-61.
14. Suzuki, M.; Okuda, T.; Shiraki, K. Synergistic antiviral activity of acyclovir and vidarabine against herpes simplex virus types 1 and 2 and varicella-zoster virus. *Antiviral Res.* **2006,** 72, 157-161.
15. Koomen, G. J. Synthesis and biological properties of selected nucleoside analogues. *Recl. Trav. Chim. Pays-Bas* **1993,** 112, 51-65.
16. Seley-Radtke, K. L.; Yates, M. K. The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Res.* **2018,** 154, 66-86.
17. King, D. H. History, pharmacokinetics, and pharmacology of acyclovir. *J Am Acad Dermatol* **1988,** 18, 176-179.
18. Elion, G. B. Acyclovir: discovery, mechanism of action, and selectivity. *Journal of Medical Virology (Suppl. 1)* **1993**, 2-6.
19. De Clercq, E.; Field, H. J. Antiviral prodrugs - the development of successful prodrug strategies for antiviral chemotherapy. *Brit J Pharmacol* **2006,** 147, 1-11.
20. Mathé, C.; Gosselin, G. L-nucleoside enantiomers as antivirals drugs: a mini-review. *Antiviral Res.* **2006,** 71, 276-281.
21. Han, S. H. Telbivudine: a new nucleoside analogue for the treatment of chronic hepatitis B. *Expert Opin. Investig. Drugs* **2005,** 14, 511-519.
22. Kim, J. W.; Park, S. H.; Louie, S. G. Telbivudine: a novel nucleoside analog for chronic hepatitis B. *Ann. Pharmacother.* **2006,** 40, 472-478.
23. Kaufman, H. E.; Heidelberger, C. Therapeutic Antiviral Action of 5-Trifluoromethyl-2'-deoxyuridine in Herpes Simplex Keratitis. *Science* **1964,** 145, 585-586.
24. Whitley, R. J. The past as prelude to the future: history, status, and future of antiviral drugs. *Ann. Pharmacother.* **1996,** 30, 967-971.
25. De Clercq, E. Historical Perspectives in the Development of Antiviral Agents Against Poxviruses. *Viruses* **2010,** 2, 1322-1339.
26. De Clercq, E. Selective Anti-Herpesvirus Agents. *Antivir Chem Chemoth* **2013,** 23, 93-101.
27. Crotty, S.; Maag, D.; Arnold, J. J. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nature Medicine* **2000,** 6, 1375-1379.
28. Balzarini, J. Metabolism and mechanism of antiretroviral action of purine and pyrimidine derivatives. *Pharmacy World and Science* **1996,** 16, 113-126.
29. Hao, Z.; Cooney, D. A.; Hartman, N. R.; Perno, C. F.; Fridland, A.; DeVico, A. L.; Sarngadharan, M. G.; Broder, S.; Johns, D. G. Factors determining the activity of 2',3'-dideoxynucleosides in suppressing human immunodeficiency virus in vitro. *Molecular pharmacology* **1988,** 34, 431-435.
30. Balzarini, J.; Kang, G. J.; Dalal, M.; Herdewijn, P.; De Clercq, E.; Broder, S.; Johns, D. G. The anti-HTLV-III (anti-HIV) and cytotoxic activity of 2', 3'-didehydro-2', 3'-dideoxyribonucleosides: a comparison with their parental 2', 3'-dideoxyribonucleosides. *Molecular pharmacology* **1987,** 32, 162-167.
31. Thornton, P. J.; Kadri, H.; Miccoli, A.; Mehellou, Y. Nucleoside Phosphate and Phosphonate Prodrug Clinical Candidates. *J. Med. Chem.* **2016,** 59, 10400-10410.
32. Pradere, U.; Garnier-Amblard, E. C.; Coats, S. J.; Amblard, F.; Schinazi, R. F. Synthesis of Nucleoside Phosphate and Phosphonate Prodrugs. *Chem. Rev.* **2014,** 114, 9154-9218.

33. Hecker, S. J.; Erion, M. D. Prodrugs of Phosphates and Phosphonates. *J. Med. Chem.* **2008,** 51, 2328-2345.

34. Mehellou, Y.; Rattan, H. S.; Balzarini, J. The ProTide Prodrug Technology: From the Concept to the Clinic. *J. Med. Chem.* **2018,** 61, 2211-2226.

35. Wagner, C. R.; Iyer, V. V.; McIntee, E. J. Pronucleotides: toward the in vivo delivery of antiviral and anticancer nucleotides. *Med. Res. Rev.* **2000,** 20, 417-451.

36. Mehellou, Y.; Balzarini, J.; McGuigan, C. Aryloxy Phosphoramidate Triesters: a Technology for Delivering Monophosphorylated Nucleosides and Sugars into Cells. *Chemmedchem* **2009,** 4, 1779-1791.

37. Mehellou, Y. The ProTides Boom. *Chemmedchem* **2016,** 11, 1114-1116.

38. Sofia, M. J.; Bao, D.; Chang, W. Discovery of a *β*-D-2′-Deoxy-2′-*α*-fluoro-2′-*β*-C-methyluridine Nucleotide Prodrug (PSI-7977) for the Treatment of Hepatitis C Virus. *J. Med. Chem.* **2010,** 53, 7202-7218.

39. Siegel, D.; Hui, H. C.; Doerffler, E.; *et al*. Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo[2,1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. *J. Med. Chem.* **2017,** 60, 1648-1661.

40. Warren, T.; Jordan, R.; Lo, M.; *et al*. Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature* **2016,** 531, 381-385.

41. Cho, A.; Saunders, O. L.; Butler, T.; Zhang, L.; Xu, J.; Vela, J. E.; Feng, J. Y.; Ray, A. S.; Kim, C. U. Synthesis and antiviral activity of a series of 1'-substituted 4-aza-7,9-dideazaadenosine C-nucleosides. *Bioorganic & Medicinal Chemistry Letters* **2012,** 22, 2705-2707.

42. Grein, J.; Ohmagari, N.; Shin, D.; *et al*. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N. Engl. J. Med.* **2020,** 382, 2327-2336.

43. Bravo, J. P. K.; Dangerfield, T. L.; Taylor, D. W.; Johnson, K. A. Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication. *Molecular Cell* **2021,** 81, 1548-1552.e4.

44. Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Höbartner, C.; Cramer, P. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat Commun* **2021,** 12.

45. De Clercq, E.; Holy, A. Acyclic nucleoside phosphonates: A key class of antiviral drugs. *Nature Reviews Drug Discovery* **2005,** 4, 928-940.

46. Piperno, A.; Chiacchio, M. A.; Iannazzo, D.; Romeo, R. Synthesis and biological activity of phosphonated nucleosides: part 1. Furanose, carbocyclic and heterocyclic analogues. *Curr Med Chem* **2006,** 13, 3675-3695.

47. De Clercq, E.; Holý, A.; Rosenberg, I.; Sakuma, T.; Balzarini, J.; Maudgal, P. C. A novel selective broad-spectrum anti-DNA virus agent. *Nature* **1986,** 323, 464-467.

48. De Clercq, E.; Sakuma, T.; Baba, M.; Pauwels, R.; Balzarini, J.; Rosenberg, I.; Holy, A. Antiviral activity of phosphonylmethoxyalkyl derivatives of purine and pyrimidines. *Antiviral Res.* **1987,** 8, 261-272.

49. Wu, T.; Froeyen, M.; Kempeneers, V.; Pannecouque, C.; Wang, J.; Busson, R.; De Clercq, E.; Herdewijn, P. Deoxythreosyl phosphonate nucleosides as selective anti-HIV agents. *J. Am. Chem. Soc.* **2005,** 127, 5056-5065.

50. Liu, C.; Dumbre, S. G.; Pannecouque, C.; Huang, C.; Ptak, R. G.; Murray, M. G.; De Jonghe, S.; Herdewijn, P. Amidate Prodrugs of Deoxythreosyl Nucleoside Phosphonates as Dual Inhibitors of HIV and HBV Replication. *J. Med. Chem.* **2016,** 59, 9513-9531.

51. Mackman, R. L.; Zhang, L.; Prasad, V.; Boojamra, C. G.; Chen, J.; Douglas, J.; Grant, D.; Laflamme, G.; Hui, H.; Kim, C. U.; Parrish, J.; Stoycheva, A. D.; Swaminathan, S.; Wang, K.; Cihlar, T. Synthesis And Anti-Hiv Activity Of Cyclic Pyrimidine Phosphonomethoxy Nucleosides And Their Prodrugs: A Comparison Of Phosphonates And Corresponding Nucleosides. *Nucleosides, Nucleotides & Nucleic Acids* **2007,** 26, 573-577.

52. Ray, A. S.; Vela, J. E.; Boojamra, C. G. Intracellular Metabolism of the Nucleotide Prodrug GS-9131, a Potent Anti-Human Immunodeficiency Virus Agent. *Antimicrob. Agents Chemother.* **2008,** 52, 648-654.

53. White, K. L.; Margot, N.; Stray, K.; Yu, H.; Stepan, G.; Boojamra, C.; Mackman, R.; Ray, A.; Miller, M. D.; Cihlar, T. In *GS-9131 is a novel NRTI with activity against NRTI-resistant HIV-1*, The Conference on Retroviruses and Opportunistic Infections, Seattle, Washington, USA, 13-16 February 2017, 2017; Seattle, Washington, USA, 2017.

54. Cano-Soldado, P.; Pastor-Anglada, M. Transporters that translocate nucleosides and structural similar drugs: structural requirements for substrate recognition. *Med. Res. Rev.* **2011,** 32, 428-457.

55. Minuesa, G.; Huber-Ruano, I.; Pastor-Anglada, M.; Koepsell, H.; Clotet, B.; Martinez-Picado, J. Drug uptake transporters in antiretroviral therapy. *Pharmacol. Ther.* **2011,** 132, 268-279.

56. Merta, A.; Votruba, I.; Jindřich, J.; Holy, A.; Cihlar, T.; Rosenberg, I.; Otmar, M.; Herve, T. Y. Phosphorylation of 9-(2-phosphonomethoxyethyl) adenine and 9-(S)-(3-hydroxy-2-phosphonomethoxypropyl) adenine by AMP (dAMP) kinase from L1210 cells. *Biochem Pharmacol* **1992,** 44, 2067-2077.

57. Cihlar, T.; Chen, M. S. Identification of enzymes catalyzing two-step phosphorylation of cidofovir and the effect of cytomegalovirus infection on their activities in host cells. *Mol. Pharmacol.* **1996,** 50, 1502-1510.

58. Xiong, S.; Smith, J. L.; Chen, M. S. Effect of incorporation of cidofovir into DNA by human cytomegalovirus DNA polymerase on DNA elongation. *Antimicrob. Agents Chemother.* **1997,** 41, 594-599.

59. Balzarini, J.; Hao, Z.; Herdewijn, P.; Johns, D. G.; De Clercq, E. Intracellular metabolism and mechanism of anti-retrovirus action of 9-(2-phosphonylmethoxyethyl) adenine, a potent anti-human immunodeficiency virus compound. *Proc. Natl. Acad. Sci. USA* **1991,** 88, 1499-1503.

60. Furman, P. A.; Fyfe, J. A.; St Clair, M. H. Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **1986,** 83, 8333-8337.

61. Balzarini, J.; Herdewijn, P.; De Clercq, E. Differential patterns of intracellular metabolism of 2', 3'-didehydro-2', 3'-dideoxythymidine and 3'-azido-2', 3'-dideoxythymidine, two potent anti-human Immunodeficiency Virus Compounds. *J. Biol. Chem.* **1989,** 264, 6127-6133.

62. Pertenbreiter, F.; Balzarini, J.; Meier, C. Nucleoside Mono- and Diphosphate Prodrugs of 2', 3'-Dideoxyuridine and 2', 3'-Dideoxy-2', 3'-didehydrouridine. *Chemmedchem* **2015,** 10, 94-106.

63. Burgess, K.; Cook, D. Syntheses of nucleoside triphosphates. *Chem. Rev.* **2000,** 100, 2047-2060.

64. Kore, A. R.; Srinivasan, B. Recent Advances in the Syntheses of Nucleoside Triphosphates. *Curr. Org. Synth.* **2013,** 10, 903-934.

65. Hollenstein, M. Nucleoside Triphosphates-Building Blocks for the Modification of Nucleic Acids. *Molecules* **2012,** 17, 13569-13591.

66. Roy, B.; Depaix, A.; Perigaud, C.; Peyrottes, S. Recent Trends in Nucleotide Synthesis. *Chem. Rev.* **2016,** 116, 7854-7897.

67. Ludwig, J. A new route to nucleoside 5'-triphosphates. *Acta Biochim. Biophys. Acad. Sci. Hung.* **1981,** 16, 131-133.

68. Yoshikawa, M.; Kato, H.; Takenishi, T. A novel method for phosphorylation of nucleosides to 5'-nucleotides. *Tetrahedron Lett.* **1967,** 8, 5065-5068.

69. Wang, G.; Wan, J.; Hu, Y. Synthesis and Anti-Influenza Activity of Pyridine, Pyridazine, and Pyrimidine C-Nucleosides as Favipiravir (T-705) Analogues. *J. Med. Chem.* **2016,** 59, 4611-4624.

70. Ludwig, J.; Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **1989,** 54, 631-635.

71. Boyle, N. A.; Rajwanshi, V. K.; Prhavc, M.; Wang, G.; Fagan, P.; Chen, F.; Ewing, G. J.; Brooks, J. L.; Hurd, T.; Leeds, J. M.; Bruice, T. W.; Cook, P. D. Synthesis of 2',3'-Dideoxynucleoside 5'-α-P-Borano-β,γ-(difluoromethylene)triphosphates and Their Inhibition of HIV-1 Reverse Transcriptase. *J. Med. Chem.* **2005,** 48, 2695-2700.

72. Boyle, N. A.; Fagan, P.; Brooks, J. L.; Prhavc, M.; Lambert, J.; Cook, P. D. 2',3'-Dideoxynucleoside 5'-β,γ-(Difluoromethylene) Triphosphates With α-P-Thio or α-P-Seleno Modifications: Synthesis and Their Inhibition of HIV-1 Reverse Transcriptase. *Nucleosides, Nucleotides & Nucleic Acids* **2005,** 24, 1651-1664.

73. Wang, G.; Boyle, N.; Chen, F.; Rajappan, V.; Fagan, P.; Brooks, J. L.; Hurd, T.; Leeds, J. M.; Rajwanshi, V. K.; Jin, Y.; Prhavc, M.; Bruice, T. W.; Cook, P. D. Synthesis of AZT 5'-Triphosphate Mimics and Their Inhibitory Effects on HIV-1 Reverse Transcriptase. *J. Med. Chem.* **2004,** 47, 6902-6913.

74. Wellmar, U.; Hörnfeldt, A. B.; Gronowitz, S.; Johansson, N. G. Synthesis of Mimics to Thymidine and 5-(2"-Thienyl)-2'-Deoxyuridine Triphosphates. *Nucleosides and Nucleotides* **1996,** 15, 1059-1076.

75. Goody, R. S.; Eckstein, F. Thiophosphate analogs of nucleoside di- and triphosphates. *J. Am. Chem. Soc.* **1971,** 93, 6252-6257.

76. Eckstein, F.; Goody, R. S. Synthesis and properties of diastereoisomers of adenosine 5'-(O-1-thiotriphosphate) and adenosine 5'-(O-2-thiotriphosphate). *Biochemistry* **1976,** 15, 1685-1691.

77. Eckstein, F.; Gindl, H. Synthesis of nucleoside 5'-polyphosphorothioates. *Biochim. Biophys. Acta* **1967,** 149, 35-40.

78. Adelfinskaya, O.; Herdewijn, P. Amino acid phosphoramidate nucleotides as alternative substrates for HIV-1 reverse transcriptase. *Angew Chem Int Ed Engl* **2007,** 46, 4356-8.

79. Adelfinskaya, O.; Terrazas, M.; Froeyen, M.; Marliere, P.; Nauwelaerts, K.; Herdewijn, P. Polymerase-catalyzed synthesis of DNA from phosphoramidate conjugates of deoxynucleotides and amino acids. *Nucleic Acids Res.* **2007,** 35, 5060-5072.

80. Giraut, A.; Song, X. P.; Froeyen, M.; Marliere, P.; Herdewijn, P. Iminodiacetic-phosphoramidates as metabolic prototypes for diversifying nucleic acid polymerization in vivo. *Nucleic Acids Research* **2010,** 38, 2541-2550.

81. Song, X. P.; Bouillon, C.; Lescrinier, E.; Herdewijn, P. Iminodipropionic acid as the leaving group for DNA polymerization by HIV-1 reverse transcriptase. *ChemBioChem* **2011,** 12, 1868-1880.

82. Yang, S. Q.; Froeyen, M.; Lescrinier, E.; Marliere, P.; Herdewijn, P. 3-Phosphono-L-alanine as pyrophosphate mimic for DNA synthesis using HIV-1 reverse transcriptase. *Organic & Biomolecular Chemistry* **2011,** 9, 111-119.

83. Song, X. P.; Bouillon, C.; Lescrinier, E.; Herdewijn, P. Dipeptides as leaving group in the enzyme-catalyzed DNA synthesis. *Chem Biodivers* **2012,** 9, 2685-700.

84. Yang, S.; Froeyen, M.; Lescrinier, E.; Marliere, P.; Herdewijn, P. 3-Phosphono-L-alanine as pyrophosphate mimic for DNA synthesis using HIV-1 reverse transcriptase. *Org. Biomol. Chem.* **2011,** 9, 111-119.

85. Herdewijn, P.; Marlière, P. Redesigning the leaving group in nucleic acid polymerization. *Febs Lett* **2012,** 586, 2049-2056.

86. Balzarini, J.; Ford, A.; Maguire, N. M.; John, J.; Das, K.; Arnold, E.; Dehaen, W.; Maguire, A. Alpha-carboxynucleoside phosphonates: direct-acting inhibitors of viral DNA polymerases. *Future Med Chem* **2019,** 11, 137-154.

87. Balzarini, J.; Das, K.; Bernatchez, J. A.; Martinez, S. E.; Ngure, M.; Keane, S.; Ford, A.; Maguire, N.; Mullins, N.; John, J.; Kim, Y.; Dehaen, W.; Vande Voorde, J.; Liekens, S.; Naesens, L.; Gotte, M.; Maguire, A. R.; Arnold, E. Alpha-carboxy nucleoside phosphonates as universal nucleoside triphosphate mimics. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** 112, 3475-3480.

88. Balzarini, J.; Menni, M.; Das, K.; van Berckelaer, L.; Ford, A.; Maguire, N. M.; Liekens, S.; Boehmer, P. E.; Arnold, E.; Gotte, M.; Maguire, A. R. Guanine alpha-carboxy nucleoside phosphonate (G-alpha-CNP) shows a different inhibitory kinetic profile against the DNA polymerases of human immunodeficiency virus (HIV) and herpes viruses. *Biochem*

Pharmacol **2017,** 136, 51-61.

89. John, J.; Kim, Y.; Bennett, N.; Das, K.; Liekens, S.; Naesens, L.; Arnold, E.; Maguire, A. R.; Gotte, M.; Dehaen, W.; Balzarini, J. Pronounced Inhibition Shift from HIV Reverse Transcriptase to Herpetic DNA Polymerases by Increasing the Flexibility of alpha-Carboxy Nucleoside Phosphonates. *Journal of Medicinal Chemistry* **2015,** 58, 8110-8127.

90. Das, K.; Balzarini, J.; Miller, M. T.; Maguire, A. R.; DeStefano, J. J.; Arnold, E. Conformational States of HIV-1 Reverse Transcriptase for Nucleotide Incorporation vs Pyrophosphorolysis-Binding of Foscarnet. *Acs Chem Biol* **2016,** 11, 2158-64.

91. Bonnaffé, D.; Dupraz, B.; Ughetto-Monfrin, J.; Namane, A.; Huynh Dinh, T. Synthesis of acyl pyrophosphates. Application to the synthesis of nucleotide lipophilic prodrugs. *Tetrahedron Lett.* **1995,** 36, 531-534.

92. Bonnaffé, D.; Dupraz, B.; Ughetto-Monfrin, J.; Namane, A.; Henin, Y.; Huynh Dinh, T. Potential Lipophilic Nucleotide Prodrugs: Synthesis, Hydrolysis, and Antiretroviral Activity of AZT and d4T Acyl Nucleotides. *J. Org. Chem.* **1996,** 61, 895-902.

93. Kreimeyer, A.; André, F.; Gouyette, C.; Huynh-Dinh, T. Transmembrane Transport of Adenosine 5'-Triphosphate Using a Lipophilic Cholesteryl Derivative. *Angew Chem Int Ed Engl* **1998,** 37, 2853-2855.

94. Camarasa, M. J. Prodrugs of Nucleoside Triphosphates as a Sound and Challenging Approach: A Pioneering Work That Opens a New Era in the Direct Intracellular Delivery of Nucleoside Triphosphates. *Chemmedchem* **2018,** 13, 1885-1889.

95. Gollnest, T.; Dinis de Oliveira, T.; Schols, D.; Balzarini, J.; Meier, C. Lipophilic prodrugs of nucleoside triphosphates as biochemical probes and potential antivirals. *Nat Commun* **2015,** 6, 8716.

96. Gollnest, T.; Dinis de Oliveira, T.; Rath, A.; Hauber, I.; Schols, D.; Balzarini, J.; Meier, C. Membrane-permeable Triphosphate Prodrugs of Nucleoside Analogues. *Angew Chem Int Ed Engl* **2016,** 55, 5255-8.

97. Meier, C. Nucleoside diphosphate and triphosphate prodrugs - An unsolvable task? *Antivir Chem Chemother* **2017,** 25, 69-82.

98. Jia, X.; Schols, D.; Meier, C. Anti-HIV-Active Nucleoside Triphosphate Prodrugs. *J Med Chem* **2020,** 63, 6003-6027.

99. Jia, X.; Schols, D.; Meier, C. Lipophilic Triphosphate Prodrugs of Various Nucleoside Analogues. *J Med Chem* **2020,** 63, 6991-7007.

100. Zhao, C.; Weber, S.; Schols, D.; Balzarini, J.; Meier, C. Prodrugs of gamma-Alkyl-Modified Nucleoside Triphosphates: Improved Inhibition of HIV Reverse Transcriptase. *Angew Chem Int Ed Engl* **2020,** 59, 22063-22071.

101. Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **1992,** 256, 1783-1790.

102. Divita, G.; Rittinger, K.; Geourjon, C.; Deleage, G. Dimerization Kinetics of HIV-1 and HIV-2 Reverse Transcriptase: A Two Step Process. *J Mol Biol* **1995,** 245, 508-521.

103. Larder, B.; Purifoy, D.; Powell, K.; Darby, G. Site-specific mutagenesis of AIDS virus reverse transcriptase. *Nature* **1987,** 327, 716-717.

104. Poch, O.; Sauvaget, I.; Delarue, M.; Tordo, N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **1989,** 8, 3867-3874.

105. Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **1998,** 282, 1669-1675.

106. Martín-Hernández, A. M.; Domingo, E.; Menéndez-Arias, L. Human immunodeficiency virus type 1 reverse transcriptase: role of Tyr115 in deoxynucleotide binding and misinsertion fidelity of DNA synthesis. *EMBO J.* **1996,** 15, 4434-4442.

107. Gill, M. S. A.; Hassan, S. S.; Ahemad, N. Evolution of HIV-1 reverse transcriptase and integrase dual inhibitors: Recent advances and developments. *European Journal of Medicinal Chemistry* **2019,** 179, 423-448.

108. Kellinger, M. W.; Johnson, K. A. Nucleotide-dependent conformational change governs specificity and analog discrimination by HIV reverse transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** 107, 7734-7739.

109. Kati, W. M.; Johnson, K. A.; Jerva, L. F.; Anderson, K. S. Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **1992,** 267, 25988-25997.

110. Steitz, T. A. A mechanism for all polymerases. *Nature* **1998,** 391, 231-232.

111. Sarafianos, S. G.; Marchand, B.; Das, K.; Himmel, D. M.; Parniak, M. A.; Hughes, S. H.; Arnold, E. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* **2009,** 385, 693-713.

112. Sarafianos, S. G.; Das, K.; Clark Jr., A. D.; Ding, J.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with β-branched amino acids. *PNAS* **1999,** 96, 10027-10032.

113. Gao, H. Q.; Boyer, P. L.; Sarafianos, S. G.; Arnold, E.; Hughes, S. H. The role of steric hindrance in 3TC resistance of human immunodeficiency virus type-1 reverse transcriptase. *J Mol Biol* **2000,** 300, 403-418.

114. Meyer, P. R.; Matsuura, S. E.; So, A. G.; Scott, W. A. Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** 95, 13471-13476.

115. Larder, B. A.; Kemp, S. D. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* **1989,** 246, 1155-1158.

116. Schönland, S. O.; Lopez, C.; Widmann, T.; Zimmer, J.; Bryl, E. Premature telomeric loss in rheumatoid arthritis is genetically determined and involves both myeloid and lymphoid cell lineages. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** 100, 13471-13476.

117. Meyer, P. R.; Matsuura, S. E.; Mian, A. M.; So, A. G.; Scott, W. A. A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. *Mol. Cell* **1999,** 4, 35-43.

118. Mehellou, Y.; De Clercq, E. Twenty-Six Years of Anti-HIV Drug Discovery: Where Do We Stand and Where Do We Go? *J. Med. Chem.* **2010,** 53, 521-538.

119. Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K. Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* **1995,** 2, 407-415

120. Das, K.; Arnold, E. HIV-1 reverse transcriptase and antiviral drug resistance. Part 2. *Curr Opin Virol* **2013,** 3, 119-28.

121. Das, K.; Clark Jr, A. D.; Lewi, P. J.; Heeres, J.; De Jonge, M. R. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *J. Med. Chem.* **2004,** 47, 2550-2560.

122. Pribut, N.; Basson, A. E.; van Otterlo, W. A. L.; Liotta, D. C.; Pelly, S. C. Aryl Substituted Benzimidazolones as Potent HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors. *Acs Med Chem Lett* **2019,** 10, 196-202.

123. Esposito, F.; Corona, A.; Tramontano, E. HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions. *Mol Biol Int* **2012,** 2012, 586401.

124. Maga, G.; Radi, M.; Gerard, M.; Botta, M.; Ennifar, E. HIV-1 RT Inhibitors with a Novel Mechanism of Action: NNRTIs that Compete with the Nucleotide Substrate. *Viruses* **2010,** 2, 880-899.

125. Michailidis, E.; Marchand, B.; Kodama, E. N.; Singh, K. Mechanism of Inhibition of HIV-1 Reverse Transcriptase by 4'-Ethynyl-2-fluoro-2'-deoxyadenosine Triphosphate, a Translocation-defective Reverse Transcriptase Inhibitor. *J. Biol. Chem.* **2009,** 284, 35681-35691.

126. Michailidis, E.; Huber, A. D.; Ryan, E. M.; Ong, Y. T. 4'-Ethynyl-2-fluoro-2'-deoxyadenosine (EFdA) inhibits HIV-1 reverse transcriptase with multiple mechanisms. *J Biol Chem* **2014,** 289, 24533-24548.

127. Kawamoto, A.; Kodama, E.; Sarafianos, S. G.; Sakagami, Y. 2'-Deoxy-4'-C-ethynyl-2-halo-adenosines active against drug-resistant human immunodeficiency virus type 1 variants. *International Journal of Biochemistry and Cell Biology* **2008,** 40, 2410-2420.

128. Jochmans, D.; Deval, J.; Kesteleyn, B.; Van Marck, H. Indolopyridones Inhibit Human Immunodeficiency Virus Reverse Transcriptase with a Novel Mechanism of Action. *J. Virol.* **2006,** 80, 12283-12292.

129. Zhang, Z.; Walker, M.; Xu, W.; Hoon Shim, J. Novel Nonnucleoside Inhibitors That Select Nucleoside Inhibitor Resistance Mutations in Human Immunodeficiency Virus Type 1 Reverse Transcriptase. *Antimicrob. Agents Chemother.* **2006,** 50, 2772-2781.

130. Ehteshami, M.; Scarth, B. J.; Tchesnokov, E. P.; Dash, C. Mutations M184V and Y115F in HIV-1 Reverse Transcriptase Discriminate against "Nucleotide-competing Reverse Transcriptase Inhibitors". *J. Biol. Chem.* **2008,** 283, 29904-29911.

131. Maga, G.; Radi, M.; Zanoli, S.; Manetti, F.; Cancio, R.; Hubscher, U.; Spadari, S.; Falciani, C.; Terrazas, M.; Vilarrasa, J.; Botta, M. Discovery of non-nucleoside inhibitors of HIV-1 reverse transcriptase competing with the nucleotide substrate. *Angew Chem Int Edit* **2007,** 46, 1810-1813.

132. Radi, M.; Falciani, C.; Contemori, L.; Petricci, E. A Multidisciplinary Approach for the Identification of Novel HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors: S-DABOCs and DAVPs. *Chemmedchem* **2008,** 3, 573-593.

133. Freisz, S.; Bec, G.; Radi, M.; Wolff, P. Crystal Structure of HIV-1 Reverse Transcriptase Bound to a Non-Nucleoside Inhibitor with a Novel Mechanism of Action. *Angewandte Chemie* **2010,** 49, 1805-1808.

134. Öberg, B. Antiviral effects of phosphonoformate (PFA, foscarnet sodium). *Pharmacol Therapeut* **1989,** 40, 213-285.

135. Canestri, A.; Ghosn, J.; Wirden, M.; Marguet, F. Foscarnet salvage therapy for patients with late-stage HIV disease and multiple drug resistance. *Antivir.Ther.* **2006,** 11, 561-566.

136. Meyer, P. R.; Rutvisuttinunt, W.; Matsuura, S. E.; So, A. G.; Scott, W. A. Stable complexes formed by HIV-1 reverse transcriptase at distinct positions on the primer-template controlled by binding deoxynucleoside triphosphates or foscarnet. *J. Mol. Biol.* **2007,** 369, 41-54.

137. Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci.* **2019,** 20.

138. Ghosh, A. K.; Gemma, S. *Structure-based Design of Drugs and Other Bioactive Molecules: Tools and Strategies*. John Wiley & Sons: **2014**; p 474.

139. Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998,** 27, 249-284.

140. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **2010,** 31, 455-461.

Chapter 2

dAMP/Tenofovir-amino acid conjugates act as polymerase substrates: Implications for avoiding cellular phosphorylation in the discovery of nucleotide analogs

2.1 Introduction

Nucleoside and nucleotide inhibitors are among the leading drugs used for the treatment of HIV-infected patients.[1] AZT is the first HIV drug approved in 1987, and currently, the nucleoside analogs ABC, FTC, 3TC, and AZT and nucleotide analog TFV are active components (**Figure 2-1**) of the drug cocktails used in treating HIV infection under anti-retroviral therapy (ART).

AZT **ABC** **FTC** : R = F / **3TC** : R = H **Tenofovir**

Parent drug of **sofosbuvir** Parent drug of **remdesivir** **T-α-CNP**
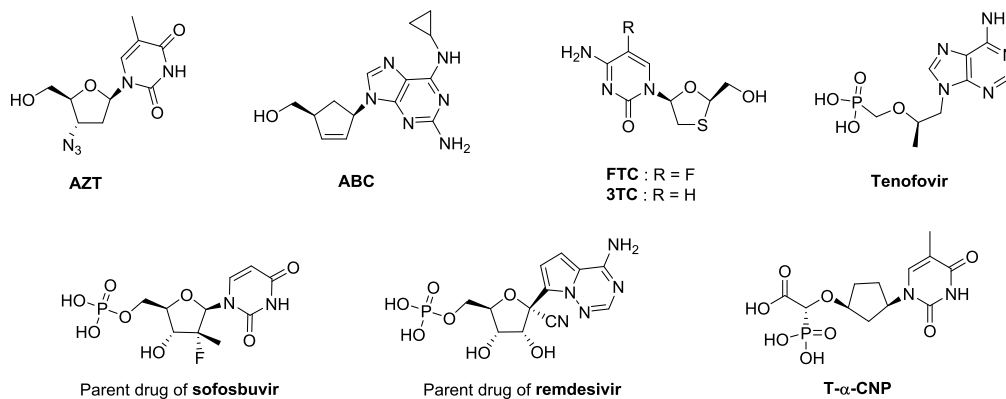
Figure 2-1. Nucleos(t)ide analogs

Ribonucleotide analogs are used for treating infections caused by the viruses which use RNA dependent RNA polymerases (RdRp) for replication/transcription of viral genes-such as HCV, influenza, and SARS-CoV-2. A nucleoside analog requires three-step phosphorylation, and a nucleotide such as TFV, the HCV drug sofosbuvir (**Figure 2-1**), or the SARS-CoV-2 drug remdesivir requires the final two phosphorylation steps; a nucleotide analog is generally delivered as a prodrug. Several NTP/dNTP analogs have been synthesized and have demonstrated efficacy *in vitro*.[2-8] However, cellular phosphorylation is a key hurdle in the successful development of most nucleotide-analogs as drugs. Chemical substitutions that can convert a nucleotide into an NTP/dNTP mimic have broad significance in designing polymerase inhibitors.

The transcription and replication of viral genes are carried out by RNA or DNA polymerases, and the polymerization reaction is accomplished by a conserved catalytic mechanism analogous to that described for *E. coli* DNA polymerase I.[9] Polymerase active sites contain highly conserved catalytic aspartate or glutamate residues that chelate two Mg^{2+} ions (**Figure 2-2**). The Mg^{2+} ion occupying site A deprotonates the primer 3'-hydroxyl, and the other Mg^{2+} ion at site B positions the triphosphate group appropriately for catalysis by chelating all three phosphates. The nucleophilic attack by 3'-OH at the α-phosphate of incoming dNTP forms a phosphodiester bond between the two groups, and β-γ pyrophosphate is released as a byproduct. Activation of a nucleophile by Mg^{2+} can also be found in other enzymes that cleave the phosphodiester bonds in a DNA backbone, such as some nucleases.[10] The chelation of the triphosphate moiety to the catalytic metal ions at the DNA and

RNA polymerase active sites can be exploited for drug design analogous to the design of HIV integrase inhibitors from the diketo acid scaffold.[11]
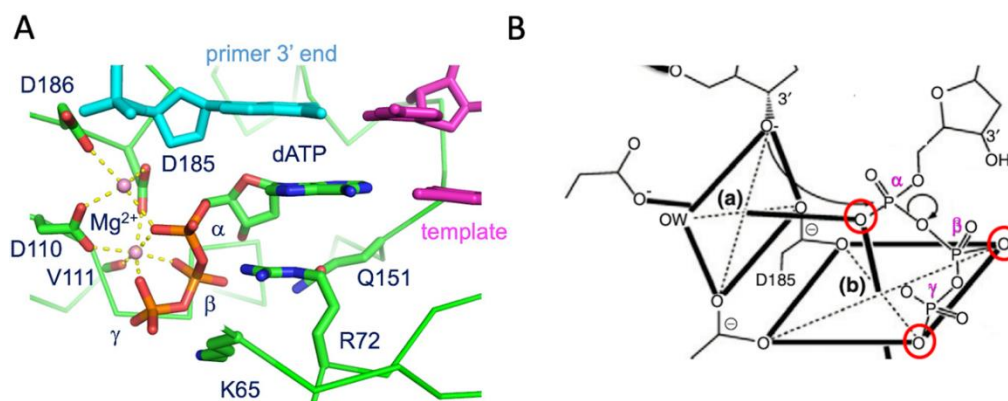
Figure 2-2. Metal chelation at the active site of HIV-1 RT. (A) Octahedral coordination of Mg^{2+} ions in HIV-1 RT/DNA catalytic complex (PDB code 4PQU); the primer strand was terminated with dideoxyguanosine to prevent incorporation. (B) A schematic representation of Mg^{2+} ion chelation in the polymerase active site of HIV-1 RT.

The enzyme RT that copies HIV-1 genomic RNA to a dsDNA at the early stage of the viral infection has been well studied biochemically and structurally. RT has been the target of almost half of the approved antiviral drugs for treating HIV infection. The nonnucleoside RT inhibitors (NNRTIs) allosterically inhibit the DNA polymerization. The nucleoside/nucleotide RT inhibitors (NRTIs) are incorporated by RT to the growing primer strand and act as chain terminators (**Figure 2-1**). The polymerase active site of HIV-1 RT is also an attractive target for designing dNTP-competitive inhibitors. Small molecule inhibitors such as INDOPY-1[12-13] and α-carboxy nucleoside phosphonates (α-CNPs)[14-20] have been discovered as nucleoside-competing RT inhibitors (NcRTIs). INDOPY-1 binding does not involve metal chelation[12], whereas α-CNPs chelate the active-site metal ions[18, 21-22]. A series of α-CNPs, such as T-α-CNP, (**Figure 2-1**) were synthesized as NcRTIs[14-20]. α-CNPs keep the natural nucleobases, but the 2'-deoxyribose moiety is substituted with a cyclopentyl ring. On this latter ring, a 3' oxygen is linked to the α-methyl carbon, which is linked to a carboxylic acid and a phosphonate group. Both groups chelate catalytic Mg^{2+} ions in the polymerase active site, and the chelation of α-CNP mimics that of the triphosphate moiety of a dNTP.

The alanine substitution has been used in the prodrug formulation of TFV, sofosbuvir, and remdesivir. In contrast, the amino-acid substitutions in the current study are introduced as the pyrophosphate analogs that would act as the leaving groups in the nucleic acid polymerization reaction. Our laboratory previously synthesized phosphoramidate conjugates of 2'-deoxynucleotides and natural amino acids (e.g., L-aspartic acid, L-histidine),[23-25] unnatural amino acids (e.g. D-aspartic acid, IDA, IDP and 3-phosphono-L-alanine),[26-28] and dipeptides.[29] It was demonstrated that several congeners within this family (e.g. 5'-L-aspartic- and 5'-L-histidinyl-phosphoramidate analogs of 2'-

dAMP) were active in a primer extension assay. For example, L-Asp-dAMP was able to mimic the natural nucleoside triphosphate dATP, and the 2'-deoxy-adenosine part is incorporated into a growing DNA strand by HIV-1 RT. However, for L-Asp-dAMP, the catalytic efficiency (V_{max}/K_m) is decreased 1300-fold compared to a dATP primarily due to an increase in K_m.[23] Only one example of a nucleotide analog conjugated to an amino acid has been described. L-Asp-PMEA (PhosphonoMethoxyEthylAdenine) has been shown to act as a chain terminator in a primer extension assay.[24] As a general principle, the conjugates can convert a nucleoside (or analog) into an active form for direct incorporation by RT or other polymerases, without needing the metabolic conversion. However, up to now, there is no structural information available for visualizing how these alternative leaving groups of the conjugates bind and what drives the catalytic incorporation at the polymerase active site.

Our current study follows the earlier chemical synthesis work to generate a series of conjugates of the amino acids L-Asp, D-Asp, L-His, L-Glu, L-Met, L-Arg and L-Phe with the nucleotide dAMP and TFV (**Scheme 2-1**). All compounds were tested for catalytic incorporation by RT and for RT inhibition. Only **2.3a** demonstrated multiple nucleotide incorporation among the dAMP conjugates, whereas several TFV conjugates demonstrated incorporation of TFV to the primer strand as a chain terminator. Crystal structures of TFV conjugates were determined in complex with HIV-1 RT/dsDNA in the catalytic competent state. The structures were analyzed in the context of their binding and modes of inhibition.

2.2 Results

2.2.1 Chemistry

The synthesis of the amino acid conjugates of dAMP and TFV was achieved according to a previously reported procedure[30] (**Scheme 2-1**). Briefly, a *N,N'*-dicyclohexylcarbodiimide (DCC) mediated coupling of an appropriate amino acid ester with dAMP (compound **2.1**) or TFV (compound **2.2**), followed by alkaline deprotection of the ester moiety furnished the desired dAMP amino acid conjugates **2.3a-g**, as well as the TFV conjugates **2.4a-g**. The following amino acids were selected for coupling : L-aspartic acid (series **a**), D-aspartic acid (series **b**), L-histidine (series **c**), L-glutamic acid (series **d**), L-methionine (series **e**), L-arginine (series **f**) and L-phenylalanine (series **g**).
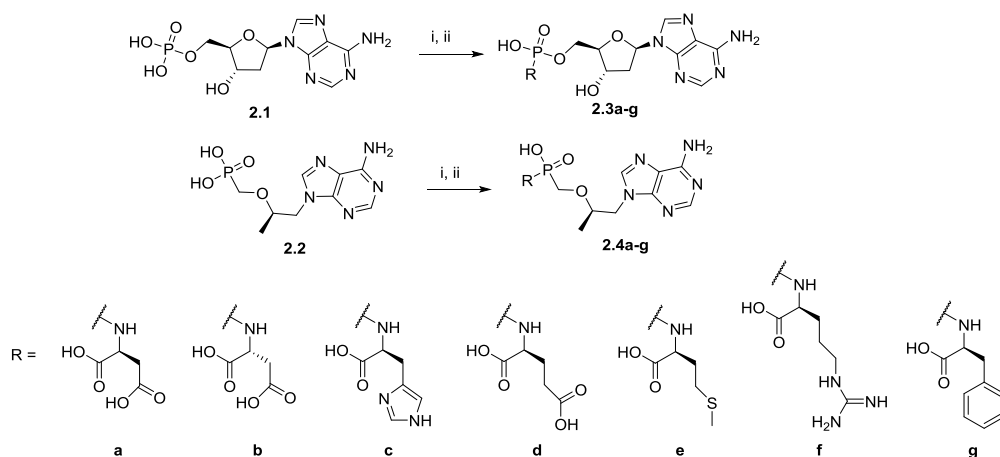
Scheme 2-1. *Reagents and conditions*. (i) amino acid ester, DCC, *t*-BuOH, H_2O, reflux, 3 h; (ii) 0.4 M NaOH, MeOH, r.t., 3 h.

2.2.2 Primer extension assay by HIV-1 RT (performed by Dr. Hoai Nguyen)

The ability of HIV-1 RT to incorporate the amino acid conjugates **2.3e-g** and **2.4a-g** into a growing DNA strand was investigated by a gel-based nucleotide incorporation assay with primer (P) and template (T) (**Figure 2-3**A-E). The primer-template duplex has 7 overhanging dT nucleotides at the 5' end of the template, and hence, this template/primer combination allows seven primer extensions. **2.3a**-**d** have been evaluated earlier in this assay, and it was demonstrated that **2.3a** was best at incorporation among dAMP analogs.[23-24] The V_{max} (rate attained when substrate concentration is much higher than K_m) of **2.3a** was only one-third of that of the natural substrate dATP, but the specificity of incorporation (V_{max}/K_m) was 1,300 times lower compared to dATP.[23] **2.3c** was close behind at 1.5 times less efficient than **2.3a**. The remaining amino acids (D-aspartic acid: **2.3b**; L-glutamic acid: **2.3d**) behaved significantly worse as substrates for RT. This indicates that the stereochemistry and the chemical composition of the amino acids are important for the amino acid-dAMP conjugates to function as substrates of HIV-1 RT. **2.3e**, **2.3f**, and **2.3g** (carrying a L-methionine, L-arginine and L-phenylalanine side chain, respectively, instead of the pyrophosphate moiety) are novel and were assessed in the current study for their ability to be incorporated in a growing DNA strand, catalyzed by HIV-1 RT. **2.3a** was selected as a reference compound, whereas dATP was included as the positive control. **2.3a** was tested at a concentration of 500 μM with 3 mM $MgCl_2$ as the metal ion source. After 1 hour of polymerase reaction, most of the primer extended to 2 or 3 bases, which is in agreement with previous data.[24] None of the novel amino acid dAMP conjugates (**2.3e**-**g**) showed noticeable incorporation when tested at concentrations of 500 μM (**Figure 2-3**A) up to 1 mM (**Figure 2-3**E), either in the presence of Mg^{2+}, or with added Mn^{2+} ions (**Figure 2-3**A-E).

The TFV conjugates with L-amino acids (**2.4a-g**) were also evaluated in the incorporation assay. When tested at a concentration of 500 μM (**Figure 2**-3B-C), P+1 incorporation was observed

for **2.4b** and **2.4d**, especially **2.4d** leading to substantially more incorporation (26.81% of P to P+1 conversion, **Table 2-1**) compared to the remaining conjugate analogs. Unlike dAMP, TFV acts as a chain terminator, and therefore, no additional nucleotide addition was observed.
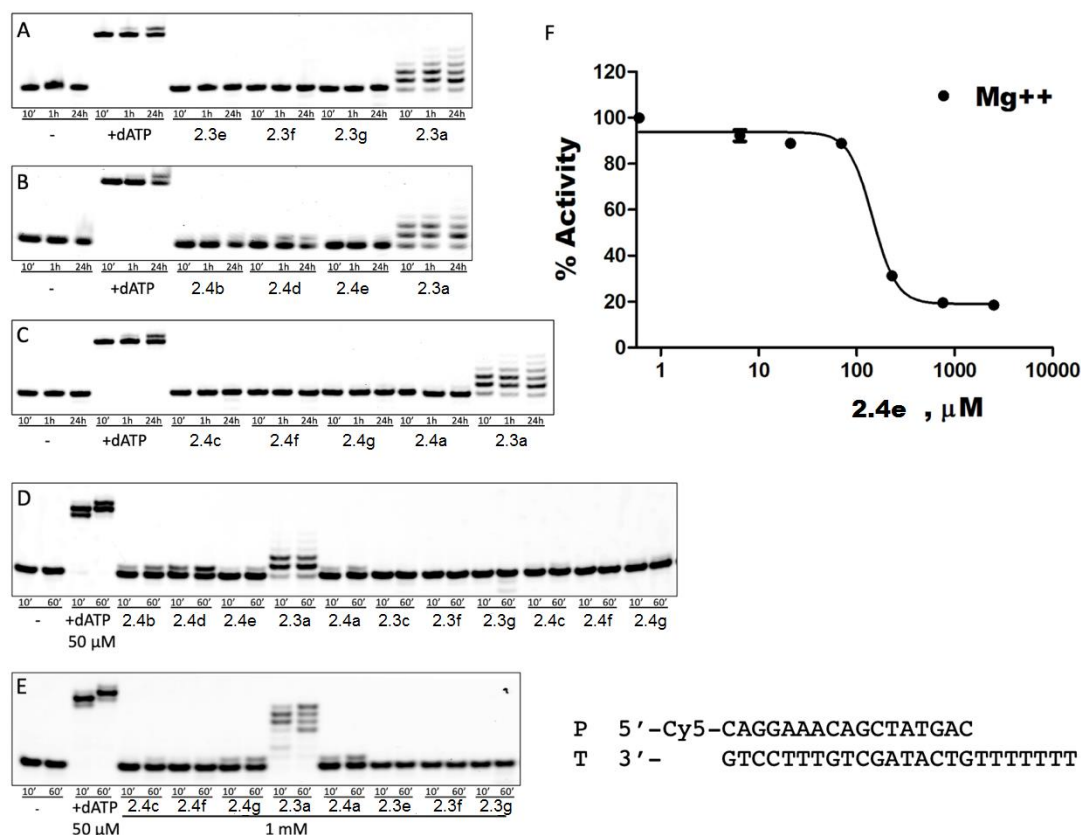
Figure 2-3. Primer extension and polymerase inhibition assays. (A,B,C) Primer extension with amino acid conjugates of 2'-dAMP and TFV by HIV-1 RT. Reaction conditions : 125 nM primer/template ; [dATP] = 50 μM ; [amino acid conjugate] = 500 μM ; [HIV-1 RT] = 0.5 μg/μL, 3 mM MgCl$_2$. (D,E) With 2 mM MnCl$_2$ added. In (E), [amino acid conjugate] = 1 mM. (F) Dose-response curve showing inhibition of HIV-1 RT in the presence of Mg^{2+} ion. The inhibitory effect of **2.4e** on dsDNA synthesis by recombinant HIV-1 RT was determined by PicoGreen-based spectrophotometric assay. Relative levels of dsDNA synthesis were calculated as a percentage of levels in reaction in the absence of compound. Data points represent the means of representative experiments run in triplicate.

The presence of 2 mM MnCl$_2$ improved the incorporation for **2.4a-b** and **2.4d-e** (**Figure 2-3**D) with the percentages of P to P+1 conversion of **2.4d** being the highest (40.78 % after 1 hour of reaction). When the concentrations of **2.4a**, **2.4c** and **2.4g** were increased from 500 μM to 1 mM, a small amount of P+1 product was also observed (**Figure 2-3**E), whereas for **2.4f**, no incorporation was observed. Overall, among the TFV conjugates, **2.4d** was the most efficiently incorporated, either with Mg^{2+} or in combination with Mn^{2+}.

Table 2-1. Summary of incorporation and inhibition data.

#	Chemical structure	Incorporation of compound in RT assay		IC$_{50}$ (μM)**
		Mg^{2+}	Mn^{2+}/Mg	
2.3a	L-Asp-dAMP	+++++	+++++	none
2.3b	D-Asp-dAMP	N/D	N/D	none
2.3c	L-His-dAMP	++++*	N/D	none
2.3d	L-Glu-dAMP	none*	N/D	none
2.3e	L-Met-dAMP	None	none	none
2.3f	L-Arg-dAMP	None	none	none
2.3g	L-Phe-dAMP	None	none	none
2.4a	L-Asp-TFV	None	++	none
2.4b	D-Asp-TFV	+	++	none
2.4c	L-His-TFV	None	+	none
2.4d	L-Glu-TFV	+	++	none
2.4e	L-Met-TFV	None	+	146
2.4f	L-Arg-TFV	None	none	none
2.4g	L-Phe-TFV	None	+	none

* Nucleic Acids Research, 35:5060-5072, 2007.   ****** at 2 mM dATP substrate. N/D = not done

2.2.3 Inhibition of HIV-1 RT (performed by Dr. Hongtao Xu)

The inhibitory effect of the compounds **2.3a-g** and **2.4a-g** on the dsDNA synthesis, catalyzed by recombinant HIV-1 RT, was determined by a PicoGreen-based spectrophotometric assay.[31] At the highest tested concentration of 2.5 mM, only **2.4e** demonstrated RT inhibition. Therefore, this compound was subjected to a full dose-response curve analysis using various concentrations of **2.4e** (**Figure 2-3**F). The IC$_{50}$ of **2.4e** was 146 μM in the presence of magnesium.

2.2.4 X-ray crystallography using RT cross-linked to DNA (performed by Dr. Sergio Martinez)

For this study, six crystal structures of **2.4b**, **2.4d**, and **2.4e** are determined in complex with HIV-1 RT/DNA either in the presence of MgCl$_2$, or MgCl$_2$ + MnCl$_2$. Crystals of RT cross-linked to a dsDNA template/primer were grown following a published protocol.[32] The crystals were soaked to form complexes with individual compounds. The diffraction data for soaked complexes were collected at synchrotron beamlines, and the structures were determined at resolutions ranging between 2.6 and 3.15 Å. The RT/DNA complex crystallized with P2$_1$ space group symmetry and contains two copies of RT/DNA in the asymmetric unit. This crystal form has been used to study several different classes of RT inhibitors, as well as resistance mutations: namely, the complexes with α-CNP,[21] dATP,[33] the allosteric inhibitor nevirapine[32-33] and the chain terminator AZT-TP before and after primer incorporation.[32] In general, the electron density for the inhibitor in one of the two copies is better than the other. Therefore, the copy with the better inhibitor density is used for the analysis and discussion unless otherwise stated. In general, RT can use either Mg^{2+} or Mn^{2+} ion for catalysis of

DNA polymerization; however, catalysis using Mg^{2+} ion requires high precision chelation groups and geometry. In contrast, Mn^{2+} ions are more forgiving. Therefore, we prepared the complexes of inhibitors in the presence of either Mg^{2+} alone, or Mn^{2+} and Mg^{2+} ions for a comprehensive understanding of the metal chelation and adaptivity of these compounds at the polymerase active site.

Structures of RT/DNA/L-Glu-TFV (compound 2.4d)

Our incorporation assay showed that the TFV moiety of **2.4d** is incorporated into the DNA primer by RT in the presence of either Mg^{2+} or Mn^{2+} ions. TFV incorporation from **2.4d** is the best of all TFV conjugates in this study (**Figure 2-3**B,D), and the incorporation is significantly enhanced with Mn^{2+}. RT/DNA crystals were soaked in the crystallization solution containing 1 mM **2.4d** for 1 hour; the crystallization solution had 5 mM $MgCl_2$. For obtaining the structure in the presence of Mn^{2+} ions, 2 mM $MnCl_2$ was added to the above soaking protocol. The crystal structures of RT/DNA/**2.4d** complex in the presence of Mg^{2+} or Mg^{2+} + Mn^{2+} ions were determined at 2.95 and 2.75 Å, respectively (**Figure 2-4** A,B). In the RT/DNA/**2.4d**/Mg^{2+} structure, the binding of **2.4d** at the polymerase active site of RT does not involve metal chelation, and in fact, no Mg^{2+} ion is present in the active site (**Figure 2-4**C). However, in the presence of Mn^{2+} ion, the compound chelates a metal ion at the active site in RT/DNA/**2.4d**/Mn^{2+} structure (**Figure 2-4**B).

Figure 2-4. (A) 2Fo - Fc electron density map at 2.95 Å resolution defines the binding of **2.4d** at the active site of the RT/DNA/**2.4d** complex in the presence of MgCl$_2$; PDB code 7AIG. No Mg^{2+} ion is present at the polymerase active site. (B) 2Fo - Fc density map at 2.75 Å resolution shows the binding of **2.4d** in the presence of Mn^{2+} (pink sphere); PDB code 7AIF. (C) **2.4d** binds at the dNTP-binding pocket. Template strand is magenta, and the primer 3' terminus is cyan. The primer strand is terminated with dideoxy-guanosine monophosphate. (D) An Mn^{2+} metal ion at site B chelates a phosphonoamidate oxygen of **2.4d**, catalytic residues D110 and D185, and the backbone carbonyl of V111. (E) Side view and (F) top view showing the binding of **2.4d** to RT/DNA complex either in the presence of Mg^{2+} (orange bonds) or added Mn^{2+} ions (green bonds). The Mn^{2+} binds at the site B. The chelation with Mn^{2+} causes better stacking of the adenine base with the DNA, brings the phosphonoamidate group closer towards the active site, and repositions the Glu sidechain. Structures were aligned in the palm domain (chain C) using the alpha carbons of residues 85-119, 151-242 and the CCP4 program SUPERPOSE.

In the structure that does not involve metal chelation, the adenine base of **2.4d** base pairs with the thymine of the first base overhang. However, the base pairing and base stacking are not optimal; adenine is tilted approximately 23° away from the plane of the thymine causing a loss of the stacking interactions with the dsDNA nucleotide guanine base at the primer terminus (**Figure 2-4**C). The sidechain of the conjugated L-Glu points toward the triphosphate-binding region of a dNTP (**Figure 2-4**A); however, L-Glu is less ordered due to lack of significant interaction primarily the lack of metal chelation. The backbone carboxyl group of the L-Glu moiety interacts with the conserved residue Arg72; both oxygens form 3.1 or 2.7 Å salt bridges with the Nε atom and the Nη2 atoms of Arg72. The Arg72 is a part of the fingers β3–β4 hairpin and Arg72 plays a critical role in the binding and positioning of dNTP at the polymerase active site.[34-35] In dNTP-bound structures, Arg72 side-chain stacks with the dNTP base and forms polar interaction with α-phosphate. In the current structures, however, the orientation of Arg72 is altered when compared to the dNTP-bound RT/DNA structures. The phosphonoamidate group forms a 3.2 Å H-bond with Nε2 atom of Gln151 (**Figure 2-4**C).

Despite weak binding of the compound at the active site and without Mg^{2+} ion chelation in the structure, **2.4d** demonstrates primer extension by RT (**Figure 2-3**B) in the presence of Mg^{2+} ions. However, the compound at 2 mM concentration did not show any noticeable RT inhibition in the PicoGreen assay. The inhibition data in agreement with the structures suggests significantly weak binding of **2.4d**, and dATP outcompetes **2.4d** in the inhibition assay.

The mode of binding of **2.4d** in the presence of Mn^{2+} ion has a significant difference from that in the presence of Mg^{2+}. The electron density reveals the presence of one Mn^{2+} ion positioned at the metal B site (**Figure 2-4**B) and chelates the catalytic aspartates Asp110 and Asp185, the backbone carbonyl of Val111, and a phosphonoamidate oxygen of **2.4d** (**Figure 2-4**D). This positional equivalent of the phosphonoamidate chelation mimics the chelation of the β–phosphate of TFV-DP.[36] For the TFV part in **2.4d**, the base pairing, base stacking, and the interaction of the acyclic linker are reminiscent of the binding of TFV-DP to RT.[36] In the structure of RT/DNA/**2.4d** in the presence of

Mn^{2+}, when compared with the structure without a metal ion at the active site, the phosphonate group is pulled towards the active site by ~1 Å to chelate the Mn^{2+} ion (**Figures 2-4**E,F). Consequently, the binding of the TFV moiety of **2.4d** is improved as the adenine base now stacks with the guanine base at the primer end. The conjugated glutamate sidechain and backbone carboxyl groups interact with the guanidinium group of the conserved residue Arg72, and the interactions improved the stability and electron density for the Glu sidechain. The side chain forms a 2.5 Å salt bridge with Nη2, and the backbone carboxylate forms a 3.0 Å interaction with Nε of Arg72. Comparison of both structures of **2.4d** shows that the Glu part is reorganized yet maintains the polar interactions with Arg72 using a different set of atoms. The structure and primer extension data suggest that binding of **2.4d** is improved in the presence of Mn^{2+} via metal chelation; however, the chelation is not optimal for catalytic incorporation. The suboptimal binding and loss of chelation of **2.4d** compared to dATP or TFV-DP are primarily responsible for the weak binding. Two different modes with suboptimal binding of **2.4d** support the conformational flexibility, which would help orient the compound favorably for catalytic incorporation by RT. **2.4d** is flexible and its two carboxyl groups and phosphonamidate are mobile in the structure. This flexibility aided with the additional flexibility of the acyclic TFV backbone helps **2.4d** attain the transient chelation events with Mg^{2+} for TFV incorporation − a structural explanation for the observed incorporation in our assay. Lacking the flexibility, **2.3d** did not show dAMP incorporation.[23-24]

Structures of RT/DNA/D-Asp-TFV (compound 2.4b)

The TFV moietyThe TFV moiety of **2.4b** is incorporated into the DNA primer by RT in the presence of either Mg^{2+} or Mn^{2+} ions; TFV incorporation is enhanced in the presence of Mn^{2+} ions. TFV incorporation from **2.4b**, however, is about half of that from **2.4d** (**Figure 2-3**B,D). In the same fashion as for **2.4d**, we soaked the crystals of HIV-1 RT/dsDNA with **2.4b** either in the presence of Mg^{2+} or Mn^{2+} + Mg^{2+}. The crystal structures of RT/DNA/**2.4b**/Mg^{2+} and RT/DNA/**2.4b**/Mn^{2+} complexes were obtained at 3.15 and 2.73 Å, respectively (**Figure 2-5**A,B). Analogous to the binding of **2.4d**, no metal chelation is involved in the binding of **2.4b** in the structure obtained only in the presence of Mg^{2+} ions (**Figure 2-5**C). In the second structure, an Mn^{2+} ion is chelating **2.4b** at the active site (**Figure 2-5**D).
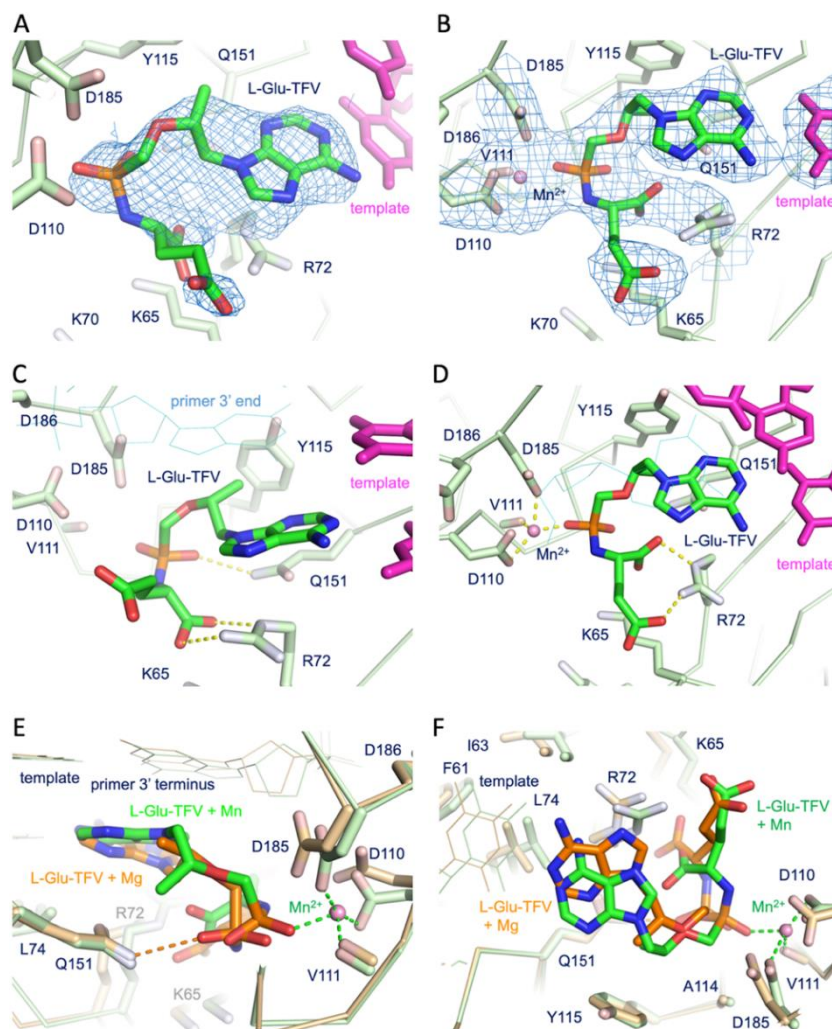
Figure 2-5. (A) 2Fo-Fc electron density map at 3.15 Å resolution showing the binding of **2.4b** at the active site of RT/DNA/**2.4b** complex in the presence of Mg^{2+}; PDB code 7AID. (B) 2Fo-Fc density map at 2.73 Å resolution showing the binding of **2.4b** in the presence of Mn^{2+} (pink sphere); PDB code 7AHX. (C) **2.4b** sits in the dNTP binding site in the presence of Mg^{2+}; no Mg^{2+}ion is present in the polymerase active site. Template strand is magenta, and the primer 3' terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (D) An Mn^{2+} metal ion at site B chelates a phosphonoamidate oxygen of **2.4b**, catalytic residues D110 and D185, and the backbone carbonyl of V111. (E) Side view and (F) Top view of **2.4b** in complex with RT in the presence of either Mg^{2+} (orange bonds) or $Mg^{2+} + Mn^{2+}$ (green bonds). The Mn^{2+} binds at the metal ion site B. The chelation with Mn^{2+} causes better stacking of the adenine base with the DNA, brings the phosphonoamidate group closer towards the active site, and repositions the D-Asp sidechain. Structures were aligned in the palm domain (chain C) using the alpha carbons of residues 85-119, 151-242 and the CCP4 program SUPERPOSE.

In either case, the adenine base-pairs with the thymine of the template first base overhang. However, the base pairing and base stacking are not optimal when **2.4b** is not involved in metal chelation, as is the case for **2.4d**. The adenine base of **2.4b** is tilted by approximately 19°, nearly the same amount for **2.4d**, away from the plane of the template thymine. The sidechain of the conjugated D-Asp points towards the triphosphate-binding site; however, unlike the triphosphate moiety of a dNTP, D-Asp is not involved in metal chelation and has poor density due to the absence of significant

interactions (**Figure 2-5**A,C). The backbone carboxyl group of the D-Asp moiety interacts with the conserved residue Arg72. One oxygen forms a 2.7 Å salt bridge with the Nε atom of Arg72. The phosphonoamidate group of **2.4b**, unlike that of **2.4d**, is not involved in any polar interaction.

In the presence of Mg^{2+}, **2.4b** demonstrated primer extension by RT despite the lack of Mg^{2+} ion chelation in our structure; however, the efficiency of primer extension from **2.4b** is less than that from **2.4d** (**Figure 2-3**B). The modes of binding of **2.4b**, which is similar to that of **2.4d**, could explain the TFV incorporation from **2.4b** by a transient chelation event with Mg^{2+}, which is again aided by two carboxyl groups and the flexible acyclic TFV backbone.

2.4b also has a significantly different mode of binding in the presence of Mn^{2+} ions compared to Mg^{2+}, which is analogous to the binding of **2.4d**. For the TFV moiety of **2.4b**, the base pairing and base stacking is improved, and the acyclic linker is repositioned to orient TFV moiety binding in the way similar to that of TFV-DP to RT[36] (**Figure 2-5**D). The conjugated D-Asp sidechain and backbone interact with the guanidinium group of Arg72, and the interaction improved the stability and electron density for the D-Asp side chain (**Figure 2-5**A,B). As for **2.4d**, the two carboxyl groups of **2.4b** interact with the guanidinium of Arg72. The side chain forms a 2.6 Å salt bridge with Nη2, and the backbone carboxylate forms one of 2.4 Å to Nε. The phosphonoamidate group is moved about 0.8 Å closer towards the Mn^{2+} ion at metal site B for chelation compared to its position in a non-chelated state (**Figure 2-5**E,F). Consequently, both carboxyl groups of the D-Asp moiety now interact with the Arg72. The TFV incorporation is very similar for L- or D-Asp conjugate in the presence of Mn^{2+} ions (**Figure 2-3**D).The longer and more flexible sidechain of Glu in **2.4d** appears to be more adaptive for the carboxyl sidechain compared to that of Asp in **2.4b**, and is possibly better able to interact as a leaving group; Arg72 interacts with the pyrophosphate leaving group of a dNTP in the process of the polymerization reaction.[34-35] There is no RT inhibition at 2 mM of either compound.

Structure of RT/DNA/L-Met-TFV (compound 2.4e)

The incorporation of the TFV moiety from **2.4e** by RT is negligible either in the presence of Mg^{2+} or Mn^{2+} ions, when compared with TFV incorporation from **2.4d** or **2.4b** (**Figure 2-3**B,D). The crystal structures of RT/DNA/**2.4e** complexes in the absence and presence of Mn^{2+} were obtained at 2.95 and 2.62 Å, respectively (**Figure 2-6**A,B). Analogous to **2.4d**, the binding of **2.4e** at the polymerase active site does not involve metal chelation when the complex was formed in the presence of Mg^{2+} ion only (**Figure 2-6**C), whereas the complex in the presence of Mn^{2+} ion involves metal chelation at the active site (**Figure 2-6**D). With Mg^{2+}, there is much less tilting of the adenine base when compared with the binding of **2.4d** or **2.4b**. Adenine is tilted only by 9° out of the plane of its complementary thymine base of the template overhang. However, the adenine base still loses stacking interaction with the guanine base at the primer end analogous to that observed for **2.4d** and **2.4b** in

the presence of Mg^{2+} (**Figure 2-6**F). Unlike the previous conjugates with carboxyl-containing sidechains, a methionine can have hydrophobic interactions and cannot participate in metal chelation. The sidechain of the conjugated L-Met only has clear density for the sulfur atom, which is weakly interacting with Arg72; the distance between the L-Met methyl group and Nη2 of Arg72 is ~ 3.4 Å (**Figure 2-6**C,D). The backbone carboxyl group of the L-Met moiety interacts with the guanidinium group of Arg72, forming a 2.3 Å salt bridge with Nη2. Two phosphonoamidate oxygens form H-bonds – a 2.7 Å H-bond with the backbone amide of Ala114 and a 2.7 Å H-bond with Nε2 of Gln151, the latter is also observed for **2.4d**.

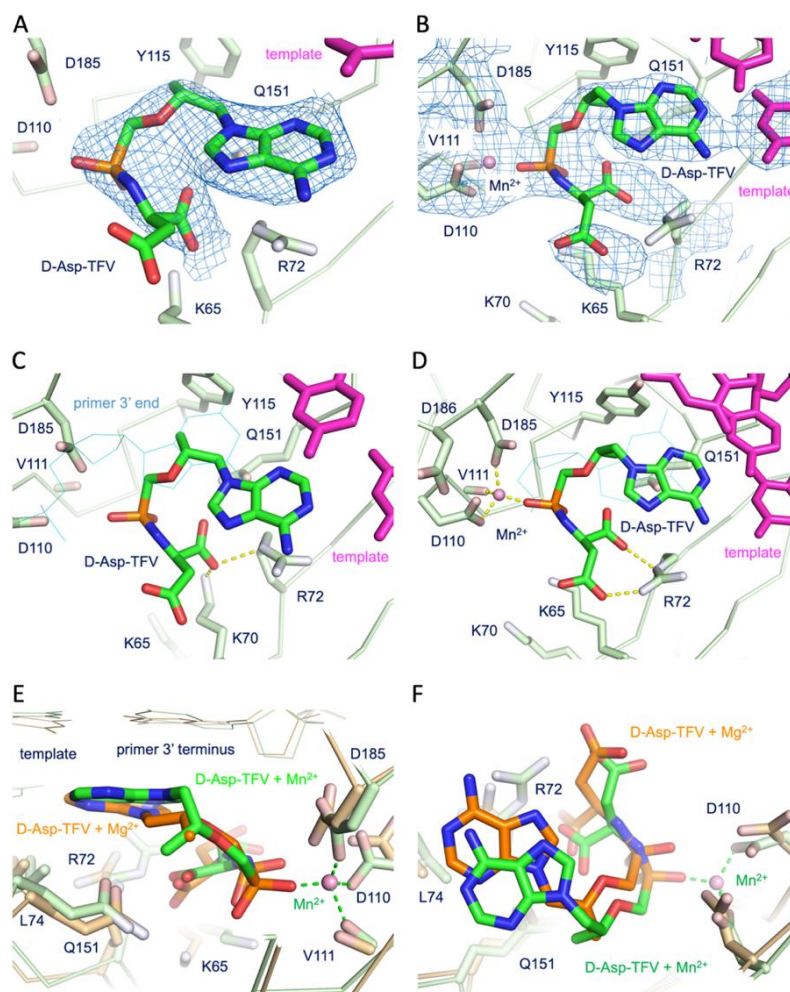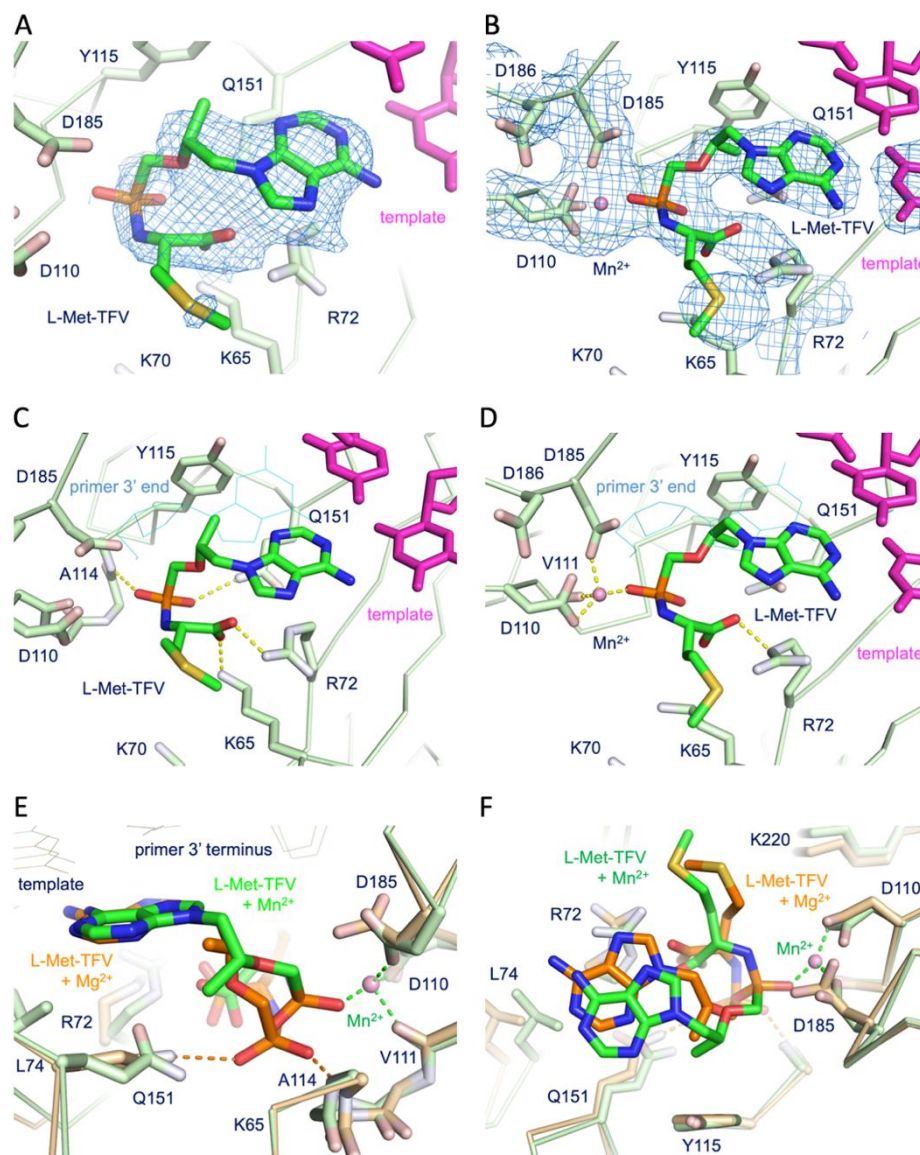**Figure 2**-**6.** (A) 2Fo-Fc electron density map at 3.15 Å resolution defines the position of **2.4e** in the active site of RT/DNA/**2.4e** complex in the presence of Mg^{2+} ion; PDB code 7AIJ. (B) 2Fo-Fc density map at 2.73 Å resolution defines the binding of **2.4e** in the presence of Mn^{2+} (pink sphere); PDB code 7AII. (C) **2.4e** sits in the dNTP-binding site in the presence of MgCl$_2$; no Mg^{2+} ion is present at the active site. Template strand is magenta, and the primer 3' terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (D) An Mn^{2+} metal ion at site B chelates a phosphonoamidate oxygen of **2.4e**, catalytic residues D110 and D185, and the backbone carbonyl of V111. (E) Side view and (F) Top view of **2.4e** in complex with RT in the presence of either Mg^{2+}

(orange bonds) or Mg^{2+}+ Mn^{2+} (green bonds). The chelation with Mn^{2+} causes better stacking of the adenine base with the DNA, brings the phosphonoamidate group closer towards the active site, and repositions the Met sidechain. Structures were aligned in the palm domain (chain C) using the alpha carbons of residues 85-119, 151-242 and the CCP4 program SUPERPOSE.

In the presence of Mg^{2+}, 500 μM **2.4e** demonstrated relatively little primer extension by RT (**Figure 2-3**B). The PicoGreen assay gave an IC_{50} of 146 μM with Mg^{2+}, the only noticeable inhibition by any of the compounds covered in this study (**Figure 2-3**F). **2.4e** appears to be a competitive inhibitor and not a good substrate for TFV incorporation.

2.4e also has a significantly different mode of binding in the presence of Mn^{2+} ions compared to Mg^{2+}. For the TFV moiety, the base pairing, base stacking, and interaction of the acyclic linker are also similar to those of **2.4b** and **2.4d**, and the binding of TFV-DP to RT[36] (**Figure 2-6**D). The sulfur and the Cε methyl group of L-Met sidechain are at a distance of ~3.4 Å from the Nη2 of Arg72 in both structures of **2.4e**. The backbone carboxyl forms a 2.7 Å salt bridge with Nε of Arg72. An indication of improved binding is that the β3-β4 fingers loop, has a better density in the Mn^{2+} structure; β3-β4 finger loop is closed down and better ordered in the RT structures with a bound dNTP. The electron density reveals the presence of one Mn^{2+} ion positioned at the metal B site and chelates the catalytic aspartates Asp110 and Asp185, the backbone carbonyl of Val111, and a phosphonoamidate oxygen of **2.4e** (**Figure 2-6**B). The phosphonoamidate group is moved about 1.9 Å, highest among all three compounds, closer towards the Mn^{2+} ion for chelation compared to its position in nonchelated state (**Figure 2-6**E,F). The key structural differences of **2.4e** from **2.4d** and **2.4b**, as discussed below, may be contributing to the better RT inhibition by **2.4e** compared to the remaining analogs. 2.**4e** has less tilting of its adenine than L-Glu or D-Asp conjugates, improving the base stacking. The phosphonoamidate group requires a large shift (1.9 Å) while going from free to the metal chelation state making the free state less favorable for catalysis. Finally, the phosphonoamidate group is more stabilized in its nonchelating state by forming two hydrogen bonds to the protein, compared to none for **2.4b** and one for **2.4d**. The TFV part of **2.4e** binds in an optimized way even in the absence of metal chelation which is contrary to that observed for the other two compounds. The structure shows a weak hydrophobic interaction of Met sidechain with the guanidinium group of Arg72. The H-bonds of **2.4b** and **2.4d** with Arg72 sidechain are solvent-exposed and can be replaced for dNTP binding with almost no energy cost. Apparently, a hydrophobic interaction of the guanidinium group of Arg72 may cost a small resistance to the adaptability of Arg72, which may contribute to the observed resistance by **2.4e**.

2.3 Discussion

Our results show the incorporation of TFV from TFV-amino acid conjugates by RT as a chain terminator, and different incorporation efficiency is observed for different amino acid conjugates.

The L-Glu conjugate that contains a flexible carboxyl sidechain provides the highest TFV incorporation, followed by L- and D-Asp conjugates. The amino acid conjugates Met that lacks the carboxyl side chain does not help TFV incorporation. It appears that the main-chain and side-chain carboxyl groups of Glu and Asp have the ability to chelate the metal ions at the active site and help to achieve catalytic incorporation of TFV. The structures show that the carboxyl groups form salt-bridges with the guanidinium side chain of Arg72. Arg72 is a highly conserved residue that plays an important role in the binding and positioning of dNTP for catalysis and pyrophosphate release.[34-35]

The crystal structures of TFV derivatives in complexes with RT show that the compounds are not bound in catalytic competent mode (**Figures 2-4 ~ 2-6**). However, structures indicate weaker and flexible binding of the compounds at the polymerase active site. We did not observe any Mg^{2+} ion involved in the binding of the inhibitors at the active site. Structures of the complexes in the presence of $MnCl_2$ reveal the presence of one Mn^{2+} ion that approximately takes the position of metal B, and the ion is chelating only one phosphonate oxygen. Compared to the binding mode of the compounds with no metal chelation, the phosphonoamidate groups of the compounds are shifted towards the active site for chelating the Mn^{2+} ion. Consequently, the binding mode of the TFV part is aligned with the binding mode of TFV of TFV-DP,[36] and the amino acid parts of our compounds are repositioned compared to the respective binding modes in absence of a metal ion.

The catalytic incorporation efficiency of the TFV from **2.4d** and **2.4b** is much less than that from TFV-DP. In the crystal structure, TFV-DP is bound in catalytic mode.[36] In contrast, the low incorporation suggests that the catalytic-relevant conformation of our compounds in the active site is a rather high-energy conformation, and our structures do not capture that conformation. Weaker binding of our compounds would support faster association (k_{on}) and dissociation (k_{off}) rates, which in combination with the internal flexibility appears to increase the probability of attaining the favorable conformation for catalytic incorporation of TFV. Despite no Mg^{2+} ion is present in the active site of the complexes in crystal structures, RT catalytically incorporates TFV from Asp/Glu conjugates suggesting that Mg^{2+} ions get involved in the dynamic process of binding of the compounds and catalytic incorporation of TFV. In fact, the bound Mn^{2+} ion present in the structures does not substantially enhance the rate of incorporation compared to Mg^{2+} ions because the observed mode of chelation with Mn^{2+} is not catalytically competent, and the binding mode would take conformational rearrangement for catalytic incorporation.

As discussed above, the compounds are weak binders compared to a dATP substrate and their H-bond interactions with Arg72 are solvent-exposed. Therefore, the compounds fail to compete with dATP binding, and consequently, we did not observe noticeable RT inhibition. One of the major differences in the modes of binding is that a dNTP binding involves three phosphate oxygens forming

four chelations with both catalytic metal ions (**Figure 2-2**B). In contrast, our compounds only have one chelation when Mn^{2+} ion is present. α-CNPs, a class of NcRTIs that show comparable metal chelation at the active site as dNTPs, have comparable RT binding affinity to that of dNTP. A crystal structure bound to RT/dsDNA shows that T-α-CNP binding replaces all four chelations of a dNTP with both Mg^{2+} ions at the active site.[22] The carboxyl group of T-α-CNP mimics the chelation of the α-phosphate oxygen of a dNTP, and two phosphonate oxygens mimic the chelation of β- and γ-phosphates. T-α-CNP in the L-enantiomer form was found to inhibit several viral polymerases, with the best IC_{50} of 0.41 μM for HIV-1 RT. The notion that octahedral coordination correlates with tighter inhibition is also supported by the study in which the phosphonate group of α-CNP is substituted with a carboxyl group. A 2.95 Å crystal structure in complex with HIV-1 RT/dsDNA, show weaker metal chelation.[18] This reduced chelation was reflected in a 50-fold higher IC_{50} to 20 μM for two carboxyl-containing compounds.

The residue Arg72 and the structural element $\beta3-\beta4$ finger loop that contains Arg72 are solvent-exposed and highly flexible. This flexibility is critical for dNTP binding. Thereby, the polar interactions between Arg72 and carboxyl groups of Glu/Asp conjugates are solvent-exposed and can be easily dissociated when a dNTP competes for binding, which could be a possible explanation for not observing noticeable inhibition by the compounds. We observed measurable but weak inhibition of polymerization only by **2.4e** that lacks a carboxyl side chain.

In comparison to TFV analogs, most dAMP-amino acid conjugates do not show incorporation. The increased flexibility of TFV linker compared to a cyclic deoxyribose ring of dAMP appears to play a role in facilitating catalytic incorporation. However, **2.3a** demonstrated multiple incorporations of dAMP molecules (**Figure 2-3**A) suggesting that the binding mode of **2.3a** is somewhat favorable for attaining the catalytic-competent mode for dAMP incorporation. The V_{max} of **2.3a** is 1/3 that of dATP, but the K_m is 403-fold higher.[23] Thus, the specificity of incorporation (V_{max}/K_m) is 1300-fold lower than for dATP. Given that the difference is almost entirely due to the higher K_m, the binding mode of **2.3a** is not optimal for catalytic incorporation by RT. The higher K_m values of our compounds also correlate with their non-productive modes of binding. No significant RT inhibition by any dAMP conjugate suggests that **2.3a** binding is also not optimum for the compound to act as a competitive inhibitor.

Our study provides the first structural perspective of nucleotide-amino acid conjugates functioning as chain terminators or competitive inhibitors. It appears that the binding of a conjugate in a catalytically competent mode is not energetically favorable and the process of association/dissociation of the molecules helps attain the chelating mode right for incorporation. We expect that the compounds with right substitutions will generate dNTP analogs for efficient

incorporation. Such studies on viral RNA polymerases may help in developing NTP analogs to avoid cellular phosphorylation as a relatively small number of nucleotide analogs are approved for the treatment of RNA viruses. Compared to the open and highly flexible dNTP-binding pocket of RT, the NTP binding pocket of viral RdRp, such as that of SARS-CoV-2,[37-38] are better formed, suggesting that the current approach may be valuable for designing RdRp inhibitors.

2.4 Conclusion

We have solved, at 2.6-3.15 Å, the structures of HIV-1 RT/DNA in complex with analogs of tenofovir diphosphate that have been modified by replacing the β and γ phosphate groups with an L- or D-amino acid conjugated through its amino group to the α-phosphate or phosphonate moiety. **2.4d** was found to work as intended, generating one primer extension in our assay. Metal chelation and electron density were seen for Mn^{2+}, but not for Mg^{2+}. However, the loss of chelation causes the phosphonoamidate group to move away from metal site B and generate new H-bond interactions with key residues or a backbone amide in the active site. A large variation in the modes of binding and chelation of the compounds provides opportunities for developing target specific inhibitors of various RNA and DNA polymerases. Structural understanding will help design metal chelators as improved (d)NTP-competing inhibitors or (d)NTP analogs that will be efficiently incorporated by a polymerase without requiring cellular phosphorylation. The broader implication of this study is to explore how the conjugates exhibit different modes of inhibition. Currently, the compounds are weakly active in the enzymatic (incorporation and inhibition) assays and, therefore, are not further profiled in antiviral assays. In addition, the chemical and metabolic stability of these types of compounds are problematic for further biological profiling, as previously shown by our group.[39] However, the structural understanding of the binding modes of these compounds will allow us to improve further their enzymatic potency.

2.5 Experimental section

Chemistry. All reagents and solvents were purchased from commercial sources and were used as obtained. Moisture-sensitive reactions were carried out in oven-dried glassware under a nitrogen or argon atmosphere. ^{1}H, ^{13}C, and ^{31}P NMR spectra were recorded on Bruker Avance 300 or 500 MHz spectrometer using tetramethylsilane as internal standard or referenced to the residual solvent signal, and 85 % H_3PO_4 for ^{31}P NMR experiments. Splitting patterns are designated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), br(broad), and dd (doublet of doublets). Coupling constants are expressed in hertz (Hz). High-resolution mass spectra (HRMS) were obtained on a quadruple orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 μL/min and spectra were obtained in positive (or negative)

ionization mode with a resolution of 15000 (fwhm) using leucine enkephalin as lock mass. Precoated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ= 254 nm). Products were purified by flash chromatography on silica gel (35-70 μm, 60 Å). All final compounds were purified by preparative RP-HPLC (C18 Phenomenex Gemini column, 110 Å, 10 μm, 21.2 mm × 250 mm) using an eluent gradient of CH_3CN with 50 mmol TEAB as eluent buffer. Final compounds were isolated as their triethylammonium salt. Prior to biological evaluation, compounds were converted into the corresponding sodium salts via sodium ion exchange resin. The exact concentration of the compounds was determined by the Lambert-Beer law. The purity of all the tested compounds was more than 95 %, as determined by HPLC analysis, using the analytical method mentioned before.

((((2R,3S,5R)-5-(6-Amino-9H-purin-9-yl)-3-hydroxytetrahydrofuran-2-yl)methoxy) (hydroxy)phosphoryl)-L-aspartic acid triethylammonium salt (2.3a) In a two-neck round bottom flask, 2'-deoxyadenosine 5'-monophosphate (100 mg, 0.30 mmol) and L-aspartic acid dimethyl ester hydrochloride (417 mg, 2.11 mmol) were dissolved in a mixture of *t*-BuOH and water (5:1, 12 mL). Et_3N (0.7 mL) was added, followed by the dropwise addition of a freshly prepared solution of DCC (626 mg, 3.02 mmol) in *t*-BuOH (1.2 mL). The reaction mixture was heated for 2~3 h at 85 °C while stirring under argon. The progress of the reaction was monitored by TLC (*i*-PrOH/H_2O/NH_3, v/v/v, 7:2:2). Upon completion of the reaction, the reaction mixture was cooled down to room temperature and the solvent was removed by rotary evaporation keeping the temperature of the water bath below 30 °C. The resulting crude residue was purified by silica gel flash chromatography, the mobile phase being a gradient of DCM/MeOH/H_2O (v/v/v) in a ratio gradually ranging from 5:1:0 to 5:2:0.25 and 5:3:0.25, affording the intermediate dimethylester as a white solid. This product was dissolved in 0.4 M NaOH in a MeOH/H_2O mixture (4:1, 10 mL) and stirred at room temperature under argon for 2~3 h. The reaction course was monitored by TLC (*i*-PrOH/H_2O/NH_3, v/v/v, 6:2:2) until disappearance of the starting material. The reaction mixture was neutralized by addition of triethylammonium acetate (1 M, pH ~ 7.5). The solvent was removed under reduced pressure, and the resulting residue was purified by silica gel flash chromatography, the mobile phase being a gradient of *i*-PrOH/H_2O/NH_3 in a ratio ranging from 10:0:0 to 10:1:1 and 7:1:1 (v/v/v). The product was finally purified by preparative HPLC with a gradient of CH_3CN in a 50 mM TEAB buffer (pH = 7.4) to yield the target compound as a colourless solid (71 mg, 53 % over two steps). ^{1}H NMR (300 MHz, D_2O) δ 8.46 (s, 1H), 8.18 (s, 1H), 6.46 (t, J = 6.9 Hz, 1H), 4.68 (s, 1H), 4.24 (s, 1H), 4.08 − 3.83 (m, 2H), 3.75 (dd, J = 13.6, 7.8 Hz, 1H), 2.87 − 2.68 (m, 1H), 2.64 − 2.28 (m, 3H). ^{13}C NMR (75 MHz, D_2O) δ 180.54 (d, J = 4.7 Hz), 178.60 (s), 155.21(s), 152.47 (s), 148.57 (s), 139.89 (s), 118.37 (s), 86.01 (d, J = 9.2 Hz), 83.44 (s), 71.15 (s), 63.82 (d, J = 5.1 Hz), 54.39 (s), 42.60 (d, J = 6.2 Hz), 38.91 (s).

^{31}P NMR (121 MHz, D$_2$O) δ 7.09 (s). HRMS: [M-H][-] calculated for C$_{14}$H$_{19}$N$_6$O$_9$P, 445.0878; found 445.0881.

(((((R)-1-(6-Amino-9H-purin-9-yl)propan-2-yl)oxy)methyl)(hydroxy)phosphoryl)-L-aspartic acid triethylammonium salt (2.4a) This compound was prepared from 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine monohydrate (100 mg, 0.33 mmol) and L-aspartic acid dimethyl ester hydrochloride (453 mg, 2.29 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (49 mg, 37 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.16 (s, 1H), 8.06 (s, 1H), 4.21 (ddd, *J* = 20.0, 14.9, 4.6 Hz, 2H), 3.86 (dd, *J* = 10.1, 5.6 Hz, 1H), 3.73 (dd, *J* = 13.4, 7.8 Hz, 1H), 3.48 (dt, *J* = 21.2, 12.6 Hz, 2H), 2.39 (ddd, *J* = 22.0, 14.3, 6.1 Hz, 2H), 1.01 (d, *J* = 6.1 Hz, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 181.45 (d, *J* = 2.9 Hz), 179.54 (s), 155.22 (s), 152.06 (s), 148.95 (s), 143.41 (s), 117.77 (s), 75.54 (d, *J* = 11.7 Hz), 66.19 (d, *J* = 145.5 Hz), 54.81 (s), 47.42 (s), 44.11 (d, *J* = 5.4 Hz), 16.01 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 17.41 (s). HRMS: [M-H][-] calculated for C$_{13}$H$_{19}$N$_6$O$_7$P, 401.0980; found 401.0981.

(((((2R,3S,5R)-5-(6-Amino-9H-purin-9-yl)-3-hydroxytetrahydrofuran-2-yl)methoxy) (hydroxy)phosphoryl)-D-aspartic acid triethylammonium salt (2.3b) This compound was prepared from 2'-deoxyadenosine 5'-monophosphate (100 mg, 0.30 mmol) and D-aspartic acid dimethyl ester hydrochloride (453 mg, 2.29 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (68 mg, 51 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.47 (s, 1H), 8.19 (s, 1H), 6.46 (t, *J* = 6.8 Hz, 1H), 4.70 (d, *J*=2.5 Hz, 1H), 4.31-4.14 (m, 1H), 3.93 (s, 2H), 3.78-3.70 (m, 1H), 2.86-2.70 (m, 1H), 2.63-2.33 (m, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 181.12 (d, *J* = 3.2 Hz), 179.36 (s), 155.27 (s), 152.47 (s), 148.45 (s), 139.84 (s), 118.30 (s), 85.97 (d, *J* = 9.2 Hz), 83.33 (s), 71.00 (s), 63.73 (d, *J* = 4.8 Hz), 55.05 (s), 43.50 (d, *J* = 7.1 Hz), 38.78 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 6.77 (s). HRMS: [M-H][-] calculated for C$_{14}$H$_{19}$N$_6$O$_9$P, 445.0878; found 445.0919.

(((((R)-1-(6-Amino-9H-purin-9-yl)propan-2-yl)oxy)methyl)(hydroxy)phosphoryl)-D-aspartic acid triethylammonium salt (2.4b) This compound was prepared from 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine monohydrate (100 mg, 0.33 mmol) and D-aspartic acid dimethyl ester hydrochloride (453 mg, 2.29 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (45 mg, 34 % over two steps) ^{1}H NMR (300 MHz, D$_2$O) δ 8.19 (s, 1H), 8.12 (s, 1H), 4.33-4.15 (m, 2H), 3.95-3.82 (m, 1H), 3.72 (dd, *J* = 13.8, 7.9 Hz, 1H), 3.60-3.36 (m, 2H), 2.45-2.22 (m, 2H), 1.04 (d, *J* = 6.3 Hz, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 181.27 (d, *J* = 3.1 Hz), 179.38 (s), 155.27 (s), 152.18 (s), 149.12 (s), 143.46 (s), 117.91 (s), 75.45 (d, *J* = 11.6 Hz), 66.06 (d, *J* = 144.0 Hz), 54.73 (s), 47.46 (s), 44.16 (d, *J* = 5.0 Hz), 15.88 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 17.34 (s). HRMS: [M-H][-] calculated for C$_{13}$H$_{19}$N$_6$O$_7$P, 401.0980; found 401.0971.

(((((2R,3S,5R)-5-(6-Amino-9H-purin-9-yl)-3-hydroxytetrahydrofuran-2-yl)methoxy) (hydroxy)phosphoryl)-L-histidine triethylammonium salt (2.3c) This compound was prepared from 2'-deoxyadenosine 5'-monophosphate (100 mg, 0.30 mmol) and L-histidine methyl ester dihydrochloride (511 mg, 2.11 mmol) according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (70 mg, 50 % over two steps). ^{1}H NMR (300 MHz, D_2O) δ 8.23 (s, 2H), 7.93 (s, 1H), 6.93 (s, 1H), 6.23 (t, J = 11.9 Hz, 1H), 4.58-4.54 (m, 1H), 4.13 (s, 1H), 3.84 (s, 2H), 3.69-3.62 (m, 1H), 2.81-2.63 (m, 3H), 2.52-2.44 (m, 1H). ^{13}C NMR (75 MHz, D_2O) δ 178.18 (d, J = 8.0 Hz), 154.68 (s), 152.09 (s), 147.99 (s), 139.22 (s), 132.30 (s), 129.21 (s), 117.76 (s), 116.16 (s), 85.45 (d, J = 9.1 Hz), 83.01 (s), 70.73 (s), 63.42 (d, J = 4.6 Hz), 55.72 (s), 38.24 (s), 29.31 (d, J = 2.8 Hz). ^{31}P NMR (121 MHz, D_2O) δ 6.41 (s). HRMS: [M-H]$^-$ calculated for $C_{16}H_{21}N_8O_7P$, 467.1198; found 467.1195.

((((((R)-1-(6-Amino-9H-purin-9-yl)propan-2-yl)oxy)methyl)(hydroxy)phosphoryl)-L-histidine triethylammonium salt (2.4c) This compound was prepared from 9-[(*R*)-2-(phosphonomethoxy) propyl]adenine monohydrate (100 mg, 0.33 mmol) and L-histidine methyl ester dihydrochloride (554 mg, 2.29 mmol) according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (60 mg, 43 % over two steps). ^{1}H NMR (300 MHz, D_2O) δ 8.18 (s, 1H), 8.13 (s, 1H), 7.53 (s, 1H), 6.73 (s, 1H), 4.32-4.17 (m, 2H), 3.96-3.91 (m, 1H), 3.79-3.72 (m, 1H), 3.53-3.32 (m, 2H), 2.77 (dd, J = 14.0 Hz, 7.9 Hz, 2H), 1.11 (d, J = 6.2 Hz, 3H). ^{13}C NMR (75 MHz, D_2O) δ 180.10 (d, J = 4.8 Hz), 154.96 (s), 151.83 (s), 148.66 (s), 142.86 (s), 134.93 (s), 119.90 (s), 119.79 (s), 117.62 (s), 75.07 (d, J = 12.4 Hz), 65.46 (d, J = 145.5 Hz), 56.18 (s), 47.39 (s), 31.72 (s), 15.57 (s). ^{31}P NMR (121 MHz, D_2O) δ 17.26 (s). HRMS: [M-H]$^-$ calculated for $C_{15}H_{21}N_8O_5P$, 423.1300; found 423.1298.

(((((2R,3S,5R)-5-(6-Amino-9H-purin-9-yl)-3-hydroxytetrahydrofuran-2-yl)methoxy) (hydroxy)phosphoryl)-L-glutamic acid triethylammonium salt (2.3d) This compound was prepared from 2'-deoxyadenosine 5'-monophosphate (100 mg, 0.30 mmol) and L-glutamic acid dimethyl ester hydrochloride (447 mg, 2.11 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (46 mg, 33 % over two steps). ^{1}H NMR (300 MHz, D_2O) δ 8.39 (s, 1H), 8.08 (s, 1H), 6.37 (t, J = 6.9 Hz, 1H), 4.59 (t, J=3.0 Hz, 1H), 4.16 (s, 1H), 3.92-3.79 (m, 2H), 3.42 (dd, J = 15.1, 6.5 Hz, 1H), 2.78-2.66 (m, 1H), 2.53-2.45 (m, 1H), 2.12 (t, J = 8.4 Hz, 2H), 1.87-1.69 (m, 2H). ^{13}C NMR (75 MHz, D_2O) δ 182.83 (s), 181.53 (d, J = 4.4 Hz), 155.83 (s), 152.86 (s), 148.86 (s), 140.15 (s), 118.81 (s), 86.30 (d, J = 9.3 Hz), 83.75 (s), 71.56 (s), 64.12 (d, J=4.9 Hz), 56.95 (s), 39.32 (s), 34.05 (s), 32.22 (d, J = 5.8 Hz). ^{31}P NMR (121 MHz, D_2O) δ 7.03 (s). HRMS: [M-H]$^-$ calculated for $C_{15}H_{21}N_6O_9P$, 459.1035; found 459.1042.

((((((R)-1-(6-Amino-9H-purin-9-yl)propan-2-yl)oxy)methyl)(hydroxy)phosphoryl)-L-glutamic acid triethylammonium salt (2.4d) This compound was prepared from 9-[(*R*)-2-

(phosphonomethoxy)propyl]adenine monohydrate (100 mg, 0.33 mmol) and L-glutamic acid dimethyl ester hydrochloride (485 mg, 2.29 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (43 mg, 31 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.23 (s, 1H), 8.19 (s, 1H), 4.35-4.19 (m, 2H), 3.95-3.87 (m, 1H), 3.59-3.39 (m, 3H), 2.17 (t, *J* = 8.3 Hz, 2H), 1.87-1.65 (m, 2H), 1.07 (d, *J* = 6.1 Hz, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 182.61 (s), 181.24 (d, *J* = 3.6 Hz), 155.20 (s), 152.08 (s), 148.96 (s), 143.30 (s), 117.76 (s), 75.39 (d, *J* = 11.7 Hz), 65.90 (d, *J* = 165.0 Hz), 56.32 (s), 47.44 (s), 33.69 (s), 32.36 (d, *J*=4.4 Hz), 15.83 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 17.42 (s). HRMS: [M-H]$^-$ calculated for C$_{14}$H$_{21}$N$_6$O$_7$P, 415.1136; found 415.1136.

((((2R,3S,5R)-5-(6-Amino-9H-purin-9-yl)-3-hydroxytetrahydrofuran-2-yl)methoxy) (hydroxy)phosphoryl)-L-methionine triethylammonium salt (2.3e) This compound was prepared from 2'-deoxyadenosine 5'-monophosphate (100 mg, 0.30 mmol) and L-methionine methyl ester hydrochloride (421 mg, 2.11 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (51 mg, 37 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.36 (s, 1H), 8.01 (s, 1H), 6.31 (t, *J* = 6.8 Hz, 1H), 4.59 (t, *J*=3.0 Hz, 1H), 4.13 (s, 1H), 3.83 (d, *J* = 3.3 Hz, 2H), 3.49-3.42 (m, 1H), 2.68 (d, *J* = 5.9 Hz, 1H), 2.51-2.44 (m, 1H), 2.26 (dd, *J* = 15.5, 7.6 Hz, 2H), 1.83 (s, 3H), 1.65-1.58 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 180.81 (d, *J* = 5.8 Hz), 155.50 (s), 152.72 (s), 148.73 (s), 140.02 (s), 118.60 (s), 86.26 (d, *J* = 9.4 Hz), 83.72 (s), 71.49 (s), 64.02 (d, *J* = 4.9 Hz), 56.47 (s), 39.35 (s), 34.86 (d, *J*=4.0 Hz), 29.38 (s), 14.20 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 6.73 (s). HRMS: [M-H]$^-$ calculated for C$_{15}$H$_{23}$N$_6$O$_7$PS, 461.1014; found 461.1019.

(((((R)-1-(6-Amino-9H-purin-9-yl)propan-2-yl)oxy)methyl)(hydroxy)phosphoryl)-L-methionine triethylammonium salt (2.4e) This compound was prepared from 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine monohydrate (100 mg, 0.33 mmol) and L-methionine methyl ester hydrochloride (457 mg, 2.29 mmol), according to the procedure used for the synthesis of compound **2.3a,** yielding the title compound as a colourless solid (42 mg, 30 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.11 (s, 1H), 8.01 (s, 1H), 4.27-4.08 (m, 2H), 3.83 (dd, *J*=9.7, 5.5 Hz, 1H), 3.62-3.41 (m, 2H), 3.34 (dd, *J* = 12.6, 9.5 Hz, 1H), 2.28 (t, *J* = 7.2 Hz, 2H), 1.89 (s, 3H), 1.75-1.50 (m, 2H), 0.99 (d, *J* = 6.2 Hz, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 180.87 (d, *J* = 4.6 Hz), 155.13 (s), 152.02 (s), 148.89 (s), 143.20 (s), 117.74 (s), 75.10 (d, *J* = 12.3 Hz), 65.77 (d, *J* = 144.75 Hz), 55.77 (s), 47.61 (s), 34.91 (d, *J*=3.3 Hz), 29.04 (s), 15.66 (s), 13.90 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 17.31 (s). HRMS: [M-H]$^-$ calculated for C$_{14}$H$_{23}$N$_6$O$_5$PS, 417.1115; found 417.1113.

(2S)-2-((((((2R,3S,5R)-5-(6-Amino-9H-purin-9-yl)-3-hydroxytetrahydrofuran-2-yl)methoxy) (hydroxy)phosphoryl)amino)-5((diaminomethylene)amino)pentanoic acid triethylammonium salt (2.3f) This compound was synthesized from 2'-deoxyadenosine 5'-monophosphate (100 mg, 0.30 mmol) and L-arginine methyl ester dihydrochloride (551 mg, 2.11 mmol), according to the procedure

used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (47 mg, 32 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.37 (s, 1H), 8.01 (s, 1H), 6.32 (t, J = 7.0 Hz, 1H), 4.64 – 4.56 (m, 1H), 4.16 (t, J = 3.0 Hz, 1H), 3.85 (dd, J = 6.5, 3.3 Hz, 2H), 3.42 (dt, J = 8.6, 5.3 Hz, 1H), 2.74-2.64 (m, 3H), 2.53-2.46 (m, 1H), 1.43-1.25 (m, 4H). ^{13}C NMR (75 MHz, D$_2$O) δ 180.95 (d, J = 6.0 Hz), 156.28 (s), 155.12 (s), 152.43 (s), 148.35 (s), 139.65 (s), 118.11 (s), 85.96 (d, J = 9.3 Hz), 83.30 (s), 71.26 (s), 63.75 (d, J=4.6 Hz), 56.03 (s), 40.66 (s), 38.95 (s), 31.57 (d, J=4.2 Hz), 23.66 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 6.83 (s). HRMS: [M-H]$^-$ calculated for C$_{16}$H$_{26}$N$_9$O$_7$P, 486.1620; found 486.1610.

(((((R)-1-(6-amino-9H-purin-9-yl)propan-2-yl)oxy)methyl)(hydroxy)phosphoryl)-L-arginine triethylammonium salt (2.4f) This compound was prepared from 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine monohydrate (100 mg, 0.33 mmol) and L-arginine methyl ester dihydrochloride (598 mg, 2.29 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (35 mg, 24 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.05 (s, 1H), 7.92 (s, 1H), 4.16-4.01 (m, 2H), 3.90-3.60 (m, 1H), 3.53-3.22 (m, 3H), 2.92 (s, 2H), 1.40 (d, J = 5.5 Hz, 4H), 0.94 (t, J=6.0 Hz, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 181.13 (d, J = 4.4 Hz), 156.38 (s), 154.99 (s), 151.94 (s), 148.74 (s), 143.03 (s), 117.55 (s), 75.16 (d, J = 12.1 Hz), 65.81 (d, J=143.3 Hz), 55.77 (s), 47.42 (s), 40.73 (s), 32.24 (d, J=2.7 Hz), 23.77 (s), 15.68 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 17.36 (s). HRMS: [M-H]$^-$ calculated for C$_{15}$H$_{26}$N$_9$O$_5$P, 442.1722; found 442.1715.

(((((2R,3S,5R)-5-(6-amino-9H-purin-9-yl)-3-hydroxytetrahydrofuran-2-yl)methoxy)(hydroxy)phosphoryl)-L-phenylalanine triethylammonium salt (2.3g) This compound was prepared from 2'-deoxyadenosine 5'-monophosphate (100 mg, 0.30 mmol) and L-phenylalanine methyl ester hydrochloride (455 mg, 2.11 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (47 mg, 33 % over two steps). ^{1}H NMR (300 MHz, D$_2$O) δ 8.32 (s, 1H), 8.02 (s, 1H), 7.17-6.95 (m, 5H), 6.30 (t, J = 6.8 Hz, 1H), 4.62-4.51 (m, 1H), 4.18-4.08 (m, 1H), 3.81-3.67 (m, 3H), 2.74-2.64 (m, 3H)), 2.54-2.46 (m, 1H). ^{13}C NMR (75 MHz, D$_2$O) δ 180.52 (d, J = 5.7 Hz), 155.16 (s), 152.38 (s), 148.34 (s), 139.60 (s), 138.00 (s), 129.21 (s), 127.88 (s), 125.98 (s), 118.30 (s), 85.88 (d, J = 9.2 Hz), 83.38 (s), 71.10 (s), 63.59 (d, J = 5.1 Hz), 58.18 (s), 40.70 (d, J = 4.8 Hz), 38.94 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 6.61 (s). HRMS: [M+H]$^+$ calculated for C$_{19}$H$_{23}$N$_6$O$_7$P, 479.1438; found 479.1439.

(((((R)-1-(6-amino-9H-purin-9-yl)propan-2-yl)oxy)methyl)(hydroxy)phosphoryl)-L-phenylalanine triethylammonium salt (2.4g) This compound was prepared from 9-[(*R*)-2-(phosphonomethoxy)propyl]adenine monohydrate (100 mg, 0.33 mmol) and L-phenylalanine methyl ester hydrochloride (494 mg, 2.29 mmol), according to the procedure used for the synthesis of compound **2.3a**, yielding the title compound as a colourless solid (57 mg, 40 % over two steps). [1]H

NMR (300 MHz, D$_2$O) δ 8.03 (s, 1H), 7.95 (s, 1H), 7.10-6.96 (m, 5H), 4.10-4.05 (m, 2H), 3.78-3.61 (m, 2H), 3.27 (dd, J = 12.6, 9.2 Hz, 1H), 3.04 (dd, J=12.6, 9.2 Hz, 1H), 2.61 (d, J = 6.1 Hz, 2H), 1.00 (d, J = 6.2 Hz). ^{13}C NMR (75 MHz, D$_2$O) δ 181.03 (d, J = 4.4 Hz), 155.33 (s), 152.24 (s), 149.04 (s), 143.28 (s), 138.45 (s), 129.65 (s), 128.19 (s), 126.29 (s), 118.03 (s), 75.46 (d, J = 12.3 Hz), 66.44 (d, J = 147.0 Hz), 58.14 (s), 47.96 (s), 41.59 (d, J=3.1 Hz), 16.23 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 16.75 (s). HRMS: [M+H]$^+$ calculated for C$_{18}$H$_{23}$N$_6$O$_5$P, 435.1540; found 435.1538.

Incorporation assay

The primer labeled with Cy5, 5'-Cy5-CAGGAAACAGCTATGAC, was annealed with the template sequence 3'-GTCCTTTGTCGATACTG**TTTTTTT**-5'. This DNA template/primer was used for tests of primer extension by HIV-RT with the various compounds in this study. Reactions were carried out at 37 °C with 125 nM primer/template, 500 μM compound or 50 μM dATP, and 0.5 μg/μL reaction of RT. The buffer was 50 mM Tris-HCl, 3 mM MgCl$_2$ with or without 2 mM MnCl$_2$, 10 mM DTT, pH 8.3. The time periods were 10 min, 1 h, and 24 h. Aliquots were taken at 10 and 60 min and quenched by adding a double volume of quenching buffer (90 % formamide, 50 mM EDTA, and 0.05 % orange G) and heated at 95 °C for 5 min. The samples were separated on a 1 mm 20 % denaturing polyacrylamide gel and gel bands were visualized using the Typhoon FLA 9500 imaging system (GE Healthcare Life Sciences). The images were processed using ImageQuant TL v8.1.0.0 (GE Healthcare). The expression and purification of the RT127A construct has been described.[32]

Inhibition of HIV-1 RT activity

Enzyme assays using recombinant HIV-1 RT were carried out in 96-well black plates by measuring the conversion of double-strand DNA from Td100/Pd18 template-primer using a PicoGreen-based spectrophotometric assay in the presence of increasing concentrations of compound as described previously.[40] Briefly, reactions containing purified recombinant HIV-1 reverse transcriptase, Td100/Pd18 DNA substrate in a buffer containing 50 mm Tris, pH 7.8, 6 mM MgCl$_2$ or 2 mM MnCl$_2$, and 50 mM NaCl were initiated by the addition of 10 μM dNTPs. DNA synthesis was carried out for 30 min at 37 °C followed by quenching with the addition of 100 mM EDTA. Quant-iT^{TM} PicoGreen reagent (Thermo Fisher Scientific Inc.) was added to quantify the amount of formed double-stranded DNA in TE buffer. The reaction mixtures were then excited at 480 nm, and fluorescence was monitored at 520 nm using a 96-well CLARIOstar microplate reader (BMG LABTECH). Dose-response curves of triplicate samples were plotted using GraphPad Prism version 5.0 (GraphPad Software, San Diego, CA) to determine the 50 % inhibitory concentration (IC$_{50}$).

RT expression, RT/DNA cross-linking, purification

RT construct RT127A used for this study was derived from a previously reported construct RT13A.[41] The mutation D498N in the RT127A construct was introduced using the methods described therein; D498 is a part of the RNase H active site. The D498N mutant RT blocks RNase H activity, but this mutant has polymerase activity comparable to wild-type RT.[42] The D498N mutant RT was expressed and purified as previously reported.[41] Briefly, the RT was expressed in BL21-CodonPlus-RIL cells, induced with 1 mM IPTG at an OD_{600} of 0.9, followed by expression at 37 °C for 3 h. The cells were sonicated at a power output of ~45 W with a Branson Sonifier SFX250 sonicator. The samples were purified using a nickel-nitrilotriacetic acid (Ni-NTA) column according to the manufacturer's recommendations (Qiagen). The final purification step was carried out using a Mono Q column, and the purified RT samples were buffer exchanged into 10 mM Tris, pH 8.0, and 75 mM NaCl.

The 27-mer DNA template (5′-ATGGAAGGCGCCCGAACAGGGACTG TG-3′), containing the PBS sequence from the HIV genome, was synthesized by Integrated DNA Technologies. The 20-mer primer (5′-ACAGTCCCTGTTCGG**G**CGCC-3′) bearing a cross-linkable thioalkyl tether (on **G** in the primer strand)[43] was custom synthesized by Midland Certified Reagent Company, using phosphoramidite that was custom synthesized by Chemgenes, and the primer was annealed to the template. The 27:20-mer dsDNA was cross-linked to RT127A at the mutated Q258C site of p66, and the cross-linked primer was extended with a ddG at the 3′-end through RT polymerization.[44] The cross-linked RT127A–DNA (AZT-terminated primer) complex was purified using Ni-NTA and heparin columns in tandem as previously described.[44] The His tag was removed by adding human rhinovirus 14 3C protease at a 1:10 ratio to the RT–DNA complex at 4 °C for 48 h. Ni-NTA beads were added to remove the cleaved His tag and His-tagged protease. An Amicon Ultra-4 Ultracel unit (30-KDa cutoff) was used to exchange the buffer into 75 mM NaCl, 10 mM Tris-HCl, pH 8.0. The protein was concentrated to 12 mg/mL using a mass extinction coefficient of $OD_{280} = 5.4$ for 1 mg/mL.

Crystallization and structure determination

Sitting drops of 0.1 μL of protein plus 0.1 μL of well solution were set up at room temperature on Swiss CI UVXPO three drop (midi), or two drop plates. Each well contained 40 μL (midi) or 85 μL of 16-19.5 % v/v PEG Smear Broad (Molecular Dimensions, MD2-250-261), 0.2 M $(NH_4)_2SO_4$, and 0.1 M Tris-HCl, pH 8.3 or 8.5. The protein solution was 10 mg/mL in 75 mM NaCl, 10 mM Tris, pH 8.0. Trays were incubated at 4 °C. Crystals began appearing in 2 or 3 days and were done growing by 2 weeks. Crystals chosen for data collection were approximately 150 mm × 100 mm × 60 mm in size. To soak compounds into crystals, compounds were first dissolved in DMSO to 100 mM. Two cryo soaking solutions were prepared: 10 or 20 % v/v glycerol, 20 % v/v PEG Smear Broad, 0.2 M $(NH_4)_2SO_4$, 0.1 M Tris-HCl pH 8.0, 5 mM $MgCl_2$ +/- 2 mM $MgCl_2$, 2 mM compound. For freezing

crystals, a nine well depression plate was used and sealed with plastic film. Crystals were soaked in 50 μL of the 10 % glycerol solution for 1−3 h, then soaked in the same volume of 20 % glycerol solution for 1−2 min, then flash frozen in liquid nitrogen. X-ray diffraction data were collected at ESRF beamlines ID29 and ID30B, and at SOLEIL beamline PX2A. Images were integrated using iMosflm[45] and scaled using AIMLESS[46] in the CCP4 v7.0 package.[47] The complexes of RT/DNA with each compound were solved starting with rigid body refinement in PHENIX,[48] using a starting model of RT/DNA (PDB code 3V4I). The crystals grew in space group P2$_1$ with two RT heterodimers in the asymmetric unit. The atomic models were rebuilt by manual model building in COOT,[49] guided by structure validation with MolProbity.[50] The models were refined in PHENIX.[48] Data and refinement statistics are in **Table 2-2**. Figures in the manuscript and supplementary information were prepared with PyMOL[51] and LigPlot$^+$,[52] respectively.

Table 2-2. Data collection and refinement statistics

compound, chemical name, divalent cations	**2.4b** D-Asp-TFV Mn^{2+}/Mg^{2+}	**2.4b** D-Asp-TFV Mg^{2+}	**2.4d** L-Glu-TFV Mn^{2+}/Mg^{2+}	**2.4d** L-Glu-TFV Mg^{2+}	**2.4e** L-Met-TFV Mn^{2+}/Mg^{2+}	**2.4e** L-Met-TFV Mg^{2+}
Crystal Parameters						
space group	P2$_1$	P2$_1$	P2$_1$	P2$_1$	P2$_1$	P2$_1$
cell dimensions a,b,c (Å), b (°)	89.4, 132.4, 138.7, 98.5	89.9, 133.5, 140.2, 98.4	89.7, 132.7, 138.8, 98.7	89.7, 133.6, 140.4, 98.4	89.5, 132.7, 139.0, 98.6	89.6, 132.9, 140.0, 98.5
molecules/AU^a	2	2	2	2	2	2
Data Collection						
Beamline, detector	ESRF ID30B Pilatus Dectris 6M	ESRF ID29 Pilatus Dectris 6MF	ESRF ID30B Pilatus Dectris 6M	ESRF ID29 Pilatus Dectris 6MF	ESRF ID30B Pilatus Dectris 6M	SOLEIL PX2A Eiger X 9M
wavelength Å	0.9763	0.9762	0.9763	0.9762	0.9763	0.9801
resolution range^b	69.8-2.73 (2.78-2.73)	53.37-3.15 (3.24-3.15)	80.14-2.75 (2.80-2.75)	54.0-2.95 (3.02-2.95)	69.78-2.62 (2.66-2.62)	80.22-2.95 (3.02-2.95)
observed/unique^c reflections	285341/84580 (14387/4484)	237495/56199 (19635/4634)	282670/82069 (16040/4508)	315909/68643 (19439/4430)	302418/96082 (15393/4715)	469134/68120 (32426/4537)
completeness %^{b}	99.7 (99.5)	99.2 (99.4)	98.5 (98.2)	99.5 (99.6)	99.7 (99.7)	99.7 (99.3)
multiplicity	3.4 (3.2)	4.2 (4.2)	3.4 (3.6)	4.6 (4.4)	3.1 (3.5)	6.9 (7.1)
R$_{merge}$ %^{b,d}	10.0 (38.8)	19.4 (168.4)	11.8 (62.9)	14.5 (124.2)	8.2 (41.3)	15.6 (148.9)
R$_{pim}$ %	6.4 (25.4)	10.4 (90.0)	7.4 (38.9)	7.5 (65.0)	5.5 (26.7)	6.4 (59.7)
I/s(I)	7.3 (2.7)	5.4 (1.5)	6.2 (1.8)	6.5 (1.5)	8.1 (2.6)	7.9 (1.3)
CC$_{1/2}$	0.991 (0.286)	0.984 (0.289)	0.993 (0.302)	0.992 (0.287)	0.986 (0.304)	0.997 (0.281)
Refinement						
resolution, Å	69.8-2.73	53.37-3.15	80.14-2.75	54.0-2.95	69.78-2.62	69.22-2.95
R$_{work}$/R$_{free}$^{e}	21.4/25.4	20.8/24.9	21.8/24.9	21.3/25.7	22.2/25.5	22.7/26.3
no. atoms	17779/17778	17670	17899	17693	17786	17688
macromolecules	17571/17570	17540	17571	17571	17571	17554
ligand	100	69	102	70	100	69
water	108	61	226	52	115	65
B-factors, Å^2						
Wilson B	64.4	78.3	65.1	75.8	66.3	84.6
all atoms	77.0	107.1	85.1	106.2	81.5	112.3
macromolecules	77.2	107.2	85.4	106.2	81.6	112.4
ligands	83.3	127.7/127.6	91.9	126.3	90.7	136.4
waters	48.6	63.0/63.9	55.7	67.4	50.9	72.2
rmsd^f						
bond length, Å	0.002	0.002	0.002	0.002	0.002	0.002
bond angle, °	0.44	0.46	0.41	0.4	0.43	0.46
Ramachandran %^{g}	96.5/3.4/0.1	96.3/3.7/0.1	96.3/3.5/0.2	96.3/3.8/0.0	96.5/3.4/0.1	96.3/3.6/0.2
PDB code	7AHX	7AID	7AIF	7AIG	7AII	7AIJ

^{a}Asymmetric unit. ^{b}Values in parentheses of resolution range, completeness, R$_{merge}$, and I/σ(I) correspond to the last resolution shell. ^{c}Friedel pairs were treated as identical reflections. ^{d}Rmerge(I) =$\Sigma_{hkl}\Sigma_j|[I(hkl)_j - I(hkl)]|/\Sigma_{hkl}I_{hkl}$, where *I(hkl)$_j$* is the measurement of the intensity of reflection *hkl* and *I(hkl)* is the average intensity.^{e}R = $\Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where R$_{free}$ is calculated without a σ cutoff for a randomly chosen 3 % of reflections, which were not used

for structure refinement, and R$_{work}$ is calculated for the remaining reflections. ^{f}Root mean square deviations from ideal bond lengths/angles. ^{g}Number of residues in favored/allowed/outlier region.

2.6 References

1. De Clercq, E.; Li, G. Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev* **2016,** 29, 695-747.
2. Gollnest, T.; de Oliveira, T. D.; Schols, D.; Balzarini, J.; Meier, C. Lipophilic prodrugs of nucleoside triphosphates as biochemical probes and potential antivirals. *Nat Commun* **2015,** 6, 8716.
3. Gollnest, T.; Dinis de Oliveira, T.; Rath, A.; Hauber, I.; Schols, D.; Balzarini, J.; Meier, C. Membrane-permeable Triphosphate Prodrugs of Nucleoside Analogues. *Angew Chem Int Ed Engl* **2016,** 55, 5255-8.
4. Jia, X.; Schols, D.; Meier, C. Anti-HIV-Active Nucleoside Triphosphate Prodrugs. *J Med Chem* **2020,** 63, 6003-6027.
5. Meier, C. Nucleoside diphosphate and triphosphate prodrugs - An unsolvable task? *Antivir Chem Chemother* **2017,** 25, 69-82.
6. Weising, S.; Sterrenberg, V.; Schols, D.; Meier, C. Synthesis and Antiviral Evaluation of TriPPPro-AbacavirTP, TriPPPro-CarbovirTP, and Their 1',2'-cis-Disubstituted Analogues. *Chemmedchem* **2018,** 13, 1771-1778.
7. Kandil, S.; Pannecouque, C.; Chapman, F. M.; Westwell, A. D.; McGuigan, C. Polyfluoroaromatic stavudine (d4T) ProTides exhibit enhanced anti-HIV activity. *Bioorg Med Chem Lett* **2019,** 29, 126721.
8. Seley-Radtke, K. L.; Yates, M. K. The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Res* **2018,** 154, 66-86.
9. Steitz, T. A.; Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** 90, 6498-502.
10. Yang, W. Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys* **2011,** 44, 1-93.
11. Grobler, J. A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; Wai, J. S.; Young, S.; Vacca, J.; Hazuda, D. J. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc Natl Acad Sci USA* **2002,** 99, 6661-6666.
12. Ruiz, F. X.; Hoang, A.; Das, K.; Arnold, E. Structural Basis of HIV-1 Inhibition by Nucleotide-Competing Reverse Transcriptase Inhibitor INDOPY-1. *J Med Chem* **2019,** 62, 9996-10002.
13. Jochmans, D.; Deval, J.; Kesteleyn, B.; Van Marck, H.; Bettens, E.; De Baere, I.; Dehertogh, P.; Ivens, T.; Van Ginderen, M.; Van Schoubroeck, B.; Ehteshami, M.; Wigerinck, P.; Gotte, M.; Hertogs, K. Indolopyridones inhibit human immunodeficiency virus reverse transcriptase with a novel mechanism of action. *J Virol* **2006,** 80, 12283-92.
14. Debarge, S.; Balzarini, J.; Maguire, A. R. Design and synthesis of alpha-carboxy phosphononucleosides. *J Org Chem* **2011,** 76, 105-26.
15. John, J.; Kim, Y.; Bennett, N.; Das, K.; Liekens, S.; Naesens, L.; Arnold, E.; Maguire, A. R.; Gotte, M.; Dehaen, W.; Balzarini, J. Pronounced Inhibition Shift from HIV Reverse Transcriptase to Herpetic DNA Polymerases by Increasing the Flexibility of alpha-Carboxy Nucleoside Phosphonates. *J Med Chem* **2015,** 58, 8110-27.
16. Keane, S. J.; Ford, A.; Mullins, N. D.; Maguire, N. M.; Legigan, T.; Balzarini, J.; Maguire, A. R. Design and synthesis of alpha-carboxy nucleoside phosphonate analogues and evaluation as HIV-1 reverse transcriptase-targeting agents. *J Org Chem* **2015,** 80, 2479-93.
17. Maguire, N. M.; Ford, A.; Balzarini, J.; Maguire, A. R. Synthesis of Guanine alpha-Carboxy Nucleoside Phosphonate (G-alpha-CNP), a Direct Inhibitor of Multiple Viral DNA Polymerases. *J Org Chem* **2018,** 83, 10510-10517.
18. Mullins, N. D.; Maguire, N. M.; Ford, A.; Das, K.; Arnold, E.; Balzarini, J.; Maguire, A. R. Exploring the role of the alpha-carboxyphosphonate moiety in the HIV-RT activity of alpha-carboxy nucleoside phosphonates. *Org Biomol Chem* **2016,** 14, 2454-65.
19. Hladezuk, I.; Chastagner, V.; Collins, S. G.; Plunkett, S. J.; Ford, A.; Debarge, S.; Maguire, A. R. Development of O-H insertion for the attachment of phosphonates to nucleosides; synthesis of alpha-carboxy phosphononucleosides. *Tetrahedron* **2012,** 68, 1894-1909.
20. Balzarini, J.; Menni, M.; Das, K.; van Berckelaer, L.; Ford, A.; Maguire, N. M.; Liekens, S.; Boehmer, P. E.; Arnold, E.; Gotte, M.; Maguire, A. R. Guanine alpha-carboxy nucleoside phosphonate (G-alpha-CNP) shows a different inhibitory kinetic profile against the DNA polymerases of human immunodeficiency virus (HIV) and herpes viruses. *Biochem Pharmacol* **2017,** 136, 51-61.
21. Balzarini, J.; Das, K.; Bernatchez, J. A.; Martinez, S. E.; Ngure, M.; Keane, S.; Ford, A.; Maguire, N.; Mullins, N.; John, J.; Kim, Y.; Dehaen, W.; Vande Voorde, J.; Liekens, S.; Naesens, L.; Gotte, M.; Maguire, A. R.; Arnold, E. Alpha-carboxy nucleoside phosphonates as universal nucleoside triphosphate mimics. *Proc Natl Acad Sci U S A* **2015,** 112, 3475-80.
22. Das, K.; Balzarini, J.; Miller, M. T.; Maguire, A. R.; DeStefano, J. J.; Arnold, E. Conformational States of HIV-1 Reverse Transcriptase for Nucleotide Incorporation vs Pyrophosphorolysis-Binding of Foscarnet. *Acs Chem Biol* **2016,** 11, 2158-64.
23. Adelfinskaya, O.; Herdewijn, P. Amino acid phosphoramidate nucleotides as alternative substrates for HIV-1 reverse transcriptase. *Angew Chem Int Ed Engl* **2007,** 46, 4356-8.
24. Adelfinskaya, O.; Terrazas, M.; Froeyen, M.; Marliere, P.; Nauwelaerts, K.; Herdewijn, P. Polymerase-catalyzed synthesis of DNA from phosphoramidate conjugates of deoxynucleotides and amino acids. *Nucleic Acids Res.* **2007,** 35, 5060-5072.

25. Terrazas, M.; Marliere, P.; Herdewijn, P. Enzymatically catalyzed DNA synthesis using L-Asp-dGMP, L-Asp-dCMP, and L-Asp-dTMP. *Chem Biodivers* **2008,** 5, 31-9.

26. Giraut, A.; Song, X. P.; Froeyen, M.; Marliere, P.; Herdewijn, P. Iminodiacetic-phosphoramidates as metabolic prototypes for diversifying nucleic acid polymerization in vivo. *Nucleic Acids Res* **2010,** 38, 2541-50.

27. Song, X. P.; Bouillon, C.; Lescrinier, E.; Herdewijn, P. Iminodipropionic acid as the leaving group for DNA polymerization by HIV-1 reverse transcriptase. *ChemBioChem* **2011,** 12, 1868-1880.

28. Yang, S.; Froeyen, M.; Lescrinier, E.; Marliere, P.; Herdewijn, P. 3-Phosphono-L-alanine as pyrophosphate mimic for DNA synthesis using HIV-1 reverse transcriptase. *Org Biomol Chem* **2011,** 9, 111-9.

29. Song, X. P.; Bouillon, C.; Lescrinier, E.; Herdewijn, P. Dipeptides as leaving group in the enzyme-catalyzed DNA synthesis. *Chem Biodivers* **2012,** 9, 2685-700.

30. Abraham, T. W.; Kalman, T. I.; McIntee, E. J.; Wagner, C. R. Synthesis and biological activity of aromatic amino acid phosphoramidates of 5-fluoro-2'-deoxyuridine and 1-beta-arabinofuranosylcytosine: evidence of phosphoramidase activity. *J Med Chem* **1996,** 39, 4569-75.

31. Singer, V. L.; Jones, L. J.; Yue, S. T.; Haugland, R. P. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. *Anal Biochem* **1997,** 249, 228-38.

32. Das, K.; Martinez, S. E.; Bauman, J. D.; Arnold, E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol* **2012,** 19, 253-9.

33. Das, K.; Martinez, S. E.; Bandwar, R. P.; Arnold, E. Structures of HIV-1 RT-RNA/DNA ternary complexes with dATP and nevirapine reveal conformational flexibility of RNA/DNA: insights into requirements for RNase H cleavage. *Nucleic Acids Res* **2014,** 42, 8125-37.

34. Das, K.; Bandwar, R. P.; White, K. L.; Feng, J. Y.; Sarafianos, S. G.; Tuske, S.; Tu, X.; Clark, A. D., Jr.; Boyer, P. L.; Hou, X.; Gaffney, B. L.; Jones, R. A.; Miller, M. D.; Hughes, S. H.; Arnold, E. Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *J. Biol. Chem.* **2009,** 284, 35092-100.

35. Sarafianos, S. G.; Pandey, V. N.; Kaushik, N.; Modak, M. J. Site-directed mutagenesis of arginine 72 of HIV-1 reverse transcriptase. Catalytic role and inhibitor sensitivity. *J Biol Chem* **1995,** 270, 19729-35.

36. Tuske, S.; Sarafianos, S. G.; Clark, A. D., Jr.; Ding, J.; Naeger, L. K.; White, K. L.; Miller, M. D.; Gibbs, C. S.; Boyer, P. L.; Clark, P.; Wang, G.; Gaffney, B. L.; Jones, R. A.; Jerina, D. M.; Hughes, S. H.; Arnold, E. Structures of HIV-1 RT-DNA complexes before and after incorporation of the anti-AIDS drug tenofovir. *Nat Struct Mol Biol* **2004,** 11, 469-74.

37. Yin, W.; Mao, C.; Luan, X.; Shen, D. D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y. C.; Tian, G.; Jiang, H. W.; Tao, S. C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Zhang, Y.; Xu, H. E. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020,** 368, 1499-1504.

38. Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of replicating SARS-CoV-2 polymerase. *Nature* **2020,** 584, 154-156.

39. Maiti, M.; Michielssens, S.; Dyubankova, N.; Maiti, M.; Lescrinier, E.; Ceulemans, A.; Herdewijn, P. Influence of the nucleobase and anchimeric assistance of the carboxyl acid groups in the hydrolysis of amino acid nucleoside phosphoramidates. *Chem. Eur. J.* **2012,** 18, 12.

40. Singh, K.; Marchand, B.; Rai, D. K.; Sharma, B.; Michailidis, E.; Ryan, E. M.; Matzek, K. B.; Leslie, M. D.; Hagedorn, A. N.; Li, Z.; Norden, P. R.; Hachiya, A.; Parniak, M. A.; Xu, H. T.; Wainberg, M. A.; Sarafianos, S. G. Biochemical mechanism of HIV-1 resistance to rilpivirine. *J Biol Chem* **2012,** 287, 38110-23.

41. Bauman, J. D.; Das, K.; Ho, W. C.; Baweja, M.; Himmel, D. M.; Clark, A. D., Jr.; Oren, D. A.; Boyer, P. L.; Hughes, S. H.; Shatkin, A. J.; Arnold, E. Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. *Nucleic Acids Res* **2008,** 36, 5083-92.

42. DeStefano, J. J.; Wu, W.; Seehra, J.; McCoy, J.; Laston, D.; Albone, E.; Fay, P. J.; Bambara, R. A. Characterization of an RNase H deficient mutant of human immunodeficiency virus-1 reverse transcriptase having an aspartate to asparagine change at position 498. *Biochim Biophys Acta* **1994,** 1219, 380-8.

43. Hou, X.; Wang, G.; Gaffney, B. L.; Jones, R. A. Synthesis of guanosine and deoxyguanosine phosphoramidites with cross-linkable thioalkyl tethers for direct incorporation into RNA and DNA. *Nucleosides Nucleotides Nucleic Acids* **2009,** 28, 1076-94.

44. Sarafianos, S. G.; Clark, A. D., Jr.; Tuske, S.; Squire, C. J.; Das, K.; Sheng, D.; Ilankumaran, P.; Ramesha, A. R.; Kroth, H.; Sayer, J. M.; Jerina, D. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Trapping HIV-1 reverse transcriptase before and after translocation on DNA. *J Biol Chem* **2003,** 278, 16280-8.

45. Powell, H. R.; Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Leslie, A. G. W. Integrating macromolecular X-ray diffraction data with the graphical user interface iMosflm. *Nat Protoc* **2017,** 12, 1310-1325.

46. Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **2013,** 69, 1204-14.

47. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **2011,** 67, 235-42.

48. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure

solution. *Acta Crystallogr D Biol Crystallogr* **2010,** 66, 213-21.

49. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** 66, 486-501.

50. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010,** 66, 12-21.

51. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

52. Laskowski, R. A.; Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* **2011,** 51, 2778-2786.

Chapter 3

Synthesis and biological evaluation of a novel class of acyclic nucleoside analogs as inhibitors of HIV-1 reverse transcriptase

3.1 Introduction

AIDS is primarily caused by HIV-1. Currently, more than 30 drugs received marketing approval by the FDA for the treatment of HIV-infected patients, and almost half of them target HIV-1 RT, highlighting the essential role of RT in the HIV life cycle.[1] The approved RT inhibitors include NRTIs and NNRTIs, which are essential to combination antiretroviral therapy (cART).[2] Although cART can transform AIDS from a lethal infection to a manageable chronic disease, there is no cure yet and HIV/AIDS still poses a threat to global health with an estimated 37.7 million people infected worldwide in 2020.[3] The effectiveness of cART is mainly limited by HIV drug resistance. Therefore, new drugs with novel mechanisms of action for HIV treatment and prevention are highly needed. NcRTIs, such as α-CNPs and indolopyridones (e.g. INDOPY-1),[4-7] displayed a different inhibition mechanism from the approved NRTIs and NNRTIs, since they inhibit RT by competing with the natural dNTP substrates.

In Chapter 2, we synthesized a series of dAMP/TFV analogs conjugated with various amino acids via a phosphoroamidate or phosphonoamidate bond. These compounds were evaluated for possible incorporation into a growing DNA strand, catalyzed by HIV-1 RT and were also investigated for their ability to function as inhibitors of RT enzymatic activity. Among the tested compounds, L-Glu-TFV (Chapter 2, **2.4d**) was the most efficiently incorporated and hence functions as a NRTI, with the advantage that it does not need cellular conversion to its diphospho-phosphonate metabolite. On the other hand, L-Met-TFV (Chapter 2, **2.4e**) acts as an inhibitor of HIV-1 RT and can therefore be considered as an NcRTI. However, when these amino acid conjugates need to be developed as NcRTIs, the chemical and metabolically lability of these compounds is hampering their further profiling in biochemical and cellular assays. The instability of these amino acid conjugates mainly results from hydrolysis of the phosphoroamidate or phosphonoamidate bond. It has been shown, via ^{31}P NMR, that L-Asp-dAMP is stable for 2 days at 25 °C and at pH 6-8. However, at 70 °C, the half-life was only 3.3 and 2.8 hours at pH 7 and 8.8, respectively.[8]

Furthermore, hydrolysis of the amino acid phosphoramidates of dAMP was influenced by the nucleobase and the carboxylic acid moieties of the amino acid group.[9] Quantum chemical analysis revealed that hydrolysis of the P-O bond was assisted by the α-carboxyl group and involved the formation of a five-membered ring intermediate. In addition, the nature of the nucleobase has an impact on the rate of hydrolysis.[10] Similarly, despite the presence of a stable phosphonate moiety of TFV, amino acid phosphonoamidates of TFV are still subjected to P-N bond hydrolysis.

In order to improve compound stability, we embarked on the synthesis of a series of acyclic nucleoside analogs in which different amino acids were conjugated with the acyclic nucleoside moiety by a stable amide linker. Each of the amino acids carried one or two free carboxylic acid

groups to exploit potential chelation with the catalytic metal ions at the active site of HIV-RT. In addition, the number of negative charges in these newly designed compounds is reduced, which will have a beneficial impact on their cellular permeability. All compounds possess a two-carbon linker between the adenine base and the amide group. This inherent flexible acyclic chain may allow for optimized interactions within the binding site of HIV-1 RT.

3.2 Results and discussion

3.2.1 Chemistry

Commercially available methyl (*R*)-3-hydroxy-2-methyl-propanoate **3.1** was selected as the starting material, since the configuration of the chiral carbon is identical to the stereochemistry of TFV. Tosylation of its primary hydroxyl group afforded intermediate **3.2** (**Scheme 3-1**). Selective N-9 alkylation of adenine with compound **3.2** in the presence of caesium carbonate (Cs$_2$CO$_3$), followed by saponification of the ester moiety yielded carboxylic acid **3.4**. Treatment of carboxylic acid **3.4** with an appropriate amino acid methyl ester hydrochloride using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDCI) and 1-hydroxy-7-azabenzotriazole (HOAt) as coupling reagents, followed by alkaline hydrolysis, furnished target compounds **3.5a-f**.
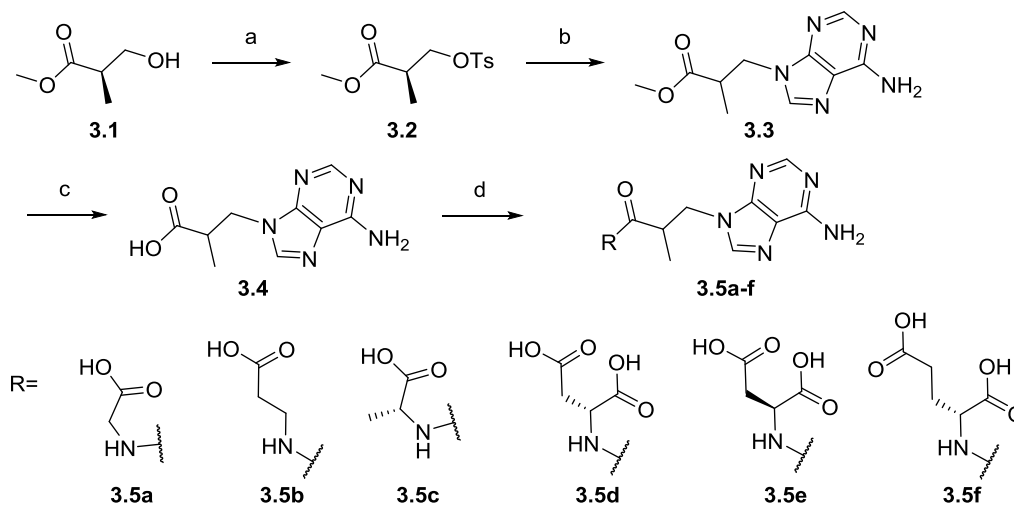
Scheme 3-1. *Reagents and conditions*: (a) TsCl, TEA, DMAP, DCM, 0 °C to r.t., 1 h，100 %; (b) adenine, Cs$_2$CO$_3$, DMF, 66 °C, 16 h, 57 %; (c) NaOH, MeOH, H$_2$O, r.t., 3 h, 85 %; (d) (i) amino acid methyl ester hydrochloride, EDCI, DIEA, HOAt, DMF, -20 °C to r.t., overnight, (ii) NaOH, MeOH, H$_2$O, r.t., 3 h，25 %-42 % over two steps.

Although the synthesis started from an optically pure starting material, NMR analysis of compounds **3.5c-f** revealed that they were isolated as diastereomeric mixtures. No efforts were done to separate the diastereomers and compounds were submitted as mixtures for X-ray crystallographic studies and biological assays. Racemization might be due to the fact that, under alkaline conditions, key intermediate **3.2** undergoes elimination followed by Michael addition, yielding a racemic mixture. To avoid racemization, a Mitsunobu reaction between alcohol **3.1** and adenine was

attempted, but, unfortunately, no desired product was formed. It has been reported that when *N*-Cbz or *N*-Boc protected serine was subjected to Mitsunobo coupling, the elimination product was dominant.[11-12] Considering the structural similarity of **3.1** with serine, the former could still undergo elimination under the applied Mitsunobu reaction conditions.

In order to avoid abovementioned racemization issues, a series of compounds lacking the *α*-methyl functionality was prepared (**Scheme 3-2**). N-9 alkylation of adenine by treatment with methyl 3-bromopropanoate **3.6** yielded intermediate **3.7**. Alkaline hydrolysis of the ester group to the corresponding carboxylic acid **3.8** was followed by coupling with suitable amino acid methyl ester hydrochlorides using EDCI and HOAt as coupling reagents, affording the desired amides. In the last step, the ester groups were hydrolyzed affording target compounds **3.9a-e**.
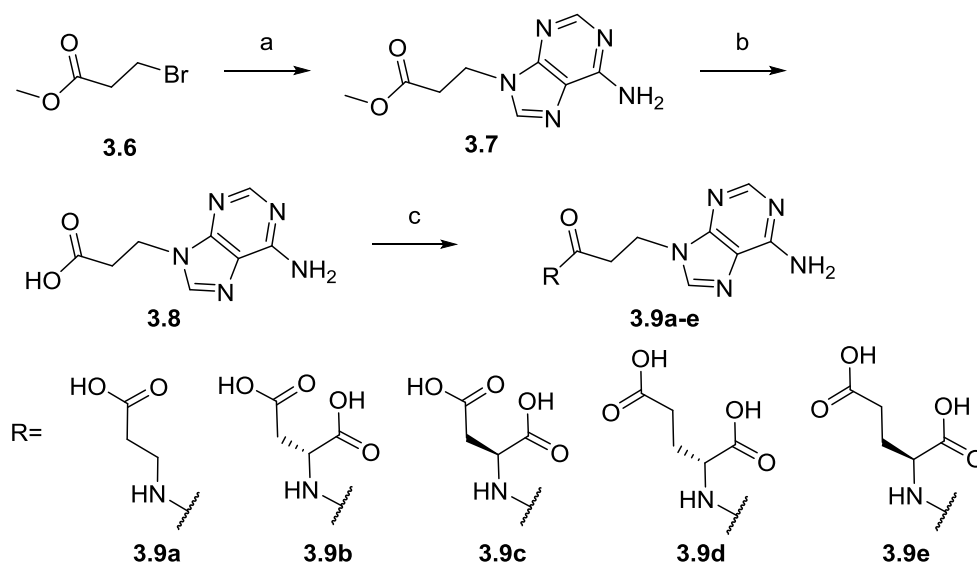
Scheme 3-2. *Reagents and conditions*: (a) adenine, K_2CO_3, DMF, 60 °C, 12 h, 71 %; (b) NaOH, MeOH, H_2O, r.t., 3 h, 88 %; (c) (i) amino acid methyl ester hydrochloride, EDCI, DIEA, HOAt, DMF, -20 °C to r.t., overnight, (ii) NaOH, MeOH, H_2O, r.t., 3 h, 17 %-29 % over two steps.

For screening in a cell-based assay, several ester prodrugs of **3.5d** as well as its analogs were prepared. Carboxylic acid **3.4** was coupled with different amino acid ester hydrochlorides using EDCI and HOAt as coupling reagents, affording the ester prodrugs **3.10a-h** (**Scheme 3-3**). The amino acid ester hydrochlorides used for the synthesis of ester prodrugs **3.10e-h** were prepared from D-aspartic acid using thionyl chloride and the corresponding alcohol following a reported procedure.[13]
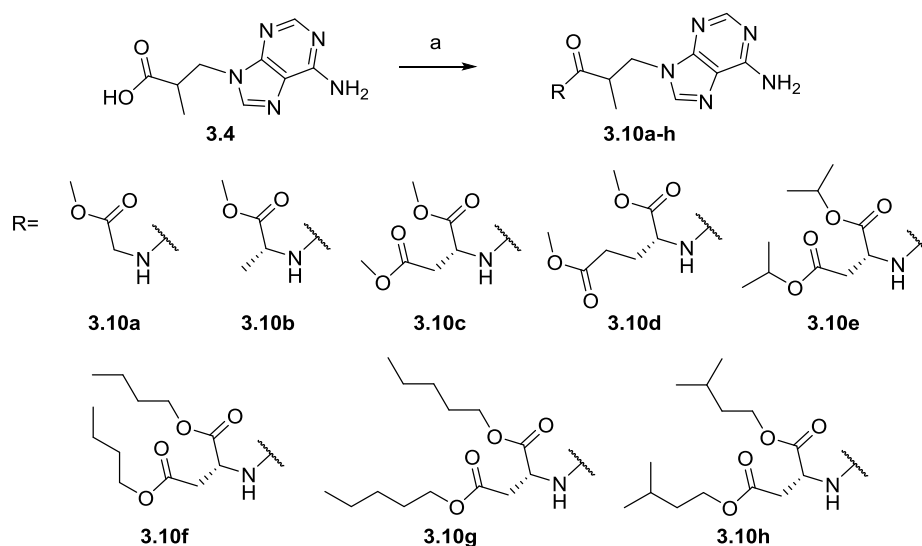
Scheme 3-3. *Reagents and conditions*: (a) amino acid ester hydrochloride, EDCI, DIEA, HOAt, DMF, -20 °C to r.t., overnight, 42 %-89 %.

3.2.2 Biological evaluation

HIV-1 RT inhibition assay (performed by Dr. Hoai Nguyen)

The inhibitory effect of the compounds **3.5a-f** and **3.9a-e** on the DNA synthesis catalyzed by HIV-1 RT was investigated by a gel-based RT inhibition assay (**Figure 3-1**). The primer-template duplex has seven dT nucleotide overhang at the 5' end of the template, so this template/primer allows seven primer extensions at most. The anti-HIV drug EFV was used as the positive control, whereas Milli-Q water was included as the negative control. dATP was used at a concentration of 5 μM with 3 mM MgCl$_2$ as the metal ion source. After 10 min of incubation for polymerization reaction in the presence of individual compounds, EFV showed the most RT inhibition when tested at the concentration of 5 or 50 μM, and no noticeable inhibition at a concentration of 0.5 μM. Among the tested compounds, **3.5a**, **3.5c**, **3.5d** and **3.5f** showed RT inhibition at the concentration of 10 mM, and **3.9d** only exerted weak inhibition for RT at the same concentration. When tested at the concentration of 5 mM, **3.5a** and **3.5d** still maintains their significant inhibition for RT, whereas **3.5c** and **3.5f** only exhibit weak inhibition for RT. However, none of the synthesized compounds show apparent inhibition at the concentration of 1 mM.
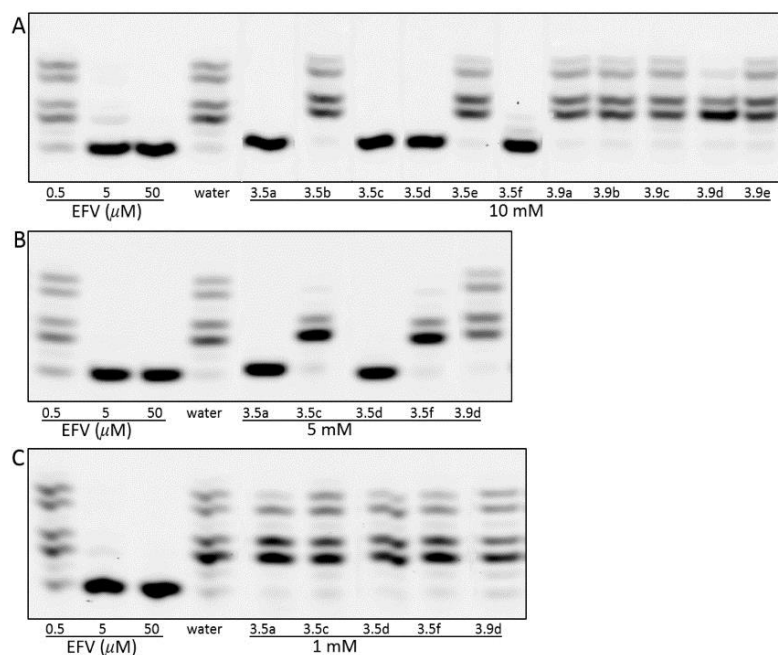
Figure 3-1. HIV-1 RT inhibition assay. Reaction conditions: 125 nM primer/template; [dATP] = 5 μM; [HIV-1 RT] = 0.06 μg/μL, 3 mM MgCl$_2$.

HIV antiviral assay (performed by the group of Prof. Dominique Schols)

Compounds **3.5a-f** and **3.9a-e**, as well as ester prodrugs **3.10a-h** were screened for *in vitro* anti-HIV activity by a luciferase assay[14] in TZM-bl cells infected with the wild type HIV-1 strains NL4.3 (CXCR4-tropic HIV-1 strain) and BaL (CCR5-tropic HIV-1 strain). Simultaneously, compounds were also evaluated for cytotoxicity in TZM-bl cells. TFV was used as a positive control and the data are summarized in **Table 3-1**. Compounds **3.5f**, **3.10c** and **3.10d** completely lacked antiviral activity and cytotoxicity (with EC$_{50}$ and CC$_{50}$ values exceeding 100 μM) and are therefore not included in the table. The majority of the compounds (compounds **3.5b-e**, **3.9a-e**, **3.10a-b** and **3.10e**) showed high micromolar *in vitro* antiretroviral activity against the NL4.3 strain, lacking activity against the BaL strain. Compounds **3.10f-h** showed activity against both HIV-1 strains, although low selectivity indexes were obtained. Given the antiviral profile (activity against one strain) and the level of activity (μM), it is unlikely that the observed antiviral activity is due to RT inhibition.

Table 3-1. Anti-HIV-1 activity and cellular cytotoxicity in TZM-bl cells[a]

Compd	EC_{50}[b] (μM)		CC_{50}[c] (μM)	SI[d]	
	HIV-1 NL4.3	HIV-1 BaL		HIV-1 NL4.3	HIV-1 BaL
3.5a	26.35	32.57	>100	>3	>3
3.5b	75.49	>100	>100	>2	-
3.5c	45.38	95.74	>100	>2	>1
3.5d	32.20	>100	>100	>3	-
3.5e	63.53	>100	>100	>2	-
3.9a	65.52	>100	>100	>2	-
3.9b	24.08	62.33	>100	>6	>2
3.9c	55.62	>100	>100	>3	-
3.9d	53.73	>100	>100	>2	-
3.9e	44.43	>100	>100	>2	-
3.10a	19.28	>100	>100	>5	-
3.10b	34.20	>100	>100	>2	-
3.10e	19.65	>100	>100	>5	-
3.10f	7.03	13.17	36.05	5	3
3.10g	1.96	1.54	6.77	4	4
3.10h	22.56	24.30	45.98	2	2
TFV	6.62	6.45	>100	>15	>15
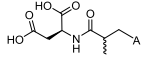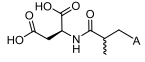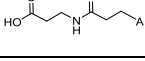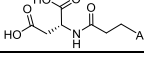
[a]Values are the average of two independent experiments. [b]Effective concentration required to reduce HIV-induced cytopathicity by 50 %. [c]Cytotoxic concentration required to reduce cell viability by 50 %. [d]Selectivity index (SI): CC_{50}/EC_{50}.

3.2.3 X-ray structures (performed by Dr. Sergio Martinez)

Four crystal structures of 3 compounds (**3.5e**, **3.9a** and **3.9b**) were determined in complex with HIV-1 RT/DNA in the presence of Mg^{2+} or Mn^{2+}. The pre-formed crystals of RT cross-linked to a dsDNA template/primer[15] were soaked in a 2 mM solution of individual compounds to form the respective complexes. The diffraction data for the complexes were collected at synchrotron beamlines, and the structures were refined at a resolution between 2.88 and 3.60 Å. The RT/DNA complex crystallized with $P2_1$ space group symmetry containing two copies of RT/DNA in the asymmetric unit. Since the electron density for the inhibitor in chain C is better than chain A, we took chain C for structural analysis and discussion. In order to gain a comprehensive understanding of the binding modes of these compounds at the polymerase active site, we prepared the complexes of inhibitors in the presence of either Mg^{2+} alone or Mn^{2+} and Mg^{2+} ions.

In the crystal structures in **Chapter 2**, no metal chelation was observed in the presence of Mg^{2+} ion alone, whereas in the presence of Mn^{2+} one metal ion was present at the active site. In contrast, herein all the three compounds (via amide oxygen atom) chelate one metal ion in the presence of Mg^{2+} alone. Although the carboxylic acid groups of the inhibitors are not involved in metal chelation as expected, they can form interactions with surrounding residues of protein. Moreover, these compounds can interact with the conserved residue Arg72 (**Table 3-2**). As a part of the fingers β3-β4 hairpin, Arg72 plays a critical role in the binding and positioning of dNTP at the polymerase active site.[16-17] In dNTP-bound structures, the side chain of Arg72 stacks with the dNTP base and forms polar interaction with α-phosphate, whereas in the current structures, the orientation of Arg72 is altered when compared to the dNTP-bound RT/DNA structures.

Table 3-2. Summary of binding information of compounds **3.5e**, **3.9a** and **3.9b**

Crystal structure complex	Chemical structure of inhibitor[a]	Amino acids involved	Interaction surface area (Å^2)[b]	Number of H-bonds	Involved chelating group of inhibitor
3.5e/RT/DNA/Mg^{2+}		K65, R72, Q151, D185	434.9	1	amide O
3.5e/RT/DNA/Mn^{2+}		R72, D113, A114, D185	452	3	amide O
3.9a/RT/DNA/Mg^{2+}		K65, R72, D110, D185	433.6	1	amide O
3.9b/RT/DNA/Mg^{2+}		R72, A114, Q151, D185	440.8	3	amide O

[a] A represents N9-adeninyl.

[b] Generated from https://www.ebi.ac.uk/pdbe/pisa.

Structure of RT/DNA/3.5e

Although **3.5e** was prepared as a racemic mixture, the electron density clearly favored binding of the (*R*)-enantiomer. The (*R*)-isomer resembles TFV in its adenine nucleobase and acyclic chain, except that the oxygen in P-C-O moiety of TFV is replaced with a carbonyl group. The amino acid L-Asp is conjugated via its nitrogen atom to this carbonyl group, yielding an amide derivative. Unlike the phosphonate chelations of the compounds in **Chapter 2**, the overall geometry of **3.5e** remains similar with Mn^{2+} or Mg^{2+}. The adenine base stays parallel and stacks with the primer terminus guanine. In the compound **3.5e**, the two-carbon linker also contains a methyl group on the carbon adjacent to the carbonyl group. The conjugated L-Asp forms contacts with the protein.

In the presence of Mg^{2+} ion, **3.5e** chelates one metal ion at the active site in the RT/DNA/**3.5e**/Mg^{2+} complex that was determined at 2.88 Å resolution (**Figure 3-2**A). One oxygen from the carbonyl group of **3.5e** chelates with the Mg^{2+} at position B in the polymerase active site, and the ion also chelates to one side-chain oxygen of the catalytic aspartate Asp185. The adenine base of **3.5e** does not base pair with the thymine base of the first template overhang nucleotide, possibly due to its stacking interactions with the dsDNA nucleotide guanine base at the primer terminus. In addition, the L-Asp side chain of **3.5e** forms a 3.2 Å salt bridge with Nε of Arg72 and a 2.6 Å interaction with the side chain of Lys65. One α-carboxyl oxygen of the L-Asp moiety forms a 2.6 Å H-bond to Nε2 of Gln151 (**Figure 3-2**B).

Metal chelation is also involved in the binding of **3.5e** in the structure obtained at 2.9 Å resolution in the presence of Mn^{2+} ions. One Mn^{2+} ion positioned at the metal B site chelates the catalytic aspartates Asp110 and Asp185, and the carbonyl oxygen of **3.5e** (**Figure 3-2**C). Due to the stronger chelation ability of Mn^{2+} ion, compound **3.5e** moves closer to the protein backbone and generates new contacts: H-bonds of 2.9 and 3.2 Å between α-carboxyl oxygens of the L-Asp and backbone amide NH of Asp113 and Ala114, respectively (**Figure 3-2**D). Although **3.5e** loses

interactions with Lys65 and Gln151 in the presence of Mn^{2+}, it forms two more H-bonds and its total interaction surface area (452.0 $\AA^2$) is higher than that in the presence of Mg^{2+} (**Table 3-2**). Because of weak binding of the compound at the active site, **3.5e** did not show noticeable RT inhibition at the concentration of 10 mM, although it showed weak antiviral activity (EC_{50} = 65.53 μM) against the wild type HIV-1 strain NL4.3 (**Table 3-1**). The inhibition data are in agreement with the structure that indicates weak binding of **3.5e**, and dATP would outcompete this compound in the RT inhibition assay.
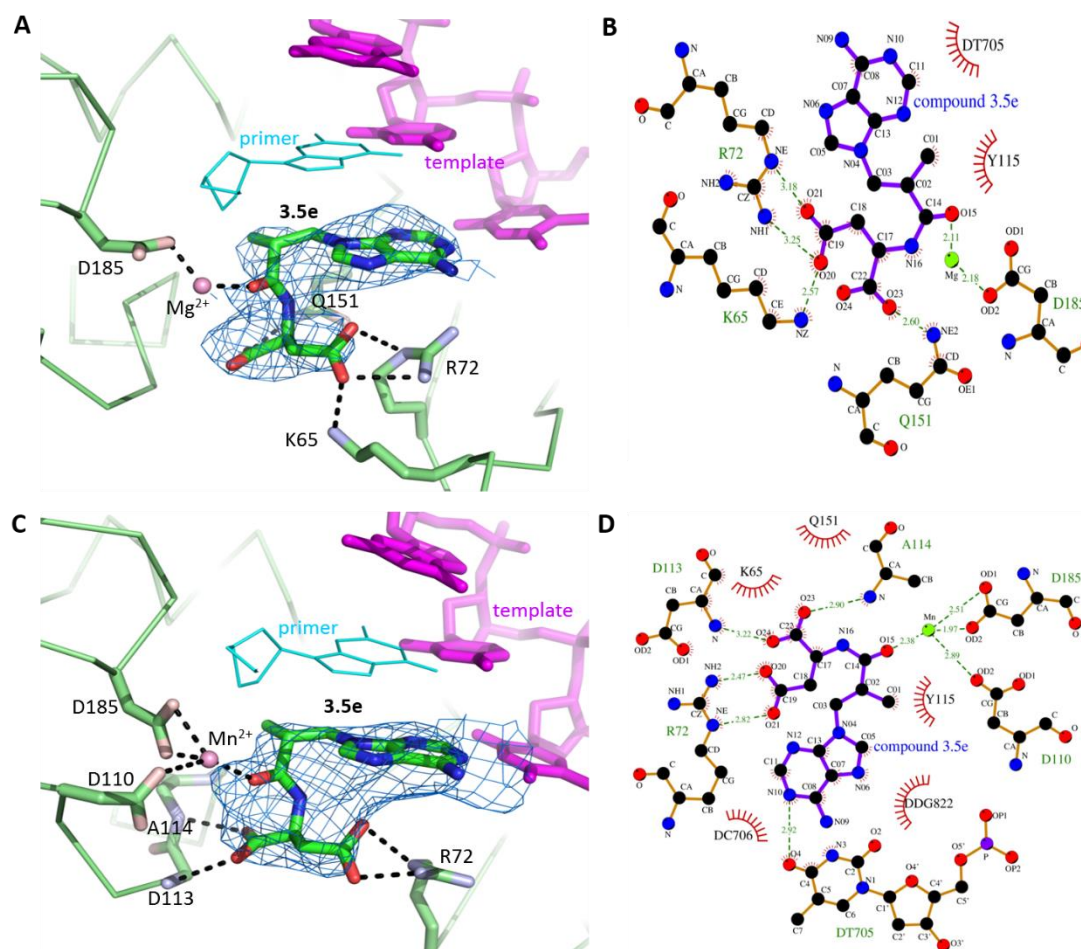
Figure 3-2. Binding of **3.5e** with HIV-1 RT/DNA from crystal structures. (A) Polder map contoured at 4.5 σ showing the binding of **3.5e** at the polymerase active site in the 2.88 Å resolution crystal structure of HIV-1 RT/DNA/**3.5e** complex in the presence of Mg^{2+} (pink sphere). The template strand is magenta, and the primer 3′ terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (B) Ligplot representation of the interaction of **3.5e** with HIV-1 RT/DNA in the presence of Mg^{2+}. (C) Polder map contoured at 6 σ showing the binding of **3.5e** at the polymerase active site in the 2.9 Å resolution crystal structure of HIV-1 RT/DNA/**3.5e** complex in the presence of Mn^{2+} (pink sphere). (D) Ligplot representation of the interaction of **3.5e** with HIV-1 RT/DNA in the presence of Mn^{2+}.

Structures of RT/DNA/3.9a and RT/DNA/3.9b

Compared with **3.5e**, compounds **3.9a** and **3.9b** lack a chiral center in their acyclic chain, and therefore they are grouped together for structural analysis. The crystal structure of RT/DNA/**3.9a** complex in the presence of Mg^{2+} was obtained at 3.05 Å. The carbonyl oxygen of the amide moiety in **3.9a** chelates with a Mg^{2+} ion (metal B) at the polymerase active site, and the ion also chelates to a side-chain oxygen of catalytic aspartate Asp110 and two side-chain oxygens of Asp185 (**Figure 3-3**A). In addition, one oxygen from the terminal carboxylic acid group of **3.9a** form 2.38 or 3.17 Å salt bridges with the Nε and the Nη2 atoms of Arg72, respectively, whereas the other oxygen atom of the carboxylic acid group forms a 3.05 Å interaction with Lys65 (**Figure 3-3**B). Despite the differences in chemical structure, **3.9a** forms the same number of H-bond as **3.5e** and its total interacting surface area (433.6 Å^2) is also similar to that of the latter (**Table 3-2**). This implies that the methyl group in the acyclic chain of **3.5e** does not have much impact on the inhibitor binding at the active site of HIV-1 RT.
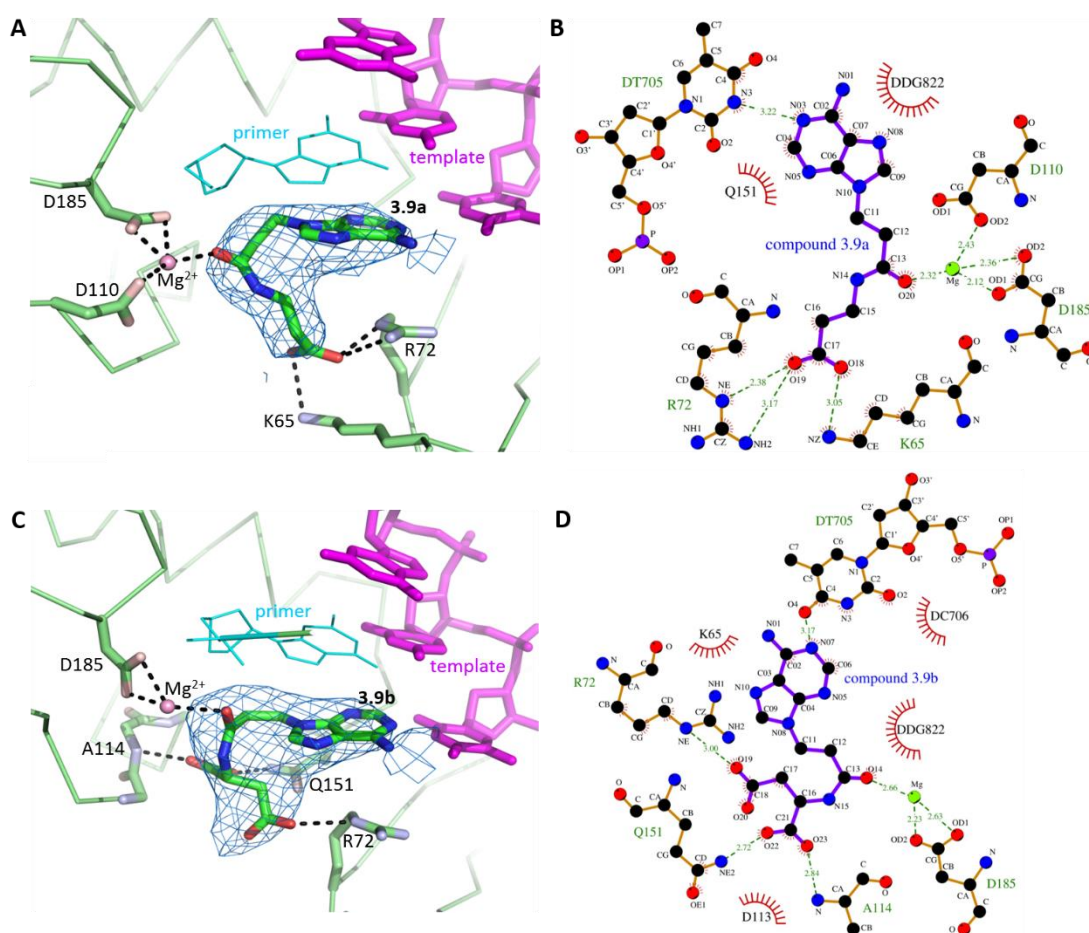
Figure 3-3. Binding of **3.9a** and **3.9b** with HIV-1 RT/DNA from crystal structures. (A) Polder map contoured at 5.9 σ showing the binding of **3.9a** at the polymerase active site in the 3.05 Å resolution crystal structure of HIV-1 RT/DNA/**3.9a** complex in the presence of Mg^{2+} (pink sphere). The template strand is magenta, and the primer 3′ terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (B) Ligplot representation of the interaction of **3.9a** with HIV-1 RT/DNA in the presence of Mg^{2+}. (C) Polder map contoured at 6.8 σ showing the binding of **3.9b** at

the polymerase active site in the 3.6 Å resolution crystal structure of HIV-1 RT/DNA/**3.9b** complex in the presence of Mg^{2+} (pink sphere). (D) Ligplot representation of the interaction of **3.9b** with HIV-1 RT/DNA in the presence of Mg^{2+}.

The crystal structure of RT/DNA/**3.9b** complex in the presence of Mg^{2+} was obtained at 3.6 Å. The adenine base of **3.9b** base pairs with the template thymine of the first base overhang, and metal chelation is also involved in this structure. One Mg^{2+} ion positioned at the metal B site chelates two oxygens from the catalytic aspartate Asp185, and the carbonyl oxygen of **3.9b** (**Figure 3-3**C). Although losing the interaction with Lys65, **3.9b** maintains contact with Arg72 by forming a 3.0 Å salt bridge. Furthermore, **3.9b** has one additional α-carboxylic acid group, when compared to **3.9a**, making it possible to engage in two additional H-bonds — a 2.84 Å H-bond with the backbone amide NH of Ala114 and a 2.72 Å H-bond with Nε2 of Gln151, the latter is also observed for **3.5e** (**Figure3-3**D). Compared with **3.5e**, **3.9b** has no methyl group in the two-carbon linker and it contains the amino acid D-Asp rather than L-Asp. The compound **3.9b** forms three H-bonds and its interaction surface area (440.8 Å^2) is higher than that of **3.5e** (**Table 3-2**). When tested at 10 mM, the two compounds did not display obvious RT inhibition, although compound **3.9a** displayed anti-HIV activity against wild type NL4.3 strain with an EC_{50} value of 65.52 μM, and compound **3.9b** exhibited an increased anti-HIV activity (24.08 and 62.33 μM for the NL4.3WT and BaL strain, respectively) (**Table 3-1**).

3.3 Conclusion

In an effort to improve compound stability, a series of acyclic nucleoside analogs were synthesized with an amide group instead of a phosphonoamidate moiety. X-ray crystallographic analysis demonstrated that these compounds bind into the active site of HIV-1 RT. However, their weak binding affinity resulted in a lack of RT inhibition and antiviral efficacy. Overall, these data highlight the importance of the phosphonate group in chelation of the metal ions.

3.4 Experimental section

Chemistry All reagents and solvents were purchased from commercial sources and were used as obtained. Moisture-sensitive reactions were carried out in oven-dried glassware under a nitrogen or argon atmosphere. ^{1}H and ^{13}C spectra were recorded on Bruker Avance 300 or 500 MHz spectrometer using tetramethylsilane as internal standard or referenced to the residual solvent signal. Splitting patterns are designated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), br(broad), and dd (doublet of doublets). Coupling constants are expressed in hertz (Hz). High-resolution mass spectra (HRMS) were obtained on a quadruple orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 μL/min and spectra were obtained in positive (or negative) ionization mode with a resolution of 15000 (FWHM)

using leucine enkephalin as lock mass. Pre-coated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ= 254 nm). Products were purified by flash chromatography on silica gel (35-70 μm, 60 Å). All final compounds were purified by preparative RP-HPLC (C18 Phenomenex Gemini column, 110 Å, 10 μm, 21.2 mm × 250 mm) using an eluent gradient of CH_3CN with 50 mmol TEAB as eluent buffer. Final compounds were isolated as their triethylammonium salts. Prior to biological evaluation, compounds were converted into the corresponding sodium salts via a sodium ion exchange resin. The exact concentration of the compounds was determined by the Lambert-Beer law. All tested compounds had a purity above 95 % as determined by HPLC analysis using the analytical method mentioned before.

Methyl (R)-2-methyl-3-(tosyloxy)propanoate (3.2) To an ice-cooled solution of methyl (*R*)-3-hydroxy-2-methylpropanoate (765 mg, 6.48 mmol) in DCM (15 mL) were added TsCl (1.236 g, 6.48 mmol), Et_3N (0.9 mL, 6.48 mmol) and DMAP (396 mg, 3.24 mmol). After stirring at room temperature for 1 h, the reaction was quenched with a saturated aqueous NH_4Cl solution and the organic layer was dried over Na_2SO_4. Evaporation of the solvent furnished the crude product as a yellow oil (1.764 g, 100 %) that was used for next step without further purification.

Methyl 3-(6-amino-9H-purin-9-yl)-2-methylpropanoate (3.3) A mixture of adenine (796 mg, 5.89 mmol) and cesium carbonate (1.24 g, 6.48 mmol) in dry DMF (15 mL) was stirred at 66 °C for 1 h. A solution of compound **3.2** (1.764 g, 6.48 mmol) in dry DMF (6 mL) was added in one portion and the mixture was stirred at 66 °C under a nitrogen atmosphere for 16 h. The mixture was evaporated *in vacuo* and the residue was purified by silica gel chromatography to give the title compound as a white solid (0.79 g, 57 %). ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.35 (s, 1H), 7.83 (s, 1H), 5.91 (s, 2H), 4.41-4.14 (m, 2H), 3.65 (s, 3H), 3.27-3.06 (m, 1H), 1.26 (d, *J* = 7.2 Hz, 3H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 174.55(s), 155.67 (s), 153.21 (s), 150.48 (s), 141.33 (s), 119.29(s), 52.24 (s), 46.21 (s), 39.99 (s), 15.26 (s). HRMS: $[M+H]^+$ calculated for $C_{10}H_{13}N_5O_2$, 236.1142; found 236.1139.

3-(6-Amino-9H-purin-9-yl)-2-methylpropanoic acid (3.4) The methyl ester **3.3** (500 mg, 2.125 mmol) was suspended in methanol (10 mL) and a 1 M NaOH solution (3 mL) was added dropwise. The resulting mixture was stirred for 3 h at room temperature. Then, the reaction mixture was neutralized with a 1N HCl solution until pH~7. MeOH was evaporated under reduced pressure. A 1 N HCl solution was added to the residue until pH~4. After cooling down in the fridge overnight, the precipitate was collected by filtration, yielding the title compound as a white powder (400 mg, 85 %). ^{1}H NMR (300 MHz, DMSO) δ 12.58 (s, 1H), 8.14 (s, 1H), 8.07 (s, 1H), 7.25 (s, 2H), 4.40-4.13 (m, 2H), 3.23-2.91 (m, 1H), 1.05 (d, *J* = 7.1 Hz, 3H). ^{13}C NMR (75 MHz, DMSO) δ 175.19 (s), 156.09 (s), 152.58 (s), 149.82 (s), 141.26 (s), 118.74 (s), 45.54 (s), 39.23 (s), 14.73 (s). HRMS: $[M+H]^+$ calculated for $C_9H_{11}N_5O_2$, 222.0985; found 222.0985.

(3-(6-Amino-9H-purin-9-yl)-2-methylpropanoyl)glycine (3.5a) To a -20 °C solution of the acid **3.4** (100 mg, 0.452 mmol) and EDCI (107 mg, 0.543 mmol) in dry DMF (1 mL), was added dropwise DIEA (90 μL, 0.543 mmol) under an argon atmosphere. After 20 min, a solution of HOAt (74 mg, 0.543 mmol) in dry DMF (1 mL) was added, and the resulting light-yellow solution was allowed to stir for an additional 20 min. Meanwhile, a suspension of glycine methyl ester hydrochloride (61 mg, 0.484 mmol) and DIEA (80 μL, 0.484 mmol) in dry DMF (1 mL) was prepared and added dropwise. The solution was then stirred overnight under argon and monitored by TLC. After consumption of the starting material, DMF was evaporated under reduced pressure and the crude residue was purified by silica gel chromatography (EA/MeOH, v/v, 30:1 to 20:1 to 10:1). The product obtained was dissolved in a 0.4 M NaOH (in MeOH/H_2O, 4:1, 5 mL) and stirred at room temperature under argon for 2-3 h. When TLC analysis indicated the disappearance of the starting material, the reaction mixture was neutralized by addition of TEAA (1 M, pH ~ 7.5). The solvent was removed under reduced pressure and the resulting residue was purified by silica gel chromatography (DCM/MeOH/H_2O, v/v/v, 10:1:0 to 10:1:0.1 to 5:1:0.2). The product was finally purified by preparative HPLC with a gradient of CH_3CN in 50 mM TEAB buffer (pH = 7.4) to yield the target compound as a colourless solid (49 mg, 39 % over two steps). ^{1}H NMR (300 MHz, D_2O) δ 7.91 (s, 1H), 7.83 (s, 1H), 4.14 (ddd, J = 19.8, 14.2, 7.3 Hz, 2H), 3.49 (dd, J = 59.3, 17.3 Hz, 2H), 2.90 (p, J = 6.2 Hz, 1H), 1.08 (d, J = 6.9 Hz, 3H). ^{13}C NMR (75 MHz, D_2O) δ 175.58 (s), 175.15 (s), 154.66 (s), 151.64 (s), 147.96 (s), 141.76 (s), 117.43 (s), 45.94 (s), 42.79 (s), 40.41 (s), 13.96 (s). HRMS: [M+H]$^+$ calculated for $C_{11}H_{15}N_6O_3$, 279.1206; found 279.1203.

3-(3-(6-Amino-9H-purin-9-yl)-2-methylpropanamido)propanoic acid (3.5b) This compound was obtained as a colourless solid (67 mg, 25 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.4** (200 mg, 0.904 mmol) and β-alanine methyl ester hydrochloride (135 mg, 0.967 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.03 (s, 1H), 7.82 (s, 1H), 4.27-4.04 (m, 2H), 3.21-3.10 (m, 2H), 2.89-2.65 (m, 1H), 2.25-1.95 (m, 2H), 1.01 (d, J = 6.9 Hz, 3H). ^{13}C NMR (75 MHz, MeOD) δ 178.73 (s), 175.66 (s), 157.30 (s), 153.69 (s), 150.70 (s), 142.98 (s), 119.94 (s), 59.96 (s), 47.63 (s), 42.25 (s), 37.50 (d, J = 7.2 Hz), 15.85 (s). HRMS: [M-H]$^-$ calculated for $C_{12}H_{16}N_6O_3$, 291.1211; found 291.1216.

(3-(6-Amino-9H-purin-9-yl)-2-methylpropanoyl)-D-alanine (3.5c) This compound was obtained as a colourless solid (56 mg, 42 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and D-alanine methyl ester hydrochloride (68 mg, 0.484 mmol). ^{1}H NMR (300 MHz, D_2O) δ 7.97 (d, J = 9.1 Hz, 1H), 7.85 (d, J = 8.7 Hz, 1H), 4.30-4.06 (m, 2H), 3.88 (dq, J = 30.8, 7.2 Hz, 1H), 2.96 − 2.84 (m, 1H), 1.14-0.83 (m, 6H). ^{13}C NMR (75 MHz, D_2O) δ 178.94 (d, J = 8.3 Hz), 174.29 (d, J = 14.0 Hz), 154.60 (d,

J = 3.8 Hz), 151.60 (d, *J* = 11.1 Hz), 147.87 (d, *J* = 8.3 Hz), 141.68 (d, *J* = 11.2 Hz), 117.35 (s), 50.47 (s), 50.16 (s), 46.18 (s), 45.69 (s), 40.55 (s), 40.28 (s), 17.08 (s), 16.74 (s), 14.17 (s), 13.64 (s). HRMS: [M+H]$^+$ calculated for $C_{12}H_{17}N_6O_3$, 293.1362; found 293.1358.

(3-(6-Amino-9H-purin-9-yl)-2-methylpropanoyl)-D-aspartic acid (3.5d) This compound was obtained as a colourless solid (113 mg, 37 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.4** (200 mg, 0.904 mmol) and D-aspartic acid dimethyl ester hydrochloride (191 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 8.11 (d, *J* = 2.2 Hz, 1H), 7.97 (d, *J* = 5.1 Hz, 1H), 4.45-4.12 (m, 3H), 3.12-2.89 (m, 1H), 2.42 (dtd, *J* = 26.0, 15.7, 7.5 Hz, 2H), 1.21-1.03 (m, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 178.09 (d, *J* = 19.5 Hz), 177.75 (d, *J* = 3.8 Hz), 174.63 (d, *J*= 12.8 Hz), 154.93 (s), 151.96 (d, *J* = 5.1 Hz), 148.38 (d, *J* = 2.6 Hz), 142.17 (d, *J* = 8.1 Hz), 117.70 (s), 52.78 (d, *J* = 21.2 Hz), 45.90 (d, *J*=7.5 Hz), 40.60 (d, *J*= 7.5 Hz), 39.60 (d, *J*=27 Hz), 14.27 (d, *J*= 44.2 Hz). HRMS: [M+H]$^+$ calculated for $C_{13}H_{16}N_6O_5$, 337.1255; found 337.1254.

(3-(6-Amino-9H-purin-9-yl)-2-methylpropanoyl)-L-aspartic acid (3.5e) This compound was obtained as a colourless solid (81 mg, 27 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.4** (200 mg, 0.904 mmol) and L-aspartic acid dimethyl ester hydrochloride (191 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 8.02 (d, *J* = 4.7 Hz, 1H), 7.91 (d, *J* = 4.6 Hz, 1H), 4.42-4.01 (m, 3H), 3.05-2.93 (m, 1H), 2.70-2.11 (m, 2H), 1.18-0.98 (m, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 178.36 (s), 177.91 (d, *J* = 8.0 Hz), 174.85 (d, *J* = 19.4 Hz), 154.84 (d, *J* = 3.9 Hz), 151.76 (d, *J* = 8.9 Hz), 148.18 (d, *J* = 7.1 Hz), 141.99 (d, *J* = 13.1 Hz), 117.65 (d, *J* = 1.5 Hz), 52.74 (s), 52.29 (s), 46.16 (s), 45.73 (s), 40.50 (d, *J* = 16.2 Hz), 39.35 (s), 38.89 (s), 14.36 (s), 13.66 (s). HRMS: [M-H]$^-$ calculated for $C_{13}H_{16}N_6O_5$, 335.1109; found 335.1087.

(3-(6-Amino-9H-purin-9-yl)-2-methylpropanoyl)-D-glutamic acid (3.5f) This compound was obtained as a colourless solid (91 mg, 29 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.4** (200 mg, 0.904 mmol) and D-glutamic acid dimethyl ester hydrochloride (205 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 7.97 (d, *J* = 14.1 Hz, 1H), 7.84 (d, *J* = 7.1 Hz, 1H), 4.30-4.05 (m, 2H), 3.99 − 3.72 (m, 1H), 3.03 − 2.89 (m, 1H), 2.11 − 1.83 (m, 2H), 1.62 − 1.30 (m, 2H), 1.13 (dd, *J* = 14.2, 6.8 Hz, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 181.76 (s), 181.28 (s), 178.19 (s), 177.86 (s), 175.03 (d, *J* = 4.7 Hz), 154.74 (d, *J* = 5.3 Hz), 151.75 (d, *J* = 14.6 Hz), 147.92 (d, *J* = 14.9 Hz), 141.75 (d, *J* = 14.5 Hz), 117.51 (s), 54.85 (d, *J* = 8.3 Hz), 46.35 (s), 45.79 (s), 40.59 (d, *J* = 14.8 Hz), 33.58 (s), 28.00 (d, *J* = 18.7 Hz), 14.32 (s), 13.64 (s). HRMS: [M+H]$^+$ calculated for $C_{14}H_{19}N_6O_5$, 351.1417; found 351.1406.

Methyl 3-(6-amino-9H-purin-9-yl)propanoate (3.7) Adenine (1 g, 7.4 mmol) was dissolved in dry DMF (10 mL), followed by addition of K$_2$CO$_3$ (1.023 g, 7.4 mmol) and methyl 3-bromopropanoate 3.6 (735 μL, 6.73 mmol). The reaction mixture was heated at 60 °C for 12 h. After

cooling to room temperature, the mixture was concentrated under reduced pressure and the residue was purified by silica gel chromatography (DCM/MeOH, v/v, 30:1 to 20:1 to 10:1) to yield the title compound as a white powder (1.06 g, 71 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.35 (s, 1H), 7.90 (s, 1H), 5.87 (s, 2H), 4.50 (t, J = 6.2 Hz, 2H), 3.67 (s, 3H), 2.94 (t, J = 6.2 Hz, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 171.02 (s), 155.66 (s), 153.18 (s), 150.05 (s), 141.35 (s), 119.90 (s), 52.17 (s), 39.60 (s), 34.10 (s).

3-(6-Amino-9H-purin-9-yl)propanoic acid (3.8) This compound was obtained as a white powder (414 mg, 88 %) according to the procedure used for the synthesis of compound **3.4**, starting from the ester **3.7** (500 mg, 2.26 mmol). ^{1}H NMR (300 MHz, DMSO) δ 12.44 (s, 1H), 8.09 (s, 1H), 8.04 (s, 1H), 7.17 (s, 2H), 4.29 (t, J = 6.7 Hz, 2H), 2.82 (t, J = 6.7 Hz, 2H). ^{13}C NMR (75 MHz, DMSO) δ 171.96 (s), 155.86 (s), 152.28 (s), 149.39 (s), 140.87 (s), 118.69 (s), 39.09 (s), 33.60 (s).

3-(3-(6-Amino-9H-purin-9-yl)propanamido)propanoic acid (3.9a) This compound was obtained as a colourless solid (68 mg, 27 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.8** (187 mg, 0.904 mmol) and β-alanine methyl ester hydrochloride (135 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 7.95 (s, 1H), 7.87 (s, 1H), 4.32 (t, J = 5.7 Hz, 2H), 3.16 (t, J = 6.8 Hz, 2H), 2.67 (t, J = 5.7 Hz, 2H), 2.12 (t, J = 6.8 Hz, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 179.63 (s), 171.91 (s), 154.78 (s), 152.00 (s), 148.21 (s), 141.95 (s), 117.75 (s), 40.17 (s), 36.37 (s), 36.27 (s), 35.38 (s). HRMS: [M+H]$^{+}$ calculated for C$_{11}$H$_{14}$N$_6$O$_3$, 279.1200; found 279.1201.

(3-(6-Amino-9H-purin-9-yl)propanoyl)-D-aspartic acid (3.9b) This compound was obtained as a colourless solid (72 mg, 25 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.8** (187 mg, 0.904 mmol) and D-aspartic acid dimethyl ester hydrochloride (191 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 7.88 (s, 2H), 4.37-4.17 (m, 3H), 2.72 (t, J = 6.5 Hz, 2H), 2.52-2.29 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 178.30 (s), 177.98 (s), 171.32 (s), 154.92 (s), 151.93 (s), 148.22 (s), 142.10 (s), 117.87 (s), 52.96 (s), 39.97 (s), 39.61 (s), 35.13 (s). HRMS: [M+H]$^{+}$ calculated for C$_{12}$H$_{14}$N$_6$O$_5$, 323.1098; found 323.1093.

(3-(6-Amino-9H-purin-9-yl)propanoyl)-L-aspartic acid (3.9c) This compound was obtained as a colourless solid (83 mg, 29 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.8** (187 mg, 0.904 mmol) and L-aspartic acid dimethyl ester hydrochloride (191 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 7.97 (s, 1H), 7.94 (s), 4.35 (t, J = 6.6 Hz, 2H), 4.27 (dd, J = 9.5, 4.0 Hz, 1H), 2.76 (t, J = 6.6 Hz, 2H), 2.54-2.32 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 178.34 (s), 178.03 (s), 171.41 (s), 155.01 (s), 151.98 (s), 148.31 (s), 142.17 (s), 117.96 (s), 52.93 (s), 40.00 (s), 39.54 (s), 35.14 (s). HRMS: [M+H]$^{+}$ calculated for C$_{12}$H$_{14}$N$_6$O$_5$, 323.1098; found 323.1092.

(3-(6-Amino-9H-purin-9-yl)propanoyl)-D-glutamic acid (3.9d) This compound was obtained as a colourless solid (66 mg, 22 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.8** (187 mg, 0.904 mmol) and D-glutamic acid dimethyl ester hydrochloride (205 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 8.03 (s, 1H), 7.94 (s), 4.40 (t, *J*=6.5 Hz, 2H), 3.92 (t, *J*=6.0Hz, 1H), 2.80 (t, *J*=6.5 Hz, 2H), 1.92-1.64 (m, 4H). ^{13}C NMR (75 MHz, D$_2$O) δ 181.50 (s), 178.09 (s), 171.55 (s), 154.66 (s), 152.04 (s), 147.96 (s), 142.07 (s), 117.96 (s), 55.05 (s), 40.12 (s), 35.24 (s), 33.87 (s), 28.33 (s). HRMS: [M+H]$^+$ calculated for C$_{13}$H$_{16}$N$_6$O$_5$, 337.1255; found 337.1253.

(3-(6-Amino-9H-purin-9-yl)propanoyl)-L-glutamic acid (3.9e) This compound was obtained as a colourless solid (53 mg, 17 % over two steps) according to the procedure used for the synthesis of compound **3.5a**, starting from the acid **3.8** (187 mg, 0.904 mmol) and L-glutamic acid dimethyl ester hydrochloride (205 mg, 0.967 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 8.13 (s, 1H), 8.00 (s, 1H), 4.45 (dd, *J* = 12.2, 5.9 Hz, 2H), 3.95-3.86 (m, 1H), 2.90-2.73 (m, 2H), 1.88-1.59 (m, 4H). ^{13}C NMR (75 MHz, D$_2$O) δ 180.81 (s), 177.98 (s), 171.61 (s), 154.85 (s), 151.69 (s), 148.22 (s), 142.09 (s), 117.87 (s), 54.71 (s), 40.11 (s), 35.18 (s), 33.11 (s), 27.79 (s). HRMS: [M+H]$^+$ calculated for C$_{13}$H$_{16}$N$_6$O$_5$, 337.1255; found 337.1251.

Methyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)glycinate (3.10a) To a -20 °C solution of the acid (100 mg, 0.452 mmol) and EDCI (107 mg, 0.543 mmol) in dry DMF (1 mL), was added dropwise DIEA (90 μL, 0.543 mmol) under an argon atmosphere. After 20 min, a solution of HOAt (74 mg, 0.543 mmol) in dry DMF (1 mL) was added, and the resulting light-yellow solution was allowed to stir for an additional 20 min. Meanwhile, a suspension of glycine methyl ester hydrochloride (61 mg, 0.484 mmol) and DIEA (80 μL, 0.484 mmol) in dry DMF (1 mL) was prepared and added dropwise. The solution was then stirred overnight under Ar and the reaction was monitored by TLC. After consumption of the starting material, DMF was evaporated under reduced pressure and the residue was purified by silica gel chromatography (EA/MeOH, v/v, 30:1 to 20:1 to 10:1). The product was further purified by preparative HPLC with a gradient of MeOH in milliQ water to yield the target compound as a colourless solid (84 mg, 64 %). ^{1}H NMR (300 MHz, MeOD) δ 8.24 (s, 1H), 8.03 (s, 1H), 4.36 (ddd, *J* = 19.1, 13.9, 7.3 Hz, 2H), 3.86 (s, 2H), 3.68 (s, 3H), 3.19 − 3.00 (m, 1H), 1.24 (d, *J* = 7.0 Hz, 3H). ^{13}C NMR (75 MHz, DMSO) δ 173.97 (s), 170.31 (s), 156.06 (s), 152.51 (s), 141.22 (s), 51.80 (s), 45.66 (s), 40.69 (s), 39.18 (s), 15.70 (s). HRMS: [M+H]$^+$ calculated for C$_{12}$H$_{17}$N$_6$O$_3$, 293.1362; found 293.1361.

Methyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)-D-alaninate (3.10b) This compound was obtained as a colourless solid (79 mg, 57 %) according to the procedure used for the synthesis of compound **3.10a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and D-alanine methyl ester

hydrochloride (68 mg, 0.484 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.18 (d, J = 0.9 Hz, 1H), 7.94 (s, 1H), 4.43 – 4.31 (m, 1H), 4.31 – 4.16 (m, 2H), 3.61 (d, J = 22.4 Hz, 3H), 3.12 – 2.89 (m, 1H), 1.29 – 1.12 (m, 6H). ^{13}C NMR (75 MHz, MeOD) δ 175.92 (d, J = 3.4 Hz), 174.34 (d, J = 12.6 Hz), 157.30 (s), 153.66 (d, J = 7.5 Hz), 150.60 (s), 143.17 (d, J = 9.6 Hz), 119.90 (s), 52.70 (s), 49.47 (s), 47.60 (d, J = 3.6 Hz), 41.89 (s), 17.12 (s), 15.75 (d, J = 9.9 Hz). HRMS: [M+H]$^+$ calculated for $C_{13}H_{19}N_6O_3$, 307.1519; found 307.1511.

Dimethyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)-D-aspartate (3.10c) This compound was obtained as a colourless solid (98 mg, 60 %) according to the procedure used for the synthesis of compound **3.10a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and D-aspartic acid dimethyl ester hydrochloride (96 mg, 0.484 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.21 (s, 1H), 7.98 (d, J = 2.7 Hz, 1H), 4.77 – 4.58 (m, 1H), 4.51 – 4.17 (m, 2H), 3.63 (dd, J = 17.8, 13.7 Hz, 6H), 3.23 – 3.00 (m, 1H), 2.79 (ddd, J = 24.5, 16.8, 5.8 Hz, 2H), 1.21 (t, J = 7.3 Hz, 3H). ^{13}C NMR (75 MHz, MeOD) δ 175.83 (d, J = 5.6 Hz), 172.20 (t, J = 8.0 Hz), 157.22 (s), 153.65 (d, J = 2.7 Hz), 150.54 (s), 143.09 (s), 119.88 (s), 53.03 (d, J = 2.7 Hz), 52.43 (s), 50.11 (d, J = 10.5 Hz), 47.54 (s), 41.80 (d, J = 7.1 Hz), 36.51 (d, J = 5.7 Hz), 15.77 (d, J = 6.6 Hz). HRMS: [M+H]$^+$ calculated for $C_{15}H_{21}N_6O_5$, 365.1573; found 365.1564.

Dimethyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)-D-glutamate (3.10d) This compound was obtained as a colourless solid (88 mg, 51 %) according to the procedure used for the synthesis of compound **3.10a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and D-glutamic acid dimethyl ester hydrochloride (102 mg, 0.484 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.20 (d, J = 4.3 Hz, 1H), 7.97 (d, J = 7.4 Hz, 1H), 4.39 (dt, J = 16.0, 8.1 Hz, 2H), 4.22 (dd, J = 13.9, 5.1 Hz, 1H), 3.69 – 3.55 (m, 6H), 3.20 – 3.02 (m, 1H), 2.33 (t, J = 7.3 Hz, 1H), 2.16 – 1.94 (m, 2H), 1.79 (tdd, J = 19.5, 12.6, 6.6 Hz, 1H), 1.27 – 1.14 (m, 3H). ^{13}C NMR (75 MHz, MeOD) δ 176.07 (d, J = 3.4 Hz), 174.39 (d, J = 8.0 Hz), 173.21 (d, J = 16.6 Hz), 157.22 (d, J = 4.0 Hz), 153.65 (d, J = 7.4 Hz), 150.52 (s), 143.05 (s), 119.88 (s), 52.75 (d, J = 9.9 Hz), 52.19 (s), 47.57 (s), 41.85 (d, J = 6.3 Hz), 30.74 (d, J = 14.4 Hz), 27.31 (s), 15.83 (d, J = 12.9 Hz). HRMS: [M+H]$^+$ calculated for $C_{16}H_{23}N_6O_5$, 379.1730; found 379.1722.

Diisopropyl D-aspartate hydrochloride To a suspension of D-aspartic acid (1 g, 7.5 mmol) in anhydrous *iso*-propanol (30 mL) was added dropwise thionyl chloride (3.8 mL, 52 mmol) at 0 °C under an argon atmosphere. The mixture was allowed to stir at room temperature and then refluxed for 8 h. After evaporation of the solvents, the solid residue was triturated with diethyl ether. The white solid product was then filtered and washed with diethyl ether yielding the title compound (1.69 g, 89 %). ^{1}H NMR (300 MHz, DMSO) δ 8.73 (s, 3H), 4.90 (ddt, J = 18.7, 12.5, 6.2 Hz, 2H), 4.18 (t, J = 5.6 Hz, 1H), 2.91 (qd, J = 17.3, 5.6 Hz, 2H), 1.16 (td, J = 5.8, 3.8 Hz, 12H). ^{13}C NMR (75 MHz,

DMSO) δ 168.64 (s), 167.73 (s), 69.93 (s), 68.58 (s), 48.47 (s), 34.38 (s), 21.40 (dd, J = 12.4, 5.3 Hz). HRMS: [M+H]$^+$ calculated for $C_{10}H_{20}NO_4$, 218.1392; found 218.1390.

Diisopropyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)-D-aspartate (3.10e) This compound was obtained as a colourless solid (80 mg, 42 %) according to the procedure used for the synthesis of compound **3.10a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and diisopropyl D-aspartate hydrochloride (123 mg, 0.484 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.26 (d, J = 25.5 Hz, 1H), 7.83 (d, J = 1.7 Hz, 1H), 6.28 (d, J = 10.2 Hz, 2H), 4.97 (dddt, J = 43.5, 18.7, 12.5, 6.2 Hz, 2H), 4.82 – 4.66 (m, 1H), 4.54 – 4.17 (m, 2H), 3.08 (dd, J = 9.5, 6.2 Hz, 1H), 2.95 – 2.47 (m, 2H), 1.30 – 1.12 (m, 15H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.53 (d, J = 11.0 Hz), 170.23 (dd, J = 17.5, 5.9 Hz), 155.72 (s), 152.87 (d, J = 11.1 Hz), 150.01 (s), 141.54 (d, J = 5.0 Hz), 119.55 (s), 69.79 (d, J = 5.4 Hz), 68.85 (d, J = 5.1 Hz), 49.10 (s), 48.60 (s), 46.49 (d, J = 13.0 Hz), 41.06 (s), 36.45 (d, J = 8.0 Hz), 22.11 – 21.49 (m), 15.95 (d, J = 18.7 Hz). HRMS: [M+H]$^+$ calculated for $C_{19}H_{29}N_6O_5$, 421.2199; found 421.2188.

Dibutyl D-aspartate hydrochloride This compound was obtained as a colourless solid (2 g, 95 %) according to the procedure used for the synthesis of diisopropyl D-aspartate hydrochloride, starting from D-aspartic acid (1 g, 7.5 mmol) and anhydrous *n*-butanol (30 mL). ^{1}H NMR (300 MHz, DMSO) δ 8.73 (s, 3H), 4.27 (t, J = 5.4 Hz, 1H), 4.13 – 3.99 (m, 4H), 2.98 (qd, J = 17.3, 5.5 Hz, 2H), 1.58 – 1.46 (m, 4H), 1.28 (dq, J = 14.3, 7.1 Hz, 4H), 0.84 (td, J = 7.4, 2.9 Hz, 6H). ^{13}C NMR (75 MHz, DMSO) δ 169.32 (s), 168.41 (s), 65.68 (s), 64.69 (s), 48.53 (s), 34.24 (s), 30.08 (d, J = 8.2 Hz), 18.61 (d, J = 10.4 Hz), 13.65 (d, J = 4.3 Hz). HRMS: [M+H]$^+$ calculated for $C_{12}H_{24}NO_4$, 246.1705; found 246.1709.

Dibutyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)-D-aspartate (3.10f) This compound was obtained as a colourless solid (75 mg, 37 %) according to the procedure used for the synthesis of compound **3.10a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and dibutyl D-aspartate hydrochloride (136 mg, 0.484 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.24 (d, J = 31.2 Hz, 1H), 7.83 (d, J = 2.6 Hz, 1H), 4.82 (td, J = 8.4, 4.7 Hz, 1H), 4.51 – 4.36 (m, 1H), 4.30 – 4.18 (m, 1H), 4.15 – 3.96 (m, 4H), 3.14 – 3.02 (m, 1H), 3.00 – 2.79 (m, 2H), 1.56 (ddd, J = 18.8, 10.4, 6.1 Hz, 4H), 1.38 – 1.19 (m, 7H), 0.88 (ddd, J = 9.0, 7.4, 4.5 Hz, 6H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 174.12 (d, J = 18.9 Hz), 171.76 – 171.11 (m), 156.28 (s), 153.25 (d, J = 15.1 Hz), 150.43 (s), 142.02 (s), 119.97 (s), 78.14 (s), 77.72 (s), 77.29 (s), 66.36 (d, J = 4.3 Hz), 65.58 (s), 49.46 (s), 48.95 (s), 47.05 (d, J = 17.6 Hz), 41.51 (d, J = 6.8 Hz), 36.66 (d, J = 10.0 Hz), 31.04 (d, J = 4.7 Hz), 19.56 (d, J = 4.6 Hz), 16.55 (s), 16.23 (s), 14.22 (s). HRMS: [M+H]$^+$ calculated for $C_{21}H_{33}N_6O_5$, 449.2512; found 449.2504.

Dipentyl D-aspartate hydrochloride This compound was obtained as a light-yellow solid (2.2 g, 95 %) according to the procedure used for the synthesis of diisopropyl D-aspartate hydrochloride,

starting from D-aspartic acid (1 g, 7.5 mmol) and anhydrous *n*-pentanol (30 mL). ^{1}H NMR (300 MHz, DMSO) δ 8.67 (s, 3H), 4.29 (t, J = 5.4 Hz, 1H), 4.14 – 3.98 (m, 4H), 2.97 (dd, J = 10.3, 5.5 Hz, 2H), 1.61 – 1.48 (m, 4H), 1.25 (d, J = 2.2 Hz, 8H), 0.84 (t, J = 6.4 Hz, 6H). ^{13}C NMR (75 MHz, DMSO) δ 169.19 (s), 168.31 (s), 65.87 (s), 64.88 (s), 48.42 (s), 34.13 (s), 27.50 (dd, J = 16.3, 9.5 Hz), 21.74 (d, J = 3.4 Hz), 13.83 (s). HRMS: [M+H]$^+$ calculated for $C_{14}H_{28}NO_4$, 274.2018; found 274.2015.

Dipentyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)-D-aspartate (3.10g) This compound was obtained as a colourless solid (88 mg, 41 %) according to the procedure used for the synthesis of compound **3.10a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and dipentyl D-aspartate hydrochloride (150 mg, 0.484 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.22 (d, J = 35.3 Hz, 1H), 7.83 (d, J = 0.9 Hz, 1H), 4.84 (dd, J = 8.2, 3.9 Hz, 1H), 4.50 – 4.37 (m, 1H), 4.30 – 4.18 (m, 1H), 4.15 – 3.92 (m, 4H), 3.16 – 3.02 (m, 1H), 3.00 – 2.80 (m, 2H), 1.57 (dq, J = 14.5, 7.1 Hz, 3H), 1.35 – 1.19 (m, 11H), 0.88 (t, J = 6.1 Hz, 5H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.72 (s), 173.45 (s), 170.75 (dd, J = 11.2, 4.0 Hz), 155.84 (s), 152.73 (d, J = 14.6 Hz), 149.82 (s), 141.30 (d, J = 5.6 Hz), 119.37 (s), 66.02 (d, J = 4.5 Hz), 65.25 (s), 48.90 (s), 48.37 (s), 46.42 (d, J = 18.9 Hz), 40.86 (d, J = 8.4 Hz), 36.08 (d, J = 9.8 Hz), 27.97 (dd, J = 19.7, 5.5 Hz), 22.20 (d, J = 2.2 Hz), 15.96 (s), 15.63 (s), 13.89 (s). HRMS: [M+H]$^+$ calculated for $C_{23}H_{37}N_6O_5$, 477.2825; found 477.2817.

Diisopentyl D-aspartate hydrochloride To a suspension of D-aspartic acid (1 g, 7.5 mmol) in anhydrous amyl alcohol (30 mL) was added thionyl chloride (3.8 mL, 52 mmol) dropwise at 0 °C under an argon atmosphere. The mixture was allowed to stir at room temperature for 12 h. The suspension was then heated at 50 °C until a clear solution was obtained. After evaporation, the crude yellow liquid was triturated with hexane and kept at -78 °C overnight. A jellytype white precipitate was obtained and the hexane was immediately decanted carefully under cold condition. Hexane was added again and kept at -78 °C until a precipitate was formed, and the above process was repeated several times to remove the impurities. The combined hexane layers were evaporated to one third of the original volume and kept again at -78 °C and the above process was repeated to increase the final yield. Finally, the white solid product was washed several times with diethyl ether to obtain as hydrochloride salt (810 mg, 35 %). ^{1}H NMR (300 MHz, DMSO) δ 8.69 (s, 3H), 4.32 (t, J = 5.4 Hz, 1H), 4.22 – 4.06 (m, 4H), 3.00 (dd, J = 9.2, 5.5 Hz, 2H), 1.64 (tt, J = 13.1, 6.7 Hz, 2H), 1.48 (dd, J = 13.5, 6.7 Hz, 4H), 0.93 – 0.83 (m, 12H). ^{13}C NMR (75 MHz, DMSO) δ 169.20 (s), 168.30 (s), 64.29 (s), 63.37 (s), 48.42 (s), 36.61 (d, J = 10.0 Hz), 34.13 (s), 24.31 (d, J = 18.6 Hz), 22.20 (d, J = 9.4 Hz). HRMS: [M+H]$^+$ calculated for $C_{14}H_{28}NO_4$, 274.2018; found 274.2013.

Diisopentyl (3-(6-amino-9H-purin-9-yl)-2-methylpropanoyl)-D-aspartate (3.10h) This compound was obtained as a colourless solid (81 mg, 38 %) according to the procedure used for the synthesis of compound **3.10a**, starting from the acid **3.4** (100 mg, 0.452 mmol) and diisopentyl D-

aspartate hydrochloride (150 mg, 0.484 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.27 (d, J = 19.5 Hz, 1H), 7.79 (d, J = 4.7 Hz, 1H), 6.00 (d, J = 8.8 Hz, 2H), 4.84 − 4.67 (m, 1H), 4.51 − 4.30 (m, 2H), 4.10 (dd, J = 14.0, 7.0 Hz, 4H), 3.09-2.99 (m, 1H), 2.97-2.83 (m, 2H), 1.63 − 1.55 (m, 2H), 1.46 (dd, J = 6.8, 3.6 Hz, 4H), 1.23 (dd, J = 14.4, 7.0 Hz, 3H), 0.87 (dd, J = 3.8, 1.9 Hz, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.98 (d, J = 14.3 Hz), 171.52 (s), 171.50 − 170.87 (m), 156.13 (s), 153.46 (d, J = 9.6 Hz), 150.61 (s), 142.09 (d, J = 4.2 Hz), 120.14 (s), 65.25 (d, J = 6.6 Hz), 64.45 (d, J = 2.7 Hz), 49.13 (d, J = 38.6 Hz), 47.07 (d, J = 19.9 Hz), 41.62 (d, J = 3.6 Hz), 37.68 (dd, J = 6.7, 4.0 Hz), 36.64 (d, J = 8.9 Hz), 25.60 (s), 22.98 (d, J = 6.4 Hz), 16.40 (d, J= 23.3 Hz). HRMS: [M+H]$^+$ calculated for C$_{23}$H$_{37}$N$_6$O$_5$, 477.2825; found 477.2812.

HIV-1 RT inhibition assay

This assay was developed based on the gel analysis of primer extension by DNA polymerase.[18] RT reaction was carried out using a primer labeled with Cy5 flurophore (5′-Cy5-CAGGAAACAGCTATGAC) and a template (3'-GTCCTTTGTCGATACTG**TTTTTTT**-5'). Both oligos were purchased from Integrated DNA Technologies (IDT, Leuven). EFV was used as positive control. The enzyme reaction mixture contained 125 nM primer-template complex, 50 mM Tris−HCl pH 8.3, 3 mM MgCl$_2$, 10 mM DTT, EFV (0.5, 5 and 50 μM) or synthesized compounds (1, 5 and 10 mM), and 0.06 μg RT (RT139A in house enzyme) per μL reaction. The primer and template were pre-annealed at 1:2 molar ratio by heating up at 95 °C for 5 min followed by slow cooling down to room temperature. The reaction mixture without dATP was pre-incubated at 37 °C for 20 min. The reaction was initiated by adding the dATP and quenched after 10 min by adding a double volume of quenching buffer (90 % formamide, 50 mM EDTA, and 0.05 % orange G) and heated at 95 °C for 5 min. The results were analyzed by separation on a 15 % denaturing polyacrylamide gel. Gel electrophoresis were performed according to the protocol described previously.[19] Gel bands were visualized using the Typhoon FLA 9500 imaging system (GE Healthcare). The images were processed using ImageQuant TL v8.1.0.0 (GE Healthcare).

HIV antiviral assay

The anti-HIV activity of synthesized compounds was tested in TZM-bl cells[14] which were seeded in transparent 96-well plates at 1×10^4 cells per well in DMEM (Dulbecco's modified Eagle's medium; Life Technologies, Waltham, MA, USA) with 10 % fetal bovine serum (FBS) and 10 mM HEPES. Subsequently, compounds were added, and the cell/compound mixture was incubated at 37 °C. After 30 min, the virus (HIV-1 X4 NL4.3 and HIV-1 R5 BaL) was added at 100 pg p24 HIV-1 Ag per well. After 48 h of incubation, the assay plates were analyzed. For the analysis, steadylite plus substrate solution (PerkinElmer, Waltham, MA, USA) was added to the assay plates. The luminescent signal of the lysed cell suspension was analyzed in white 96-well plates on a SpectraMax L

luminescence microplate reader (Molecular Devices, Sunnyvale, CA, USA) after a 10 min incubation period in the dark. Luciferase activity induced by HIV-1 Tat protein expression was measured as an assessment of the amount of HIV replication.

For experimental details of **RT expression, RT/DNA cross-linking, purification & Crystallization and structure determination** in this chapter, please refer to Experimental section of Chapter 2.

3.5 References

1. Xavier RUIZ, F.; Arnold, E. Evolving understanding of HIV-1 reverse transcriptase structure, function, inhibition, and resistance. *Current Opinion in Structural Biology* **2020,** 61, 113-123.
2. Tsibris, A. M.; Hirsch, M. S. Antiretroviral therapy in the clinic. *J. Virol.* **2010,** 84, 5458-5464.
3. *UNAIDS Global HIV & AIDS statistics - 2020 fact sheet.* http:// www.unaids.org/en/resources/fact-sheet (accessed Aug 11, 2021).
4. Balzarini, J.; Das, K.; Bernatchez, J. A.; Martinez, S. E.; Ngure, M.; Keane, S.; Ford, A.; Maguire, N.; Mullins, N.; John, J.; Kim, Y.; Dehaen, W.; Vande Voorde, J.; Liekens, S.; Naesens, L.; Gotte, M.; Maguire, A. R.; Arnold, E. Alpha-carboxy nucleoside phosphonates as universal nucleoside triphosphate mimics. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** 112, 3475-3480.
5. Das, K.; Balzarini, J.; Miller, M. T.; Maguire, A. R.; DeStefano, J. J.; Arnold, E. Conformational States of HIV-1 Reverse Transcriptase for Nucleotide Incorporation vs Pyrophosphorolysis-Binding of Foscarnet. *Acs Chem Biol* **2016,** 11, 2158-64.
6. Mullins, N. D.; Maguire, N. M.; Ford, A.; Das, K.; Arnold, E.; Balzarini, J.; Maguire, A. R. Exploring the role of the alpha-carboxyphosphonate moiety in the HIV-RT activity of alpha-carboxy nucleoside phosphonates. *Org Biomol Chem* **2016,** 14, 2454-65.
7. Balzarini, J.; Ford, A.; Maguire, N. M.; John, J.; Das, K.; Arnold, E.; Dehaen, W.; Maguire, A. Alpha-carboxynucleoside phosphonates: direct-acting inhibitors of viral DNA polymerases. *Future Med Chem* **2019,** 11, 137-154.
8. Adelfinskaya, O.; Terrazas, M.; Froeyen, M.; Marliere, P.; Nauwelaerts, K.; Herdewijn, P. Polymerase-catalyzed synthesis of DNA from phosphoramidate conjugates of deoxynucleotides and amino acids. *Nucleic Acids Res.* **2007,** 35, 5060-5072.
9. Maiti, M.; Michielssens, S.; Dyubankova, N.; Maiti, M.; Lescrinier, E.; Ceulemans, A.; Herdewijn, P. Influence of the nucleobase and anchimeric assistance of the carboxyl acid groups in the hydrolysis of amino acid nucleoside phosphoramidates. *Chem. Eur. J.* **2012,** 18, 12.
10. Michielssens, S.; Maiti, M.; Maiti, M.; Dyubankova, N. Reactivity of Amino Acid Nucleoside Phosphoramidates: A Mechanistic Quantum Chemical Study. *J. Phys. Chem. A* **2012,** 116, 644-652.
11. Cherney, R. J.; Wang, L. Efficient Mitsunobu reactions with N-phenylfluorenyl or N-trityl serine esters. *J Org Chem* **1996,** 61, 2544-2546.
12. Guo, H. M.; Wu, Y. Y.; Niu, H. Y.; Wang, D. C.; Qu, G. R. Synthesis of Acyclic Nucleosides with a Chiral Amino Side Chain by the Mitsunobu Coupling Reaction. *J Org Chem* **2010,** 75, 3863-3866.
13. Maiti, M.; Maiti, M.; Rozenski, J.; De Jonghe, S.; Herdewijn, P. Aspartic acid based nucleoside phosphoramidate prodrugs as potent inhibitors of hepatitis C virus replication. *Org. Biomol. Chem.* **2015,** 13, 5158-5174.
14. Montefiori, D. C. Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol. Biol.* **2009,** 485, 395-405.
15. Das, K.; Martinez, S. E.; Bauman, J. D.; Arnold, E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol* **2012,** 19, 253-9.
16. Sarafianos, S. G.; Pandey, V. N.; Kaushik, N.; Modak, M. J. Site-directed mutagenesis of arginine 72 of HIV-1 reverse transcriptase. Catalytic role and inhibitor sensitivity. *J Biol Chem* **1995,** 270, 19729-35.
17. Das, K.; Bandwar, R. P.; White, K. L.; Feng, J. Y.; Sarafianos, S. G.; Tuske, S.; Tu, X.; Clark, A. D., Jr.; Boyer, P. L.; Hou, X.; Gaffney, B. L.; Jones, R. A.; Miller, M. D.; Hughes, S. H.; Arnold, E. Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *J. Biol. Chem.* **2009,** 284, 35092-100.
18. Creighton, S.; Bloom, L. B.; Goodman, M. F. Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol.* **1995,** 262, 232-256.
19. O'Flaherty, D. K.; Guengerich, F. P. Steady-state kinetic analysis of DNA polymerase single-nucleotide incorporation products. *Curr Protoc Nucleic Acid Chem.* **2014,** 59, 7.21.1-7.21.13.

Chapter 4

Exploring the dNTP binding site of HIV-1 reverse transcriptase for inhibitor design

This chapter is based on the following publication:

Weijie Gu, Sergio Martinez, Abhimanyu K. Singh, Hoai Nguyen, Jef Rozenski, Dominique Schols, Piet Herdewijn, Kalyan Das,* Steven De Jonghe* Exploring the dNTP Binding Site of HIV-1 Reverse Transcriptase for Inhibitor Design. *European Journal of Medicinal Chemistry*, **2021**, 225，113785. https://doi.org/10.1016/j.ejmech.2021.113785.

4.1 Introduction

The latest statistics from UNAIDS demonstrated that AIDS, caused by HIV, is still a prevalent disease, especially in low-income countries.[1] There are two main types of HIV strains, called HIV-1 and HIV-2. HIV-1 is more virulent, more efficiently transmitted and is mainly responsible for the global AIDS pandemic.[2] Nowadays, there are more than 30 anti-HIV drugs approved by FDA. Based on their target in different key stages of HIV life cycle, they can be divided into four classes, namely entry/fusion inhibitors, RTIs, integrase inhibitors (INIs) and protease inhibitors (PIs).[3-4] The enzyme RT is responsible for viral DNA synthesis during HIV replication, and hence plays a central role in the HIV life cycle.

The approved RTIs used for HIV/AIDS treatment are divided into two classes based on their mechanisms of RT inhibition. The first class is known as NRTIs, while the other type is classified as NNRTIs. An NRTI must be phosphorylated into its biologically active form via cellular kinases and binds at the polymerase active site of RT as a dNTP mimic. Typically, after incorporation at the 3'-end of the DNA primer strand, an NRTI, that lacks the 3'-hydroxyl group, is unable to initiate the nucleophilic attack on the incoming dNTP, and therefore acts as a chain terminator. In contrast, NNRTIs bind to an allosteric hydrophobic pocket, also termed as NNRTI-binding pocket (NNIBP), adjacent to the polymerase active site.

Nowadays, HIV infected patients can receive a combination therapy, also known as antiretroviral therapy (ART)[5] in which NRTIs and NNRTIs are both important components. However, long term usage of drugs often leads to the emergence of drug resistance, and therefore, the search for new drugs which are effective against existing drug-resistant mutants and act by distinct mechanisms of inhibition is of urgent need. A new drug candidate should have a novel mode of binding such that it can circumvent drug resistance. RT has been investigated as target for new classes of drugs, like RNase H inhibitors[6-9] and different types of NcRTIs such as INDOPYs[10-12] and α-CNPs.[13-17]

The α-CNPs, such as T-α-CNP (Chapter 2, Figure 2-1), compete with the natural nucleotide substrate dTTP and thereby inhibit the RT polymerase activity, without being incorporated to the growing DNA strand. X-ray crystallographic studies revealed that α-CNPs mimic dNTP binding to RT with the nucleobase and cyclopentyl moieties oriented in a manner analogous to the nucleobase and the 2'-deoxyribose moieties of the natural substrate, respectively, with the carboxymethylphosphonate group chelating catalytic Mg^{2+} ions at the polymerase active site.[14]

Recently, we have reported a series of novel TFV analogs, in which the phosphonate moiety of TFV was conjugated with various amino acids as a phosphonoamidate.[18] Those compounds were designed as dNTP analogs, in which the amino acid parts were expected to mimic triphosphate part

of a dNTP and hence they would not require metabolic phosphorylation for binding to HIV-1 RT. These TFV-amino acid conjugates were evaluated for possible incorporation into a growing DNA strand, catalyzed by HIV-1 RT and for their ability to function as inhibitors of RT polymerization activity. TFV was most efficiently incorporated from L-Glu-TFV (Chapter 2, **2.4d**) and hence functions as an NRTI, whereas no incorporation was observed for L-Met-TFV (Chapter 2, **2.4e**); L-Met-TFV acted as an NcRTI. Crystal structures of these TFV-amino acid conjugates were determined in complex with HIV-1 RT/dsDNA.

Despite the promise of these TFV-amino acid conjugates as NcRTI, the chemical and metabolic lability of the phosphonoamidate linker hampers further profiling, as mentioned in Chapter 3. Therefore, in this study, we embarked on the design and synthesis of a novel series of ANP analogs (**Figure 4-1**) lacking a phosphonoamidate moiety. Instead, these new derivatives contain either an amine (compounds **4.1-4.4** and **4.7**) or ether bridge (compounds **4.5-4.6**) for making the compounds chemically stable. In order to exploit possible metal chelation, different chelating moieties (such as a phosphonate or carboxylic acid moiety) were introduced. Seven stable and novel ANPs were designed that could potentially form additional interactions at the polymerase active site of HIV-1 RT. Moreover, variation of the chain length and stereochemistry at chiral centers allowed to investigate their impact on the mode of binding to HIV-1 RT. New synthetic routes to get access to these structurally diverse ANPs are elaborated. To understand their binding mode, crystal structures of seven compounds in complexes with HIV-1 RT/dsDNA were determined.

4.1 (RMC-230)

4.2 (RMC-233): (*R*)-enantiomer, n = 2
4.3 (RMC-247): (*S*)-enantiomer, n = 2
4.4 (RMC-257): (*S*)-enantiomer, n = 3

4.5 (RMC-259): n = 2
4.6 (RMC-264): n = 3

4.7 (RMC-282)
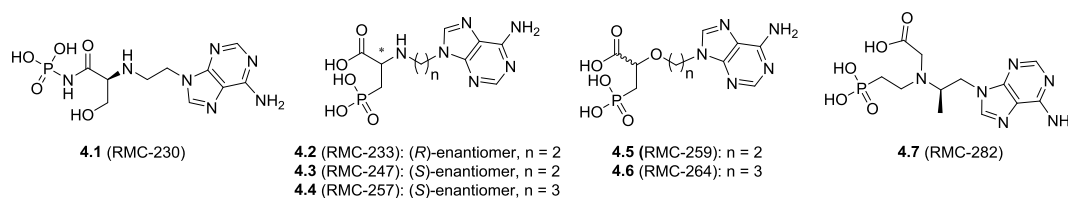
Figure 4-1. Designed and synthesized ANPs in this study

4.2 Results and discussion

4.2.1 Chemistry

The synthesis of intermediates **4.10a** and **4.10b** is depicted in **Scheme 4-1** according to a reported procedure.[19] Briefly, alkylation of adenine with 1-bromo-2,2-diethoxyethane **4.8a** or 3-bromo-1,1-diethoxypropane **4.8b** under alkaline conditions afforded acetals **4.9a** and **4.9b**, respectively. Acidic hydrolysis of the acetal moiety furnished the corresponding aldehydes **4.10a** and **4.10b**, respectively.
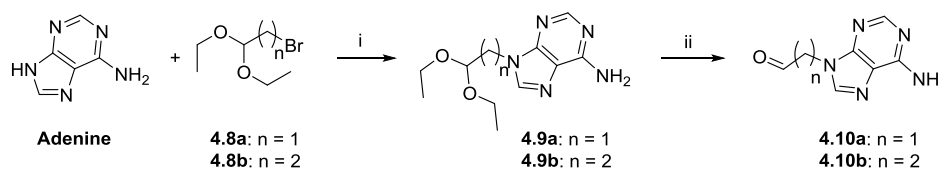
Scheme 4-1. Synthesis of intermediates **4.10a** and **4.10b**. *Reagents and conditions*: (i) K_2CO_3, DMF, 90 °C, 24 h, 38 % for **4.9a** and 45 % for **4.9b**; (ii) 1 N HCl, 90 °C, 1 h, 100 %.

The synthesis of compound **4.1** is shown in **Scheme 4-2**. The key step is a reductive amination (Borch reaction)[20] between aldehyde **4.10a** and L-serinamide hydrochloride **4.11** in the presence of sodium cyanoborohydride, yielding compound **4.12**. The primary hydroxyl group of intermediate **4.12** was protected as a *tert*-butyldimethylsilyl (TBS) ether yielding intermediate **4.13** in good yield. Subsequent reaction with tetrabenzyl pyrophosphate and lithium *tert*-butoxide as a strong base afforded phosphoroamidate **4.14**. Finally, tetrabutylammonium fluoride (TBAF)-mediated cleavage of the TBS protecting group, followed by cleavage of the benzyl protecting groups by treatment with a Lewis acid, afforded target compound **4.1**.
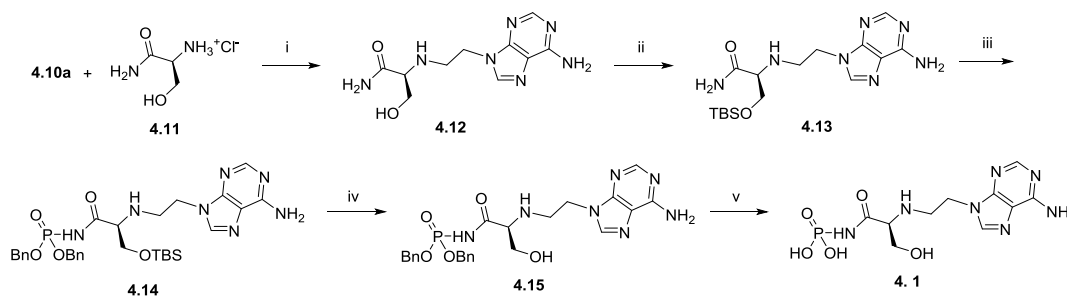
Scheme 4-2. Synthesis of compound **4.1**. *Reagents and conditions*: (i) $NaCNBH_3$, 3 Å molecular sieves, MeOH, r.t., overnight, 62 %; (ii) TBS-Cl, imidazole, DMF, r.t., overnight, 70 %; (iii) tetrabenzyl pyrophosphate, *t*-BuOLi, THF, r.t., 2 h, 59 %; (iv) TBAF, THF, 0 °C→r.t., 3 h, 60 %; (v) 2,6-lutidine, TMSBr, DCM, 0 °C→r.t., overnight, 32 %.

The preparation of compounds **4.2** and **4.3** is illustrated in **Scheme 4-3**. *N*-Boc-L-serine **4.16a** and *N*-Boc-D-serine **4.16b** were cyclized to the corresponding lactones **4.17a** and **4.17b**, respectively, under Mitsunobu reaction conditions, using diisopropyl azodicarboxylate (DIAD) and triphenylphosphine. Nucleophilic addition of trimethyl phosphite to **4.17a** and **4.17b** yielded (*R*)- and (*S*)-methyl *N*-Boc-2-amino-3-(dimethylphosphono)propanoate **4.18a** and **4.18b**, respectively, in excellent yield.[21] Acidic cleavage of the Boc protecting groups with 4 M HCl in dioxane yielded the amine hydrochlorides **4.19a** and **4.19b**, which were subjected to a reductive amination with aldehyde **4.10a** furnishing the secondary amines **4.20a** and **4.20b**, respectively. Hydrolysis of the phosphonate ester groups with TMSBr gave compounds **4.21a** and **4.21b**, respectively. Finally, the methyl ester functionalities were hydrolyzed to the corresponding carboxylic acids with 1M NaOH, affording compounds **4.2** and **4.3**, respectively.
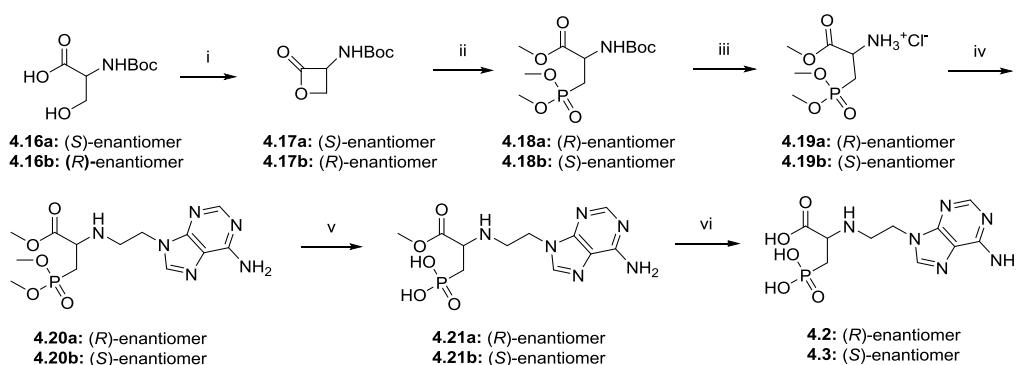
Scheme 4-3. Synthesis of compounds **4.2** and **4.3**. *Reagents and conditions*: (i) DEAD, PPh$_3$, THF, -78 °C→r.t., 2.5 h, 57 % for **4.17a** and **4.17b**; (ii) P(OMe)$_3$, 70 °C, 42 h, 79 % for **4.18a** and 76 % for **4.18b**; (iii) 4 M HCl in dioxane, 0 °C→r.t., 1 h; (iv) **4.10a**, NaCNBH$_3$, MeOH, r.t., overnight, 70 % for **4.20a** and 66 % for **4.20b** over two steps; (v) TMSBr, 2,6-lutidine, DCM, 0 °C→r.t., overnight, 89 % for **4.21a** and 82 % for **4.21b**; (vi) 1 M NaOH, MeOH, r.t., 3 h, 40 % for compound **4.2** and 68 % for compound **4.3**.

For the synthesis of intermediate **4.22**, applying a similar procedure as for the synthesis of compounds **4.2** and **4.3** failed. Therefore, a two-step reductive amination reaction was performed (**Scheme 4-4**). Condensation of aldehyde **4.10b** with amine **4.19b** in the presence of TEA and anhydrous sodium sulfate yielded the corresponding imine, which was subsequently reduced with sodium cyanoborohydride furnishing amine **4.22**, albeit only in 25 % yield. This suggests a diminished reactivity of intermediate **4.10b** in the reductive amination reaction, when compared to compound **4.10a**. Cleavage of the phosphonate ester groups with TMSBr, followed by saponification of the methyl ester group of **4.23** afforded compound **4.4**.
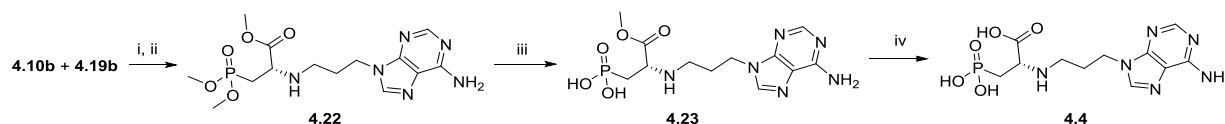
Scheme 4-4. Synthesis of compound **4.4**. *Reagents and conditions*: (i) TEA, Na$_2$SO$_4$, THF, r.t., 16 h; (ii) MeOH, NaCNBH$_3$, HOAc, r.t., 1 h, 25 % over two steps; (iii) 2,6-lutidine, TMSBr, DCM, 0 °C→r.t., overnight, 67 %; (iv) 1 M NaOH, MeOH, r.t., 3 h, 28 %.

The synthesis of compounds **4.5** and **4.6** is shown in **Scheme 4-5**. Key intermediates **4.33a** and **4.33b** were prepared by modifying the procedure reported by Hocková.[22-23] The two compounds were prepared as racemic mixtures. Oxidative cleavage of 3-benzyloxy-1,2-propanediol **4.24** with sodium periodate yielded (benzyloxy)acetaldehyde **4.25**. Nucleophilic addition of the lithium salt of diisopropyl methylphosphonate **4.26** to aldehyde **4.25** yielded diisopropyl (3-(benzyloxy)-2-hydroxypropyl) phosphonate **4.27**. The secondary hydroxyl group of **4.27** was alkylated with ethyl bromoacetate furnishing ethyl ester **4.28**. Reduction of the ethyl ester moiety was achieved with borane-THF affording alcohol **4.30a**. In order to have access to the propanol derivative **4.30b**, alcohol **4.27** was alkylated with ethyl 3-bromopropanoate employing a similar procedure as for the synthesis of **4.30a**. However, even after increasing the reaction temperature to 60 °C and extension of the reaction time to 24 hours, the desired product was not formed. As ethyl 3-bromopropanoate displayed

a lower reactivity in the nucleophilic substitution reaction than ethyl bromoacetate and the tendency of elimination at higher temperature or in the presence of a strong base, forced us to explore an alternative strategy. Therefore, alcohol **4.27** was alkylated with the more reactive allyl bromide yielding allyl derivative **4.29**. Hydroboration of the alkene moiety with 9-borabicyclo[3.3.1]nonane (9-BBN) in THF, followed by oxidation with H_2O_2/NaOH successfully gave the propanol derivative **4.30b**. Alkylation of 6-chloropurine with primary alcohols **4.30a** and **4.30b** under Mitsunobu conditions afforded derivatives **4.31a** and **4.31b**, respectively. Nucleophilic substitution of the chlorine by ammonia was achieved by heating compounds **4.31a** and **4.31b** in methanolic ammonia in an autoclave affording adenine derivatives **4.32a** and **4.32b**, respectively. Debenzylation by hydrogenolysis (Pd/C) gave hydroxymethyl derivatives **4.33a** and **4.33b**, respectively. Oxidation of the primary hydroxyl group using the TEMPO/$NaClO_2$/NaClO system[24] yielded carboxylic acids **4.34a** and **4.34b.** Finally, the isopropyl ester moieties were cleaved off with TMSBr affording target compounds **4.5** and **4.6**, respectively.
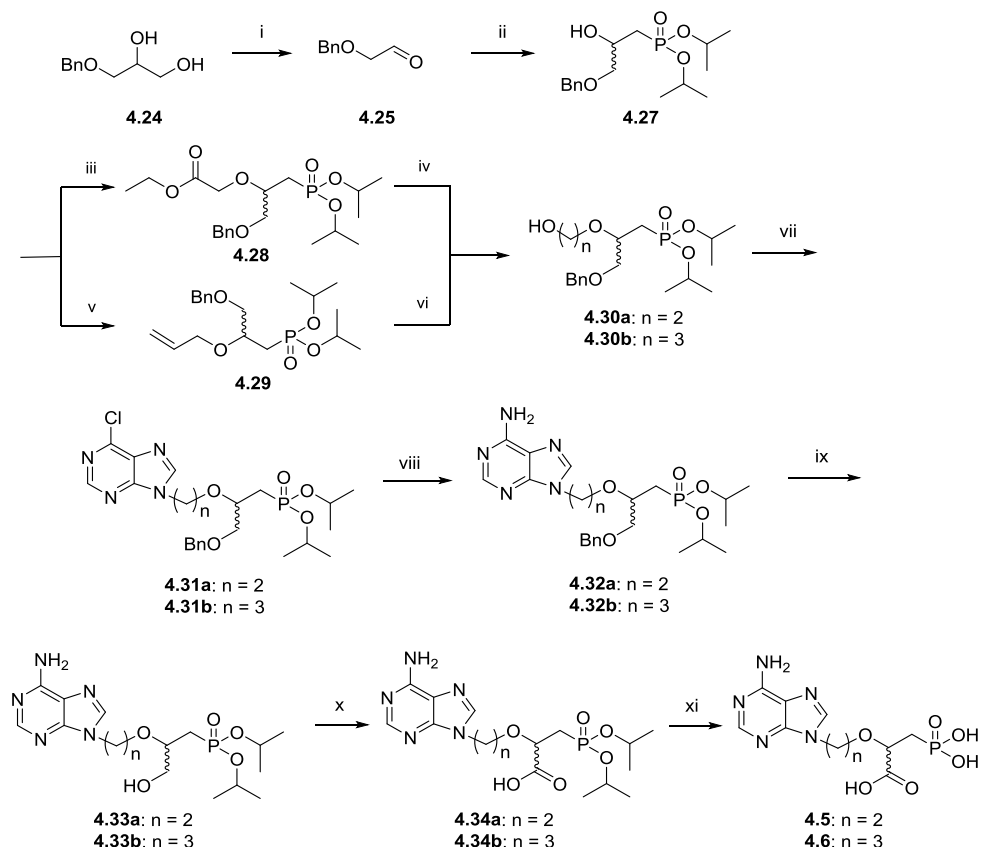
Scheme 4-5. Synthesis of compound **4.5** and **4.6**. *Reagents and conditions*: (i) $NaIO_4$, DCM, H_2O, r.t., 15 min; (ii) 2.5 M *n*-BuLi in hexane, diisopropyl methylphosphonate **4.26**, THF, -78 °C, 30 min, 56 % over two steps; (iii) NaH, TEA, THF, -40 °C, 30 min, then ethyl bromoacetate, r.t., 6 h, 77 %; (iv) 1M BH_3 in THF, -20 °C→r.t., 3 days, 84 % for **4.30a**; (v) NaH, TEA, THF, -40 °C, 30 min, then allyl bromide, r.t., 6 h, 80 %; (vi) 9-BBN, THF, 0 °C→r.t., 12 h, then EtOH, 4 N NaOH, 30 % H_2O_2, 0 °C→r.t., 2 h, 63 % over steps for **4.30b**. (vii) 6-chloropurine, PPh_3, DIAD, THF, -30 °C→r.t., 2 days, 81 % for **4.31a** and 75 % for **4.31b**; (viii) 7 M NH_3 in MeOH, 70 °C, 30 h, 72 % for **4.32a** and 72 % for **4.32b**; (ix) H_2, 10 % Pd/C, MeOH, r.t., 2 days, 88 % for **4.33a** and 71 % for **4.33b**; (x) TEMPO, $NaClO_2$, 5 % NaClO, 0.67 M PBS buffer, MeCN, r.t., 24 h, 36 % for **4.34a** and 31 % for

4.34b; (xi) 2,6-lutidine, TMSBr, DCM, r.t., 2 days, 37 % for compound **4.5** and 36 % for compound **4.6**.

For the preparation of intermediate **4.41**, needed for the synthesis of compound **4.7**, the procedure reported by Hocková *et al.*[25] was followed (**Scheme 4-6**). An aza-Michael addition of diethyl vinylphosphonate **4.35** to (*R*)-2-amino-1-propanol **4.36** gave the desired phosphonate **4.37** in excellent yield. Unfortunately, alkylation of the amino group of compound **4.37** with methyl bromoacetate yielded mainly lactone **4.40**, whereas the desired compound **4.41** was only isolated in minor amounts. To avoid this intramolecular esterification reaction, the primary hydroxyl group was protected as a *tert*-butyldiphenylsilyl (TBDPS) group, yielding compound **4.38**. Alkylation with methyl bromoacetate smoothly provided methyl ester **4.39** in high yield. Unfortunately, upon cleavage of the TBDPS protecting group, an intramolecular cyclization again occurred affording lactone **4.40**.
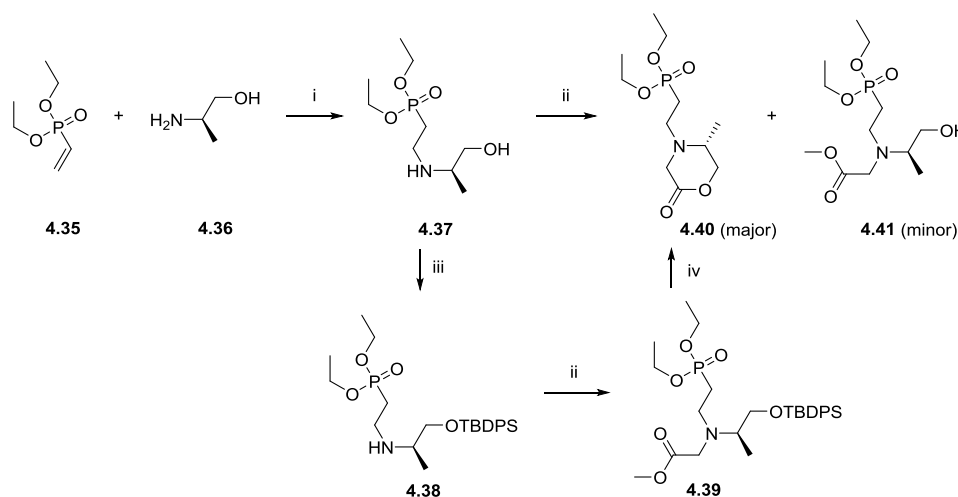
Scheme 4-6. Undesired cyclization of intermediate **4.41**. *Reagents and conditions*: (i) H_2O, r.t., 2 days, 94 %; (ii) methyl bromoacetate, K_2CO_3, CH_3CN, 0 °C→r.t., overnight, 95 % for **4.39** and 50 % for **4.40**; (iii) TBDPS-Cl, imidazole, DMF, r.t., overnight, 80 %; (iv) TBAF, THF, 0 °C→r.t., 2 h, 56 %.

Because of these synthetic problems, a new route was designed in which the isopropylamino group was linked to the N9-position of 6-chloropurine at an early stage of the synthesis (**Scheme 4-7**).
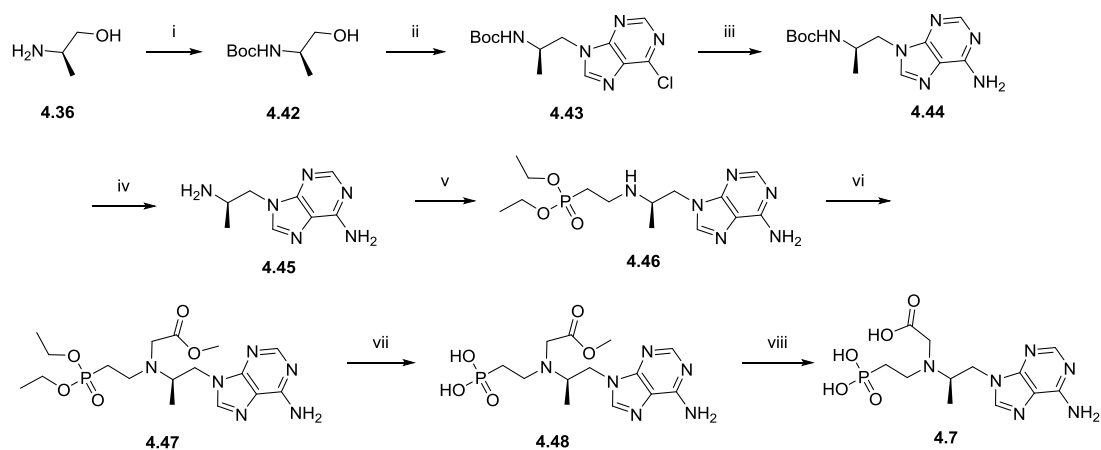
Scheme 4-7. Synthesis of compound **4.7**. *Reagents and conditions*: (i) Boc$_2$O, TEA, MeOH, 0 °C→r.t., 12 h, 100 %; (ii) 6-chloropurine, PPh$_3$, DIAD, DCM, 0 °C→r.t., 6 h, 36 %; (iii) 7 M NH$_3$ in MeOH, 70 °C, 30 h, 65 %; (iv) 4 M HCl in dioxane, 0 °C→r.t., 1 h, 87 %; (v) **4.35**, H$_2$O, r.t., 2 days, 26 %; (vi) methyl bromoactate, K$_2$CO$_3$, CH$_3$CN, 0 °C→r.t., 40 h, 51 %; (vii) 2,6-lutidine, TMSBr, DCM, 0 °C→r.t., overnight, 71 %; (viii) 1 M NaOH, MeOH, r.t., 3 h, 52 %.

The amino group of (*R*)-2-aminopropan-1-ol **4.36** was protected with a Boc group, furnishing compound **4.42**. Mitsunobu reaction of intermediate **4.42** with 6-chloropurine, using DIAD and triphenylphosphine, yielded compound **4.43**. The 6-chloropurine derivative **4.43** was converted into adenine derivative **4.44** by a nucleophilic aromatic substitution in 7N methanolic ammonia. Acidic cleavage of the Boc-protected group generated amine **4.45**. An aza Michael reaction of diethyl vinylphosphonate **4.35** with amine **4.45** gave the addition product **4.46**. Alkylation of the secondary amine with methyl bromoacetate generated methyl ester **4.47**. Lewis acid-mediated cleavage of the phosphonate esters, followed by alkaline hydrolysis of carboxylic acid ester moiety, furnished compound **4.7**.

4.2.2 HIV-1 RT inhibition assay (performed by Dr. Hoai Nguyen)

Compounds **4.1**-**4.7** were evaluated for their inhibitory effect on the primer extension, catalyzed by HIV-1 RT, using a gel-based *in vitro* RT inhibition assay. The template-primer duplex has a seven dT nucleotides overhang at the 5' end of the template. Hence, this template-primer allows for seven primer extensions. Since compounds **4.1-4.7** are designed as NcRTIs, the anti-HIV drug EFV, an NNRTI, was selected as positive control,[26] while Milli-Q water served as negative control. The natural substrate dATP was used at a concentration of 5 μM with 3 mM MgCl$_2$ as the metal ion source. After 10 min of polymerase reaction, EFV showed nearly complete RT inhibition when tested at concentrations of 5 or 50 μM, and no RT inhibition was observed at 0.5 μM. Among the newly synthesized ANP analogs, a clear inhibition of the RT mediated primer extension was observed for compounds **4.1** and **4.4-4.7** at a concentration of 10 mM. In contrast, compounds **4.2** and **4.3** did not show any inhibition, when tested at 10 mM (**Figure 4-2**A). At 5 mM, compounds **4.1** and **4.4-4.7** still demonstrated clear RT inhibition, especially more prominent for compounds **4.5** and **4.7** (**Figure 4-**

2B). However, none of the ANP analogs exerted an inhibitory effect on the RT catalyzed polymerization at a concentration of 1 mM (**Figure 4-2**C). Overall, the fact that these compounds are endowed with RT inhibitory activity, but only at higher concentrations, suggest that they have only weak binding affinity for RT and function as poor competitors of the natural dATP substrate.
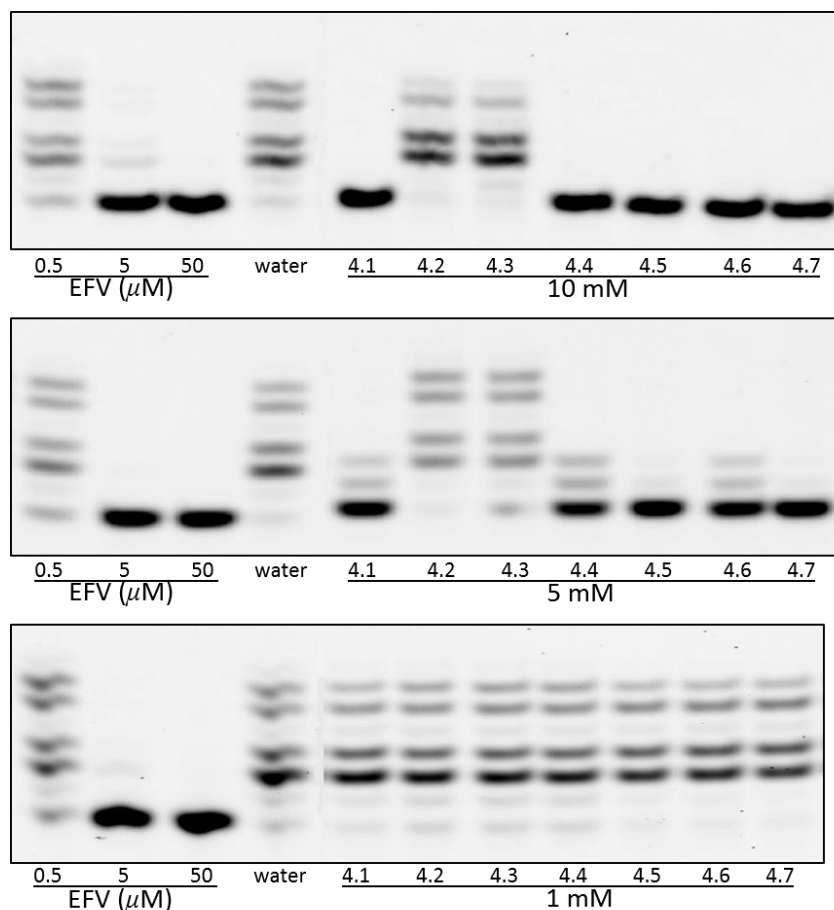
Figure 4-2. HIV-1 RT inhibition assay. Reaction conditions: 125 nM primer/template; [dATP] = 5 μM; [HIV-1 RT] = 0.06 μg/μL, 3 mM MgCl$_2$.

4.2.3 X-ray structure analysis (performed by Dr. Sergio Martinez and Dr. Abhimanyu K. Singh)

Crystal structures of seven compounds were determined in complex with HIV-1 RT/DNA in the presence of Mg^{2+} or Mn^{2+} ions. The pre-formed crystals of RT cross-linked to a dsDNA template/primer[27] were soaked in a 2 mM solution of individual compounds to form respective complexes. The diffraction data from the inhibitor-soaked crystals were collected at synchrotron beamlines, and the structures were refined at resolution between 2.95 and 3.45 Å (for data collection and refinement statistics, **Table 4-1**). The RT/DNA complex crystallizes with P2$_1$ space group symmetry and contains two copies of RT/DNA in the asymmetric unit. In general, the electron density for the inhibitor in one of the two copies is better than for the other. Thus, the copy with the better inhibitor density is used for the analysis and discussion, unless otherwise stated. RT can use either Mg^{2+} or Mn^{2+} ions for catalysis of DNA synthesis, but catalysis using Mg^{2+} ion requires high precision

chelation groups and geometry. In contrast, Mn^{2+} ion is more forgiving. Therefore, we prepared the complexes of inhibitors in the presence of either Mg^{2+} alone or Mn^{2+} and Mg^{2+} ions for a comprehensive understanding of the binding mode of these compounds at the polymerase active site.

Table 4-1. Data collection and refinement statistics

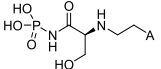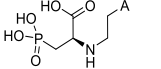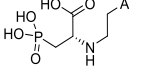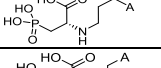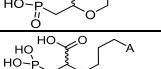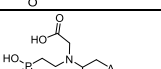| compound, divalent cation	4.1 (RMC230) Mg^{2+}	4.2 (RMC233) Mg^{2+}	4.3 (RMC247) Mn^{2+}	4.4 (RMC257) Mg^{2+}	4.5 (RMC259) Mg^{2+}	4.6 (RMC264) Mg^{2+}	4.7 (RMC282) Mg^{2+}
PDB code	7OTA	7OTK	7OTN	7OTX	7OTZ	7OUT	7OT6
Crystal Parameters							
space group	P2$_1$	P2$_1$	P2$_1$	P2$_1$	P2$_1$	P2$_1$	P2$_1$
cell dimensions a,b,c (Å), b (°)	89.7, 132.1, 138.5, 98.4	89.6, 132.5, 139.3, 98.5	89.8, 132.4, 138.6, 98.5	89.8, 132.9, 139.6, 98.8	89.8, 133.1, 140.4, 97.6	89.6, 132.6, 139.4, 98.7	89.5, 132.7, 139.7, 98.5
molecules/AU^a	2	2	2	2	2	2	2
Data Collection							
Beamline, detector	SOLEIL PX1 Eiger X 16M	SOLEIL PX2A Eiger X 9M	SOLEIL PX2A Eiger X 9M	SOLEIL PX2A Eiger X 9M	SOLEIL PX2A Eiger X 9M	SOLEIL PX2A Eiger X 9M	SOLEIL PX2A Eiger X 9M
wavelength Å	0.9786	0.9801	0.9801	0.9801	0.9801	0.9801	0.9801
resolution range^b	79.99-3.00 (3.07-3.00)	61.88-2.95 (3.02-2.95)	95.23-3.40 (3.53-3.40)	53.93-3.45 (3.58-3.45)	96.19-3.10 (3.18-3.10)	53.81-3.20 (3.30-3.20)	95.70-3.20 (3.30-3.20)
observed/unique^c reflections	425510/64004 (29585/4504)	452149/67827 (31471/4588)	278514/44271 (29873/4613)	280070/42820 (29593/4505)	381481/59442 (29799/4578)	322541/53240 (27748/4595)	328486/53347 (29211/4601)
completeness %^{b}	100.0 (100.0)	100.0 (100.0)	100.0/(99.9)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
multiplicity	6.6 (6.6)	6.7 (6.9)	6.3 (6.5)	6.5 (6.6)	6.4 (6.5)	6.1 (6.0)	6.2 (6.3)
R$_{merge}$^{b,d}	0.12 (2.322)	0.107 (1.80)	0.188 (2.328)	0.205 (2.707)	0.14 (1.84)	0.134 (2.097)	0.168 (1.685)
R$_{pim}$	0.05 (0.977)	0.045 (0.737)	0.082 (0.988)	0.087 (1.136)	0.06 (0.788)	0.059 (0.922)	0.074 (0.722)
I/s(I)	8.6 (1.0)	10.5 (1.2)	7.3 (1.1)	7.1 (1.2)	9.4 (1.3)	7.8 (1.2)	7.5 (1.2)
CC$_{1/2}$	0.997 (0.274)	0.999 (0.281)	0.979 (0.350)	0.971 (0.257)	0.996 (0.330)	0.995 (0.328)	0.997 (0.275)
Refinement							
resolution, Å	53.50-3.00	61.11-2.95	95.23-3.40	53.93-3.45	96.19-3.10	53.81-3.20	95.70-3.20
R$_{work}$/R$_{free}$^{e}	23.1/26.9	22.7/27.1	21.4/25.4	21.4/25.1	21.0/25.7	21.9/26.5	22.3/25.8
no. of atoms	17624	17639	17644	17624	17620	17622	17619
macromolecules	17597	17613	17597	17597	17597	17597	17594
ligand	24	23	46	24	23	24	25
water	3	3	1	3	0	1	0
B-factors, Å^2							
Wilson B	100.1	92.5	103.0	109.1	85.8	102.4	91.6
all atoms	134.72	118.61	133.86	150.53	115.09	145.08	117.55
macromolecules	134.68	118.56	133.69	150.43	115.01	144.95	117.46
ligands	178.4	172.2	197.5	241.9	175.5	247.7	178.1
waters	30.0	30.0	134.6	30.0	0	30.0	0
rmsd^f							
bond length, Å	0.002	0.004	0.002	0.002	0.004	0.004	0.002
bond angle, °	0.433	0.666	0.484	0.447	0.800	0.771	0.498
Ramachandran %^{g}	96.2/3.7/0.1	95.8/4.1/0.1	96.2/3.7/0.1	96.4/3.4/0.2	96.2/3.6/0.2	96.0/3.8/0.2	96.0/3.8/0.2

^{a}Asymmetric unit. ^{b}Values in parentheses of resolution range, completeness, R$_{merge}$, and I/σ(I) correspond to the last resolution shell. ^{c}Friedel pairs were treated as identical reflections. ^{d}Rmerge(I) =$\Sigma_{hkl}\Sigma_j|[I(hkl)_j - I(hkl)]|/\Sigma_{hkl}I_{hkl}$, where $I(hkl)_j$ is the measurement of the intensity of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity. ^{e}R = $\Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where R$_{free}$ is calculated without a σ cutoff for a randomly chosen 3% of reflections, which were not used for structure refinement, and R$_{work}$ is calculated for the remaining reflections. ^{f}Root mean square deviations from ideal bond lengths/angles. ^{g}Number of residues in favored/allowed/outlier region.

In our previously reported crystal structures,[18] no metal chelation was observed in the presence of Mg^{2+} ion alone, whereas one Mn^{2+} ion was involved in the chelation with inhibitors at the polymerase active site. Interestingly, in the current study, except for compound **4.3** which did not show good diffraction data in the presence of Mg^{2+}, all the other compounds chelate one Mg^{2+} or one Mn^{2+} ion. Because the modes of chelation with Mg^{2+} or Mn^{2+} ion are very similar, the structures of all compounds, except for compound **4.3**, are reported with a bound Mg^{2+} ion, and the structure of compound **4.3** is reported with a bound Mn^{2+} ion at the polymerase active site. It is noteworthy that all these compounds interact with the conserved residue Arg72 (**Table 4-2**). As a part of the fingers β3-β4 hairpin, Arg72 plays a critical role in the binding and positioning of dNTP at the polymerase active site.[28-29] In dNTP-bound structures, the side chain of Arg72 stacks with the dNTP base and

forms polar interaction with α-phosphate, whereas in the current structures, the orientation of Arg72 is altered when compared to the dNTP-bound RT/DNA structures.

Table 4-2. Summary of binding information of compounds in this study

Crystal structures	Chemical structure of inhibitor[a]	Involved amino acids	Interacting surface area (Å^2)[b]	Number of H-bonds	Chelating group of inhibitor
RT/DNA/Mg^{2+}/**4.1**		K65, R72, D110, D113, A114	464.8	4	-C=O
RT/DNA/Mg^{2+}/**4.2**		K65, R72, Q151, D185	451.8	3	-COOH
RT/DNA/Mn^{2+}/**4.3**		K65, R72, D110, A114, Q151, Q185	455.9	4	-COOH
RT/DNA/Mg^{2+}/**4.4**		R72, D110, Q151, D185	459.1	2	-COOH
RT/DNA/Mg^{2+}/**4.5**		R72, D185	439.7	2	-COOH
RT/DNA/Mg^{2+}/**4.6**		K65, R72, D110, Q151, D185	464.5	2	-COOH
RT/DNA/Mg^{2+}/**4.7**		K65, R72, D110, D185	508.3	2	-COOH

[a] A represents N9-adeninyl.

[b] Generated from https://www.ebi.ac.uk/pdbe/pisa.

Structure of compound 4.1 with RT/DNA

Among the synthesized compounds, only compound **4.1** possesses a hydroxymethyl group. The structure of RT/DNA/**4.1**/Mg^{2+} was determined at 3.0 Å resolution. The electron density reveals the presence of one Mg^{2+} ion (metal B) at the polymerase active site. One oxygen from the carbonyl group of compound **4.1** chelates with the Mg^{2+} ion, and the ion also chelates two side-chain oxygens of the catalytic aspartate Asp110 (**Figure 4-3**A). The adenine base of compound **4.1** base pairs with the thymine base of the first overhang nucleotide. However, base pairing and base stacking are not optimal; adenine is tilted approximately 19° away from the plane of the thymine causing a loss of the stacking interactions with the dsDNA nucleotide guanine base at the primer terminus. In addition, the two phosphonate oxygens of **4.1** are involved in strong polar interactions: 2.55 and 2.90 Å salt bridge interactions with the Nζ atom of Lys65 and the Nη2 atom of Arg72, respectively. Another phosphonate oxygen of **4.1** forms a 2.63 Å H-bond with the backbone amide NH of Asp113, and the hydroxymethyl group of **4.1** forms a 3.0 Å H-bond with the backbone amide NH of Ala114 (**Figure 4-3**B).
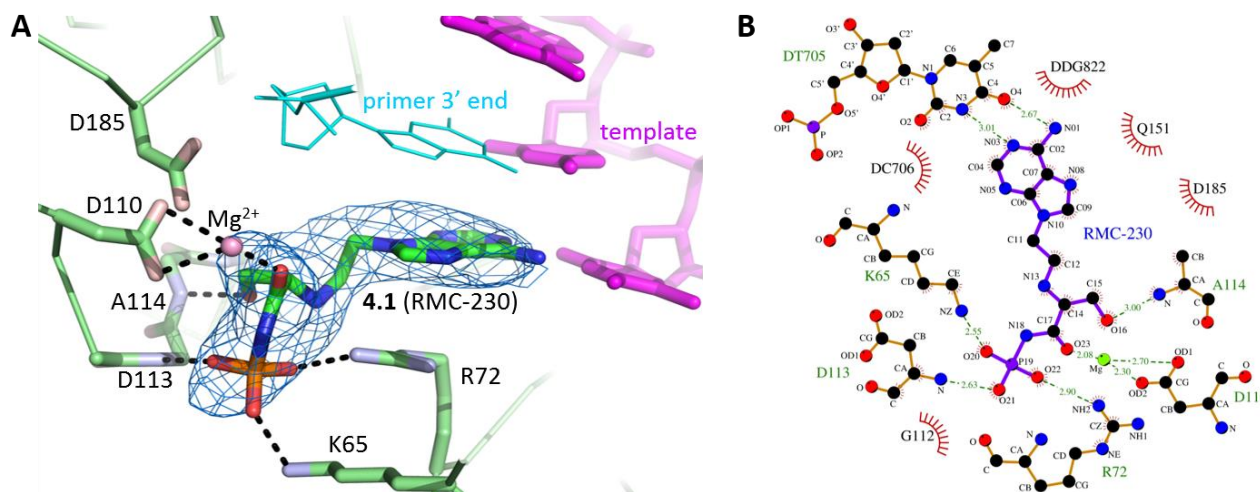
Figure 4-3. Binding mode of the compound **4.1** (RMC-230) to HIV-1 RT/DNA. (A) Polder map contoured at 7σ showing the binding of **4.1** at the polymerase active site in the 3.0 Å resolution crystal structure of HIV-1 RT/DNA/**4.1** complex in the presence of Mg^{2+} (pink sphere); PDB code 7OTA. The template strand is magenta, and the primer 3′ terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (B) Ligplot representation of the interaction of **4.1** with HIV-1 RT/DNA in the presence of Mg^{2+}.

Although compound **4.1** forms the highest number of H-bonds (four H-bonds), its metal chelation geometry is not optimal. With the aid of PDBePISA server (http://pdbe.org/pisa/), we calculated the total interacting surface area of the individual compound with the surrounding atoms of protein (chain A of HIV-1 RT) and dsDNA (chain T and P). The interacting surface area, also called as buried surface area, indicates the surface of a ligand involved in the interface formation. For compound **4.1**, the total interaction surface area is 464.8 $Å^2$ (**Table 4-2**). Due to weak binding, compound **4.1** showed noticeable RT inhibition only at higher concentrations (5 or 10 mM), but lost its inhibitory effect at lower concentrations (**Figure 4-2**). These inhibition data are in agreement with the structural data that suggest weak binding of **4.1**, and hence, the natural substrate dATP outcompetes **4.1** in the inhibition assay as both compete for binding at the dNTP-binding pocket.

Structures of compounds 4.2, 4.3 and 4.4 with RT/DNA

Compounds **4.2**, **4.3** and **4.4** all contain a secondary amine linker. Compounds **4.2** and **4.3** are enantiomers, while compound **4.4** has a longer acyclic chain. The structure of the RT/DNA/**4.2** complex in the presence of Mg^{2+} was obtained at 2.95 Å resolution. The adenine base pairs with the thymine of the template first base overhang. Metal chelation is involved in the binding of **4.2** in the structure obtained in the presence of Mg^{2+} ions. One Mg^{2+} ion positioned at the metal B site chelates the catalytic aspartate Asp185 and two carboxyl oxygens of **4.2** (**Figure 4-4**A). Different from our previously reported structures, the metal ion B chelates the carboxylic acid group of **4.2** rather than the phosphonate moiety. The phosphonate group interacts with the side chain of Lys65 and the guanidinium group of the conserved residue Arg72, forming a 3.23 Å interaction with Lys65, and a

2.86 Å interaction with Nε of Arg72 (**Figure 4-4**B). In addition, one 2.86 Å H-bond is formed between sidechain of residue Gln151 and the NH in acyclic chain of **4.2**. This compound forms three H-bonds and its total interacting surface area (451.8 Å^2) is also less than that of **4.1** (**Table 4-2**).
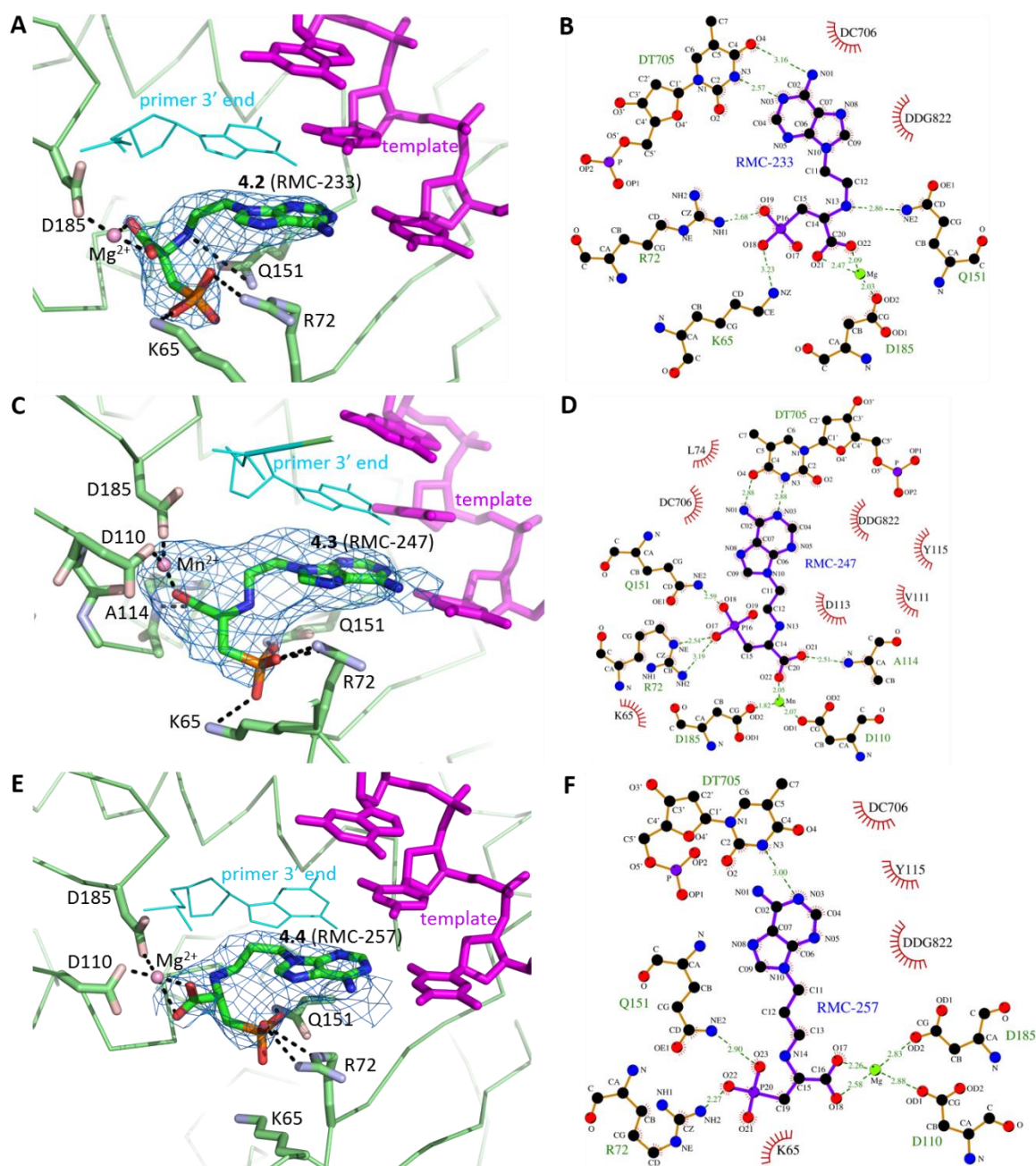
Figure 4-4. Binding modes of **4.2** (RMC-233), **4.3** (RMC-247), and **4.4** (RMC-257) to HIV-1 RT/DNA. (A) Polder map contoured at 6σ showing the binding of **4.2** at the polymerase active site in the 2.95 Å resolution crystal structure of HIV-1 RT/DNA/**4.2** complex in the presence of Mg^{2+} (pink sphere); PDB code 7OTK. The template strand is magenta, and the primer 3′ terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (B) Ligplot representation of the interaction of **4.2** with HIV-1 RT/DNA in the presence of Mg^{2+}. (C) Polder map contoured at 6σ showing the binding of **4.3** at the polymerase active site in the 3.4 Å resolution crystal structure of HIV-1 RT/DNA/**4.3** complex in the presence of Mn^{2+} (pink sphere); PDB code 7OTN. (D) Ligplot representation of the interaction of **4.3** with HIV-1 RT/DNA in the presence of Mn^{2+}. (E) Polder map contoured at 5.3σ showing the binding of **4.4** at the polymerase active site in the 3.45 Å resolution

crystal structure of HIV-1 RT/DNA/**4.4** complex in the presence of Mg^{2+} (pink sphere); PDB code 7OTX. (F) Ligplot representation of the interaction of **4.4** with HIV-1 RT/DNA in the presence of Mg^{2+}.

Due to the poor diffraction from the crystals of RT/DNA compex with **4.3** in the presence of Mg^{2+}, we could obtain the structure of RT/DNA/**4.3** complex only in the presence of Mn^{2+}, which was determined at 3.4 Å resolution. The binding of compound **4.3** in the polymerase active site also involves Mn^{2+} chelation. One oxygen from the carboxylic acid group of **4.3** chelates with the Mn^{2+} (metal B) at the polymerase active site, and this cation also chelates the catalytic aspartates Asp110 and Asp185 (**Figure 4-4**C). With the Mn^{2+} ion, the adenine base is tilted by 13° out of the plane of its complementary thymine base of the template overhang. The other oxygen from the carboxyl group of **4.3** forms a 2.51 Å H-bond with the backbone NH of Ala114. In addition, one oxygen from the phosphonate group of **4.3** forms 2.54 and 3.19 Å salt bridge interactions with the Nε atom and the Nη2 atoms of Arg72, respectively. One other phosphonate oxygen forms a 2.59 Å H-bond with Nε2 of Gln151 (**Figure 4-4**D).

Despite many structural differences, **4.3** forms the same number of H-bonds as **4.1**. The compound **4.3** has a slightly higher interacting surface area (455.9 $Å^2$) than **4.2** (**Table 4-2**). Possibly because of the stronger metal chelation ability of Mn^{2+}, **4.3** moves closer to the polymerase active site and forms more interactions when compared to its enantiomer **4.2**. Both compounds **4.2** and **4.3** were devoid of inhibition even at a 10 mM concentration, suggesting that none of the enantiomers was able to outcompete dATP for binding.

The crystal structure of RT/DNA/**4.4** complex in the presence of Mg^{2+} was obtained at 3.45 Å. The adenine base of **4.4** only partially base pairs with the thymine of the first base overhang. In the structure of RT/DNA/**4.4** in the presence of Mg^{2+}, the chelating Mg^{2+} ion positioned at the metal B site and chelates one oxygen from each of the catalytic aspartates Asp110 and Asp185, and two carboxyl oxygens of **4.4** (**Figure 4-4**E). The phosphonate group of **4.4** interacts with residue Gln151 and the conserved residue Arg72. One oxygen forms a 2.90 Å H-bond with Nε atom of Gln151, and another oxygen engages in a 2.27 Å salt bridge with the Nη2 atoms of Arg72 (**Figure 4-4**F).

Compound **4.4** is a close analog of **4.3** with an additional methylene group in the acyclic chain. The compound **4.4** has a slightly higher interacting surface area (459.1 $Å^2$) than compound **4.2** (**Table 4-2**). In the biochemical assay, compound **4.4** showed better RT inhibition than **4.3**, when tested at 5 or 10 mM (**Figure 4-2**).

Structures of compounds 4.5 and 4.6 with RT/DNA

Compounds **4.5** and **4.6** contain an ether linker. Although **4.6** has a longer acyclic chain than **4.5**, the two compounds have similar binding modes. The crystal structure of RT/DNA/**4.5** complex

in the presence of Mg^{2+} was obtained at 3.1 Å resolution (**Figure 4-5**A). Although **4.5** was prepared as a racemic mixture, only the (*S*)-enantiomer can fit the electron density map at the active site of HIV-1 RT, demonstrating that the (*S*)-enantiomer of **4.5** is a better binder than the (*R*)-enantiomer in the active site of HIV-1 RT. The (*S*)-enantiomer of **4.5** can be considered as a structural analog of **4.3** by substitution of the amino group in the acyclic chain of the latter with an oxygen linker. Interestingly, the metal chelation geometries of compounds **4.5** and **4.3** are different. In addition, compound **4.5** is more potent as HIV-1 RT inhibitor when compared to compound **4.3**, as evaluated in the gel-based primer extension assay.

In the crystal structure of RT/DNA/**4.5**/Mg^{2+} complex, the two oxygens of the carboxylic acid moiety of **4.5** chelate with a Mg^{2+} ion (metal B) at the polymerase active site. The Mg^{2+} ion also chelates a side-chain oxygen of catalytic aspartate Asp185 (**Figure 4-5**A). In addition, the two oxygens of the phosphonate group of **4.5** form 2.66 Å and 3.17 Å salt bridges with the Nε and the Nη2 atoms of Arg72, respectively (**Figure 4-5**B). Although **4.5** forms only two H-bonds and its interacting surface area (439.7 Å^2) is lower than that of **4.3** (**Table 4-2**), the former shows better RT inhibition, which might be due to better metal chelation with the two carboxylic acid oxygens.

Figure 4-5. Binding modes of the compounds **4.5** (RMC-259) and **4.6** (RMC-264) to HIV-1 RT/DNA. (A) Polder map contoured at 6.7 σ showing the binding of **4.5** at the polymerase active site in the 3.1 Å resolution crystal structure of HIV-1 RT/DNA/**4.5** complex in the presence of Mg^{2+} (pink sphere); PDB code 7OTZ. The template strand is magenta, and the primer 3′ terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (B) Ligplot representation of the

interaction of **4.5** with HIV-1 RT/DNA in the presence of Mg^{2+}. (C) Polder map contoured at 5.4 σ showing the binding of **4.6** at the polymerase active site in the 3.2 Å resolution crystal structure of HIV-1 RT/DNA/**4.6** complex in the presence of Mg^{2+} (pink sphere); PDB code 7OUT. (D) Ligplot representation of the interaction of **4.6** with HIV-1 RT/DNA in the presence of Mg^{2+}.

As an analog of **4.5** with an additional carbon in the acyclic chain, **4.6** still has a similar mode of binding as **4.5**. Only the (*S*)-enantiomer of **4.6** could be modeled to the electron density in the crystal structure of RT/DNA/**4.6** complex that was determined at 3.2 Å resolution. Metal chelation involves one more atom from Asp110 when compared with that in the structure of RT/DNA/**4.5**/Mg^{2+} complex. Two phosphonate oxygens of **4.6** form 2.62 and 2.93 Å H-bond interactions with the side-chain atoms of Lys65 and Arg72, respectively (**Figure 4-5**C). One other phosphonate oxygens of **4.6** also forms a 2.25 Å H-bond with Nε2 of Gln151 (**Figure 4-5**D). Compound **4.6** forms the same number of H-bonds as **4.5**, but it has a slightly higher interacting surface area (464.5 $Å^2$) than **4.5** (**Table 4-2**). However, poor base-pairing makes compound **4.6** only weakly bind at the polymerase active site, which might explain the slightly lower RT inhibition of **4.6** at a concentration of 5 mM, when compared to compound **4.5** (**Figure 4-2**B).

Structure of compound 4.7 with RT/DNA

Compound **4.7** is the only analog that contains a tertiary amine linker, and it resembles TFV at its adenine nucleobase and acyclic chain. The structure of RT/DNA/**4.7**/Mg^{2+} complex was determined at 3.2 Å resolution. Just like the abovementioned structures, the adenine base pairs with the thymine of the template first base overhang. However, base pairing and base stacking are not optimal. The adenine base of **4.7** is tilted by approximately 22° away from the plane of the template thymine causing a loss of the stacking interactions with the dsDNA nucleotide guanine base at the primer terminus. The binding of **4.7** at the polymerase active site also involves metal chelation when the complex was formed in the presence of Mg^{2+} ions. One Mg^{2+} ion at metal site B chelates two side-chain oxygens of catalytic aspartates Asp185 and one side-chain oxygen of Asp110, and both carboxyl oxygens of **4.7** (**Figure 4-6**A). Furthermore, two of the phosphonate oxygens of **4.7** forms 2.38 and 3.09 Å salt bridges with the N ζ atom of Lys65, and one of the two phosphonate oxygens forms a 3.10 Å salt bridge with the Nε atom of Arg72 (**Figure 4-6**B). Compound **4.7** forms three H-bonds, and its total interacting surface area (504.8 $Å^2$) is the highest among the synthesized compounds (**Table 4-2**). In agreement with the structure, this compound shows RT inhibition, when evaluated at 5 or 10 mM, and is the most potent inhibitor among the compounds reported in this study. However, this compound did not display RT inhibition when the concentration was reduced to 1 mM.

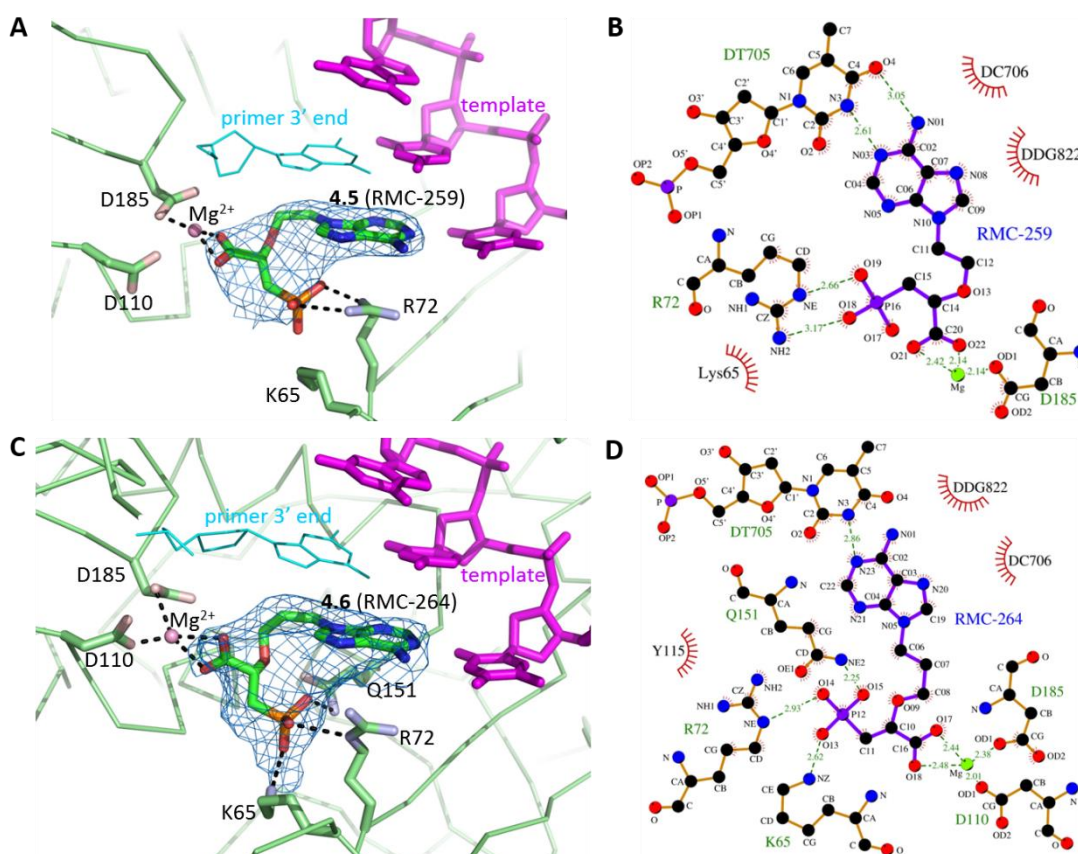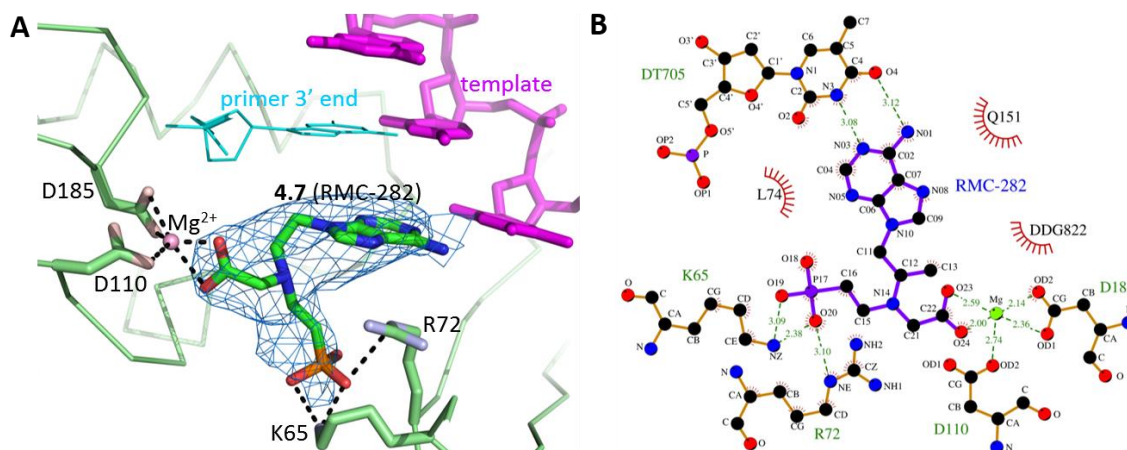**Figure 4-6**. Binding mode of the compound **4.7** (RMC-282) to HIV-1 RT/DNA. (A) Polder map contoured at 5.7σ showing the binding of **4.7** at the polymerase active site in the 3.4 Å resolution crystal structure of HIV-1 RT/DNA/**4.7** complex in the presence of Mg^{2+} (pink sphere); PDB code 7OT6. The template strand is magenta, and the primer 3′ terminus is cyan. The primer strand is terminated with dideoxyguanosine monophosphate. (B) Ligplot representation of the interaction of **4.7** with HIV-1 RT/DNA in the presence of Mg^{2+}.

Overall, all the newly synthesized ANP analogs are weak binders compared to the natural substrate dATP. The residue Arg72 and the *β3-β4* finger loop that contains Arg72 are solvent-exposed and highly flexible, which is vital for dNTP binding. Thus, even though nearly all these compounds can interact with the highly conserved residue Arg72, their polar interactions with Arg72 are solvent-exposed and can be easily dissociated when a natural dNTP competes for binding, which might be the reason for their weak RT inhibition.

4.3 Conclusion

Mimicking metal chelation at the polymerase active site has broad implications for designing inhibitors of viral DNA and RNA polymerases. Previously, we reported a systematic structural study of dAMP/TFV-amino acid conjugates as dATP mimics.[18] In the current study, we have introduced phosphonate and carboxylic acid groups as metal chelating moieties. Seven new ANP analogs with structural variations of the chain length, chirality and linker atom were designed as NTP mimics. Several challenging synthetic routes for the preparation of these ANP analogs were developed successfully. These synthetic explorations paved the way for synthesis of other novel ANPs. X-ray crystal structures were solved to get insight into their binding modes at the polymerase active site of HIV-1 RT. The ANP analogs bind at the same site as the natural dATP substrate,[30] but with a significantly reduced binding affinity. As a result, these ANP analogs only show inhibitory activity in an HIV-1 RT inhibition assay at high concentrations (mM range), because dATP outcompetes our compounds. Detailed structural information indicated that their binding modes are different from previously reported compounds by forming new interactions at the active site.[18] The carboxylic acid moiety, rather than the phosphonate group, is involved in metal chelation at the polymerase active

site. As expected, the base-pairing and base-stacking of the TFV part is almost conserved in the binding of the different inhibitors. Rather than chelating with the active-site metal ions, the phosphonate group in the inhibitors primarily interact with the sidechains of Lys64, Arg72, and Gln151, although the extent of interactions of the phosphonate group varies among the compounds. Novel interactions in the ligand-protein binding offer opportunities to develop more potent HIV-1 RT inhibitors. Our structure-based approach is employed to find different chemical scaffolds that can effectively chelate the active site metal ions. The active site metal ion chelation of compounds **4.1**-**4.7** at the polymerase active sire of RT are less extensive when compared to the NcRTI α-CNPs.[14] An understanding of the metal chelation possibilities by different chemical substitutions may help designing new class of polymerase inhibitors.

4.4 Experimental section

Chemistry All reagents and solvents were purchased from commercial sources and were used as obtained. Moisture-sensitive reactions were carried out in oven-dried glassware under a nitrogen or argon atmosphere. ^{1}H, ^{13}C, and ^{31}P NMR spectra were recorded on Bruker Avance 300 or 500 MHz spectrometer using tetramethylsilane as internal standard or referenced to the residual solvent signal, and 85 % H_3PO_4 for ^{31}P NMR experiments. Splitting patterns are designated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), br(broad), and dd (doublet of doublets). Coupling constants are expressed in hertz (Hz). High-resolution mass spectra (HRMS) were obtained on a quadruple orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 μL/min and spectra were obtained in positive (or negative) ionization mode with a resolution of 15000 (FWHM) using leucine enkephalin as lock mass. Optical rotations were recorded with a polarimeter Model 341 using a Na gas lamp with a wavelength of 589 nm and a standard temperature of 20 °C. The concentrations are expressed in g/100mL and the specific rotations ($[\alpha]_D$) are reported in deg·mL·dm^{-1}·g^{-1}. Pre-coated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ= 254 nm). Products were purified by flash chromatography on silica gel (35-70 μm, 60 Å). All final compounds were purified by preparative RP-HPLC (C18 Phenomenex Gemini column, 110 Å, 10 μm, 21.2 mm × 250 mm) using an eluent gradient of CH_3CN with 50 mmol TEAB as eluent buffer. Final compounds were isolated as their triethylammonium salt. Prior to biological evaluation, compounds were converted into the corresponding sodium salts via sodium ion exchange resin. The exact concentration of the compounds was determined by the Lambert-Beer law. The purity of all the tested compounds was more than 95 % by HPLC analysis.

9-(2,2-Diethoxyethyl)-9H-purin-6-amine (4.9a) A mixture of adenine (1 g, 7.4 mmol) and K_2CO_3 (1.714 g, 8.88 mmol) in DMF (15 mL) was stirred at 90 °C in the presence of 2-bromo-1,1-

dimethoxyethane **4.8a** (1.357 mL, 8.88 mmol). After 24 h, the reaction was filtered, and the solvent was removed under reduced pressure. The crude mixture was purified by silica gel column chromatography (DCM/MeOH, v/v, 60:1 to 30:1) to yield the title compound as a white solid (707 mg, 38 %). ^{1}H NMR (300 MHz, DMSO) δ 8.13 (s, 1H), 8.03 (s, 1H), 7.19 (s, 2H), 4.82 (t, $J = 5.4$ Hz, 1H), 4.20 (d, $J = 5.4$ Hz, 2H), 3.67-3.35 (m, 4H), 1.00 (t, $J = 7.0$ Hz, 6H). ^{13}C NMR (75 MHz, DMSO) δ 156.14 (s), 152.67 (s), 149.95 (s), 141.48 (s), 118.63 (s), 99.68 (s), 62.46 (s), 45.51 (s), 15.29 (s).

2-(6-Amino-9H-purin-9-yl) acetaldehyde (4.10a) Compound **4.9a** (200 mg, 0.8 mmol) was dissolved in 1N HCl (5 mL) and refluxed at 90 °C for 1 h. The mixture was neutralized with 10 N NaOH. The solvent was evaporated under reduced pressure, and the crude residue was used for the next step without purification. ^{1}H NMR (300 MHz, DMSO) δ 9.70 (s, 1H), 8.55 (s, 1H), 8.47 (s, 1H), 5.12 (t, $J = 5.2$ Hz, 1H), 4.18 (d, $J = 5.1$ Hz, 2H). ^{13}C NMR (75 MHz, DMSO) δ 195.61 (s), 150.11 (s), 148.59 (s), 144.67 (s), 144.46 (s), 117.41 (s), 87.33 (s), 52.87 (s), 49.72 (s).

(S)-2-((2-(6-Amino-9H-purin-9-yl)ethyl)amino)-3-hydroxypropanamide (4.12) Aldehyde **4.10a** (302 mg, 1.2 mmol), 3 Å molecular sieves (3 g) and a trace of bromocresol green were suspended in methanol (10 mL). NaBH$_3$CN (95 mg, 1.44 mmol) was added, followed by the dropwise addition of TEA. After 15 min, the color turned yellow and TEA was added dropwise to restore the blue color. The solution was stirred overnight. A saturated aqueous solution of NaHCO$_3$ was added to quench the reaction. The solvent was evaporated under reduced pressure, and the crude residue was purified by silica gel chromatography (EA/MeOH, v/v, 20:1 to 10:1 to 5:1) to yield the title compound as a white powder (275 mg, 62 %). ^{1}H NMR (300 MHz, MeOD) δ 8.22 (s, 1H), 8.21(s, 1H), 4.37 (t, $J = 5.5$ Hz, 2H), 3.76-3.61 (m, 2H), 3.24 (t, $J = 4.8$ Hz, 1H), 3.07 (t, $J = 5.5$ Hz, 2H). ^{13}C NMR (75 MHz, MeOD) δ 178.02 (s), 157.46 (s), 153.91 (s), 150.91 (s), 143.55 (s), 120.03(s), 65.39 (s), 63.92 (s), 57.13 (s), 45.38 (s). HRMS: [M+H]$^+$ calculated for C$_{10}$H$_{15}$N$_7$O$_2$, 266.1360; found 266.1364.

(S)-2-((2-(6-Amino-9H-purin-9-yl)ethyl)amino)-3-((tert-butyldimethylsilyl)oxy) propanamide (4.13) Compound **4.12** (358 mg, 1.35 mmol) and imidazole (138 mg, 2.03 mmol) were dissolved in DMF (4 mL) and TBS-Cl (306 mg, 2.03 mmol) was added. The mixture was stirred at room temperature overnight. A saturated aqueous solution of NaHCO$_3$ (30 mL) was added. The mixture was extracted with ethyl acetate (3×15 mL). The combined organic layers were washed with brine, dried over anhydrous MgSO$_4$. The solvents were evaporated *in vacuo* and the crude residue was purified by silca gel chromatography (DCM/MeOH, v/v, 20:1 to 10:1 to 5:1) to yield the title compound as a white solid (357 mg, 70 %). ^{1}H NMR (300 MHz, DMSO) δ 8.12 (s, 2H), 7.29-7.14 (m, 4H), 4.21 (t, $J = 6.0$ Hz, 2H), 3.60 (t, $J=3.0$ Hz, 2H), 3.06 (t, $J = 6.0$ Hz, 1H), 2.99-2.81 (m, 2H), 2.08 (s, 1H), 0.78 (s, 9H), -0.04 (s, 6H). ^{13}C NMR (75 MHz, DMSO) δ 174.23 (s), 156.20 (s), 152.52

(s), 149.86 (s), 141.43 (s), 118.66 (s), 64.16 (s), 63.33 (s), 47.11 (s), 43.49 (s), 25.97 (s), 18.16（s), -5.26 (d, J = 2.7 Hz). HRMS: $[M+H]^+$ calculated for $C_{16}H_{29}N_7O_2Si$, 380.2225; found 380.2220.

Dibenzyl (N-(2-(6-amino-9H-purin-9-yl)ethyl)-O-(tert-butyldimethylsilyl)-L-seryl) phosphoramidate (4.14) To a solution of **4.13** (216 mg, 0.57 mmol) in dry THF (20 mL), was added a solution of *t*-BuOLi (89.3 mg, 1.12 mmol) in THF (10 mL). The resulting reaction mixture was stirred under argon for 3 h at room temperature. Then, a solution of tetrabenzyl pyrophosphate (368 mg, 0.684 mmol) in THF (10 mL) was added dropwise over 5 min and the resulting mixture was stirred under argon for 2 h. The solvent was evaporated under vacuum and the residue was purified by silica gel flash chromatography (DCM/MeOH, v/v, 80:1 to 40:1 to 20:1), affording the phosphonate ester as a yellowish oil (214 mg, 59 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.95 (s, 1H), 8.45 (s, 1H), 8.02-7.89 (m, 10H), 7.05 (s, 2H), 5.76 (t, J = 9.0 Hz, 4H), 4.85 (t, J=6.0 Hz, 2H), 4.30 (t, J=6.0 Hz, 2H), 3.70 (dd, J = 12.0, 6.0 Hz, 2H), 3.55 (t, J=6.0 Hz, 1H), 1.36 (s, 9H), 0.57 (s, 6H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 174.08 (s), 156.00 (s), 153.27 (s), 150.36 (s), 140.96 (s), 135.98 (s), 128.78 (s), 128.30 (s), 119.97 (s), 69.63 (dd, J = 11.3, 5.5 Hz), 64.63 (d, J = 8.1 Hz), 62.60 (s), 48.03 (s), 44.58 (s), 25.94 (s), 18.32 (s), -5.35 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ -1.65 (s). HRMS: $[M+H]^+$ calculated for $C_{30}H_{42}N_7O_5PSi$, 640.2827; found 640.2852.

Dibenzyl ((2-(6-amino-9H-purin-9-yl)ethyl)-L-seryl)phosphoramidate (4.15) To an ice-cooled solution of **4.14** (214 mg, 0.335 mmol) in dry THF (10 mL) was added dropwise 1M TBAF in THF (669 μL, 0.669 mmol). The reaction mixture was then stirred at room temperature for 3 h. The solvent was evaporated under vacuum and the residue was purified by silica gel chromatography (DCM/MeOH, v/v, 30:1 to 10:1 to 5:1) to yield the title compound as a white solid (105 mg, 60 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.17 (s, 1H), 7.86 (s, 1H), 7.30-7.12 (m, 10H), 6.99 (s, 2H), 5.06 (s, 2H), 5.03 (s, 2H), 4.07 (s, 2H), 3.73 (s, 2H), 3.29 (s, 1H), 2.89 (s, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 175.29 (s), 155.83 (s), 152.84 (s), 149.74 (s), 141.29 (s), 135.58 (s), 135.49 (s), 128.54 (s), 128.00 (s), 119.16 (s), 69.51 (s), 64.74 (s), 61.92 (s), 47.25 (s), 43.97 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ -1.25 (s). HRMS: $[M+H]^+$ calculated for $C_{24}H_{28}N_7O_5P$, 526.1962; found 526.1956.

((2-(6-Amino-9H-purin-9-yl)ethyl)-L-seryl)phosphoramidic acid (4.1) To a solution of dibenzyl ester **4.15** (95 mg, 0.18 mmol) and 2,6-lutidine (233 μL, 2 mmol) in dry DCM (5 mL) was added dropwise bromotrimethylsilane (119 μL, 0.9 mmol) at 0 °C. The reaction mixture was stirred at room temperature overnight and quenched with a 1 M TEAB solution (4 mL). The solvent was removed under reduced pressure. The residue was partitioned between water and EA/ether (1:1). The aqueous layer was lyophilized, and the residue was first purified by silica gel flash chromatography (*i*-PrOH/NH$_3$/H$_2$O, v/v/v, 10:1:1 to 5:1:1). Further purification using preparative reverse phase HPLC with a gradient of CH$_3$CN and a 0.05 M TEAB solution gave the title compound as a white foam (20

mg, 32 %). ^{1}H NMR (300 MHz, D$_2$O) δ 8.10 (s, 1H), 8.08 (s, 1H), 4.29 (t, J = 5.3 Hz, 2H), 3.66 (d, J = 4.5 Hz, 2H), 3.19 (t, J = 4.5 Hz, 1H), 3.04(t, J= 5.3 Hz, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 174.47 (s), 155.25 (s), 152.17 (s), 148.53 (s), 142.50 (s), 63.57 (s), 62.10 (s), 46.28 (s), 45.92 (s), 43.50 (s), 8.29 (s). ^{31}P NMR (121 MHz, D$_2$O) δ -4.72 (s). HRMS: [M+H]$^+$ calculated for C$_{10}$H$_{16}$N$_7$O$_5$P, 346.1023; found 346.1016.

tert-**Butyl (S)-(2-oxooxetan-3-yl)carbamate (4.17a)** A solution of triphenylphosphine (1.28 g, 4.87 mmol) in anhydrous THF (25 mL) was cooled down to −78 °C. DEAD (0.77 mL, 4.87 mmol) was added dropwise over a period of 10 min. A solution of (Boc)-L-Ser **4.16a** (1 g, 4.87 mmol) in anhydrous THF (5 mL) was added dropwise over a period of 15 min. After stirring for 20 min at −78 °C, the temperature was slowly raised to room temperature. The mixture was stirred for 2.5 h and the solvent was evaporated *in vacuo*. Purification of the crude residue by silica gel flash chromatography using hexane/AcOEt (65:35) as the mobile phase, afforded the title compound as a white solid (516 mg, 57 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.37 (d, J = 6.9 Hz, 1H), 5.09 (d, J = 4.8 Hz, 1H), 4.52 − 4.33 (m, 2H), 1.45 (s, 9H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 169.35 (s), 154.43 (s), 81.36 (s), 66.91 (s), 59.64 (s), 28.48 (s).

Methyl (R)-2-((tert-butoxycarbonyl)amino)-3-(dimethoxyphosphoryl)propanoate (4.18a) Trimethyl phosphite (3 mL, 25 mmol) and *tert*-butyl (*S*)-(2-oxooxetan-3-yl)carbamate **4.17a** (372 mg, 1.987 mmol) were combined and heated at 70 °C in an oil bath under N$_2$ for 42 h. Upon cooling, the mixture was concentrated *in vacuo* to give a thick oil. The residue was purified by flash chromatography (5 % MeOH/CH$_2$C1$_2$) affording the title compound as a clear oil (491 mg, 79 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.65 (d, J = 7.7 Hz, 1H), 4.69-4.36 (m, 1H), 3.77-3.65 (m, 9H), 2.33 (dd, J = 17.3, 5.4 Hz, 2H), 1.40 (s, 9H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 171.59 (d, J = 9.2 Hz), 155.13 (s), 80.38 (s), 52.87 (d, J = 2.3 Hz), 52.76 (d, J=6.8 Hz), 49.30 (t, J=3.0 Hz), 28.54 (s), 26.53 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 30.17 (s). HRMS: [M+Na]$^+$ calculated for C$_{11}$H$_{22}$NO$_7$P, 334.1032; found 334.1031.

Methyl (R)-2-amino-3-(dimethoxyphosphoryl)propanoate hydrochloride (4.19a) Compound **4.18a** (374 mg, 1.2 mmol) was dissolved in 4 M HCl solution in dioxane (3 mL) at 0 °C. The ice-bath was removed, and the mixture was stirred at room temperature for 1 h. When TLC indicated that the reaction was completed, the solvent was evaporated *in vacuo* and the title compound was directly used for the subsequent reductive amination. ^{1}H NMR (300 MHz, D$_2$O) δ 4.79-4.48 (m, 1H), 3.91 (s, 3H), 3.87 (s, 3H), 3.83 (s, 3H), 2.84-2.56 (m, 2H). ^{31}P NMR (121 MHz, CDCl$_3$) δ 28.82 (s). ^{13}C NMR (75 MHz, D$_2$O) δ 168.56 (d, J = 7.0 Hz), 66.25 (s), 53.77 (s), 53.48 (s), 47.75 (s), 25.07 (s), 23.12 (s). HRMS: [M+H]$^+$ calculated for C$_6$H$_{14}$NO$_5$P, 212.0682; found 212.0681.

Methyl (R)-2-((2-(6-amino-9H-purin-9-yl)ethyl)amino)-3-(dimethoxyphosphoryl)propanoate (4.20a) This compound was obtained as a white solid (259 mg, 70 %) according to the reductive amination procedure used for the synthesis of intermediate **4.12**, starting from aldehyde **4.10a** (211 mg, 1.19 mmol) and amine **4.19a** (251 mg, 1.19 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.20 (s, 1H), 7.95 (s, 1H), 6.69 (s, 2H), 4.33-4.03 (m, 2H), 3.82-3.41 (m, 10H), 3.18-2.73 (m, 2H), 2.27-1.77 (m, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.61 (d, J = 14.2 Hz), 155.85 (s), 152.68 (s), 149.82 (s), 141.30 (s), 119.31 (s), 56.13 (d, J = 4.0 Hz), 52.56 (d, J=7.0 Hz), 52.35 (d, J=7.0 Hz), 52.18 (s), 46.89 (s), 43.79 (s), 29.64 (s), 27.74 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 30.60 (s). HRMS: [M+H]$^+$ calculated for C$_{13}$H$_{21}$N$_6$O$_5$P, 373.1384; found 373.1382.

(R)-(2-((2-(6-Amino-9H-purin-9-yl)ethyl)amino)-3-methoxy-3-oxopropyl)phosphonic acid (4.21a) This compound was obtained as a white solid (212 mg, 89 %) according to the last step used for the synthesis of **4.1**, starting from methyl ester **4.20a** (259 mg, 0.696 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 7.83 (s, 1H), 7.76 (s, 1H), 4.15 (t, J = 5.7 Hz, 2H), 3.71-3.51 (m, 4H), 3.05 (t, J=6.0 Hz, 2H), 1.76-1.63 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 174.23 (s), 154.63 (s), 151.69 (s), 148.07 (s), 141.84 (s), 117.68 (s), 57.25 (s), 52.47 (s), 45.45 (s), 42.78 (s), 31.00 (s), 29.29 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 17.32 (s). HRMS: [M+H]$^+$ calculated for C$_{11}$H$_{17}$N$_6$O$_5$P, 345.1071; found 345.1066.

(R)-2-((2-(6-Amino-9H-purin-9-yl)ethyl)amino)-3-phosphonopropanoic acid (4.2) To a solution of methyl ester **4.21a** (100 mg, 0.29 mmol) in MeOH (3 mL) was added dropwise a 1 M NaOH solution (2 mL) at 0 °C. The reaction mixture was stirred at room temperature for 3 h. The solvent was removed under reduced pressure. The residue was purified by silica gel flash chromatography (*i*-PrOH/NH$_3$/H$_2$O, v/v/v, 10:1:1 to 5:1:1) to give the title compound. Further purification was performed by preparative reverse phase HPLC with a gradient of CH$_3$CN and a 0.05 M TEAB solution to afford the title compound as a white foam (38 mg, 40 %). ^{1}H NMR (300 MHz, D$_2$O) δ 8.10 (s, 1H), 8.08 (s, 1H), 4.45 (t, J = 6.2 Hz, 2H), 3.66-3.45 (m, 1H), 3.43-3.27 (m, 2H), 2.02-1.46 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 176.40 (d, J = 14.9 Hz), 155.26 (s), 152.32 (s), 148.77 (s), 142.22 (s), 118.26 (s), 60.51 (d, J = 5.4 Hz), 45.98 (s), 41.69 (s), 30.96 (s), 29.33 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 16.93 (s). HRMS: [M-H]$^-$ calculated for C$_{10}$H$_{15}$N$_6$O$_5$P, 329.0769; found 329.0784.

tert-Butyl (R)-(2-oxooxetan-3-yl)carbamate (4.17b) This compound was obtained as a white solid (520 mg, 57 %) according to the procedure used for the synthesis of compound **4.17a**, starting from (tert-butoxycarbonyl)-D-serine (1 g, 4.87 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.32 (d, J = 5.1 Hz, 1H), 5.10 (d, J = 5.8 Hz, 1H), 4.51-4.29 (m, 2H), 1.45 (s, 9H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 169.42 (s), 154.58 (s), 81.58 (s), 66.74 (s), 59.62 (s), 28.31 (s).

Methyl (S)-2-((tert-butoxycarbonyl)amino)-3-(dimethoxyphosphoryl)propanoate (4.18b) This compound was obtained as a white solid (654 mg, 76 %) according to the procedure used for the

synthesis of compound **4.18a**, starting from tert-butyl (*R*)-(2-oxooxetan-3-yl)carbamate **4.17b** (520 mg, 2.78 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.69 (d, *J* = 7.9 Hz, 1H), 4.46-4.27 (m, 1H), 3.60-3.52 (m, 9H), 2.20 (dd, *J* = 17.4, 5.5 Hz, 2H), 1.26 (s, 9H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 171.18 (d, *J* = 10.0 Hz), 154.98 (s), 79.71 (s), 60.04 (s), 52.35 (d, *J* = 5.3 Hz), 52.24 (d, *J*=2.3 Hz), 48.76 (d, *J*=3.8 Hz), 28.03 (s), 27.77(s), 25.90 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 30.08 (s). HRMS: [M+Na]$^+$ calculated for C$_{11}$H$_{22}$NO$_7$P, 334.1026; found 334.1044.

Methyl (S)-2-amino-3-(dimethoxyphosphoryl) propanoate (4.19b) This compound was isolated as a hydrochloride salt according to the procedure used for the synthesis of **4.19a**, starting from Boc-protected amine **4.18b** (300 mg, 0.96 mmol). Since this compound was the enantiomer of **4.19a**, it was detected by TLC plate and HRMS only, and compared with **4.19a** without further structural characterization. HRMS: [M+H]$^+$ calculated for C$_6$H$_{14}$NO$_5$P, 212.0682; found 212.0685.

Methyl (S)-2-((2-(6-amino-9H-purin-9-yl)ethyl)amino)-3-(dimethoxyphosphoryl)propanoate (4.20b) This compound was obtained as a white solid (236 mg, 66 %) according to the procedure used for the synthesis of compound **4.20a**, starting from amine **4.19b** (203 mg, 0.96 mmol) and aldehyde **4.10a** (170 mg, 0.96 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.24 (s, 1H), 7.99 (s, 1H), 6.52 (s, 2H), 4.35-4.07 (m, 2H), 3.75-3.44 (m, 10H), 3.26-2.74 (m, 2H), 2.23-1.92 (m, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.67 (d, *J* = 13.5 Hz), 155.82 (s), 152.84 (s), 150.04 (s), 141.44 (s), 119.51 (s), 56.28 (d, *J* = 4.5 Hz), 52.66 (d, *J*=6.0 Hz), 52.45 (d, *J*=6.8 Hz), 52.28 (s), 47.03 (s), 43.94 (s), 29.80 (s), 27.91 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 30.60 (s). HRMS: [M+H]$^+$ calculated for C$_{13}$H$_{21}$N$_6$O$_5$P, 373.1384; found 373.1353.

(S)-(2-((2-(6-Amino-9H-purin-9-yl)ethyl)amino)-3-methoxy-3-oxopropyl)phosphonic acid (4.21b) This compound was obtained as a white solid (179 mg, 82 %) according to the procedure used for the synthesis of **4.21a**, starting from methyl ester **4.20b** (236 mg, 0.63 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.12 (s,1H), 8.06 (s, 1H), 4.21 (t, *J* = 6.0 Hz, 2H), 3.62-3.46 (m, 4H), 3.07-2.79 (m, 2H), 1.99-1.57 (m, 2H). ^{13}C NMR (75 MHz, MeOD) δ 175.85 (d, *J* = 14.6 Hz), 157.20 (s), 153.51 (s), 150.76 (s), 143.45 (s), 120.01 (s), 58.97 (d, *J* = 3.3 Hz), 52.49 (s), 47.94 (s), 44.82 (s), 33.43 (s), 31.69 (s). ^{31}P NMR (121 MHz, MeOD) δ 18.68 (s). HRMS: [M-H]$^-$ calculated for C$_{11}$H$_{17}$N$_6$O$_5$P, 343.0925; found 343.0884.

(S)-2-((2-(6-Amino-9H-purin-9-yl)ethyl)amino)-3-phosphonopropanoic acid (4.3) This compound was obtained as a white solid (65 mg, 68 %) according to the procedure used for the synthesis of **4.2**, starting from methyl ester **4.21b** (100 mg, 0.29 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 7.89 (s, 1H), 7.80 (s, 1H), 4.39 (t, *J* = 12.7 Hz, 2H), 3.74-3.53 (m, 2H), 3.48-3.28 (m, 1H), 2.10-1.57 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 174.03 (d, *J* = 15.3 Hz), 155.05 (s), 152.45 (s), 148.60 (s), 142.02 (s), 118.09 (s), 60.29 (d, *J* = 5.6 Hz), 45.79 (s), 41.20 (s), 30.00 (s), 28.37 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 16.81 (s). HRMS: [M-H]$^-$ calculated for C$_{10}$H$_{15}$N$_6$O$_5$P, 329.0769; found 329.0724.

9-(3,3-Dimethoxypropyl)-9H-purin-6-amine (4.9b) This compound was obtained as a white solid (791 mg, 45 %) according to the procedure used for the synthesis of **4.9a**, starting from adenine (1 g, 7.4 mmol) and 3-bromo-1,1-dimethoxypropane **4.8b** (1.1 mL, 8.14 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.36 (s, 1H), 7.79 (s, 1H), 6.03 (s, 2H), 4.34 (t, J = 6.0 Hz, 1H), 4.28 (t, J = 6.9 Hz, 2H), 2.19 (dd, J = 12.0, 6.0 Hz, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 155.71 (s), 153.14 (s), 150.25 (s), 140.87 (s), 119.72 (s), 102.21 (s), 53.48 (s), 39.94 (s), 32.77 (s).

Methyl (S)-2-((3-(6-amino-9H-purin-9-yl)propyl)amino)-3-(dimethoxyphosphoryl) propanoate (4.22) Compound **4.9b** (406 mg, 1.53 mmol) was dissolved in 1N HCl (10 mL) and refluxed at 90 °C for 1 h. The mixture was neutralized with 10 N NaOH. The solvent was evaporated under reduced pressure, and the crude residue of **4.10b** was used for the next step without purification. To a solution of the above residue (292 mg, 1.53 mmol) in THF (6 mL) was added amine **4.19b** (339 mg, 1.60 mmol) and Na$_2$SO$_4$ (7 g). The mixture was stirred at ambient temperature for 16 h. The mixture was then filtered through a pad of Celite® and the solids were rinsed with anhydrous THF (~8 mL). The filtrate was concentrated *in vacuo*, and the crude residue was taken up in anhydrous MeOH (8 mL). NaCNBH$_3$ (96 mg, 1.53 mmol) and acetic acid (0.2 mL) were then added and the reaction mixture was stirred for 1 h at room temperature. The mixture was concentrated *in vacuo* and to the crude residue was added a saturated aqueous NaHCO$_3$ solution. The mixture was concentrated *in vacuo*. The crude residue was purified by silica gel chromatography (EA/MeOH, v/v, 20:1 to 10:1 to 5:1) to give the title compoound as a white solid (150 mg, 25 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.32 (s, 1H), 7.96 (s, 1H), 6.07 (s, 2H), 4.34 (t, J = 6.7 Hz, 2H), 3.83-3.64 (m, 9H), 3.64-3.44 (m, 1H), 2.69-2.37 (m, 2H), 2.24-2.09 (m, 2H), 2.08-1.94 (m, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.99 (d, J = 14.1 Hz), 155.69 (s), 152.99 (s), 150.26 (s), 141.37 (s), 119.88 (s), 56.38 (d, J = 3.8 Hz), 52.79 (d, J=6.0 Hz), 52.52 (d, J=6.0 Hz), 52.29 (s), 44.19 (s), 41.37 (s), 29.97 (d, J = 5.5 Hz), 28.11 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 30.80 (s). HRMS: [M+H]$^+$ calculated for C$_{14}$H$_{23}$N$_6$O$_5$P, 387.1540; found 387.1540.

(S)-(2-((3-(6-Amino-9H-purin-9-yl)propyl)amino)-3-methoxy-3-oxopropyl)phosphonic acid (4.23) This compound was obtained as a white solid (93 mg, 67 %) according to the procedure used for the synthesis of compound **4.20a**, starting from methyl ester **4.22** (150 mg, 0.39 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.16 (s, 1H), 8.15 (s, 1H), 4.31 (t, J = 6.8 Hz, 2H), 3.80-3.62 (m, 4H), 2.70 (m, 2H), 2.16-2.04 (m, 2H), 2.03-1.66 (m, 2H). ^{13}C NMR (75 MHz, MeOD) δ 174.39 (d, J = 16.1 Hz), 157.32 (s), 153.69 (s), 150.71 (s), 143.03 (s), 120.09 (s), 58.80 (d, J = 3.9 Hz), 52.87 (s), 45.33 (s), 42.42 (s), 32.32 (s), 30.62 (s), 29.94 (s). ^{31}P NMR (121 MHz, MeOD) δ 17.60 (s). HRMS: [M-H]$^-$ calculated for C$_{12}$H$_{19}$N$_6$O$_5$P, 357.1082; found 357.1089.

(S)-2-((3-(6-Amino-9H-purin-9-yl)propyl)amino)-3-phosphonopropanoic acid (4.4) This compound was obtained as a white solid (25 mg, 28 %) according to the procedure used for the

synthesis of **4.2**, starting from methyl ester **4.23** (93 mg, 0.26 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 8.10 (d, *J* = 1.4 Hz, 2H), 4.29 (t, *J* = 6.9 Hz, 2H), 3.61-3.39 (m, 1H), 2.95 (t, *J* = 11.1 Hz, 2H), 2.23 (p, *J* = 7.1 Hz, 2H), 2.10-1.50 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 175.08 (d, *J* = 17.5 Hz), 155.14 (s), 152.17 (s), 148.57 (s), 142.11 (s), 118.21 (s), 60.05 (d, *J* = 5.6 Hz), 44.06 (s), 40.86 (s), 30.40 (s), 28.79 (s), 26.69 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 16.33 (s). HRMS: [M-H]$^-$ calculated for C$_{11}$H$_{17}$N$_6$O$_5$P, 343.0925; found 343.0911.

2-(Benzyloxy)acetaldehyde (4.25) To a vigorously stirred suspension of chromatographic grade silica gel (5 g) in DCM (30 mL) was added dropwise an aqueous solution (5 mL) of sodium periodate (0.763 g, 3.567 mmol) yielding a flaky suspension. 3-Benzyloxy-1,2-propanediol **4.24** (0.5 g, 2.744 mmol) in DCM (3 mL) was then added, and the reaction was monitored by TLC until disappearance of the initial product (15 min). The mixture was then filtered, and the silica gel was thoroughly washed with DCM (3×5 mL). Evaporation of the solvent gave the crude product, which is used as such for the next step without any purification. TLC: Rf = 0.45 (hexane/EA 2:1) ^{1}H NMR (300 MHz, CDCl$_3$) δ 9.71 (d, *J* = 0.8 Hz, 1H), 7.35 – 7.28 (m, 5H), 4.62 (s, 2H), 4.09 (d, *J* = 0.8 Hz, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 200.49 (s), 128.72 (s), 128.46 (s), 128.33 (s), 128.14 (s), 75.44 (s), 73.83 (s).

Diisopropyl (3-(benzyloxy)-2-hydroxypropyl)phosphonate (4.27) To a stirred solution of diisopropyl methylphosphonate **4.26** (1.041 g, 5.488 mmol) in dry THF (10 mL), was added a solution of *n*-BuLi in hexane (2.634 mL, 2.5 M) at -78 °C under argon. The reaction mixture was stirred for 15 min and then a solution of aldehyde **4.25** (824 mg, 5.488 mmol) in dry THF (3 mL) was added dropwise. After 30 min of stirring, the temperature was increased to -20 °C and the reaction mixture was neutralized with a concentrated HCl solution to pH= 7. The solvent was evaporated under vacuum and the residue was purified by silica gel chromatography (DCM/MeOH, v/v, 100:1 to 50:1 to 30:1) to yield the title compound as a colorless oil (1.009 g, 56 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 7.42-7.16 (m, 5H), 4.78-4.64 (m, 2H), 4.56 (s, 2H), 4.18 (s, 1H), 3.83 (s, 1H), 3.50 (d, *J* = 5.1 Hz, 2H), 2.16-1.81 (m, 2H), 1.31 (d, *J* = 6.1 Hz, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 137.92 (s), 128.24 (s), 127.57 (s), 73.94 (d, *J* = 15.2 Hz), 73.20 (s), 70.39 (t, *J* = 6.5 Hz), 65.74 (d, *J* = 3.6 Hz), 32.41 (s), 30.54 (s), 23.88 (d, *J* = 3.6 Hz). ^{31}P NMR (121 MHz, CDCl$_3$) δ 28.38 (s). HRMS: [M+H]$^+$ calculated for C$_{16}$H$_{27}$O$_5$P, 331.1669; found 331.1666.

Ethyl 2-((1-(benzyloxy)-3-(diisopropoxyphosphoryl)propan-2-yl)oxy)acetate (4.28) To a cooled (-40 °C) mixture of 60 % NaH (60 mg, 1.5 mmol), dry THF (5 mL) and dry TEA (266 μL, 1.9 mmol) was added **4.27**. After 30 min of vigorous stirring, ethyl bromoacetate (182 μL, 1.64 mmol) was added and the temperature of the mixture was allowed to rise slowly to room temperature. After stirring for 6 h, diethyl ether was added, and the mixture was washed with brine. The organic layer was dried over anhydrous MgSO$_4$ and then evaporated. The crude residue was purified by silica gel

chromatography (DCM/MeOH, v/v, 100:1 to 50:1) to yield the title compound as a colorless oil (437 mg, 77 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 7.39-7.16 (m, 5H), 4.65 (s, 2H), 4.57-4.45 (m, 2H), 4.26 (dd, J = 11.4, 7.7 Hz, 2H), 4.20-4.05 (m, 2H), 3.93 (s, 1H), 3.75-3.55 (m, 2H), 2.20-1.97 (m, 2H), 1.32-1.17 (m, 15H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 170.53 (s), 138.16 (s), 128.37 (s), 127.63 (s), 75.50 (s), 73.34 (s), 73.05 (d, J=8.3 Hz), 70.34 (s), 68.15 (s), 60.71 (s), 31.20 (s), 29.33 (s), 24.04 (s), 14.20 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 26.47 (s). HRMS: [M+H]$^+$ calculated for C$_{20}$H$_{33}$O$_7$P, 417.2037; found 417.2038.

Diisopropyl (3-(benzyloxy)-2-(2-hydroxyethoxy)propyl)phosphonate (4.30a)

Intermediate ester **4.28** (437 mg, 1.05 mmol) was cooled to -20 °C and a solution of borane in THF (6 mL, 1 M) was added under argon atmosphere. The reaction mixture was stirred at room temperature for 3 days. Methanol was added at -20 °C and the solution was concentrated after the evolution of hydrogen had stopped. The residue was purified by silica gel flash chromatography (DCM/MeOH, v/v, 100:1 to 50:1) to afford the title compound as a colorless oil (330 mg, 84 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 7.42-7.21 (m, 5H), 4.83-4.62 (m, 2H), 4.55 (s, 2H), 4.02-3.91 (m, 1H), 3.87-3.77 (m, 1H), 3.72-3.41 (m, 5H), 2.09-1.92 (m, 2H), 1.32 (t, J = 5.2 Hz, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 137.89 (s), 128.37 (s), 127.68 (s), 127.58 (s), 74.14 (s), 73.34 (s), 72.57 (s), 72.25 (s), 70.33 (d, J = 16.9 Hz), 61.45 (s), 31.57 (s), 29.66 (s), 23.97 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 28.50 (s). HRMS: [M+H]$^+$ calculated for C$_{18}$H$_{31}$O$_6$P, 375.1931; found 375.1928.

Diisopropyl (3-(benzyloxy)-2-(2-(6-chloro-9H-purin-9-yl)ethoxy)propyl)phosphonate (4.31a)

A solution of PPh$_3$ (497 mg, 1.89 mmol) in dry THF (8 mL) was cooled at -30 °C under argon atmosphere. DIAD (367 μL, 1.81 mmol) was added slowly. The mixture was stirred for 30 min, and then was added to a solution of 6-chloropurine (275 mg, 1.78 mmol) and **4.30a** (306 mg, 0.88 mmol) in dry THF (5 mL) at -30 °C. The resulting mixture was slowly warmed to room temperature and stirred for 48 h. The solvent was evaporated *in vacuo* and the residue was purified by silica gel flash chromatography (DCM/MeOH, v/v, 100:1 to 60:1 to 30:1) to give the title compound as a yellow oil (345 mg, 81 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.70 (s, 1H), 8.41 (s, 1H), 7.35-7.18 (m, 5H), 4.77-4.60 (m, 2H), 4.55-4.36 (m, 4H), 4.00 (t, J = 4.2 Hz, 2H), 3.96-3.83 (m, 1H), 3.61-3.37 (m, 2H), 1.95 (dd, J = 18.5, 6.1 Hz, 2H), 1.33-1.25 (m, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 151.77 (s), 151.63 (s), 150.66 (s), 146.75 (s), 137.67 (s), 131.43 (s), 128.37 (s), 127.73 (s), 127.52 (s), 75.08 (s), 73.29 (s), 72.80 (d, J = 11.0 Hz), 70.34 (t, J = 6.4 Hz), 67.90 (s), 44.38 (s), 31.24 (s), 29.35 (s), 24.01 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 26.17 (s). HRMS: [M+H]$^+$ calculated for C$_{23}$H$_{32}$ClN$_4$O$_5$P, 511.1871; found 511.1877.

Diisopropyl (2-(2-(6-amino-9H-purin-9-yl)ethoxy)-3-(benzyloxy)propyl)phosphonate (4.32a)

A solution of **4.31a** (579 mg, 1.14 mmol) in a 7 M methanolic ammonia solution (50 mL) was stirred

in a sealed tube at 70 °C for 30 h. The solvent was evaporated, and the residue was purified by silica gel chromatography (DCM/MeOH, v/v, 60:1 to 30:1 to 10:1) to afford the title compound as a yellowish oil (402 mg, 72 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.33 (s, 1H), 8.05 (s, 1H), 7.33-7.15 (m, 5H), 6.89 (s, 2H), 4.77-4.57 (m, 2H), 4.45 (s, 2H), 4.37 (dd, J = 9.5, 4.7 Hz, 2H), 3.95 (t, J = 4.8 Hz, 2H), 3.92-3.83 (m, 1H), 3.60-3.40 (m, 2H), 2.08-1.90 (m, 2H), 1.32-1.21 (m, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 155.83 (s), 152.59 (s), 149.61 (s), 141.46 (s), 137.68 (s), 128.19 (s), 127.51 (s), 127.42 (s), 119.12 (s), 74.76 (s), 73.10 (s), 72.23 (d, J = 10.5 Hz), 70.12 (t, J=5.3 Hz), 67.96 (s), 43.65 (s), 30.94 (s), 29.06 (s), 23.83 (d, J = 4.0 Hz). ^{31}P NMR (121 MHz, CDCl$_3$) δ 26.44 (s). HRMS: [M+H]$^+$ calculated for C$_{23}$H$_{34}$N$_5$O$_5$P, 492.2370; found 492.2370.

Diisopropyl (2-(2-(6-amino-9H-purin-9-yl)ethoxy)-3-hydroxypropyl)phosphonate (4.33a) A solution of **4.32a** (402 mg, 0.818 mmol), 10 % palladium on charcoal (210 mg) in methanol (12 mL) was stirred under a hydrogen atmosphere (atmospheric pressure) at room temperature for 48 h. The mixture was filtered through a pad of Celite$^®$. The catalyst was washed with hot water (30 mL) and hot methanol (30 mL). The solvents were evaporated *in vacuo* yielding the crude title product as a yellowish oil (288 mg, 88 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.14 (s, 1H), 7.92 (s, 1H), 6.91 (s, 2H), 5.47 (s, 1H), 4.64-4.46 (m, 2H), 4.38-4.08 (m, 2H), 3.96-3.75 (m, 2H), 3.72-3.54 (m, 2H), 3.46 (dd, J = 12.0, 6.0 Hz, 1H), 1.86 (dd, J = 18.0, 6.0 Hz, 2H), 1.15 (t, J = 6.0 Hz, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 156.04 (s), 152.74 (s), 149.86 (s), 141.68 (s), 119.23 (s), 77.55 (s), 70.68 (d, J = 6.7 Hz), 68.55 (s), 64.40 (d, J = 9.1 Hz), 44.35 (s), 30.85 (s), 28.98 (s), 24.20 (d, J = 4.2 Hz). ^{31}P NMR (121 MHz, CDCl$_3$) δ 26.86 (s). HRMS: [M+H]$^+$ calculated for C$_{16}$H$_{28}$N$_5$O$_5$P, 402.1901; found 402.1898.

2-(2-(6-Amino-9H-purin-9-yl)ethoxy)-3-(diisopropoxyphosphoryl)propanoic acid (4.34a) A solution of **4.33a** (144 mg, 0.359 mmol), TEMPO (17 mg, 0.108 mmol) and NaClO$_2$ (81 mg, 0.717 mmol) in a mixture of CH$_3$CN (3 mL) and a 0.67 M aqueous sodium phosphate buffer (3 mL, pH~6.7) was vigorously stirred at room temperature. A 5 % aqueous solution of sodium hypochlorite (1 mL) was added. The mixture was stirred at room temperature until the reaction was complete (~24 h) and the solvents were evaporated. Purification by silica gel flash chromatography (DCM/MeOH, v/v, 10:1 to 6:1 to 3:1) afforded the title compound as a yellow-brown oil (54 mg, 36 %). ^{1}H NMR (300 MHz, DMSO) δ 8.32 (s, 0.5H), 8.11 (s, 0.5H), 7.19 (s, 1H), 4.57-4.42 (m, 2H), 4.28 (t, J = 15.3 Hz, 1H), 3.97-3.68 (m, 4H), 2.26-1.88 (m, 2H), 1.26-1.06 (m, 12H). ^{13}C NMR (75 MHz, DMSO) δ 173.92 (s), 155.80 (s), 152.17 (s), 149.47 (s), 141.44 (s), 118.51 (s), 75.50 (d, J = 3.9 Hz), 69.47 (t, J=5.3 Hz), 67.35 (s), 42.96 (s), 31.60 (s), 29.73 (s), 23.66 (s). ^{31}P NMR (121 MHz, DMSO) δ 27.00 (s). HRMS: [M+H]$^+$ calculated for C$_{16}$H$_{26}$N$_5$O$_6$P, 416.1693; found 416.1699.

2-(2-(6-Amino-9H-purin-9-yl)ethoxy)-3-phosphonopropanoic acid (4.5) This compound was obtained as a white solid (16 mg, 37 %) according to the procedure used for the synthesis of **4.20a**, starting from *iso*-propyl ester **4.34a** (54 mg, 0.13 mmol). [α]$^{20}_D$ = + 0.5° (c = 0.2, MeOH). ^{1}H NMR

(300 MHz, D$_2$O) δ 8.28 (s, 1H), 8.11 (s, 1H), 4.49-4.23 (m, 2H), 4.00-3.89 (m, 1H), 3.88-3.67 (m, 2H), 1.90-1.72 (m, 2H). ^{13}C NMR (75 MHz, D$_2$O) δ 180.10 (d, J = 14.1 Hz), 155.25 (s), 152.06 (s), 148.81 (s), 143.16 (s), 118.12 (s), 78.34 (d, J = 4.7 Hz), 67.37 (s), 43.56 (s), 33.47 (s), 31.73 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 19.50 (s). HRMS: [M+H]$^+$ calculated for C$_{10}$H$_{12}$N$_5$O$_6$P, 330.0598; found 330.0610.

Diisopropyl (2-(allyloxy)-3-(benzyloxy)propyl)phosphonate (4.29) This compound was obtained as a colorless oil (410 mg, 80 %) according to the procedure used for the synthesis of **4.28**, starting from allyl bromide (144 μL, 1.667 mmol) and alcohol derivative **4.27** (459 mg, 1.389 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 7.43-7.17 (m, 5H), 6.00-5.85 (m, 1H), 5.35-5.08 (m, 2H), 4.77-4.61 (m, 2H), 4.57 (s, 2H), 4.12 (dt, J = 5.7, 1.3 Hz, 2H), 3.97-3.84 (m, 1H), 3.67-3.50 (m, 2H), 2.05 (ddd, J = 18.7, 6.3, 4.4 Hz, 2H), 1.34-1.24 (m, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 138.31 (s), 135.04 (s), 128.36 (s), 127.64 (s), 116.91 (s), 73.65 (s), 73.33 (s), 72.37 (d, J = 9.5 Hz), 71.08 (s), 70.11 (t, J = 6.7 Hz), 31.22 (s), 29.35 (s), 24.10 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 27.18 (s). HRMS: [M+H]$^+$ calculated for C$_{19}$H$_{31}$O$_5$P, 371.1982; found 371.1982.

Diisopropyl (3-(benzyloxy)-2-(3-hydroxypropoxy)propyl)phosphonate (4.30b) To a solution of allyl ether **4.29** (410 mg, 1.11 mmol) in dry THF (1 mL) was added 9-BBN (2.21 mL, 0.5 M in THF). The resulting solution was stirred at room temperature for 12 h. Ethanol (1 mL), 4 N NaOH (1 mL) and 30 % H$_2$O$_2$ (1 mL) were added sequentially at 0 °C. The reaction mixture was stirred at room temperature for 2 h and quenched by adding a saturated aqueous NaHCO$_3$ solution. The mixture was extracted with EA, and the combined organic layers were dried over anhydrous MgSO$_4$ and concentrated *in vacuo*. The crude residue was further purified by silica gel chromatography (DCM/MeOH, v/v, 100:1 to 60:1 to 30:1) to obtain the title compound as a colorless oil (0.27 g, 63 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 7.40-7.21 (m, 5H), 4.79-4.60 (m, 2H), 4.56 (s, 2H), 3.89-3.63 (m, 5H), 3.53 (d, J = 5.1 Hz, 2H), 1.98 (dt, J = 15.3, 7.8 Hz, 2H), 1.83-1.66 (m, 2H), 1.30 (dd, J = 6.1, 1.8 Hz, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 138.06 (s), 128.44 (s), 127.73 (s), 74.31 (s), 73.40 (s), 72.29 (d, J = 13.5 Hz), 70.36 (dd, J = 12.0, 6.5 Hz), 67.89 (s), 59.81 (s), 32.43 (s), 31.30 (s), 29.40 (s), 24.06 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 27.59 (s). HRMS: [M+H]$^+$ calculated for C$_{19}$H$_{33}$O$_6$P, 389.2087; found 389.2081.

Diisopropyl (3-(benzyloxy)-2-(3-(6-chloro-9H-purin-9-yl)propoxy)propyl)phosphonate (4.31b) This compound was obtained as a yellow oil (679 mg, 75 %) according to the procedure used for the synthesis of **4.31a**, starting from 6-chloropurine (536 mg, 3.467 mmol) and alcohol derivative **4.30b** (670 mg, 1.725 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.72 (s, 1H), 8.42 (s, 1H), 7.39-7.24 (m, 5H), 4.81-4.61 (m, 2H), 4.57 (s, 2H), 4.54-4.40 (m, 2H), 3.92-3.76 (m, 1H), 3.66-3.46 (m, 4H), 2.26-2.09 (m, 2H), 2.08-1.88 (m, 2H), 1.38-1.24 (m, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 151.81 (s), 151.48

(s), 150.56 (s), 146.60 (s), 137.80 (s), 131.67 (s), 128.30 (s), 127.64 (s), 127.60 (s), 74.36 (s), 73.25 (s), 72.17 (d, J = 11.5 Hz), 70.17 (t, J = 7.2 Hz), 65.64 (s), 41.37 (s), 31.12 (s), 29.50 (s), 29.23 (s), 23.95 (d, J = 2.9 Hz). ^{31}P NMR (121 MHz, CDCl3) δ 26.75 (s). HRMS: [M+H]$^+$ calculated for $C_{24}H_{34}ClN_4O_5P$, 525.2028; found 525.2026.

Diisopropyl (2-(3-(6-amino-9H-purin-9-yl)propoxy)-3-(benzyloxy)propyl)phosphonate (4.32b) This compound was obtained as a yellowish oil (478 mg, 73 %) according to the procedure used for the synthesis of **4.32a**, starting from 6-chloropurine derivative **4.31b** (679 mg, 1.293 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.33 (s, 1H), 7.98 (s, 1H), 7.42-7.17 (m, 5H), 6.84 (s, 2H), 4.82-4.63 (m, 2H), 4.55 (s, 2H), 4.37 (s, 2H), 3.93-3.78 (m, 1H), 3.67-3.41 (m, 4H), 2.25-1.95 (m, 4H), 1.39-1.23 (m, 12H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 155.83 (s), 152.64 (s), 149.83 (s), 141.24 (s), 137.90 (s), 128.27 (s), 127.58 (s), 119.59 (s), 74.30 (s), 73.21 (s), 72.07 (d, J = 9.9 Hz), 70.12 (t, J = 6.7 Hz), 65.73 (s), 40.68 (s), 31.09 (s), 29.78 (s), 29.21 (s), 23.93 (d, J = 3.7 Hz). ^{31}P NMR (121 MHz, CDCl$_3$) δ 26.98 (s). HRMS: [M+H]$^+$ calculated for $C_{24}H_{36}N_5O_5P$, 506.2527; found 506.2527.

Diisopropyl (2-(3-(6-amino-9H-purin-9-yl)propoxy)-3-hydroxypropyl)phosphonate (4.33b) This compound was obtained as a yellowish oil (280 mg, 71 %) according to the procedure used for the synthesis of **4.33a**, starting from benzyl derivative **4.32b** (478 mg, 0.946 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 7.87 (s, 1H), 7.84 (s, 1H), 4.42 (dq, J = 12.0, 6.0 Hz, 2H), 4.14-3.98 (m, 2H), 3.57-3.23 (m, 5H), 1.98-1.86 (m, 2H), 1.80 (dd, J = 18.0, 6.0 Hz, 2H), 1.05 (d, J = 6.0 Hz, 12H). ^{13}C NMR (75 MHz, D$_2$O) δ 154.77 (s), 151.71 (s), 148.21 (s), 141.92 (s), 117.79 (s), 75.21 (d, J = 2.9 Hz), 72.15 (d, J = 6.7 Hz), 66.24 (s), 62.88 (d, J = 12.0 Hz), 40.73 (s), 29.04 (s), 27.20 (s), 22.90 (d, J = 3.7 Hz). ^{31}P NMR (121 MHz, D$_2$O) δ 28.73 (s). HRMS: [M+H]$^+$ calculated for $C_{17}H_{30}N_5O_5P$, 416.2057; found 416.2061.

2-(3-(6-Amino-9H-purin-9-yl)propoxy)-3-(diisopropoxyphosphoryl)propanoic acid (4.34b) This compound was obtained as a yellow-brown oil (48 mg, 31 %) according to the procedure used for the synthesis of **4.34a**, starting from alcohol derivative **4.33b** (150 mg, 0.36 mmol). ^{1}H NMR (300 MHz, DMSO) δ 8.24 (s, 1H), 8.12 (s, 1H), 7.21 (s, 1H), 4.62-4.48 (m, 2H), 4.26 (t, J=6.0 Hz, 2H), 3.83 (t, J=12.0 Hz, 2H), 3.58 (t, J=6.0 Hz, 1H), 2.29-1.89 (m, 4H), 1.21 (d, J = 6.0 Hz, 12H). ^{13}C NMR (75 MHz, DMSO) δ 174.40 (s), 155.86 (s), 152.25 (s), 149.44 (s), 141.36 (s), 118.81 (s), 75.12 (s), 69.57 (dd, J = 10.0, 6.4 Hz), 66.12 (s), 40.22 (s), 31.54 (s), 29.67 (s), 29.43 (s), 23.67 (dd, J = 9.0, 4.0 Hz). ^{31}P NMR (121 MHz, DMSO) δ 27.09 (s). HRMS: [M+H]$^+$ calculated for $C_{17}H_{28}N_5O_6P$, 430.1850; found 430.1852.

2-(3-(6-Amino-9H-purin-9-yl)propoxy)-3-phosphonopropanoic acid (4.6) This compound was obtained as a white solid (14 mg, 36 %) according to the procedure used for the synthesis of **4.5**, starting from *iso*-propyl ester **4.34b** (48 mg, 0.112 mmol). [α]$^{20}_D$ = + 1.0° (c = 0.2, MeOH). ^{1}H NMR (300 MHz, D$_2$O) δ 8.21 (s, 1H), 8.14 (s, 1H), 4.42-3.25 (m, 2H), 4.01-3.81 (m, 1H), 3.55-3.42 (m,

1H), 3.30-3.20 (m, 2H), 2.16-2.00 (m, 2H), 1.91 (dd, J = 16.4, 6.8 Hz, 2H). ^{13}C NMR (75 MHz, D_2O) δ 180.28 (d, J = 14.6 Hz), 155.28 (s), 152.06 (s), 148.66 (s), 142.87 (s), 118.31 (s), 77.57 (d, J = 5.0 Hz), 65.80 (s), 40.79 (s), 33.19 (s), 31.44 (s), 29.01 (s). ^{31}P NMR (121 MHz, D_2O) δ 20.49 (s). HRMS: [M-H]$^-$ calculated for $C_{11}H_{16}N_5O_6P$, 344.0765; found 344.0771.

Diethyl (R)-(2-((1-hydroxypropan-2-yl)amino)ethyl)phosphonate (4.37) A solution of (*R*)-2-aminopropan-1-ol **4.36** (500 mg, 6.66 mmol) and diethyl vinylphosphonate **4.35** (625 mg, 3.81 mmol) in H_2O (10 mL), was stirred for 2 days at room temperature. The solvents were evaporated *in vacuo* and the crude residue was purified by silica gel flash chromatography (DCM/MeOH, 10:1), yielding the title compound as a light-yellow oil (855 mg, 94 %). ^{1}H NMR (300 MHz, CDCl3) δ 4.11-3.89 (m, 4H), 3.47 (dd, J = 10.8, 3.9 Hz, 1H), 3.21 (dd, J = 10.8, 7.3 Hz, 1H), 3.08-2.83 (m, 3H), 2.82-2.61 (m, 2H), 1.89 (dt, J = 18.1, 7.2 Hz, 2H), 1.23 (t, J = 7.1 Hz, 6H), 0.94 (d, J = 6.5 Hz, 3H). ^{13}C NMR (75 MHz, CDCl3) δ 65.39 (s), 61.68 (t, J=6.0 Hz), 54.28 (s), 40.38 (d, J = 3.6 Hz), 27.47 (s), 25.62 (s), 16.75 (s), 16.37 (d, J = 5.9 Hz). ^{31}P NMR (121 MHz, CDCl3) δ 31.20 (s). HRMS: [M+H]$^+$ calculated for $C_9H_{22}NO_4P$, 240.1359; found 240.1340.

Diethyl (R)-(2-((1-((tert-butyldiphenylsilyl)oxy)propan-2-yl)amino)ethyl)phosphonate (4.38) To a solution of alcohol **4.37** (875 mg, 3.657 mmol) and imidazole (374 mg, 5.486 mmol) in DMF (10 mL) was added TBDPS-Cl (1.123 mL, 4.389 mmol). The resulting mixture was stirred at room temperature overnight and quenched by adding a saturated aqueous $NaHCO_3$ solution. The mixture was extracted with ethyl acetate. The combined organic layers were washed with brine, dried over anhydrous $MgSO_4$ and concentrated *in vacuo*. The crude residue was then purified by silica gel chromatography (DCM/MeOH, v/v, 60:1 to 30:1) to obtain the title compound as a colorless oil (1.397 g, 80 %). ^{1}H NMR (300 MHz, CDCl3) δ 7.70-7.57 (m, 4H), 7.42-7.31 (m, 6H), 4.20-3.98 (m, 4H), 3.61-3.44 (m, 2H), 3.03-2.74 (m, 3H), 2.06 (s, 1H), 1.95 (dt, J = 18.3, 7.6 Hz, 2H), 1.30 (t, J = 7.1 Hz, 6H), 1.04 (s, 9H), 0.98 (d, J = 6.4 Hz, 3H). ^{13}C NMR (75 MHz, CDCl3) δ 135.64 (d, J = 1.4 Hz), 133.63 (s), 129.76 (s), 127.77 (s), 68.03 (s), 61.64 (d, J = 6.4 Hz), 54.49 (s), 40.91 (d, J = 2.8 Hz), 28.14 (s), 26.97 (s), 26.30 (s), 19.34 (s), 17.09 (s), 16.53 (d, J = 5.8 Hz). ^{31}P NMR (121 MHz, CDCl3) δ 30.95 (s). HRMS: [M+H]$^+$ calculated for $C_{25}H_{40}NO_4PSi$, 478.2537; found 478.2532.

Methyl (R)-N-(1-((tert-butyldiphenylsilyl)oxy)propan-2-yl)-N-(2-(diethoxyphosphoryl)ethyl)glycinate (4.39) To an ice-cooled solution of secondary amine **4.38** (700 mg, 1.465 mmol) and K_2CO_3 (263 mg, 1.905 mmol) in dry CH_3CN (6 mL) was added methyl bromoacetate (166.5 μL, 1.759 mmol). The resulting mixture was stirred at room temperature overnight. The mixture was diluted with water and extracted with ethyl acetate. The combined organic layers were washed with brine, dried over anhydrous $MgSO_4$ and concentrated *in vacuo*. The residue was purified by silica gel chromatography (DCM/MeOH, v/v, 150:1 to 100:1 to 60:1) affording the title compound as a colorless oil (766 mg, 95 %). ^{1}H NMR (300 MHz, CDCl3) δ 7.74-7.55 (m, 4H), 7.47-7.32 (m, 6H),

4.05 (p, *J* = 7.3 Hz, 4H), 3.68-3.59 (m, 4H), 3.58-3.32 (m, 3H), 3.10-2.75 (m, 3H), 2.00-1.78 (m, 2H), 1.28 (t, *J* = 7.1 Hz, 6H), 1.07 (d, *J* = 6.9 Hz, 3H), 1.04 (s, 9H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.04-172.78 (m), 135.74 (d, *J* = 1.9 Hz), 129.81 (s), 127.81 (s), 66.67 (s), 61.56 (d, *J* = 6.4 Hz), 58.32 (s), 52.51 (s), 52.07 (d, *J* = 64.8 Hz), 46.25 (s), 26.99 (s), 25.41 (s), 19.30 (s), 16.56 (d, *J* = 6.0 Hz), 14.28 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 30.83 (s). HRMS: [M+H]$^+$ calculated for C$_{28}$H$_{44}$NO$_6$PSi, 550.2748; found 550.2727.

Diethyl (R)-(2-(5-methyl-2-oxomorpholino)ethyl)phosphonate (4.40) This by-product was obtained as a colorless oil (233 mg, 50 %) according to the procedure used for the synthesis of **4.39**, starting from intermediate **4.37** (400 mg, 1.672 mmol), K$_2$CO$_3$ (300 mg, 2.173 mmol) and methyl bromoacetate (222 μL, 2.006 mmol). Alternatively, to an ice-cooled solution of intermediate **4.39** (766 mg, 1.393 mmol) in dry THF (10 mL) was added TBAF (3 mL, 1 M in THF). The resulting mixture was stirred at room temperature for 2 h. The solvent was evaporated *in vacuo* and the resulting residue was purified by silica gel chromatography (DCM/MeOH, v/v, 50:1 to 30:1 to 15:1) affording the title compound as a colorless oil (218 mg, 56 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 4.16 (dd, *J* = 10.8, 3.4 Hz, 1H), 4.06-3.83 (m, 5H), 3.41 (dd, *J* = 17.6, 2.3 Hz, 1H), 3.10 (dd, *J* = 17.6, 2.4 Hz, 1H), 2.90-2.65 (m, 2H), 2.65-2.47 (m, 1H), 1.76 (dt, *J* = 19.2, 7.9 Hz, 2H), 1.24-1.13 (m, 6H), 1.01-0.92 (m, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 167.49 (s), 73.33 (s), 61.50 (d, *J* = 6.5 Hz), 51.83 (s), 50.58 (s), 46.22 (s), 23.60 (s), 21.74 (s), 16.24 (d, *J* = 5.9 Hz), 12.25 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 29.94 (s). HRMS: [M+H]$^+$ calculated for C$_{11}$H$_{22}$NO$_5$P, 280.1308; found 280.1315.

tert-Butyl (R)-(1-hydroxypropan-2-yl)carbamate (4.42) To a solution of (*R*)-(-)-2-amino-1-propanol **4.36** (2.00 g, 26.6 mmol) and triethylamine (4.27 mL, 30.6 mmol) in MeOH (20 mL) at 0 °C was added di-*tert*-butyl dicarbonate (6.42 g, 29.4 mmol). After stirring at room temperature for 12 h, the solvent was evaporated *in vacuo*. The residue was diluted with DCM. After washing with water, the organic phase was evaporated to dryness furnishing the title compound (4.60 g, 100 %), which was used without further purification in the next step. ^{1}H NMR (300 MHz, CDCl$_3$) δ 4.91 (d, *J* = 6.3 Hz, 1H), 3.70 (s, 1H), 3.56 (dt, *J* = 10.8, 5.6 Hz, 1H), 3.45 (dd, *J* = 11.5, 5.6 Hz, 2H), 1.39 (s, 9H), 1.10 (d, *J* = 6.8 Hz, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 156.91 (s), 80.15 (s), 67.46 (s), 49.05 (s), 28.99 (s), 17.92 (s).

tert-Butyl (R)-(1-(6-chloro-9H-purin-9-yl)propan-2-yl)carbamate (4.43) Alcohol derivative **4.42** (765 mg, 4.36 mmol) was dissolved in dry DCM (30 mL). To this solution was added sequentially 6-chloropurine (1.011 g, 6.54 mmol) and PPh$_3$ (1.83 g, 6.98 mmol). Finally, DIAD (1.37 mL, 6.98 mmol) was added dropwise at 0 °C. The mixture was stirred at room temperature for 6 h. The solvents were evaporated *in vacuo*, and the residue was purified by silica gel flash chromatography (DCM/MeOH, v/v, 100:1 to 60:1 to 30:1) to yield the title compound as a yellow oil

(486 mg, 36 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.49 (s, 1H), 7.89 (s, 1H), 4.98 (d, J = 6.0 Hz, 1H), 4.40-4.22 (m, 2H), 4.11-3.95 (m, 1H), 1.31 (s, 9H), 1.15 (d, J = 6.0 Hz, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 161.13 (s), 152.51 (s), 152.08 (s), 151.91 (s), 142.81 (s), 121.19 (s), 79.80 (s), 48.39 (s), 47.05 (s), 28.30 (s), 18.21 (s). HRMS: [M+H]$^+$ calculated for C$_{13}$H$_{18}$ClN$_5$O$_2$, 312.1222; found 312.1211.

tert*-Butyl (*R*)-(1-(6-amino-9H-purin-9-yl)propan-2-yl)carbamate *(4.44) This compound was obtained as a white solid (501 mg, 65 %) according to the procedure used for the synthesis of **4.32a**, starting from 6-chloropurine derivative **4.43** (824 mg, 2.649 mmol). ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.31 (s, 1H), 7.78 (s, 1H), 6.22 (s, 2H), 5.11 (s, 1H), 4.29 (dd, J = 14.0, 4.6 Hz, 2H), 4.12-4.00 (m, 1H), 1.33 (s, 9H), 1.17 (d, J = 5.5 Hz, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 155.87 (s), 153.07 (s), 150.58 (s), 141.08 (s), 119.47 (s), 79.84 (s), 48.26 (s), 47.27 (s), 28.35 (s), 18.26 (s). HRMS: [M+H]$^+$ calculated for C$_{13}$H$_{20}$N$_6$O$_2$, 293.1720; found 293.1716.

(R)*-9-(2-Aminopropyl)-9H-purin-6-amine *(4.45) This compound was obtained as a white solid (168 mg, 87 %) according to the procedure used for the synthesis of **4.19a**, starting from Boc-protected intermediate **4.44** (294 mg, 1.01 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.17 (s, 1H), 8.08 (s, 1H), 4.12 (qd, J = 14.0, 6.3 Hz, 2H), 3.46-3.33 (m, 1H), 1.09 (d, J = 6.5 Hz, 3H). ^{13}C NMR (75 MHz, MeOD) δ 157.35 (s), 153.75 (s), 150.95 (s), 143.11 (s), 120.06 (s), 51.52 (s), 47.99 (s), 19.83 (s). HRMS: [M+H]$^+$ calculated for C$_8$H$_{12}$N$_6$, 193.1196; found 193.1189.

Diethyl (*R*)-(2-((1-(6-amino-9H-purin-9-yl)propan-2-yl)amino)ethyl)phosphonate *(4.46) To a solution of amine derivative **4.45** (355 mg, 1.847 mmol) in H$_2$O (10 mL) was added diethyl vinylphosphonate **4.35** (286 μL, 1.847 mmol). After stirring at room temperature for 2 days, the solvent was evaporated *in vacuo*. The residue was purified by silica gel chromatography (DCM/MeOH, 10:1 to 5:1) to afford the title compound as a white solid (173 mg, 26 %). ^{1}H NMR (300 MHz, MeOD) δ 8.18 (s, 1H), 8.13 (s, 1H), 4.16 (d, J = 6.0 Hz, 2H), 4.11-3.95 (m, 4H), 3.28 (dt, J = 3.3, 1.6 Hz, 1H), 3.19 (dd, J = 12.5, 6.2 Hz, 1H), 3.00-2.67 (m, 2H), 1.93 (dt, J = 18.4, 7.6 Hz, 2H), 1.26 (t, J = 7.1 Hz, 6H), 1.04 (d, J = 6.5 Hz, 3H). ^{13}C NMR (75 MHz, MeOD) δ 157.28 (s), 153.70 (s), 151.05 (s), 143.47 (s), 119.89 (s), 63.28 (d, J = 6.6 Hz), 53.59 (s), 49.65 (s), 41.48 (s), 27.52 (s), 25.68 (s), 18.02 (s), 16.70 (d, J = 6.0 Hz). ^{31}P NMR (121 MHz, MeOD) δ 30.98 (s). HRMS: [M+H]$^+$ calculated for C$_{14}$H$_{25}$N$_6$O$_3$P, 357.1798; found 357.1797.

Methyl (*R*)-N-(1-(6-amino-9H-purin-9-yl)propan-2-yl)-N-(2-(diethoxyphosphoryl)ethyl) glycinate *(4.47) To an ice-cooled mixture of amine derivative **4.46** (164 mg, 0.46 mmol) and K$_2$CO$_3$ (83 mg, 0.60 mmol) in dry CH$_3$CN (5 mL) was added methyl bromoacetate (52 μL, 0.55 mmol). After stirring at room temperature for 40 h, the solvent was evaporated *in vacuo*. The residue was purified by silica gel chromatography (DCM/MeOH, v/v, 60:1 to 40:1 to 20:1) to obtain yellowish solid (100

mg, 51 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.32 (s, 1H), 8.07 (s, 1H), 5.82 (s, 2H), 4.21-3.94 (m, 6H), 3.60 (s, 2H), 3.40-3.20 (m, 3H), 2.95-2.72 (m, 2H), 2.40-2.21 (m, 1H), 1.70 (dt, *J* = 18.9, 8.1 Hz, 2H), 1.27 (t, *J* = 7.1 Hz, 6H), 1.10 (d, *J* = 6.8 Hz, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 172.16 (s), 155.48 (s), 152.85 (s), 150.27 (s), 142.16 (s), 119.43 (s), 77.58 (s), 77.16 (s), 76.74 (s), 61.75 (d, *J* = 6.5 Hz), 56.88 (s), 51.92 (s), 51.30 (s), 46.92 (s), 45.12 (s), 26.45 (s), 24.64 (s), 16.54 (d, *J* = 6.0 Hz), 13.16 (s). ^{31}P NMR (121 MHz, CDCl$_3$) δ 30.07 (s). HRMS: [M+H]$^+$ calculated for C$_{17}$H$_{29}$N$_6$O$_5$P, 429.2010; found 429.2006.

(R)-(2-((1-(6-Amino-9H-purin-9-yl)propan-2-yl)(2-methoxy-2-oxoethyl)amino)ethyl)
phosphonic acid (4.48) This compound was obtained as a white solid (62 mg, 71 %) according to the procedure used for the synthesis of **4.21a**, starting from iso-propyl ester **4.47** (100 mg, 0.233 mmol). ^{1}H NMR (300 MHz, D$_2$O) δ 8.06 (s, 1H), 8.05 (s, 1H), 4.09 (ddd, *J* = 20.0, 14.6, 7.4 Hz, 2H), 3.39 (s, 3H), 3.25 (s, 2H), 3.07-2.85 (m, 1H), 2.80-2.58 (m, 2H), 1.65-1.39 (m, 2H), 1.00 (d, *J* = 6.6 Hz, 3H). ^{13}C NMR (75 MHz, D$_2$O) δ 174.02 (s), 154.86 (s), 151.61 (s), 148.42 (s), 142.83 (s), 117.65 (s), 55.43 (s), 51.72 (s), 50.87 (s), 45.71 (s), 44.92 (s), 27.61 (s), 25.91 (s), 10.61 (s). ^{31}P NMR (121 MHz, D$_2$O) δ 22.15 (s).

(R)-N-(1-(6-Amino-9H-purin-9-yl)propan-2-yl)-N-(2-phosphonoethyl)glycine (4.7) This compound was obtained as a white solid (31 mg, 52 %) according to the procedure used for the synthesis of **4.2**, starting from methyl ester **4.48** (62 mg, 0.167 mmol). ^{1}H NMR (500 MHz, D$_2$O) δ 8.21 (s, 1H), 8.18 (s, 1H), 4.57 (ddd, *J* = 22.1, 14.9, 6.7 Hz, 2H), 4.13-4.00 (m, 1H), 3.76 (dd, *J* = 62.9, 16.6 Hz, 2H), 3.55-3.31 (m, 2H), 1.88-1.75 (m, 2H), 1.32 (d, *J* = 6.8 Hz, 3H). ^{13}C NMR (126 MHz, D$_2$O) δ 171.35 (s), 155.44 (s), 152.63 (s), 149.06 (s), 142.35 (s), 118.21 (s), 59.36 (s), 52.55 (s), 50.34 (s), 44.28 (s), 25.22 (s), 24.24 (s), 11.14 (s). ^{31}P NMR (202 MHz, D$_2$O) δ 15.50 (s). HRMS: [M-H]$^-$ calculated for C$_{12}$H$_{19}$N$_6$O$_5$P, 357.1082; found 357.1071.

For experimental details of **HIV-1 RT inhibition assay** in this chapter, please refer to Experimental section of Chapter 3.

For experimental details of **RT expression, RT/DNA cross-linking, purification & Crystallization and structure determination** in this chapter, please refer to Experimental section of Chapter 2.

4.5 References

1. *UNAIDS Global HIV & AIDS statistics - 2020 fact sheet*. http:// www.unaids.org/en/resources/fact-sheet (accessed Aug 11, 2021).
2. S Fauci, A. HIV and AIDS: 20 years of science. *Nat. Med.* **2003,** 9, 839-843.
3. Atta, M. G.; De Seigneux, S.; Lucas, G. M. Clinical pharmacology in HIV therapy. *Clin. J. Am. Soc. Nephrol.* **2019,** 14, 435-444.
4. Park, K. H.; Kim, M.; Bae, S. E.; Lee, H. J.; Kim, K.-C.; Choi, B. S.; KIm, Y. B. Study on suitable analysis method for HIV-1 non-catalytic integrase inhibitor. *Virology Journal* **2021,** 18, 1-9.
5. Pirrone, V.; Thakkar, N.; Jacobson, J. M.; Wigdahl, B. Combinatorial approaches to the prevention and treatment of

HIV-1 infection. *Antimicrob. Agents Chemother.* **2011,** 55, 1831-1842.

6. Kankanala, J.; Kirby, K. A.; Huber, A. D.; Casey, M. C.; Wilson, D. J.; Sarafianos, S. G.; Wang, Z. Design, synthesis and biological evaluations of N-hydroxy thienopyrimidine-2,4-diones as inhibitors of HIV reverse transcriptase-associated RNase H. *Eur. J. Med. Chem.* **2017,** 141, 149-161.

7. Kirby, K. A.; Myshakina, N. A.; Christen, M. T.; Chen, Y. L.; Schmidt, H. A.; Huber, A. D.; Xi, Z. Y.; Kim, S.; Rao, R. K.; Kramer, S. T.; Yang, Q. Y.; Singh, K.; Parniak, M. A.; Wang, Z. Q.; Ishima, R.; Sarafianos, S. G. A 2-Hydroxyisoquinoline-1,3-Dione Active-Site RNase H Inhibitor Binds in Multiple Modes to HIV-1 Reverse Transcriptase. *Antimicrob Agents Ch* **2017,** 61.

8. Wang, L.; Tang, J.; Huber, A. D.; Casey, M. C.; Kirby, K. A.; Wilson, D. J.; Kankanala, J.; Parniak, M. A.; Sarafianos, S. G.; Wang, Z. Biphenylmethyl-3-hydroxypyrimidine-2,4-diones potently and selectively inhibited HIV reverse transcriptase-associated RNase H. *Eur. J. Med. Chem.* **2018,** 156, 680-691.

9. Boyer, P. L.; Smith, S. J.; Zhao, X. Z.; Das, K.; Gruber, K.; Arnold, E.; Burke, T. R.; Hughes, S. H. Developing and Evaluating Inhibitors against the RNase H Active Site of HIV-1 Reverse Transcriptase. *J Virol* **2018,** 92.

10. Jochmans, D.; Deval, J.; Kesteleyn, B.; Van Marck, H. Indolopyridones Inhibit Human Immunodeficiency Virus Reverse Transcriptase with a Novel Mechanism of Action. *J. Virol.* **2006,** 80, 12283-12292.

11. Zhang, Z.; Walker, M.; Xu, W.; Hoon Shim, J. Novel Nonnucleoside Inhibitors That Select Nucleoside Inhibitor Resistance Mutations in Human Immunodeficiency Virus Type 1 Reverse Transcriptase. *Antimicrob. Agents Chemother.* **2006,** 50, 2772-2781.

12. Ruiz, F. X.; Hoang, A.; Das, K.; Arnold, E. Structural Basis of HIV-1 Inhibition by Nucleotide-Competing Reverse Transcriptase Inhibitor INDOPY-1. *Journal of Medicinal Chemistry* **2019,** 62, 9996-10002.

13. Debarge, S.; Balzarini, J.; Maguire, A. R. Design and Synthesis of alpha-Carboxy Phosphononucleosides. *J Org Chem* **2011,** 76, 105-126.

14. Balzarini, J.; Das, K.; Bernatchez, J. A.; Martinez, S. E.; Ngure, M.; Keane, S.; Ford, A.; Maguire, N.; Mullins, N.; John, J.; Kim, Y.; Dehaen, W.; Vande Voorde, J.; Liekens, S.; Naesens, L.; Gotte, M.; Maguire, A. R.; Arnold, E. Alpha-carboxy nucleoside phosphonates as universal nucleoside triphosphate mimics. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** 112, 3475-3480.

15. Keane, S. J.; Ford, A.; Mullins, N. D.; Maguire, N. M.; Legigan, T.; Balzarini, J.; Maguire, A. R. Design and synthesis of alpha-carboxy nucleoside phosphonate analogues and evaluation as HIV-1 reverse transcriptase-targeting agents. *J Org Chem* **2015,** 80, 2479-93.

16. Das, K.; Balzarini, J.; Miller, M. T.; Maguire, A. R.; DeStefano, J. J.; Arnold, E. Conformational States of HIV-1 Reverse Transcriptase for Nucleotide Incorporation vs Pyrophosphorolysis-Binding of Foscarnet. *Acs Chem Biol* **2016,** 11, 2158-64.

17. Balzarini, J.; Ford, A.; Maguire, N. M.; John, J.; Das, K.; Arnold, E.; Dehaen, W.; Maguire, A. Alpha-carboxynucleoside phosphonates: direct-acting inhibitors of viral DNA polymerases. *Future Med Chem* **2019,** 11, 137-154.

18. Gu, W.; Martinez, S.; Nguyen, H.; Xu, H.; Herdewijn, P.; De Jonghe, S.; Das, D. Tenofovir-Amino Acid Conjugates Act as Polymerase Substrates - Implications for Avoiding Cellular Phosphorylation in the Discovery of Nucleotide Analogues. *J. Med. Chem.* **2021,** 64, 782-796.

19. Palazzolo, M. A.; Nigro, M. J.; Iribarren, A. M.; Lewkowicz, E. S. A Chemoenzymatic Route To Prepare Acyclic Nucleoside Analogues. *European Journal of Organic Chemistry* **2016,** 2016, 921-924.

20. Richard F. Borch; Mark D. Bernstein; Durst, H. D. The Cyanohydridoborate Anion as a Selective Reducing Agent. *J. Am. Chem. Soc.* **1971,** 93, 2897-2904.

21. Edward C. R. Smith; Loretta A. McQuaid; Jonathan W. Paschal; DeHoniesto., J. An Enantioselective Synthesis of D-(-)-and L-(+)-2-Amino-3-phosphonopropanoic Acid. *J. Org. Chem.* **1990,** 55, 4472-4474.

22. Hockova, D.; Holy, A.; Masojidkova, M.; Keough, D. T.; de Jersey, J.; Guddat, L. W. Synthesis of branched 9-[2-(2-phosphonoethoxy)ethyl]purines as a new class of acyclic nucleoside phosphonates which inhibit Plasmodium falciparum hypoxanthine-guanine-xanthine phosphoribosyltransferase. *Bioorganic & medicinal chemistry* **2009,** 17, 6218-6232.

23. Hockova, D.; Holy, A.; Andrei, G.; Snoeck, R.; Balzarini, J. Acyclic nucleoside phosphonates with a branched 2-(2-phosphonoethoxy)ethyl chain: Efficient synthesis and antiviral activity. *Bioorganic & medicinal chemistry* **2011,** 19, 4445-4453.

24. Kaiser, M. M.; Jansa, P.; Dracinsky, M.; Janeba, Z. A novel type of acyclic nucleoside phosphonates derived from 2-(phosphonomethoxy)propanoic acid. *Tetrahedron* **2012,** 68, 4003-4012.

25. Hockova, D.; Keough, D. T.; Janeba, Z.; Wang, T. H.; de Jersey, J.; Guddat, L. W. Synthesis of Novel N-Branched Acyclic Nucleoside Phosphonates As Potent and Selective Inhibitors of Human, Plasmodium falciparum and Plasmodium vivax 6-Oxopurine Phosphoribosyltransferases. *Journal of Medicinal Chemistry* **2012,** 55, 6209-6223.

26. Barnard, J. P.; Huber, K. D.; Sluis-Cremer, N. Nonnucleoside Reverse Transcriptase Inhibitor Hypersusceptibility and Resistance by Mutation of Residue 181 in HIV-1 Reverse Transcriptase. *Antimicrob Agents Ch* **2019,** 63.

27. Das, K.; Martinez, S. E.; Bauman, J. D.; Arnold, E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol* **2012,** 19, 253-9.

28. Sarafianos, S. G.; Pandey, V. N.; Kaushik, N.; Modak, M. J. Site-directed mutagenesis of arginine 72 of HIV-1 reverse transcriptase. Catalytic role and inhibitor sensitivity. *J Biol Chem* **1995,** 270, 19729-35.

29. Das, K.; Bandwar, R. P.; White, K. L.; Feng, J. Y.; Sarafianos, S. G.; Tuske, S.; Tu, X.; Clark, A. D., Jr.; Boyer, P. L.; Hou, X.; Gaffney, B. L.; Jones, R. A.; Miller, M. D.; Hughes, S. H.; Arnold, E. Structural basis for the role of the K65R

mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *J. Biol. Chem.* **2009,** 284, 35092-100.

30. Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **1998,** 282, 1669-1675.

Chapter 5

Design and synthesis of compounds targeting a novel P-site pocket of HIV-1 RT discovered by a crystallographic fragment screening study

This chapter is based on the following manuscript: Abhimanyu K. Singh,[1] Sergio E. Martinez,[1] Weijie Gu, Hoai Nguyen, Dominique Schols, Piet Herdewijn, Steven De Jonghe, Kalyan Das[*] Sliding of HIV-1 reverse transcriptase over DNA creates a transient P pocket – Targeting P-pocket by fragment screening. DOI:10.21203/rs.3.rs-685322/v1. (Under review in *Nature Communications*)

5.1 Introduction

Antiretroviral therapy (ART) is the current HIV treatment regimen which involves a combination of three or more HIV drugs from at least two different drug classes. Nowadays, nearly 30 drugs in five major classes have been approved for HIV treatment, namely entry/attachment inhibitors, NRTIs, NNRTIs, integrase inhibitors (INIs) and protease inhibitors (PIs). They target key steps of the HIV life cycle. Since its introduction, ART has achieved significant success in controlling HIV infection, making HIV a chronic condition. However, the development of HIV drug resistance and adverse effects from long-term use of these drug regimens limit the efficacy of ART. Therefore, there is an urgent need to find new anti-HIV drug candidates with novel mechanism of action, increased potency and reduced side effects.

Reverse transcription of viral RNA into dsDNA is a central step in the HIV-1 life cycle.[1] In this process, HIV-1 RT binds its substrate with the primer 3' end nucleotide occupying the priming site (P site). A complementary dNTP of first template overhang binds the nucleotide-binding site (N site) and the nucleotide part is catalytically incorporated by HIV-1 RT at the 3'-end of the primer accompanied by the release of a pyrophosphate. Following nucleotide incorporation every time, the growing DNA strand is translocated from the N site to the P site to accommodate the next dNTP. Currently approved RT inhibitors include NRTIs and NNRTIs. NRTIs need to be phosphorylated into their active forms and then competitively bind to the N site. Because NRTIs are incorporated into a growing DNA strand as modified nucleotides and block the incorporation of natural nucleotides, they are DNA-chain terminators. In contrast, NNRTIs allosterically block DNA polymerization by binding to a pocket adjacent to the N site. HIV-1 RT is responsible for the catalytic incorporation of an NRTI. RT mutations lead to the development of NRTI and NNRTI resistance.

Polymerases can also be inhibited directly by small molecules that prevent (d)NTP binding at the N site. NcRTIs have been investigated as potential drug candidates. Unlike a NRTI, NcRTIs do not require cellular phosphorylations. NcRTIs competitively bind at the N site and can be classified as metal-dependent inhibitors such as α-CNPs[2-5] and metal-independent inhibitors such as INDOPYs.[6-7] The discovery of novel druggable sites that are highly conserved is of great importance for developing new classes of drugs with resilience to existing drug-resistance mutations.

Fragment screening is a well-established approach for the discovery of hit compounds.[8] Various techniques, such as surface plasmon resonance, ligand-detected NMR spectroscopy, thermal shift assay and X-ray crystallography, are being used for fragment screening. In particular, X-ray crystallography emerged as a powerful technique, because of its sensitivity and the provision of precise structural information on the interaction between the fragment and its biological target.[9] Once detailed structural information is obtained, structure-based design approaches can be utilized to

design and synthesize novel and potent inhibitors. Molecular docking is one of the most frequently used structure-based design strategies due to their low cost and wide range of applications in the analysis of molecular recognition events, such as molecular interactions, binding energy as well as conformational changes.[10]

Recently, our lab performed a fragment screening of 300 drug-like fragments against HIV-1 RT, using the state-of-the-art XChem facility at Diamond Light Source. It led to the identification of two fragment hits, binding at the P-pocket that is flanked by highly conserved residues. Using the X-ray structure of hit 166 in complex with HIV-1 RT/dsDNA, 84 new compounds were designed that may engage in additional interactions at the P-pocket. A molecular docking study was conducted to investigate their binding affinity and orientations. Five new analogs of the fragment hit 166 were prepared, and the synthesis of another two selected compounds was attempted.

5.2 Results and discussion
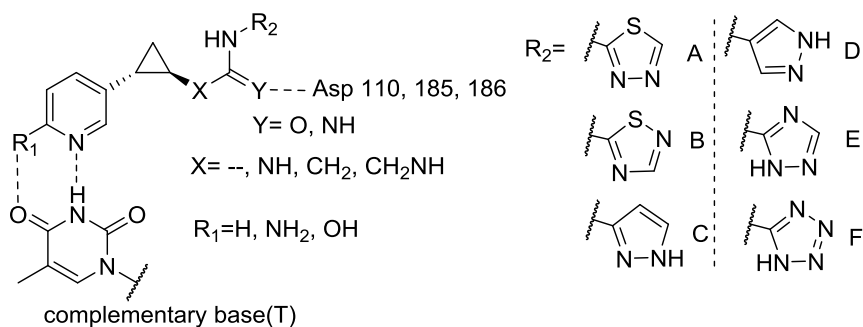
5.2.1 Fragments design and docking study

By analysis of the binding mode of fragment **166**, we envisioned the opportunity of this fragment to form H-bond with the template thymine base. Therefore, the phenyl ring was replaced by a pyridine moiety. In addition, an amino or hydroxyl group was introduced at the 5'-position of the pyridine ring so that two H-bonds might be formed, in analogy to the natural base-pairing. At the same time, we maintained the amide bond of **166**, and made variations of the thiazole moiety with different five-membered heteroaromatics. Eighteen (F01-F18) analogs of **166** were modeled into HIV-RT by accommodating the abovementioned structural variations (**Table 5-1**).

In order to investigate the favorability of binding of the designed fragments in the P pocket, we conducted a molecular docking study. Initially, rigid docking was performed, followed by flexible docking using AutoDock Vina (ver. 1.1.1).[11] For flexible docking, we defined the catalytic residues D110, D185 and D186 of HIV-1 RT as being flexible. Analysis of docking results revealed that ten fragments (F03-F08, F10-F13) have the potential to form the desired H-bond with the template thymine base. In addition, the newly introduced heteroaromatics can engage in additional H-bond formation with the surrounding residues and the DNA primer. Specifically, fragments F03, F08, F10-11 formed one H-bond with the 3'-hydroxyl group of the overhang primer DC821, while F04-F06 formed one H-bond with the backbone NH of residue M230. Moreover, F07 and F12-13 formed H-bonds with both M230 and the primer 3'-end nucleotide. However, all the above-mentioned fragments did not engage in interactions with the catalytic residues D110, D185 and D186.

The guanidinium group of the side-chain of arginine interacts with carboxylic acid groups via salt bridges, as found in many crystal structures of enzyme complexes with oxoanionic substrates and

simple guanidinium salts.[12] To explore the possibility of forming such type of interactions with the catalytic residues, we decided to substitute the amide linkage for an amidine or guanidine linker and make small variation on the linker length to increase flexibility, resulting in the design of the fragments F19-F84 (**Table 5-1**).
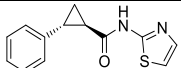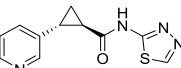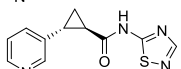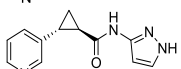
Table 5-1. Designed fragments.

HN–R₂

X–Y --- Asp 110, 185, 186

Y= O, NH

X= --, NH, CH₂, CH₂NH

R₁=H, NH₂, OH

R₂= A, B, C, D, E, F

complementary base(T)
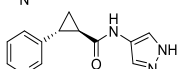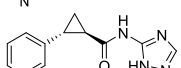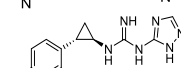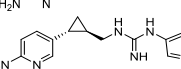
Frag.	R₁	R₂	X	Y	Frag.	R₁	R₂	X	Y	Frag.	R₁	R₂	X	Y
F01	H	A	--	O	F29	NH₂	E	--	NH	F57	H	C	CH₂	NH
F02	H	B	--	O	F30	NH₂	F	--	NH	F58	H	D	CH₂	NH
F03	H	C	--	O	F31	OH	A	--	NH	F59	H	E	CH₂	NH
F04	H	D	--	O	F32	OH	B	--	NH	F60	H	F	CH₂	NH
F05	H	E	--	O	F33	OH	C	--	NH	F61	NH₂	A	CH₂	NH
F06	H	F	--	O	F34	OH	D	--	NH	F62	NH₂	B	CH₂	NH
F07	NH₂	A	--	O	F35	OH	E	--	NH	F63	NH₂	C	CH₂	NH
F08	NH₂	B	--	O	F36	OH	F	--	NH	F64	NH₂	D	CH₂	NH
F09	NH₂	C	--	O	F37	H	A	NH	NH	F65	NH₂	E	CH₂	NH
F10	NH₂	D	--	O	F38	H	B	NH	NH	F66	NH₂	F	CH₂	NH
F11	NH₂	E	--	O	F39	H	C	NH	NH	F67	OH	A	CH₂	NH
F12	NH₂	F	--	O	F40	H	D	NH	NH	F68	OH	B	CH₂	NH
F13	OH	A	--	O	F41	H	E	NH	NH	F69	OH	C	CH₂	NH
F14	OH	B	--	O	F42	H	F	NH	NH	F70	OH	D	CH₂	NH
F15	OH	C	--	O	F43	NH₂	A	NH	NH	F71	OH	E	CH₂	NH
F16	OH	D	--	O	F44	NH₂	B	NH	NH	F72	OH	F	CH₂	NH
F17	OH	E	--	O	F45	NH₂	C	NH	NH	F73	H	A	CH₂NH	NH
F18	OH	F	--	O	F46	NH₂	D	NH	NH	F74	H	B	CH₂NH	NH
F19	H	A	--	NH	F47	NH₂	E	NH	NH	F75	H	C	CH₂NH	NH
F20	H	B	--	NH	F48	NH₂	F	NH	NH	F76	H	D	CH₂NH	NH
F21	H	C	--	NH	F49	OH	A	NH	NH	F77	H	E	CH₂NH	NH
F22	H	D	--	NH	F50	OH	B	NH	NH	F78	H	F	CH₂NH	NH
F23	H	E	--	NH	F51	OH	C	NH	NH	F79	NH₂	A	CH₂NH	NH
F24	H	F	--	NH	F52	OH	D	NH	NH	F80	NH₂	B	CH₂NH	NH
F25	NH₂	A	--	NH	F53	OH	E	NH	NH	F81	NH₂	C	CH₂NH	NH
F26	NH₂	B	--	NH	F54	OH	F	NH	NH	F82	NH₂	D	CH₂NH	NH
F27	NH₂	C	--	NH	F55	H	A	CH₂	NH	F83	NH₂	E	CH₂NH	NH
F28	NH₂	D	--	NH	F56	H	B	CH₂	NH	F84	NH₂	F	CH₂NH	NH

Molecular docking demonstrated that fragments F19-F24 with an amidine linker can form H-bonds with the template thymine and with the residue D185, whereas F26-F30 form H-bonds with thymine only. Fragments (F43, F45-46 and F52) with a guanine linker can only form H-bonds with template thymine, whereas F47 and F48 can also interact with D185. Interestingly, although fragments (F62-F64) with a longer guanidine linker only form H-bonds with thymine, their counterparts (F60, F66, F79-F84) can also form one additional interaction with the residue D186. For these fragments, the newly introduced heteroaromatics can also form a H-bond with the primer grip/primer 3'-nucleotide. Both F81 and F82 can form one H-bond with the 3'-hydroxyl end of the

primer strand and the backbone of residue L228, respectively, and the former can form an additional H-bond with G231, while maintaining interaction with the template thymine.

Considering the binding affinity, stability, structural variation and ease of synthesis of the designed fragments, we selected fragments F47 and F81 for synthesis. The detailed docking results and drug-like properties of hit 166 and selected fragments are shown in **Table 5-2**. Docking score of selected fragments were improved compared with that of two fragment screen hits. In addition, the docked modes of F47 and F81 suggested more interacting surface areas than that of two hits. Furthermore, selected fragments comply Lipinski's Rule of Five.[13]

Table 5-2. Docking results and drug-like properties of hit 166 and selected fragments via Autodock Vina[11]

Fragment No.	Chemical structure	Affinity (kcal/mol)	Number of H bond[a]	M.W.	cLogP[b]	Interacting surface area (Å²)[c]
166		-5.9	2	244.31	2.68	339.5
F1		-6.7	0	246.29	0.86	347.8
F2		-6.2	1	246.29	0.87	332.8
F3		-6.7	2	228.26	0.35	348.4
F4		-6.7	2	228.26	0.63	367.7
F5		-6.4	2	229.24	0.31	377.2
F47		-6.6	3	258.29	-0.12	378.9
F81		-7.6	4	271.33	0.78	438.8

[a] Calculated from LigPlot+.[14]

[b] Calculated from http://www.vcclab.org/lab/alogps/.

[c] Generated from https://www.ebi.ac.uk/pdbe/pisa/

The docking results and binding mode of selected fragments were analysed using PyMol.[15] As can be seen in **Figure 5-1**A, F47 not only forms two H-bonds with template DT706, but also interacts with residue D185. It can also form two additional H-bonds with primer DC821 and surrounding residue Y115. With an additional carbon atom in the guanidine linker, F81 becomes more flexible and adopts a conformation at the active site of HIV-1 RT that allows for more interactions, when compared with F47 (**Figure 5-1**B). Similar to F47, F81 can interact with template DT706, primer DC821 and residue Y115 as well. However, F81 interacts with catalytic residue D186 rather than with D185, that is involved in the binding of F47. In addition, F81 can form two H-bonds with surrounding residues L228 and G231, respectively.
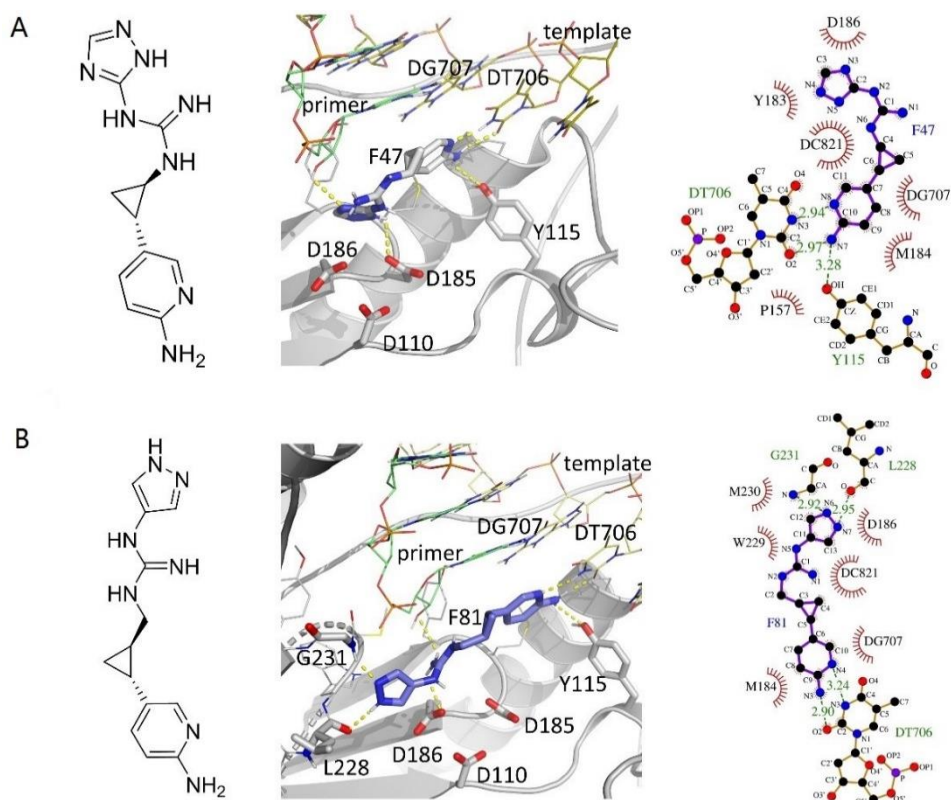
Figure 5-1. Interactions of (A) F47 and (B) F81 with HIV-1 RT/dsDNA. Conformation having best free energy of binding is shown in each case. H-bond interactions are depicted in yellow dashed lines; docking pose is generated in PyMol,[15] and schematic 2D representation is obtained with LigPlot+.[14]

5.2.2 Chemistry

Based on the abovementioned docking results and synthetic feasibility, we firstly proceeded to synthesize compounds **5.5a-e** for structural studies to examine if the pyridyl ring can form a H-bond with thymine. The synthesis of the *trans*-cyclopropane-containing carboxylic acid intermediate **5.4** which is required for the preparation of these compounds is depicted in **Scheme 5-1**. The *trans*-cyclopropane-containing compounds were all synthesized as racemic mixtures. Palladium-mediated coupling of the commercially available 3-bromopyridine **5.1** with ethyl acrylate afforded the ester **5.2** in good yield. The α,β-unsaturated ester **5.2** was transformed into *trans*-cyclopropane **5.3** via a Corey-Chaykovsky reaction.[16] Hydrolysis of the ester moiety of compound **5.3** under basic condition yielded the corresponding carboxylic acid **5.4**. Subsequent coupling with the appropriate amines using EDCI and HOBt afforded the final compounds **5.5a-e**.
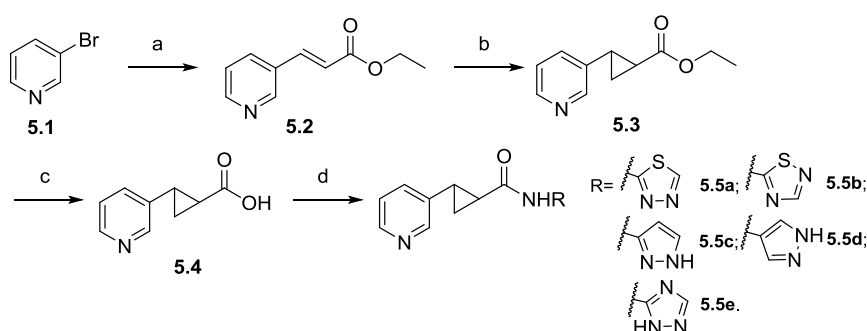
Scheme 5-1. *Reagents and conditions*: (a) ethyl acrylate, Pd(OAc)$_2$, K$_2$CO$_3$, PPh$_3$, DMF, 100 °C, 20 h, 75 %; (b) Me$_3$SOI, NaH, DMSO, r.t., 1 h, 26 %; (c) NaOH, MeOH, H$_2$O, 60 °C, 12 h, 49 %; (d) RNH$_2$, EDCI, DIEA, HOAt, DMF, -20 °C to r.t., overnight, 20 %-52 %.

Another two compounds with a guanidine moiety were selected for synthesis based on the abovementioned docking study. Although the synthetic route shown in **Scheme 5-1** gives a short access to the cyclopropyl derivatives, its main disadvantage is the low yield of the Corey-Chaykovsky reaction (20-30 %). Therefore, an alternative synthetic route was established, and which was based upon a known sequence of reactions,[17] that reported yields up to 82 %. The synthesis of target fragment F47 is illustrated in **Scheme 5-2.** Commercially available 5-bromopyridin-2-amine **5.6** was coupled with ethyl acrylate via a palladium-catalyzed Heck reaction yielding the ester intermediate **5.7** in good yield. Protection of the amino group of compound **5.7** as a PMB group generated intermediate **5.8**. Hydrolysis of the ester moiety under basic conditions gave the corresponding carboxylic acid. Amide coupling with *N,O*-dimethylhydroxylamine using EDCI afforded intermediate **5.9**. Subsequent transformation into the *trans*-cyclopropane derivative **5.10** was achieved via a Corey-Chaykovsky reaction with the sulfur ylide derived from trimethylsulfoxonium iodide.[16] Alkaline hydrolysis of the amide moiety in compound **5.10** under basic condition afforded the corresponding carboxylic acid **5.11**. Curtius rearrangement of acid **5.11** produced the amine **5.12**. Reaction of this amine with *N*-benzyloxycarbonyl isothiocyanate **5.13** yielded a thiourea intermediate. Coupling with 1,2,4-triazol-5-amine using EDCI, allowed the formation of the substituted guanidine **5.14**.[18] Treatment of compound **5.14** with TFA in order to remove the Cbz and PMB protecting groups did not yield the desired compound, based on mass spectral analysis. Alternatively, when a reductive hydrogenation using a Pd/C catalyst was applied, only starting material was recovered.
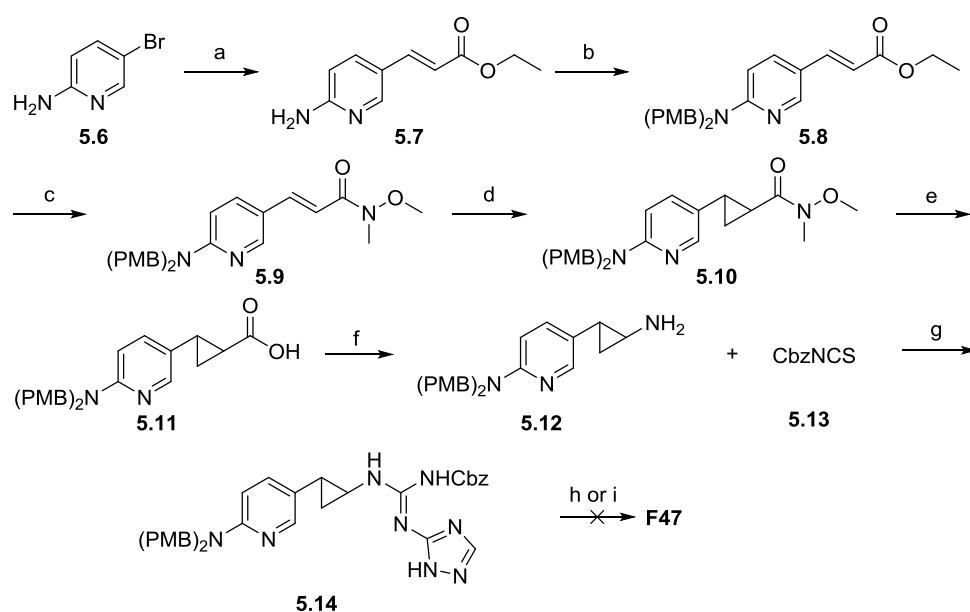
Scheme 5-2. Synthesis of F47. *Reagents and conditions*: (a) ethyl acrylate, Pd(OAc)$_2$, DIEA, P(*o*-tol)$_3$, DMF, 100 °C, 20 h, 74 %; (b) PMB-Cl, NaH, DMF, 0 °C, 1 h, 71 %; (c) (i) 2M NaOH in EtOH/H$_2$O, r.t., 24 h; (ii) *N,O*-dimethylhydroxylamine hydrochloride, EDCI, DMAP, DCM, r.t., 2 h, 40 % over two steps; (d) Me$_3$SOI, NaH, DMSO, 0 °C to r.t., 4 h, 82 %; (e) *t*-BuOK, EtOH/H$_2$O, r.t., 24 h, 62 %; (f) (i) DPPA, TEA, benzene, 80 °C, 1 h; (ii) H$_2$O, 80 °C, 30 min, 43 % over two steps; (g) (i) DCM, 0 °C to r.t., 4 h; (ii) 1,2,4-triazol-5-amine, EDCI, DIEA, DCM, 0 °C, 1 h, then r.t.,10 h, 32 % over two steps; (h) TFA, DCM, r.t., 5 h; (i) 10 % Pd/C, MeOH, r.t., 24 h.

In parallel, the synthesis of **F81** was also undertaken (**Scheme 5-3**). Reduction of amide **5.10** with LiAlH$_4$ yielded the primary alcohol **5.15**. Mesylation of the hydroxyl group, followed by reaction with sodium azide and a Staudinger reduction furnished amine **5.16**.[19] Amine **5.16** was converted to the 1,3-substituted guanidine **5.17** using the same procedure as for the synthesis of **5.14**. Unfortunately, upon treatment of compound **5.17** with TFA, the desired compound was not formed, as demonstrated by mass spectral analysis of the crude residue.
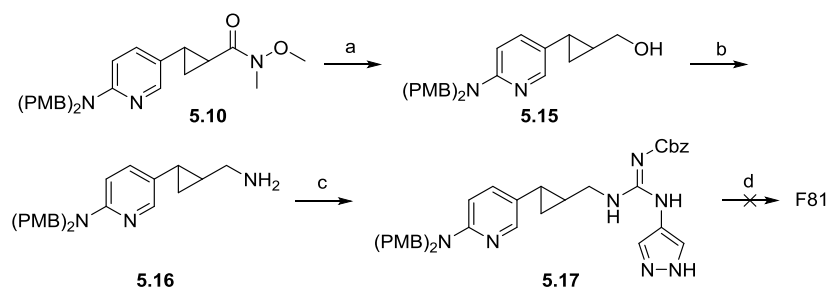
Scheme 5-3. Synthesis of F81. *Reagents and conditions*: (a) LiAlH$_4$, THF, 0 °C, 2 h, 72 %; (b) (i) MsCl, TEA, DCM, 0 °C, 2 h; (ii) NaN$_3$, DMF, 60 °C, 6 h; (iii) PPh$_3$, THF, H$_2$O, r.t., 12 h, 25 % over three steps; (c) (i) **5.13**, DCM, 0 °C to r.t., 4 h; (ii) pyrazol-4-amine, EDCI, DIEA, DCM, 0 °C, 1 h, then r.t., 10 h, 63 % over two steps; (d) TFA, DCM, r.t., 5 h.

5.2.3 Biological evaluations

In vitro **RT inhibition assay** (performed by Dr. Hoai Nguyen)

The inhibitory effect of compounds **5.5a-e** on the DNA synthesis catalyzed by HIV-1 RT was evaluated by a gel-based RT inhibition assay (**Figure 5-2**). The primer-template duplex has seven dT nucleotide overhang at the 5' end of the template, so this template/primer allows seven primer extensions at most. The anti-HIV drug EFV was used as the positive control, whereas Milili-Q water was included as the negative control. dATP was used at a concentration of 5 μM with 3 mM MgCl$_2$ as the metal ion source. After 10 min of incubation for polymerization reaction in the presence of individual compounds, ETV showed pronounced RT inhibition when tested at the concentration of 5 or 50 μM. Among the tested compounds, only **5.5c** and **5.5e** showed weak RT inhibition at the concentration of 10 mM and 5 mM.
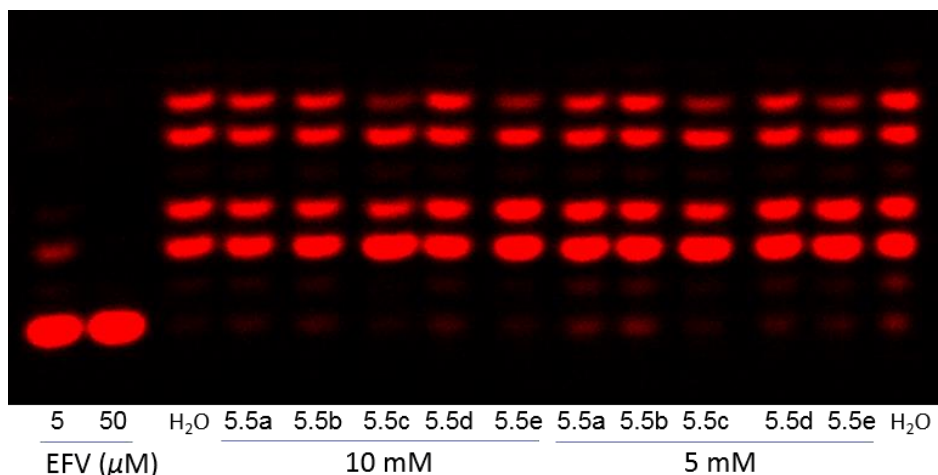
Figure 5-2. HIV-1 RT inhibition assay. Reaction conditions: 125 nM primer/template; [dATP] = 5 μM; [HIV-1 RT] = 0.06 μg/μL, 3 mM MgCl$_2$.

HIV antiviral assay (performed by the group of Prof. Dominique Schols)

Compounds **5.5a-e** were tested *in vitro* for their antiviral activity against HIV in MT-4 cells infected with HIV-1 NL4.3 and HIV-2 ROD. Their cytotoxic effect on the MT-4 cell line was also evaluated. TFV was used as a positive control. Possibly due to the weak binding, none of these compounds showed antiviral activity or cytotoxicity even at the highest concentration tested (100 μM).

5.3 Conclusion

A series of new compounds were designed based on hit 166. Molecular docking studies demonstrated that the introduction of a 2-aminopyridine ring and a guanidine moiety could improve the interaction between the ligands and HIV-1 RT/dsDNA. Five initial ligands were synthesized, and the synthesis of another two ligands has been attempted. In addition, the synthesized compounds were evaluated for their RT inhibition, and screened for their antiviral activity against HIV. Unfortunately, no significant activity or cytotoxicity was observed.

5.4 Experimental section

Molecular docking study. To obtain possible conformations and orientations for the ligands at the binding site, docking studies were performed using Autodock Vina (ver. 1.1.1).[11] We utilized the crystal structure of HIV-1 RT/dsDNA complexed with hit 166 , which was obtained from previous fragments screening in our lab. Then, the co-crystallized ligand and water molecules were removed, only chain C and dsDNA (chain E, F) were kept, and the remaining protein coordinate was then saved and converted to pdbqt format using Autodock Tools (1.5.6).[20] The chemical structures of designed ligands were drawn using ChemBioDraw Ultra 14.0. All the ligands were energy minimized by MM2 method with Chem3D Pro 14.0 and were then saved as pdb fromat which was then converted to pdbqt format using Autodock Tools. The docking parameters were set as follow: size_x = 22; size_y = 22;

119

size_z = 18; center_x = -129.599; center_y = -2.922; center_z = 13. 379; exhaustiveness = 100. The other parameters were left as default. Finally, the conformations with the most favorable free energy of binding were selected. The interactions of complex protein-ligand conformations, including hydrogen bonds and bond lengths were analyzed using PyMol.[15]

Chemistry. All reagents and solvents were purchased from commercial sources and were used as obtained. Moisture-sensitive reactions were carried out in oven-dried glassware under a nitrogen or argon atmosphere. [1]H and ^{13}C spectra were recorded on Bruker Avance 300 or 500 MHz spectrometer using tetramethylsilane as internal standard or referenced to the residual solvent signal. Splitting patterns are designated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), br(broad), and dd (doublet of doublets). Coupling constants are expressed in hertz (Hz). High-resolution mass spectra (HRMS) were obtained on a quadruple orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 μL/min and spectra were obtained in positive (or negative) ionization mode with a resolution of 15000 (FWHM) using leucine enkephalin as lock mass. Pre-coated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ= 254 nm). Products were purified by flash chromatography on silica gel (35-70 μm, 60 Å).

Ethyl (E)-3-(pyridin-3-yl)acrylate (5.2) To a mixure of 3-bromopyridine **5.1** (1.336 mL, 8.36 mmol), K_2CO_3 (2.31 g, 16.72 mmol), PPh_3 (44 mg, 0.168 mmol), $Pd(OAc)_2$ (18.8 mg, 0.084 mmol) in anhydrous DMF (20 mL), was added ethyl acrylate (806 μL, 12.54 mmol). The mixture was heated and stirred at 120 °C for 20 h under an atmosphere of argon. The progress of the reaction was monitored by TLC. After completion of the reaction, the reaction mixture was cooled to room temperature and filtered. The filtrate was extracted with ethyl acetate and water. The organic layer was washed with water and brine, dried with sodium sulphate and evaporated under reduced pressure. The residue was finally purified by silica gel flash chromatography (gradient hexane/EA, v/v, 10:1 to 6:1) to obtain the title compound as a yellow oil (1.388 g, 75 %). ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.75 (d, J = 2.0 Hz, 1H), 8.61 (dd, J = 4.8, 1.4 Hz, 1H), 7.85 (dt, J = 7.9, 1.8 Hz, 1H), 7.68 (d, J = 16.1 Hz, 1H), 7.34 (dd, J = 7.9, 4.8 Hz, 1H), 6.52 (d, J = 16.1 Hz, 1H), 4.29 (q, J = 7.1 Hz, 2H), 1.35 (t, J = 7.1 Hz, 3H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 165.72 (s), 150.52 (s), 149.33 (s), 140.39 (s), 133.73 (s), 129.79 (s), 123.30 (s), 120.07 (s), 60.25 (s), 13.87 (s). HRMS: [M+H]$^+$ calculated for $C_{10}H_{11}NO_2$, 178.0862; found 178.0854.

(trans)-Ethyl-2-(pyridin-3-yl)cyclopropane-1-carboxylate (5.3) Trimethyl sulfoxonium iodide (1.863 g, 8.47 mmol) was added portionwise to a suspension of 60 % NaH (338.6 mg, 8.47 mmol) in dry DMSO (20 mL) at room temperature over a period of 20 min. and stirred for 1 h till the formation of a clear solution. A solution of (*E*)-ethyl 3-(pyridin-3-yl) acrylate **5.2** (1 g, 5.64 mmol) in dry DMSO

(5 mL) was added dropwise and stirred at room temperature for 20 min. After completion, the reaction mixture was poured into ice water (200 mL) and extracted with EA (3×50 mL). The combined organic layers were washed with ice water (150 mL), brine (150 mL), dried over anhydrous Na$_2$SO$_4$, filtered and evaporated. The crude residue was purfied by silica gel chromatography (gradient hexane/EA, v/v, 6:1 to 3:1) to yield the title compound as a yellowish oil (278 mg, 26 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.46 (dd, J = 3.0, 1.8 Hz, 2H), 7.40 – 7.32 (m, 1H), 7.21 (dd, J = 7.9, 4.8 Hz, 1H), 4.19 (q, J = 7.1 Hz, 2H), 2.52 (ddd, J = 9.4, 6.4, 4.2 Hz, 1H), 1.93 (ddd, J = 8.5, 5.3, 4.3 Hz, 1H), 1.65 (dt, J = 9.3, 5.0 Hz, 1H), 1.38 – 1.32 (m, 1H), 1.29 (t, J = 7.2 Hz, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 172.97 (s), 148.63 (s), 147.98 (s), 135.73 (s), 133.25 (s), 123.37 (s), 61.02 (s), 23.91 (s), 23.55 (s), 16.73 (s), 14.32 (s). HRMS: [M+H]$^+$ calculated for C$_{11}$H$_{13}$NO$_2$, 192.1019; found 192.1011.

(trans)-2-(Pyridin-3-yl)cyclopropane-1-carboxylic acid (5.4) A solution of NaOH (640 mg, 16 mmol) in water (8 mL) was added to a solution of the ester intermediate **5.3** (244 mg, 1.28 mmol) in MeOH (30 mL) at room temperature. The reaction mixture was stirred at 60 °C overnight, then the pH of the solution was adjusted to 6 with 12 M HCl, and the mixture was concentrated under vacuum to dryness. The resulting solid was triturated with MeOH (20 mL), and the resulting solids were removed by filtration. The filtrate was concentrated under vacuum to give the crude product as a light-yellow solid. This material was recrystallized from a 5:1 mixture of EA/MeOH (6 mL), affording the title compound as a white solid (102 mg, 49 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 10.42 (s, 1H), 8.47 (d, J = 7.2 Hz, 2H), 7.46 (d, J = 7.6 Hz, 1H), 7.29 (dd, J = 7.3, 5.3 Hz, 1H), 2.58 (dd, J = 10.6, 7.3 Hz, 1H), 1.83 (ddd, J = 13.4, 8.4, 4.5 Hz, 2H), 1.35 (dd, J = 11.4, 6.9 Hz, 1H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 176.55 (s), 146.85 (s), 146.08 (s), 136.98 (s), 134.98 (s), 124.01 (s), 24.33 (s), 23.56 (s), 16.87 (s). HRMS: [M+H]$^+$ calculated for C$_9$H$_{10}$NO$_2$, 164.0711; found 164.0698.

(trans)-2-(Pyridin-3-yl)-N-(1,2,4-thiadiazol-5-yl)cyclopropane-1-carboxamide (5.5a) To a solution of the acid **5.4** (52 mg, 0.319 mmol) and EDCI (73.4 mg, 0.383 mmol) in dry DMF (2 mL), DIEA (63.3 μL, 0.383 mmol) was added dropwise under argon atmosphere, maintaining the temperature at -20 °C. After 20 min, a solution of HOAt (53 mg, 0.383 mmol) in dry DMF (2 mL) was added, and the resulting light-yellow solution was allowed to stir for an additional 20 min. Meanwhile, a suspension of 5-amino-1,2,4-thiadiazole (35.6 mg, 0.341 mmol) in dry DMF (2 mL) was prepared and added dropwise. The solution was then stirred overnight under argon and monitored by TLC. After consumption of the starting material, DMF was evaporated under reduced pressure and the crude residue was purified by silica gel choromatography (gradient DCM/MeOH, v/v, 50:1 to 40:1 to 30:1) to yield the target compound as a colourless solid (16 mg, 20 %). ^{1}H NMR (300 MHz, MeOD) δ 8.55 – 8.40 (m, 2H), 8.37 (s, 1H), 7.67 (d, J = 7.7 Hz, 1H), 7.44 (dd, J = 7.5, 5.1 Hz, 1H), 2.85 – 2.70 (m, 1H), 2.43 – 2.29 (m, 1H), 1.92 (dt, J = 9.2, 4.7 Hz, 1H), 1.66 (ddd, J = 8.1, 6.6, 5.2

Hz, 1H). ^{13}C NMR (75 MHz, DMSO) δ 174.79 (s), 171.74 (s), 158.60 (s), 148.33 (s), 147.75 (s), 135.48 (s), 133.24 (s), 123.50 (s), 24.97 (s), 24.30 (s), 16.49 (s). HRMS: [M+H]$^+$ calculated for C$_{11}$H$_{10}$N$_4$OS, 147.0648; found 147.0642.

(trans)-2-(Pyridin-3-yl)-N-(1,3,4-thiadiazol-2-yl)cyclopropane-1-carboxamide (5.5b) This compound was obtained as a white solid **5.4** (26 mg, 45 %) according to the procedure used for the synthesis of **5.5a**, starting from the acid (38 mg, 0.233 mmol) and 2-amino-1,3,4-thiadiazole (26 mg, 0.25 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.97 (s, 1H), 8.45 (s, 1H), 8.41 (d, J = 4.8 Hz, 1H), 7.60 (d, J = 7.8 Hz, 1H), 7.37 (dd, J = 7.5, 5.1 Hz, 1H), 2.78 – 2.58 (m, 1H), 2.43 – 2.23 (m, 1H), 1.84 (dt, J = 9.6, 5.0 Hz, 1H), 1.61 – 1.45 (m, 1H). ^{13}C NMR (75 MHz, MeOD) δ 171.49 (s), 160.04 (s), 149.18 (s), 148.26 (s), 147.74 (s), 136.94 (s), 135.08 (s), 124.60 (s), 25.44 (s), 24.83 (s), 17.49 (s). HRMS: [M+H]$^+$ calculated for C$_{11}$H$_{10}$N$_4$OS, 147.0648; found 147.0644.

(trans)-N-(1H-Pyrazol-3-yl)-2-(pyridin-3-yl)cyclopropane-1-carboxamide (5.5c) This compound was obtained as a white solid (31 mg, 43 %) according to the procedure used for the synthesis of **5.5a**, starting from the acid **5.4** (52 mg, 0.319 mmol) and 3-aminopyrazole (28.9 mg, 0.341 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.42 (d, J = 1.9 Hz, 1H), 8.34 (dd, J = 4.9, 1.5 Hz, 1H), 8.01 (d, J = 2.9 Hz, 1H), 7.64 – 7.56 (m, 1H), 7.32 (dd, J = 7.5, 4.9 Hz, 1H), 5.97 (d, J = 3.0 Hz, 1H), 3.19 (dd, J = 4.8, 2.5 Hz, 1H), 2.70 – 2.55 (m, 1H), 1.80 – 1.70 (m, 1H), 1.56 (ddd, J = 8.4, 6.6, 4.5 Hz, 1H). ^{13}C NMR (75 MHz, MeOD) δ 170.68 (s), 160.85 (s), 148.90 (s), 148.17 (s), 138.11 (s), 135.61 (s), 130.78 (s), 125.19 (s), 102.92 (s), 26.13 (s), 23.98 (s), 18.13 (s). HRMS: [M+H]$^+$ calculated for C$_{12}$H$_{12}$N$_4$O, 229.1084; found 229.1074.

(trans)-N-(1H-Pyrazol-4-yl)-2-(pyridin-3-yl)cyclopropane-1-carboxamide (5.5d) This compound was obtained as a white solid (38 mg, 52 %) according to the procedure used for the synthesis of **5.5a**, starting from the acid (52 mg, 0.319 mmol) and 4-aminopyrazole (28.9 mg, 0.341 mmol). ^{1}H NMR (300 MHz, MeOD) δ 8.45 – 8.34 (m, 2H), 7.76 (d, J = 12.0 Hz, 2H), 7.57 (d, J = 7.8 Hz, 1H), 7.36 (dd, J = 7.6, 4.8 Hz, 1H), 2.62 – 2.46 (m, 1H), 2.06 (ddd, J = 6.2, 5.1, 4.1 Hz, 1H), 1.79 – 1.62 (m, 1H), 1.45 – 1.31 (m, 1H). ^{13}C NMR (75 MHz, MeOD) δ 169.31 (s), 146.99 (s), 146.23 (s), 136.76 (s), 133.87 (s), 124.66 (s), 123.43 (s), 120.74 (s), 24.94 (s), 21.89 (s), 14.96 (s). HRMS: [M+H]$^+$ calculated for C$_{12}$H$_{12}$N$_4$O, 229.1084; found 229.1076.

(trans)-2-(Pyridin-3-yl)-N-(1H-1,2,4-triazol-5-yl)cyclopropane-1-carboxamide (5.5e) This compound was obtained as a white solid (17 mg, 23 %) according to the procedure used for the synthesis of **5.5a**, starting from the acid **5.4** (52 mg, 0.319 mmol) and 4-amino-1,2,4-triazole (29.3 mg, 0.341 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.53 (d, J = 1.8 Hz, 1H), 8.50 (dd, J = 4.7, 1.3 Hz, 1H), 7.49 (s, 1H), 7.47 – 7.40 (m, 1H), 7.30 – 7.15 (m, 1H), 6.99 (s, 2H), 3.28 (ddd, J = 8.5, 5.2, 4.3 Hz, 1H), 2.77 (ddd, J = 9.4, 6.8, 4.2 Hz, 1H), 1.91 (dt, J = 9.5, 4.9 Hz, 1H), 1.63 (ddd, J = 8.4, 6.8,

4.6 Hz, 1H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 173.27 (s), 156.91 (s), 150.96 (s), 148.92 (s), 148.43 (s), 134.86 (s), 133.49 (s), 123.53 (s), 26.60 (s), 23.65 (s), 18.79 (s). HRMS: [M+H]$^+$ calculated for C$_{11}$H$_{11}$N$_5$O, 230.1036; found 230.1030.

Ethyl (E)-3-(6-aminopyridin-3-yl)acrylate (5.7) To a mixure of 5-bromopyridin-2-amine **5.6** (1 g, 5.78 mmol), P(*o*-tol)$_3$ (176 mg, 0.58 mmol), Pd(OAc)$_2$ (64 mg, 0.29 mmol) and DIEA (1.147 mL, 6.94 mmol) in anhydrous DMF (20 mL), was added ethyl acrylate (677 μL, 6.36 mmol). The mixture was heated and stirred at 110 °C for 20 h under an atmosphere of argon. The progress of the reaction was monitored by TLC. After completion of the reaction, the solvent was evaporated under reduced pressure. The residue was purified by silica gel flash chromatography (gradient hexane/EA, v/v, 2:1 to 1:1) to obtain the title compound as a yellow solid (818 mg, 74 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.17 (d, J = 2.0 Hz, 1H), 7.62 (dd, J = 8.6, 2.2 Hz, 1H), 7.56 (d, J = 16.0 Hz, 1H), 6.49 (d, J = 8.6 Hz, 1H), 6.24 (d, J = 16.0 Hz, 1H), 4.87 (s, 2H), 4.23 (dd, J = 14.3, 7.1 Hz, 2H), 1.31 (t, J = 7.1 Hz, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 167.51 (s), 159.81 (s), 150.33 (s), 141.92 (s), 136.06 (s), 121.20 (s), 115.23 (s), 108.98 (s), 60.68 (s), 14.66 (s). HRMS: [M+H]$^+$ calculated for C$_{10}$H$_{12}$N$_2$O$_2$, 193.0971; found 193.0971.

Ethyl (E)-3-(6-(bis(4-methoxybenzyl)amino)pyridin-3-yl)acrylate (5.8) Sodium hydride (2.5 g of a 60 % dispersion in mineral oil, 62.4 mmol) was added in small portions to an ice-cooled solution of compound **5.7** (4.0 g, 20.8 mmol) in DMF (70 mL) maintained under nitrogen. The resulting suspension was stirred for 10 min at 0 °C, and then 1-(chloromethyl)-4-methoxybenzene (6.178 mL, 45.76 mmol) was added. The reaction mixture was stirred for 1 h at 0 °C and was then quenched by the careful addition of water (5 mL). The resulting mixture was concentrated under vacuum, and the residue was partitioned between EA (3×100 mL) and water (150 mL). The combined organic layers were washed with brine, dried over Mg$_2$SO$_4$, filtered, and concentrated under vacuum. The crude residue was purified by silica gel flash chromatography (gradient hexane/EA, v/v, 20:1 to 10:1 to 6:1) to furnish the title compound as a yellow solid (6.344 g, 71 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.31 (d, J = 2.2 Hz, 1H), 7.64 − 7.54 (m, 2H), 7.14 (d, J = 8.5 Hz, 4H), 6.84 (d, J = 8.7 Hz, 4H), 6.48 (d, J = 9.0 Hz, 1H), 6.20 (d, J = 15.9 Hz, 1H), 4.73 (s, 4H), 4.24 (q, J = 7.1 Hz, 2H), 3.78 (s, 6H), 1.32 (t, J = 7.1 Hz, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 167.55 (s), 159.47 (s), 158.91 (s), 150.29 (s), 142.06 (s), 135.45 (s), 129.61 (s), 128.39 (s), 119.23 (s), 114.15 (s), 113.66 (s), 106.28 (s), 77.58 (s), 77.16 (s), 76.74 (s), 60.33 (s), 55.35 (s), 50.31 (s), 14.47 (s). HRMS: [M+H]$^+$ calculated for C$_{26}$H$_{28}$N$_2$O$_4$, 433.2122; found 433.2117.

(E)-3-(6-(Bis(4-Methoxybenzyl)amino)pyridin-3-yl)-N-methoxy-N-methylacrylamide (5.9) A solution of KOH (2.47 g, 44 mmol) in water (15 mL) was added to a solution of intermediate **5.8** (6.345 g, 14.67 mmol) in EtOH (15 mL) at 25 °C. The reaction mixture was stirred at this temperature

for 24 h. The pH of the solution was then adjusted to 6.0 with 12 M HCl. The resulting precipitate was filtered off and air-dried to give the crude carboxylic acid as an off-white solid, which was used for next step without purification. EDCI (4.217 g, 22 mmol) and 4-(dimethylamino)pyridine (2.688 g, 22 mmol) were added sequentially to a solution of the above carboxylic acid and *N,O*-dimethylhydroxylamine hydrochloride (1.342 g, 22 mmol) in DCM (120 mL) at 25 °C. The mixture was stirred at this temperature for 2 h and was then partitioned between water (200 mL) and DCM (2 × 100 mL). The combined organic layers were washed with water (3 × 100 mL), dried over Na_2SO_4, filtered, and concentrated under vacuum. The residue was purified by silica gel flash chromatography (gradient hexane/EA, v/v, 5:1 to 3:1 to 2:1) to furnish the title compound as a yellow oil (2.595 g, 40 % over two steps). ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.37 (d, J = 2.1 Hz, 1H), 7.70 – 7.60 (m, 2H), 7.14 (d, J = 8.5 Hz, 4H), 6.83 (dd, J = 12.1, 8.0 Hz, 5H), 6.48 (d, J = 9.0 Hz, 1H), 4.73 (s, 4H), 3.78 (s, 6H), 3.73 (s, 3H), 3.29 (s, 3H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 167.61 (s), 159.25 (s), 158.89 (s), 150.07 (s), 140.91 (s), 135.66 (s), 129.76 (s), 128.41 (s), 119.92 (s), 114.14 (s), 111.39 (s), 106.14 (s), 61.84 (s), 55.36 (s), 50.37 (s), 32.63 (s). HRMS: [M+H]$^+$ calculated for $C_{26}H_{29}N_3O_4$, 448.2231; found 448.2231.

(trans)-2-(6-(Bis(4-Methoxybenzyl)amino)pyridin-3-yl)-N-methoxy-N-methylcyclopropane-1-carboxamide (5.10) Sodium hydride (0.2 g of a 60 % dispersion in mineral oil, 5 mmol) was added to an ice-cooled solution of trimethylsulfoxonium iodide (1.1 g, 5 mol) in DMSO (15 mL). The resulting solution was warmed to room temperature and was stirred for 1 h. Compound **5.9** (1.118 g, 2.5 mmol) was added, and the reaction mixture was stirred at room temperature for 4 h. The mixturewas partitioned between a saturated aqueous NH_4Cl solution (100 mL) and EtOAc (3 × 100 mL). The combined organic layers were washed with water (3 × 50 mL), dried over Na_2SO_4, filtered, and concentrated under vacuum. The residue was purified by silica gel chromatography (gradient hexane/EA, v/v, 3:1 to 1:1) to give the title compound as a light-yellow oil (944 mg, 82 %). ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.06 (d, J = 2.1 Hz, 1H), 7.14 (d, J = 8.6 Hz, 5H), 6.83 (d, J = 8.6 Hz, 4H), 6.40 (d, J = 8.7 Hz, 1H), 4.68 (s, 4H), 3.78 (s, 6H), 3.71 (s, 3H), 3.23 (s, 3H), 2.46 – 2.34 (m, 1H), 1.62 – 1.51 (m, 1H), 1.21 (ddd, J = 8.2, 6.4, 3.6 Hz, 2H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 173.60 (s),158.95 (s), 157.86 (s), 146.76 (s), 136.01 (s), 130.72 (s), 128.60 (s), 124.08 (s), 114.25 (s), 106.03 (s), 62.06 (s), 60.70 (s), 55.56 (s), 50.56 (s), 23.38 (s), 20.69 (s), 15.72 (s). HRMS: [M+H]$^+$ calculated for $C_{27}H_{31}N_3O_4$, 462.2387; found 462.2387.

(trans)-2-(6-(Bis(4-Methoxybenzyl)amino)pyridin-3-yl)cyclopropane-1-carboxylic acid (5.11) A suspension of compound **5.10** (600 mg, 1.3 mmol) and potassium *tert*-butoxide (786 mg, 7 mmol) in ether (9 mL) and water (46 μL) was stirred at room temperature for 16 h. The mixture was concentrated *in vacuo*, diluted with water (10 mL) and acidified by the slow addition of concentrated

HCl till pH~6, and the aqueous mixture was extracted with DCM (3×15 mL). The combined organic layers were washed with brine (20 mL), dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford the title compound as a light-yellow oil (333 mg, 0.8 mmol, 62 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 11.11 (s, 1H), 8.05 (d, J = 1.8 Hz, 1H), 7.11 (d, J = 8.5 Hz, 4H), 7.06 (d, J = 2.2 Hz, 1H), 6.82 (d, J = 8.5 Hz, 4H), 6.40 (d, J = 8.8 Hz, 1H), 4.66 (s, 4H), 3.75 (s, 6H), 2.58 – 2.31 (m, 1H), 1.75 (dt, J = 8.6, 4.5 Hz, 1H), 1.62 – 1.47 (m, 1H), 1.33 – 1.20 (m, 1H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 178.95 (s), 158.70 (s), 157.47 (s), 146.16 (s), 135.93 (s), 130.09 (s), 128.29 (s), 123.00 (s), 114.02 (s), 106.27 (s), 55.27 (s), 50.52 (s), 24.08 (s), 23.15 (s), 16.26 (s). HRMS: [M+H]$^+$ calculated for $C_{25}H_{27}N_2O_4$, 419.1971; found 419.1960.

5-((trans)-2-Aminocyclopropyl)-N,N-bis(4-methoxybenzyl)pyridin-2-amine (5.12) To an ice-cooled solution of the carboxylic acid **5.11** (750 mg, 1.79 mmol) and DIEA (343 μL, 1.97 mmol, 1.1 equiv) in anhydrous THF (10 mL), was added dropwise diphenyl phosphoryl azide (424 μL, 1.97 mmol, 1.1 equiv) under a nitrogen atmosphere. The reaction mixture was stirred for 5 h. The solvents were evaporated, and the residue was dissolved in toluene (6 mL) and stirred under reflux for 1 h. The reaction mixture was cooled down to 0 °C and a 1 M HCl solution (3 mL) was added. The mixture was heated to 50 °C and stirred for 10 min. The obtained solution was poured into a 1 M NaOH (10 mL) and stirred for 1 h at room temperature. EA (15 mL) was added and the aqueous phase was washed thoroughly with EA (3×15 mL). The combined organic phase was dried over anhydrous MgSO$_4$ and concentrated under reduced pressure. The resulting residue was purified by silica gel chromatography (gradient DCM/MeOH/TEA, v/v/v, 100:0:0 to 100:1:0 to 50:1:0.1) to afford the title compound as a light-yellow oil (298 mg, 43 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 7.98 (d, J = 2.3 Hz, 1H), 7.12 (d, J = 8.5 Hz, 4H), 7.01 (dd, J = 8.7, 2.5 Hz, 1H), 6.82 (d, J = 8.6 Hz, 4H), 6.36 (d, J = 8.7 Hz, 1H), 4.66 (s, 4H), 3.76 (s, 6H), 2.43 (dt, J = 7.2, 3.7 Hz, 1H), 1.74 (ddd, J = 9.1, 5.8, 3.1 Hz, 1H), 0.98 – 0.77 (m, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 158.62 (s), 157.17 (s), 145.93 (s), 135.21 (s), 130.62 (s), 128.34 (s), 125.31 (s), 113.95 (s), 105.68 (s), 55.29 (s), 50.27 (s), 34.00 (s), 23.04 (s), 16.96 (s). HRMS: [M+H]$^+$ calculated for $C_{24}H_{28}N_3O_2$, 390.2181; found 390.2169.

(Benzyloxycarbonyl) isothiocyanate (5.13) A solution of potassium isothiocyanate (0.5 g, 5.14 mmol) in anhydrous tetrachloroethene (11 mL) was stirred vigorously after which 18-crown-6 (0.06 g, 0.21 mmol) and CbzCl (0.55 mL, 3.9 mmol) were added. After stirring at room temperature for 15 min, the mixture was allowed to reflux at 85 °C overnight. TLC analysis (DCM: hexane, 1:4) showed complete consumption of the CbzCl (Rf= 0.45) and the formation of a more polar major product (Rf = 0.3) along with a faint side product (Rf = 0.1). The reaction was then diluted with hexane (20 mL), cooled on ice and then filtered through a pad of Celite$^®$ to remove salts, which were rinsed with a 1:1 solution of DCM: hexane. The solvent was removed under vacuum yielding the crude product. [1]H-

NMR analysis showed the formation of the desired product, with spectral data that were consistent with literature values. Impurities were also present. Since literature stated that these impurities do not interfere with subsequent use and are easily removed in later chromatographic purifications, this product was carried on. The crude Cbz-NCS was dissolved in dry DCM to make a 0.5 M solution. This was flushed with nitrogen, sealed and kept in the fridge at 4 °C to be used as a stock solution.

1-((trans)-2-(6-(Di(4-methoxybenzyl)amino)pyridin-3-yl)cyclopropyl)-2-benzyl carbamate-3-(1H-1,2,4-triazol-5-yl)guanidine (**5.14**) A solution of Cbz-NCS **5.13** (1.2 mL, 0.5 M in DCM) in DCM (25 mL) was cooled to 0 °C before adding alkylamine **5.12** (227 mg, 0.583 mmol). The ice bath was removed, and the solution was stirred for 4 h under nitrogen. The solution was washed with 1 % HCl, water, and brine and dried over Na_2SO_4. The solvent was removed under reduced pressure, yielding the crude carbamoyl thiourea as a white solid. This crude carbamoyl thiourea (0.583 mmol), 1,2,4-triazol-5-amine (74 mg, 0.875 mmol), and diisopropylethylamine (102 μL, 0.583 mmol) were added to anhydrous dichloromethane (25 mL) and cooled to 0 °C. EDCI (168 mg, 0.875 mmol) was added, and the solution was stirred under nitrogen. After 1 h the ice bath was removed, and the solution was stirred for an additional 10 h at room temperature. In cases where TLC indicated unreacted starting material, addition of more amine and EDCI resulted in increased yields. The reaction mixture was washed with 1 % HCl, water, and brine and dried with Na_2SO_4. The solvent was removed under reduced pressure, and the residue was purified by silica gel chromatography (gradient hexane/EA, v/v, 2:1 to 1:1 to 1:2 to 1:3) to obtain the title compound as a light-yellow solid (117 mg, 32 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.01 (d, J = 2.0 Hz, 1H), 7.40 (s, 1H), 7.29 (dd, J = 10.7, 7.2 Hz, 6H), 7.09 (d, J = 8.5 Hz, 4H), 6.76 (d, J = 8.5 Hz, 4H), 6.37 (d, J = 8.8 Hz, 1H), 4.92 (d, J = 12.2 Hz, 1H), 4.66 (d, J = 7.8 Hz, 4H), 4.51 (s, 1H), 3.74 (s, 6H), 3.58 – 3.41 (m, 1H), 3.05 (s, 1H), 2.07 – 1.98 (m, 1H), 1.40-1.20 (m, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 158.68 (s), 157.57 (s), 149.45 (s), 146.36 (s), 136.10 (s), 134.77 (s), 130.41 (s), 128.76 (s), 128.55 (s), 128.32 (s), 128.28 (s), 122.43 (s), 114.01 (s), 105.79 (s), 67.85 (s), 55.31 (s), 50.32 (s), 23.09 (s), 17.72 (s). HRMS: [M+H]$^+$ calculated for $C_{35}H_{37}N_8O_4$, 633.2938; found 633.2933.

Deprotection of intermediate 5.14 The protected intermediate **5.14** (117 mg, 0.185 mmol) was treated with TFA (5 mL) and anisole (2 mL) at room temperature for 5 h. To remove nonpolar impurities, the residue was partitioned between ether (15 mL) and H_2O (15 mL) and after separation, the ether layer further extracted with an additional H_2O (5 mL). The aqueous layers were combined and concentrated *in vacuo*. Analysis of this crude residue by HRMS did not show the formation of the desired compound.

((trans)-2-(6-(Bis(4-Methoxybenzyl)amino)pyridin-3-yl)cyclopropyl)methanol (5.15) To an ice-cooled solution of **5.10** (2.36 g, 5.113 mmol) in THF (40 mL) was added LiAlH$_4$ (14.6 mL, 1M

in THF, 14.572 mmol). The mixture was stirred for 2 h, after which time the reaction mixture was quenched by the sequential addition of water (0.5 mL), 15 % NaOH (0.5 mL), and water (1.5 mL). After stirring for an additional 2 h, the suspension was filtered, and the filtrate was washed with additional THF (50 mL). The combined organic layers were concentrated *in vacuo*, and the resulting residue was purified by silica gel chromatography (gradient hexane/EA, v/v, 4:1 to 2:1) to afford the title compound as a light-yellow oil (1.492 g, 72 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.01 (d, J = 2.3 Hz, 1H), 7.12 (d, J = 8.7 Hz, 4H), 7.04 (dd, J = 8.7, 2.4 Hz, 1H), 6.82 (d, J = 8.7 Hz, 4H), 6.37 (d, J = 8.7 Hz, 1H), 4.66 (s, 4H), 3.76 (s, 6H), 3.55 (d, J = 6.8 Hz, 2H), 2.12 (s, 1H), 1.73 − 1.62 (m, 1H), 1.31 (dq, J = 11.2, 6.7 Hz, 1H), 0.81 (dd, J = 7.3, 6.1 Hz, 2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 158.64 (s), 157.26 (s), 146.18 (s), 135.30 (s), 130.57 (s), 128.34 (s), 125.48 (s), 113.99 (s), 105.87 (s), 66.51 (s), 55.32 (s), 50.33 (s), 23.93 (s), 18.12 (s), 12.47 (s). HRMS: [M+H]$^+$ calculated for C$_{25}$H$_{28}$N$_2$O$_3$, 405.2173; found 405.2166.

5-((trans)-2-(Aminomethyl)cyclopropyl)-N,N-bis(4-methoxybenzyl)pyridin-2-amine (5.16) To an ice-cooled solution of **5.15** (1.2 g, 2.97 mmol, 1.0 eq.), DIEA (1.033 mL, 5.93 mmol, 2.0 eq.) in DCM (30 mL) was added methane sulfonyl chloride (276 μL, 3.56 mmol, 1.2 eq.). The resulting mixture was stirred at this temperature for 2 h. After TLC revealed the complete consumption of the starting material, the reaction was quenched with water (20 mL) and extracted with DCM (3 × 20 mL). The combined organic layers were dried over MgSO$_4$, filtered, and concentrated under reduced pressure. The residue was dissolved in DMF (30 mL) and NaN$_3$ (386 mg, 5.93 mmol, 2.0 eq.) was added at room temperature. The reaction mixture was stirred at 60 °C for 6 h, followed by cooling to 25 °C. The mixture was treated with water (40 mL) and extracted with ethyl acetate (3 × 50 mL). The combined organic phases were dried over MgSO$_4$, filtered, and concentrated under reduced pressure to give the crude azide. The azide was dissolved in THF (20 mL), H$_2$O (5 mL) and PPh$_3$ (1.56 g, 5.93 mmol, 2.0 eq.) was added at room temperature. After stirring for 12 h, the reaction was concentrated to remove most of the THF. The resulting suspension was diluted with 1 M HCl and extracted with diethyl ether (3 × 30 mL). The aqueous layer was then basified by adding 1 M NaOH till pH~12 and extracted with EA (3 × 40 mL). The combined organic extracts were dried over MgSO$_4$, filtered, and concentrated under reduced pressure to get the crude amine, which was purified by silica gel chromatography (gradient DCM/MeOH/TEA, v/v/v, 100:0:0 to 100:1:0 to 50:1:0.1) to afford the title compound as a light-yellow oil (302 mg, 25 %). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.03 (d, J = 2.1 Hz, 1H), 7.11 (d, J = 8.4 Hz, 4H), 7.03 (dd, J = 8.7, 2.3 Hz, 2H), 6.80 (d, J = 8.4 Hz, 4H), 6.35 (d, J = 8.8 Hz, 1H), 4.65 (s, 4H), 3.73 (s, 6H), 3.35 (s, 2H), 2.69 (d, J = 6.8 Hz, 2H), 1.69 − 1.58 (m, 1H), 1.17 (td, J = 12.6, 6.8 Hz, 1H), 0.85 − 0.68 (m, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 158.51 (s), 157.08 (s), 146.11 (s), 135.13 (s), 130.47 (s), 128.22 (s), 125.55 (s), 113.84 (s), 105.61 (s), 55.14 (s), 50.15 (s),

45.90 (s), 23.87 (s), 18.62 (s), 13.01 (s). HRMS: [M+H]$^+$ calculated for $C_{25}H_{30}N_3O_2$, 404.2338; found 404.2332.

1-(((trans)-2-(6-(Di(4-methoxybenzyl)aminopyridin-3-yl)cyclopropyl)methyl)-2-benzyl carbamate-3-(3H-pyrazol-4-yl)guanidine (5.17) This compound was obtained as a light-yellow solid (210 mg, 63 %) according to the procedure used for the synthesis of compound **5.14**, starting from Cbz-NCS **5.13** (1.2 mL, 0.5 M in DCM), alkylamine **5.12** (227 mg, 0.583 mmol), pyrazol-4-amine (63 mg, 0.755 mmol), diisopropylethylamine (96 μL, 0.533 mmol) and EDCI (145 mg, 0.755 mmol). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.02 (d, J = 2.3 Hz, 1H), 7.43 (dd, J = 7.5, 1.4 Hz, 2H), 7.39 – 7.29 (m, 5H), 7.14 (d, J = 8.6 Hz, 4H), 7.05 (dd, J = 8.4, 2.2 Hz, 1H), 6.83 (d, J = 8.6 Hz, 4H), 6.38 (d, J = 8.8 Hz, 1H), 5.71 (s, 2H), 5.38 (d, J = 1.7 Hz, 1H), 5.18 (s, 2H), 4.67 (s, 4H), 3.78 (s, 6H), 1.72 – 1.62 (m, 1H), 0.93-0.83 (m, 3H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 158.71 (s), 157.48 (s), 146.53 (s), 142.37 (s), 136.64 (s), 135.63 (s), 130.59 (s), 128.59 (s), 128.50 (s), 128.40 (s), 128.26 (s), 124.59 (s), 114.04 (s), 105.81 (s), 89.60 (s), 67.62 (s), 55.37 (s), 50.30 (s), 19.15 (s), 14.28 (s), 12.84 (s). HRMS: [M+H]$^+$ calculated for $C_{37}H_{40}N_7O_4$, 646.3142; found 646.3148.

Deprotection of intermediate 5.17 The protected intermediate **5.17** (210 mg, 0.325 mmol) in DCM (4 mL) was treated with TFA (5 mL) and anisole (3 mL) at room temperature for 5 h. To remove nonpolar impurities, the residue was partitioned between ether (15 mL) and H$_2$O (15 mL) and after separation, the ether layer further extracted with an additional H$_2$O (5 mL). The aqueous layers were combined and concentrated *in vacuo*. Analysis of this crude residue by HRMS did not show the formation of the desired compound.

HIV antiviral assay. CD4$^+$CXCR4$^+$MT-4 cells (1 $\times$ 10^6 cells/mL; 50 μL) were seeded in cell culture medium in a 96-well plate (Falcon, BD Biosciences, Erembodegem, Belgium) and pre-incubated with 5-fold dilutions of test compound (100 μL) at 37 °C for 30 min, in duplicate. Virus (50 μL; HIV strains NL4.3 and ROD) was added according to a 50 % cell culture infectious dose (CCID$_{50}$) which was determined by titration of the stock. After an incubation period of 5 days, virus-induced cytopathic effect (CPE) was scored with light microscopy and IC$_{50}$ was calculated using the spectrophotometric MTS viability staining assay (Cell-Titer96 Aqueous One Solution Pro-liferation Assay kit; Promega, Leiden, The Netherlands). Absorbance was measured using the Versamax microplate reader and SoftMax Prosoftware (Molecular Devices, Sunnyvale, Ca, USA). To determine the compounds potential cellular cytotoxicity, the assays were performed without the addition of virus.

For experimental details of **HIV-1 RT inhibition assay** in this chapter, please refer to Experimental section of Chapter 3.

5.5 References

1. Gilboa, E.; Mitra, S. W.; Goff, S.; Baltimore, D. A detailed model of reverse transcription and tests of crucial aspects. *Cell* **1979,** 18, 93-100.

2. Balzarini, J.; Das, K.; Bernatchez, J. A.; Martinez, S. E.; Ngure, M.; Keane, S.; Ford, A.; Maguire, N.; Mullins, N.; John, J.; Kim, Y.; Dehaen, W.; Vande Voorde, J.; Liekens, S.; Naesens, L.; Gotte, M.; Maguire, A. R.; Arnold, E. Alpha-carboxy nucleoside phosphonates as universal nucleoside triphosphate mimics. *Proc Natl Acad Sci U S A* **2015,** 112, 3475-80.

3. Balzarini, J.; Ford, A.; Maguire, N. M.; John, J.; Das, K.; Arnold, E.; Dehaen, W.; Maguire, A. Alpha-carboxynucleoside phosphonates: direct-acting inhibitors of viral DNA polymerases. *Future Med Chem* **2019,** 11, 137-154.

4. Das, K.; Balzarini, J.; Miller, M. T.; Maguire, A. R.; DeStefano, J. J.; Arnold, E. Conformational States of HIV-1 Reverse Transcriptase for Nucleotide Incorporation vs Pyrophosphorolysis-Binding of Foscarnet. *Acs Chem Biol* **2016,** 11, 2158-64.

5. Mullins, N. D.; Maguire, N. M.; Ford, A.; Das, K.; Arnold, E.; Balzarini, J.; Maguire, A. R. Exploring the role of the alpha-carboxyphosphonate moiety in the HIV-RT activity of alpha-carboxy nucleoside phosphonates. *Org Biomol Chem* **2016,** 14, 2454-65.

6. Jochmans, D.; Deval, J.; Kesteleyn, B.; Van Marck, H.; Bettens, E.; De Baere, I.; Dehertogh, P.; Ivens, T.; Van Ginderen, M.; Van Schoubroeck, B.; Ehteshami, M.; Wigerinck, P.; Gotte, M.; Hertogs, K. Indolopyridones inhibit human immunodeficiency virus reverse transcriptase with a novel mechanism of action. *J Virol* **2006,** 80, 12283-92.

7. Ruiz, F. X.; Hoang, A.; Das, K.; Arnold, E. Structural Basis of HIV-1 Inhibition by Nucleotide-Competing Reverse Transcriptase Inhibitor INDOPY-1. *J Med Chem* **2019,** 62, 9996-10002.

8. Fragment-Based Drug Design: Tools, Practical Approaches, and Examples. In *Methods in Enzymology*, Kuo, L. C., Ed. Elsevier, Inc.: San Diego, CA, **2011**; Vol. 493.

9. Bauman, J. D.; Patel, D.; Dharia, C.; Fromer, M. W.; Ahmed, S.; Frenkel, Y.; Vijayan, R. S.; Eck, J. T.; Ho, W. C.; Das, K.; Shatkin, A. J.; Arnold, E. Detecting allosteric sites of HIV-1 reverse transcriptase by X-ray crystallographic fragment screening. *J Med Chem* **2013,** 56, 2738-46.

10. Kalyaanamoorthy, S.; Chen, Y. P. Structure-based drug design to augment hit discovery. *Drug Discov Today* **2011,** 16, 831-9.

11. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **2010,** 31, 455-461.

12. F. P. Schmidtchen; Berger, M. Artificial Organic Host Molecules for Anions. *Chem. Rev.* **1997,** 97, 1609-1646.

13. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods* **2000,** 44, 235-249.

14. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, Design and Selection* **1995,** 8, 127-132.

15. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

16. Toy, P. H.; Dhanabalasingam, B.; Newcomb, M.; Hanna, I. H.; Hollenberg, P. F. A Substituted Hypersensitive Radical Probe for Enzyme-Catalyzed Hydroxylations: Synthesis of Racemic and Enantiomerically Enriched Forms and Application in a Cytochrome P450-Catalyzed Oxidation. *J. Org. Chem.* **1997,** 62, 9114-9122.

17. Giannetti, A. M.; Zheng, X.; Skelton, N. J.; Wang, W.; Bravo, B. J.; Bair, K. W.; Baumeister, T.; Cheng, E.; Crocker, L.; Feng, Y.; Gunzner-Toste, J.; Ho, Y. C.; Hua, R.; Liederer, B. M.; Liu, Y.; Ma, X.; O'Brien, T.; Oeh, J.; Sampath, D.; Shen, Y.; Wang, C.; Wang, L.; Wu, H.; Xiao, Y.; Yuen, P. W.; Zak, M.; Zhao, G.; Zhao, Q.; Dragovich, P. S. Fragment-based identification of amides derived from trans-2-(pyridin-3-yl)cyclopropanecarboxylic acid as potent inhibitors of human nicotinamide phosphoribosyltransferase (NAMPT). *J Med Chem* **2014,** 57, 770-92.

18. Linton, B. R.; Carr, A. J.; Orner, B. P.; Hamilton, A. D. A Versatile One-Pot Synthesis of 1,3-Substituted Guanidines from Carbamoyl Isothiocyanates. *J. Org. Chem.* **2000,** 65, 1566-1568.

19. Wong, Y. C.; Ke, Z.; Yeung, Y. Y. Lewis Basic Sulfide Catalyzed Electrophilic Bromocyclization of Cyclopropylmethyl Amide. *Org Lett* **2015,** 17, 4944-7.

20. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009,** 30, 2785-2791.

Chapter 6. General discussion and perspectives

6.1 General discussion

Nucleosides are important biomolecules that function as precursors to nucleotides that are involved in various processes such as nucleic acid biosynthesis, cellular signalling and metabolic energy transactions.[1] Structural modifications of the nucleobase and/or sugar moiety of natural nucleosides have afforded numerous nucleoside analogs possessing a plethora of biological properties, although they are mostly known for their antiviral activity. Despite the fact that the first antiviral nucleoside analog, idoxuridine, was approved in 1963,[2] using nucleoside analogs as antivirals had not become widespread until AZT was approved as the first HIV drug to treat AIDS. To date, nearly 30 antiviral nucleos(t)ide analogs have been licensed for clinical use. As the largest class of antiviral small molecules, nucleoside analogs are the cornerstone of antiviral chemotherapy and they are being used in the treatment of viral infections caused by DNA viruses (such as VZV, HSV and HBV) RNA viruses (e.g. HCV, influenza virus, Ebola virus and coronavirus) and retroviruses (such as HIV).[3]

To exert their antiviral activity, nucleoside analogs generally need three phosphorylation steps via host cell kinases to generate the corresponding nucleoside triphosphate (NTP) analogs that act as substrate mimics of viral DNA or RNA polymerases. Their catalytic incorporation into viral DNA or RNA leads to chain termination. Because of the high substrate specificity of cellular kinases, these biotransformations often proceed inefficiently, resulting in a weak or complete lack of antiviral effect of many nucleoside analogs. Especially the first phosphorylation step, affording the nucleoside monophosphate (NMP), is often inefficient and rate-limiting, as is the case for anti-HIV drug d4T. Moreover, the clinical efficacy of nucleoside analogs suffers from additional problems including variable oral bioavailability, poor biological half-lives and the development of drug resistance.[4-5] To overcome some of these issues, prodrugs of NMP and acyclic nucleoside phosphonates (ANPs) have been developed successfully, such as the approved antivirals sofosbuvir, remdesivir and tenofovir alafenamide (TAF), which can bypass the rate-limiting first phosphorylation step and thereby only need two phosphorylations (instead of three for general nucleoside analogs) to become biologically active.

In the three-step activation process of nucleoside analogs, the second or third step may also be a hurdle for some nucleoside analogs such as AZT, d4U and ddU.[6] As a consequence, it is still desirable to develop NTP analogs, which can circumvent all phosphorylation steps and act as direct inhibitors or substrates of viral polymerases. Furthermore, although a number of antiviral drugs received marketing approval, their effectiveness might be limited by toxicity and drug resistance after

long-term use. Therefore, it is also urgently needed to find new antiviral candidates more resilient to resistance and endowed with less adverse effects.

Extensive studies in our group had previously demonstrated that various amino acid phosphoramidates of dAMP behave as dATP mimics and act as direct substrates of DNA polymerases without requiring any metabolic phosphorylations. However, few examples of amino acid conjugates of ANPs have been described before.[7-11] Even though, molecular modelling was performed to explain the possible binding modes of these compounds, no crystal structures are available yet to experimentally determine how they bind and what drives the catalytic incorporation at the polymerase active site. In the current study, amino acid conjugates of dAMP and TFV, were prepared following a previously reported procedure.[12] Compounds were evaluated for their potential incorporation in a growing DNA strand, catalyzed by HIV-1 RT, as well as their ability to function as HIV-1 RT inhibitors. A systematic X-ray structural analysis was also done, as described in **Chapter 2**.

Although none of the newly synthesized amino acid conjugates of dAMP showed noticeable incorporation, most of the amino acid conjugates of TFV were recognized successfully as substrates by HIV-1 RT, and the TFV parts were catalytically incorporated into the primer strand, functioning as chain terminators. It was observed that the different amino acid conjugates have a different incorporation efficiency, with the L-Glu conjugate providing the highest TFV incorporation, followed by L- and D-Asp conjugates. The crystal structures of the TFV conjugates in complex with RT/DNA indicate that the carboxyl groups form salt bridges with the guanidinium group of the highly conserved residue Arg72, which plays an essential role in dNTP binding and positioning.[13-14] No metal chelation was observed in the binding of these compounds in the presence of Mg^{2+} alone, whereas one Mn^{2+} positioned in the metal B site was involved in metal chelation in the presence of Mn^{2+}. Although no ordered Mg^{2+} ion was seen at the active site in crystal structures, RT still incorporated TFV from Glu/Asp conjugates suggesting that Mg^{2+} ions participate in the dynamic process of compound binding and TFV incorporation. However, these compounds are weak binders compared with the natural substrate dATP, and their interactions with Arg72 are solvent-exposed and can be dissociated easily when a dNTP competes for binding. As a result, these compounds cannot compete with dATP binding, and thereby fail to show significant RT inhibition.

In **Chapter 3**, we described the synthesis, biological evaluation and X-ray structural study of a series of tenofovir analogs in which a stable amide linker was introduced to replace the labile phosphonoamidate bond of the TFV-amino acid conjugates. Initially, we decided to prepare optically pure TFV analogs, from which the methyl group in the acyclic chain has the same stereochemistry as in TFV. Although an optically pure starting material was used and Mitsunobu reaction was also attempted for the synthesis of initially designed compounds, NMR spectra indicate that the final

compounds were diastereomeric mixtures. To circumvent racemization, a series of compounds lacking a methyl group in the acyclic chain was synthesized. *In vitro* RT inhibition assay demonstrates that several compounds are endowed with very weak RT inhibition. Moreover, several ester prodrugs to improve cell permeability were prepared. All compounds were tested for antiviral activity against HIV in a cell-based assay. Most analogs showed very weak anti-HIV-1 activity in TZM-bl cells. A more pronounced cellular cytotoxicity was observed for several of ester prodrugs.

In **Chapter 4**, the synthesis, biological evaluation and binding mode analysis of a series of new ANP analogs were described. Structural variations of the chain length, stereochemistry and linker atom were introduced to study influences on activity and binding. To prepare these structurally diverse compounds, several synthetic routes were established, in which a Borch or Mitsunobu reaction played a key role. For compounds with different chain length, the intermediates showed a different reactivity, so various reactions and synthetic routes were evaluated accordingly. These ANP analogs were tested for their inhibitory activity against RT using a gel-based *in vitro* RT inhibition assay. It was demonstrated that most compounds were only weak inhibitors of RT. They only displayed activity at low millimolar concentrations, when a low concentration of the natural substrate dATP (5 μM) was used. Since these compounds are very polar and only show very weak RT inhibition, we did not test their activity in a cell-based assay. X-ray crystallographic analysis revealed that these ANP analogs have distinct binding modes when compared with those from Chapters 2 and 3, by forming new interactions at the polymerase active site. The carboxylic groups were involved in metal chelation with one Mg^{2+} ion, whereas the phosphonate groups in these compounds mainly interact with side chains of surrounding residues such as Lys64, Arg72 and Gln151. Although these compounds can chelate with one metal ion, the chelation geometry is far from optimal considering the two-metal ion octahedral coordination of dATP or T-α-CNP. In addition, despite the fact that these compounds form salt bridges with the highly conserved residue Arg72 that is solvent-exposed and essential for dNTP binding, their polar interactions with Arg72 can be easily dissociated once a dNTP binding. Consequently, all these compounds are unable to outcompete dATP for binding. To find compounds with more favorable chelation and inhibition, further explorations are desirable.

In the last part of this thesis (**Chapter 5**), the design, docking study and synthesis of new fragments based on the X-ray structure of a fragment hit 166 in complex with HIV-1 RT/dsDNA is described. Initially, the phenyl moiety of hit 166 was replaced with a 5'-amino or hydroxyl pyridine ring enabling the formation of two hydrogen bonds with the template thymine base. In addition, various five-membered heteroaromatics instead of the thiazole moiety were introduced, as a docking study suggested that they might form additional hydrogen bonds with surrounding residues. Furthermore, the introduction of an amidine or guanidine linker as a replacement of amide linker was

also envisioned, with the purpose of forming additional interactions with the catalytic residues D110, D185 and D186.

Because of the straightforward chemistry, several fragments containing a pyridyl ring and an amide linker were prepared for a preliminary structural study. Moreover, in order to increase binding affinity, the synthesis of two additional fragments with a 5'-amino-pyridine ring and a guanidine moiety was attempted. A Corey-Chaykovsky reaction was utilized for the preparation of *trans*-cyclopropane derivatives. Initially, the yield was low for the cyclopropanation, possibly due to the hydrolysis of the ester group in basic condition. Therefore, an alternative synthetic route was elaborated that yielded the *trans*-cyclopropane derivatives in high yields. As key intermediates of selected fragments, substituted guanidine derivatives were prepared successfully via challenging synthetic routes, but the following deprotection failed because of an unexpected cyclization.

Among the synthesized fragments, two compounds showed weak RT inhibition, but due to their weak binding at the P pocket of HIV-RT, they did not display any antiviral activity or cytotoxicity at the highest tested concentration of 100 μM. However, one cryoEM structure obtained showed that the introduced pyridine moiety can form one H-bond with template thymine as expected.

6.2 Future perspectives

To develop NTP mimics that do not require metabolic phosphorylation, a series of new TFV-amino acid conjugates was synthesized in **Chapter 2**. *In vitro* HIV-1 RT inhibition assay demonstrated that L-Met-TFV acted as a NcRTI. Although this compound only displayed weak inhibition of HIV-1 RT as determined in a biochemical assay, it can be used as a lead for further structural modifications of the nucleobase base, the acyclic chain the phosphonate moiety and the amino acid part (**Figure 6-1**). However, since the main liability of this type compound class is the instability of the phosphonamidate bond and the fact that the phosphonate moiety is essential for metal chelation, the focus of further modifications should be on phosphonate moiety taking into account both chemical stability and metal chelation. Not only bioisosterism but also scaffold hopping approach[15] can be applied for new compounds design. These further explorations might generate antiviral compounds based on new scaffolds, better affinity and enhanced stability.
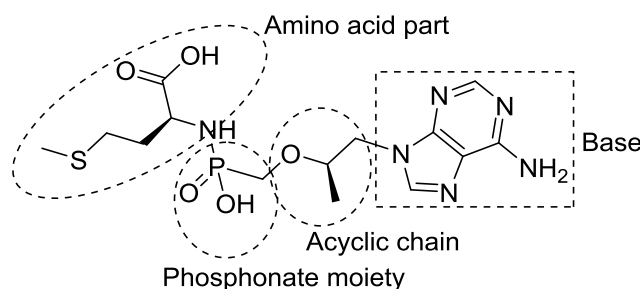

Figure 6-1. Further modifications sites of L-Met-TFV

The dAMP/TFV-amino acid conjugates from **Chapter 2** were also evaluated as potential inhibitors of *E. coli* RNA polymerase. Interestingly, several compounds showed inhibitory activity at micromolar concentrations in an enzymatic assay. However, because of their chemical lability and high polarity, these compounds were not tested in cell-based antibacterial assays. Among these compounds, D-Asp-TFV (**Chapter 2**, **2.4b**) displayed *E. coli* RNA polymerase inhibition with an IC_{50} value of 9.3 μM. As future work, we will carry out a structural study to confirm its activity and investigate its binding mode at polymerase active site. In addition, when its activity is confirmed with detailed structural information obtained, we can use structure-based drug design approaches to design and synthesize new compounds to improve binding affinity and stability. This may result in a new series of antibacterial agents.

In order to increase the stability of TFV-amino acid conjugates, the labile phosphonoamidate moiety was replaced by a more stable amide linker in **Chapter 3**. Several compounds show weak inhibitory activity in a gel-based RT inhibition assay and a cell-based anti-HIV assay. Systematic modifications can be done on these compounds to further improve their antiviral activity. With the crystal structures obtained, a docking study or virtual screening can be performed for the design of novel compounds, prior to start wet chemistry. Taking into consideration docking results and synthesis feasibility, the most promising compounds can be selected for chemical synthesis followed by biological evaluations.

It is worth noting that one of these compounds (Chapter 3, compound **3.5d**) showed weak activity (IC_{50} ~ 8 μM) and several structure-closely related analogs lack activity against SARS-CoV-2 RNA polymerase using a gel-based enzymatic assay, indicating that this compound may be further modified to find compounds with better activity. Since the cryogenic-electron microscopy (cryo-EM) structure of SARS-CoV-2 RdRp in complex with RNA/remdesivir has been reported,[16] this information can be used to perform a systematic docking study to design new compounds acting as potential NTP-competitive inhibitors. Compounds with favorable binding affinities can then be selected for synthesis and subjected to biochemical and antiviral evaluation.

ANP analogs have two main structural advantages. The phosphonate moiety is chemically and metabolically stable and hence is resistant to enzymatic degradation *in vivo*. In addition, the flexibility of the acyclic chain allows the molecules to adopt different conformations to bind targeted polymerases. Therefore, in an effort to gain metal chelation, a phosphonate moiety was reintroduced yielding seven ANP analogs in **Chapter 4**. Unfortunately, X-ray structure analysis indicated that the carbonyl or carboxyl group rather than phosphonate group of these compounds was involved in metal chelation. In addition, these compounds only show weak RT inhibition. Thus, further modifications are needed to discover compounds with enhanced activity. Based on the binding modes and synthesis

experience of the previously synthesized compounds, new ANP analogs with two phosphonate groups are proposed for future work (**Figure 6-2**). These compounds might chelate two metal ions at the polymerase active site, making them potentially more favorable as NTP mimics without requiring metabolic conversion and acting as direct inhibitors of viral polymerases. Their synthesis can be achieved using a combination of synthetic routes developed by us, and procedures reported in the literature.
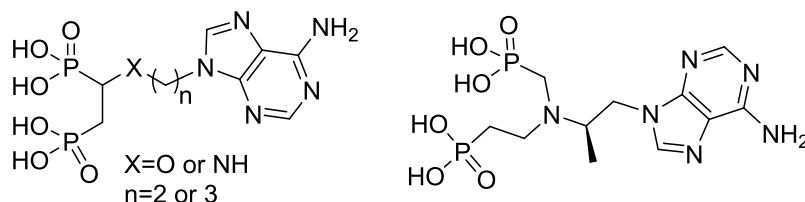

Figure 6-2. Proposed new ANP analogs with two phosphonate groups

Although nucleoside phosphonates are mostly known as antiviral agents, they exhibit a wide range of other biological activities such as antitumoral, antibacterial and antiparasitic.[17] Therefore, it is worthwhile to screen the newly synthesized ANP analogs from this doctoral thesis for these biological activities as well. All these compounds are very polar making it difficult to enter cells. Therefore, various prodrugs, such as ester prodrugs, can be prepared to increase their cellular permeability. Moreover, to gain insight into the binding mode of these new ANP analogs at the active site of the targeted protein, structural studies (e.g. using X-ray crystallography, cryo-EM) can be performed.

In **Chapter 5**, the preparation of two selected fragments (F47 and F81) was attempted using challenging synthetic routes. Unfortunately, upon removal of the protecting group in the last step, no desired compound was formed, based on mass spectral analysis of the crude residue. The exact structural identification of this side product will require its isolation and purification, allowing for structural assignment by NMR.

Although we failed to obtain the desired fragments F47 and F81, the synthetic explorations are valuable for the synthesis of other fragments. For example, following a reported synthetic route,[18] the previously obtained intermediate **5.15** can be used for the synthesis of F79 and F97 (**Figure 6-3**). Briefly, a Mitsunobu reaction between the alcohol derivative **5.15** and the commercially available agent **6.1** affords intermediate **6.2**. The appropriate amines 1,3,4-thiadiazol-2-amine and (2-thiazolyl)methylamine are subsequently guanylated with the guanylating agent **6.2** furnishing the Boc-protected guanidines **6.3** and **6.4**, respectively. After PMB and Boc deprotection using TFA, the desired compounds F79 and F97 are finally obtained as TFA salts. It is noteworthy that the two compounds are synthesized as racemic mixtures.
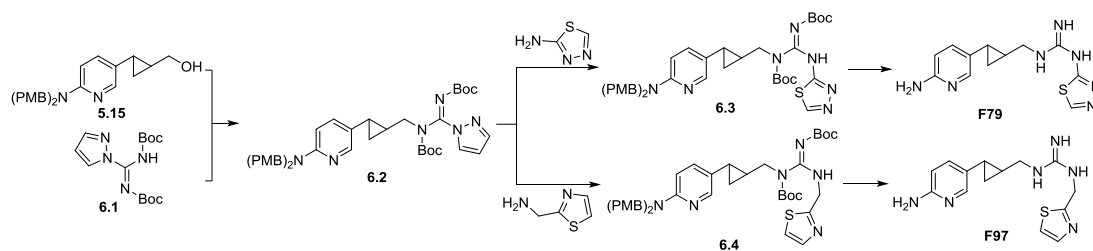
Figure 6-3. Proposed synthetic route towards F79 and F97.

We obtained one cryo-EM structure of fragment F04 (Chapter 5, compound **5.5d**) in complex with HIV-1 RT/DNA which demonstrated that the nitrogen atom of the pyridine ring formed a hydrogen bond with the template thymine as predicted by molecular docking. However, we have not yet confirmed that the amino and guanidine groups of these fragments can also form the desired interactions. Therefore, as a future work plan, further structural studies need to be performed to check if two hydrogen bonds are formed with the template thymine and if the guanidine and heteroaromatic moieties of fragments F79 and F97 can engage in additional interactions as shown by the docking study. At the same time, the two compounds can be screened not only for their antiviral activity, but also for antitumoral and antibacterial activity considering that structurally related guanidine-containing compounds were reported to have antitumoral and antibacterial activity.[18-19]

6.3 References

1. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry, 6th Edition*. W.H. Freeman and Company, New York, USA: **2012**.
2. De Clercq, E. In Search of Selective Antiviral Chemotherapy. *Clin Microbiol Rev* **1997,** 10, 674-693.
3. Ami, E.-i.; Ohrui, H. Intriguing Antiviral Modified Nucleosides: A Retrospective View into the Future Treatment of COVID-19. *ACS Medicinal Chemistry Letters* **2021,** 12, 510-517.
4. Van Rompay, A. R.; Johansson, M.; Karlsson, A. Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases. *Pharmacol. Therapeut.* **2000,** 87, 189-198.
5. Asahchop, E. L.; Wainberg, M. A.; Sloan, R. D.; Tremblay, C. L. Antiviral Drug Resistance and the Need for Development of New HIV-1 Reverse Transcriptase Inhibitors. *Antimicrob Agents Ch* **2012,** 56, 5000-5008.
6. Gollnest, T.; Dinis de Oliveira, T.; Rath, A.; Hauber, I.; Schols, D.; Balzarini, J.; Meier, C. Membrane-permeable Triphosphate Prodrugs of Nucleoside Analogues. *Angew Chem Int Ed Engl* **2016,** 55, 5255-8.
7. Adelfinskaya, O.; Herdewijn, P. Amino acid phosphoramidate nucleotides as alternative substrates for HIV-1 reverse transcriptase. *Angew Chem Int Ed Engl* **2007,** 46, 4356-8.
8. Adelfinskaya, O.; Terrazas, M.; Froeyen, M.; Marliere, P.; Nauwelaerts, K.; Herdewijn, P. Polymerase-catalyzed synthesis of DNA from phosphoramidate conjugates of deoxynucleotides and amino acids. *Nucleic Acids Res.* **2007,** 35, 5060-5072.
9. Giraut, A.; Song, X. P.; Froeyen, M.; Marliere, P.; Herdewijn, P. Iminodiacetic-phosphoramidates as metabolic prototypes for diversifying nucleic acid polymerization in vivo. *Nucleic Acids Research* **2010,** 38, 2541-2550.
10. Song, X. P.; Bouillon, C.; Lescrinier, E.; Herdewijn, P. Iminodipropionic acid as the leaving group for DNA polymerization by HIV-1 reverse transcriptase. *ChemBioChem* **2011,** 12, 1868-1880.
11. Yang, S. Q.; Froeyen, M.; Lescrinier, E.; Marliere, P.; Herdewijn, P. 3-Phosphono-L-alanine as pyrophosphate mimic for DNA synthesis using HIV-1 reverse transcriptase. *Organic & Biomolecular Chemistry* **2011,** 9, 111-119.
12. Abraham, T. W.; Kalman, T. I.; McIntee, E. J.; Wagner, C. R. Synthesis and biological activity of aromatic amino acid phosphoramidates of 5-fluoro-2'-deoxyuridine and 1-beta-arabinofuranosylcytosine: evidence of phosphoramidase activity. *J Med Chem* **1996,** 39, 4569-75.
13. Sarafianos, S. G.; Pandey, V. N.; Kaushik, N.; Modak, M. J. Site-directed mutagenesis of arginine 72 of HIV-1 reverse transcriptase. Catalytic role and inhibitor sensitivity. *J Biol Chem* **1995,** 270, 19729-35.
14. Das, K.; Bandwar, R. P.; White, K. L.; Feng, J. Y.; Sarafianos, S. G.; Tuske, S.; Tu, X.; Clark, A. D., Jr.; Boyer, P. L.; Hou, X.; Gaffney, B. L.; Jones, R. A.; Miller, M. D.; Hughes, S. H.; Arnold, E. Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *J. Biol. Chem.*

2009, 284, 35092-100.

15. Hu, Y.; Stumpfe, D.; Bajorath, J. Recent Advances in Scaffold Hopping. *Journal of Medicinal Chemistry* **2016,** 60, 1238-1246.

16. Yin, W.; Mao, C.; Luan, X.; Shen, D; *et al*. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020,** 368, 1499-1504.

17. Groaz, E.; De Jonghe, S. Overview of Biologically Active Nucleoside Phosphonates. *Frontiers in Chemistry* **2021,** 8.

18. Maccari, G.; Sanfilippo, S.; De Luca, F.; Deodato, D.; Casian, A.; Dasso Lang, M. C.; Zamperini, C.; Dreassi, E.; Rossolini, G. M.; Docquier, J.-D.; Botta, M. Synthesis of linear and cyclic guazatine derivatives endowed with antibacterial activity. *Bioorganic & Medicinal Chemistry Letters* **2014,** 24, 5525-5529.

19. Sidoryk, K.; Świtalska, M.; Jaromin, A.; Cmoch, P.; Bujak, I.; Kaczmarska, M.; Wietrzyk, J.; Dominguez, E. G.; Żarnowski, R.; Andes, D. R.; Bańkowski, K.; Cybulski, M.; Kaczmarek, Ł. The synthesis of indolo[2,3-b]quinoline derivatives with a guanidine group: Highly selective cytotoxic agents. *European Journal of Medicinal Chemistry* **2015,** 105, 208-219.

Acknowledgements, Personal contributions and Conflict of interest statments

2. Personal contribution

Weijie Gu (WG) synthesized and characterized all final compounds in the thesis. Sergio Martinez (SM) contributed to the soaking and crystallization of all synthesized compounds with RT/DNA. SM and Kalyan Das (KD) are responsible for X-ray data collection, refinement and analysis in Chapter 2~3. SM and Abhimanyu K. Singh are responsible for data collection, refinement and analysis in Chapter 4. Hoai Nguyen performed the gel incorporation assay in Chapter 2 and RT inhibition assay in Chapters 3~5. Hongtao Xu and WG conducted the PicoGreen inhibition assay. Steven De Jonghe and KD contributed to extensively proofreading this thesis.

3. Conflict of interest statements

No conflicts of interest are declared.

Curriculum Vitae

Weijie Gu was born in the Anhui Province, China in 1991. He started studying Pharmacy in September 2010, at Anhui University of Chinese Medicine, where he obtained his Bachelor of Medicine (B.M.) degree in June 2014. Then he was enrolled in China Pharmaceutical University to gain further experience in Medicinal Chemistry under the supervision of Prof. Hai Qian and received his Master of Science (M.S.) degree in June 2017. On the 25th Spetember 2017, he joined Prof. Kalyan Das's group and started his doctoral training at the Rega Institute, KU Leuven.

Publications

1. **W. Gu**, S. Martinez, H. Nguyen, *et al*. Tenofovir-amino acid conjugates act as polymerase substrates – Implications for avoiding cellular phosphorylation in the discovery of nucleotide analogs [J]. *J. Med. Chem*. **2021**, 64, 1, 782–796.

2. **W. Gu**, S. Martinez, A. K. Singh, *et al*. Exploring the dNTP binding site of HIV-1 reverse transcriptase for inhibitor design. *Eur. J. Med. Chem*, **2021**, 225, 113785. DOI: 10.1016/j.ejmech.2021.113785.

3. A. K. Singh,[1] S. E. Martinez,[1] **W. Gu**, H. Nguyen, D. Schols, P. Herdewijn, S. De Jonghe, K. Das. Sliding of HIV-1 reverse transcriptase over DNA creates a transient P pocket – Targeting P-pocket by fragment screening. DOI:10.21203/rs.3.rs-685322/v1. (Under review in *Nature Communications*)

4. G. Li, M. Xu, T. Yue, **W. Gu**, *et al*. Life-long passion for antiviral research and drug development: 80th birthday of Prof. Dr. Erik De Clercq [J]. *Biochemical Pharmacology*. **2021**, 185, 114485.

5. G. Li, T. Yue, P. Zhang, **W. Gu**, *et al*. Drug Discovery of Nucleos(t)ide Antiviral Agents: Dedicated to Prof. Dr. Erik De Clercq on Occasion of His 80th Birthday [J]. *Molecules*. **2021**, 26(4), 923.

6. P. Herdewijn, **W. Gu**, E. De Clercq. Anti-DNA Virus Agents. In *Burger's Medicinal Chemistry and Drug Discovery* [B], **2021**, pp 1-41.

7. W. Huang, H. Qian, **W. Gu**. Preparation of the 2,4-disubstituted pyridine and their medicine application for the treatment of cancer [P]. CN108069938A, **2018**.

8. **W. Gu**, Y. Dai, W. Shi, *et al*. Discovery of novel 2-substituted-4-(2-fluorophenoxy) pyridine derivatives possessing pyrazolone and triazole moieties as dual c-Met/VEGFR-2 receptor tyrosine kinase inhibitors [J]. *Bioorg. Chem*. **2017**, 72, 116-122.

9. J. Li, **W. Gu**, X. Bi, *et al*. Design, synthesis, and biological evaluation of thieno[2,3-d] pyrimidine derivatives as novel dual c-Met and VEGFR-2 kinase inhibitors [J]. *Bioorg. Med. Chem*. **2017**, 25(24), 6674-6679.

10. H. Qiang, **W. Gu**, Y. Dai, *et al*. Design, synthesis and biological evaluation of 4-aminopyrimidine-5-cabaldehyde oximes as dual inhibitors of c-Met and VEGFR-2 [J]. *Bioorg. Med. Chem*. **2016**, 24, 3353-3358.

11. Z. Li, J. Yang, **W. Gu**, *et al*. Discovery of a novel oxime ether scaffold as potent and orally bioavailable free fatty acid receptor 1 agonists [J]. *RSC Adv.*, **2016**, 6, 46356-46365.

Conferences/Workshops attended

1. 15th International Conference on Therapeutic Targets and Medicinal Chemistry 2021 (ICTTMC 2021), Amsterdam, The Netherlands, January 21-22, 2021 (e-Poster presentation, won *Best Presentation Award*)

2. 4th Spring Symposium DDPS, Leuven, Belgium, May 3, 2019.

3. Janssen Lectures in Medicinal Chemistry Editon 2018-2019.