UNIVERSIDAD
DE GRANADA

Programa de Doctorado en Farmacia

Departamento de Química Farmacéutica y Orgánica

Memoria de Tesis Doctoral

Síntesis de Nucleósidos y su Evaluación Biológica como Inhibidores Selectivos de la Retrotransposición de LINE-1 de Mamíferos

Doctorando

Guillermo Bañuelos Sánchez

Directores

Dr. Juan Antonio Tamayo Torres

Dr. José Luis García Pérez

Tutor

Dr. Francisco Franco Montalbán

Editor: Universidad de Granada. Tesis Doctorales
Autor: Guillermo Bañuelos Sánchez
ISBN: 978-84-1306-261-7
URI: http://hdl.handle.net/10481/56499

Agradecimientos

Largo y arduo ha sido el camino que hasta aquí me ha traído, pero sobre todo enriquecedor, por lo que me dispongo a agradecer por tanto recibido.

Quiero agradecer en primer lugar a mis directores de tesis, Juan A. Tamayo Torres y José Luis García Pérez, el haberme brindado la oportunidad de llevar a cabo este trabajo de investigación. A Juan debo agradecerle estar donde estoy por aquella proposición, hace 10 años ya y que no olvidaré, de aprender a trabajar en un laboratorio de química orgánica y abrirme la puerta al maravilloso mundo de la síntesis orgánica. A José le agradezco descubrirme el intrigante y apasionante LINE-1, así como su paciencia y comprensión ante un ignorante de la biología molecular (un poco menos ahora). A ambos les he de agradecer también la confianza depositada en mí, la generosidad, la igualdad de trato, el respeto y consideración de mis opiniones y sobre todo sus conocimientos y visión de la ciencia, que tanto me han enriquecido como investigador.

Agradecer también al grupo BIO-250 de la Universidad de Granada, del que formo parte, así como del Departamento de Química Farmacéutica y Orgánica por su apoyo y los conocimientos aportados. En especial agradezco a Francisco Franco Montalbán, tutor de la presente tesis, sus sabios consejos y enseñanzas tanto científicas como formativas durante todo mi paso por el grupo de investigación, que tanto me han aportado. También a José Antonio Gómez Vidal, la Junta de Andalucía y el Fondo Social Europeo por el contrato de Garantía de Juvenil que ha permitido mi financiación durante parte de esta tesis.

A todos mis compañeros del laboratorio 8 de GENYO agradezco vuestro cariño, acogiéndome desde el primer día como uno más y dedicándome tanto el tiempo y esfuerzo por los que os debo gran parte de esta tesis. Sois un grupo maravilloso.

Agradezco a todos los compañeros que han pasado por laboratorio del grupo BIO-250, en especial a Fernando y Javi, quienes me enseñaron algunas de las primeras técnicas de laboratorio, siendo el segundo el responsable de ser el adicto al café que soy. Quiero agradecer en particular a María Dolores, más que compañera, amiga del principio al fin de este camino, que ha estado en los buenos y en los peores momentos siempre que lo he necesitado, y con quien he crecido como investigador gracias a su inteligencia.

Agradezco a Antonio, compañero además de amigo, los tres meses inolvidables que compartimos en Edimburgo, así como a todos los compañeros de laboratorio del MRC.

Además, agradezco a Thomas Widmann el realizar la estancia Internacional y por la formación recibida.

También quiero agradecer por haberme dado tanto a mis padres y mi hermano, a quienes todo debo, por quienes soy quien soy, os quiero. Imagino que estaréis orgullosos de mí (modestia aparte), pero más lo estoy yo de tener esta familia.

Por último, agradezco a Encarni, mi pareja, tanta paciencia, tanto apoyo, tanta compresión, tanta madurez, tanto amor mostrados durante todo este camino, sin los cuales no estaría aquí. Gracias.

ÍNDICE

Síntesis de Nucleósidos y su Evaluación Biológica como Inhibidores Selectivos de la Retrotransposición de LINE-1 de Mamíferos

CAPÍTULO 1: SÍNTESIS DE NUCLEÓSIDOS Y SU EVALUACIÓN BIOLÓGICA COMO INHIBIDORES SELECTIVOS DE LA RETROTRANSPOSICIÓN DE LINE-1 DE MAMÍFEROS

RESUMEN

Los retrotransposones LINE-1 (Long Insterspersed Element class I) son los únicos elementos transponibles autónomos activos en la actualidad. Suponen el 17% del genoma humano, si bien la mayoría de las copias son inactivas debido a mutaciones y reordenamientos. Sin embargo, el genoma humano contiene aproximadamente 80-100 copias activas de LINE-1 y su movilización produce un impacto sobre nuestro genoma. Los LINE-1s se movilizan mediante un mecanismo de copia-pega que implica la reversotranscripción del RNA mensajero de LINE-1. En humanos, la retrotransposición de LINE-1 durante el desarrollo embrionario temprano genera inserciones heredables que pueden resultar potencialmente mutagénicas y producir un desorden genético. Por otro lado, los LINE-1 son también activos en células cancerígenas y cerebro humano, produciendo un impacto sobre nuestro genoma somático. Los elementos LINE-1 son por lo general sobreexpresados en tumores humanos y la acumulación de nuevas inserciones tiene la capacidad de aumentar la malignidad y potencial metastásico de las células cancerígenas. Por ello, aunque el papel de la retrotransposición en tejidos somáticos es aún desconocido, tiene el potencial de influir en el origen y progresión del cáncer, así como de la biología del cerebro. Con el objetivo de descubrir más sobre el papel de la actividad de LINE-1 en células cancerígenas y cerebrales, una estrategia de pérdida de función podría resultar muy informativa. En el presente trabajo y para explorar si la inhibición de la reversotranscriptasa de LINE-1 puede ser usada como una eficaz estrategia de pérdida de función, hemos ensayado compuestos con estructura nucleosídica sintetizados y adquiridos comercialmente como inhibidores selectivos de la retrotransposición de LINE-1. Tras analizar 33 nucleósidos mediante un ensayo de retrotransposición, hemos identificado 3 inhibidores selectivos y no tóxicos de la retrotransposición de LINE-1 que no muestran actividad frente otros retroelementos (retrotransposones LTR). Además, estos compuestos inhiben eficientemente LINE-1s humanos y de ratón, y podrían ser empleados para determinar el impacto de la actividad de LINE-1 en modelos de

ratón a través de una estrategia de pérdida de función. Cabe destacar que estos compuestos podrían potencialmente usarse como tratamiento coadyuvante de tumores con una alta expresión de LINE-1.

ABSTRACT

Long interspersed element class 1 (LINE-1 or L1) is the only active autonomous transposable element in humans. It comprises 17% of our genomic mass, although most L1 copies are inactive due to mutations and rearrangements. However, an average human genome contains 80-100 active L1s, and their mobilization continue to impacts our genome. L1s move by a copy and paste mechanism that involves the reverse transcription of an intermediate L1 RNA. In humans, L1 mobilization during early embryogenesis led to the accumulation of heritable insertions that sporadically can be mutagenic and generate a genetic disorder. Surprisingly, L1s are also active in cancer cells and the human brain, impacting our somatic genome. L1 elements are frequently overexpressed in human tumors, and the accumulation of new insertions can potentially increase the malignancy and metastasis potential of tumor cells. Thus, although the role of somatic retrotransposition is currently unknown, it has the potential to influence cancer origin and progression, as well as brain biology. To learn more on the role of L1 activity in cancer cells and the brain, Loss of Function (LOF) strategies will be very informative. Here, and to explore whether inhibition of the Reverse Transcriptase of L1 can be used as an effective LOF strategy, we tested currently available and de novo synthesized nucleoside related compounds as selective L1 inhibitors. After analysing 33 nucleoside structures using an L1 retrotransposition assay, we identified three non-toxic selective L1 inhibitors that exhibit no activity against other retroelements (LTR-containing retrotransposons). Notably, these compounds can efficiently inhibit human and mouse LINE-1s, and could be used to determine the impact of L1 activity in mouse models, in a LOF approach. Importantly, these compounds also have the potential to be effective in the treatment of those cancers characterized by a high L1 expression.

1. Introducción

1 INTRODUCTION

1.1 HUMAN GENOME COMPOSITION, TRANSPOSABLE ELEMENTS AND THEIR GENOMIC IMPACT

1.1.1 Human genome and repeated sequences

The sequencing of the human genome revealed new features that changed our view on genomics. Although early on researchers had observed the presence of repetitive sequences on the human genome [1], [2], it was not until the finalization of the Human Genome Project when we realized about the astonishing abundance of repeated DNA sequences in our genome. Indeed, while less than 5 % of the genome is made of coding sequences (i.e., exons), repeated sequences account for more than 50% of the human genome [3] More recently, additional studies revealed that up to two-thirds of our genome is made of repeated sequences [4] (**Figure 1**). Among repeated sequences, transposable elements comprise for 45% of the human genome. Repetitive sequences are present in both prokaryotic and eukaryotic organisms [5], [6], and these sequences can be classified as follows [3]:

- Transposable Element-derived repeats or interspersed repeats.
- Processed pseudogenes: copies of genes generated by the genomic insertion of their mRNA sequences through a retrotransposition process.
- Simple sequence repeats: repetition of nucleotides (for instance $(A)_n$, $(CA)_n$, $(CGG)_n$).
- Segmental duplications: highly similar sequences of 10-300 kb in size, duplicated from one region to a different one by recombination.
- Tandem repeats. Telomeres, centromeres, ribosomal gene clusters and short arms of acrocentric chromosomes.

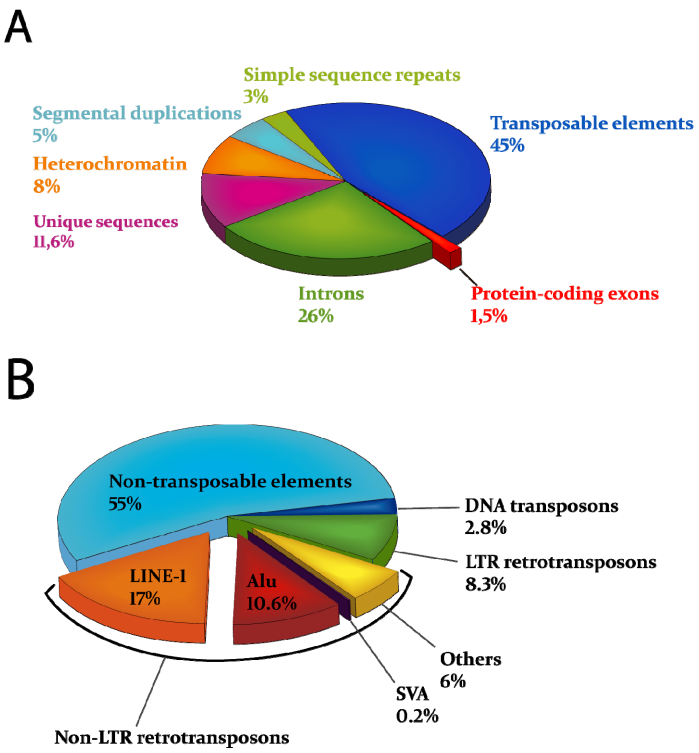


Figure 1. *Human genome composition.* ***A.*** *Global view of the main components of the human genome (adapted from [7]).* ***B.*** *Content of the human genome in transposable elements (adapted from [8]).*

In all living organisms, DNA has two vital roles: on one hand, it contains the information needed for the synthesis of macromolecules, and on the other hand DNA is responsible to transfer this information to descendants. Because repeated sequences are not conventional genes, what are those repeat sequences doing in our genome? The presence of such repeated sequences was and still is disconcerting due to the lack of contribution to phenotype or regulatory function, so it was firstly considered not only as "junk" DNA [9], but also as "selfish" or "parasitic" DNA that use the cell enzymatic machinery in order to survive within genomes [10], [11]. However, that point of view has been evolving over the years with new evidences that prove the interaction of repeated DNA sequences with our genome (reviewed in [12], [13]). For instance, the evolution of human glycophorin gene family through several gene duplication events is related with the recombination of Alu elements, a type of repeated DNA sequence [14]. In addition to the possibility of contributing to the evolution of human genes, the expansion of repetitive DNA sequences has been related with the origin of several human

developmental and degenerative disorders (reviewed in [15]), neurological and neuromuscular diseases, as well as with the progression and severity of them (reviewed in [16]).

1.1.2 Transposable elements: the human mobilome

The largest group of repeated DNA sequences in the human genome are transposable elements (TEs), which, as a whole, make up the 45% of our genome [3]. TEs are a type of repeated DNA sequences that are able to move and/or duplicate their sequences from one position within the genome to a new one, either through a "cut-and-paste" or "copy-and-paste" mechanism. The combination of all TEs present in a genome is called *the mobilome*.

The first evidence of the existence of such repeated DNA sequences was provided by Barbara McClintock in 1950, when she discovered that variegation in maize kernels and leaves was associated with the mobility of an Activator/Dissociation TE in maize (reason why she was awarded with the Nobel Prize in Physiology or Medicine in 1983) [17]. McClintock established that TEs promote genetic variety and also suggested that TEs might be involved in the regulation of gene expression if transposition took place at or near the locus of a gene [18]. The latter view was not widely accepted and the discussion about the role of TEs in genomes continues at present. Since these seminal findings by McClintock, many others TEs have been identified in different species, and virtually any genome sequenced contained TE-derived sequences.

In eukaryote genomes, and based on its mobilization mechanism, TEs are classified into class I and class II elements. The former, also known as retrotransposons, spread in genomes using a "copy-and-paste" mechanism, in which a new copy of the TE is generated by reverse transcription of an intermediate ARN. By contrast, class II (also known as DNA transposons), generally mobilize using a "cut-and-paste" mechanism, excising the sequence of the TE from the initial genomic position and inserting it into a new genomic loci. Both class I and class II elements are further subclassified as autonomous or non-autonomous, depending on whether they encode or not, respectively, for the enzymatic

machinery required for its mobilization [19], [20]. **Figure 2 and table 1** collect some features of main TE types.

To illustrate the diversity of TEs in genomes, below I will use the human genome as an example to describe the different types of TEs found in our DNA, which are also present in most vertebrate genomes.

1.1.2.1 DNA transposons

DNA transposons comprise around 3% of the human genome and are active in many organisms including bacteria (reviewed in [21]), maize (Ac/Dc aforementioned) and others plant species (reviewed in [22]), the nematode *Caenorhabditis elegans* [23] and *Drosophila* [24] among others. However DNA transposons are no longer active in the human genome (due to the accumulation of mutations [3]) and most mammals, with the exception of some bat species [25], [26]. Although inactivated in human genome, DNA transposons have significantly interacted with it in the past. For instance, RAG1 and RAG2 proteins involved in V(D)J recombination during lymphocyte development, have their origin in an ancient DNA transposon [27], [28].

As the enzyme that mediates the transposition reaction, termed Transposase, cannot recognize active from inactive copies, over time inactive transposon copies are mostly transposed, making their accumulation in genomes a less efficient process. Therefore, the mobilization of DNA transposons is inactivated trough evolution, although DNA Transposons can survive in nature using horizontal transfer [29]–[31]. Although the mechanism responsible for of horizontal transfer DNA Transposons it is not completely known, some events may involve external parasites [32] or viruses [33] as vehicules.

Structurally, DNA transposons are composed by a gene coding for the enzyme transposase that is flanked by two terminal inverted repeat sequences (TIRs). The 5´TIR harbour an RNA polymerase II promoter, generating the mRNA of transposase. Upon translation, the transposase access the nucleus where transposition take place through a "cut-and-paste" mechanism. During transposition, the integrase and DNA binding activities of the transposase

recognize TIRs sequences and excise the transposase gene from its original genomic location, and subsequently inserts it into a new genomic loci. After a transposition event, *Target Site Duplications* (TSDs) flaking the inserted sequence are formed, because of the staggered cut of transposase at the target site. Thus, transposition is generally a non-replicative process. Furthermore, those sequences accidentally flanked by TIRs or containing a mutated transposase gene, would behave as non-autonomous DNA transposons, which could use the transposase from an active element to mobilize themselves.

Based on their sequence, enzymology and structural characteristics, DNA transposons are further subclassified in subclass I and subclass II. The former is composed of "cut-and-paste" DNA transposons in the strict sense and can be further divided into a variety of families being the most representative in humans the hAT (e.g. MER1-Charlie) and Tc-1/*mariner* groups (e.g. MER2-Tigger). Subclass II includes transposons with a replicative process via a rolling-circle mechanism in the case of *Helitron* elements or via self-synthesizing DNA in the case of *Politons*/*Mavericks* [3], [19]–[21], [34]–[38].

1.1.2.2 Retrotransposons

Retrotransposons are copy-and-paste elements, whose sequences are first transcribed into mRNA, translated, and the encoded proteins have the capacity to reverse-transcribe their mRNA into DNA, which is ultimately inserted elsewhere in the genome. Their mobilization is known as retrotransposition, and is a replicative process. As DNA transposons, new retrotransposon insertions are often flanked by variable sized TSDs.

Retrotransposons can be subdivided into two main groups: Long Terminal Repeat (LTR)-containing retrotransposons, also known as endogenous retroviruses (ERVs), and non-LTR retrotransposons, which contain a poly(A) tail at the end of their sequence. Furthermore, both LTR and non-LTR retrotransposons could be subclassified in autonomous and non-autonomous elements (which retrotranspose using the proteins encoded by active/autonomous elements.

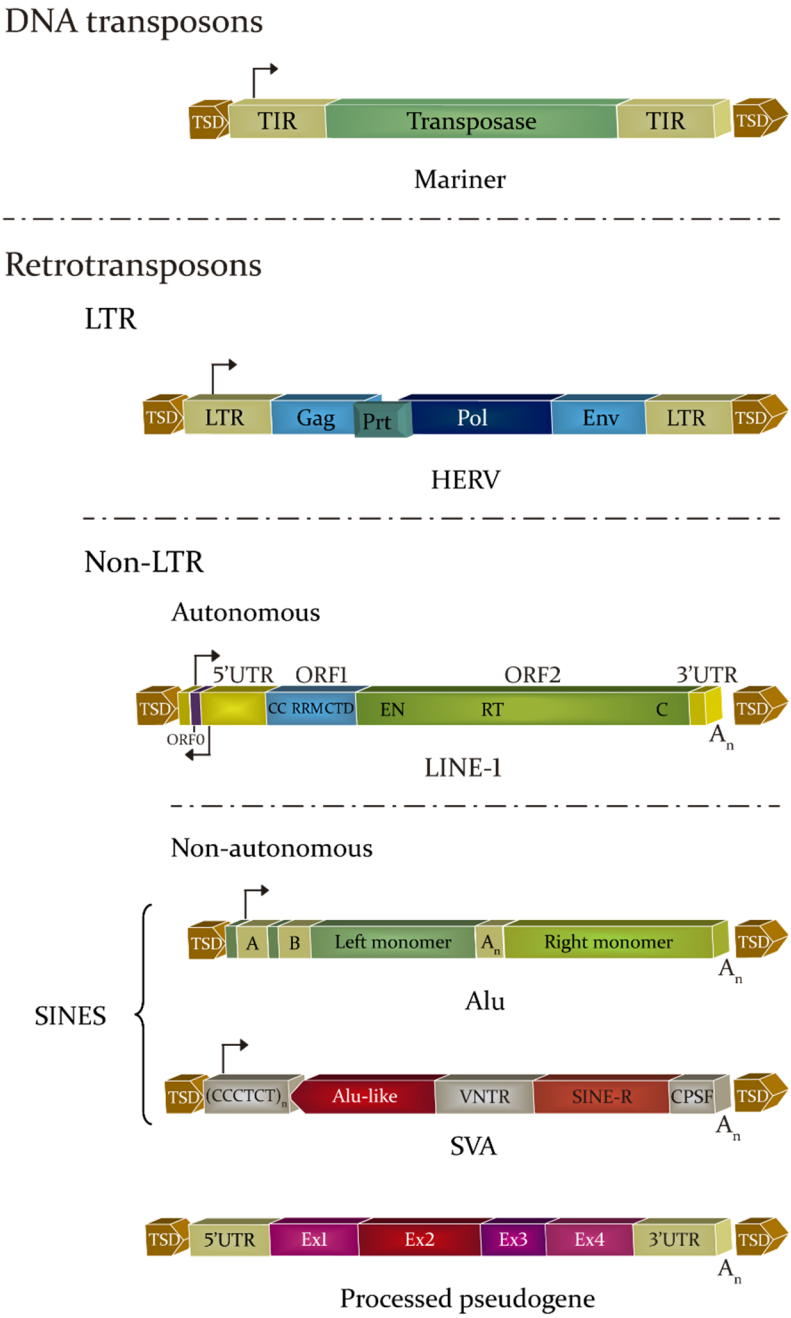


Figure 2. *Structure and classification of human transposable elements. TEs are subdivided into DNA transposons (class II) and retrotransposons (class I). The latter are in turn subclassified into LTR retrotransposons and non-LTR retrotransposons which can be further divided into autonomous and non-autonomous elements. Black arrows represent promoter activity. Red rectangle tagged with Alu-like represent antisense orientation of this domain. Abbreviations: TSD, target site duplication; TIR, terminal inverted repeat; LTR, long terminal repeat; Gag, group-specific antigen gen; Prt, protease; Pol, polymerase (reverse transcriptase); Env, envelope gene; UTR, untranslated region; ORF, open reading frame; CC, coiled-coil; RRM, RNA recognition motif; CTD, carboxyl-terminal domain; EN, endonuclease domain; RT, reverse transcriptase domain; C, cysteine-rich domain; An, A-rich domain or poly A tail; A and B, domains with promoter activity; VNTR, variable number target repeats; CPFS, cleavage polyadenylation specific factor; Ex, exon. Figures are not represented in scale (adapted from references within text).*

TRANSPOSABLE ELEMENT	% OF GENOME	LENGTH	COPY NUMBER	ACTIVE
DNA TRANSPOSONS	~3%	2-3 kb	~450,000	No
RETROTRANSPOSONS				
AUTONOMOUS				
HERVs	~8%	6-10 kb	~450,000	Unknown
LINEs	~21%	~6 kb	~850,000	Yes
NON-AUTONOMOUS				
SINEs				
Alu	~10%	~300 bp	~1,009,000	Yes
SVA	0,2%	700-4,000 bp	~2,700	Yes
Processed pseudogenes	<1%	Variable	~11,000	Yes

Table 1. *Overview of some key features of human transposable elements. It should be noticed that a slightly different classification, in relation to figure 2, is depicted here. Percentage of human genome, length, copy number and current activity are listed for each human TE. LINE's row includes the data from the three LINE families (L1, L2 and L3). Data are collected from references within text.*

1.1.2.2.1 LTR retrotransposons

In humans, endogenous retroviruses (known as HERVs, for human endogenous retroviruses) are the most representative group of LTR retrotransposons, comprising about 8% of our genome. Phylogenetic studies have determined that ERVs colonized genomes before placental-mammalian radiation (~70 million years ago) [39], [40].

It is not really known whether exogenous retrovirus was integrated by eukaryotic genomes becoming ERVs; or the latter was released from cells, acquiring an envelope gene, giving rise to exogenous retrovirus; or a combination of both hypothesis. The first hypothesis is the stronger one, and is based on experiments in mice where implantation of previously infected embryos with a retrovirus (murine leukemia virus or MLV) led to the expansion of MLV-derived ERVs [41]. The second hypothesis, which seems more speculative, comes from a phylogenetic analysis of the gene Ribonuclease HI, which is encoded by LTR-containing elements in Eubacteria and Eukarya [42].

From a mechanistic point, HERVs are also similar to exogenous retrovirus, and even reverse transcription takes place inside cytoplasmic virus-like particles [43], [44]. It has been assumed that HERVs no longer retrotranspose in humans, at least in the germline [3]. The presence of polymorphic HERV-K elements in the human genome seems to indicate that they became inactive after the speciation of humans and chimpanzees (~6 million years ago) [45], [46]. Indeed, an intact and polymorphic HERV-K was recently identified, and this element had an intact sequence and could potentially undergo retrotransposition [47]. Furthermore, the infectious capacity of a reanimated HERV-K provirus has been demonstrated [48], [49]. In rodents, ERVs are still active, and elements such as IAP and MusD elements have been associated with the generation of several genetic disorders in mice [50]; similarly, LTR-retrotransposons are also active in yeast (i.e., Ty1 and Ty3 LTR-retrotransposons [43], [44].

Structurally, HERVs resemble exogenous retroviruses: they possess ORFs that encode capsid and matrix proteins that surround the retroviral RNA (*gag*), a protease that cleaves the viral polyproteins encoded by *gag* into smaller functional proteins and a protein with reverse transcriptase and integrase activity (*pol*). However the sequence that encodes the viral envelope protein (*env*) is non-functional or completely absent, thus making HERVs not allowed to exit the cell. The flanking LTRs (300-1000 bp) encompass regulatory elements required for the transcription and maturation of HERV RNAs, including a RNA polymerase II promoter [3], [51]–[54]. Based on their sequences, four groups of ERVs can be distinguished:

-Complete ERVs that contain intact ORFs, including *env*, which are similar to exogenous retrovirus and are virtually able to retrotranspose.

-Deficient ERVs that have one or more retrotransposition-required ORFs mutated, generally the *env* gene.

-ERVs which coding sequences are substituted by non-coding DNA flanked by LTRs.

-"Solo LTR" ERVs, which are made of just one LTR, and that are likely generated by homologous recombination between LTRs. This group is more prominent.

Although considered as fossils, HERVs are not exempt from influencing our genome. DNA hypomethylation during embryogenesis allows HERV-K to be expressed, conferring an immunoprotective effect against reinfection of some endo and exogenous retroviruses [55]. A domesticated *env* gene from a HERV-W is the responsible for the formation of syncytin, required protein for placenta morphogenesis [56], [57]. Additionally ERVs have generated multiple interferon-inducible enhancers, thereby contributing to innate immunity [58].

1.1.2.2.2 Non-LTR retrotransposons

As ERVs, non-LTR retrotransposons mobilize through an RNA intermediate, but in contrast to ERVs, retrotransposition occurs by a process termed target-primed reverse transcription (TPRT), in which reverse transcription occurs in the nucleus. AS other TEs, ther are autonomous and non-autonomous non-LTR elements [59], [60].

1.1.2.2.2.1 LINE-1

Long interspersed element class 1, LINE-1 or L1, is the only autonomous active retrotransposon in the human genome. As an autonomous element, it is able to encode the enzymes required for its own mobilization in *cis*, which are also responsible for the mobilization of non-autonomous non-LTR elements and other cellular mRNAs generating processed pseudogenes in *trans*. Beside LINE-1s, the human genome also harbour LINE-2 and LINE-3 elements (accounting for around 4% of our genome), although they are completely inactive [3]. In this Thesis, I have researched mostly on human LINE-1s, and this the structure of LINE-1 elements, their retrotransposition mechanism, their regulation and the consequences of its mobilization will be discussed below in detail.

1.1.2.2.2.2 SINEs

Short interspersed elements are unable to retrotranspose by themselves so all genome copies of these elements have been generated using the proteins

encoded by L1, what is in agreement with the typical L1 hallmarks (i.e., TSDs, a polyA tail and consensus insertion site mediated by L1 EN) [38]. We can distinguish two types: Alu and SVA.

Alu

Alu elements are non-autonomous non-LTR elements with a recognition site for *AluI* restriction enzyme as a common feature [61]. Their origin is placed about 65 million years ago, coinciding with primate radiation [62]. Regarding the copy number, Alu elements are the most abundant retroelement in human genome, making up around 11% of it [3]. These elements are primate specific [63] and their origin resides in the 7SL RNA, component of the signal recognition particle [64], [65] through joining two free monomers together [66], [67].

Alu are non-coding sequences and as a non-autonomous elements, once incorporated to the LINE-1 ribonucleoprotein particle (see below), Alu elements retrotranspose by a TPRT mechanism taking advantage of LINE-1 ORF2 proteins (ORF2p) [68], process known as *trans*-mobilization. Furthermore, although not required for its mobilization, LINE-1 ORF1 proteins (ORF1p) enhance Alu retrotransposition [69]. Unlike LTR and LINE-1 retrotransposons, Alu elements are more prone to be inserted in the proximity of genes and also in intronic regions of them [70].

A typical full-length Alu element consists in two 7SL derived monomers, left and right, with similar sequences and an adenosine-rich sequence between them. Left monomer contains A and B boxes, which exhibit promoter activity [71] for RNA polymerase III [72], whereas the right monomer, lacking those boxes, ends in a poly(A) signal. The terminator signal for RNA polymerase III (four or more consecutive T residues) is located within the genomic region downstream of the Alu element [72]. Most ancient subfamilies of Alu elements are Alu J and Alu S, while Alu Y subfamily is the youngest one (reviewed in [73]), the main contributor of Alu retrotransposition in humans and seemingly, the only active Alu subfamily [74] although just a limited number of elements, known as “master” or “source” elements [75], [76], are still retrocompetent. Furthermore, Alu insertions have

been proved to be polymorphic with respect to insertion presence/absence in genomes throughout human population, thus being useful as genetic marks in phylogenetic studies [77], [78]. It is believed that somatic Alu activity happens in human genomes although there is no definite evidence of that. Studies have focused on germline and some cell-based assays, both showing the capacity of Alu elements to retrotranspose, which indicate that somatic tissues may support Alu retrotransposition [79], [80].

SVA

SINE-VTNR-Alu (SVA) are non-autonomous primate-specific retrotransposons and the youngest TEs in humans. They have been present in primate genomes for the last 25 million years [81], [82] and now comprise 0.2% of the human genome. Like Alu, the non-autonomous nature of SVA elements relegates them to be retrotranscribed *in trans* employing the enzymatic machinery of LINE-1 proteins upon integration on a LINE-1 ribonucleoprotein particle [83], [84].

Its name is a reflection of their composite constitution (**Figure 2**): a hexameric CCCTCT repeat; two antisense Alu-like element; a stretch of GC-rich variable nucleotide tandem repeats (VNTRs); a SINE-R with certain homology with the *env* and right LTR of a HERV-K; a binding site for cleavage polyadenylation specific factor (CPSF); a polyadenylation signal (reviewed in [82]). Upon SVA insertion, generation of TSDs flank the element.

The absence of an internal RNA polymerase III internal promoter, together with the presence of a poly(A) signal, multiple RNA polymerase III terminators and the fact that SVA elements are too long to be transcribed by the mentioned enzyme, makes SVA likely to be transcribed by RNA polymerase II, although it have not been proved yet. SVA transcription might start in an internal promoter or in an upstream promoter and can ends in its own poly(A) signal or bypass it until transcription reaches a downstream poly(A) signal [81], [82]. SVA remains active based on polymorphic insertions throughout the human population [85] and reported cases of insertion-created diseases (will be further discussed).

1.1.2.2.2.3 Processed pseudogenes

LINE-1 enzymes are able to retrotranscribe and integrate cellular mRNAs into the genome giving rise to processed pseudogenes (PPs) whose sequences lack the introns present in their precursors [86]. From 8000 to 20000 copies of PPs have been identified in human genomes [87], [88] being the most successful sequences those coming from ribosomal protein mRNAs and housekeeping genes [89]. New inserted sequences terminate with a poly(A) and are flanked by TSDs and, along with introns, lack a promoter so they are, at least in principle, not possible to be transcribed.

Detection of processed pseudogene polymorphism has been published, demonstrating that PPs insertions are still happening at an important rate [90]–[92]. Retrotranscription of processed mRNAs may result in an evolutionary advantage, such as the TRIM5/CypA fusion protein stemmed from the insertion of CypA mRNA into the TRIM5 gene in owl monkeys, conferring them HIV-1 resistance [93].

1.1.2.2.2.4 Group II intron

The mitochondrial and chloroplast genomes of some eukaryotes, and also bacterial DNA, harbour group II introns. Nevertheless, they are not present in nuclear genomes. They are mobile elements with two main features that enable the spread of group II introns: are autocatalytic RNA (ribozyme) able to carry out self-splicing and also include an intron-encoded protein with various activities being their reverse transcriptase activity the most remarkable one. Group II intron mobilization occurs via TPRT. The relevance of these retroelements is the possibility that they could be the ancestors of LINE-1, as well as telomerase and other reverse transcriptases, as it has been suggested (reviewed in [94]–[98]).

1.1.3 Human LINE-1

Firstly identified just as a DNA repeated sequence [99], the long interspersed element-1 (LINE-1 or L1), which belongs to non-LTR retrotransposons, is the only active autonomous transposable element in humans. An average human genome contains 500000 copies of L1 elements, accounting the 21% of our

genomic mass [3]. Furthermore, its activity has led to the generation of another 11% of our genome by reversetranscription *in trans* of other cellular mRNAs, as previously mentioned. However, most LINE-1 copies are "fossils" due to 5' truncation or ORFs internal mutations and actually just 80-100 copies per cell are still active. Interestingly, most of the new insertions have been caused by a limited number of L1 elements which are named "hot LINE-1" that belongs to the L1Hs (for human specific) subfamily, also known as L1PA1, most of them included in the L1-Ta (for transcribed-active) subgroup [100]. Those full-length LINE-1 elements that retain the ability to retrotranspose autonomously are called retrocompetents (RC-L1).

1.1.3.1 Structure of a retrocompetent LINE-1 element

Briefly, a full-length L1 element is about 6.0 kb in length and consists in a 5'UTR, three non-overlapping open reading frames (i.e. ORF0, ORF1 and ORF2) and a 3'UTR. Upon insertion, target-site duplications (TSDs) of variable size are generated. **Figure 3** shows most features of a RC-L1 element.

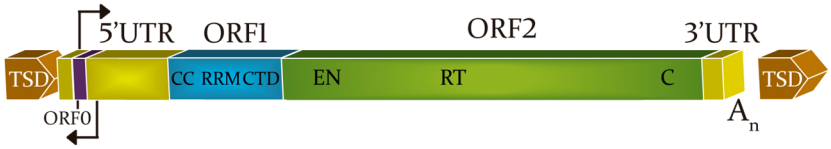


***Figure 3**. Structure of a full-length LINE-1 element. Main features are depicted. Black arrows represent promoter activity. Abbreviations: TSD, target site duplication; UTR, untranslated region; ORF, open reading frame; CC, coiled coil; RRM, RNA recognition motif; CTD, carboxyl-terminal domain; EN, endonuclease domain; RT, reverse transcriptase domain; C, cysteine-rich domain; An, poly A tail. Features are not represented in scale.*

1.1.3.1.1 5'-untranslated region

With 910 bp, the LINE-1 5'UTR harbours an internal RNA polymerase II sense promoter that conducts transcription. The deletion of the first 100 bp leads to a significant drop in L1 expression, highlighting the crucial role of that stretch and concluding that L1 promoter activity resides within that sequence, probably starting from the first base pair. Additionally, other experiments proved that the first 668 bp also have an important role on L1 expression [101]. Moreover, chimeric antisense transcripts of L1 and cellular genes have been detected, as a consequence of an antisense promoter (ASP) activity, located within the L1 5'UTR (around

nucleotides 399-467), which is only able to transcribe contiguous genes that are in the same orientation [102]. Although the origin and role of the ASP is not known, it has been stablished that the transcripts produced from both sense and antisense promoters can form small interfering RNAs (siRNAs), a double-strand RNA that inhibit L1 retrotransposition via RNA interference [103]. Despite that both sense and antisense promoters are active, plasmid-based experiments on pluripotent cells showed that the activity of the ASP is lower than the exhibited by the internal promoter [104].

Relatively recent is the discovery of a previously unknown primate-specific open reading frame, named ORF0. It is included within the 5'UTR region but in the opposite strand to ORF1 and ORF2, concretely comprised between nucleotides 452-236 from the 5' end. Translated ORF0 protein increases L1 activity (but it is not strictly required for L1 retrotransposition), possibly through its binding to negatively charged nucleic acid due to the many positive charges it presents [105].

Different transcription factors are involved in the regulation of L1 expression. The 5'UTR has binding sites for both activating and repressing transcription factors. To the former group belong YY1 [106], [107], RUNX3 [108] and SOX11 [109] whereas SOX2 [110] and MeCP2 [111] belong to the latter group.

1.1.3.1.2 Open reading frame 1

ORF1 protein (ORF1p) [112] is about 40 kDa and 338 amino acids and its expression is considerably higher (from 27:1 to 47:1 ratios) than ORF2 protein (ORF2p), based on *in vitro* experiments using L1 constructs [113]. It exhibits nucleic acid chaperone activity [114] and RNA and single-stranded DNA binding properties, both required for L1 retrotransposition [115]–[117]. Its crystal structure has been elucidated [118], appreciating its previously known trimeric structure and the nucleic acid strand attached to the ORF1p surface rather than encompassed by them. The N-terminus contains a coiled-coil (CC) motif that allows the trimerization process through the interaction between them, stabilized by two chloride ions [119]. Alongside the C-terminal domain (CTD) of ORF1p, the central region acts as an RNA recognition motif (RRM) [117], binding RNA in a non-sequence-specific manner in experiments carried out in mouse ORF1p [120].

It has been stablished that ORF1p are responsible for the cytoplasmic formation of L1 ribonucleoprotein particles (RNPs) in conjunction with ORF2p and L1 RNA [121]–[124]. These RNPs are required for L1 retrotransposition [125] and will be further discussed.

1.1.3.1.3 Open reading frame 2

Spaced by a 63 bp sequence in length from the ORF1 end, ORF2 encodes a 150 kDa protein that possesses endonuclease (EN) [126] and reverse transcriptase (RT) [127] activities, both strictly required for a successful retrotransposition event [116].

The role of the EN domain (which has been crystalized [128]) is to nick one strand of the DNA at the new insertion site, characterized by the consensus sequence 5'-TTTT/A-3' (the slash represents the cleavage site) and some derivatives, leaving a 3' OH group which will serve as a primer for the nascent insertion [126], [129], [130].

The RT domain is in charge of producing cDNA at the insertion site using an RNA template. The catalytic site, which is highly conserved through species [131], requires the F(Y)ADD motif (700-704) to be active [126], [127]. It needs to coordinate with two divalent cations, obtaining higher RT activity when Mg^{2+} is used instead of Mn^{2+} [132], what is in agreement with viral RTs (reviewed in [133], [134]) as well as group II intron RT [135], LTR-retrotransposon RTs [136], [137] and DNA polymerases (reviewed in [138]). Remarkably, L1 RT shows a *cis*-preference for its own mRNA, it means that it preferentially associates to the mRNA of L1 to use it as the template for reverse transcription [139].

A cysteine-rich region with not fully understood function is found at the C-terminus domain although it has been suggested that acts as a zin-knuckle [140] and studies showed that this domain displays RNA binding properties [141]. Mutations within this domain lead to a decrease on L1 retrotransposition since it affects the ability to form the RNPs [124].

1.1.3.1.4 3'-untranslated region

The L1 3'UTR is approximately 206 bp in length and is made up of a polypurine stretch that presumably adopt a guanosine quadruplex structure, based on experiments made with rat L1 [142]. Experiments have proved that this region is not required for an optimal L1 retrotransposition [116].

A weak polyA signal is located within the 3'UTR and transcription often does not stop at this point and RNA polymerase II continues until a canonical polyA signal is found downstream, thus generating chimeric transcripts [143].

In a region overlapping the end of ORF2 and the beginning of 3'UTR has been found a nuclear RNA export factor 1 (NXF1 or TAP) binding site that may tag L1 mRNA to be transported from the nucleus to the cytoplasm [144].

1.1.3.2 LINE-1 retrotransposition cycle

Briefly, human L1 retrotransposition starts in the nucleus from a retrocompetent element located somewhere within the genome, then a cytoplasmic step take place before a final step in the nucleus where a new copy of the element will be inserted elsewhere (**Figure 4**).

Upon full-length L1 transcription from its sense promoter, the bicistronic mRNA (i.e. it encodes for two proteins, ORF1p and ORF2p) is exported to the cytoplasm where is translated into the aforementioned proteins by an unconventional mechanism, according to which ORF1 translation terminates and then it reinitiates to translate ORF2 [145]. Then, several ORF1p molecules and just one or a few ORF2p molecules bind to the L1 mRNA they were encoded by (*cis*-preference), leading to the RNP assembly. It is in this step when other mRNAs, such as SVA, Alu and other cellular RNAs can "hijack" the L1 enzymatic complex to carry out its own retrotransposition (retrotransposition *in trans*) [68], [83], [84]. This complex accesses the nucleus in an unknown nuclear import way and in a cell-division independent manner, although this cell state enhances retrotransposition and has been proposed that it could be a post-translational regulatory mechanism [146], [147].

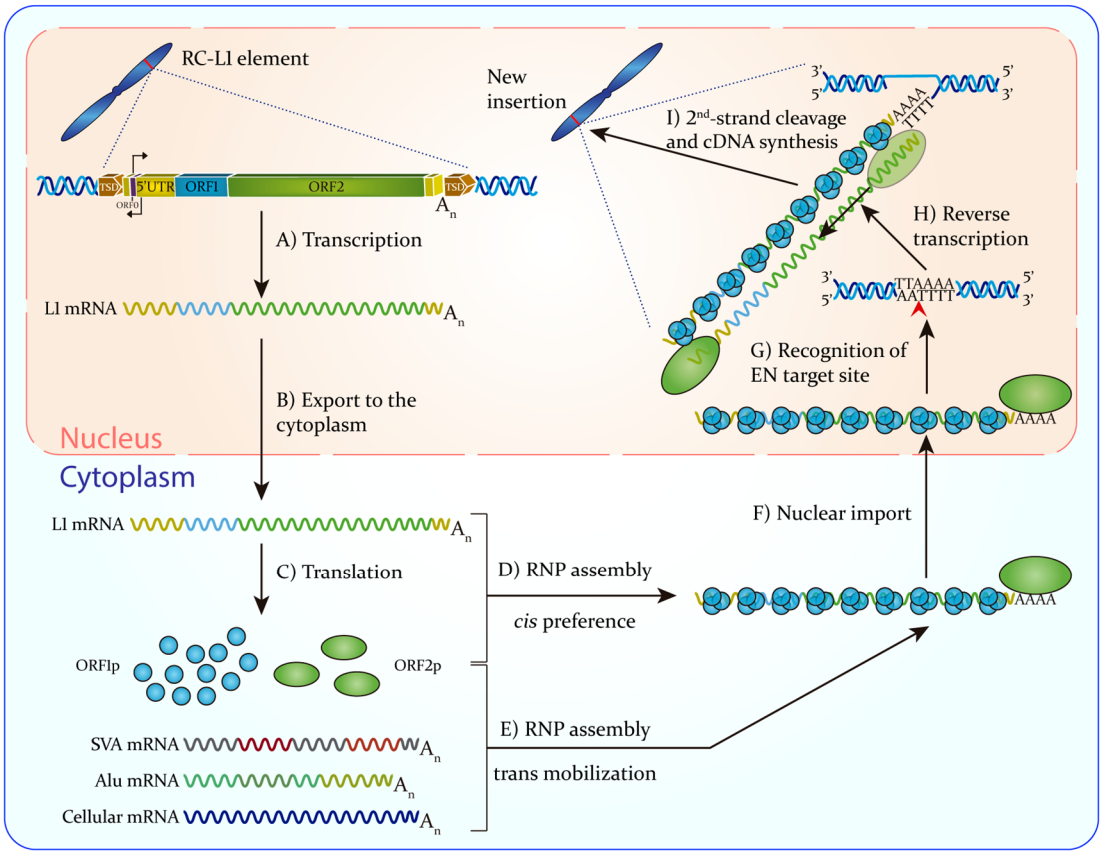


Figure 4*. LINE-1 retrotransposition cycle. A retrocompetent L1 element (coloured boxes) is transcribed (**A**) and its mRNA (coloured wavy line) is exported to the cytoplasm (**B**) where it is translated to its proteins, ORF1p (blue circles) and ORF2p (green ellipses) (**C**). These proteins have cis preference for binding to the mRNA they were encoded by, forming a ribonucleoprotein particle (**D**), although they are able to form RNPs with other mRNAs (**E**), mainly SVA and Alu elements. Once the RNP assembly has taken place, it is imported to the nucleus (**F**) where the EN recognizes a consensus sequence in the DNA and then cleaves one strand (**G**). From this point, the RT synthesizes the complementary DNA according to the template strand (**H**). Finally, a second cleavage in the opposite strand is produced and the cDNA synthesized (**I**), generating a new insertion with its characteristic TSDs. Items are not depicted in scale. Adapted from references within text.*

In contrast to retroviruses and LTR-retrotransposons, in which cDNA synthesis is hold within the cytoplasm to be introduced then into the nucleus and finally inserted into the genome [134], LINE-1 insertion takes place via target-site primed reverse transcription (TPRT). This means that reverse transcription and insertion happen at the site of integration [59], [60]. According to this mechanism, L1 EN produces a single-strand break (SSB) at the consensus, but not strict, sequence 5'-TTTT/A-3'. As a result, a hydroxyl (OH) group from a thymidine is released, what allows L1 mRNA polyA tail annealing to the T-rich genomic DNA at the insertion site, thereby priming ORF2p-mediated reverse transcription. Once reverse transcription ends, cellular RNase H2 degrades the cDNA:RNA hybrid

[148]–[150]. Second-strand nick and second-strand cDNA synthesis still remains unclear but it has been proposed that cellular DNA repair mechanism is involved in those processes. The final step implies the completion and ligation of both ends of the new insertion. Since second-strand nick is produced a few nucleotides downs tream of the first-strand cleavage site, the final step creates the TSDs and these serves as a characteristic feature of L1 insertion.

L1 retrotransposition can undergo an alternative pathway (termed endonuclease-independent retrotransposition) in which reverse transcription takes advantage of an already existing DNA lesions, starting from that point, without participation of L1 EN [151], [152].

1.1.3.3 Regulatory mechanisms of LINE-1 retrotransposition

Despite the potential benefits derived from L1 and, by extension, other TEs retrotransposition, the negative impact of such events might be devastating (both positive and negative impact effects of L1 retrotransposition will be further discussed). Regarding this, it is clear that cells had to adopt mechanisms to control retrotransposition processes, what has been described as an *arm race*. In addition to the previously mentioned transcription factors that affect L1 expression interacting with the L1 5'UTR, there are many regulatory mechanisms at different levels to control L1 retrotransposition in order to maintain genome integrity (reviewed in [153]–[156]). It has to be noted that all the mechanism do not act neither at the same time, nor in all cell types. Here, some of the most important mechanisms are commented.

The first process that restricts L1 retrotransposition is the TPRT mechanism by itself since it produces 5' truncations, internal deletions or inversions so most of the new L1 copies are not full-length elements, therefore not able to undergo retrotransposition. In addition, full-length copies are prone to mutations and DNA recombination processes giving rise to even less RC-L1 elements [157]–[159]. The presence of several splicing sites in L1 mRNA makes possible the elimination of a portion from 5'UTR or ORFs sequences [160]. Moreover, L1 A-rich stretch contains numerous polyA signals what may cause premature polyadenylation reducing the retrotransposition rate [161].

1.1.3.3.1 Transcriptional control

Promoter DNA methylation constitutes a main regulatory mechanism of gene repression through DNA methyltransferases that target cytosines (in mammals) from CpG islands. First during gametogenesis and after in early embryogenesis, methylation patterns are erased and then, during embryo development, different waves of methylation take place, leading to each specific somatic cell line (reviewed in [162], [163]). Many CpG islands are present within L1 5'UTR sequence and are susceptible to be subjected to methylation during embryo development. In consequence, L1 expression in somatic tissues is inactivated but active in the early embryo [164]–[166]. In fact, when somatic cells are reprogrammed into induced pluripotent stem cells (iPSCs), L1 promoter CpG islands methylation level decreases so that L1 expression rises. However, in the same study using iPSCs, it was also detected that new insertions are rapidly silenced by methylation [167]. DNMT3A and DNMT3B are in charge of *de novo* methylation of CpG islands, with the assistance of DNMT3L whereas DNMT1 has a methylation maintenance role. In mice, L1 overexpression is detected if one of these enzymes loss its function what causes sterility because of spermatogenesis failure [168], [169].

An enrichment in histone H3 methylated at Lys9 (H3K9me3), a repressive chromatin mark associated with heterochromatin formation, is detected close to non-LTR elements in mouse ESCs, repressing the expression of such transposable elements [170].

1.1.3.3.2 Post-transcriptional control

L1 transcripts are subjected to different RNA silencing mechanisms that regulate L1 retrotransposition. This mechanism, termed RNA interference (RNAi), acts degrading RNAs at a post-transcriptional level, in a sequence-dependent manner, through three types of small silencing RNAs (ssRNAs): PIWI-interacting RNAs (piRNAs), microRNAs (miRNAs) and small interfering RNAs (siRNAs) (reviewed in [153], [171]).

The piRNAs are small non-coding single-stranded RNAs (24-32 nucleotides in length) expressed in the germline from intergenic clusters throughout the

genome. These type of ssRNAs are built with repetitive elements and recruit PIWI proteins (members of AGO protein family). The resulting piRNA-induced silencing complex (piRISC) is able to degrade L1 RNA (and mouse LTR retrotransposons) due to two features of piRISC: piRNA is complementary to TEs transcripts and AGO protein displays endonucleolytic enzyme activity, so the former guides the latter towards its target. In male mice piRNA regulatory mechanism is crucial since piRISC failure leads to sterility because of gametogenesis arrest. Another potential role of PIWI proteins located within the nucleus is their involvement in DNA methylation of TEs (reviewed in [172]).

The presence of a sense and antisense promoter in L1 elements may lead to the formation of double-stranded RNA that, once processed in the cytoplasm by Dicer (a ribonuclease from RNase III family), give rise to siRNAs. Upon its loading onto AGO2 as single-stranded RNA, the resulting siRISC targets the cleavage of L1 mRNA in a similar way that piRISC does [103]. In contrast to mouse male germline, mouse female germline control mechanism of TEs goes through siRNA [173].

The small non-coding miRNAs (~21-24 nt in length) are produced by the Drosha-DGCR8 complex (named the Microprocessor). Primary transcripts of miRNAs (pri-miRNAs) are transcribed principally from intronic and intergenic regions, in general by RNA polymerase II, as long RNAs that acquire a hairpin structure. Still in the nucleus, pri-miRNAs are potentially bound to DGCR8 because of its two double-stranded RNA-binding domains (that is, DGCR8 recognizes the hairpin), to be then cleaved by Drosha RNase III activity into miRNA precursors (pre-miRNA). Once in the cytoplasm, Dicer mediates maturation of pre-miRNAs to short double-stranded RNAs. Finally one strand, the miRNA, is loaded onto AGO proteins (resulting in miRISC) degrading mRNAs as previously explained [174], [175]. This is the case for mir-128, which directly binds L1 RNA in the ORF2 region, thus repressing L1 retrotransposition in cancer cells and iPSCs in which L1 promoter is hypomethylated, providing an alternative regulatory mechanism when methylation fails [176].

DGCR8 also binds mRNAs that contains hairpin structures such as the one present in the 5'UTR region of L1 transcripts and other TEs. If this region is

engineered removed, retrotransposition rate decreases and, consequently, analogously happens when Drosha processes L1 mRNA [177]. RNase L also restricts L1 retrotransposition degrading L1 transcript either recognizing the 5'UTR stem loop or the dsRNA originated from the sense and antisense promoter activity, although it is not completely clear [178].

Moloney leukemia virus 10 (MOV10) is also involved in controlling L1 retrotransposition mechanism. It exhibits RNA helicase activity and RNA binding properties [179] and acts in the nucleus preventing complementary DNA synthesis [180] and also colocalizes with ORF1p in the cytoplasm sequestering RNPs but its precise mechanism has not been determined yet [181]. Experiments have shown that MOV10 overexpression produces a fall in L1 transcripts while the opposite effect is observed when MOV10 is inhibited [182].

Its mechanism has not been determined yet but it is required a MOV10-RNase H2 association for a successful inhibition [183].

Zinc-finger Antiviral Protein (ZAP), known for its activity against RNA retroviruses, also colocalizes with ORF1p and MOV10 in the cytoplasm and decreases L1 mRNA preventing translation [184], [185].

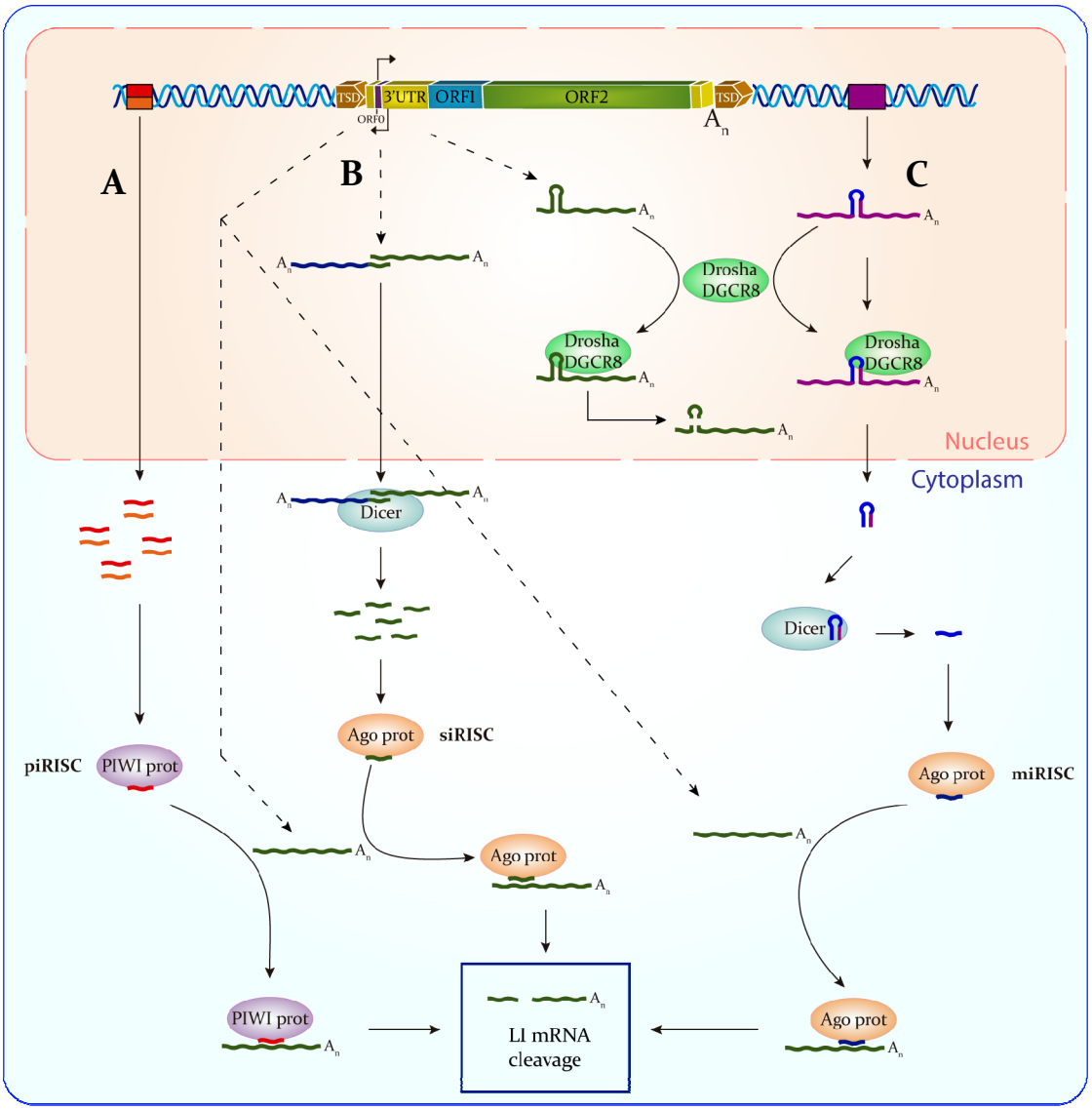


Figure 5. *LINE-1 small silencing RNA regulatory mechanisms (post-transcriptional control).* ***A.*** *PIWI-interacting RNAs: piRNAs are transcribed from an intergenic cluster (red-orange box), then are exported and processed into de cytoplasm (red and orange wavy lines) where they recruits PIWI proteins (purple ellipse) leading to the piRISC. The piRNA within the piRISC is complementary to the LINE-1 transcripts (green wavy lines), binding to them and degrading them by the endonucleose activity of the PIWI proteins.* ***B.*** *Small interfering RNAs: sense and antisense promoters contained in the retrotransposon could generate a double-stranded RNA (central discontinued arrow, green wavy line represents LINE-1 transcript sequence, blue wavy line represents genomic transcript sequence) because of their complementarity, which is processed by Dicer (blue ellipse) in the cytoplasm and loaded onto Ago proteins (brown ellipse). The resulting siRISC binds and cleaves LINE-1 transcripts in a similar way that piRISC does. LINE-1 RNA can form a hairpin (right discontinued arrow) which is recognized by DROSHA-DGCR8 (green ellipse) that degrades it.* ***C.*** *MicroRNAs: are transcribed from intronic or intergenic regions (purple box) and present a hairpin (blue and purple wavy line) recognizable by DROSHA-DGCR8 and subsequently processed by it and Dicer leading to miRNAs (blue wavy line) which are loaded into Ago. The resulting miRISC cleaves LINE-1 transcripts. Items are not depicted in scale. Adapted from references within text.*

1.1.3.3.3 Preintegrational control

At this point the cytidine deaminase apolipoprotein B mRNA editing complex (APOBEC) family and activation-induced cytidine deaminase (AID) proteins play an important inhibiting L1 retrotransposition as well as LTR-retrotransposons and retroviruses. In particular, APOBEC3 family (A-D, F, G and H) together with APOBEC1 and AIDs catalyse deamination of cytidine residues to deoxyuridines in the L1 cDNA that is being synthesized. As a corollary, the second strand will contain deoxyadenosines instead of deoxyguanosines and these mutations may inactivate new L1 copies [186]–[191]. Catalytically inactive APOBEC3 mutants also restrict L1 retrotransposition in a deamination-independent manner through a mechanism according to which in the cytoplasm those enzymes bind ORF1p not only complicating the RNP import into the nucleus, but also decreasing L1 RT activity [192]–[194].

SAMHD1 is an enzyme with deoxynucleoside triphosphate triphosphohydrolase activity and inhibits reverse transcription of retroviruses in non-dividing cells due to the depletion of the required dNTPs [195], [196]. Similarly, SAMHD1 also restricts L1 retrotransposition. In Aicardi-Goutières syndrome (AGS) SAMHD1 is mutated and, in consequence, retrotransposition rate increases in cell culture but surprisingly, AGS patients do not present more insertions when compared with controls. On the other hand, in diving cells SAMHD1 L1 inhibition is also observed and, in addition, a mutated SAMHD1 still inhibits L1 retrotransposition [197]–[199] so a different mechanism, not based on its deoxynucleoside triphosphate triphosphohydrolase activity, has to be involved. To answer this question, it has been proposed that SAMHD1 affects L1 retrotransposition either through inhibition of L1 ORF2p [197] or driving L1 RNP sequestration by stress granules [200].

The 3'-5' DNA exonuclease TREX1 degrades cDNA and has been related to downregulation of L1 retrotransposition since its overexpression inhibits L1 retrotransposition whereas in TREX1 mouse knockout cells L1 cDNA accumulates [201].

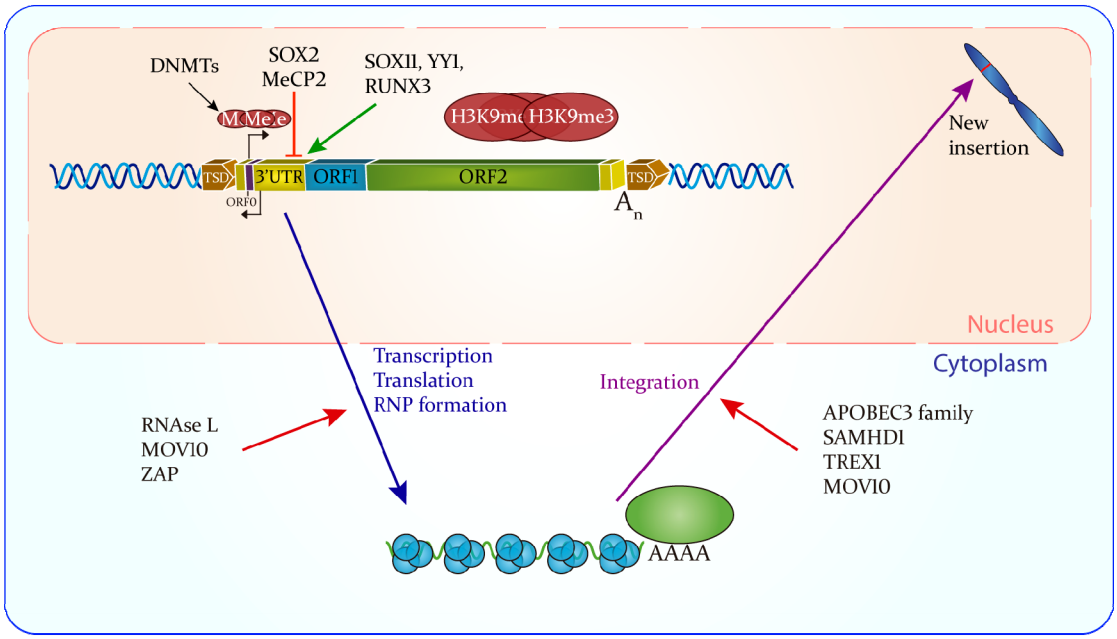


***Figure 6**. LINE-1 transcriptional, post-transcriptional (different from small silencing RNAs) and preintegrational regulatory mechanisms. Red arrows, bars or ellipses represent inhibitory processes, green arrow represents activating processes. Items are not depicted in scale. Adapted from references within text.*

1.1.3.4 LINE-1 cell-type and timing expression pattern

While it is true that LINE-1 elements are present in every human cell, its expression does not take place in all cell types. In light of L1 elements constitute around 17% of the human genome, its propagation have to be derived from the accumulation over generations of new heritable insertions, what could only happen if such insertions occur in the germline or embryonic cells. Indeed, there are many evidences of retrotransposition events taking place in both situations.

The first reported case was two unrelated patients with haemophilia A caused by an L1 insertion in exon 14 of the factor VIII gene. This mutation was present in none of their parents so the only possible explanation is that L1 can retrotranspose in germ cells or during embryo development [202]. Later investigations have stablished that TEs are able to retrotranspose not all along germline specification and embryo development but during some stages of them and in some somatic tissues.

1.1.3.4.1 Retrotransposition in the germline

Endogenous L1 mRNA and its ORF1 products are detected in mouse embryonic oocytes [203] as well as human oocytes support engineered L1

retrotransposition [204] coinciding with a short-term hypomethylation state [205]. On the other hand, primitive spermatogonia in mouse embryonic testis [203], leptotene and zygotene spermatocytes of postnatal mouse testis, round spermatids of the adult mouse testis [206] and human male germline [207] exhibit L1 expression, also demonstrated in transgenic mouse experiments [208].

1.1.3.4.2 Retrotransposition in the early embryo

Many studies have evidenced that L1 retrotransposition occurs during early development. Mouse blastocyst [209] and syncytiotrophoblast cells (placenta) [203] express both L1 mRNA and ORF1p, as well as mouse embryonic stem cells (mESCs) [210] and human embryonic stem cells (hESCs), which also support new L1 insertions [104], [167], [211]. On this matter, an outstanding proof of retrotransposition during early embryo was disclosed in a chroroidemia (a recessive X-linked disease, so it only affects males) patient with the CHM gene from one allele mutated by an L1 insertion and whose mother was a somatic and germline mosaic for the L1 insertion. Consequently, it was clearly demonstrated that retrotransposition had to occur during mother embryo development [212]. In view of the fact that human induced pluripotent stem cells (iPSCs) and mouse induced pluripotent stem cells (miPSCs) [167], [213], [214] also express L1 elements, these results taken together suggest that there is a need of L1 and other TE retrotransposition in maintaining pluripotency [55], [215], [216] and host cells must correctly balance TEs mobilization to avoid its potential genome damage and regulatory mechanisms that repress such process.

Keeping in mind the aforementioned findings, retrotransposition in the germline is less frequent, although insertions are heritable, whereas L1 mobilization in embryo is consistently higher leading to a somatic mosaicism [217].

1.1.3.4.3 Retrotransposition in adult tissues

L1 retrotransposition is normally repressed in somatic cells. For that reason, the relatively recent uncovered L1 mobilization in mammalian brain has been drawing an increasing attention. The first evidence was collected from rat neural progenitor cells (NPCs), which are multipotent cells able to undergo

differentiation into specific neuronal and glial cell types, using an engineered human L1 retrotransposon and also from the brain of transgenic mice *in vivo*, concluding that L1 can retrotranspose during early stages of neuronal differentiation, thus creating a somatic mosaicism within the brain [110]. A later research revealed that engineered L1 retrotransposition also occurs *in vitro* using human NPCs and more important, it demonstrated that hippocampus and other brain tissues accumulate endogenous L1 insertions when compared with other somatic tissues from the same patient [218]. Other studies corroborated human brain L1 retrotransposition employing single-neuron sequencing [199], [219], [220] together with Alu and SVA retrotransposons. These insertions seem to tend to accommodate within genes that are differentially expressed in the brain [221]. In an attempt to give a biological explanation to somatic retrotransposition in the brain, genomic mosaicism present in brain cells has been suggested to be the cause of neural plasticity and interindividual brain diversity even between monozygotic twins. Whether brain L1 retrotransposition is the cause of mental disorders remains unclear.

Despite the tight regulation of L1 retrotransposition in somatic cells a recent research showed that although normally inactive, somatic cells may express L1 elements in a higher level than previously thought depending on cell-type, locus and individual among other factors for each L1 element. To explain this, it has been hypothesized that L1 repression by host cell mechanisms might fail preventing L1 retrotransposition, hence a small subset of elements (or even one) could be impacting the cell genome in a low-frequency manner but in the long-term it can be catastrophic [222]–[224]. Supporting this idea, it has been stablished that two independent colorectal cancers were initiated by an L1 insertion in normal colon tissues that had mutated one allele of the adenomatous polyposis coli (APC) gene, a tumour suppressor, whereas the other allele is frequently altered by point mutations [225], [226]. On the other hand, notably important is the high expression observed in many tumour cells due to hypomethylation of a subset of L1 elements [227], thereby intensifying genomic instability what, in turn, may contribute to cancer progression and metastasis. The extent of the L1

retrotransposition impact causing tumorigenesis and its high expression in cancer will be discussed later.

1.1.3.5 The impact of LINE-1 retrotransposition

TEs-mediated retrotransposition events are impacting human genome, changing its functioning and structure in a relatively high rate. In humans, the L1 retrotransposition rate has been estimated as 1 for every 20-200 births (depending on the method used in the analysis) [228], [229] but more active seem to be Alu elements with approximately 1 out of 20 births [230]. SVAs are the less active since its retrotransposition rate has been established in 1 insertion for every 900 births [229]. Altogether, these *de novo* insertions mediated by L1, Alu and SVA elements are responsible for about the 0.27% of total mutations in humans [231]. Despite that and given that exon-coding and regulatory sequences just suppose a small proportion in human genomes, the chance of the new insertions to be introduced within those regions would be in theory highly improbable but, intriguingly, an important proportion of new insertions land in or close to genes. However, an L1 insertion *per se* is not enough to cause a disease unless the other allele exhibits a loss-of-function mutation (excepting sex chromosomes) and, even more, thanks to the low efficiency of L1 retrotransposition, most times the new L1 insertion is 5'-truncated (99% for L1 but just 16% for Alu whereas most SVA are full lenght) and just comprises a small fragment of the retroelement located in a region full of repeated sequences. Nonetheless, the harmful effects of TEs transposition have become patent throughout the last two decades either altering gene expression, causing structural variation or disease-causing mutations (reviewed in [8], [232]–[238]).

1.1.3.5.1 Structural effects on the host genome

Genomic DNA is subjected to target site deletions in the proximity of TEs insertion sites, ranging from 1 bp to ~130 kb in cultured cells but less frequently in whole organisms. EN independent retrotransposition is especially involved in this process since the majority of such events are characterized by deletions. These kind of deletions are known to cause disease like a case of pyruvate dehydrogenase complex deficiency originated by a 46 kb L1 insertion-mediated deletion in the

pyruvate dehydrogenase complex, component X (PDHX) gene [239] or a deletion of the human leukocyte antigen A mediated by an SVA insertion that led to leukemia [240].

TEs, as well as repetitive elements, are prone to undergo recombination between non-allelic homologous elements. As a result, distinct genomic rearrangements can be found such as deletions, inversions and duplications. Alu elements are responsible for the most part of recombination-mediated deletions in human genomes although those mediated by L1 are larger but, in both cases, it has been stablished the relationship between recombination-mediated deletions with the development of diseases. In a similar manner, the majority of the chromosomal inversions that have been detected in our genome have their origin in recombination processes involving L1 and Alu. Segmental duplications comprise around 5% of the human genome and share high similarity with Alu, reason why it is thought that recombination events between Alu elements have led to the accumulation of segmental duplications.

During L1 and SVA transcription, it is possible that RNA polymerase keeps working when it reaches polyA signal, therefore transcribing 3' flanking sequence until it finds another polyA downstream (3' transduction), and it is also possible, but less frequent, that transcription starts from a promoter upstream of the TE (5' transduction). In both situations, once retrotranscribed the new insertion will contain the 3' or 5' flanking sequence of the original location.

When retrotranscription is taking place, L1 enzymatic machinery can stop to transcribe other cellular mRNAs (template switch), principally U6 RNA, arising as U6/L1 chimeras in the new insertion.

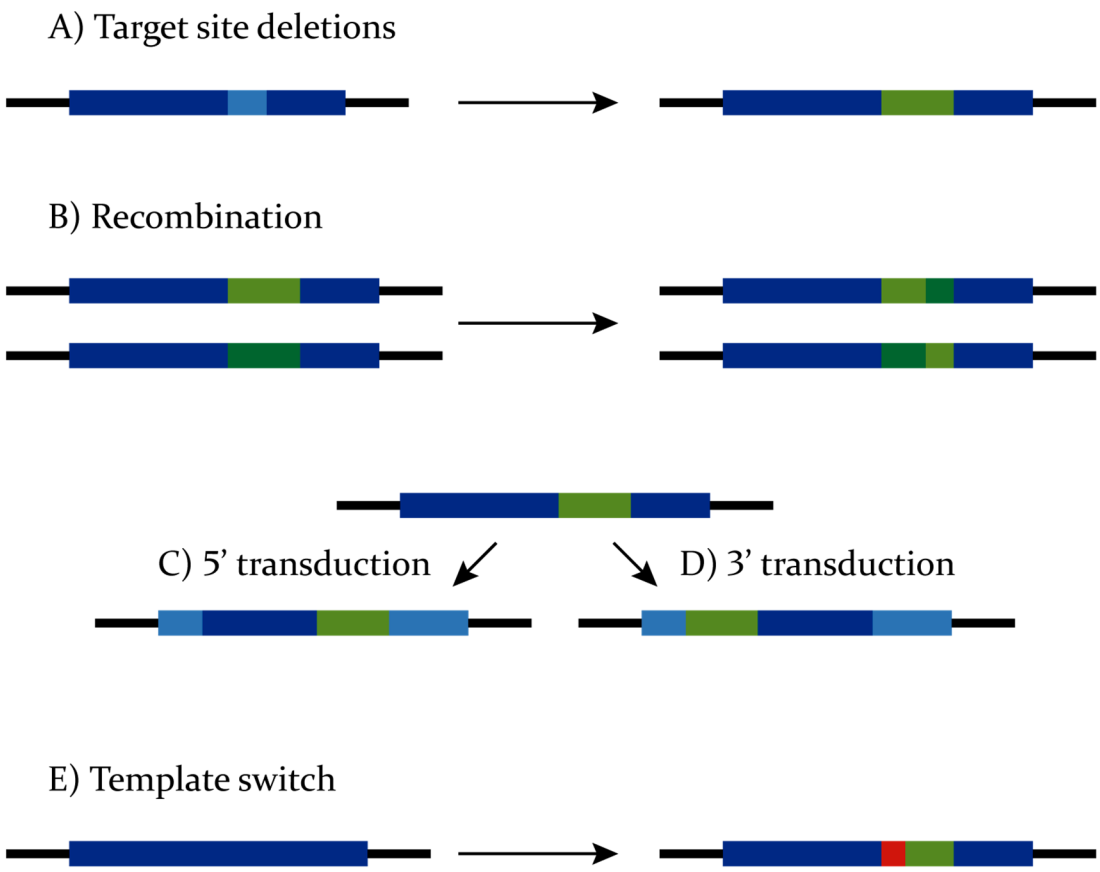


Figure 7. *Structural effects of LINE-1 retrotransposition. Green rectangles represent LINE-1 fragments or full copies; blue rectangles represent DNA sequences different from L1; red rectangle represents a retrotranscribed mRNA different from L1. Items are not depicted in scale. Adapted from references within text.*

The activity of ORF2p has been shown to generate DSBs [241], what suppose a source of genomic instability since they are highly mutagenic and tend to recombination. It is not completely known in what extent L1 EN-mediated DSBs impact human genomes as the experiment was carried out with much higher levels of ORF2p than in normal conditions although, in any case, their harmful effects are not contemptible.

1.1.3.5.2 Functional effects on the host genome

Newly inserted TEs can alter gene function in diverse ways. Disruption of coding exons by TE insertion usually leads to a loss-of-function as exemplified by the colon cancer developed upon insertion of a L1 element in the tumour suppressor gene APC [225], [226] or the haemophilia A case presented above

[202]. Because of their phenotypic consequences these are the best studied impacts of TEs mobilization and the easiest to be detected.

Either inserted in exonic or non-exonic regions, new transcripts of genes positioned upstream or downstream the new insertion might be generated from the sense and antisense promoters of L1 and produce the so called transcriptional interference what means that such promoter activity of TEs supresses or activates the transcription of nearby genes. Similarly, retrotranscribed elements can introduce stop codons or polyA signals into genes, thus producing premature termination of transcription.

In addition, splice acceptor and donor sites harboured by TEs can modify the splicing of the host gene, avoiding the inclusion of an exon (exon skipping) or including fragments of the TE in the mRNA what may conduct to exonization (i.e. the inclusion of an intronic or intergenic sequence into a gene that becomes exonic upon insertion) what might result in an evolutionary advantage. It can also alter the frameshift with the consequent loss of function.

Since new inserted TE copies are epigenetically repressed, the condensed heterochromatin formed with that purpose may also represses the transcription of nearby genes.

Considering the abovementioned effects, negative repercussions derived from TEs mobilization is clearly understood. However, in some cases it could be evolutionary beneficial for organisms if retrotransposition events drive to new proteins or isoforms of host genes. As an example, mouse Dicer1, whose role is to produce strictly needed small regulatory RNAs in some cell types, presents an intragenic insertion of an LTR-retrotransposon what cause the production of a truncated protein. This protein is transcribed from the ERV promoter and exhibits enhanced enzymatic activity when compared with the normal one due to the absence of an autoinhibitory helicase domain [242]. The 3' transduced exons or regulatory sequences during TEs mobilization can also arise new genes by the inclusion or duplication (exon shuffling) of them into or near to a gene so when it is transcribed and translated, the new inserted exon will form part of the protein

either inactivating it, increasing its activity or developing a new enzymatic role. It is the case of the acyl-malonyl condensing enzyme 1 (AMACl) in humans, which was formed through retrotransposition of SVA elements containing 3' flanking sequences [243].

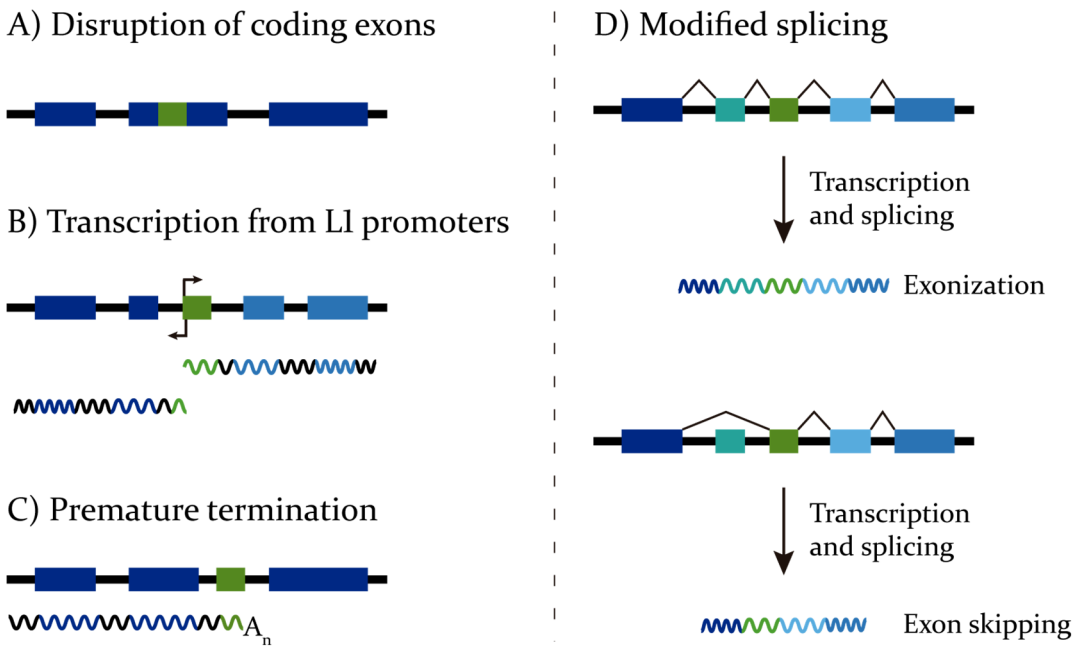


Figure 8. *Functional effects of LINE-1 retrotransposition. Green rectangles or wavy lines represent LINE-1 elements; blue rectangles and wavy lines represent sequences different from L1. Items are not depicted in scale. Adapted from references within text.*

1.1.3.6 LINE-1 and disease

The relationship between L1 retrotransposition and disease can be considered from two standpoints: on the one hand, L1 as a cause of disease and genome instability through insertional mutagenesis mainly during early embryo but also important when it happens in the germ line and non-cancer somatic cells; on the other hand, L1 as an overexpressed element in many types of cancer, thus potentially increasing the progression and metastasis ability of the tumor. However, it has not been elucidated the real extent of high L1 retrotransposition impact in cancer cells. Regarding the first issue, it is clearly difficult to therapeutically address disease-producing L1 insertions, but with respect to the second one it is more interesting since it is possible to face the problem of high L1 retrotransposition in human cancers. Disease-causing TEs mobilization within human genomes has been extensively reviewed [232], [233], [237], [244], [245].

1.1.3.6.1 LINE-1 as a cause of disease

Depending on the cell type, the consequences of L1 and non-LTR retrotransposition have different significances. If a new insertion take place in the germline, the offspring might inherit the new copy, carrying it in every cell of the organism. New insertions produced during early embryogenesis will give rise to a range of cell types derived from the one that had supported the retrotransposition event owning the insertion in the adult organism (e.g. if it happens in an endoderm cell, liver, stomach, pancreas, lungs among others will contain the insertion). In non-cancer somatic cells a new insertion just affect one cell. In light of these differences, it is obvious that most disease-causing L1 insertions have their origin in the germline or the early embryo development stages where it can be spread to a myriad of cells. It is also evident that retrotransposition in non-cancer somatic cells "just" can result in tumorigenesis either affecting the expression of tumor suppressor genes (TSGs) or proto-oncogenes.

To date, four tumors might be attributed to somatic LINE-1 retrotransposition. Two of them have been previously mentioned for being responsible for causing colorectal cancer after mutating the APC gene (TSG) [225], [226]. Suppression of tumorigenicity 18 (ST18) gene has been recently found mutated by an intronic L1 insertion, hindering its expression, in a case of liver cancer so it is thought to be the etiological cause [246]. Similarly recent is the finding of the phosphatase and tensin homolog (PTEN) gene, a TSG, with an L1 exonic insertion in an endometrial carcinoma which is considered to have driven the tumorigenesis [247]. Despite that, these studies are not extensively accepted as a proof of tumorigenesis driven by L1 retrotransposition and more information has to be sought in order to definitely determine whether L1 insertions can give rise to cancer because of two considerations: the precise timing of retrotransposition is practically impossible to define, since it could have happened right after tumor development, and it is not easy to determine whether a particular insertion in a certain TSG or proto-oncogene is enough to cause cancer.

On the other hand, it is possible that L1-mediated tumorigenesis is being underrated due to some factors. Firstly, there still are undiscovered pathways for

tumor triggering, therefore a somatic TE insertion could impact unknown TSGs or proto-oncogenes. In addition, a new insertion might change a host gene into a TSG or a proto-oncogene. In both cases the association between TE insertions and tumorigenesis could be overlooked.

Germline and early embryonic L1 insertions mean a much higher source of disease. A total of 123 cases of human diseases are attributed to TEs insertions, of which 29 belong to L1 retrotransposition and 94 cases through Alu (76), SVA (13) and other mRNAs (5) mobilization *in trans* (for a complete list see [233], [248]). Compared with gene regulatory regions, most part of the new insertions in these cases landed into coding exons what likely led to disease. The different TEs-mediated pathologies range from hereditary cancer and other hereditary diseases, leukaemia, X-linked diseases to haemophilia A and B, chronic pancreatitis, cystic fibrosis, retinitis pigmentosa, mental disorders, among many others.

Last, but not least, new studies link the L1 derepression during aging, caused by SIRT6 (a longevity regulating protein and a L1 repression factor) failure in senescent cells [249], with the activation of the type-I interferon (IFN-I) response which lead to inflammation, a hallmark of aging. Thus, it is suggested that L1 retrotransposition in senescent cells contributes to the maintenance of this state and, as a corollary, to age-related pathologies [250], [251].

1.1.3.6.2 LINE-1 in cancer

Much more relevance entails the high L1 activity observed in genomes of many human cancer cell types, what produces new insertions in an already unstable genome, thereby increasing the aggressiveness of the tumor. Epithelial cancers seem to be characterized by L1 retrotransposition, finding new insertions (combining L1 and non-autonomous non-LTR retrotransposons and other mRNAs) when compared with matched normal tissues. The majority of these insertions belong to 5' truncated L1 sequences and around 25% of them have 3' transduced sequences. Remarkably, most, if not all, retrotransposition events in tumors are due to a few source elements whose activity likely fluctuate during tumor development [252].

In 2010, a total of 9 insertions were found in 20 non-small lung cancer cell lines [253]. Since then, a growing number of studies have been stablishing the high L1 activity in human epithelial cancers: 72 insertions were found in 16 colorectal cell lines [254] but brain cancer cells did not seem to exhibit L1 mobilization; in another study 143 insertions were identified in 5 colorectal cell lines (one of them accumulating 106 insertions), 28 insertions in 7 prostate cancer cell lines and 22 insertions in 8 ovarian cancer cell lines whereas no L1 activity was detected in blood and brain cancer cells [255]; 12 insertions in 19 samples of liver cancer [246]; 83 insertions, all of them originated from the same L1 source, in 92 cases of colorectal cancer [256]; 695 insertions in 200 cancer samples coming from 11 tumor types, namely rectal adenocarcinoma (READ), colon adenocarcinoma (COAD), lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), ovarian carcinoma (OV), kidney clear cell carcinoma (KIRC), uterine corpus endometrioid carcinoma (UCEC), head and neck squamous cell carcinoma (HNSC), breast carcinoma (BRCA), acute myeloid leukaemia (LAML) and glioblastoma multiforme (GBM) of which lung, colon, head and neck and uterine gather the vast majority of such insertions whereas little to no L1 activity was observed in the rest cancer types [247]; 2756 insertions in 290 samples across 12 different tumors, concretely lung, colon, breast, bladder, bone, head and neck, renal, prostate, brain and pancreas cancers, mesothelioma and melanoma of which as previously, lung, colon, head and neck cancers are more prone to L1 retrotransposition [252]; 5108 insertions in 43 samples of esophageal cancer [257]; 465 insertions in 20 pancreatic ductal adenocarcinoma genomes [258]; 104 insertions in a total of 18 samples (4 colorectal cancer, 7 pancreatic carcinomas and 7 gastric cancers) [259]; 62 insertions in 9 samples of esophageal carcinoma [260].

Lung and colorectal cancers exhibit a higher L1 retrotransposition level than oesophageal, ovarian, uterine, head and neck, prostate, gastric and pancreatic cancers although the number of somatic insertions can considerably vary across samples. Lower L1 mobilization is detected in liver, kidney, bone, breast and testicular cancers. Surprisingly, no L1 insertions are found in brain cancers while some normal brain tissues present L1 activity.

L1 promoter hypomethylation is a hallmark of many cancer types as a consequence of the genome-wide hypomethylation in cancer [227] what, in turn, explains the raising of L1 activity in tumor cells as evidenced by many groups [225], [246], [252]–[254], [261]. Whilst L1 retrotransposition in human cancers (at least in epithelial tissues) has become a fact, the precise role of such impact in human tumors either driving tumorigenesis or destabilizing an already deregulated genome, or both, is still uncertain. In any case, L1 hypomethylation has been linked to cancer prognosis: genomic instability [262] and poor prognosis in lung [263], [264], oesophageal [265], [266], colon [267]–[269], breast [270], [271], hepatocellular [272]–[274], intrahepatic cholangiocarcinoma [275], oropharyngeal [276] and ovarian [277] cancers.

Hypomethylation of L1 promoter has been related to other negatives consequences in cancer such as lymph node metastasis in breast [278] and oesophageal cancers [266] and enhancement of the tumor promoter progression CDK6 (participates in the regulation of the cell cycle and promote proliferation) expression in oesophageal cancer [279]; activation of MET, RAB3IP, CHRM3 proto-oncogenes in colorectal cancer metastasis [280]; stimulation of proliferation and invasion in colorectal [281] and breast cancer [282] through activation of ETS-1, a transcription factor involved in cancer cell proliferation, metastasis and invasion.

Furthermore, it has been found that some tumor cell types are able to release microvesicles containing L1 and Alu transcripts (together with other DNA and RNA sequences), what may result in horizontal gene transfer to normal cells [283].

Epigenetic changes produced by new L1 insertions can activate oncogenes or inactivate tumor suppressor genes which is the case of the metastasis suppressor gene TFPI-2, whose expression is epigenetically silenced through the expression of an L1 chimeric transcript in breast and colon cancers [284].

1.1.3.7 How can LINE-1 retrotransposition be measured?

Multiple techniques can be applied to the study of LINE-1 biology. Here, and because of its relevance for this dissertation, the more important cell-based assay will be commented. The experiment was developed in 1996 by the Kazazian's group [116] and provide a quantitative method that can be used to measure L1 retrotransposition with different purposes. The assay has been modified to be used to elucidate the function of each L1 domain, to discover aspects of L1 biology, to study which genomic L1 copies are active, to develop animal models among many others.

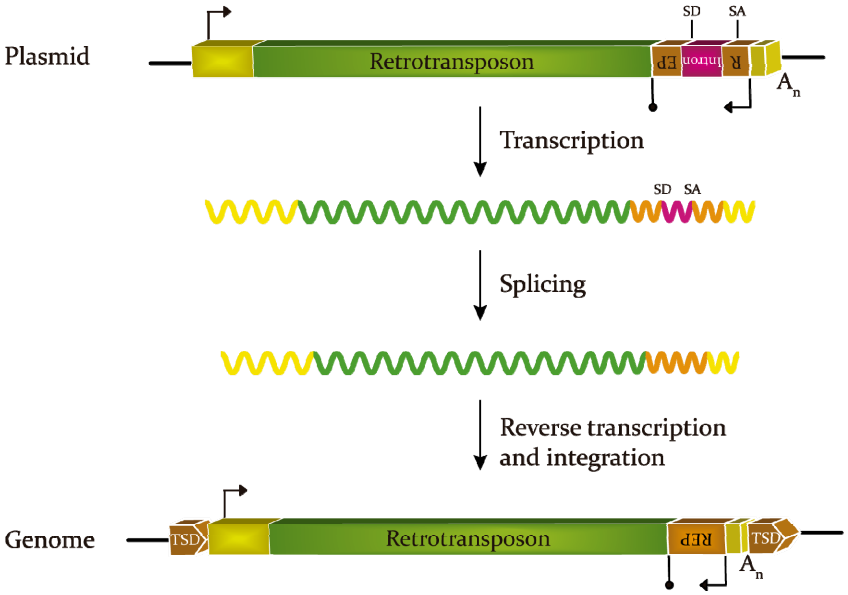


Figure 9. *Representative scheme of the retrotransposition assay. Into a retrotransposon contained in an expression vector has been previously cloned a reporter gene (in antisense orientation, orange boxes labelled R and EP) that includes an intron (pink box) in the transcriptional sense of the retrotransposon promoter (left black arrow). Upon transcription from this promoter, splicing and reverse transcription and genomic integration processes, the intron has been removed and the reporter gene (orange box) is allowed to be expressed from its own promoter (face down right black arrow). Lollipops indicate transcriptional terminator, yellow boxes depict regulatory sequences (UTR or LTR), and coloured wavy lines represent mRNA. Abbreviations: SD, splice donor; SA, splice acceptor; TSD, target site duplication, A_n, polyA signal.*

In brief, this assay consists in the transfection of a plasmid containing an active L1 copy tagged with a reporter gene (REP) which can only be expressed after a retrotransposition event (**Figure 9**). Originally, the transfected plasmid contained a RC-L1 copy and the neomycin resistance gene (*mneoI*) [285] but it has evolved to use different retrotransposons tagged with different reporter genes (e.g. firefly luciferase [286], blasticidin resistance gene [151], [287], green fluorescent

protein [288]). More detailed, *mneoI* was cloned in the antisense orientation into the 3'UTR of an active L1 copy. The reporter gene has its own promoter and polyA signal but it is interrupted by intron 2 of the human γ-globin gene in the sense orientation what avoids its translation into a functional protein. The retrotransposon tagged with the REP was cloned into pCEP4 vector in which transcription is controlled by CMV promoter to ensure an efficient and high-level expression. Once the transcription from the CMV promoter ends, the mRNA is spliced (removing the intron), reverse transcribed and integrated into the cell genome. Now, since REP lacks the intron, upon transcription from REP promoter its gene product can be generated. Importantly, despite of the high frequency of 5'-truncation during L1 retrotransposition, the REP is practically always retrotranscribed since retrotransposition process starts from the 3'UTR region, so results are generally not compromised.

Summarizing, if retrotransposition takes place the REP gene product will be properly synthesized and transfected cells will become resistant to neomycin whereas in the absence of retrotransposition the cells will remain sensitive to neomycin.

1.1.3.8 Inhibiting LINE-1 retrotransposition using small-molecules inhibitors

Some biochemical and cell-based assays have been carried out using small-molecules inhibitors, transcription factors, small silencing RNAs, among others, attempting to repress L1 retrotransposition through inhibition of L1 reverse transcriptase. Based on one of the therapeutic approaches used in the treatment of HIV patients, nucleoside reverse-transcriptase inhibitors (NRTIs) and non-nucleoside reverse-transcriptase inhibitors (NNRTIs) have also been used to inhibit L1 RT. HIV replicates through reverse transcription and integration of its single-stranded RNA. During HIV reverse transcription, previously triphosphorilated NRTIs by cell enzymes can be incorporated into the nascent strand, instead of the endogenous ligand and, due to the lack of a free 3'-OH, the RT is not able to successively add nucleosides and, as a consequence, reverse

transcription stops. In contrast, NNRTIs bind to an allosteric site of the HIV RT, thus changing its conformation what prevent the RT enzymatic activity.

To the best of my knowledge, the NNRTI nevirapine was the first small-molecule inhibitor found to repress L1 retrotransposition. In 2003 researchers reported that whole cell extracts from some human and murine cancer cell lines were able to reverse transcribe an RNA template but in the presence of nevirapine cDNA synthesis was inhibited, demonstrating the L1 RT activity inhibition by nevirapine. In addition, in the presence of nevirapine a set of human and mouse progenitor and tumorigenic cell lines exhibited a decrease in the proliferation rate compared with untreated cells. This effect was discarded to be produced by apoptosis or non-specific cell death but because of cell cycle arrest. Furthermore, nevirapine seemed to promote differentiation in the two progenitor cell lines assayed. Despite all these interesting results, researchers did not use a system that clearly demonstrated that the observed effects in cells were derived from L1 RT inhibition [289]. The same research group published in 2005 another research article in which efavirenz (NNRTI) was assayed, in addition to nevirapine, showing similar results regarding proliferation and differentiation using tumor cell lines distinct from the previously employed. Moreover, they also showed that both NNRTIs reduced the volume of human tumor xenografts in mice. The same results were obtained when an siRNA system targeting L1 was performed what supports the hypothesis that the effects on tumor cells are mediated by L1 RT inhibition although a direct effect was again not completely demonstrated. In addition no experiments were developed to prove the ability of efavirenz to inhibit L1 RT activity [290].

The first report (2008) of L1 RT inhibition in cell culture was carried out using a plasmid containing an L1 element tagged with GFP, as previously depicted, to test the inhibitory abilities of lamivudine, zidovudine, stavudine, tenofovir (NRTIs) and nevirapine. With this reporter system, GFP expression is inversely proportional to the degree of inhibition. All compounds showed inhibitory activity with the exception of nevirapine in contrast with previous data [291].

In 2010 a research group evaluated abacavir (NRTI) in prostate cancer cell lines exhibiting high expression levels of L1. The drug produced a reduction in cell growth due to cell cycle arrest what led to senescence, inhibited migration and invasion of tumor cells, altered gene expression patterns and upregulated L1 mRNA expression. However no experiments were developed to prove the ability of abacavir to inhibit L1 RT activity [292].

The RT inhibitory effect of four triphosphate NRTIs (AZT, stavudine, zalcitabine and lamivudine) and three NNRTIs (efavirenz, nevirapine and delavirdine) was tested in 2011 in an enzymatic assay as well as in cell culture assay transfecting cells with a plasmid containing an RC-L1 element tagged with the *mneoI* gene. Authors concluded that NRTIs are more efficient RT inhibitors than NNRTIs (nevirapine had no effect on L1 reversetranscription inhibition) [293].

More recently, lamivudine and stavudine have been used to treat SIRT6 deficient mice and aged mice, which accumulate L1 cDNA within the cytoplasm triggering the IFN-I response. Treatment resulted in a reduction of the sterile inflammation and increase of the lifespan of those mice, suggesting L1 RT as a target in aging-associated pathologies [250], [251]. Similarly, IFN-1 response is triggered in Aicardi-Goutières syndrome (an autoimmune disease that recognizes endogenous nucleic acids as nonself) because of the accumulation of L1 cDNA. A clinical trial in human patients to whom NRTis were administrated showed an amelioration of the symptoms of such disease [294].

Considering all the data aforementioned, it is clear that some NRTIs and NNRTIs are good inhibitors of L1 RT activity, thereby repressing retrotransposition either in enzymatic or in cell culture assays. These drugs seem to reduce the cell growth of some cancer cell lines that manifest L1 expression among other effects such as promoting cell differentiation. Whether these observed effects are owed to inhibition of L1 retrotransposition remains unknown. Some NNRTIs do not inhibit L1 RT but possess anticancer activity what probably indicates the presence of an underlying mechanism not related with L1. Likewise, the effects obtained by NRTIs treatment of cancer cells might not be mediated by L1 inhibition but through the other mechanisms in which nucleosides act as antimetabolites.

Human genome was thought to be a static, an unchangeable genetic storage system that allows organisms to preserve such important information. DNA repeats were considered as "junk" or parasites as well as TEs when they were discovered. Nowadays, however, we now know that genomes are subjected to the random "bombardment" of transposable elements, which are a source of structural variation, genomic instability and disease. It has also changed our notion that every cell in an organism shares the same DNA sequence. The role of TEs in living organisms is often overlooked, but it is widely clear their impact causing diseases and worsening those cancer characterized by high L1 expression levels as well as its deregulation (usually derepression) in some pathologies and aging. Furthermore, some reports indicate that it could be stablished a relationship between L1 promoter hypomethylation and patient prognosis in cancer. Additionally, the L1 expression pattern is really intriguing, ¿why does LINE-1 retrotranspose in selected tissues? ¿has it any role in living organisms or it is a mere parasite?

In light of all of this, it becomes necessary to develop a non-toxic therapeutic tool to selectively target LINE-1 reverse transcriptase (what represent the general objective of the present dissertation) with the aim of seeking the role of L1 retrotransposition and also a potential new therapeutic strategy for the treatment of cancer and other pathologies that show high levels of LINE-1 expression. Given the similarities between HIV and LINE-1 reversetranscription through a reverse transcriptase enzyme, nucleoside reverse transcriptase inhibitors seem to be the best choice for a first approach, as previously demonstrated. Because of that, the next sections are going to describe the nucleoside features, their mechanisms of action and their synthesis.

1.2 INHIBIDORES DE LA TRANSCRIPTASA INVERSA DE VIH. CLASIFICACIÓN, MECHANISM OF ACTION Y SÍNTESIS DE NUCLEÓSIDOS

1.2.1 VIH e inhibidores de la reverso transcriptasa

El virus de la inmunodeficiencia humana (VIH) parece haber surgido alrededor de 1920 en lo que actualmente es la República Democrática del Congo a través de la infección y adaptación a su nuevo hospedador del virus de inmunodeficiencia de simios (VIS) [295]. Sin embargo, hasta mediados o finales de la década de 1970, cuando se empezaron a observar un elevado número de casos en EE.UU. de lo que posteriormente se llamaría síndrome de inmunodeficiencia adquirida (SIDA), no empezó a llamar la atención internacional. Fruto de una intensa investigación se identificó en 1983 el retrovirus VIH como causa etiológica de la epidemia de SIDA [296]–[301]. En 1987 la Food and Drug Administration (FDA) aprobó el AZT, que había sido desarrollado anteriormente con la intención de ser usado como anticancerígeno [302], como primer fármaco para el tratamiento del SIDA, enfermedad que según los datos registrados se ha cobrado unos 35 millones de vidas desde el inicio de la epidemia y que a día de hoy presentan unos 37 millones de personas [303]. Como consecuencia de una extensa investigación se han sintetizado una amplísima variedad de estructuras químicas con la intención de impedir la replicación del virus que tradicionalmente son clasificados en (ampliamente revisado en [304]):

-Inhibidores de la proteasa (PIs): inhiben la enzima encargada de hidrolizar las largas cadenas polipeptídicas codificadas por los genes del virus en proteínas funcionales. Fármacos actualmente aprobados por la FDA: atazanavir, duranavir, fosamprenavir, ritonavir, saquinavir y tipranavir.

-Inhibidores de la fusión (FIs): impiden la fusión de la envoltura vírica con la membrana celular de los linfocitos CD4, evitando de este modo la

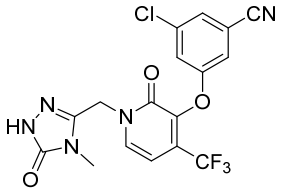
infección de dichas células. Fármacos actualmente aprobados por la FDA: enfuvirtida.

-Inhibidores de la transcriptasa inversa no análogos de nucleósidos (NNRTIs).

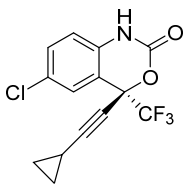
-Inhibidores de la transcriptasa inversa análogos de nucleósidos y nucleótidos (NRTIs).

1.2.1.1 Inhibidores de la transcriptasa inversa no análogos de nucleósidos (NNRTIs)

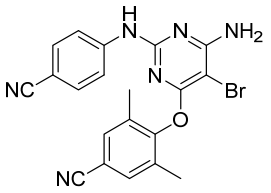
Son compuestos liposolubles que se comportan como inhibidores no competitivos. Se unen al sitio alostérico (localizado a unos 10-15 Å del sitio activo) de la reversotranscriptasa de VIH-1 [305]–[307] provocando con esta unión un cambio conformacional en la enzima y con ello una pérdida de la actividad catalítica de su centro activo. A diferencia de los NRTIs, los NNRTIs no precisan de otras enzimas para ser activos (no necesitan ser fosforilados) pero por otro lado no son activos frente a RT de VIH-2 [308] al no estar conservada la región correspondiente al sitio alostérico al comparar la estructura con la RT de VIH-1 [309]. Dado que los NNRTIs no guardan, en general, una relación estructural con ligandos endógenos, las reacciones adversas derivadas del uso de estos fármacos son menores que las comparadas con los NRTIs. Fármacos actualmente aprobados por la FDA: doravirina, efavirenz, etravirina, nevirapina y rilpivirina (**Figura 10**).



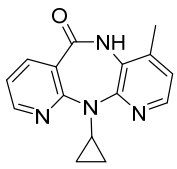
Doravirina (DOR)



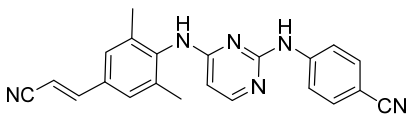
Efavirenz (EFV)



Etravirina (ETR)



Nevirapina (NVP)



Rilpivirina (RPV)

Figura 10. *Fármacos NNRTIs permitidos por la FDA.*

1.2.1.2 Inhibidores de la transcriptasa inversa análogos de nucleósidos y nucleótidos (NRTIs)

Todos los compuestos comercializados pertenecientes a este grupo tienen en común una estructura de 2',3'-didesoxinucleósidos. Estas moléculas hidrófilas acceden a la célula mediante transportadores de membrana específicos de nucleósidos (revisado en [310]) pero para ser activos, estos compuestos han de ser fosforilados por enzimas celulares (nucleósido kinasas, nucleósido monofosfato kinasa y nucleósido difosfato kinasa) para formar los correspondientes 2',3'-didesoxinucleótidos, motivo por el que todos han de ser considerados profármacos [311], [312]. Estos se unen al sitio activo de la reversotranscriptasa (RT) de forma similar a como lo hacen los nucleótidos naturales, siendo los derivados de pirimidina los más estudiados. El dominio catalítico está formado por los D110, D185 y D186 (D185 y D186 forman parte del motivo YXDD, altamente conservado en reversotranscriptasas de virus) coordinados con dos cationes Mg^{2+} a través de los grupos carboxilo de sus cadenas laterales y que a su vez se coordinan con los oxígenos de los grupos fosfatos del nucleótido a añadir. Los residuos V111, D113 y A114 también se coordinan de igual manera, aunque no forman parte del dominio catalítico y en función del nucleótido pueden no participar en esta unión. A los grupos fosfatos se unen también los grupos amino de las cadenas laterales de la K65 y la R72 (β y γ-fosfatos respectivamente). La Y115 forma un puente de hidrógeno con el hidroxilo 3' (en caso de los nucleótidos naturales) o con el grupo azida en el AZT. Los grupos metileno de la Q151 provocan interacciones hidrofóbicas con el C2' de la desoxi-D-ribosa. Estas interacciones pueden variar ligeramente en función del nucleótido sintético empleado. La colocación del nucleótido está además asistida por interacciones de apilamiento con la base precedente en la nueva hebra y con los puentes de hidrógeno que forma con la base en la hebra opuesta [305], [313]–[317].

Dado que se unen al sitio activo de la enzima actúan inhibiéndola al competir con los ligandos naturales endógenos. También producen una acumulación de mutaciones en la progenie viral, al introducir los nucleótidos no

naturales, reduciéndose así la capacidad de infección. Sin embargo, su principal papel es el de actuar como terminadores de cadena ya que al carecer en su estructura de un grupo 3'-OH la enzima RT es incapaz de añadir el siguiente nucleótido (no puede formarse el enlace 3'-5' fosfodiester) durante el proceso de integración del material genético del virus en el genoma del hospedador [311], [318].

La similitud estructural de los NRTIs con ligandos naturales implica que en función de la dosis pueden interaccionar con diversas enzimas celulares esenciales tales como ADN y ARN polimerasas, kinasas, ADN metiltransferasas, reductasa de ribonucleótido, nucleosidasas y timidilato sintasa. Producto en gran parte de estas interacciones son las reacciones adversas padecidas durante el tratamiento con NRTIs [311]. Fármacos actualmente aprobados por la FDA: abacavir, emtricitabina, lamivudina, fumarato de disoproxilo de tenofovir, zidovudina (**Figura 11**).

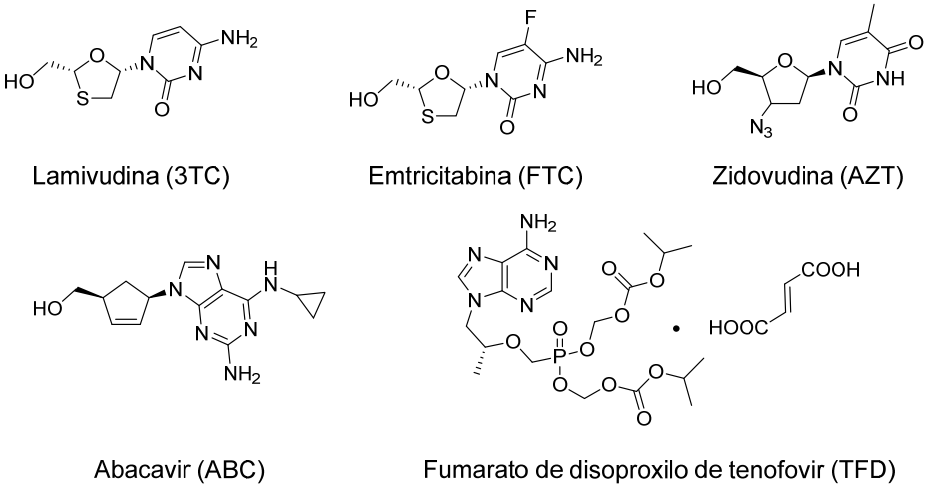


Figura 11. *Fármacos NRTIs permitidos por la FDA.*

Relaciones generales estructura-actividad de los NRTIs

La aparición del VIH llevó a la comunidad científica a hacer un gran esfuerzo en el desarrollo de nuevos fármacos para luchar contra la enfermedad, dentro de los cuales los nucleósidos han recibido buena parte de la atención. Por ello se ha alcanzado una gran variabilidad estructural en este tipo de compuestos y muchas estructuras diferentes poseen actividades similares frente a VIH, si bien varían dependiendo de la línea celular utilizada para su evaluación. Debido a lo anterior, el establecimiento de unas relaciones estructura-actividad de forma general no es posible y se han establecido para familias de compuestos como se comentará

brevemente a continuación. En general, los potenciales compuestos activos serían aquellos con configuración β (tanto serie D como L), si bien hay casos de α-nucleósidos que también presentan actividad anti-VIH incluso mayor que sus correspondientes β-isómeros, siendo prácticamente todos los casos nucleósidos derivados de oxatiolano o dioxolano. El efecto de algunos α-nucleósidos podría deberse a que en el citoplasma de las células las bases nitrogenadas de estos nucleósidos pueden ser escindidas de la fracción derivada del anillo de furanosa y actuar como base nitrogenada o bien ser unido, mediante el metabolismo celular, a un monosacárido endógeno y actuar como un nucleósido. Por otro lado, no es descartable que los α-nucleósidos puedan interaccionar con diversas enzimas relacionadas con el metabolismo de los ácidos nucleicos, bloqueando de esta forma dichas rutas enzimáticas. Dos aspectos comunes a todos los NRTIs es la necesidad de carecer de un grupo 3'-OH (como se ha explicado anteriormente) y la de poseer un grupo 5'-OH ya que han de ser fosforilados para poder ser activos, de modo que las sustituciones de dicho grupo por grupos azido o amino conducen a una pérdida total de la actividad antiretroviral [319]. Salvo las aquí comentadas, no se han establecido relaciones estructura-actividad definitivas para este grupo de fármacos y los datos al respecto solo son comparables dentro de un mismo grupo estructural. Por este motivo, las relaciones estructura-actividad específicas de cada familia de nucleósidos serán comentadas cuando dichas familias sean presentadas.

1.2.2 Otras aplicaciones terapéuticas de análogos de nucleósidos y nucleótidos

Además de usarse los análogos de bases nitrogenadas, nucleósidos y nucleótidos en el tratamiento de distintas enfermedades víricas, a estos compuestos se les han encontrado otras aplicaciones terapéuticas.

Pueden encontrarse fármacos derivados de nucleósidos, nucleótidos y bases nitrogenadas usados para un variado espectro de patologías (revisados en [311], **Figura 12**). El alopurinol, derivado de la hipoxantina, actúa inhibiendo la xantina oxidasa y es utilizado para tratar la hiperuricemia. La azatioprina (profármaco derivado de la 6-mercaptopurina) y la cladribina se usan como inmunsupresores.

La teofilina, análogo de la metilxantina, se usa por su actividad inhibidora de fosfodiesterasas en el tratamiento de asma y enfermedad pulmonar obstructiva crónica. La decitabina y azacitidina son empleadas en el tratamiento del síndrome mielodisplásico al actuar como moduladores epigenéticos mediante la inhibición de metiltransferasas. Recientemente se ha demostrados que los NRTIs tienen propiedades antiinflamatorias que podrían ser utilizadas clínicamente [320] o en el tratamiento del síndrome de Aicardi-Goutières [321].

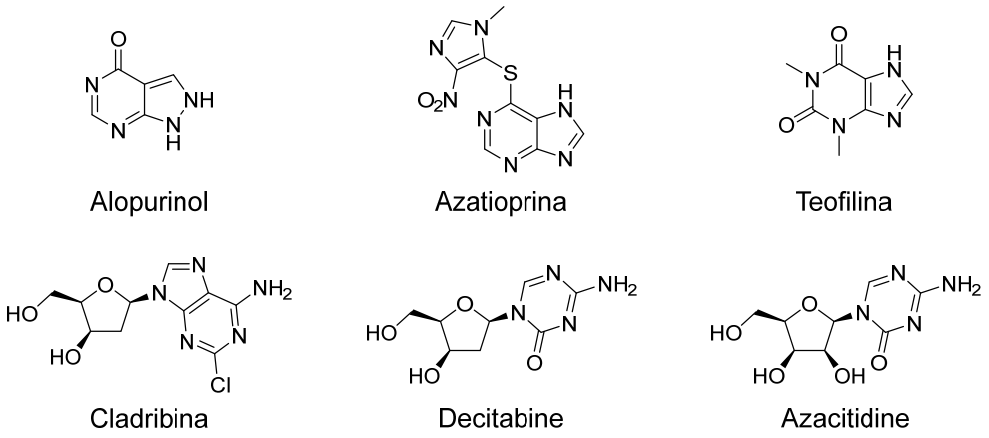


Figura 12. *Estructuras de fármacos derivados de nucleósidos y bases nitrogenadas.*

1.2.2.1 Cáncer

Nucleósidos con diversas estructuras han sido utilizados en el tratamiento del cáncer desde hace varias décadas. De hecho, como se ha mencionado anteriormente, la síntesis del AZT buscaba encontrar un fármaco anticancerígeno. La utilidad de los derivados de nucleósidos en el tratamiento de esta patología radica en su capacidad citotóxica al actuar como antimetabolitos en la síntesis de los ácidos nucleicos. Las células cancerígenas se caracterizan, entre otros aspectos, por su alta tasa de división. Con cada división, el material genético de la célula ha de ser duplicado por lo que hay una alta actividad enzimática relacionada con la síntesis de ácidos nucleicos fruto de una sobreexpresión, en células cancerígenas, de los genes responsables de dicha ruta metabólica.

Los análogos de nucleósidos son incorporados en estas reacciones enzimáticas actuando como inhibidores competitivos y reduciendo así la eficacia de la síntesis de ácidos nucleicos endógenos. Por otro lado, los análogos de

nucleósidos trifosfato interfieren con ADN polimerasas, pudiendo incorporarse a las nuevas hebras durante el proceso de replicación genómica, deteniendo el proceso y activando los sistemas celulares de reparación del ADN. También pueden ser utilizados por las enzimas que reparan daños en el ADN como nucleótidos endógenos. Cuando cualquiera de las dos situaciones ocurre, el proceso de división celular se bloquea mientras se intenta reparar el daño. Si la maquinaria de reparación es incapaz de revertir la situación, la célula entra normalmente en apoptosis.

Las células somáticas no cancerígenas presentan un metabolismo de ácidos nucleicos mucho menor comparadas con las cancerígenas por lo que estas terapias implican cierto grado de selectividad aunque no por ello están exentas de reacciones secundarias al tratamiento (revisado en [311], [322], [323]).

1.2.3 Características estructurales de los nucleósidos: conformación y configuración

Dejando a un lado las modificaciones que pueden realizarse tanto en la base nitrogenada como en el anillo de furanosa (serán comentadas más adelante) de los nucleósidos, tanto la conformación como la configuración de los centros quirales presentes en la molécula tienen un impacto en su actividad biológica. Estas características se han estudiado principalmente para los nucleósidos [319], [324]–[328] pero pueden extrapolarse a los nucleótidos dado que la fosforilación, en general, no afecta a la conformación de estos compuestos [327], [329]. Se utilizará durante el presente trabajo la numeración estándar para nucleósidos (**Figura 13**). En el caso en el que la furanosa no estuviera unida a la base nitrogenada, la numeración correcta sería C1, C2, etc., pero para evitar confusiones se utilizará C1', C2', etc., aunque el derivado de furanosa no esté formando un nucleósido.

Figura 13. *Numeración empleada para designar las posiciones en nucleósidos pirimidínicos (izquiera) y púricos (derecha).*

1.2.3.1 Configuración absoluta y relativa

La notación D-L es un sistema anterior a las reglas de Cahn-Ingold-Prelog (*R* y *S,* actualmente vigentes) para describir dos enantiómeros cuando aún no podía determinarse la configuración absoluta de los estereocentros. Sin embargo, su uso se ha mantenido para el caso de los monosacáridos (y por extensión a los nucleósidos) y aminoácidos [330], [331] por dos motivos:

-Su brevedad sin pérdida de información: D-ribosa en lugar de (2*R*,3*R*,4*S*,5*R*)-5-(hidroximetil)tetrahidrofuran-2,3,4-triol.

-Permite diferenciar claramente aquellos monosacáridos (D) y aminoácidos (L) presentes en la naturaleza de sus enantiómeros (que no lo están). El uso de la notación (*R*, *S*) complica claramente una identificación rápida: en todos los ribonucleósidos de la serie D, el C4' tiene la misma configuración, pero si se usa el sistema (*R*, *S*) la notación para la configuración del C4' es (*R*) para los 2'-desoxinucleósidos y (*S*) para los 2',3'-didesoxinucleósidos (**Figura 14**).

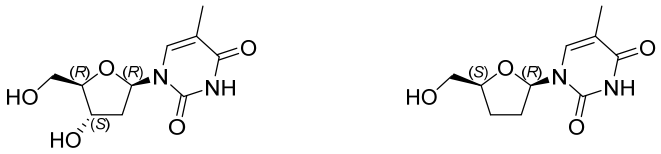


Figura 14. *Configuraciones absolutas de los estereocentros de los nucleósidos presentes en la naturaleza (izquierda) y de 2',3'-didesoxinucleósidos (derecha).*

Por otro lado, para denotar la configuración del C1', si bien puede usarse la notación (*R*, *S*), se suele emplear α y β para describir la configuración relativa de dicho carbono con respecto al C4'. De esta forma se denominan α- y β-nucleósidos a aquellos en los que la base nitrogenada queda en el plano contrario (*trans*) o en el mismo (*cis*), respectivamente, con respecto al grupo hidroximetilo. De esta forma, la configuración del C4' junto con la del C1' da lugar a una amplia gama de compuestos nucleosídicos (**Figure 15**). Los nucleósidos se encuentran en la naturaleza como β-D-nucleósidos normalmente, aunque se han encontrado α-D-nucleósidos como por ejemplo los que forman parte de las vitaminas del grupo B12 [332] y algunos poseen actividad biológica [333], [334].

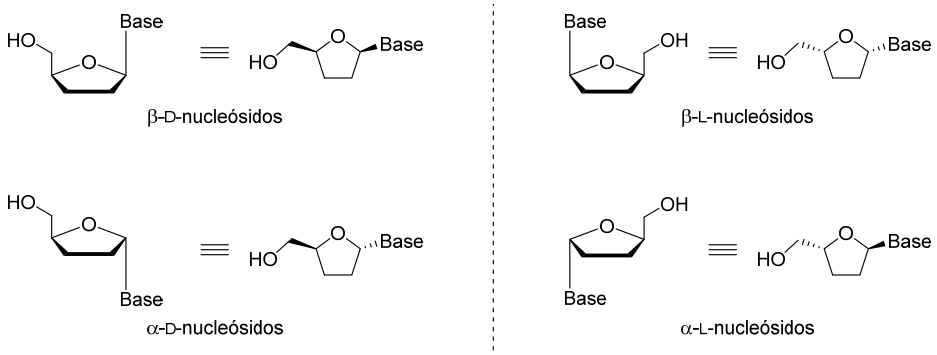


Figura 15. *Configuraciones relativas de los D- y L-2',3'-didesoxinucleósidos.*

1.2.3.2 Conformación de los nucleósidos

Ángulos de los enlaces

La unión entre el C1' de la furanosa y la base nitrogenada (N1, N3 o un carbono) es un enlace simple y por tanto tiene libertad de giro, de igual forma que ocurre con el enlace C4'-C5'. El ángulo del primero se denomina *chi* (χ) y el del segundo *gamma* (γ). Aunque todas coexisten, debido a factores esteroelectrónicos algunas conformaciones son predominantes con respecto a otras.

En función del ángulo de giro del enlace glicosídico puede hablarse de dos conformaciones principales: *anti* cuando el H6 en las bases pirimidínicas o el H8 en las bases púricas queda sobre el anillo de furanosa, y *syn* si dichos átomos quedan orientados en sentido contrario a la furanosa **(Figure 16)** Los α-nucleósidos adoptan predominantemente la conformación *anti*. En el caso de los β-nucleósidos, si bien la conformación *anti* también es la que adquieren de forma preferente, los derivados de purina presentan en un rango de giro sobre el enlace χ más amplio que los de pirimidinas. En la naturaleza, únicamente se ha encontrado la guanosina en conformación *syn* formando parte de la estructura del ADN-Z [334]–[338].

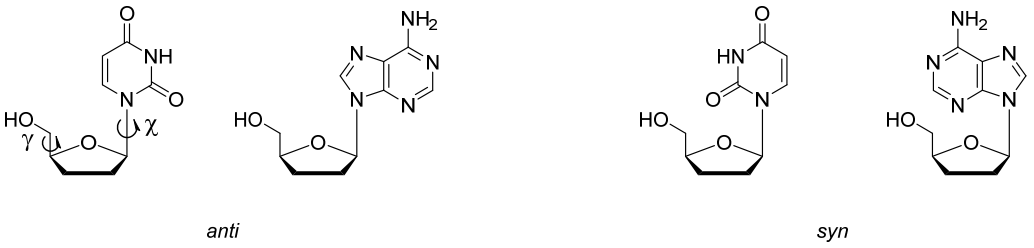


Figura 16. *Conformaciones syn y anti de los 2',3'-didesoxinucleósidos en función del ángulo de enlace C1'-N1 (χ).*

Con respecto al enlace C4'-C5', de las distintas conformaciones posibles permitidas por la libertad de giro de este enlace, la más habitual en todo tipo de nucleósidos es la denominada *gauche-gauche*, aquella en la que el grupo hidroxilo queda sobre la base de furanosa (**Figure 17**) [334]–[338].

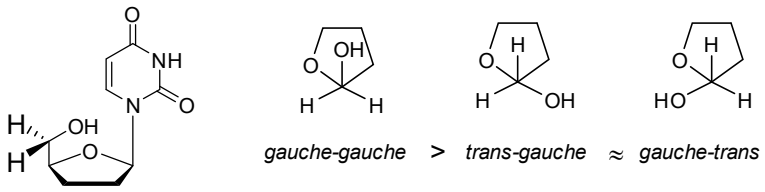


Figura 17. *Conformaciones de los 2',3'-didesoxinucleósidos en función del ángulo de enlace C4'-C5' (γ).*

La mayor abundancia de las conformaciones *anti* y *gauche-gauche* en nucleósidos se ha atribuido a los distintos puentes de hidrógeno intramoleculares que son capaces de formar cuando adquieren dichas conformaciones. Aunque el grupo carbonilo (C2) de la base nitrogenada podría formar un enlace de hidrógeno con el grupo hidroxilo (C5') de la furanosa, parece que las fuerzas de repulsión son predominantes dando lugar a una población mayor de moléculas con la conformación *anti* y en los casos en los que el nucleósido tiene un grupo hidroxilo en la posición C3', este forma un enlace de hidrógeno con el carbonilo en C2 [336]. Sin embargo, es el enlace de hidrógeno que se forma entre el H6 en bases pirimidínicas o el H8 en bases púricas y el O5' (C-H···O) el que determina la mayor proporción de conformaciones *anti* y *gauche-gauche*. Además, pueden establecerse enlaces entre los hidrógenos de la base nitrogenada (H6 o H8) y el O4' y de este a su vez con el hidroxilo en C5' (**Figura 18**). Este tipo de interacciones [339], aunque débiles, se ha demostrado con estudios de rayos X [340]–[343] y estudios computacionales [344]–[356].

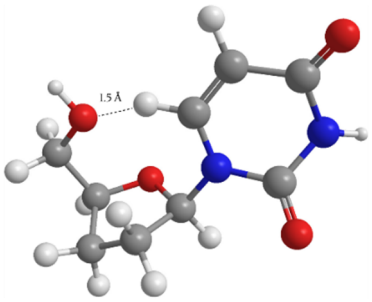


Figura 18. *Estructura 3D de la conformación más estable de 2',3'-didesoxiuridina. Las bolas rojas representan oxígeno y las azules nitrógeno.*

Conformación del anillo de furanosa

El anillo de furanosa, lejos de ser totalmente plano, suele presentar uno o más átomos fuera del plano, hecho que se debe a efectos estereoelectrónicos. Si un átomo se encuentra desplazado hacia donde se encuentran la base y el C5' (en β-nucleósidos) la conformación que posee es denominada *endo* mientras que en la dirección contraria *exo* (**Figura 19**). El C2' y el C3' son los que, en general, se encuentran desplazados con respecto al plano de la furanosa y normalmente lo hacen en sentidos opuestos. En los β-nucleósidos, si la base está en conformación *anti*, existe una preferencia por el C2' a adquirir la conformación *endo* y el C3'-*exo*. Esta es por tanto la situación más habitual. Si la rotación del enlace tiende a la conformación *syn* disminuye la abundancia de C2'-*endo* a la vez que aumenta la de C3'-*endo* y con ello la C2'-*exo*. En los α-nucleósidos los efectos estereoelectrónicos son menores y por tanto las conformaciones son más intercambiables [334], [336], [357]. Como curiosidad, el ADN-A contiene nucleótidos con la conformación C3'-*endo,* mientras que en el ADN-B es la conformación C2'-*endo* [358].

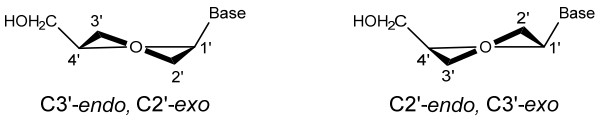


Figura 19. *Conformaciones mayoritarias del anillo de furanosa en 2',3'-didesoxinucleósidos.*

Tautomerismo en nucleósidos

Las bases nitrogenadas, en solución acuosa, se encuentran en un equilibrio entre sus diferentes tautómeros: amino-imino para citosina y adenina, ceto-enol para timina y guanina (**Figura 20**). Los equilibrios están desplazado hacia las formas amino y ceto, respectivamente, y salvo en ciertos casos concretos (ver [359] y referencias) tiene poca significancia a nivel biológico.

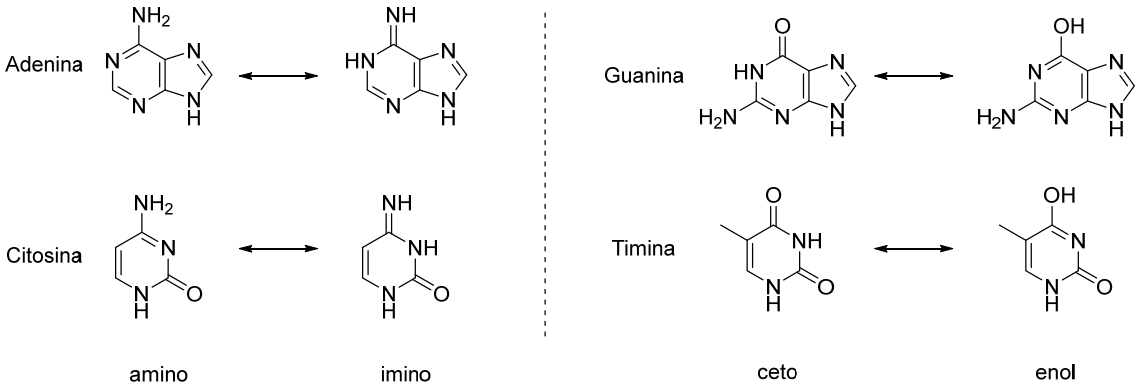


Figura 20. *Tautomerismo de las bases nitrogenadas.*

1.2.4 Tipos de estructuras químicas y síntesis de nucleósidos

Distintas familias de nucleósidos han demostrado su gran versatilidad y eficacia en el tratamiento de diversas enfermedades víricas y también de tumores. Como consecuencia, a lo largo del siglo XX se llevó a cabo una intensa investigación que ha generado una ingente cantidad de estructuras derivadas de los nucleósidos naturales y que a su vez han requerido el desarrollo de nuevas rutas sintéticas, enriqueciendo así el campo de la química orgánica. Las modificaciones pueden realizarse sobre el anillo de furanosa o en los sustituyentes de la misma, sobre la base nitrogenada o en la forma en que la furanosa y la base se unen.

1.2.4.1 Modificaciones sobre la furanosa

Las modificaciones sobre el anillo de furanosa se basan principalmente en un cambio de heterociclo: cambiando el oxígeno por otro heteroátomo, cambiando un carbono por otro heteroátomo o una combinación de ambos. Menos frecuentemente, también pueden encontrarse otros ciclos compuestos por un número diferente de átomos (piranosa, biciclos, anillos aromáticos...) (revisado en [360]–[363]). Los principales heterociclos empleados en la síntesis de nucleósidos están representados en la **figura 21**. Las rutas sintéticas que conducen a la obtención de los nucleósidos suelen tener como primer paso la transformación de un producto natural, por lo general un monosacárido, lo que facilita y asegura obtener la molécula final con la configuración deseada, además de ser económicamente accesibles. Sin embargo, cabe mencionar que algunas rutas

implican la síntesis del anillo derivado de furanosa mediante ciclaciones. Por otra parte, se ha ido desarrollando una amplia batería de métodos biocatalíticos, basados en enzimas o en microorganismos, para transformar no solo el anillo de furanosa y sus sustituyentes, sino también la base nitrogenada y la reacción de acoplamiento de esta con la furanosa. Al margen del coste económico de algunas enzimas, estos métodos son sencillos, eficientes y respetuosos con el medio ambiente sin olvidar uno de los aspectos más importantes como es su aplicación en síntesis enantioselectiva. Salvo ejemplos concretos estos procedimientos no serán comentados (revisado en [364], [365]).

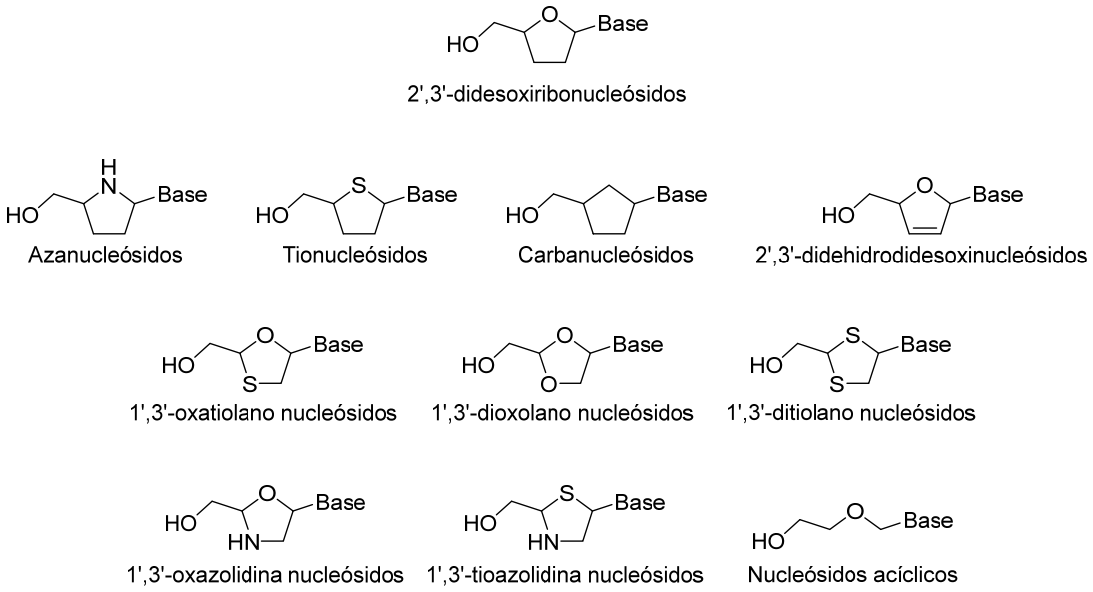


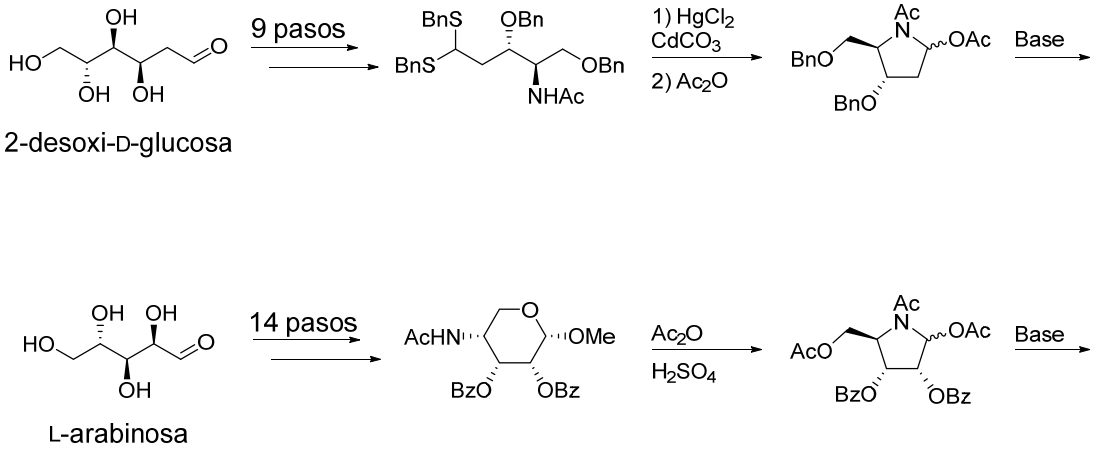
Figura 21. *Selección de estructuras derivadas de furanosa más utilizadas en la síntesis de nucleósidos. Los ciclos con dos heteroátomos pueden encontrarse también con las posiciones de los mismos intercambiadas.*

Dada la cantidad de reacciones descritas hasta la fecha para la obtención de nucleósidos derivados de furanosa, esta sección se centrará en la síntesis de heterociclos derivados de la esta. Por el mismo motivo, en relación con la síntesis de estos compuestos, se describirán aquellas reacciones que conlleven la obtención de nucleósidos con estructuras similares a los inhibidores de RT, dentro de las cuales se dará preferencia a las que representan método más general de síntesis. Añadir que se incluirán preferentemente rutas sintéticas que parten de compuestos quirales y que por tanto pueden conducir con relativa facilidad a productos finales enantioméricamente puros, aunque haya diversos métodos de resolución de enantiómeros. Esto es debido a que muchas veces dichos métodos de resolución

no son satisfactorios al no ser aplicables por restricciones económicas o simplemente porque no son capaces de resolver la mezcla de enantiómeros o el exceso enantiomérico resultante no es apto para su correcta evaluación biológica. No se comentarán síntesis específicas de fármacos NRTIs ya que no suelen ser síntesis generales aplicables en la síntesis de otros nucleósidos al estar orientadas a una producción a nivel industrial.

1.2.4.1.1 Síntesis de derivados de pirrolidina

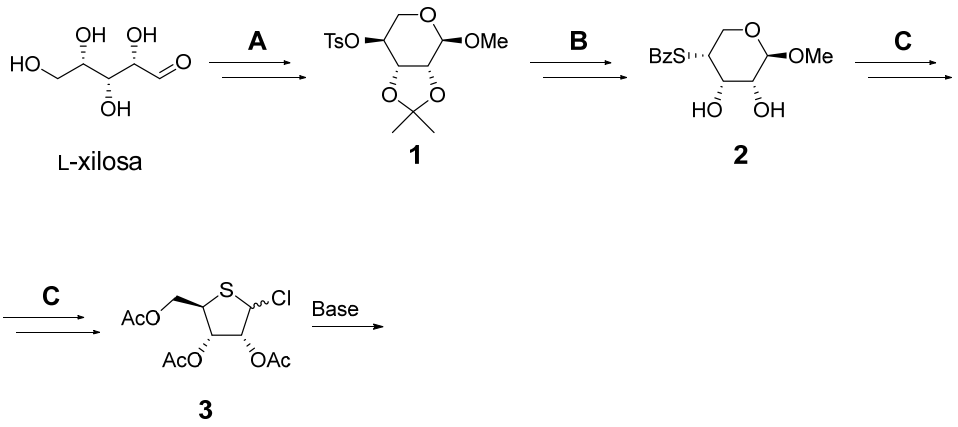
Los análogos de nucleósidos derivados de pirrolidina o azanucleósidos tienen un nitrógeno en el lugar del oxígeno de los nucleósidos naturales. Para su obtención suelen usarse monosacáridos, de cadena abierta [366] o cíclicos [367], con un grupo amino protegido que en última instancia puede ser desprotegido (revisado en [368]) (**Esquema 1**).



Esquema 1. *Síntesis de azanucleósidos.*

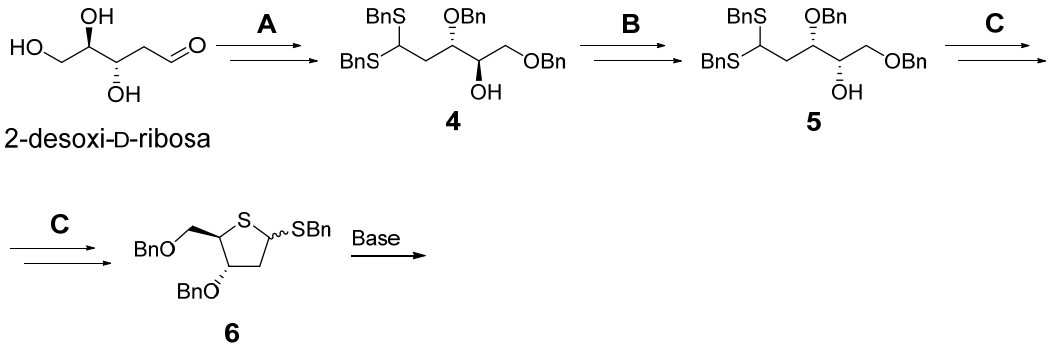
1.2.4.1.2 Síntesis de derivados de tiolano

Los nucleósidos derivados de tiolano surgen del cambio del oxígeno de la furanosa por azufre y pueden ser obtenidos a partir de diferentes azúcares como la L-xilosa. Tras la metilación del hidroxilo anomérico y la consecuente ciclación de la molécula, se hace reaccionar con 2,2-dimetoxipropano y el hidroxilo restante se tosila obteniendo **1**. La reacción S_N2 con tiobenzoato potásico y a continuación hidrólisis del grupo isopropiliden da lugar a **2**. Este compuesto es tratado seguidamente con anhídrido acético, ácido acético y ácido sulfúrico concentrado (obteniéndose así el anillo de tiolano) para finalmente ser tratado con cloruro de acetilo formándose **3** que será acoplado a la base nitrogenada (**Esquema 2**) [369].



***Esquema 2**. Síntesis de tionucleósidos a partir de L-xilosa. **A**: 1) HCl/ MeOH; 2) Ag_2CO_3; 3) DMP, HCl/ dioxano; 4) TsCl. **B**: 1) PhCOSK/ DMF; 2) AcOH. **C**: 1) Ac_2O/ AcOH, H_2SO_4; 2) MeCOCl/ HCl.*

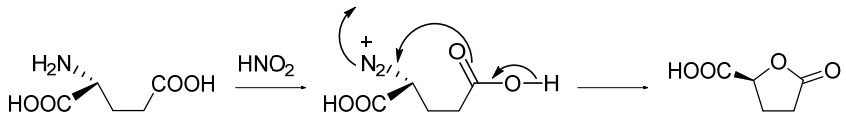
De forma alternativa pueden obtenerse los tionucleósidos a partir de 2-desoxi-D-ribosa. La metilación de esta molécula genera la furanosa correspondiente que es protegida con grupos bencilo para hacerlo reaccionar con bencilmercaptano y ácido clorhídrico, lo que abre el anillo de furanosa para dar lugar a **4**. Mediante la reacción de Mitsunobu se invierte la configuración del hidroxilo en posición 4 que, previa desprotección (en la que se forma **5**), es mesilado. Finalmente se hace reaccionar con yoduro sódico y carbonato de bario para dar lugar al compuesto **6** que es susceptible de reaccionar con una base nitrogenada para formar distintos nucleósidos (**Esquema 3**) [370].



***Esquema 3**. Síntesis de tionucleósidos a partir de 2-desoxi-D-ribosa. **A**: 1) HCl/ MeOH; 2) NaH, Bu_4NI, BrBn/ THF; 3) BnSH/ HCl. **B**: 1) PPh_3, $PhCO_2H$, DEAD/ THF; 2) MeONa/ MeOH. **C**: 1) MsCl/ Py; 2) NaI, $BaCO_3$.*

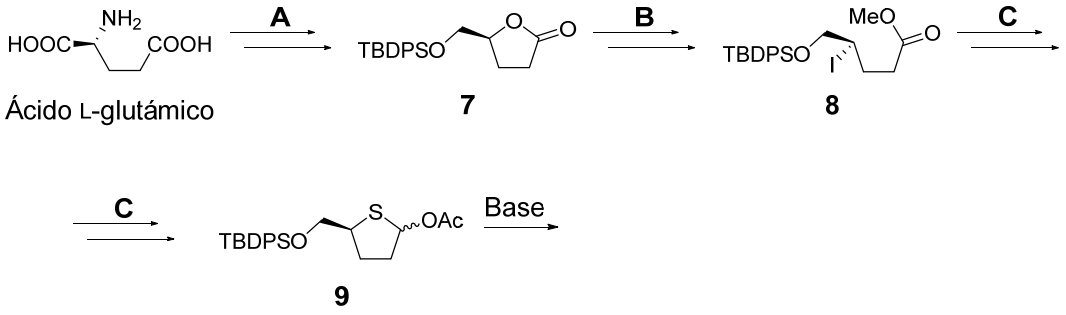
Empleando el ácido L-glutámico como sustrato de partida en la síntesis de derivados de tiolano, el primer paso es la reacción de diazotación. El compuesto resultante sufre una ciclación interna de forma espontánea, reteniendo la

configuración por la participación del grupo α-carboxilo, formándose así una γ-carboxi-γ-butirolactona (**Esquema 4)** [371].



Esquema 4. *Mecanismo de reacción por el cual se forma la γ-carboxi-γ-butirolactona.*

El grupo carboxilo es reducido y el hidroxilo resultante convenientemente protegido (**7**). El ciclo se abre en medio básico, se esterifica el carboxilo libre y se intercambia el hidroxilo por yodo con inversión de la configuración (**8**). El yodo es a su vez intercambiado por tioacetato, de nuevo con inversión de la configuración. La reducción con hidruro de diisobutilamonio (DIBAL-H), además de desproteger el azufre reduce el metiléster a aldehído, produciéndose una ciclación de forma espontánea. El lactol resultante es acetilado (**9)** y puede ser empleado en la reacción de acoplamiento con bases nitrogenadas para formar nucleósidos (**Esquema 5**) [372].

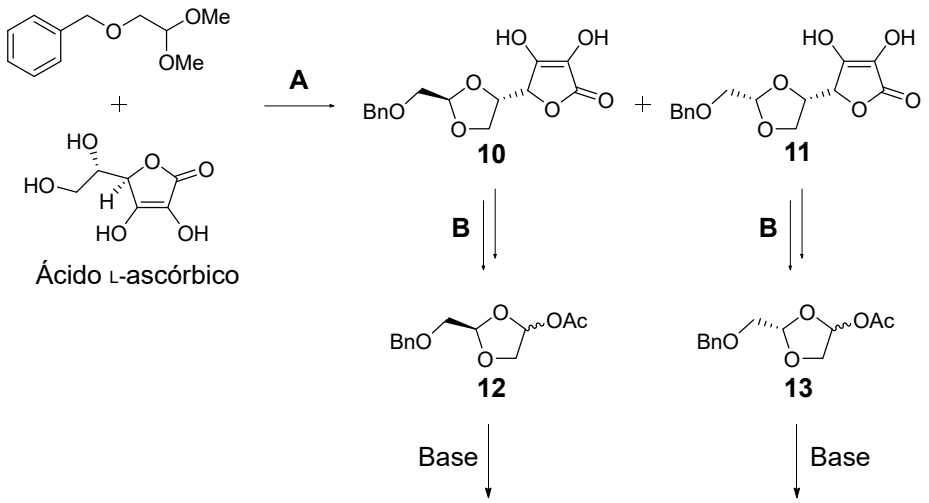


Esquema 5. *Síntesis de tionucleósidos a partir de ácido L-glutámico.* ***A:*** *1) $NaNO_2$/ HCl; 2) $BH_3{\cdot}SMe_2$; 3) TBDPSCl, imidazol.* ***B:*** *1) NaOH; 2) Me_2SO_4; 3) I_2, Ph_3P, imidazol.* ***C:*** *1) Bu_4NOH, MeOH, AcSH; 2) DIBAL-H; 3) Ac_2O, TEA.*

1.2.4.1.3 Síntesis de derivados de 1,3-dioxolano

Para la síntesis de derivados de dioxolano existen diversas rutas. Dos de ellas, reportadas por el mismo grupo, parten del ácido-L-ascórbico o del D-manitol que presentan como gran ventaja la obtención de D- y L-nucleósidos derivados de dioxolano. Según las condiciones de la reacción de acoplamiento con la base nitrogenada, pueden formarse α- y β-nucleósidos de modo que por estas rutas se pueden obtener una gama muy amplia de estructuras [373], [374]. La síntesis que utiliza como producto de partida el ácido L-ascórbico es más efectiva según sus

autores: este compuesto se hace reaccionar con el dimetilacetal de benciloxiacetaldehído para dar lugar a los correspondientes benciloximetilacetales derivados del ácido L-ascórbico **10** y **11**. La separación de estos estereoisómeros se lleva a cabo por cristalización y son tratados con peróxido de hidrógeno que degrada el ciclo proveniente del ácido L-ascórbico. Mediante la reacción de Wolfe, estos compuestos son oxidados y finalmente se obtienen los acetatos **12** y **13** mediante descarboxilación oxidativa utilizando tetracetato de plomo (**Esquema 6**).



Esquema 6. *Síntesis de derivados de dioxolano.* ***A:*** *TsOH.* ***B:*** *1) H_2O_2, K_2CO_3; 2) $RuCl_3$, NaOCl, $BnEt_3NCl$; 3) $Pb(OAc)_4$.*

Relaciones estructura-actividad de los nucleósidos derivados de dioxolano

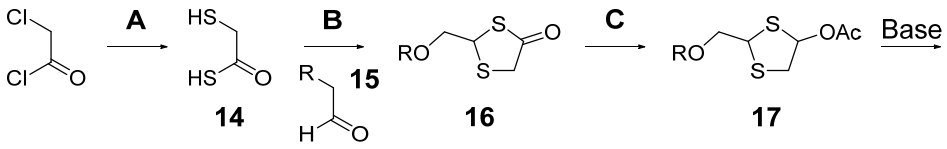
Se han sintetizado una amplia gama de nucleósidos con esta estructura, tanto β-D, como β-L y α-L con el objetivo de buscar características estructurales que mejoren la potencia de estos compuestos frente a VIH. De la evaluación de estos compuestos se extrae que gran parte de ellos poseen una importante capacidad antiretroviral, entre los que la β-L-5-fluoro-1,3-dioxolanilcitidina muestra la más potente dentro de los derivados pirimidínicos y la β-L-2,6-diamino-1,3-dioxolanilpurina entre los púricos (**Figura 22**). En general, sustituyentes pequeños y electrón atrayentes en la posición 5 de bases pirimidínicas aumenta la actividad mientras que un aumento del tamaño o de la capacidad electrón donante del sustituyente lo disminuyen. La sustitución en bases púricas de la posición 6 por un

grupo amino aumenta en general la actividad mientras que los grupos hidroxilo, cloro, metoxi o metilamino producen el efecto contrario. Con algunas excepciones, los L-nucleósidos derivados de 1,3-dioxolano son más activos frente a VIH que sus correspondientes D-nucleósidos [375], [376].

Figura 22. *Estructuras de nucleósidos derivados de 1,3-dioxolano.*

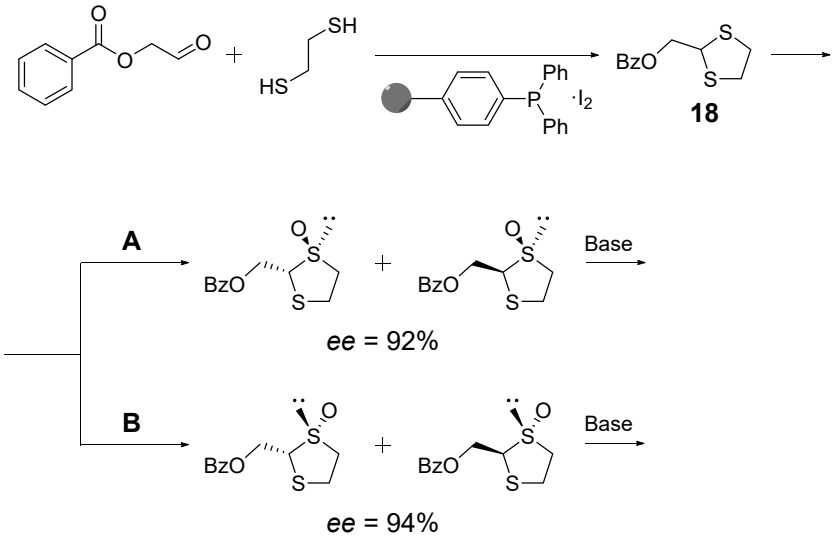
1.2.4.1.4 Síntesis de derivados de ditiolano

Las rutas sintéticas que conducen a los nucleósidos derivados de ditiolano implican pocas reacciones, pero no pueden realizarse partiendo de productos quirales y requieren procesos de resolución de enantiómeros. A modo de ejemplo, haciendo reaccionar cloruro de cloroacetilo e hidrógeno sulfuro de sodio se obtiene el ditiol **14** que reacciona con el aldehído **15** (donde R son diversos grupos protectores) para formar el ciclo de ditiolano **16**. La reducción con DIBAL-H y acetilación del lactol da lugar a **17**. Una vez obtenidas las mezclas racémicas de los isómeros *cis* y *trans*, cada mezcla de enantiómeros es resuelta empleando un HPLC provisto de una columna quiral (**Esquema 7**) [377].



Esquema 7. *Síntesis de derivados de ditiolano.* ***A:*** *NaSH.* ***B:*** *$ZnCl_2$.* ***C:*** *1) DIBAL-H; 2) AcCl.*

La oxidación de un azufre también puede ser empleado para resolver mezclas racémicas. De este modo, el compuesto **18**, obtenido mediante la reacción entre benzoiloxietanal y etanoditiol (en presencia de un complejo de yodo y difenilfosfina soportado en poliestireno que actúa tanto de ácido de Lewis como de agente deshidratante), es oxidado a sus correspondientes sulfóxidos (empleando los correspondientes auxiliares quirales) que pueden ser separados por métodos convencionales y empleados en la síntesis de nucleósidos (**Esquema 8**) [378].

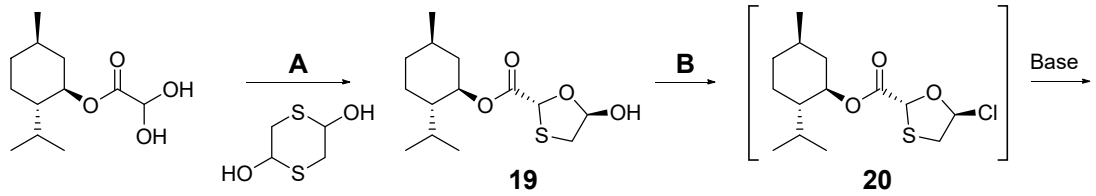


Esquema 8. Síntesis de derivados de ditiolano. **A:** dietil D-tartrato/Ti(*Oi*Pr)$_4$/*t*BuOOH. **B:** dietil L-tartrato/Ti(*Oi*Pr)$_4$/*t*BuOOH.

1.2.4.1.5 Síntesis de derivados de 1,3-oxatiolano

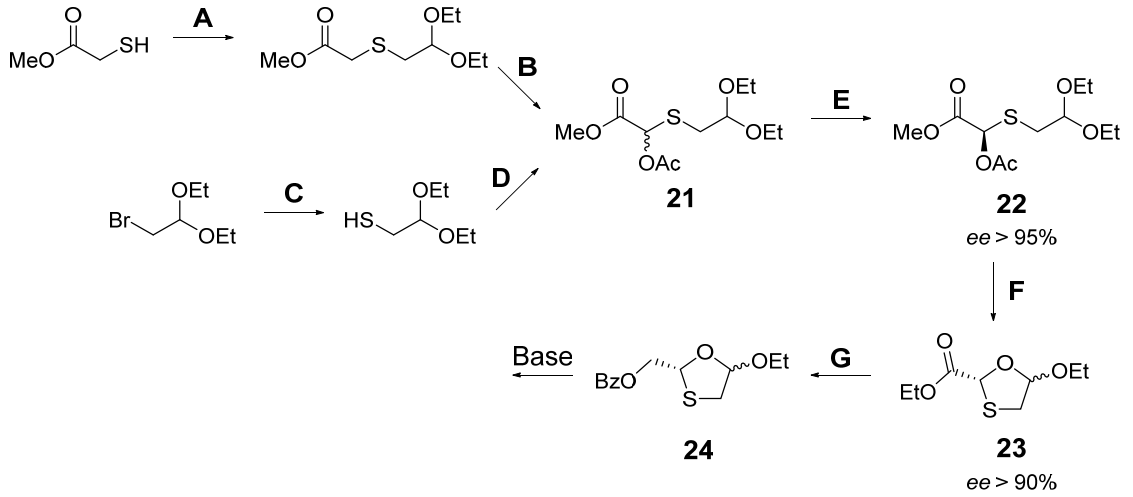
La síntesis de los derivados de oxatiolano tiene un especial interés ya que lamivudina y emtricitabina, dos fármacos con alta eficacia clínica empleados en el tratamiento del VIH, tienen la mencionada estructura. Dado que ambos pertenecen a la serie L-nucleósidos, la investigación sobre la síntesis de nucleósidos derivados de oxatiolano ha ido encaminada a obtener dicha serie. La síntesis quiral de los derivados de oxatiolano implica un número de pasos elevado y en consecuencia un rendimiento a menudo insatisfactorio. Por ello se han desarrollado síntesis a partir de productos de partida aquirales y diversos métodos de resolución de enantiómeros para tal propósito como, por ejemplo, mediante la unión de un intermedio a L-mentol o norefedrina y cristalización fraccionada [379]. Posteriormente se aplicó también el L-mentol como auxiliar quiral, mejorando los resultados anteriores, introduciéndolo en otro punto de la ruta sintética. Para ello, el glioxilato de L-mentol [380] se hace reaccionar con ditianodiol formándose los cuatro diastereoisómeros posibles del compuesto **19**, el cual es el único que cristaliza en hexano conteniendo una pequeña cantidad de TEA como catalizador. Además, el resto de diastereoisómeros pueden interconvertirse en **19** dado que en ciertos disolventes todos los estereoisómeros están en equilibrio. El intercambio del hidroxilo libre por cloro para obtener **20** permite el acoplamiento de la base

(**Esquema 9**). Una vez acoplada la base nitrogenada, para eliminar la fracción perteneciente al L-mentol se reduce con borohidruro sódico [381].



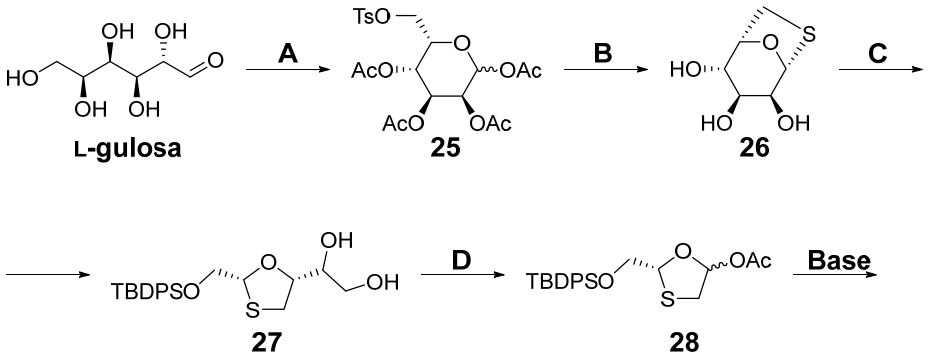
Esquema 9. *Síntesis de derivados de oxatiolano usando un auxiliar quiral.* ***A:*** *1) tolueno calentado a reflujo; 2) ditianodiol; 3) cristalización con hexano y TEA.* ***B:*** *$SOCl_2$.*

También se han usado enzimas que son capaces de transformar uno solo de los enantiómeros formados en una ruta sintética, como la lipasa de *Pseudomonas fluorescens*. Esta enzima, en metil *tert*-butil éter como disolvente, transforma mayoritariamente el enantiómero ***S*-21** desacetilándolo mientras que presenta poca o ninguna actividad sobre el enantiómero ***R*-21** (compuesto **22**) [382]. La mezcla racémica **21** se puede obtener por dos vías: la primera comienza con la reacción de la sal sódica de metil 2-mercaptoacetato con dietilacetal de bromoacetaldehído, oxidación del azufre para obtener el sulfóxido correspondiente que en presencia de anhídrido acético y acetato sódico sufre un reajuste de Pummerer dando lugar a **21**; la segunda vía parte también de dietilacetal de bromoacetaldehído que es transformado a éster xantato con etilxantato de potasio, este en tiol y finalmente se obtiene **21** por tratamiento del tiol con glioxilato de metilo y a continuación con Ac_2O. El intermedio **22** en medio ácido en etanol sufre una transesterificación que desencadena una ciclación espontánea (**23**). Para preparar el precursor de nucleósidos **24**, basta con reducir el éster y finalmente proteger el hidroxilo con benzoilo (**Esquema 10**) [383].



Esquema 10. *Síntesis de derivados de 1,3-oxatiolano con resolución enzimática.* ***A:*** *1) Na; 2) $BrCH_2CH(OEt)_2$.* ***B:*** *1) m-CPBA; 2) Ac_2O, NaOAc.* ***C:*** *1) KSC(S)OEt; 2) etilendiamina.* ***D:*** *1) glioxilato de metilo; 2) Ac_2O, DMAP.* ***E:*** *lipasa de Pseudomonas fluorescens.* ***F:*** *HCl, EtOH.* ***G:*** *1) $LiBH_4$, iPrOH (cat.); 2) BzCl, Py.*

La síntesis de 1,3-oxatiolanos a partir de un producto de partida quiral (L-gulosa) requiere 13 pasos para obtener la serie L [384], la cual es la que mayor interés tiene ya que lamivudina y emtricitabina, pertenecientes a esta serie, son compuestos más potentes que sus enantiómeros, al igual que ocurre con la mayoría de los nucleósidos derivados de 1,3-oxatiolano. Cabe mencionar que, si bien la L-gulosa puede adquirirse de casas comerciales, su precio es muy elevado y suele sintetizarse a partir de otros productos económicamente más asequibles como D-manosa (8 pasos) [385], o la L-gulono-1,4-lactona (3 pasos) [386]. Por otro lado, utilizando la D-manosa es posible sintetizar D-nucleósidos derivados de 1,3-oxatiolano [387]. Partiendo de la L-gulosa (**Esquema 11**), esta es selectivamente tosilada en su hidroxilo primario para a continuación proteger el resto de hidroxilos libres mediante acetilación (**25**). El intercambio del grupo acetilo por bromo y reacción con etilxantato de potasio y desacetilación da como resultado la formación de **26**. Sobre este se emplea $NaIO_4$ para producir la ruptura por oxidación del enlace entre los hidroxilos vecinales en *cis* y el aldehído resultante se reduce. El diol resultante se protege como acetónido para poder proteger en forma de sililéter el hidroxilo primario restante. De nuevo el diol vecinal (**27**) se rompe por oxidación y el aldehído que se forma es oxidado a ácido, el cual es transformado en acetilo (**28**) para poder llevar a cabo la reacción de acoplamiento.



Esquema 11. *Síntesis de derivados de 1,3-oxatiolano a partir de L-gulosa.* ***A:*** *1) TsCl; 2) Ac_2O.* ***B:*** *1) HBr/AcOH; 2) O-etilxantato potásico; 3) NH_4OH.* ***C:*** *1) $NaIO_4$; 2) $NaBH_4$; 3) p-TsOH, acetona; 4) TBDPSCl, imidazol; 5) p-TsOH (cat.).* ***D:*** *1) $Pb(OAc)_4$; 2) PDC; 3) $Pb(OAc)_4$.*

Relaciones estructura-actividad de los nucleósidos derivados de oxatiolano

Tras el hallazgo de la lamivudina y emtricitabina, ambos con estructura de β-L-1,3-oxatiolano, como compuestos muy potentes y con poca toxicidad para el tratamiento del VIH, han aparecido grandes familias de α- y β-nucleósidos tanto de la serie D como L. Curiosamente, la lamivudina y la emtricitabina tienen mayor actividad que sus correspondientes D-isómeros e igual pasa con otros derivados. La mayoría de derivados de uracilo no poseen actividad, pero sí los derivados en la posición 5 de la citosina, siendo la 5-bromocitidina la más activa, a la vez que no presentan toxicidad. Entre la serie D o la L, hay igualdad en cuanto número de derivados activos. Los derivados de purina exhiben buenas actividades (α y β) sin toxicidad entre los que destaca el derivado de inosina. Algunos sustituyentes conducen a un aumento de la actividad como 6-cloropurina, 2-fluoroadenina o N^6-metiladenina (**Figura 23**). En estos derivados todos aquellos nucleósidos de la serie D son más potentes que los de la serie L [384], [388].

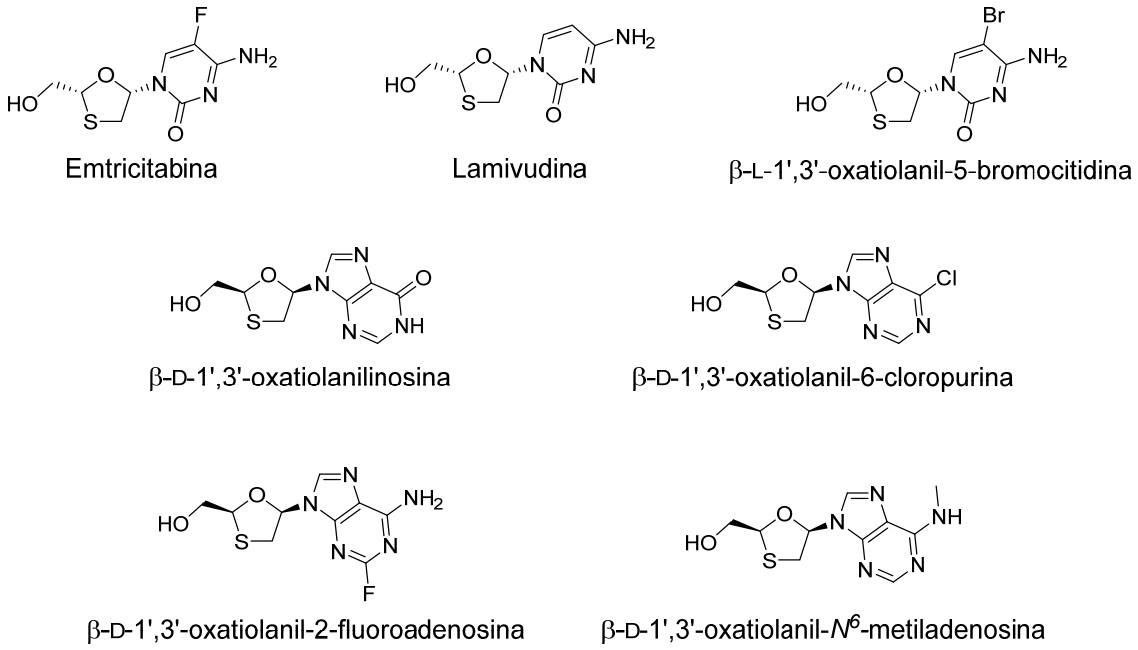


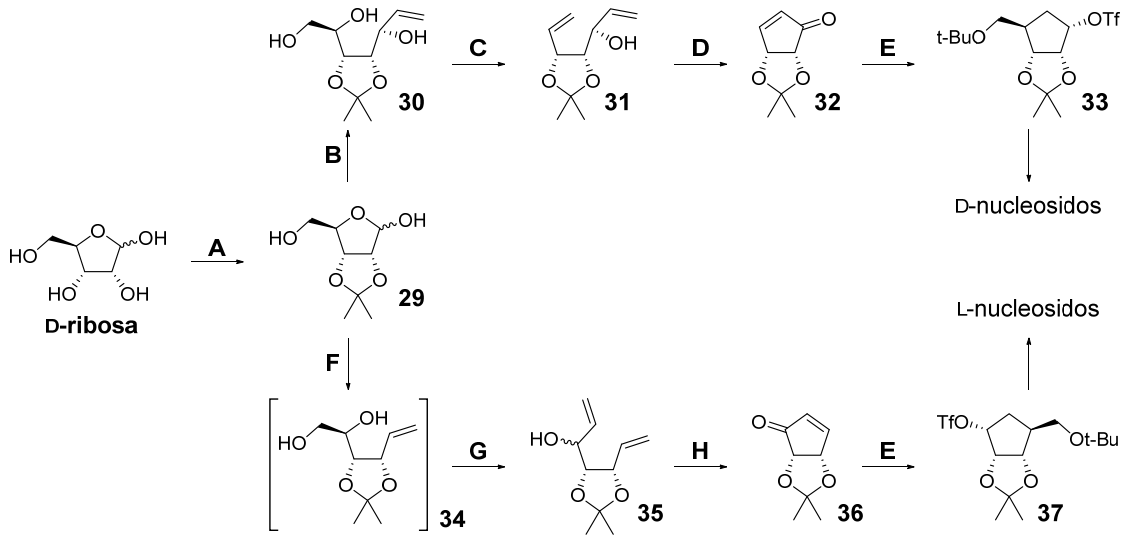
Figura 23. *Estructuras de nucleósidos derivados de 1,3-oxatiolano.*

1.2.4.1.6 Síntesis de derivados de carbociclos

Los nucleósidos derivados de carbociclos o carbanucleósidos son aquellos en los que el oxígeno del anillo derivado de furanosa ha sido sustituido por un grupo metileno (revisados en [389], [390]). Como en el resto de derivados, aunque pueden encontrarse síntesis y actividades biológicas interesantes en otros tamaños de ciclos, los más comunes son aquellos con anillo de ciclopentano o ciclopenteno. Los métodos sintéticos de estas estructuras son muy amplios e incluyen métodos de resolución de mezclas racémicas, síntesis a partir de productos de partida quirales, cicloadiciones, entre otros.

La D-manosa, el ácido D-isoascórbico, la D-ribonolactona, la D-ribosa son algunas de las fuentes quirales que se han utilizado en la síntesis de estos nucleósidos. Partiendo de la última (**Esquema 24**) pueden obtenerse las dos ciclopentenonas **32** y **36,** precursores de nucleósidos carbocíclicos de la serie D y L, respectivamente. Una vez protegida la D-ribosa como diacetónido **29**, la reacción de este con bromuro de vinilmagnesio produce la alquilación con apertura del ciclo para dar lugar a **30**. La ruptura oxidativa de los hidroxilos vecinales vuelve a formar el anillo de furanosa, la cual es sometida a una reacción de Wittig para formar el dieno **31**. El uso sobre dicho compuesto del catalizador de Grubbs permite que tenga lugar una RCM (ring-closing metathesis), cerrándose el ciclo y formando un

ciclopentenol que es oxidado a la D-ciclopentenona **32**. Para obtener su enantiómero, precursor de la serie contraria de nucleósidos, el compuesto **29** se hace reaccionar con bromuro de metiltrifenilfosfonio para dar el intermedio **34** como producto de la reacción de Wittig. La ruptura del diol con metaperyodato sódico y subsiguiente reacción con el magnesiano permite obtener el dieno **35**. Aplicando de nuevo el catalizador de Grubbs tiene lugar la RCM que seguida de una oxidación resulta en la formación de la L-ciclopentenona **36** [391]. Los compuestos **32** y **36** dan un solo isómero cuando son sometidos a una adición conjugada con di-*tert*-butoximetilcuprato de litio. A continuación, el carbonilo es reducido estereoselectivamente y el hidroxilo protegido con triflato (**33** y **37**), lo que activa esta posición para el acoplamiento con la base nitrogenada. La estereoselectividad de la adición conjugada y la reducción se explican por los impedimentos estéricos y efectos electrónicos del acetónido [392].

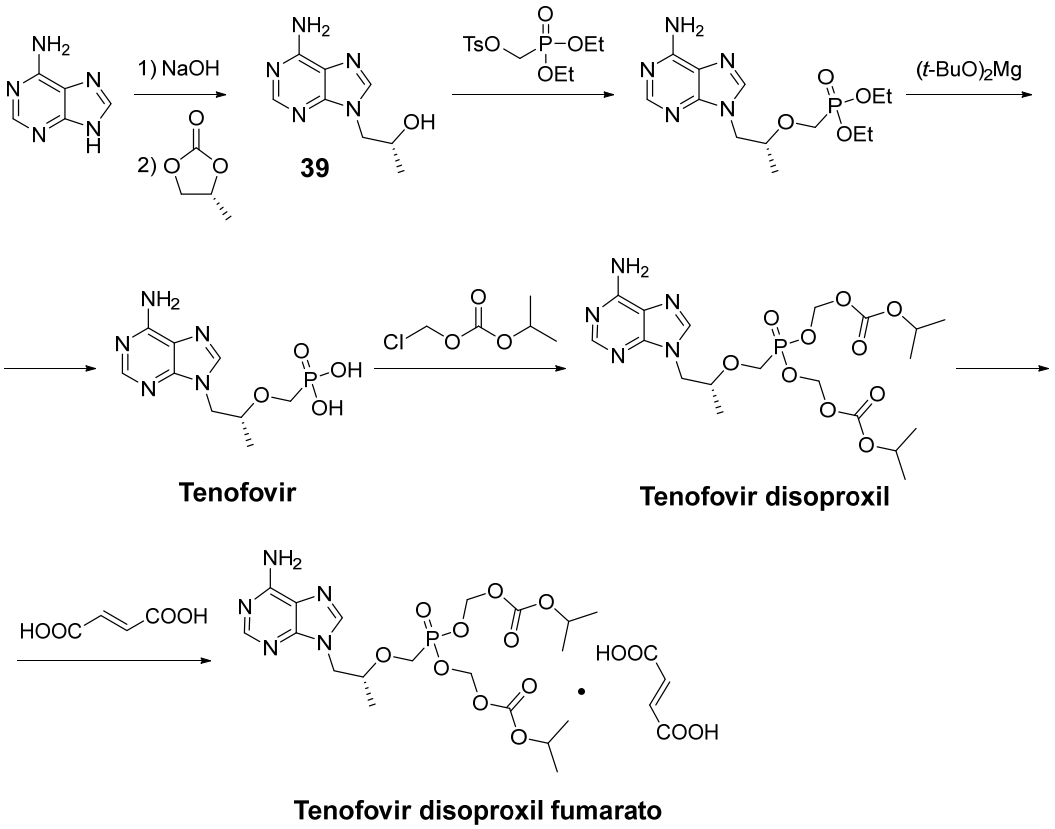


Esquema 12. *Síntesis de carbanucleósidos.* ***A:*** *acetona, H_2SO_4.* ***B:*** *bromuro de vinilmagnesio.* ***C:*** *1) $NaIO_4$, DCM, H_2O; 2) NaH, DMSO, CH_3PPh_3Br.* ***D:*** *1) Catalizador de Grubbs; 2) MnO_2.* ***E:*** *1) $(t\text{-}BuOCH_2)_2CuLi$, t-BuOMe/THF; 2) DIBAL-H; 3) Tf_2O.* ***F:*** *CH_3PPh_3Br, t-BuOK.* ***G:*** *1) $NaIO_4$, DCM, H_2O; 2) Bromuro de vinilmagnesio.* ***H:*** *1) Catalizador de Grubbs; 2) MnO_2.*

Al margen de las diferentes estructuras derivadas de furanosa mencionadas anteriormente, sobre esta fracción de los nucleósidos caben infinitas modificaciones mediante introducción de sustituyentes en prácticamente todas las posiciones y no se comentarán (ampliamente revisado en [360], [361]).

1.2.4.1.7 Síntesis de derivados acíclicos

Estos compuestos constan de una base nitrogenada con una cadena lineal funcionalizada que emula parte de la estructura de la furanosa (revisados en [393], [394]). A este grupo pertenecen importantes fármacos como aciclovir (VHS), cidofovir (CMV) y tenofovir (VIH). El **esquema 13** muestra una de las posibles síntesis del tenofovir que además puede aplicarse a cualquier base nitrogenada. La ruta parte de la adenina, aunque puede intercambiarse por cualquier otra base púrica o pirimidínica. La reacción de la base nitrogenada con el *R*-propilencarbonato permite añadir la cadena lateral (**39**) con el metilo en la configuración deseada. Para formar el tenofovir a partir de este compuesto hay que introducir el ácido fosfónico mediante la reacción con *p*-tosiloximetilfosfonato de dietilo y posterior hidrólisis de los ésteres con terbutóxido de magnesio. Finalmente, y para obtener la forma administrable, se alquilan los hidroxilos libres con clorometilisopropil carbonato y se hace reaccionar con el ácido fumárico para formar la sal correspondiente (**tenofovir disoproxil fumarato**) que cristaliza fácilmente [395].

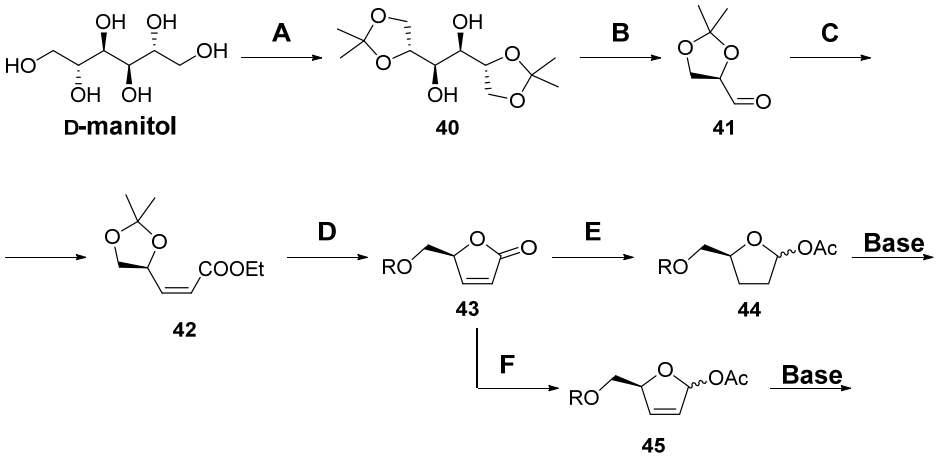


Esquema 13. *Síntesis de derivados acíclicos.*

1.2.4.1.8 Síntesis de derivados de (2',3'-didehidro)-2',3'-didesoxi-D-ribosa

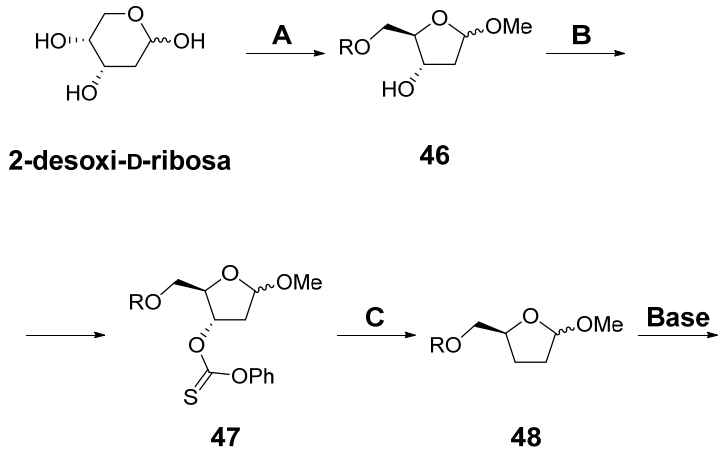
Diferentes rutas sintéticas tanto quirales como aquirales se han desarrollado para obtener nucleósidos con estructura de 2',3'-didesoxi-D-ribosa y los estrechamente relacionados derivados de 2',3'-didehidro-2',3'-didesoxi-D-ribosa. Las rutas que permiten obtener nucleósidos ópticamente puros de estas familias parten del ácido L-glutámico (será visto en Resultados y Discusión), D-manitol, D-ribosa o 2-desoxi-D-ribosa (revisadas en [396]).

Empleando el D-manitol se pueden obtener ambas familias mencionadas anteriormente [397]. Para ello se forma el diacetónido del manitol (**40**) con 2,2-dimetoxipropano en presencia de cloruro de estaño (o acetona en presencia de cloruro de zinc) y se fragmenta en dos aldehídos iguales (**41**) con tetraacetato de plomo. Estos son sometidos a una reacción de Wittig con (etoxicarbonilmetilen)trifenilfosforano de la que resulta el compuesto **42**, el cual en presencia de cantidades catalíticas de ácido sulfúrico cicla para dar la lactona α,β-insaturada **43**. El hidroxilo primario es protegido con grupos sililo, acetilo o tritilo. A partir de **43**, para obtener el precursor de nucleósidos derivados de 2',3'-didesoxi-D-ribosa (**44**) hay que reducir el doble enlace por hidrogenación, reducir con DIBAL-H la lactona a lactol y finalmente acetilar el hidroxilo anomérico. Para obtener el precursor de nucleósidos derivados de 2',3'-didehidro-2',3'-didesoxi-D-ribosa (**45**) solo son necesarias las dos últimas reacciones (**Esquema 14**).



Esquema 14. *Síntesis de 2',3'-didesoxi-D-ribosa y 2',3'-didehidro-2',3'-didesoxi-D-ribosa a partir de D-manitol.* ***A:*** *2,2-DMP, $SnCl_2$.* ***B:*** *$Pb(OAc)_4$.* ***C:*** *Ph_3PCHCO_2Et.* ***D:*** *1) H_2SO_4 (cat.); 2) RCl (TBDPSCl, AcCl, TrCl).* ***E:*** *1) H_2/Pd-C; 2) DIBAL-H; 3) Ac_2O, TEA.* ***F:*** *1) DIBAL-H, 2) Ac_2O, TEA.*

Empleando la D-ribosa se requieren cinco pasos de reacción para obtener el precursor de nucleósidos derivados de 2',3'-didesoxi-D-ribosa mientras que partiendo de 2-desoxi-D-ribosa se reduce a cuatro [398]. Esta última se hace reaccionar con metanol en presencia de cantidades catalíticas de cloruro de hidrógeno para formar el correspondiente metilfuranósido, cuyo hidroxilo primario es protegido con cloruro de 4-bifenilcarbonilo (**46**). El hidroxilo secundario libre se hace reaccionar con clorotiocarbonato de fenilo y dimetilaminopiridina (**47**) para a continuación poder llevar a cabo una desoxigenación de Barton con hidruro de tributilestaño y azobisisobutironitrilo (AIBN), lo que permite obtener el precursor **48** (**Esquema 15**).



Esquema 15. *Síntesis de 2',3'-didesoxi-D-ribosa a partir de 2-desoxi-D-ribosa.* ***A:*** *1) HCl, MeOH; 2) RCl.* ***B:*** *PhOC(S)Cl, DMAP.* ***C:*** *$(n\text{-}Bu)_3SnH$, AIBN.*

Relaciones estructura-actividad de los 2',3'-didesoxinucleósidos

Basándose en la actividad de 2',3'-didesoxiadenosina (**ddA**), 2',3'-didesoxiguanosina (**ddG**), 2',3'-didesoxicitidina (**ddC**, zalcitabina), 2',3'-didesoxitimidina (**ddT**), 2',3'-didesoxiinosina (**ddI**, didanosina), 2,6-diaminopurina-2',3'-didesoxirribósido (**ddAPR**), 5-flour-2',3'-didesoxicitidina (**5FddC**) y 2',3'-didesoxiazacitidina (**ddAC**) (**Figura 24)** se pudo determinar que, aunque todos los derivados eran activos, los derivados de citosina presentaban las actividades más potentes frente a VIH. Sin embargo, estos mismos compuestos exhibían los mayores valores de toxicidad en las células ensayadas. Los 5-metil y 2-bromo derivados de ddC no poseen actividad, de igual forma que ocurre con el 8-bromo derivado de ddA [399].

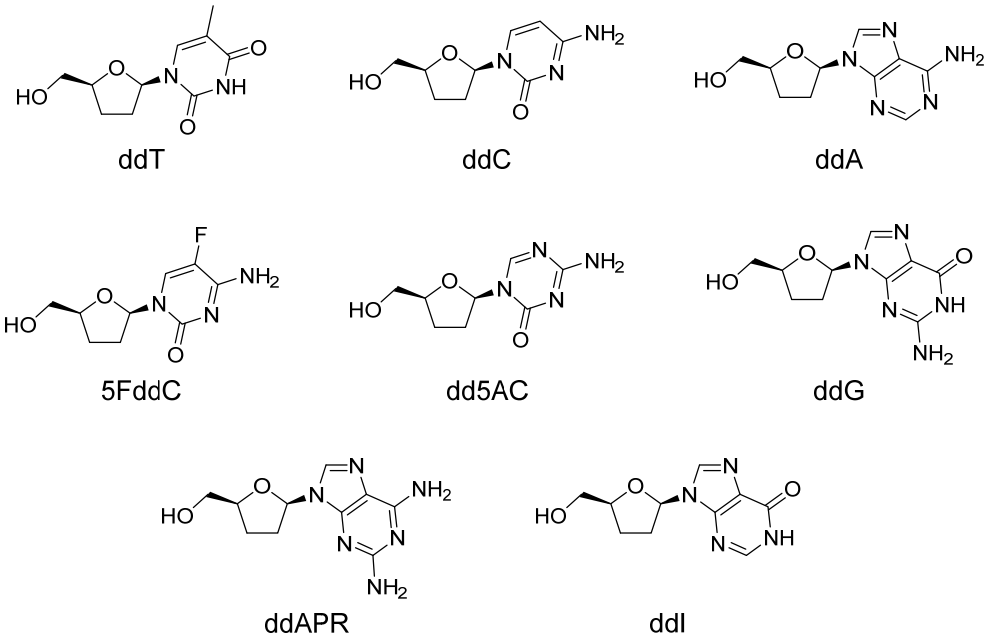


Figura 24. *Estructuras de 2',3'-didesoxinucleósidos.*

Relaciones estructura-actividad de los 2',3'-didehidro-2',3'-didesoxinucleósidos

Entre los derivados de este tipo, se han evaluado la 2',3'-didehidro-2',3'-didesoxicitidina (**d4C** ó ddddC) y la 2',3'-didehidro-2',3'-didesoxitimidina (**d4T** ó ddddT) mostrando una potencia casi igual a sus derivados ddC y ddT (stavudina) mientras que por otra parte los derivados de 2',3'-didehidro-2',3'-didesoxiuridina (**d4U**), 2',3'-didehidro-2',3'-didesoxiadenosina (**d4A**), 2',3'-didehidro-2',3'-didesoxiguanosina (**d4G**) y 2,6-diaminopurina-2',3'-didehidro-2',3'-didesoxirribósido (**d4APR**) carecen de actividad [399] (**Figura 25**).

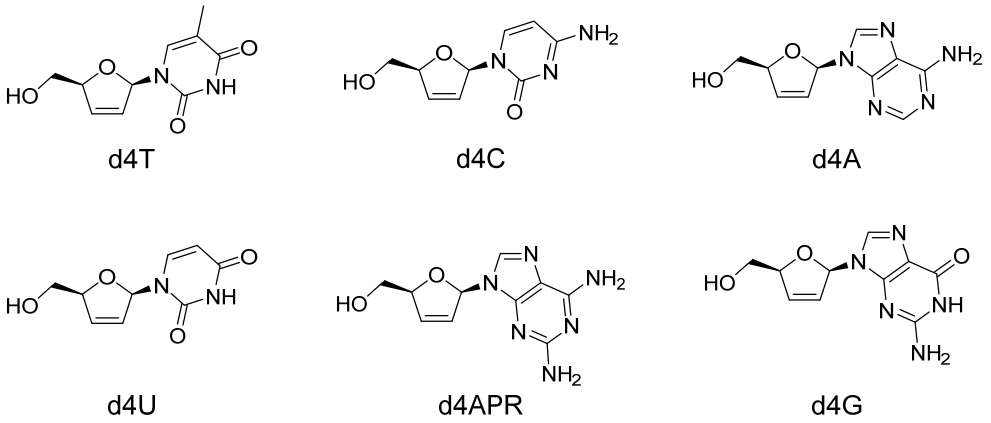


Figura 25. *Estructuras de 2',3'-didehidro-2',3'-didesoxinucleósidos.*

Relaciones estructura-actividad de los 3'-azido-2',3'-didesoxinucleósidos

La potente actividad del **AZT** (3'-azido-2',3'-didesoxitimidina o AzddT) frente a VIH abrió la puerta a la familia de los 3'-azido derivados. Aun siendo el AZT el más potente de esta familia, los derivados 3'-azido-2',3'-didesoxiuridina (**AzddU**) y 3'-azido-2',3'-didesoximetilcitidina (**AzddMeC**) también muestran una potencia elevada, así como los **derivados 3-(3-oxo-1-propenil)** (**Figura 26**) de AzddT, AzddU, AzddBrU, AzddFC, AzddIU, AzddC, AzddFU y algunos derivados 5-metil y 5-etil (cadenas alifáticas más largas como 5-propil o 5-bromovinil conducen a una pérdida de actividad). Los derivados de **AzddA** y **AzddI** pierden la actividad que poseen sus derivados carentes del grupo azido. Algunos de estos compuestos han sido sometidos a una reducción de su grupo azido a amino lo que ha provocado un importante descenso en su actividad al igual que ocurre cuando es sustituido por yodo [319], [399].

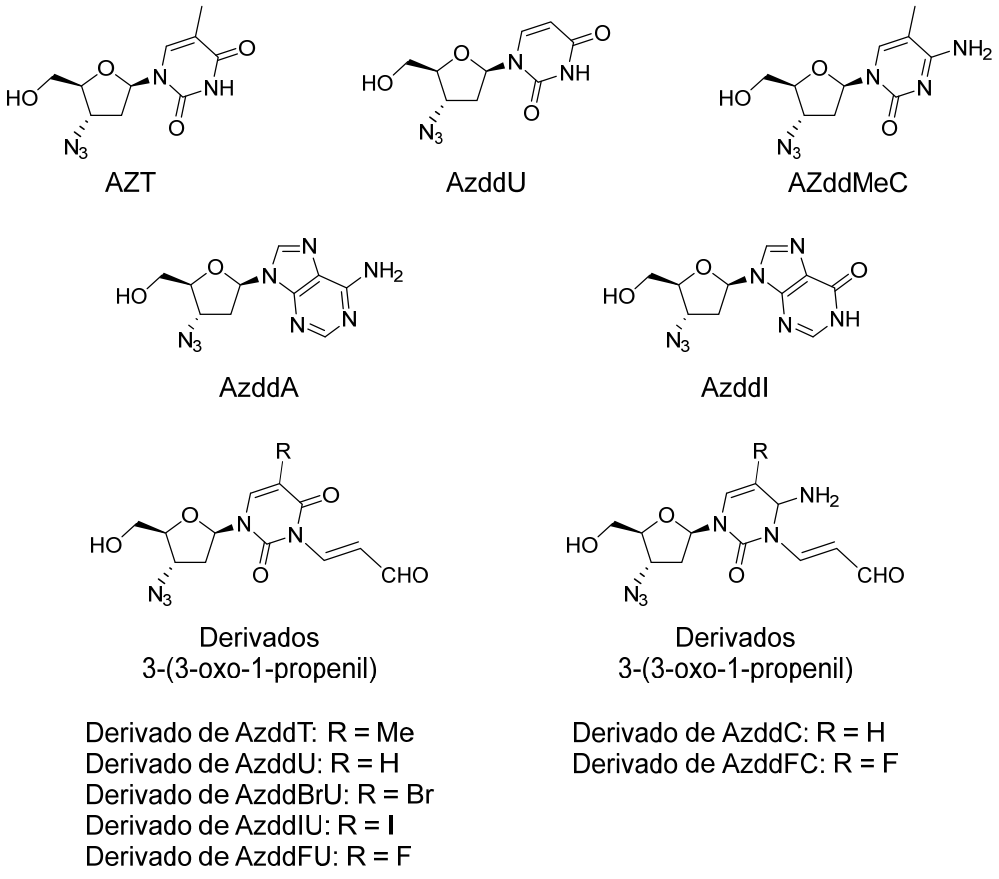


Figura 26. *Estructuras de 3'-azido-2',3'-didesoxinucleósidos.*

Relaciones estructura-actividad de los 3'-fluoro-2',3'-didesoxinucleósidos

Estos compuestos presentan también una marcada actividad anti-VIH. De hecho, el compuesto **FddT** es más potente que AzddT aunque también más tóxico. De igual forma, la actividad de los derivados **FddU**, **FddG**, **FddAPR** y **FddClU** (**Figura 27**) es similar a la de sus 3'-azido derivados, siendo la actividad del último la más alta entre los miembros de esta familia de compuestos [399].

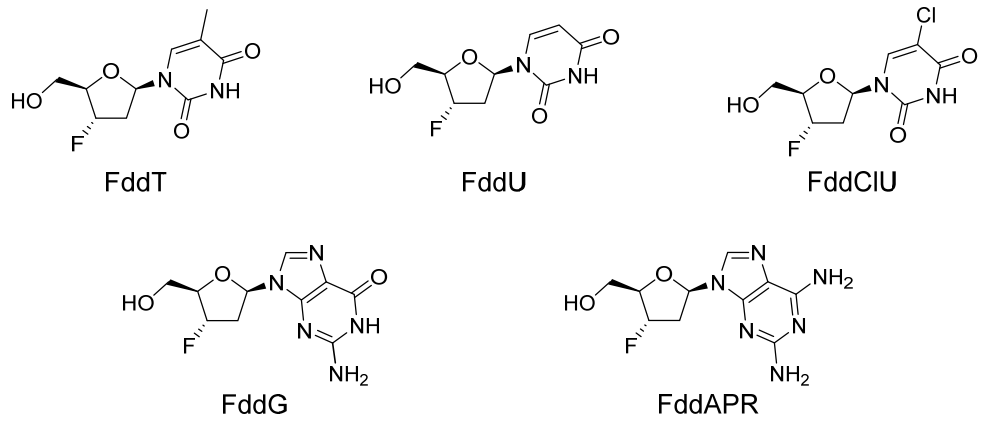


Figura 27. *Estructuras de 3'-fluoro-2',3'-didesoxinucleósidos.*

1.2.4.2 Modificaciones sobre la base nitrogenada

Las modificaciones sobre la base nitrogenada pueden realizarse previamente a la reacción de acoplamiento con el derivado de furanosa o bien una vez unido. En el primer caso se evita la formación de productos secundarios en caso de que los reactivos puedan interaccionar con otros grupos funcionales de la molécula de nucleósido. Estas modificaciones están bastante limitadas ya que las bases nitrogenadas o sus análogos deben ser capaces de establecer puentes de hidrógeno con las enzimas con las que interaccionan. Se mencionarán resumidamente las modificaciones más frecuentes (revisado en [400]).

Las principales modificaciones que se han hecho en bases pirimidínicas incluyen la sustitución de algún hidrógeno de la base por un halógeno, un grupo metoxi o una cadena alifática no mayor de tres átomos de carbono y que puede tener algún sustituyente. Por otra parte, los carbonos 5 y 6 de las bases pueden ser reemplazados por nitrógeno dando lugar al azauracilo y a la azacitosina por ejemplo. Un tipo de modificación que ha llevado a la obtención de un elevado número de nucleósidos es la alquilación aromática o alifática, la acetilación o la

protección con otros grupos de un nitrógeno de la molécula. Algunas bases nitrogenadas son susceptibles de ser transformadas en otras, como la citosina en uracilo, aunque en la práctica estas reacciones no suelen llevarse a cabo al ser las bases asequibles comercialmente.

Las bases púricas suelen ser transformadas en la posición 8 incluyendo un halógeno o una cadena alifática. El C2 se puede funcionalizar, normalmente con un halógeno o un grupo amino, al igual que ocurre con el C6 (grupo ceto en guanina y amino en adenina) y muchas veces ambos. Los nitrógenos de la base pueden ser modificados de forma similar a las bases pirimidínicas.

Lo anterior se aplica a derivados de bases naturales. Sin embargo, mucho se ha investigado con el uso de ciclos heteroaromáticos como imidazol, pirazol, oxazol, isoxazol, piridina, entre otros, en sustitución de las bases púricas y pirimidínicas naturales modificadas.

1.2.4.3 Unión entre la base nitrogenada y el derivado de furanosa

El enlace entre el C1' de la pentosa y la base nitrogenada puede formarse con distintos átomos presentes en esta última. Por lo general, la posición más favorecida para la formación del enlace glicosídico es el N1 o el N9, según sean bases pirimidínicas o púricas respectivamente. Dependiendo de las condiciones de reacción se formarán también isómeros N3 en bases pirimidínicas o N7 o NH_2 (en el C6) de algunas bases púricas.

Si la unión se produce por un C de la base nitrogenada se obtienen los denominados C-nucleósidos o pseudonucleósidos (revisado en [363]). Estos compuestos presentan como gran ventaja su mayor estabilidad ya que el enlace C-C no es hidrolizable por las enzimas celulares como sí ocurre el enlace N-glicosídico. La pseudouridina (**Figura 28**) es un ejemplo de C-nucleósido que además está presente en la naturaleza, siendo el nucleósido modificado más habitual encontrado en el ARN. A este grupo pertenecen también muchos análogos de nucleósidos con anillo heteroaromático no análogo de bases naturales, como los mencionados anteriormente, ya que frecuentemente se unen a través de un carbono de este anillo.

Merecen una mención los nucleósidos isoméricos o isonucleósidos, nucleósidos en los que el enlace entre el azúcar y la base nitrogenada se produce por un átomo distinto al C1', generalmente el C2' aunque también el C3' (revisados en [401]). Estos compuestos han mostrado una elevada potencia antiviral.

Por último, mencionar también los 1'-homonucleósidos, aquellos en los que la base está unida a la furanosa a través de un grupo metileno que actúa de espaciador (revisado en [402]). Estos compuestos también presentan actividad antiviral.

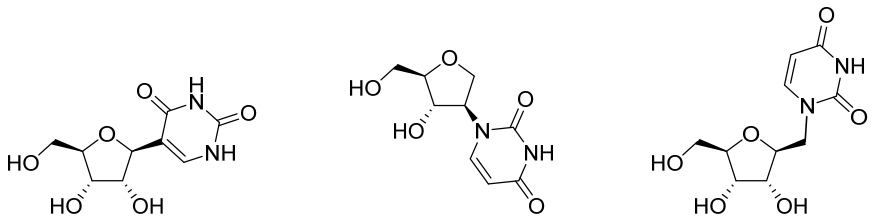


Figura 28. *Estructuras de C-nucleósidos, isonucleósidos y 1'-homonucleósidos.*

1.2.5 Acoplamiento entre la base nitrogenada y el derivado de furanosa

La síntesis de nucleósidos puede realizarse, simplificando, de dos formas: la base nitrogenada se forma a partir de la ciclación de una cadena lineal que está unida al derivado de la furanosa (o viceversa) o más comúnmente se lleva a cabo una reacción de acoplamiento (síntesis convergente) entre la base nitrogenada (adecuadamente activada) y la fracción derivada de furanosa (convenientemente funcionalizada). Este último es el más extendido y en general los métodos actuales se basan en la síntesis de (silyl-)Hilbert-Johnson de *N*-glicósidos que data de 1930 [403]. Las modificaciones realizadas posteriormente hicieron que el nombre cambiara a reacción de Vorbrüggen en honor a sus aportaciones en la optimización de dicho método [404]. Otros métodos actuales de síntesis convergente de nucleósidos se basan en *N*-glicosilaciones a través de estrategias como la activación mediante electrófilos (distintos de ácidos de Lewis) de sustituyentes en C1' como PhSH/NBS o S(O)Ph/Tf_2O. Un tercer método emplea halógenos como

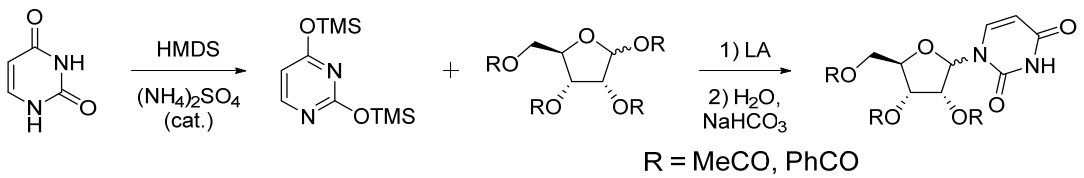
sustituyentes en la posición anomérica que son sustituidos por la base nitrogenada mediante una S_N2 o a través de un ión oxonio formado por un ácido de Lewis (revisado en [405], [406]).

Los métodos alternativos a la reacción de Hilbert-Johnson se basan en lo que se denomina método de sales metálicas en las que se emplean sales de plata y mercurio de las bases nitrogenadas y se hacen reaccionar con monosacáridos peracetilados salvo en su posición C1' en la que el sustituyente es un halógeno (bajos rendimientos y difícil eliminación de las sales, sobre todo en el caso del mercurio que además resulta tóxico al ser evaluado en sistemas biológicos) y la reacción de fusión en la que un monosacárido peracetilado y un heterociclo ácido son calentados hasta que se alcanza la temperatura de fusión y reaccionan entre sí (bajos rendimientos) (revisado en [405], [406]).

1.2.5.1 Reacción de Vorbrüggen

Esta reacción de glicosilación se basa en el tratamiento de un monosacárido peracetilado (ofrece mejores rendimientos que el análogo benzoilado [407]) con una base nitrogenada sililada y un ácido de Lewis (LA en adelante) (**Esquema 16)** (revisado en [405], [408]). La reacción también es aplicable cuando otros sustituyentes están presentes en el derivado de furanosa (sobre todo se aplica al hidroxilo 5'). Además, el grupo saliente en C1 de la furanosa no necesariamente tiene que ser un grupo acetilo, si bien es el grupo que mejores rendimientos ofrece, y puede sustituirse por otros como metoxi, cloro o bromo.

Como LA se han empleado diversos tipos, así como combinaciones de ellos. Algunos de ellos son los siguientes, dentro de los cuales los tres primeros son los más comúnmente empleados: $SnCl_4$, TMSOTf, $EtAlCl_2$, $TMSClO_4$, $TiCl_4$, $Ti(OiPr)_4$, $TiCl_2(OiPr)_2$, $BF_3{\cdot}Et_2O$, $Zn(OTf)_2$, $ZnCl_2$, TMSCl, TMSI, $AgClO_4$ [405], [409], [410].



Esquema 16. *Esquema general de la reacción de Vorbrüggen.*

El acoplamiento suele llevarse a cabo en acetonitrilo (ACN) o 1,2-dicloroetano (1,2-DCE) anhidros y conduce a la formación de N^1-isómeros para pirimidinas de forma preferente y mezcla de regioisómeros en el caso de purinas, aunque los N^9-isómeros son mayoritarios [409]. Los rendimientos de las reacciones de acoplamiento con pirimidinas son mayores que los de las reacciones con purinas. Por otro lado, los anillos derivados de furanosa reaccionan más rápidamente que los de piranosa.

Las bases silanizadas (Silyl-Hilbert-Johnson) no solo reaccionan más rápido, sino que además son lipófilas y por tanto solubles en los disolventes orgánicos habitualmente empleados en esta reacción, permitiendo una reacción homogénea a diferencia de las hidrófilas bases sin silanizar (Hilbert-Johnson). Debido a las propiedades electrón donantes del silicio, las bases silanizadas son más nucleófilas que las correspondientes alquiloxi que se empleaban anteriormente.

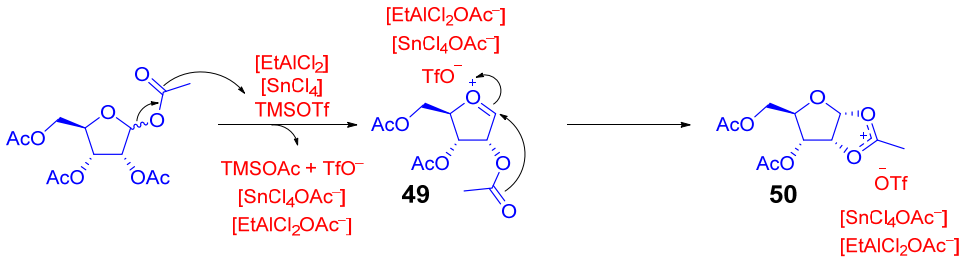
La configuración del enlace glicosídico tras el acoplamiento viene determina por la configuración y tamaño de los sustituyentes en C2', hecho que se explica por el mecanismo de reacción. Brevemente, y como reglas generales, la ausencia de grupo en C2' conlleva la formación de α y β-nucleósidos en proporciones similares, mientras que la presencia de grupos C2' conduce el acoplamiento en la orientación contraria a este, con una mayor proporción de un isómero o incluso de forma exclusiva de manera más marcada cuanto mayor es el grupo director en C2' [411].

1.2.5.1.1 Condiciones de reacción y mecanismo de acoplamiento.

La reacción de Vorbrüggen implica 3 procesos reversibles (revisado en [405], [406], [408], [412]–[414]):

1) Formación del catión electrófilo derivado de la furanosa (**50**). El LA es el responsable de la eliminación del grupo acetilo (**49**), con la subsiguiente formación de una sal de 1,2-aciloxonio (**50**) en función del LA empleado (**Esquema 17**). Sobre esta estructura, la base silanizada con hexametildisilazano (HMDS) únicamente puede llevar a cabo el ataque nucleofílico sobre la posición C1' de la furanosa por la cara β o superior, obteniéndose de forma muy mayoritaria β-

nucleósidos. Los α-nucleósidos se forman en poca o ninguna proporción, aunque hay ciertos factores que aumentan el ataque por la cara α: a) si la base tiene grupos muy polarizados o cargados negativamente lo que permitiría la interacción de la base con el catión por la cara α; b) si el derivado de furanosa contiene grupos polares en su cara α; c) si la furanosa no contiene sustituyentes en C2' capaces de formar el ión aciloxonio (como 2'-desoxinucleósidos), caso en el que la base reaccionaría con un análogo del compuesto **49** con el que tiene teóricamente las mismas probabilidades de hacerlo por la cara α que por la β; d) cuando el catión se forma sobre el plano de la furanosa, lo que implica que el ataque puede realizarse solo por la cara α, formándose así de forma muy mayoritaria α-nucleósidos.



Esquema 17. *Formación del catión electrófilo derivado de la furanosa.*

2) Formación de un complejo σ entre la base silanizada y el LA. La formación de estos complejos tiene lugar con el N1 de las bases pirimidínicas y N6 de las púricas, ya que aglomeran la mayor concentración electrónica (**Figura 29**). Dicha formación depende la basicidad y acidez de aquellos e implica una disminución de los moles de LA disponibles para formar el catión 1,2-aciloxonio. Por ello, para llevar a cabo la reacción los equivalentes del LA deben ser mayores cuanto más básica es la base nitrogenada (como citosina) y cuanto mayor sea la acidez del LA (como el $SnCl_4$) de modo que siempre pueda haber LA libre. Además, estos complejos son los responsables de la formación de los N^3-nucleósidos derivados de pirimidinas y regioisómeros distintos de N9 derivados de purinas, ambos no naturales (lo que no significa que carezcan de interés). Cabe mencionar que los complejos σ son predominantes en disolventes no polares como 1,2-DCE mientras que en disolventes polares (como ACN) se reduce su formación al competir este con el LA para formar tales complejos.

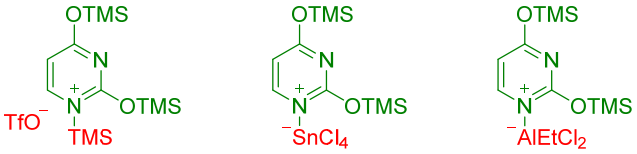
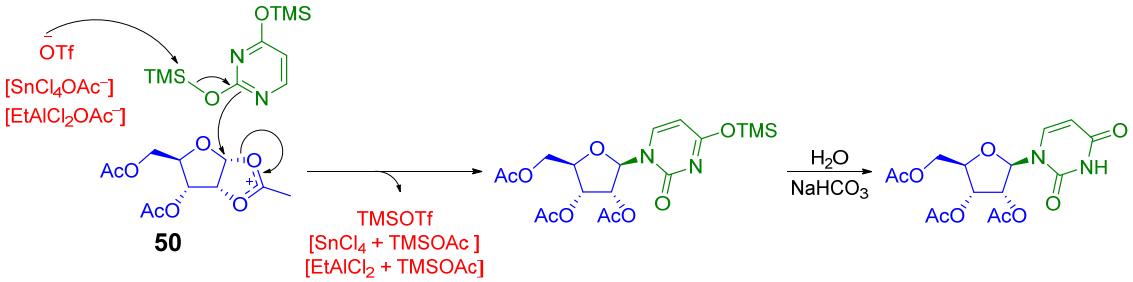
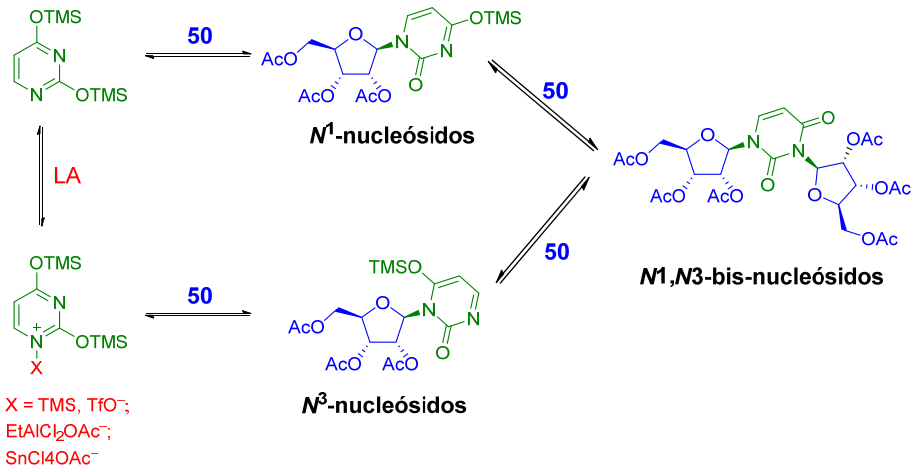


Figura 29. *Complejos σ entre las bases silanizadas y los ácidos de Lewis.*

3) Reacción entre la base silanizada y el catión 1,2-aciloxonio **50**. La base silanizada, sin formar complejo σ con el LA, es la que reacciona con **50** por ataque nucleofílico para obtener los N^1-nucleósidos (**Esquema 18**), mientras que la formación de N^3-nucleósidos se debe a que los complejos σ tienen bloqueado el N1 de la base y solo pueden reaccionar por el N3, que al estar más impedido reacciona más lentamente. Durante el ataque nucleofílico el LA se regenera a través del TMS saliente del oxígeno en C2 de la base. El sililéter restante es fácilmente hidrolizado en solución acuosa de $NaHCO_3$. Por otra parte, ambos nucleósidos pueden reaccionar con **50** para formar un N^1,N^3-bis-nucleósido. Todas estas reacciones están en equilibrio (**Esquema 19**). Por tanto, y dado que al inicio de la reacción los complejos σ son mayoritarios, los N^3-nucleósidos son los productos cinéticos de la reacción mientras que los N^1-nucleósidos son los productos termodinámicos. Puesto que el LA es necesario para la transición entre los regioisómeros, las reacciones en disolventes apolares favorecen que haya más ácido libre necesario para llevar a cabo la transformación de los N^3-nucleósidos a N^1-nucleósidos.



Esquema 18. *Reacción entre la base silanizada y el catión 1,2-aciloxonio.*

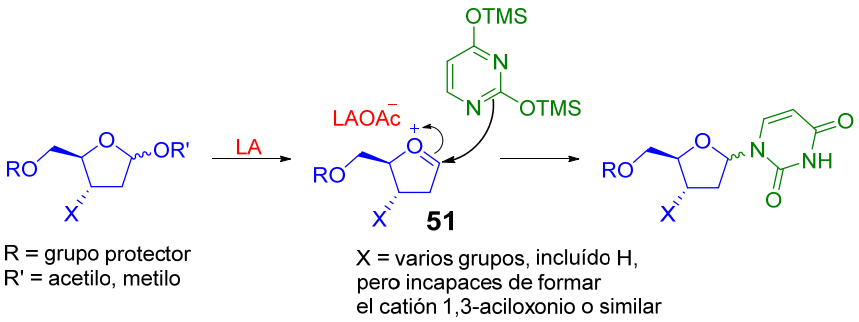


Esquema 19. *Reacciones en equilibrio en la síntesis de nucleósidos por la reacción de Vorbrüggen.*

1.2.5.1.2 Mecanismos de acoplamiento con otros derivados de furanosa

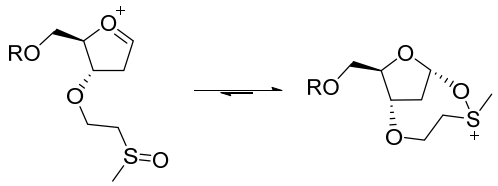
La reacción de Vorbrüggen no solo es aplicable a derivados peracetilados o perbenzilados de la furanosa, sino que también puede realizarse con otras estructuras. Los rendimientos suelen ser igualmente buenos pero existen otras implicaciones derivadas del mecanismo de reacción (revisado en [405], [406], [415]).

Los acoplamientos entre bases silanizadas y 2'-desoxinucleósidos (y por extensión 2',3'-didesoxinucleósidos) pueden llevarse a cabo al igual que en los casos anteriores, si bien la ausencia de un grupo en C2' que imponga restricciones estéricas y/o electrónicas condiciona la reacción a una pérdida de la estereoselectividad que se ha visto previamente con los derivados peracetilados y perbenzoilados, entre otros. Los sustituyentes en C3' y C4' están a demasiada distancia como para contribuir a la estereoselectividad. En general esta es la situación habitual para 2-desoxinucleósidos y aplicable a la gran mayoría de sustituyentes en C3' ya que al no ser posible la formación del intermedio **50** ni de ningún derivado 1,3-aciloxonio, la base nitrogenada reacciona con el oxonio **51** (**Esquema 20**) formándose mezclas de α/β isómeros independientemente del LA utilizado (sin incluir LA quirales).



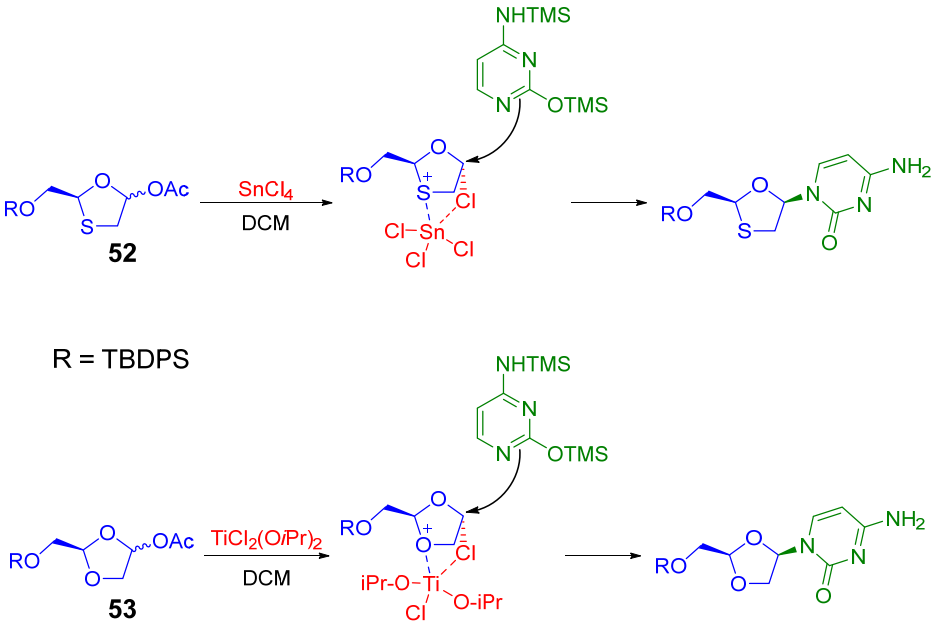
Esquema 20. *Mecanismo de acoplamiento de derivados de 2'-desoxirribosa y 2',3'-didesoxirribosa con las bases silanizadas.*

Existen algunos casos concretos en los que el sustituyente en C3' confiere cierta estereoselectividad β/α a la reacción como el caso del flúor (5:1) posiblemente debido a repulsiones causadas por la alta electronegatividad del flúor o del 3'-*O*-etilmetil sulfóxido (8:1) que estabiliza el oxonio **51** (**Esquema 21**). En teoría todo grupo capaz de formar estructuras como esta aumentarían la estereoselectividad de la reacción.



Esquema 21. *Formación del intermedio que estabiliza el oxonio* ***51.***

Los 1,3-heterociclos derivados de furanosa con más de un heteroátomo como oxatiolano y dioxolano presentan ciertas particularidades en cuanto a su comportamiento con los LA. Tanto el oxatiolano **52** como el dioxolano **53** en presencia de una base silanizada y TMSOTf da como resultado la formación de nucleósidos sin estereoselectividad. Sin embargo **52** y **53** en presencia de $SnCl_4$ y $TiCl_2(OiPr)_2$, respectivamente, y una base silanizada da lugar a la formación casi en exclusiva de β-isómeros. En ambos casos se debe a la formación de un complejo entre el heteroátomo y el metal del LA que se coordina también a través de uno de sus ligandos con el Cl' del derivado de furanosa por la cara α, impidiendo o dificultando el acceso de la base por esta (**Esquema 22**).



Esquema 22. *Mecanismo de la reacción de acoplamiento con 1,3-oxatiolano y 1,3-dioxolano.*

2. Objetivos

2 OBJETIVOS:

La movilidad en humanos del retroelemento LINE-1 tiene importantes implicaciones en la etiología y progresión de tumores. Por otra parte, la función de la movilidad del LINE-1 en ciertos tejidos, si la hubiera, es desconocida, así como su papel, si lo tuviera, en ciertas enfermedades autoinmunes y células senescentes. La inhibición de la reversotranscriptasa de LINE-1, como enzima clave en el proceso de retrotransposición, puede ser útil en el tratamiento del cáncer, así como para estudiar el papel que tiene la retrotransposición en ciertos tejidos o situaciones patológicas a través de una estrategia de pérdida de función. Sin embargo, los estudios previos sobre la materia no han sido suficientes para determinar la selectividad de los inhibidores de RT evaluados ni su toxicidad y, en algunos casos mencionados anteriormente, tampoco se ha demostrado directamente que los efectos observados se deban a la inhibición de la RT del L1.

Por ello el objetivo global de la presente Tesis Doctoral es la síntesis de compuestos nucleosídicos y su evaluación como inhibidores potentes y selectivos de la reversotranscriptasa del elemento LINE-1, para lo que se plantean los siguientes objetivos específicos:

A. Síntesis de nucleósidos inhibidores de LINE-1.

A.1. Síntesis de nucleósidos derivados de 2,3-desoxirribosa.

A.2. Síntesis de nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa.

A.3. Síntesis de nucleósidos derivados de 3-fluoro-2-desoxirribosa.

B. Evaluación de la inhibición de la movilidad de retrotransposones.

B.1. Evaluación de la inhibición de la movilidad de LINE-1 humano.

B.2. Evaluación de la inhibición de la movilidad de LINE de pez cebra.

B.3. Evaluación de la inhibición de la movilidad de LINE de ratón.

B.4. Evaluación de la inhibición de la movilidad de retrotransposones LTR de ratón.

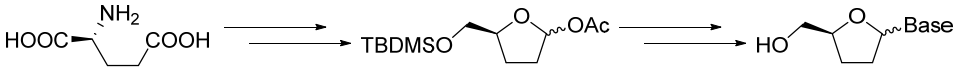
C. Evaluación de la toxicidad de los compuestos ensayados.

3. Resultados y Discusión

3 RESULTADOS Y DISCUSIÓN

3.1 SÍNTESIS DE NUCLEÓSIDOS INHIBIDORES DE LINE-1

3.1.1 Síntesis de nucleósidos derivados de 2,3-didesoxi-D-ribosa



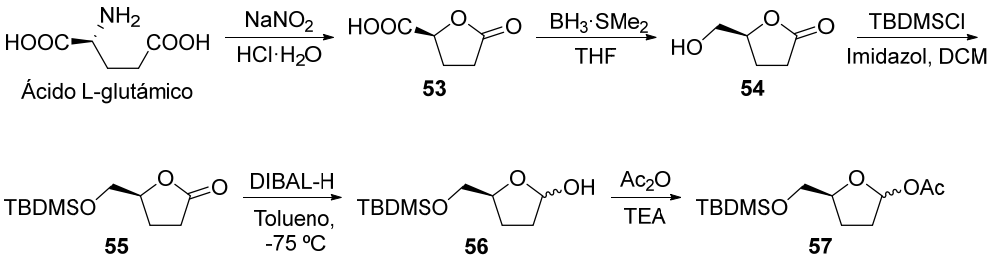
Ácido L-glutámico

Todos los fármacos pertenecientes al grupo de NRTIs son nucleósidos derivados de 2,3-didesoxi-D-ribosa, la mayoría de ellos pirimidínicos. Debido a ello, nuestro primero objetivo fue sintetizar análogos nucleosídicos de dicha familia. Entre las posibles rutas sintéticas reportadas se escogió la ruta que emplea el ácido L-glutámico [416] como material quiral de partida por lo que, teniendo conocimiento de las reacciones a las que es sometido, se pueden obtener los productos acetilados precursores de nucleósidos con la configuración deseada. Además, esta síntesis implica pocos pasos y buenos rendimientos. Alternativamente la ruta puede partir de otros compuestos quirales como D-manitol o D-ribonolactona (como se ha mencionado anteriormente) aunque requieren un mayor número de reacciones.

3.1.1.1 Síntesis del precursor de nucleósidos derivados de 2,3-didesoxi-D-ribosa

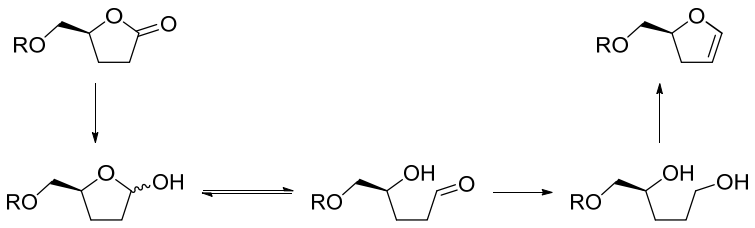
El ácido L-glutámico fue sometido a una desaminación en solución acuosa con ácido nitroso generado *in situ* mediante la reacción entre el nitrito sódico y el ácido clorhídrico añadido lentamente a la reacción. En primer lugar, se produce la diazotación y consecuente formación de la sal de diazonio correspondiente. Debido a la inestabilidad de estos compuestos, se produce una desaminación y ciclación de la molécula de forma espontánea, como se mostró en el **esquema 4**, manteniéndose la configuración. La purificación consistió en una agitación con resina de intercambio catiónico (Dowex 50WX8, 200-400 mesh, H^+ form) que

capta el ácido L-glutámico sobrante para obtener γ-carboxi-γ-butirolactona **53** como un sólido cristalino. Este fue reducido con un complejo de borano y sulfuro de dimetilo en THF anhidro dando lugar al alcohol correspondiente (**54**) en forma de sirupo amarillento. El crudo se hizo reaccionar con cloruro de *tert*-butildimetilsilano en presencia de imidazol en diclorometano (DCM) para proteger el hidroxilo libre. La lactona protegida **55** se obtuvo con un rendimiento del 48% desde el ácido L-glutámico, similar al reportado (51%).



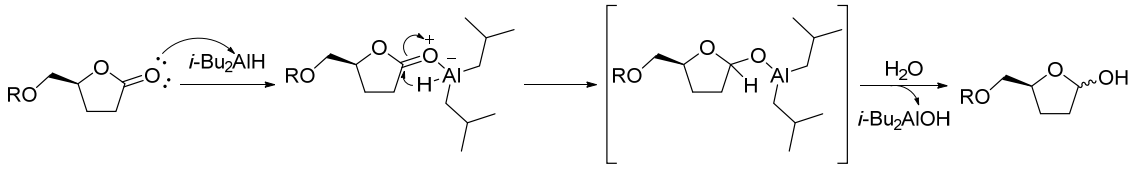
Esquema 23. *Síntesis del precursor de nucleósidos derivados de 2',3'-didesoxi-D-ribosa.*

La reducción con hidruro de diisobutilaluminio (DIBAL-H) en tolueno de la lactona a lactol (**56**) y su posterior acetilación con anhídrido acético en trietilamina permitió obtener el precursor de nucleósidos derivados de 2',3'-didesoxi-D-ribosa (**57**) (**Esquema 23**). La reducción de lactonas a lactoles implica un equilibrio entre su forma cíclica y abierta como hidroxialdehído, sobre el que un exceso de agente reductor genera el diol correspondiente, el cual puede ciclar para dar lugar a un 2,3-dihidrofurano (**Esquema 24**). El uso del DIBAL-H frente a otros agentes reductores (como $LiAlH_4$) permite obtener el lactol con mayor rendimiento al suponer unas condiciones de reacción más suaves, lo que impide la reducción a diol [417], [418]. Adicionalmente, la reacción se lleva cabo a bajas temperaturas para desplazar el equilibrio a la forma cíclica (lactol). De hecho, la reacción de reducción de **55** se llevó a cabo a -78 °C ya que a esta temperatura el intermedio tetrahédrico es estable (**Esquema 25**). Sin embargo, tras el procesado de la reacción y acetilación se pudo comprobar mediante RMN que el compuesto presente de forma mayoritaria correspondía con el diol diacetilado y el 2,3-didehidrofurano. El uso de hidruro de tri-*tert*-butóxido de aluminio y litio [419], [420] manteniendo el resto de condiciones no produjo cambios en la lactona.



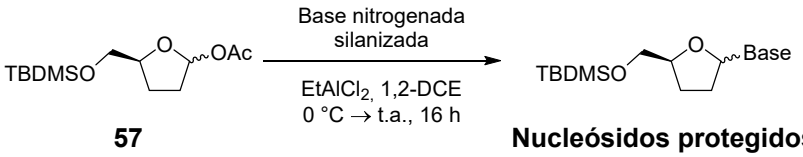
Esquema 24. *Mecanismo de apertura de lactoles y formación del 2,3-didehidrofurano.*

La apertura de lactonas cuando son reducidas a lactoles es un problema recurrente en bibliografía por lo que se han reportado métodos para reducir la formación de productos no deseados en la reacción. En nuestro caso, la adición de dimetilamino piridina, trietilamina y anhídrido acético sobre la reacción de reducción con DIBAL-H a -78 °C [421], manteniendo la agitación y dejando que la temperatura subiera lentamente, condujo a la formación en exclusiva del compuesto deseado **57** con un rendimiento del 67% desde **55** (33% desde el ácido L-glutámico) tras su purificación por columna de sílica flash, la cual se realizó con un 1% de TEA para reducir la descomposición de **55** en contacto con la sílica.



Esquema 25. *Mecanismo de reducción de lactonas con DIBAL-H.*

3.1.1.2 Acoplamiento entre el precursor 57 y las bases nitrogenadas



Esquema 26. *Reacción de acoplamiento entre **57** y las bases nitrogenadas silanizadas.*

Para formar los nucleósidos protegidos, el compuesto **57** fue sometido a la reacción de Vorbrüggen (**Esquema 26**) con un total de 9 bases nitrogenadas, tanto naturales como modificadas (**Figura 30**): uracilo, timina, 6-azauracilo, 5-fluorouracilo, 5-clorouracilo, 5-bromouracilo, citosina, 5-fluorocitosina y 5-azacitosina. Dichas bases deben previamente ser activadas mediante silanización para poder llevar a cabo el acoplamiento con el precursor **57**. Debido a las propiedades electrón donantes del silicio, la silanización de las bases provoca un

aumento de la densidad electrónica en la base nitrogenada favoreciendo así su reactividad. Además, garantiza una buena solubilidad en el disolvente de la reacción, favoreciendo de nuevo la reacción. Para silanizar las bases nitrogenadas, estas se hicieron reaccionar con hexametildisilazano (HMDS) en presencia de cantidades catalíticas de sulfato amónico (**Esquema 27**) calentando la suspensión resultante a reflujo hasta que se obtuvo una solución (1-2 horas), lo que indicaba que la reacción había llegado a su fin, momento en el que se procedió a evaporar el resto de HDMS y algunos subproductos del mismo mediante vacío.

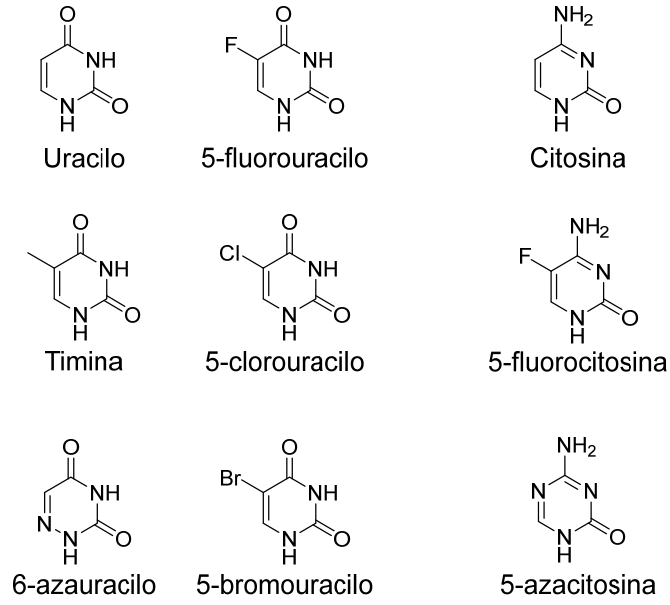
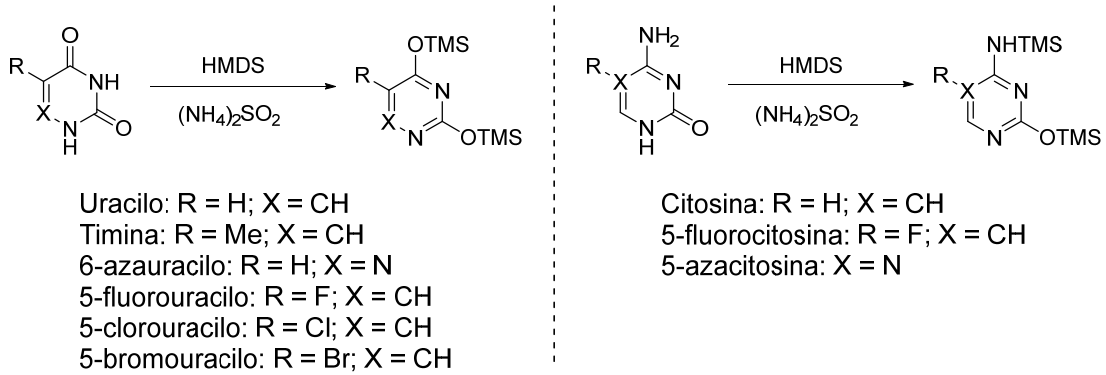


Figura 30. *Bases nitrogenadas empleadas en la presente Tesis Doctoral.*



Esquema 27. *Silanización de las bases nitrogenadas.*

Existen múltiples condiciones para llevar cabo la reacción de Vorbrüggen entre el compuesto acetilado **57** y las bases silanizadas (**Figura 26**). Básicamente, y como se ha mostrado anteriormente, estas condiciones se pueden resumir en la elección del disolvente (ACN, 1,2-DCE o DCM) y del ácido de Lewis (LA).

A fin de determinar las condiciones más favorables (en nuestro caso) de entre las disponibles para realizar la reacción de Vorbrüggen, se llevaron a cabo distintas reacciones de acoplamiento entre el uracilo silanizado y el compuesto **57** combinando, siempre en atmósfera inerte, ACN y 1,2-DCE como disolventes anhidros y $SnCl_4$, TMSOTf y $EtAlCl_2$ como LAs. Además, aunque la adición del LA se hizo siempre a 0 °C, cada una de las condiciones de reacción fue llevada lentamente a tres temperaturas diferentes: -20 °C, temperatura ambiente, 60 °C. Una muestra de cada condición fue tomada a las 16-18 horas de reacción, procesada y analizada mediante cromatografía líquida acoplada a espectrometría de masas (LC-MS) de baja resolución para detectar la presencia de nucleósidos (**Tabla 2**).

Disolvente	Ácido de Lewis	Formación de nucleósidos
ACN	$SnCl_4$ (2, 5 y 10 eq)	No
	TMSOTf (2 y 5 eq)	No
	$EtAlCl_2$ (2 y 5 eq)	**Sí**
1,2-DCE	$SnCl_4$ (2, 5 y 10 eq)	No
	TMSOTf (2 y 5 eq)	No
	$EtAlCl_2$ (2 y 5 eq)	**Sí**

Tabla 2. *Condiciones de acoplamiento evaluadas.*

Independientemente del disolvente empleado, el uso de $SnCl_4$ (tanto puro como en disolución) o TMSOTf (recién destilado) como LAs a distintas concentraciones no permitió la detección de nucleósidos por LC-MS ni mediante cromatografía en capa fina (TLC). Si bien el uso de $SnCl_4$ y TMSOTf en la reacción de Vorbrüggen está ampliamente reportado en bibliografía, en nuestro caso no solo conllevó la ausencia de formación de nucleósidos, sino que además produjo la desaparición de los productos de partida (por TLC). Sin embargo, todas las condiciones de reacción en las que se utilizó $EtAlCl_2$ (1 M en tolueno) condujeron a la formación de nucleósidos. Los rendimientos obtenidos en ACN y 1,2-DCE resultaron ser prácticamente similares. Independientemente del disolvente y del LA empleado, las distintas temperaturas a las cuales fueron sometidas las reacciones de acoplamiento no parecieron tener ningún efecto significativo sobre el rendimiento de la reacción.

En vista de los resultados, puesto que el $EtAlCl_2$ resultó ser el único LA que permitió la obtención de nucleósidos se escogió este para realizar el resto de reacciones de acoplamiento entre las bases nitrogenadas silanizadas y el compuesto acetilado **57**. Este LA además resulta más fácilmente manejable en comparación con los anteriores, no requiere ningún tipo de purificación previa a su uso y según Okabe et al., de entre los LAs empleados, el uso de $EtAlCl_2$ lleva a la formación de β-nucleósidos en una mayor proporción que los α-nucleósidos [416]. En cuanto al disolvente de elección, dado que ninguno ofreció mejores resultados sobre el otro, se escogió el 1,2-DCE que facilita la posterior extracción con DCM. Tras la adición de $EtAlCl_2$ a la reacción a 0 °C, se decidió dejar que alcanzase lentamente temperatura ambiente al no haber variaciones en cuanto a rendimiento se refiere con las distintas temperaturas ensayadas.

Una vez escogidas las condiciones que mejores resultados ofrecían, se procedió a la formación del resto de nucleósidos protegidos con las bases pirimidínicas silanizadas arriba mencionadas. Las bases derivadas de citosina a menudo se encuentran con su grupo amino primario protegido para reducir la reactividad de dicho grupo en la reacción de Vorbrüggen y obtener menor proporción de regioisómeros no deseados. Sin embargo, se optó en este caso por no utilizar dichas bases para evitar un paso extra de desprotección al estar reportado que la mejora de rendimiento producida cuando se emplean derivados de citosina protegidos no es muy significativa [422].

En la mayoría de ellas el producto de partida **57** desapareció completamente (TLC) dando lugar a los α- y β-nucleósidos protegidos correspondientes. Estos estereoisómeros fueron detectados con LC-MS de baja de resolución mientras que el análisis por TLC no permitió dicha observación. El desarrollo repetido (3-6 veces) de la TLC empleando una fase móvil menos polar que la anterior permitió la detección mediante TLC de los estereoisómeros de todas las reacciones de acoplamiento. Inicialmente la purificación mediante columna de sílica flash (FCC) resultó insatisfactoria al no ser capaz de resolver la mencionada mezcla de estereoisómeros en la mayoría de los casos. Sin embargo, el uso de TLC preparativas (Merck) permitió en todos los casos la separación de las mezclas de

estereoisómeros (50-150 mg dependiendo de la base nitrogenada) desarrollando repetidas veces la TLC preparativa. Este sistema permitió obtener puros los nucleósidos protegidos representados en la **Figura 31**.

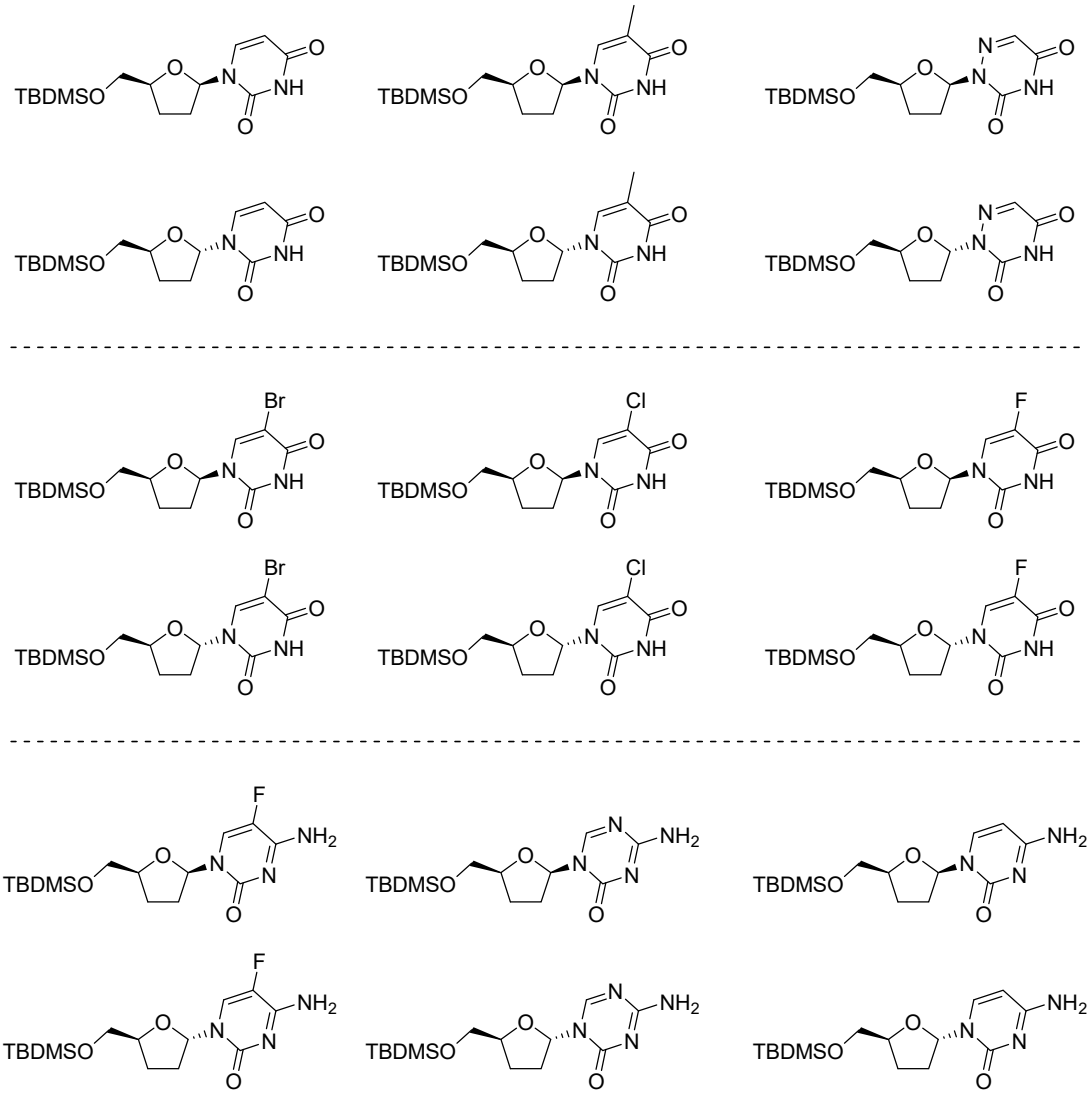


Figura 31. *Nucleósidos protegidos obtenidos derivados de 2,3-didesoxi-D-ribosa.*

En algunos casos (uracilo, timina, citosina y 5-fluorocitosina) se detectó la formación de otros regioisómeros mediante LC-MS de baja resolución, presumiblemente N^3-nucleósidos, aunque no pudieron aislarse en cantidades suficientes para caracterizarlos y continuar con la desprotección. Cabe mencionar que se llevaron a cabo estas mismas reacciones con bases nitrogenadas púricas previamente silanizadas como guanina, adenina y 2,6-dicloropurina, pero en todos los casos se observó la formación de múltiples regiosiómeros (4-6) que resultaron imposibles de purificar. Estas mezclas fueron sometidas a la desprotección de

grupo sililo en 5' con la intención de que fuera posible su separación. Sin embargo, la mezcla de regioisómeros de los nucleósidos desprotegidos resultó ser igualmente inseparable.

La determinación de la configuración del enlace glicosídico se realizó mediante espectros 1D-NOESY (generalmente sobre el nucleósido protegido) irradiando los protones H1' y H4' (**Figura 32**). En los casos en los que había una ausencia de señal positiva en ambos estereoisómeros entre dichos protones (H1'-H4'), se tomó en cuenta la intensidad de otras señales como H1'-H5' y H4'-H6 para establecer la configuración, además de otros datos disponibles en bibliografía en el caso de los productos que habían sido previamente descritos.

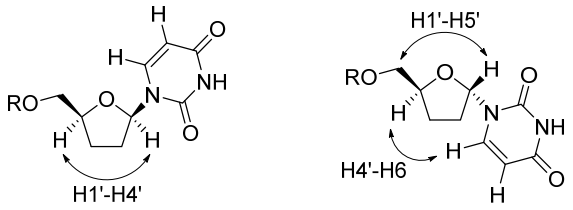


Figura 32. *Interacciones a través del espacio (NOE) empleadas en la determinación de la configuración del enlace glicosídico.*

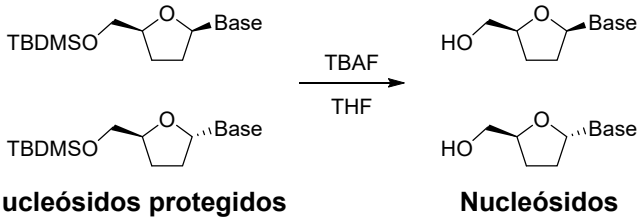
La **Tabla 3** muestra los rendimientos de las reacciones de acoplamiento realizadas con cada base y la proporción de estereoisómeros obtenidos tras la purificación de los nucleósidos protegidos. Cabe mencionar que el método de purificación mediante TLC preparativa tiene como objetivo obtener en un alto grado de pureza estos compuestos, pero a cambio el rendimiento disminuye comparado con otros métodos como FCC. Algunos de estos nucleósidos protegidos han sido previamente sintetizados (ver Materiales y Métodos).

Base nitrogenada	Configuración	Rendimiento	Proporción α/β
Uracilo	α	18%	1 : 1.3
	β	23%	
Timina	α	13%	1 : 1.7
	β	22%	
6-azauracilo	α	23%	1 : 1.3
	β	30%	
5-fluorouracilo	α	15%	1 : 1.9
	β	28%	
5-clorouracilo	α	18%	1 : 1.8

	β	33%	
5-bromouracilo	α	16%	1 : 1.3
	β	21%	
Citosina	α	21%	1 : 1.1
	β	23%	
5-fluorocitosina	α	29%	1 : 1.1
	β	32%	
5-azacitosina	α	22%	1 : 1
	β	22%	

Tabla 3. *Rendimientos y proporciones de estereoisómeros protegidos obtenidos en las reacciones de acoplamiento entre el compuesto* ***57*** *y las distintas bases nitrogenadas.*

3.1.1.3 Desprotección del hidroxilo en posición 5' de los nucleósidos protegidos derivados de 2,3-didesoxi-D-ribosa



Esquema 28. *Reacción de desprotección de los nucleósidos derivados de 2,3-didesoxi-D-ribosa.*

Para obtener estructuras que puedan evaluarse en cultivo celular, los nucleósidos protegidos que se habían sintetizado fueron sometido a la desprotección de su grupo hidroxilo en 5' (**Esquema 28**). Uno de los métodos más eficientes de desprotección de sililos es a través de iones fluoruro (revisado en [423]), de modo que se procedió a la ruptura del enlace O-Si empleando fluoruro de tratrabutilamonio (TBAF) sólido en THF anhidro [424]. A pesar de lo extendido que está este método, en nuestro caso, tras el procesado de la reacción, no solo no generó los nucleósidos desprotegidos, sino que además produjo la descomposición del material de partida. Debido a ellos se exploraron alternativas para la consecución de la desprotección como el uso de ácido *p*-toluensulfónico [416], fluoruro de cesio a reflujo [425], Dowex 50WX8 [426], fluoruro potásico [427] o un complejo HF·Py [428]–[431]. Únicamente las condiciones que empleaban ácido *p*-toluensulfónico y HF·Py condujeron a la obtención de nucleósidos desprotegidos, pero con bajos rendimientos y descomposición de los compuestos presentes en la reacción. Teniendo en cuenta que la cantidad de agua presente en el reactivo TBAF afecta a la reacción [432], procedimos a añadir tamiz molecular

(3 Å) a esta reacción con similares resultados a los anteriores. Finalmente, el uso de TBAF (1 M en THF) anhidro en THF permitió obtener los nucleósidos desprotegidos en 15-20 minutos con rendimientos cuantitativos. El procesado habitual de estas reacciones implica extraer con agua para eliminar los restos de TBAF (altamente tóxico) y subproductos, método no aplicable a nuestro caso puesto que los nucleósidos son solubles en agua. La FCC no logró eliminar los productos indeseados por lo que se recurrió al uso reportado en bibliografía de Dowex 50WX8 y carbonato cálcico, que captan el tetrabutilamonio y el fluoruro respectivamente [433]. Este sistema no redujo el contenido en sales de tetrabutilamonio incluso cuando el proceso fue repetido sucesivamente, por lo que se recurrió al uso de columnas de extracción en fase sólida C8 (SPE C8). De entre las casas comerciales que fueron probadas, solo las SPE C8 de Supelco eliminaron la totalidad de sales de tetrabutilamonio presentes en la reacción. Por lo que sabemos, es la primera vez que se hace uso de este sistema para eliminar exitosamente sales de tetrabutilamonio. Para ello, el crudo de la reacción fue disuelto en agua e introducido en la SPE previamente equilibrada. Empleando agua como fase móvil se recogieron tubos que fueron analizados en LC-MS de baja resolución para confirmar la presencia de los nucleósidos deseados y descartar la presencia de sales de tetrabutilamonio, obteniéndose puros y con un rendimiento cuantitativo los nucleósidos representados en la **Figura 33**. La **Tabla 4** recoge los rendimientos globales (desde ácido L-glutámico) para cada nucleósido sintetizado. Algunos de estos nucleósidos protegidos han sido previamente sintetizados (ver Materiales y Métodos)

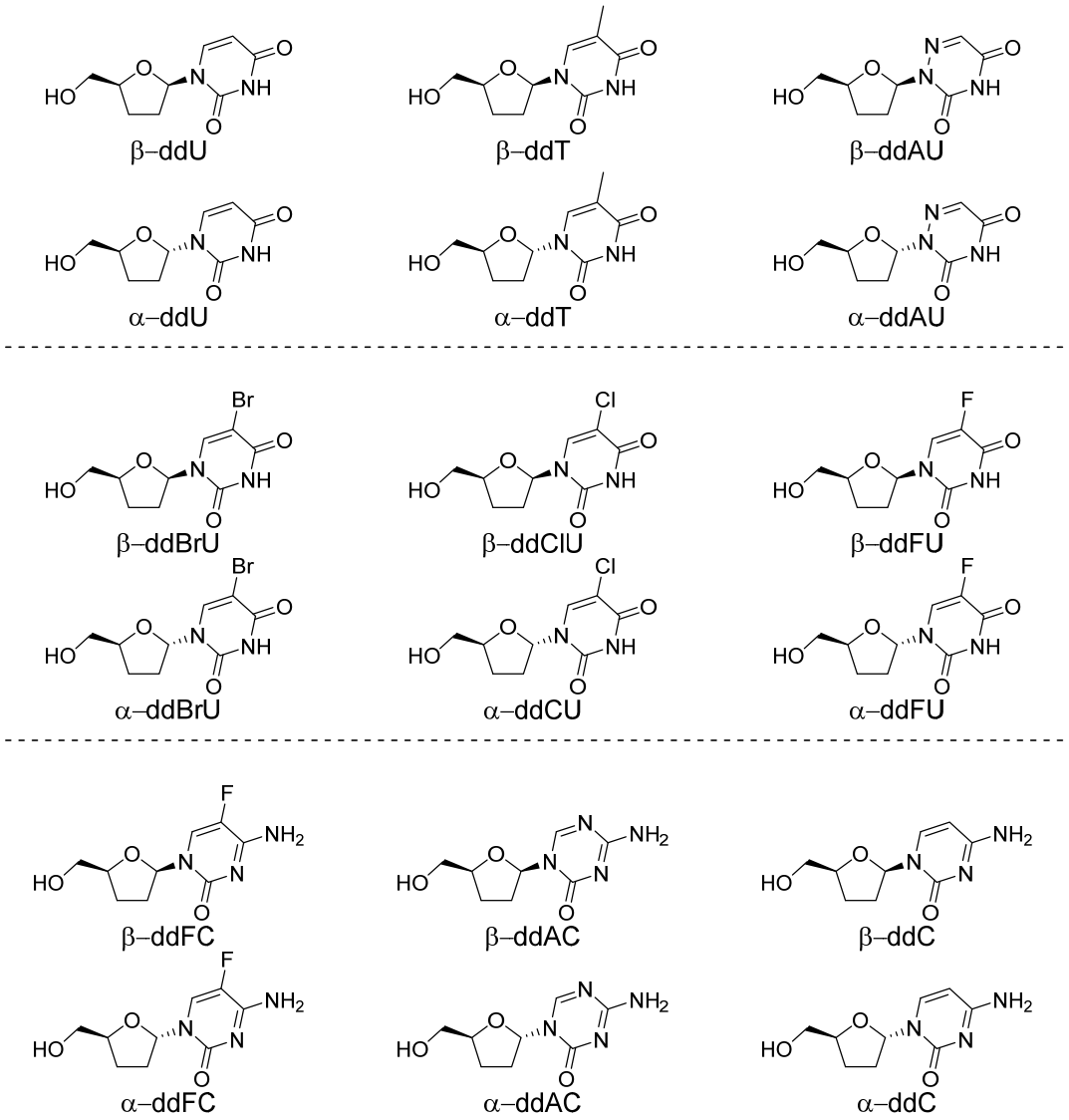


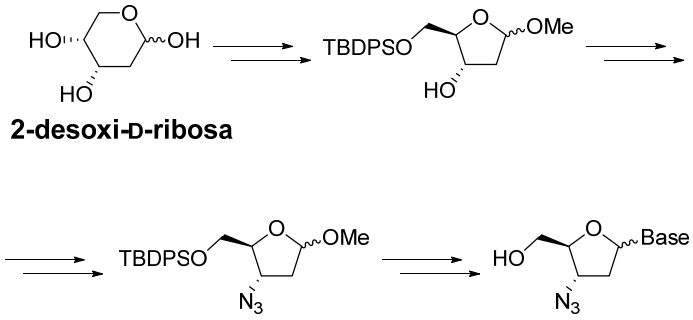
Figura 33. *Nucleósidos desprotegidos obtenidos derivados de 2,3-didesoxi-D-ribosa.*

Base nitrogenada	Configuración	Rendimiento
Uracilo	α	6%
	β	8%
Timina	α	4%
	β	7%
6-azauracilo	α	8%
	β	10%
5-fluorouracilo	α	5%
	β	9%
5-clorouracilo	α	6%
	β	11%
5-bromouracilo	α	5%
	β	7%
Citosina	α	7%
	β	8%
5-fluorocitosina	α	10%

	β	11%
5-azacitosina	α	7%
	β	7%

Tabla 4. *Rendimientos globales de los nucleósidos desprotegidos obtenidos.*

3.1.2 Síntesis de nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa

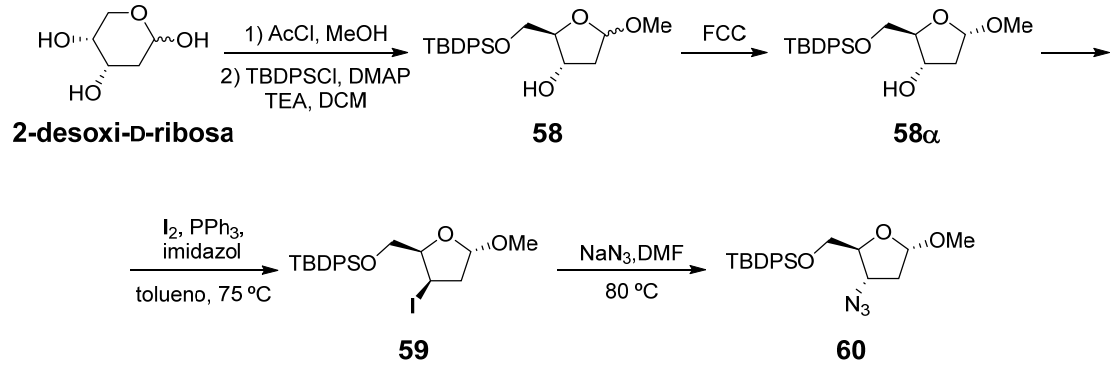


Basándonos en la alta actividad inhibitoria del AZT y otros derivados frente a la actividad de la RT de VIH nos propusimos como segundo objetivo sintetizar nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa. De entre las diversas formas de introducir el grupo azido en el anillo de 2,3-didesoxi-D-ribosa, escogimos la más corta disponible, aunque modificando las condiciones, que además utiliza como material de partida un compuesto quiral y ofrece buenos rendimientos [434].

3.1.2.1 Síntesis del precursor de nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa

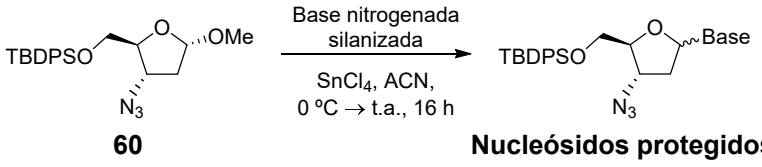
La ruta (**Figura 29**) parte de la 2-desoxi-D-ribosa que fue tratada ácido clorhídrico en metanol y a continuación protegido su hidroxilo en posición 5 empleando cloruro de *terc*-butildifenilsilano (TBDPSCl), trifenilfosfina e imidazol en DMF. Debido a que el rendimiento que se obtuvo fue bajo (40%, desde 2-desoxi-D-ribosa), se exploraron otras opciones. Entre estas, la sustitución de ácido clorhídrico por cloruro de acetilo [435] y realizar la silanización en DCM y en presencia de DMAP y TEA [436] supuso la obtención de la mezcla de anómeros **58** con un rendimiento del 60%, que si bien produjo una mejora, seguía sin ser la reportada (70%) [434]. Los anómeros fueron separados mediante FCC y obtenidos en una ratio (α/β) 1.9:1. El anómero **58α** fue el que se utilizó para continuar esta

ruta sintética, mientras que el anómero **58β** fue destinado a la síntesis de derivados 3-fluoro-2,3-didesoxi-D-ribosa por los motivos que se verán más adelante. Para obtener el compuesto con el grupo azida en la configuración correcta, es necesario llevar a cabo dos reacciones S_N2 consecutivas, la primera de intercambio de hidroxilo por yodo y la segunda de este último por azida. Para esto el compuesto **58α** se trató con trifenilfosfina, yoduro de metilo y azocarboxilato de dietilo (DEAD) en tolueno calentando la reacción a 110 °C obteniéndose **59** con un rendimiento del 42% (siendo el reportado 40%). Sin embargo, la combinación de yodo, trifenilfosfina e imidazol en tolueno [437] aumentó el rendimiento de la reacción hasta el 92%, lo que supuso una importante mejora de la ruta sintética. Tras la inversión de la configuración en el C3 tras el intercambio de hidroxilo por yodo, se procedió a la S_N2 de **59** con azida sódica en DMF para formar el precursor de nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa **60** (54%, similar al 52% reportado. Rendimiento del 30% desde 2-desoxi-D-ribosa) que contiene el grupo azida con la configuración deseada.



Esquema 29. *Síntesis del precursor de nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa.*

3.1.2.2 Acoplamiento entre el precursor 60 y las bases nitrogenadas



Esquema 30. *Reacción de acoplamiento entre **60** y las bases nitrogenadas silanizadas.*

Una vez obtenido el precursor de nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa **60**, se procedió a la reacción de acoplamiento (**Esquema 30**) entre este y las 9 bases nitrogenadas previamente mostradas y silanizadas como se

ha visto. Las condiciones empleadas en la síntesis de nucleósidos derivados de 2,3-didesoxi-D-ribosa no condujeron a la obtención de tales compuestos en este caso o lo hicieron con muy bajo rendimiento, como tampoco lo hizo el empleo de TMSOTf en acetonitrilo anhidro [438]. En cambio, el uso de $SnCl_4$ (1 M en DCM) en ACN anhidro [439] logró llevar a cabo la reacción de acoplamiento obteniéndose mezclas de estereoisómeros como pudo detectarse mediante LC-MS de baja resolución. Aunque se obtuvieron los nucleósidos protegidos deseados, el rendimiento (TLC) no resultó satisfactorio. La silanización de la base con bis(trimetilsilil)acetamida (BSA) a reflujo en ACN anhidro y la adición de **60** y $SnCl_4$ sobre la reacción anterior, una vez enfriada a 0 °C [440], generó los nucleósidos protegidos en un rendimiento aparentemente mayor (TLC) que con el proceso anteriormente mencionado. Ninguna de las mezclas obtenidas (**Figura 34**) fue posible separarla en sus estereoisómeros puros mediante TLC o FCC independientemente de la fase móvil empleada. Debido a ello se procedió a la desprotección de los nucleósidos obtenido a fin de que su separación fuera factible una vez hecho.

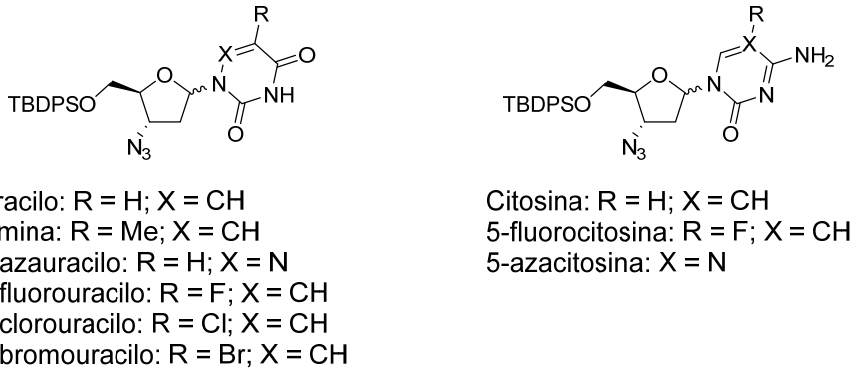
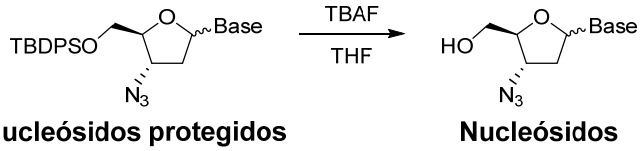


Figura 34. *Estructuras de las mezclas de estereoisómeros de los nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa obtenidos*

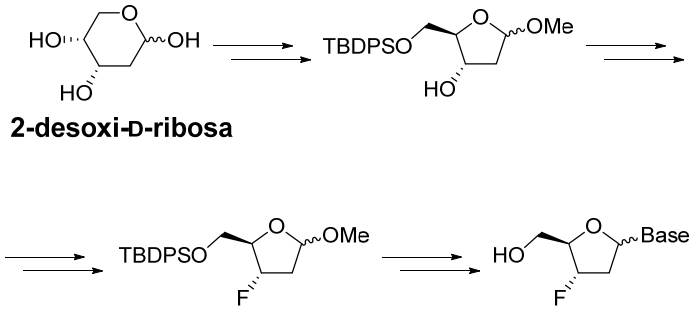
3.1.2.3 Desprotección del hidroxilo en posición 5' de los nucleósidos protegidos derivados de 3-azido-2,3-didesoxi-D-ribosa



Esquema 31. *Reacción de desprotección de los nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa.*

La desprotección de estos derivados transcurrió como se esperaba con las condiciones empleadas en la anterior familia y se obtuvieron las mezclas de estereoisómeros desprotegidos con un rendimiento cuantitativo (**Esquema 31**). Estas mezclas de nucleósidos desprotegidos sí pudieron ser separadas en su respectivos estereoisómeros mediante TLC a través del uso de distintas fases móviles, aunque aún no han podido ser caracterizados.

3.1.3 Síntesis de nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa

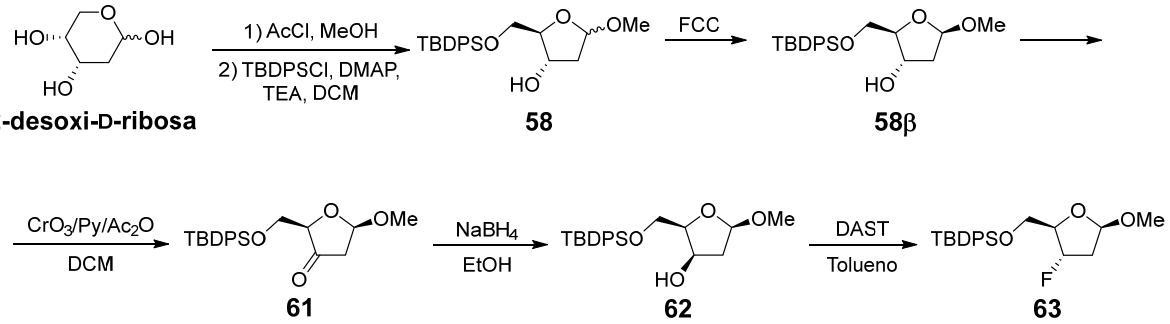


Los nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa han mostrado tener una alta actividad inhibidora de RT de VIH (revisado en [441]). De hecho, el cambio de azida por flúor en la molécula de AZT conlleva un aumento de la actividad pero también de la toxicidad [442], [443] y en general esta familia de compuestos ha demostrado ser útil como terminadores de cadena [444]. Además, comparte con la ruta anterior los primeros pasos de reacción. Por todo ello, nos propusimos sintetizar nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa.

3.1.3.1 Síntesis del precursor de nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa

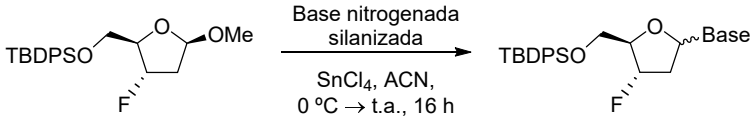
Aprovechando la ruta sintética anterior, que permitía obtener nucleósidos derivados de 3-azido-2,3-didesoxi-D-ribosa, se escogió para la síntesis del precursor de nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa una ruta con la que comparte los primeros pasos (**Esquema 32**) [445]. La metilación y subsiguiente protección del hidroxilo libre en C5 de manera análoga a la anterior condujo a la obtención de la mezcla de anómeros **58**. Una vez separados, se tomó el anómero **58β** para continuar la ruta. Aunque el grupo hidroxilo puede intercambiarse por

flúor, este quedaría con una configuración indeseada, siendo el objetivo que dicho grupo tenga la misma configuración que el hidroxilo en C3' de los nucleósidos naturales. Para ello, el hidroxilo en C3 fue a continuación oxidado con un complejo de óxido de cromo (VI)/piridina/anhídrido acético (1:2:1) recién preparado en DCM (oxidación de Collins con anhídrido acético) que resulta especialmente útil en la síntesis de nucleósidos ya que implica unas condiciones de reacción suaves, sobre todo con la adición de anhídrido acético que además acelera la reacción al facilitar la reducción del cromo (VI) [446], [447]. Estas condiciones aportaron un rendimiento significativamente mayor (TLC) que el obtenido mediante el uso de clorocromato de piridinio (PCC). De esta forma se obtuvo el compuesto **61** que debido a su reportada inestabilidad fue sometido sin purificar a una reducción estereoselectiva con borohidruro sódico en etanol. La adición de hidruro al C3 se produce por la cara α del derivado de 3-fluoro-2,3-didesoxi-D-ribosa al estar menos impedida obteniéndose rápidamente **62** de forma muy mayoritaria o casi en exclusiva, con total desaparición del producto de partida. Con el grupo hidroxilo en C3 con la configuración apropiada, se llevó a cabo el intercambio de dicho grupo por flúor con inversión de la configuración. Esta reacción se realiza con trifluoruro de dietilaminoazufre (DAST) como donador de fluoruros y permite obtener el precursor de nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa **63** aunque con un rendimiento del 25% (frente al 43% reportado) para esta última reacción, lo que hace deseable explorar otras opciones que permitan sintetizar **63** con mejores rendimientos (como utilizar una disolución comercial de DAST 1 M en DCM).



Esquema 32. *Síntesis del precursor de nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa.*

3.1.3.2 Acoplamiento entre el precursor 63 y las bases nitrogenadas



63 **Nucleósidos protegidos**

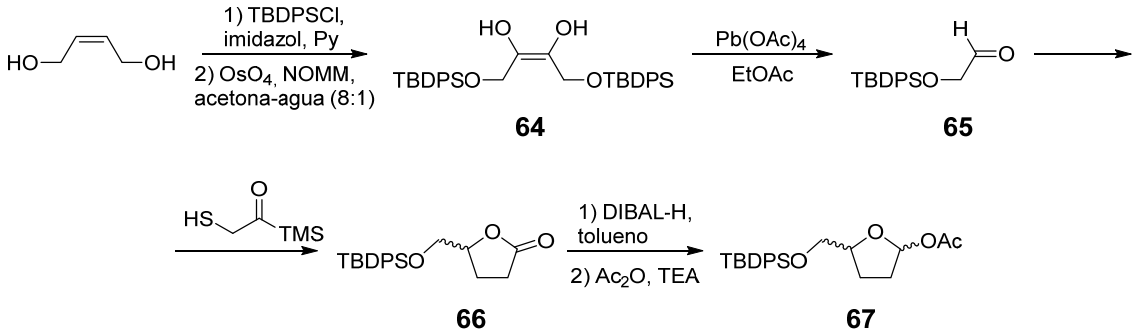
Esquema 33. *Reacción de acoplamiento entre **63** y las bases nitrogenadas silanizadas.*

Para formar los nucleósidos derivados de 3-fluoro-2,3-didesoxi-D-ribosa se llevó a cabo la reacción de acoplamiento (**Esquema 33**) en las mismas condiciones que para los derivados de 3-azido-2,3-didesoxi-D-ribosa, las cuales condujeron a la obtención de nucleósidos protegidos.

3.1.4 Síntesis de nucleósidos derivados de 1,3-oxatiolano

Habida cuenta de la potente actividad inhibidora de RT de VIH por parte de los L-nucleósidos derivados de 1,3-oxatiolano, especialmente emtricitabina y lamivudina, se planteó su síntesis. La síntesis quiral fue abordada en nuestro laboratorio pero requiere más de 20 pasos de reacción [384] con un rendimiento global demasiado bajo para nuestros propósitos, de modo que se intentó por otras rutas. Todas las rutas quirales precisan un número muy alto de reacciones, de modo que se escogió una ruta que permitía obtener mezclas racémicas para posteriormente resolverlas.

Esta síntesis (**Esquema 34**) comienzó transformando 2-buten-1,4-diol en el disililéter correspondiente con TBDPSCl para a continuación fragmentar la molécula por ozonolisis obteniéndose el aldehído **64**. Sin embargo, este aldehído se obtuvo con mejores rendimientos añadiendo un paso intermedio en el que el alqueno se dihidroxiló con tetróxido de osmio (**65**) y posteriormente fue sometido a una fragmentación con tetraacetato de plomo. El aldehído **64** se hace reaccionar con ácido tioglicólico para dar lugar a la tiolactona **66**, que en presencia de DIBAL-H y posterior acetilación permitió obtener el precursor de nucleósidos derivado de 1,3-oxatiolano **67** como una mezcla racémica.



Esquema 34. *Síntesis del precursor de nucleósidos derivados de 1,3-oxatiolano.*

Obtenido el precursor de nucleósidos derivados de 1,3-oxatiolano como una mezcla racémica, se procedió a la aplicación de algunas técnicas de resolución de enantiómeros sobre esta mezcla racémica o sobre **67**. Al haber sido infructuosos estos esfuerzos, se comentarán someramente. En primer lugar, **67** fue desprotegido con TBAF para dar el alcohol correspondiente con la intención de acetilarlo enzimáticamente con acetato de vinilo y Chirazyme L-2, c.f., C-2 [448]. Esta reacción se llevó a cabo en distintas mezclas de éteres como disolventes ya que afectan a la enantioselectividad de la reacción. Sin embargo, ninguna de las condiciones llevó a la obtención de un producto de reacción enantioméricamente puro cuando fue analizado por un cromatógrafo de gases provisto con una columna quiral, como tampoco lo hizo la desacetilación enzimática. Por otro lado, el compuesto **67** desprotegido se esterificó con anhídrido diacetil-L-tartárico y con anhídrido dibenzoil-L-tartárico [449], [450]. Aunque la reacción transcurrió sin problemas, los distintos estereoisómeros generados no pudieron ser aislados mediante FCC, al menos en las cantidades necesarias para que resultase un sistema práctico. A la vista de los resultados se hace patente la necesidad de encontrar otra estrategia que permita obtener nucleósidos derivados de esta familia con la configuración del C2 fijada.

3.2 EVALUACIÓN DE LA INHIBICIÓN DE LA MOVILIDAD DE RETROTRANSPOSONES EMPLEANDO NUCLEÓSIDOS Y DETERMINACIÓN DE LA TOXICIDAD DE LOS COMPUESTOS ENSAYADOS

Con el objetivo de encontrar compuestos nucleosídicos que inhiban selectiva y eficientemente la movilidad del retroelemento LINE-1 mediante su interacción con la RT, se evaluó la capacidad de distintas estructuras nucleosídicas para inhibir la retrotransposición empleando para ello un ensayo de retrotransposición previamente establecido [44], [116], [285], [451]. Para dichos ensayos se utilizaron células HeLa, las cuales presentan poca o ninguna expresión endógena de LINE-1 [452] pero soportan la retrotransposición de retroelementos transfectados, además de haber sido utilizadas previamente en numerosos ensayos de retrotransposición. Para determinar la capacidad inhibidora de estos compuestos se empleó el LINE-1 humano L1.3, clonado en el plásmido JM101/L1.3 (**Human L1.3**) [453]. Para determinar la selectividad de los compuestos ensayados se emplearon retrotransposones LTR de ratón: MusD, contenido en el plásmido pCMV-MusD-6neo^{TNF} (**Mouse MusD**) [454] e IAP, contenido en el plásmido pIAP-92L23neo^{TNF} (**Mouse IAP**) [455] (**Figura 35A**). Simultáneamente, se llevaron a cabo ensayos de clonabilidad/toxicidad a largo plazo, empleando para ello el plásmido pU6ineo [191] que contiene el gen de resistencia a neomicina (sin intrón) por lo que el posible efecto tóxico observado sería debido únicamente al nucleósido añadido al medio de cultivo. Estos ensayos preliminares para evaluar la capacidad inhibidora de los compuestos se llevaron a cabo a dos concentraciones, 5 y 25 μM.

Para realizar los ensayos de retrotransposición y de clonabilidad/toxicidad a largo plazo se sembraron células HeLa, las cuales fueron transfectadas con los plásmidos indicados. A las 24 horas se les cambió el medio de cultivo añadiendo los RTis indicados a las concentraciones mencionadas y fueron mantenidos durante el resto del ensayo. A las 72 horas post-transfección se comenzó la selección con G-418, la cual se aplicó durante 14 días, tras los cuales las colonias

fueron fijadas y teñidas para ser contabilizadas (**Figura 35B**). Todos los ensayos de retrotransposición realizados se basan en el mismo principio: cada retroelemento tiene clonado en su región 3'UTR el gen de resistencia a neomicina (neomicina fosfotransferasa, *mneoI* [285]) en orientación antisentido con respecto al promotor de LINE-1, pero está interrumpido por un intrón en el mismo sentido transcripcional que el promotor de LINE-1. Esto significa que el gen de resistencia a neomicina solo puede expresarse, desde su propio promotor, una vez eliminado dicho intrón mediante splicing del tránscrito generado desde el promotor de LINE-1 y subsiguiente inserción en el genoma (**Figura 35C**). Resumidamente, si se produce retrotransposición desde el retroelemento presente en el plásmido, habrá un mayor número de colonias ya que son resistentes a la neomicina (o G418) adicionada al medio de cultivo al haber incorporado en su genoma el gen de resistencia a dicho antibiótico, mientras que si se inhibe la retrotransposición habrá un menor número de colonias resistentes a neomicina al no haberlo incorporado. Los resultados son cuantitativos, de modo que el número de colonias resultantes es proporcional a la inhibición de la retrotransposición.

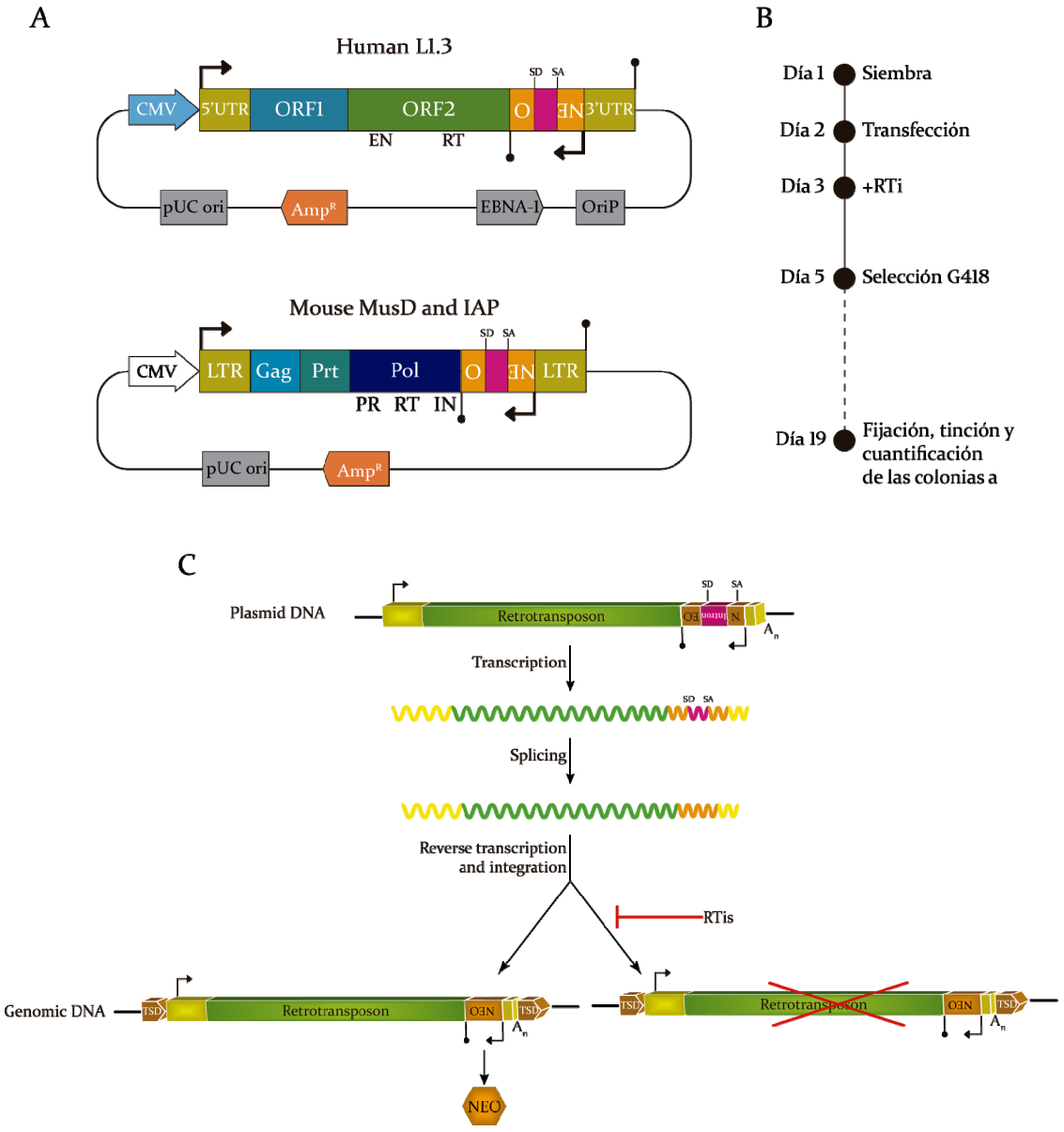


Figura 35. Ensayo de retrotransposición con el gen de resistencia a neomicina como gen reportero de la movilidad. A. *Representación de los plásmidos empleados en los ensayos de retrotransposición con el gen de resistencia a neomicina como gen reportero. Las principales características de cada elemento están representadas, pero no a escala. La expresión del retroelemento* ***IAP*** *depende de su promotor interno, a diferencia de* ***MusD*** *que lo hace desde el promotor CMV (flecha azul pálido). Ver* ***Figura 2*** *para más información sobre la estructura de los retrotransposones. Neo, gen de resistencia a neomicina; pUC ori, origen de replicación bacteriano; Amp^R, gen de resistencia a ampicilina; EBNA-1 y OriP, antígeno nuclear y origen de replicación eucariota.* ***B.*** *Esquema del desarrollo temporal de los ensayos con el gen de resistencia a neomicina como gen reportero.* ***C.*** *Esquema del proceso de retrotransposición desde el retroelemento marcado con el gen reportero presente en el plásmido. El gen de resistencia a neomicina se puede expresar tras un evento de retrotransposición. En presencia de RTis, la inhibición de la movilidad impide su expresión. Ver* ***Figura 9*** *para más información.*

Además se evaluó la toxicidad aguda mediante el ensayo MTT [456]. Para ello, se sembraron células HeLa en una placa de 96 pocillos y a las 24 horas se les

cambió el medio conteniendo el RTi indicado a las mismas concentraciones que los ensayos de retrotransposición (5 y 25 μM). A las 72 horas (**Figura 37 inferior derecha**) se añade el reactivo MTT (una sal de tetrazolio) que es transformado por enzimas oxidoreductasas dependientes de NADH a formazán, insoluble en el medio de cultivo. Esta transformación depende de la actividad metabólica de las células de modo que a mayor número de células mayor será la formación de formazán. Este compuesto es de color morado por lo que una vez disuelto en dimetilsulfóxido puede medirse su absorbancia, que será proporcional al número de células de cada pocillo.

En nuestra búsqueda de inhibidores potentes y selectivos de LINE-1, tres grupos de compuestos nucleosídicos fueron ensayados: fármacos comercializados frente a RT de VIH, compuestos con estructura nucleosídica obtenidos de fuentes comerciales y 2',3'-didesoxi-D-ribonucleósidos sintetizados.

3.2.1 Evaluación de la inhibición de la de la movilidad de retrotransposones de fármacos comercializados frente a RT de VIH (y toxicidad)

El primer grupo testado fue el de fármacos que se emplean en el tratamiento del SIDA y que ejercen su efecto mediante la inhibición de su RT, lo que en un principio podría significar mayor facilidad de encontrar una molécula inhibidora de RT de LINE-1 (además de los encontrados previamente que han sido comentados más arriba). De hecho, así se confirmó en estudios preliminares anteriores en nuestro laboratorio. Entre estos fármacos se utilizaron 6 NRTIs (**abacavir, AZT, emtricitabina, lamivudina, stavudina** y **tenofovir**) y 3 NNRTIs (**efavirenz, etravirina** y **nevirapina**) (**Figura 36**), estos últimos como controles negativos ya que al unirse a un sitio alostérico de la RT de VIH difícilmente podrían inhibir la RT de LINE-1 (de hecho, no son activos o presentan muy baja actividad frente VIH-2).

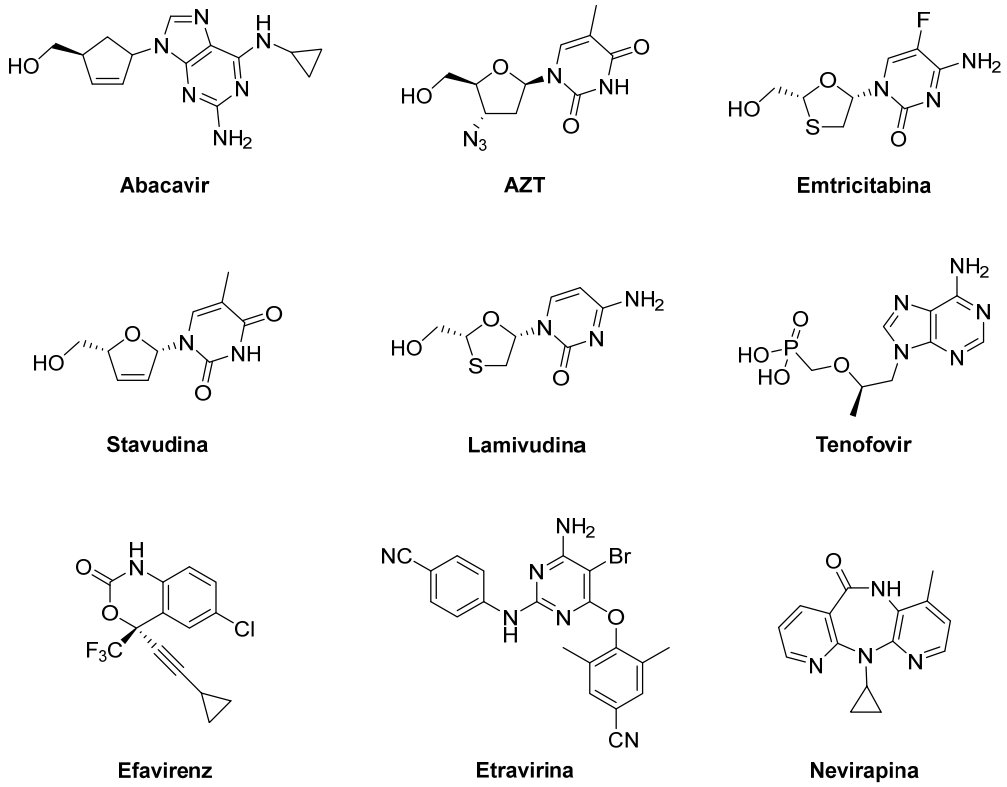


Figura 36. *Fármacos comercializados frente a RT de VIH empleados en este estudio.*

El resultado de la evaluación de estos compuestos mostró que todos los NRTIs eran activos inhibiendo la retrotransposición de Human L1.3. Sin embargo, estos datos han de ser corregidos, para obtener los valores de inhibición reales, empleando los datos de cuantificación del ensayo de clonabilidad/toxicidad a largo plazo (en adelante se simplificará a clonabilidad) con el plásmido pU6ineo (**Figura 37**). De hecho, algunos compuestos revelaron ser tóxicos para estas células en el ensayo de clonabilidad de una forma dosis dependiente. **AZT** y **efavirenz** produjeron una leve toxicidad a 5 μM y más marcada a 25 μM mientras que la **etravirina** resultó altamente tóxica a ambas concentraciones. El ensayo de MTT (**Figura 38**) para esta serie de compuestos arrojó unos resultados similares a los de clonabilidad, revelando cierta toxicidad del **tenofovir** a 25 μM.

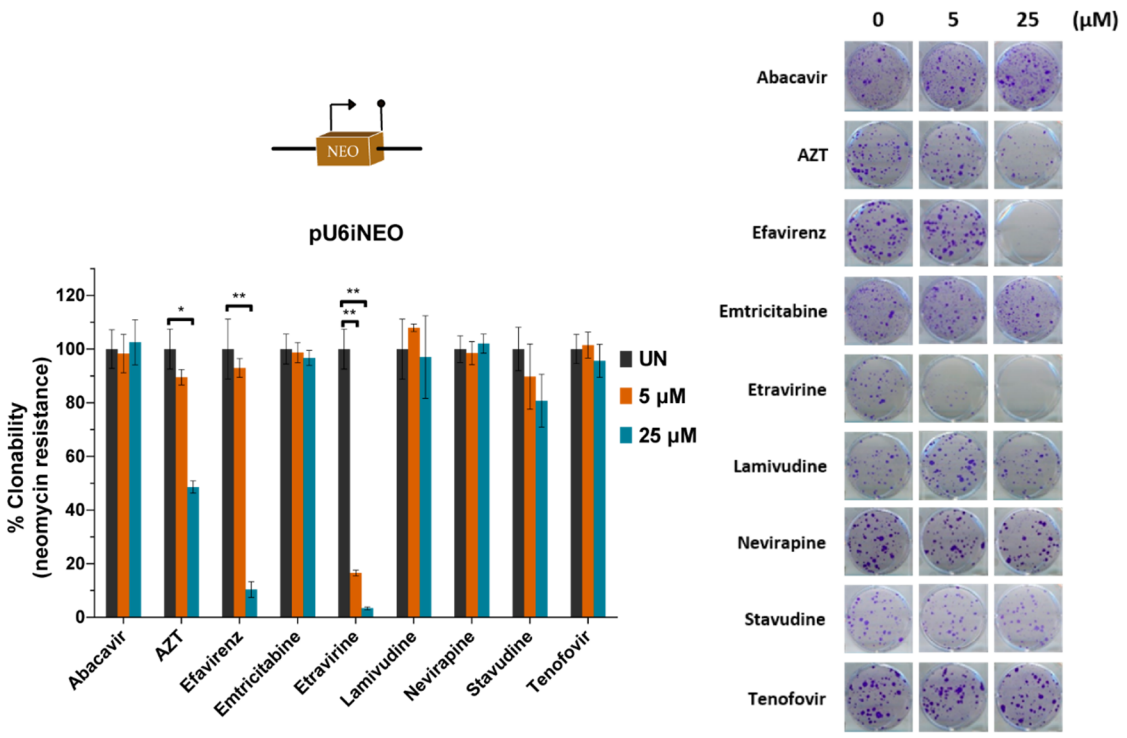


Figura 37. *Resultado y cuantificación del ensayo de clonabilidad en células HeLa de los fármacos comercializados frente a RT de VIH.*

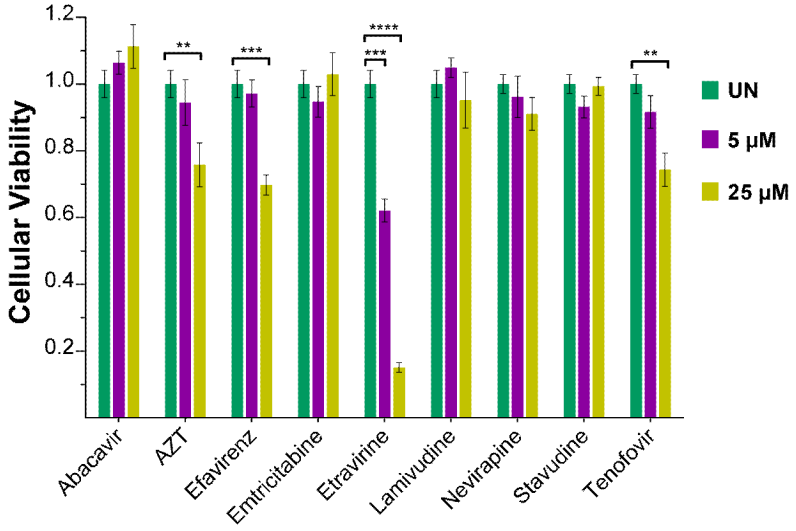


Figura 38. *Viabilidad celular de HeLa para los fármacos comercializados frente a RT de VIH mediante el ensayo MTT.*

Una vez corregidos los datos para Human L1.3 (**Figura 39**), seguía observándose una importante inhibición por parte del **abacavir** a 25 µM y de **AZT**, **emtricitabina**, **lamivudina**, **stavudina** y **tenofovir** de forma dosis dependiente, mientras que **nevirapina** produce una inhibición moderada independiente de la dosis. La **emtricitabina** y la **lamivudina** destacan entre los compuestos inhibidores por su ausencia de toxicidad, tanto a corto como a largo plazo, así como

por su potencia ya que reducen la retrotransposición un 65% y 60% respectivamente a 5 µM y 85% y 70% respectivamente a 25 µM. El **tenofovir** también mostró una potencia muy elevada con valores cercanos al 100% de inhibición, aunque también cierta toxicidad aguda. A excepción de la **nevirapina**, los NNRTIs no surtieron efecto alguno sobre la retrotransposición de Human L1.3, como por otra parte se esperaba. En general, y teniendo en cuenta las diferencias experimentales, los datos obtenidos son coherentes con los anteriormente publicados [291], [293].

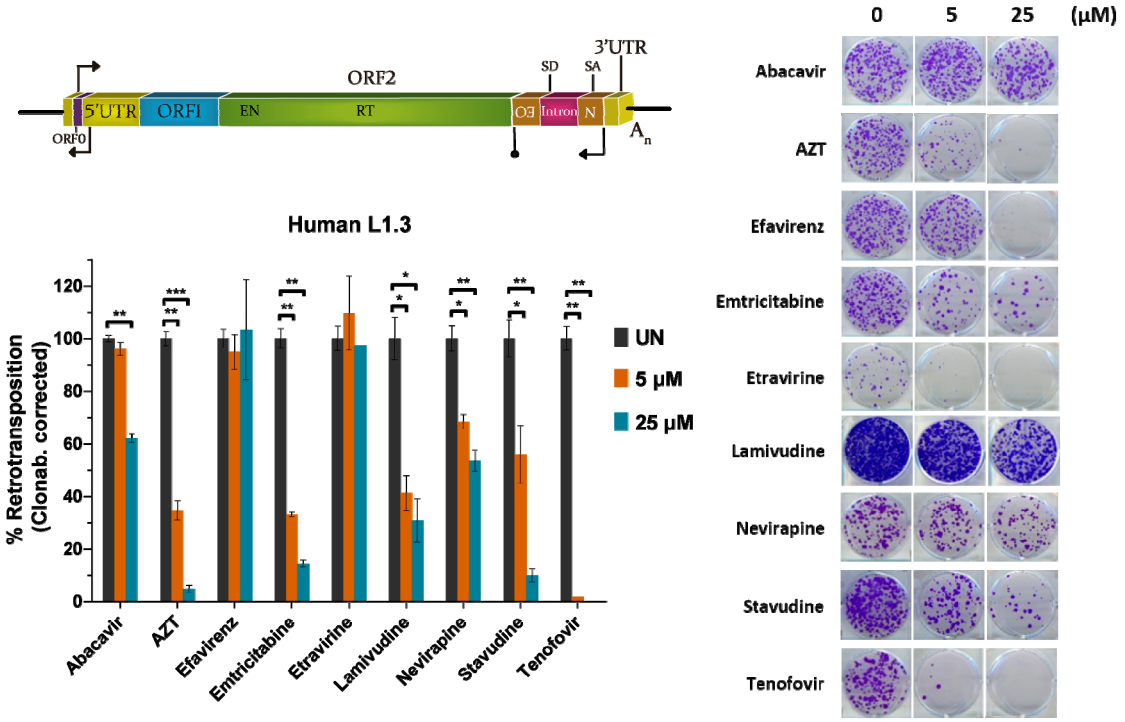


Figura 39. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los fármacos comercializados frente a RT de VIH con Human L1.3.*

Para determinar la selectividad de los compuestos anteriores para inhibir Human L1.3, se ensayaron estos mismos empleando para ello los retrotransposones LTR Mouse MusD y Mouse IAP (ERVs). Usando de nuevo los valores del ensayo de clonabilidad, se cuantificó la inhibición de la retrotransposición que se produjo usando los retroelementos mencionados (**Figura 40** y **41**). Así, se pudo determinar que **abacavir**, **nevirapina** y **stavudina** son capaces de inhibir modestamente a 25 µM la retrotransposición de Mouse MusD mientras que **AZT** y **tenofovir** mostraron una potente inhibición dosis dependiente de ambos LTRs.

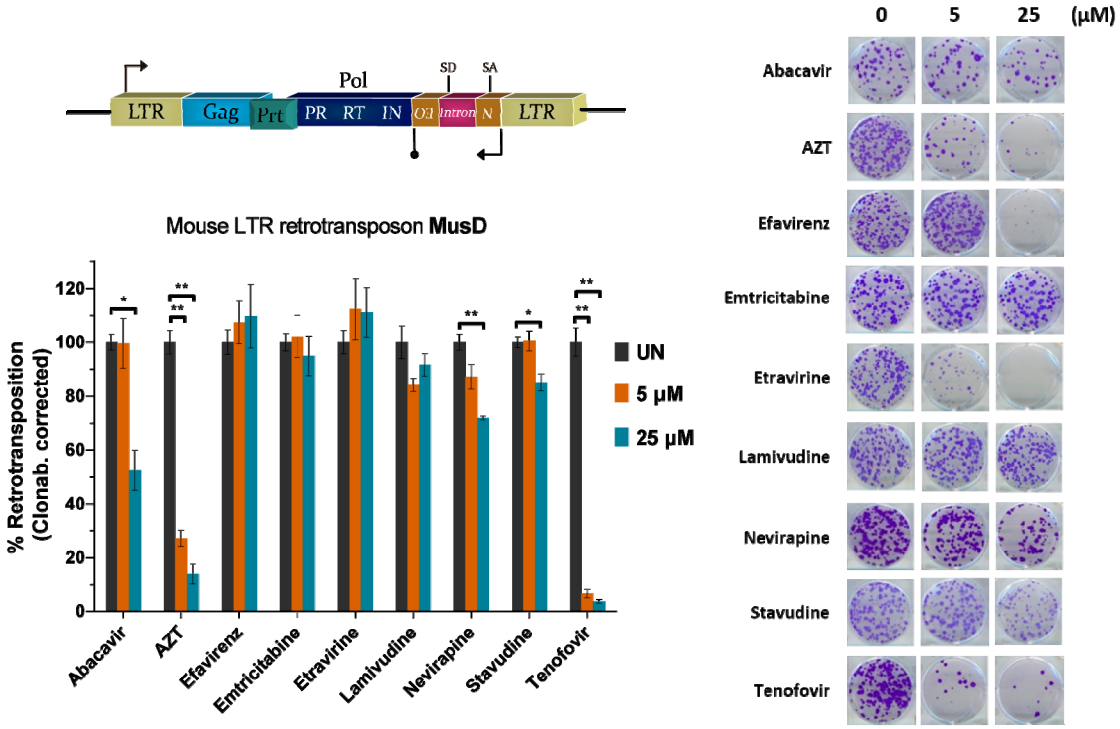


Figura 40. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los fármacos comercializados frente a RT de VIH con Mouse MusD.*

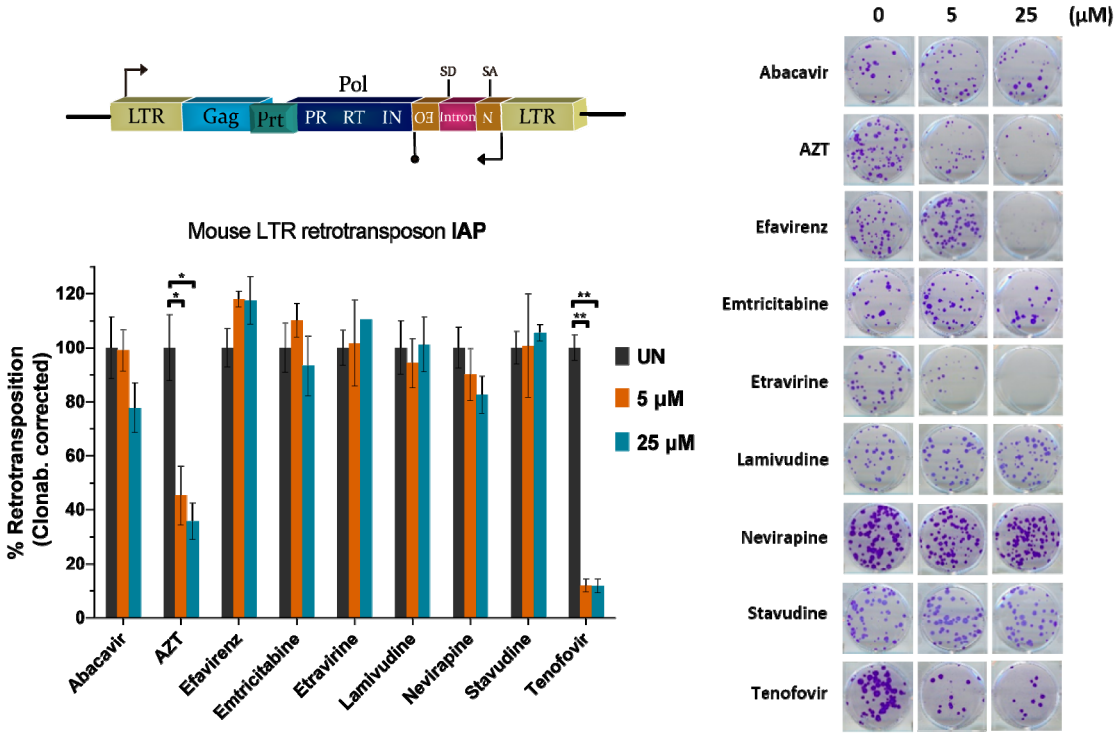


Figura 41. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los fármacos comercializados frente a RT de VIH con Mouse IAP.*

Analizado el grupo de fármacos inhibidores de RT comercializados para el tratamiento del SIDA, **AZT** y **tenofovir** mostraron ser inhibidores potentes, pero

no selectivos, de Human L1.3, además de ser tóxicos en células HeLa (sobre todo **AZT**). La **stavudina** también resultó ser relativamente selectiva al mostrar una ligera actividad significativa sobre Mouse MusD a 25 µM, además de algo tóxica. Sin embargo, **emtricitabina** y **lamivudina** destacaron de entre estos fármacos como inhibidores potentes y selectivos de Human L1.3, lo que hizo que fueran seleccionados para llevar a cabo futuros experimentos.

3.2.2 Evaluación de la inhibición de la de la movilidad de retrotransposones de compuestos con estructura nucleosídica obtenidos de fuentes comerciales (y toxicidad)

Para ampliar la variabilidad estructural del presente estudio, se adquirieron 9 compuestos con estructura nucleosídica de distintas casas comerciales (**Figura 42**).

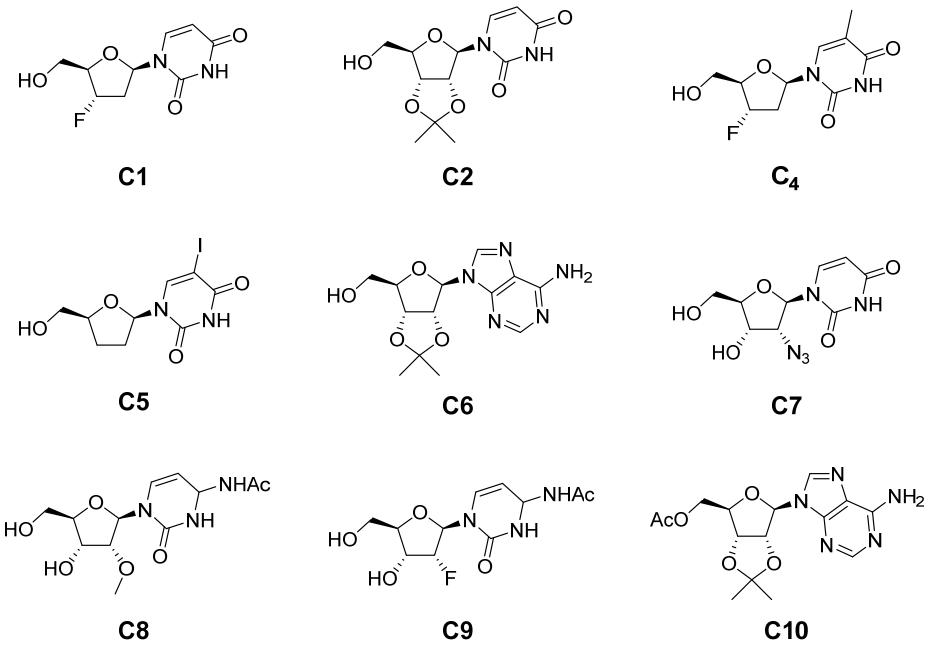


Figura 42. *Compuestos con estructura nucleosídica obtenidos de fuentes comerciales empleados en este estudio.*

Estos compuestos fueron sometidos a los mismos ensayos de retrotransposición, clonabilidad y MTT que la anterior familia de compuestos. En los ensayos de clonabilidad, **C1** resultó ser tóxico a 5 µM y **C4** altamente tóxico a

las dos concentraciones ensayadas de manera dosis no dependiente (**Figura 43**). Por otra parte, ninguno estos compuestos mostraron toxicidad en el ensayo de MTT (**Figura 44**), a la vez que sí lo hicieron **C2**, **C8** y **C9**, aunque de forma discreta. Tras emplear estos datos para corregir la cuantificación de los ensayos de retrotransposición, se pudo comprobar que **C4** produjo una inhibición dosis dependiente en Human L1.3 (**Figura 45**) y LTRs de ratón (**Figura 46** y **47**), al igual que **C9** inhibió Human L1.3 aunque en un grado sumamente menor.

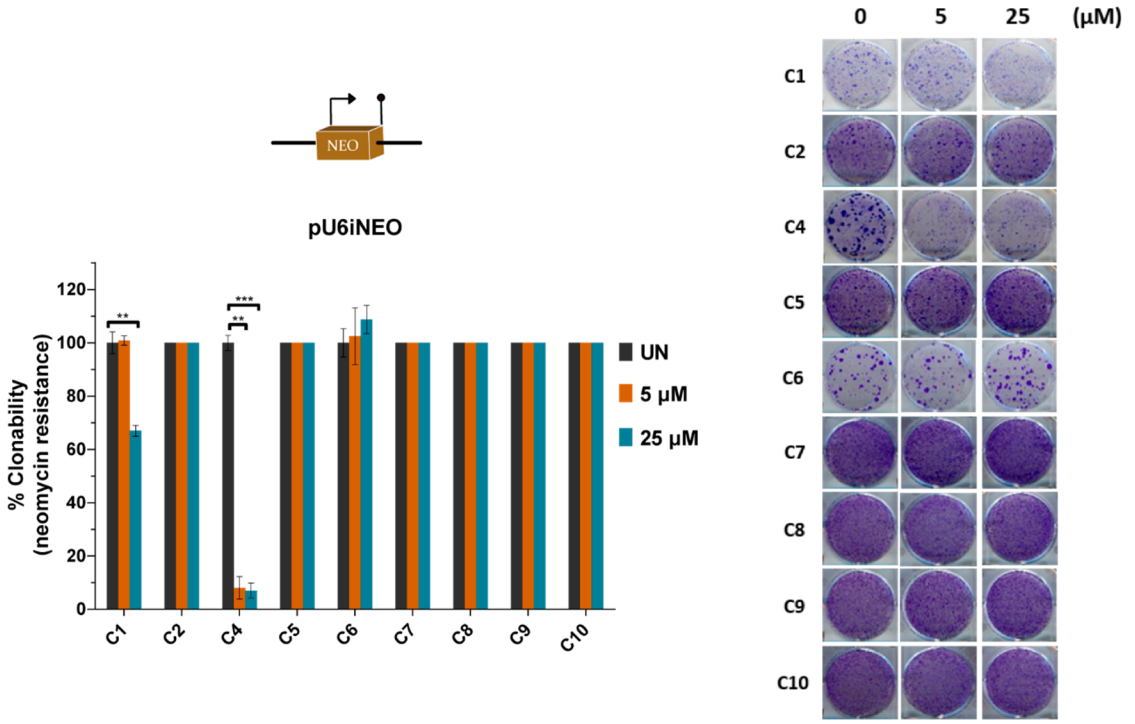


Figura 43. *Resultado y cuantificación del ensayo de clonabilidad en células HeLa de compuestos con estructura nucleosídica obtenidos de fuentes comerciales.*

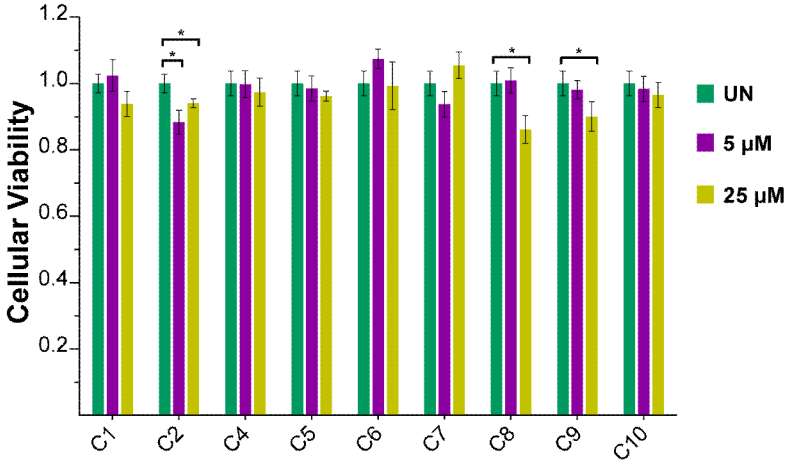


Figura 44. *Viabilidad celular de HeLa para los compuestos con estructura nucleosídica obtenidos de fuentes comerciales mediante el ensayo MTT.*

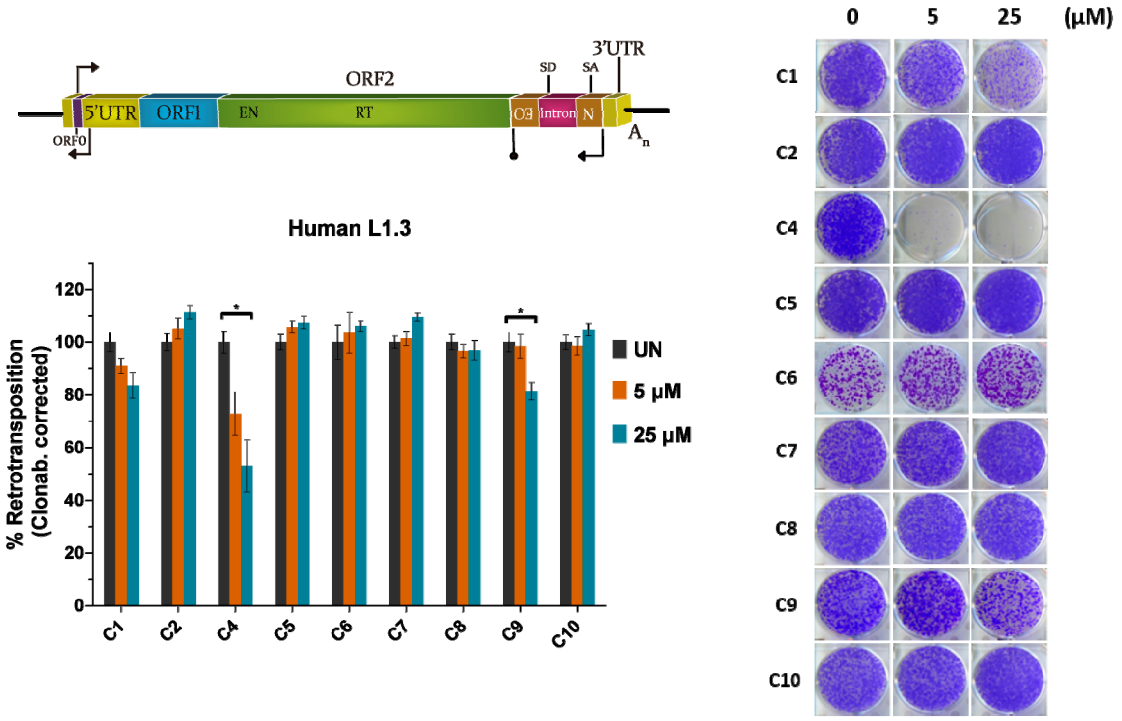


Figura 45. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los compuestos con estructura nucleosídica obtenidos de fuentes comerciales con Human L1.3.*

El compuesto **C4** (o alovudina) fue una molécula que alcanzó la fase II de ensayos clínico para ser usado en el tratamiento del SIDA como NRTI, si bien debido a su alta toxicidad fue descartado para tal propósito [457]. Esto es consistente con la observación de que inhibe Human L1.3 y LTRs de ratón, a la vez que resulta altamente tóxico. Las estructuras con 3'-OH (**C7**, **C8** y **C9**) no mostraron ningún efecto sobre la retrotransposición, probablemente debido a que al tener presente y libre dicho grupo podrían incorporarse en teoría por la RT a la hebra de DNA creciente en el punto de inserción genómico y a continuación añadir el siguiente nucleótido uniéndolo al 3'-OH de estos compuestos. Los compuestos con 2',3'-isopropilidenen (**C2**, **C6**, **C10**) en principio pueden actuar como terminadores de cadena al carecer del grupo 3'-OH. Sin embargo, su actividad depende de ser fosforilados enzimáticamente y de su posterior interacción con la RT para ser adicionados a la nueva hebra, lo que puede estar impedido estéricamente en estos casos por el grupo 2',3'-isopropiliden.

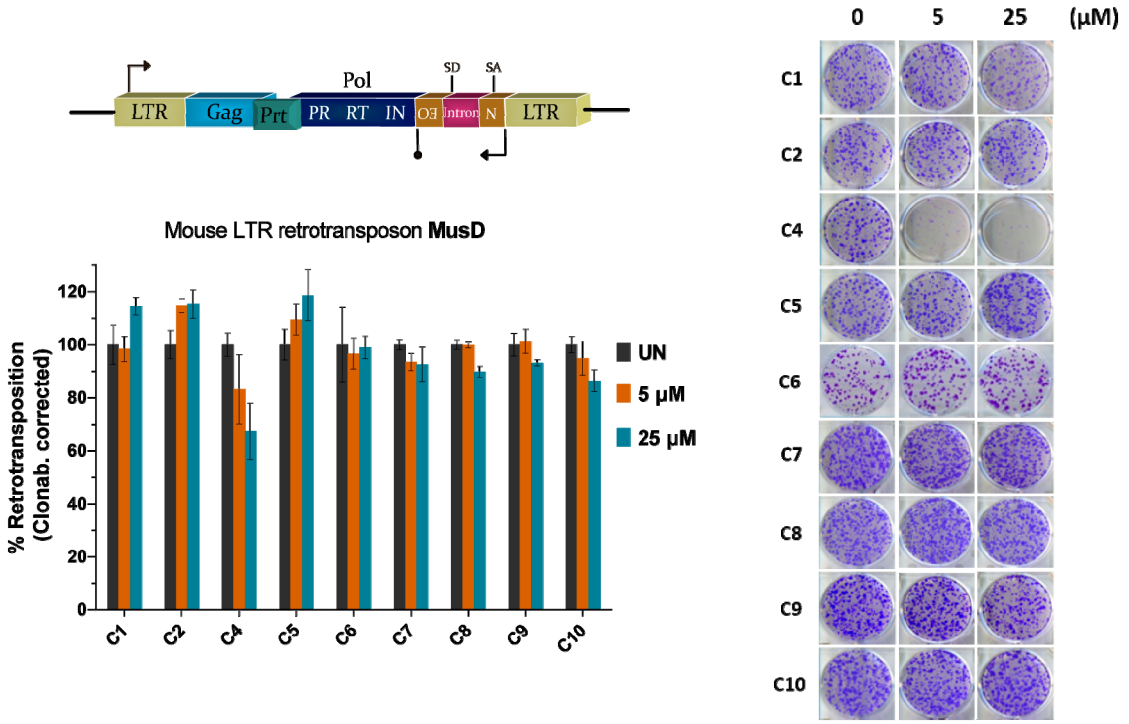


Figura 46. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los compuestos con estructura nucleosídica obtenidos de fuentes comerciales con Mouse MusD.*

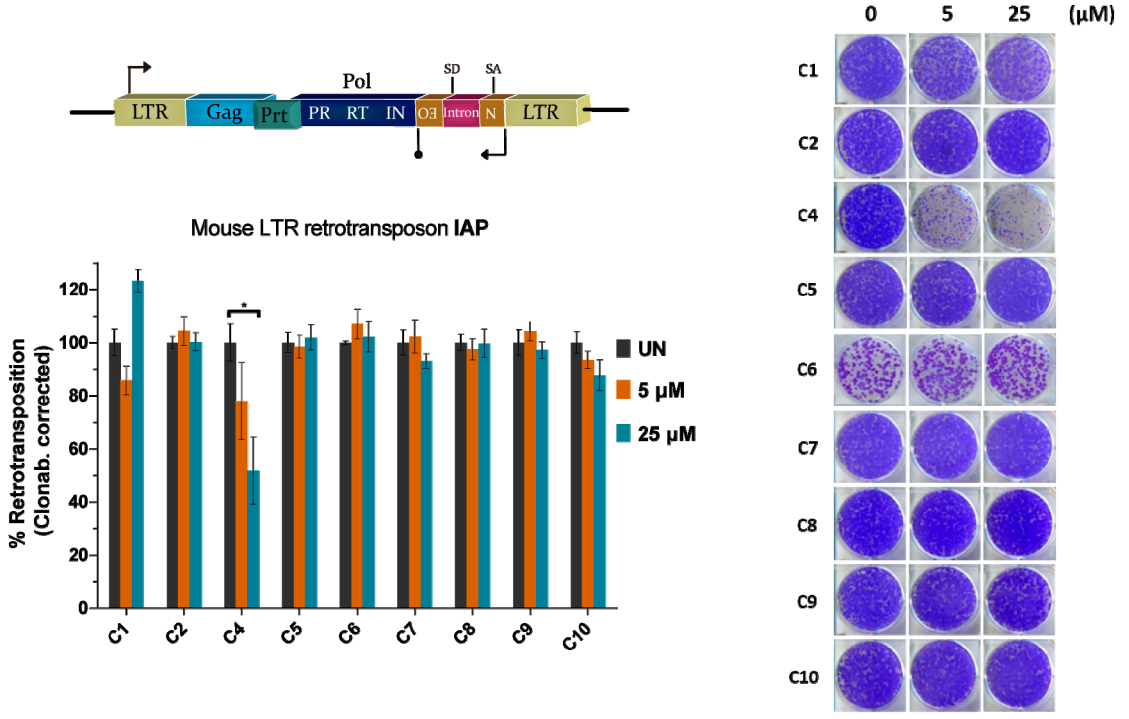


Figura 47. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los compuestos con estructura nucleosídica obtenidos de fuentes comerciales con Mouse IAP.*

Entre las moléculas de este grupo, ninguna estructura resultó útil puesto que el único inhibidor con potencia suficiente para ser interesante sería el **C4**, si bien su falta de selectividad y su importante toxicidad a largo plazo lo descartaron para futuros experimentos.

3.2.3 Evaluación de la inhibición de la de la movilidad de retrotransposones de los nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados (y toxicidad)

Con el objetivo de expandir aún más los límites del cribado llevado a cabo en este trabajo, se sintetizaron nucleósidos derivados de 2',3'-didesoxi-D-ribosa (como se ha descrito más arriba) para su evaluación como inhibidores de Human L1.3. Un total de 18 nucleósidos, 9 parejas de estereoisómeros α y β, fueron sintetizados y testados, todos conteniendo bases nitrogenadas derivadas de pirimidina (**Figura 48**). Algunos de estos compuestos han sido previamente sintetizados y han sido referenciados en Materiales y Métodos.

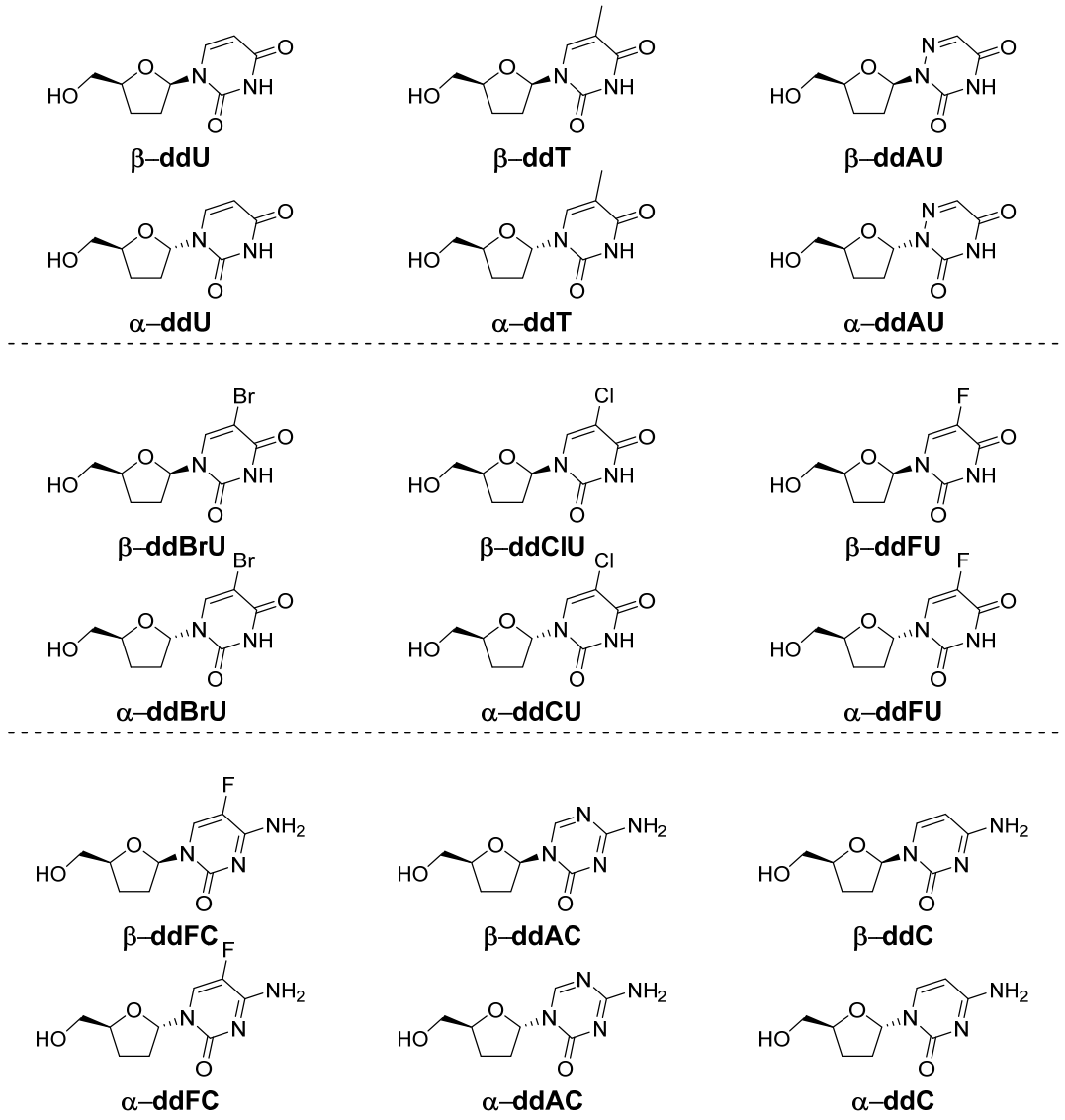


Figura 48. *Nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados empleados en este estudio.*

Los ensayos de MTT (**Figura 49**) mostraron que varios de los compuestos de esta familia exhibían toxicidad aguda en células HeLa, la cual osciló entre leve-moderada para **β-ddU** y **β-ddT** y moderada-alta para **α-ddAU**, **β-ddAU**, **α-ddFU**, **β-ddFU**, **β-ddC** y **β-ddFC**. Todos los compuestos que presentaban una toxicidad aguda moderada-alta resultaron ser también tóxicos en el ensayo de clonabilidad de una forma dosis dependiente (**Figura 50**), de nuevo con valores de toxicidad elevados. Algunos de estos efectos tóxicos han sido descritos previamente. **α-ddFU** y **β-ddFU** han demostrado ser tóxicos en células HeLa [458] y por otra parte, como se ha explicado anteriormente, los nucleósidos en el citoplasma de la célula pueden hidrolizarse en el derivado de furanosa y la base nitrogenada, dando en este caso 5-fluorouracilo. Conocido desde hace décadas por su actividad anticancerígena, el

5-fluorouracilo, a través de dos de sus metabolitos, actúa inhibiendo la timidilato sintasa por un lado, y compitiendo por el uracilo durante la síntesis de RNA [459]. En el caso de **β-ddC** (o zalcitabina), este fue el tercer fármaco aprobado para el tratamiento del SIDA, pero fue retirado por los graves efectos adversos que producía. De entre estos efectos adversos destaca la inhibición de la DNA polimerasa γ (mitocondrial) la cual tiene una elevada actividad en células tumorales por lo que tiene una acción anticancerígena, lo que a su vez explica la toxicidad en células HeLa [460]. La toxicidad de **β-ddFC** se puede explicar por un mecanismo similar al de **ddFU**, ya que la enzima citidina deaminasa transforma la citidina en uridina (y también la **ddFC** en **ddFU**), dando como resultado **β-ddFU** [461] que ejerce su toxicidad como ya se ha visto.

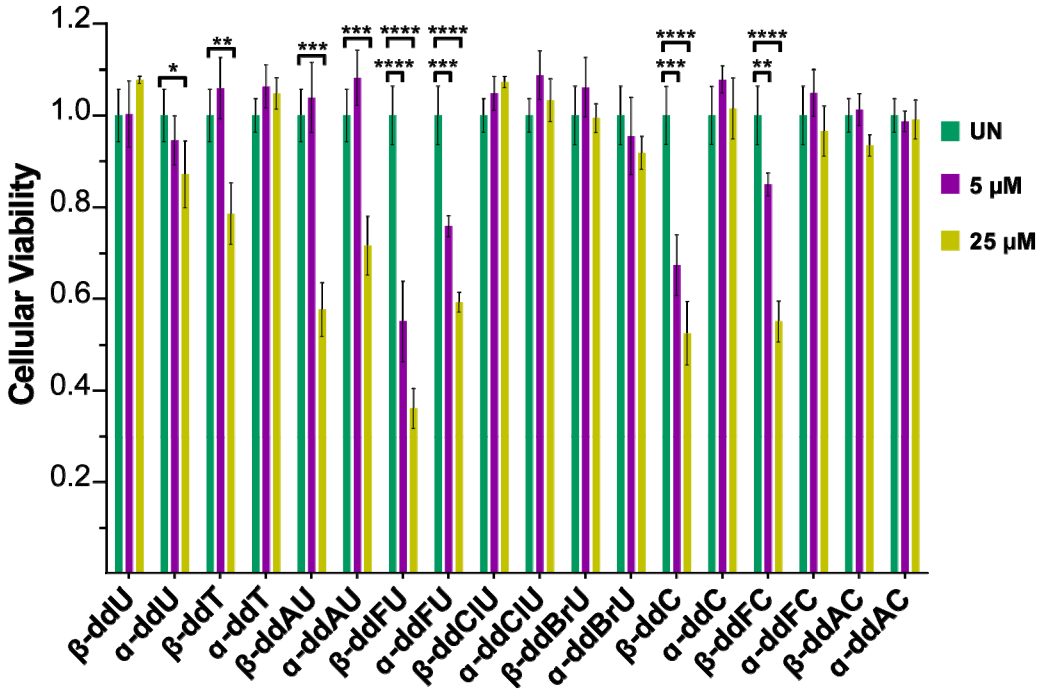


Figura 49. *Viabilidad celular de HeLa para nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados mediante el ensayo MTT.*

Mediante estos datos se corrigió la evaluación de la inhibición de la movilidad de Human L1.3 (**Figura 51**). Como resultado se puedo determinar que varios nucleósidos sintetizados inhibían la retrotransposición de dicho elemento. **β-ddT**, **α-ddFU** y **β-ddFC** tuvieron un efecto medio sobre la retrotransposición, siendo el último el que mayor inhibición produjo. El compuesto **β-ddC** parece inhibir la retrotransposición de Human L1.3 de forma moderada y con valores

similares a **β-ddFC,** si bien la disminución no es estadísticamente significativa. Sin embargo, el compuesto **β-ddAC** mostró una potente inhibición de la retrotransposición de Human L1.3 (60% a 5 μM y 95% a 25 μM) de una forma dosis dependiente. Cuando se ensayó esta familia de nucleósidos para evaluar su capacidad de inhibición de retroelementos LTR de ratón, ninguno de ellos tuvo efecto sobre la movilidad del retrotransposón LTR Mouse IAP (**Figura 53**). Curiosamente utilizando el retroelemento Mouse MusD **β-ddAU** y **β-ddFU** mostraron una leve inhibición de su retrotransposición mientras que el compuesto **β-ddAC** no tuvo efecto alguno sobre la movilidad de los retrotransposones LTR de ratón (**Figura 52**). Esta selectividad y potencia frente a Human L1.3, junto con su nula toxicidad (a las concentraciones ensayadas), hicieron que el compuesto **β-ddAC** fuera seleccionado para llevar a cabo ensayos posteriores.

Con los resultados obtenidos hasta ahora, parece que la RT de LINE-1 humano muestra cierta selectividad por unirse a nucleósidos derivados de citosina. Todos los nucleósidos activos y selectivos sobre LINE-1 humano frente a retrotransposones LTR de ratón son derivados de esta base nitrogenada: **β-ddAC, emtricitabina** y **lamivudina**. Además, **β-ddFC** y **β-ddC** también parecen ser inhibidores selectivos, si bien de forma mucho menos potente y sobre todo con una importante toxicidad que limita su uso en estudios posteriores. Por otra parte, y como cabía esperar, solo los nucleósidos con la configuración β poseen actividad inhibitoria sobre los retrotransposones en general de forma significativa, al menos con los datos obtenidos en este estudio. La falta de una estructura cristalina del dominio RT de LINE-1 por las dificultades de su que parece implicar [60], [132], ya sea humana o de otra especie, impide realizar estudios computacionales a fin de encontrar una estructura química cuyas características estructurales permitan obtener análogos nucleosídicos más potentes y selectivos. A pesar de que enzimas relativamente cercanas desde el punto de vista filogenético, como las del grupo II de intrones, han sido cristalizadas [135], [462], así como de retrotransposones LTR de levaduras Ty3 [463], estas enzimas no son adecuadas para los estudios anteriormente mencionados. Por ello, a día de hoy la única manera consistiría en desarrollar dichos análogos químicamente y evaluarlos en cultivo celular.

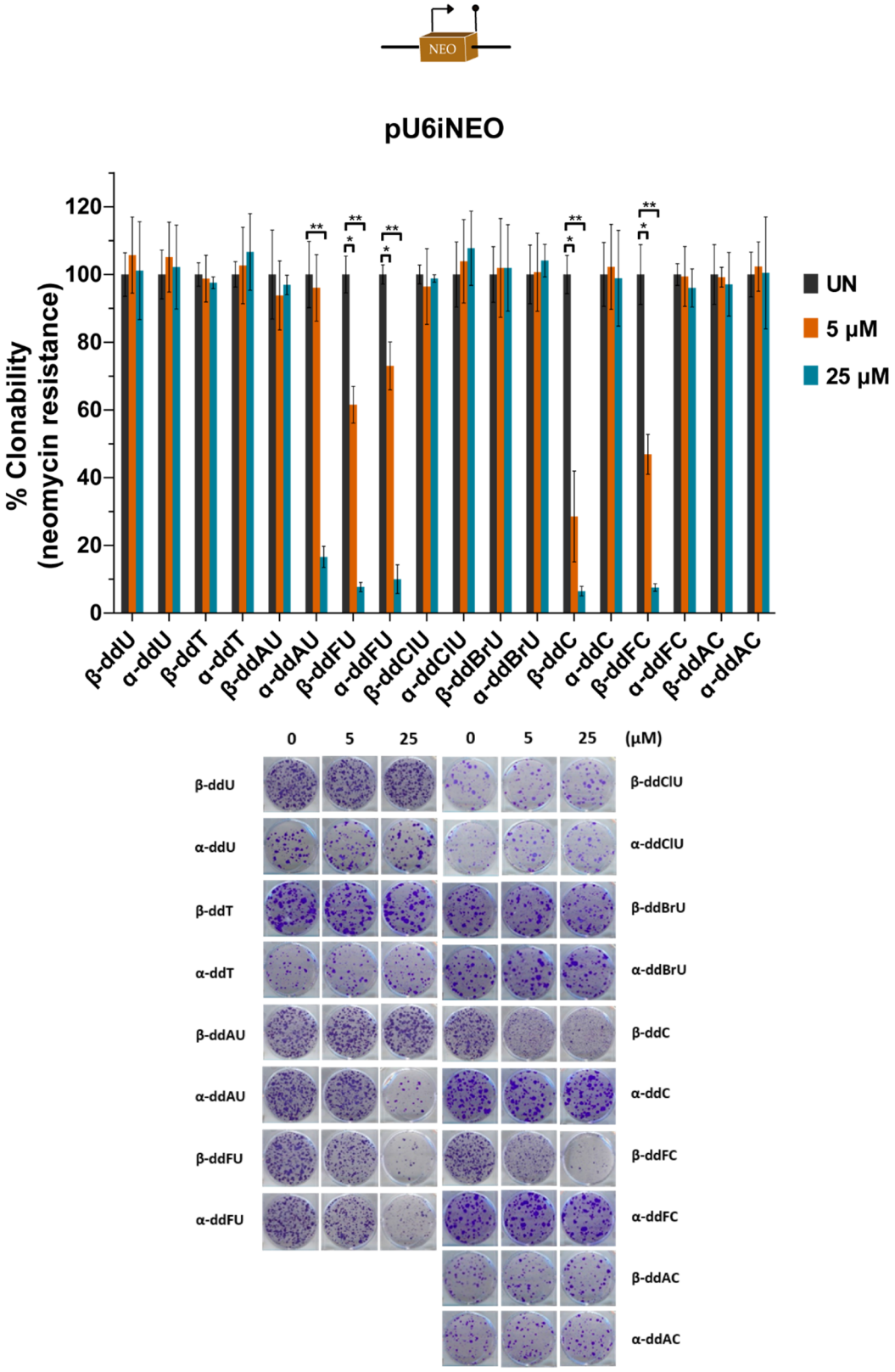


Figura 50. *Resultado y cuantificación del ensayo de clonabilidad en células HeLa de nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados.*

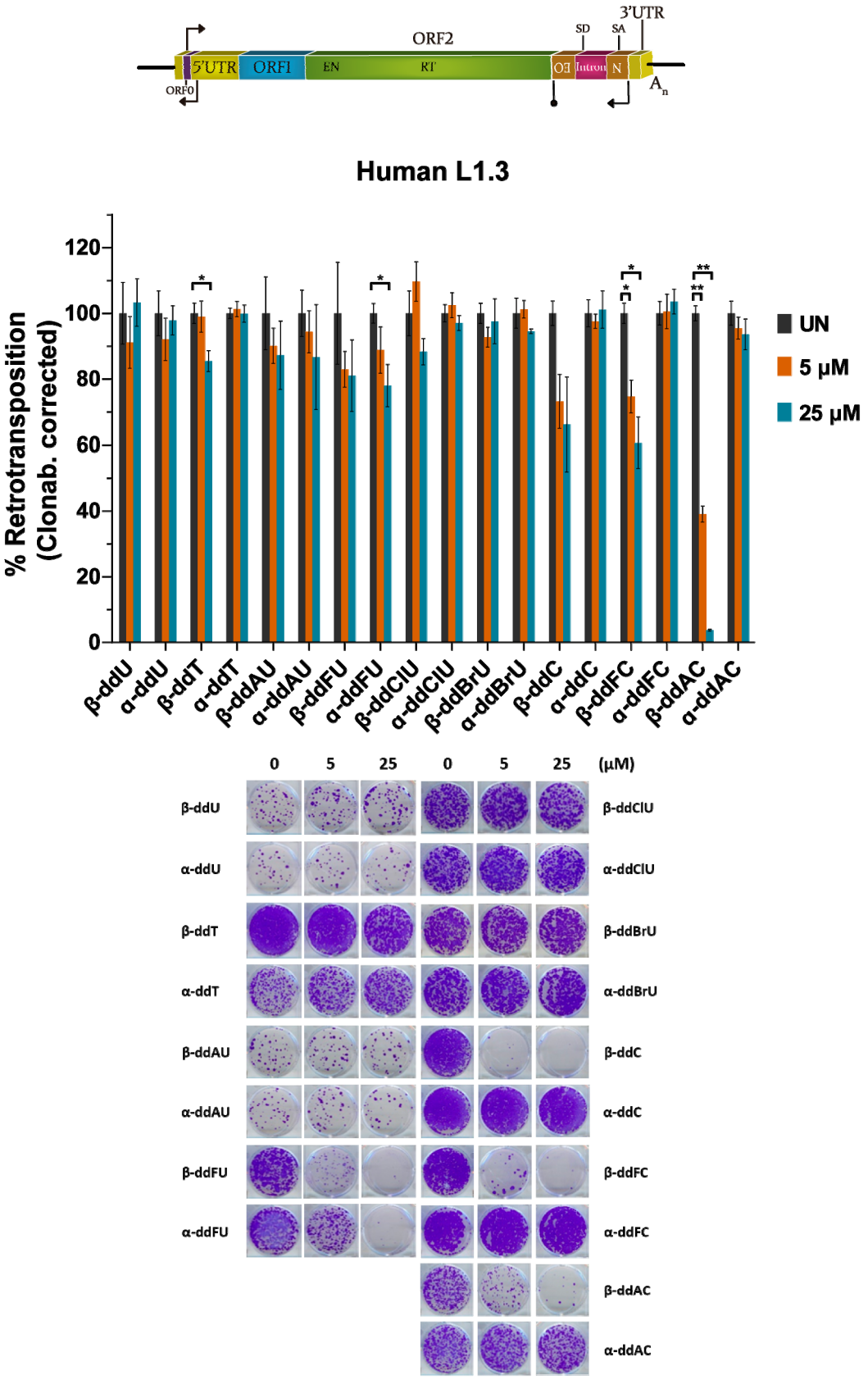


Figura 51. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados con Human L1.3.*

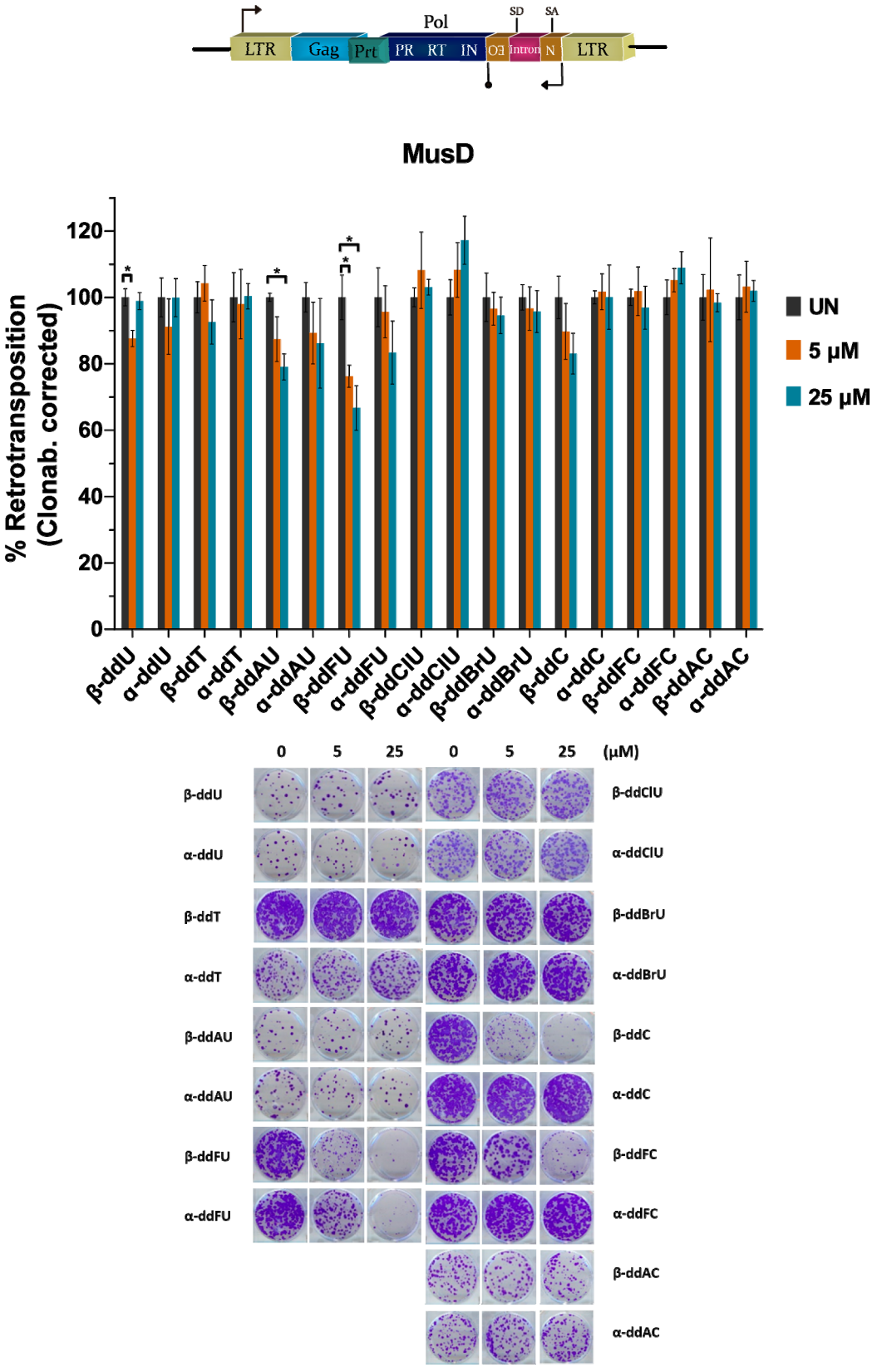


Figura 52. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados con Mouse MusD.*

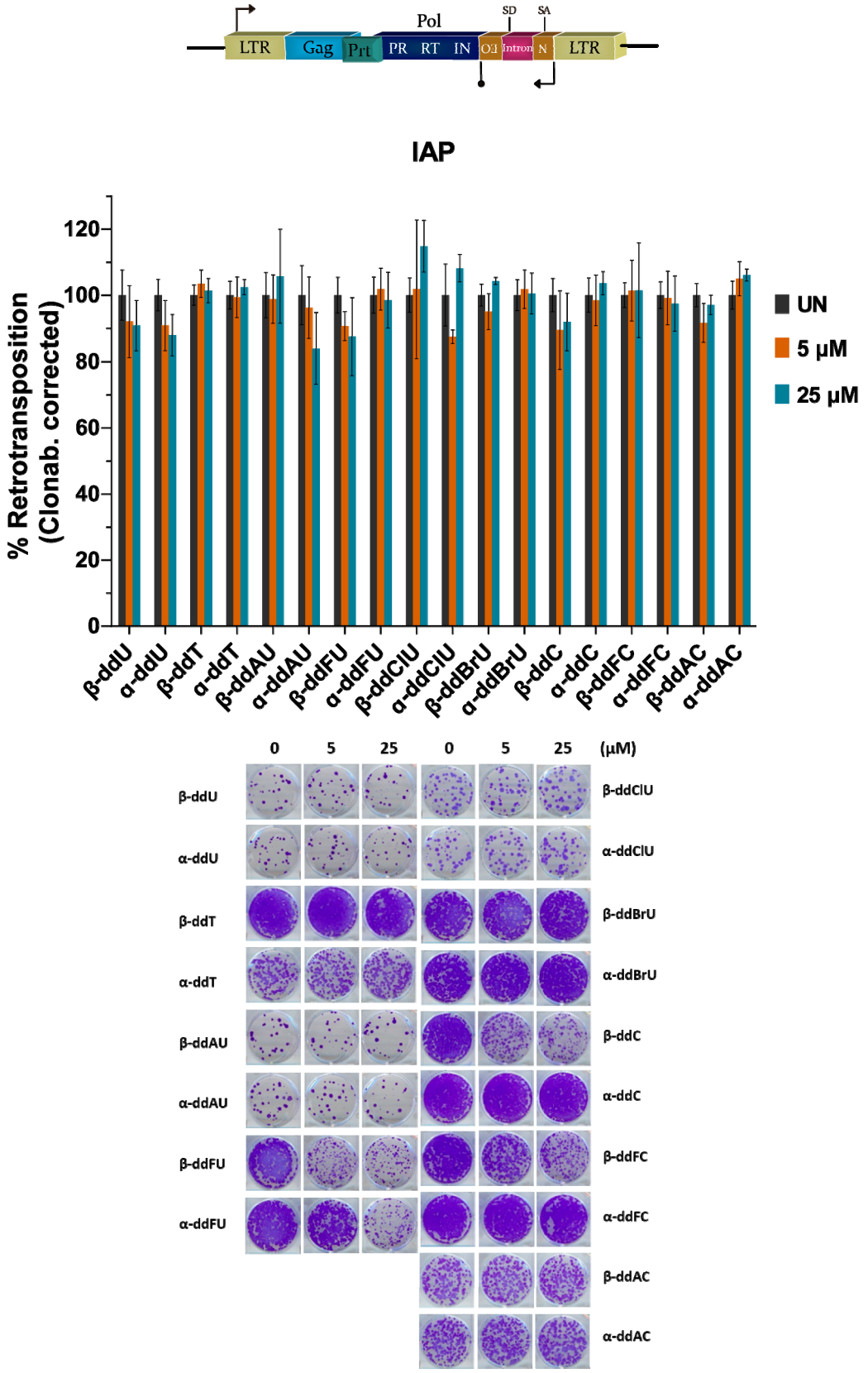


Figura 53. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados con Mouse IAP.*

En resumen, 3 familias de nucleósidos, un total de 33 compuestos (además de los 3 NNRTIs), fueron sometidos a una evaluación para determinar su capacidad de inhibir selectivamente el retroelemento transfectado Human L1.3 en células HeLa, además de evaluar la toxicidad de los mismos mediante el ensayo MTT (toxicidad aguda) y el plásmido pU6ineo (toxicidad a largo plazo/clonabilidad).

3.2.4 Determinación del alcance de la selectividad y potencia de los nucleósidos selectivos y no tóxicos.

Los compuestos **β-ddAC, emtricitabina y lamivudina** demostraron ser inhibidores potentes, selectivos y no tóxicos de la movilidad del retroelemento Human L1.3. Ninguno de ellos mostró actividad inhibitoria sobre retrotransposones LTR de ratón. Con el objetivo de completar la caracterización de **β-ddAC, emtricitabina y lamivudina** como inhibidores selectivos, estos fueron sometidos a una serie de ensayos en los que se incluyeron distintas concentraciones de RTi, células y/o retrotransposones utilizados hasta ahora.

3.2.5 Ensayos de dosis-respuesta de los nucleósidos selectivos y no tóxicos

Una vez caracterizados los 3 RTis como inhibidores potentes, selectivos y no tóxicos de la movilidad del retroelemento LINE-1 humano, nos propusimos evaluar un rango de concentraciones más amplio que el anterior a la vez que evaluamos la capacidad de **β-ddAC, emtricitabina y lamivudina** para inhibir la retrotransposición de LINE de ratón. Para ello utilizamos un ensayo de retrotransposición previamente establecido [151], [287] que emplea el gen de resistencia a blasticidina (blasticidina-S desaminasa) como gen reportero [464] de la retrotransposición (**Figura 54**). En este caso, se empleó un LINE humano L1.3 [453] contenido en el plásmido JJ101/L1.3 [465] y un LINE de ratón LIMd-LI_{Orl} [466] contenido en el plásmido JJLISM [467]. El principio de este ensayo es idéntico al de neomicina ya que el gen de resistencia a blasticidina está interrumpido por un intrón. Resumidamente, más retrotransposición se traduce en mayor número de colonias resistentes a blasticidina. La duración de este ensayo es menor a la del

ensayo con el gen de resistencia a neomicina por la rápida muerte celular que provoca la blasticidina. Brevemente, al día siguiente de la siembra se transfectaron células HeLa con los plásmidos indicados y tras 24 horas se cambió el medio de cultivo añadiendo el RTi correspondiente a la concentración indicada (0, 1, 2.5, 5, 10 y 20 µM), la cual fue mantenida hasta el final del ensayo. A los 5 días post-transfección se comenzó la selección con blasticidina, la cual duró 9 días adicionales. Transcurrido este tiempo las colonias se fijaron y tiñeron para ser contabilizadas.

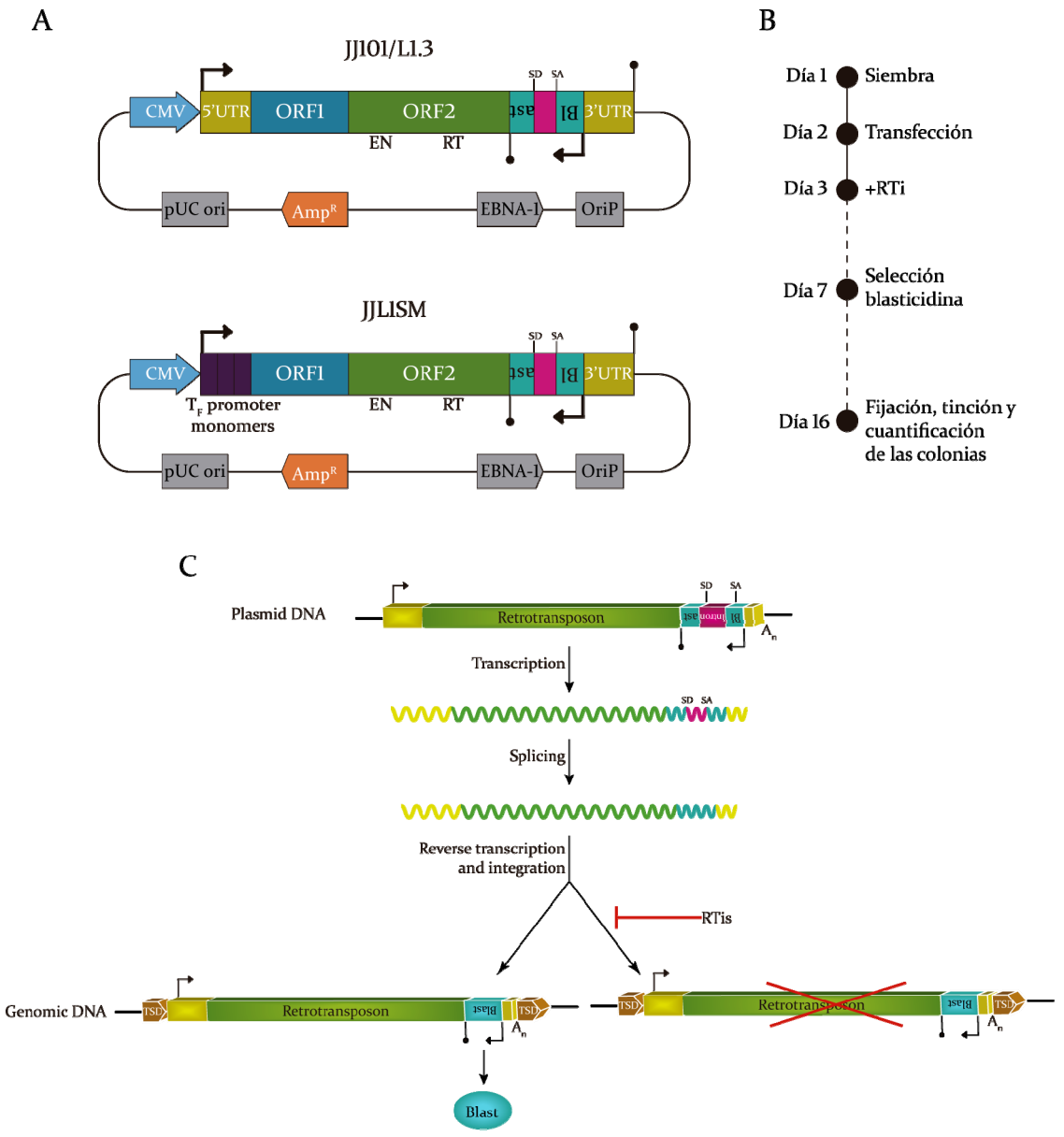


Figura 54. Ensayo de retrotransposición con el gen de resistencia a blasticidina como gen reportero de la movilidad. A. *Representación de los plásmidos empleados en los ensayos de retrotransposición con el gen de resistencia a blasticidina como gen reportero. Las principales*

características de cada elemento están representadas, pero no a escala. Ver ***Figura 2*** *para más información sobre la estructura de los retrotransposones. Blast, gen de resistencia a blasticidina; pUC ori, origen de replicación bacteriano; Amp^R, gen de resistencia a ampicilina; EBNA-1 y OriP, antígeno nuclear y origen de replicación eucariota.* ***B.*** *Esquema del desarrollo temporal de los ensayos con el gen de resistencia a blasticidina como gen reportero.* ***C.*** *Esquema del proceso de retrotransposición desde el retroelemento marcado con el gen reportero presente en el plásmido. El gen de resistencia a blasticidina se puede expresar tras un evento de retrotransposición. En presencia de RTis, la inhibición de la movilidad impide su expresión. Ver* ***Figura 9*** *para más información.*

Este ensayo evidenció la potente capacidad inhibitoria de **β-ddAC**, **emtricitabina** y **lamivudina** frente a LINE humano (**Figura 55**) y también de ratón (**Figura 56**) incluso a concentraciones muy bajas en ambos casos, aunque la inhibición de LINE humano fue más significativa. A diferencia del resto de los ensayos, en este no se observaron diferencias importantes entre los 3 RTis en cuanto a su comportamiento para inhibir la retrotransposición, ofreciendo valores similares.

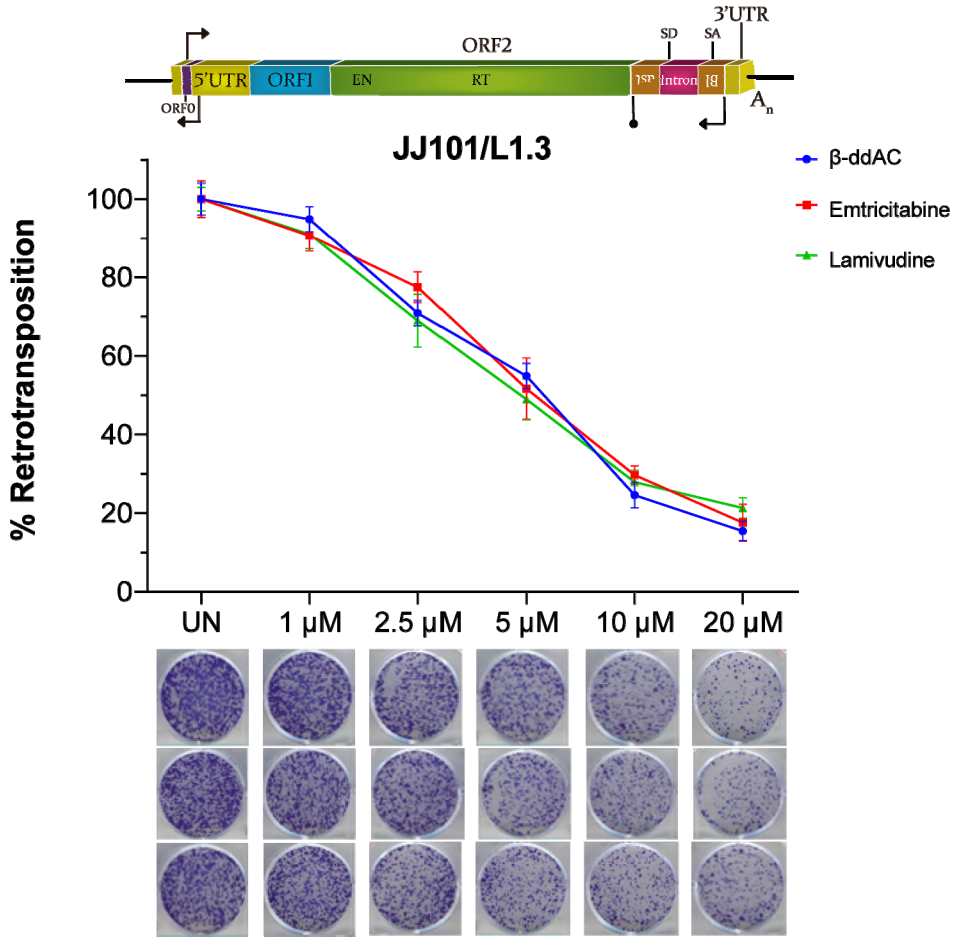


Figura 55. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados con JJ101/L1.3.*

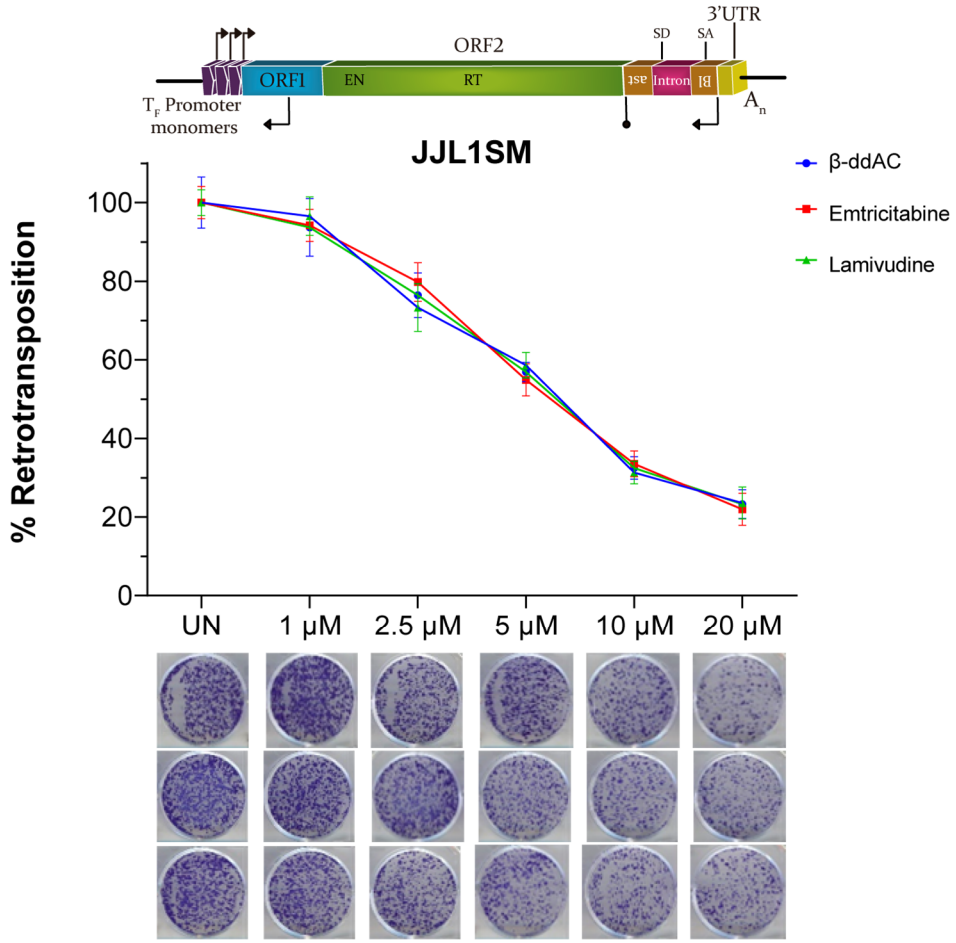


Figura 56. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los nucleósidos derivados de 2',3'-didesoxi-D-ribosa sintetizados con JJLiSM.*

A continuación, medimos la actividad de **β-ddAC**, **emtricitabina** y **lamivudina** para inhibir la retrotransposición de LINE humano en células HEK293T, pero utilizando sistema ligeramente distinto a los anteriores. En este caso, además de utilizar otro tipo celular para comprobar que la efectividad de los RTis se mantenía, se empleó el plásmido pYX014 [286] en el que se ha incluido el LINE humano $L1_{RP}$ [468] (semejante al contenido en el plásmido JM101/L1.3). En este retroelemento se ha clonado el gen de la luciferasa de luciérnaga en antisentido en su región 3'UTR y, de la misma forma que anteriormente, interrumpido por un intrón que impide la expresión del gen de la luciferasa salvo que tenga lugar un evento de retrotransposición (**Figura 57**). La luz generada por esta enzima en presencia de su sustrato, la luciferina, es cuantificable y fue usada para medir la inhibición de la retrotransposición por parte de **β-ddAC**,

emtricitabina y **lamivudina** sobre un LINE humano. Para normalizar los resultados con respecto a la transfección se empleó la luminiscencia de la luciferasa de renilla, contenida en el mismo plásmido pYX014, y como control negativo del ensayo el plásmido pYX015, el cual es idéntico al anterior salvo por dos mutaciones con cambio de sentido en ORF1 que impiden la movilidad del retrotransposón. Estos ensayos se llevaron a cabo como se ha descrito previamente [150]. Brevemente, las células HEK293T fueron sembradas y transfectadas al día siguiente con los plásmidos correspondientes. Transcurridas 6 horas desde la transfección se cambió el medio añadiendo los RTis a distintas concentraciones (0, 2.5, 5, 10, 15 y 20 μM), las cuales fueron mantenidas hasta el final del ensayo. A las 24 horas post-transfección se comenzó la selección con puromicina, que se aplicó durante 3 días adicionales tras los que se midió la actividad de la luciferasa de luciérnaga y de renilla.

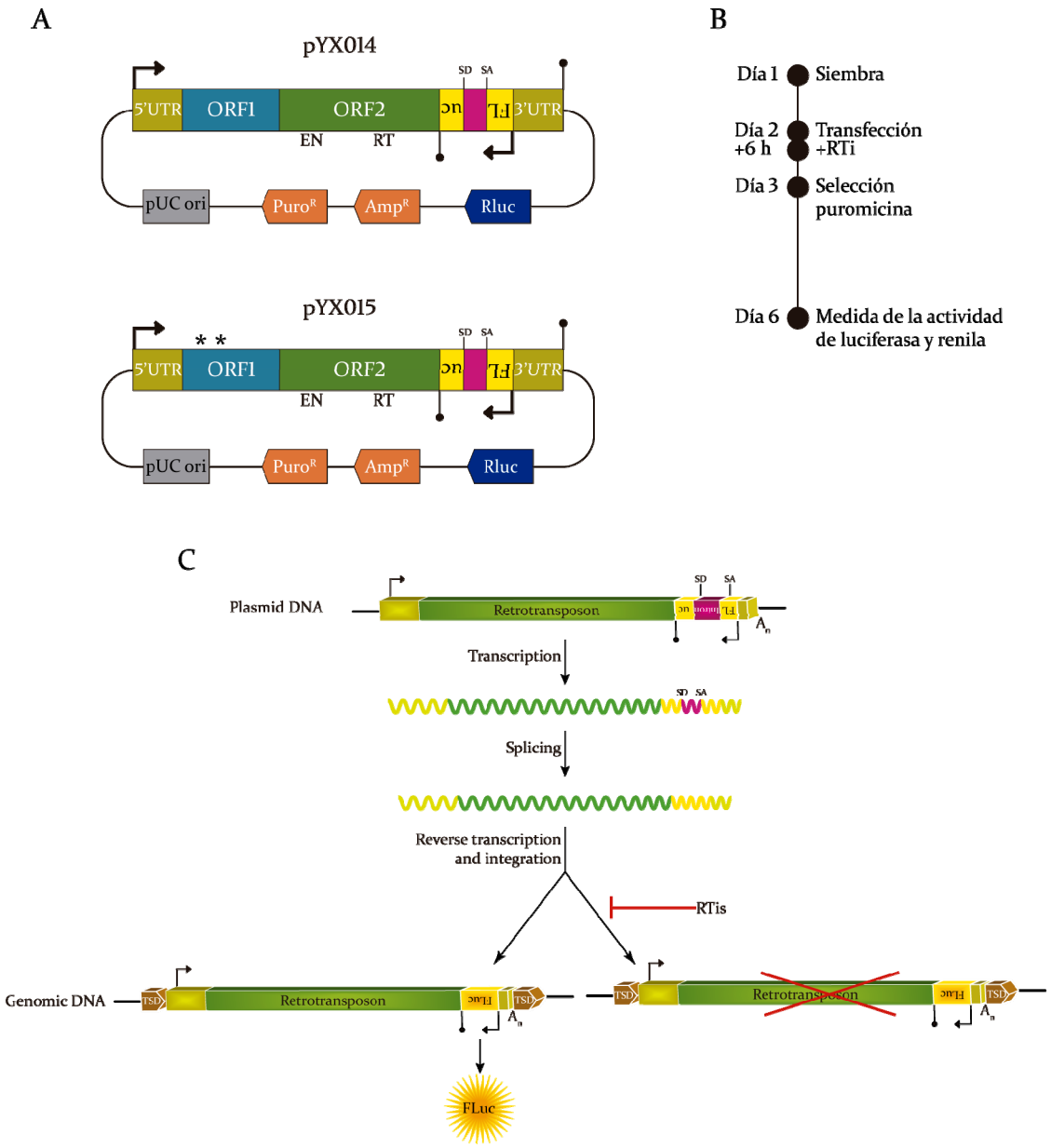


Figura 57. Ensayo de retrotransposición con el gen de la luciferasa como gen reportero de la movilidad. A. *Representación de los plásmidos empleados en los ensayos de retrotransposición con el gen de la luciferasa como gen reportero. Las principales características de cada elemento están representadas, pero no a escala. Ver* ***Figura 2*** *para más información sobre la estructura de los retrotransposones. Fluc, luciferasa de luciérnaga; pUC ori, origen de replicación bacteriano; Amp^R, gen de resistencia a ampicilina; Puro^R, gen de resistencia a puromicina; RLuc, luciferasa de renilla; EBNA-1 y OriP, antígeno nuclear y origen de replicación eucariota.* ***B.*** *Esquema del desarrollo temporal de los ensayos con el gen la luciferasa como gen reportero.* ***C.*** *Esquema del proceso de retrotransposición desde el retroelemento marcado con el gen reportero presente en el plásmido. El gen de la luciferasa se puede expresar tras un evento de retrotransposición. En presencia de RTis, la inhibición de la movilidad impide su expresión. Ver* ***Figura 9*** *para más información.*

Los resultados confirmaron de nuevo que **β-ddAC, emtricitabina** y **lamivudina** se comportaban como inhibidores de un LINE humano en otro tipo celular, empleando un sistema reportero distinto (**Figura 58**). El compuesto **β-ddAC** mostró significativamente ser el más potente de los tres, seguido de una potente inhibición por la **lamivudina** y finalmente por **emtricitabina** que en este caso inhibió la retrotransposición de una forma más moderada.

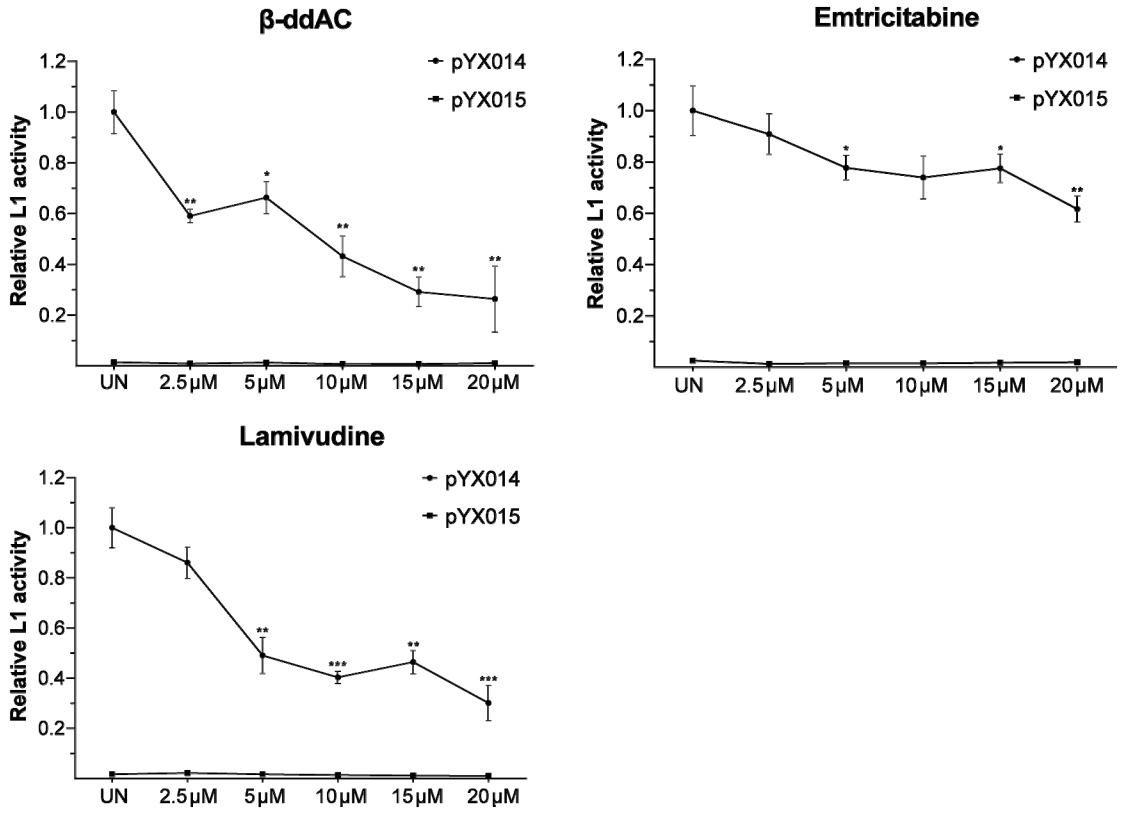


Figura 58. *Cuantificación del ensayo de retrotransposición en células HEK293T de los nucleósidos selectivos con pYX014.*

La menor actividad de los compuestos ensayados en células HEK293T puede estar motivado por distintos factores como a una menor cantidad o actividad de transportadores de membrana, una menor actividad metabólica (necesaria para transformar los nucleósidos en los nucleótidos activos) o una capacidad disminuida de estas células para expresar el retrotransposón contenido en el plásmido. Además, a diferencia de las células HeLa, las HEK293T, aunque bajos, presentan ciertos niveles de expresión endógena de LINE-1 [186], por lo que podría darse una competencia entre la reversotranscriptasa endógena y la generada por el plásmido para unirse a los nucleósidos.

3.2.6 Controles adicionales sobre la validez de los resultados

Para descartar que los resultados observados anteriormente no se debían a efectos indirectos no relacionados con la inhibición de la retrotransposición, se realizaron ensayos para determinar la estabilidad del plásmido en presencia de los RTis así como para detectar posibles cambios en los niveles de RNA de LINE-1. Además de **β-ddAC**, **emtricitabina** y **lamivudina**, se utilizaron como controles tenofovir (inhibición no selectiva, no tóxico), AZT (inhibición no selectiva, tóxico), C5 (no inhibición, no tóxico). Para realizar estos controles se transfectaron células HeLa con el plásmido JM101/L1.3 (**Figura 59**) y se pusieron en contacto con los distintos nucleósidos durante 48 horas a 25 μM, tras lo cual se extrajo DNA genómico y RNA. Para comprobar la integridad del plásmido se cuantificaron sus niveles mediante qPCR (quantitative Polymerase Chain Reaction) empleando para tal fin unos primers que amplifican una región del gen EBNA-1 (contenido en el plásmido) y otros que lo hacen de una región del gen GAPDH para normalizar los datos. Para cuantificar los niveles de RNA de LINE y descartar así que los RTis pudieran afectar a la expresión del retroelemento contenido en el plásmido se llevó a cabo una RT-qPCR (Reverse Transcription-qPCR) empleando unos primers que amplifican una región transcrita por el gen de resistencia a neomicina (una vez ha sufrido el splicing) y otros que lo hacen del tránscrito del EBNA-1 para normalizar los datos. Ni los niveles de plásmido ni los de RNA del LINE humano contenido en él sufrieron variaciones significativas en presencia de los RTis, confirmando que los resultados obtenidos anteriormente no se habían debido a una disminución de los valores mencionados.

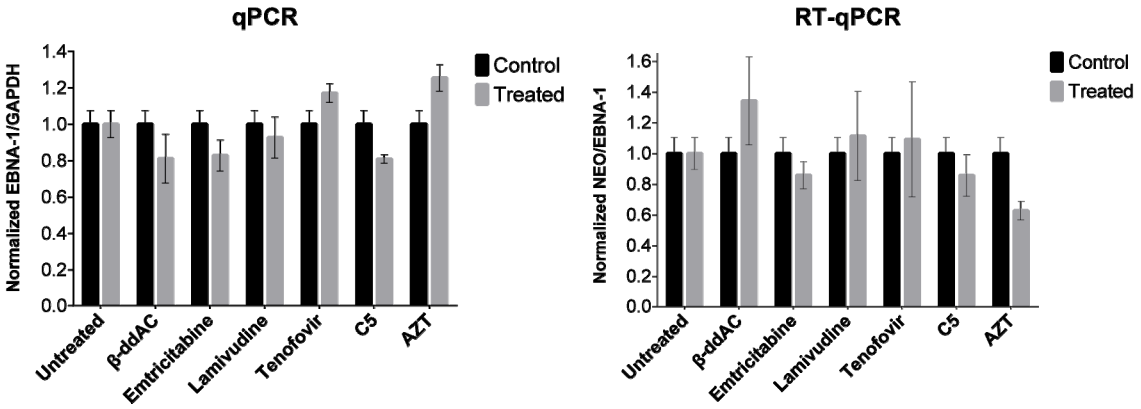


Figura 59. *Control de la estabilidad del plásmido (qPCR) y de la expresión de RNA de LINE-1 (RT-qPCR).*

A continuación, se llevaron a cabo ensayos de toxicidad con **β-ddAC**, **emtricitabina** y **lamivudina** en células PA-1, las cuales sobreexpresan de forma endógena LINE-1 [469], mediante el ensayo de MTT. Esta evaluación mostró que los 3 RTis no mostraban toxicidad en PA-1 a bajas concentraciones (5 y 25 µM) como tampoco lo hicieron a altas concentraciones (50 y 150 µM) (**Figura 60**). Finalmente, se analizó la posibilidad de que estos compuestos (junto con tenofovir, C5 y C1) provocaran cambios en los niveles de ORF1p de LINE-1 en células PA-1 (empleando p53 como normalizador). Para ello, se sembraron células PA-1 y se pusieron en contacto con los distintos RTis a 25 µM durante 96 horas, tras las cuales se midió la cantidad de proteína. El Western Blot mostró que ninguno de los compuestos provocó cambios significativos en los niveles de expresión de ORF1p. Resultados similares ofreció la misma prueba cuando fue realizada con células HeLa (**Figura 61**). De esta manera quedó demostrado que los RTis carecen de toxicidad tanto en células con sobreexpresión endógena de LINE-1 (PA-1) como en células con bajos o nulos niveles de expresión (HeLa) de LINE-1.

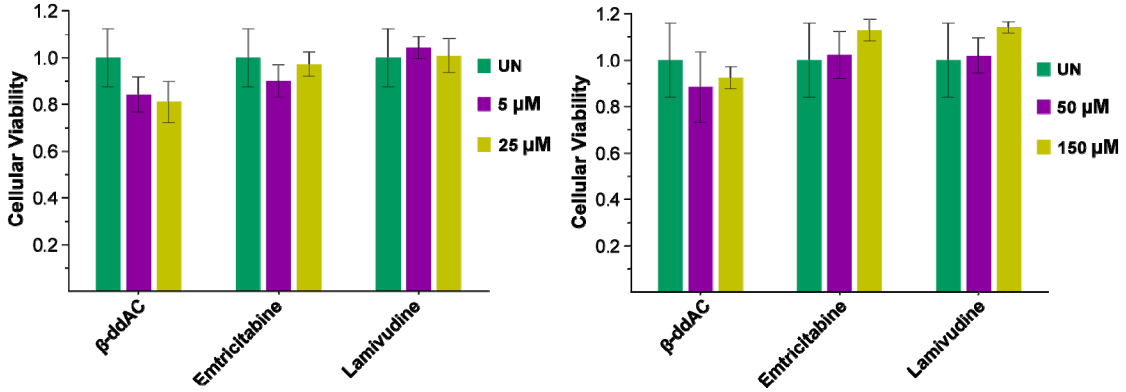


Figura 60. *Viabilidad celular de PA-1 para los nucleósidos selectivos mediante el ensayo MTT.*

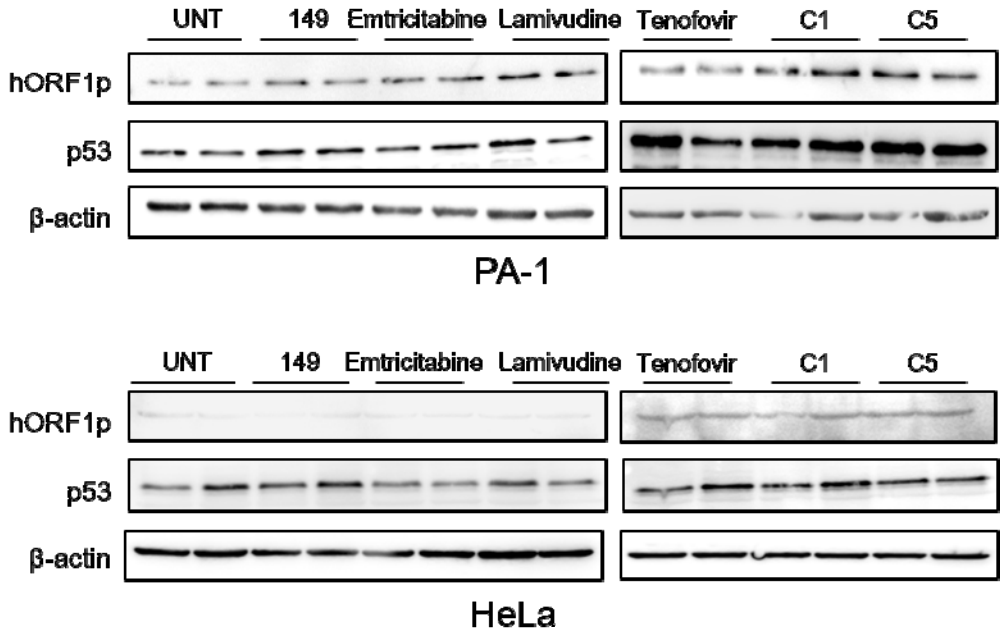


Figura 61. *Análisis de la expresión de ORF1p de LINE-1 mediante Western blot en células PA-1 y HeLa en presencia de compuestos nucleosídicos (25 μM).*

3.2.7 Evaluación de la selectividad de los nucleósidos activos y no tóxicos

Para confirmar definitivamente la selectividad de **β-ddAC**, **emtricitabina** y **lamivudina**, se repitieron los ensayos de retrotransposición con el gen de resistencia a neomicina como gen reportero, empleando Mouse MusD y Mouse IAP, pero esta vez las concentraciones a las que se llevaron a cabo los experimentos fueron 50 y 150 μM a fin de comprobar si se comportaban como inhibidores de retrotransposones LTR a concentraciones superiores. Para corregir los datos se determinó la toxicidad utilizando esas mismas concentraciones en el ensayo MTT y con pU6ineo. Bajo estas condiciones, ninguno de los compuestos produjo una disminución de la retrotransposición de Mouse MusD o Mouse IAP (**Figura 62**) ni tampoco mostraron toxicidad (**Figura 63**) en los ensayos destinados a su determinación. La ausencia de actividad inhibitoria sobre la movilidad de los retrotransposones LTR de ratón a 50 y 150 μM (muy superiores a las anteriores)

demostró definitivamente que **β-ddAC, emtricitabina** y **lamivudina** eran nucleósidos que inhibían selectivamente la movilidad de LINE-1 humano de manera potente y sin exhibir toxicidad.

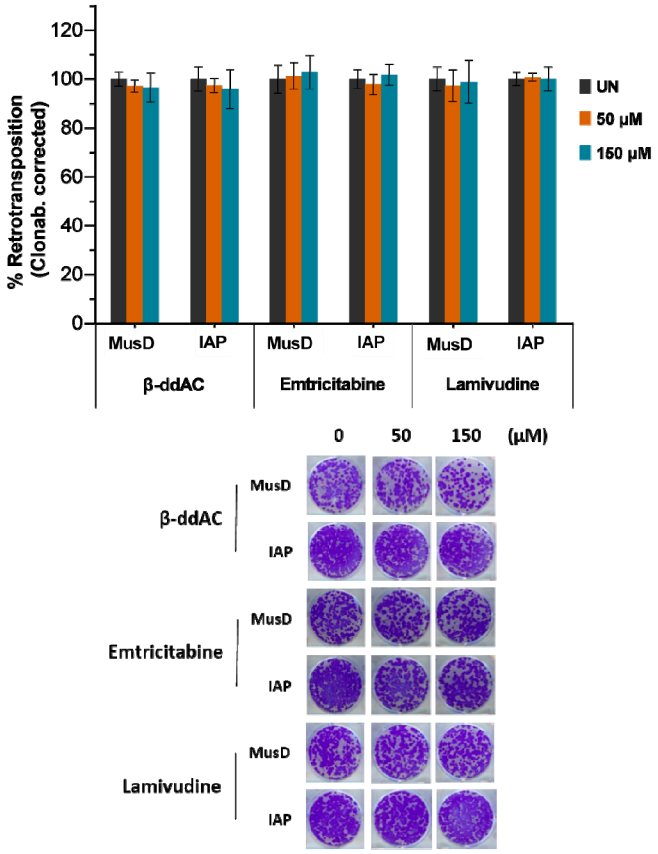


Figura 62. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los nucleósidos selectivos con Mouse MusD y Mouse IAP.*

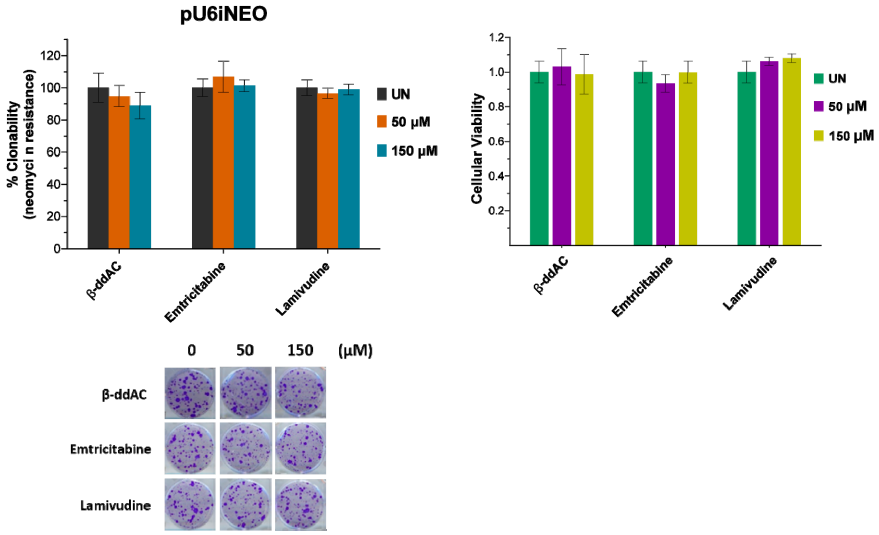


Figura 63. *Resultado y cuantificación del ensayo de clonabilidad (izquierda) y viabilidad celular (derecha) en células HeLa de los nucleósidos selectivos.*

Probada la selectividad de **β-ddAC, emtricitabina** y **lamivudina** frente a LINE-1 humano, nos propusimos confirmar el alcance de dicha selectividad midiendo su capacidad para inhibir la movilidad de LINEs de ratón, similares a los LINEs humanos. Para ello, llevamos a cabo un ensayo de retrotransposición con el gen de resistencia a neomicina en células HeLa como se ha relatado anteriormente. En este caso se emplearon 3 LINEs de ratón (igual o semejantes al contenido en el plásmido JJL1SM, **Figura 56**) de distintas familias, L1Md-A , L1Md-T_F y L1Md-G_F [470]–[472], las cuales se diferencian básicamente en su región promotora: L1Md-$L1_{Orl}$ [466], contenido en el plásmido pCEPL1SM (**Mouse L1-T_F**) [473]; L1Md-G_F21, contenido en el plásmido pCEP-TG_F21 (**Mouse L1-G_F**) [471]; L1Md-A101, contenido en el plásmido pCEP-A101 (**Mouse L1-A**) [471]. Mediante este ensayo se pudo confirmar que **β-ddAC, emtricitabina** y **lamivudina** eran capaces de inhibir también los LINEs de ratón, si bien esta inhibición fue bastante menos acusada que en el caso de Human L1.3 (**Figura 64**). En base a los datos obtenidos, **β-ddAC** resultó ser el más potente para inhibir Human L1.3 a la vez que en general fue el menos potente para los LINEs de ratón, la **emtricitabina** fue el segundo en potencia frente a Human L1.3 al igual que al inhibir LINEs de ratón y la **lamivudina**, aunque menos potente frente a Human L1.3 que los anteriores, resultó ser la más potente para inhibir LINEs de ratón, aunque esta inhibición no fue excesivamente grande.

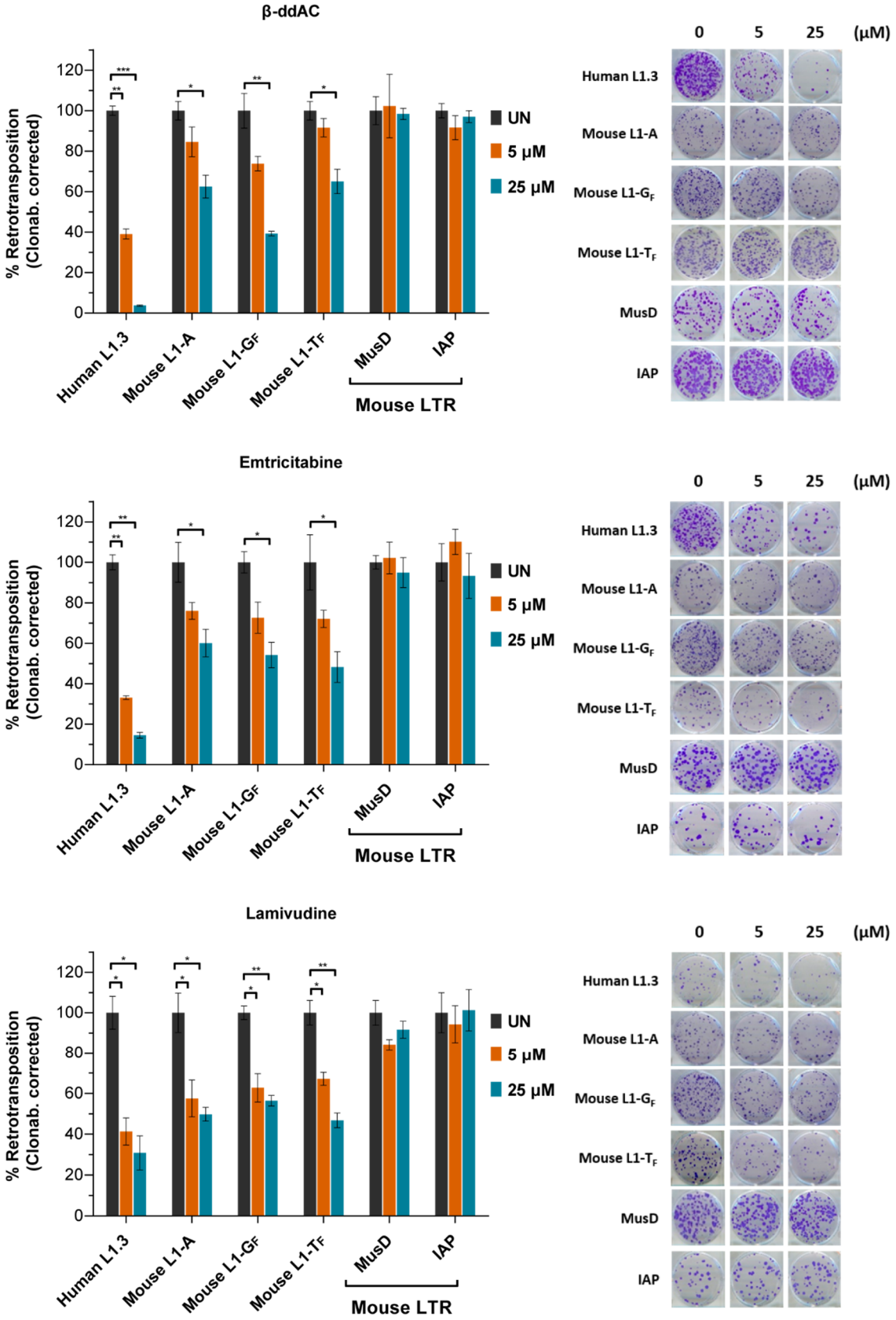


Figura 64. *Resultado y cuantificación del ensayo de retrotransposición en células HeLa de los nucleósidos selectivos con LINE-1 humano (Human L1.3), LINE-1 de ratón (Mouse L1-A, Mouse L1-G_F y Mouse L1-T_F) y retrotransposones LTR de ratón (MusD e IAP).*

Algunos estudios previos habían mostrado fehacientemente el potencial de ciertos fármacos empleados en el tratamiento del SIDA para inhibir la retrotransposición de LINE-1 humano [291], [293]. *Jones et al.*[291] llevaron a cabo el primer estudio al respecto en el que se empleó un plásmido similar a los utilizados en este estudio (contenía un LINE-1 humano con un gen reportero interrumpido por un intrón que impedía la expresión del gen reportero salvo que tuviera lugar un evento de retrotransposiciónx) pero en este caso el gen reportero de elección fue el EGFP (Enhanced Green Fluorescent Protein) [288]. En estos ensayos, realizados en células HeLa, se determinó que AZT, lamivudine, stavudine y tenofovir producían una inhibición potente sobre el mencionado retroelemento, mientras que la nevirapina no produjo inhibición. En nuestro caso, los resultados concuerdan de forma cualitativa con los del estudio de *Jones et al.* a excepción de la nevirapina que mostró cierta actividad, aunque de manera dosis no dependiente. La potencia mostrada en uno y otro estudio no son del todo comparables, a pesar de usar la misma línea celular, ya que emplean un LINE-1 distinto (LRE3), sistemas reporteros distintos y un diseño del experimento diferente. A pesar de la calidad del estudio de *Jones et al.*, este no tiene en cuenta la toxicidad producida por la presencia de los RTis en el medio de cultivo de modo que los resultados de inhibición mostrados podrían presentar dicho sesgo. De hecho, con nuestro estudio hemos demostrado que el AZT es un inhibidor (no selectivo) de la retrotransposición de LINE-1 humano y LTR de ratón, pero también es un compuesto con una alta toxicidad en las células HeLa.

Dai et al.[293] investigaron posteriormente el efecto de diversos NRTIs (AZT, lamivudina, stavudine y zalcitabina) y NNRTIs (delavirdina, efavirenz y nevirapina) sobre proteínas RT recombinantes purificadas de LINE-1 humano (Ll_{RP}), VIH-1 y LTR de levadura Ty1. En este estudio todos los NRTIs mostraron actividad inhibitoria sobre las 3 proteínas, con algo más de potencia frente a LINE-1 comparado con Ty1. Por otro lado, los NNRTIs solo fueron efectivos en la inhibición de HIV-1, como por otra parte cabía esperar, al igual que en nuestro estudio. Cabe mencionar que *Dai et al.* también observaron un leve grado de inhibición de la retrotransposición de LINE-1 por parte de la nevirapina de una

forma dosis no dependiente. Además, realizaron un ensayo celular transfectando un LINE-1 humano (Ll_{RP}) y de ratón ($LlMd\text{-}T_F$), en el que obtuvieron resultados similares a los anteriores, con los NRTIs produciendo una inhibición significativa y con poco o nulo efecto de los NNRTIs. Por otro lado, aunque corrigen los resultados obtenidos mediante un ensayo de toxicidad diferente al nuestro, estos datos no están disponibles para ser tenidos en cuenta ni se comentan en el estudio.

Nuestro trabajo confirma, refuerza y amplía considerablemente los resultados de estos dos estudios mencionados. No solo hemos demostrado de una forma inequívoca la inhibición que distintos nucleósidos producen sobre la retrotransposición de LINE-1 humano en células HeLa, entre ellos algunos ya ensayados en estos estudios, sino que hemos extendido esos ensayos a retrotransposones LTR de ratón, mostrando el efecto de hasta 33 nucleósidos en retrotransposición de esos dos tipos de retroelementos. De este modo se han podido caracterizar tres de ellos como inhibidores potentes y selectivos de la movilidad de un LINE-1 humano y tres de ratón frente a dos retrotransposones LTR de ratón. Los ensayos llevados a cabo con estos compuestos han demostrado que es posible inhibir la movilidad de LINE-1 humano y de ratón transfectados. En el caso de LINE-1 humano se ha podido mostrado dicha inhibición en al menos dos líneas celulares, HeLa y HEK293T, y con distintos sistemas reporteros de la retrotransposición (neomicina, blasticidina y luciferasa). En el caso de los LINE-1 de ratón, se produjo inhibición, aunque con menor potencia que para LINE-1 humano, en células HeLa con distintos sistemas reporteros (neomicina y blasticidina). Además, para corregir los resultados de estos ensayos y evitar sesgos, hemos determinado la toxicidad de todos los compuestos de este estudio en células HeLa a través de dos ensayos complementarios: toxicidad a largo plazo o clonabilidad mediante el plásmido pU6ineo y toxicidad a corto plazo o aguda mediante el ensayo de MTT, todos ellos a las mismas concentraciones a las que se realizaron los ensayos de retrotransposición (5 y 25 μM, y en algunos casos 50 y 150 μM). Con este último método también se ha evaluado la toxicidad de los compuestos más potentes y selectivos en células PA-1, las cuales presentan altos niveles de expresión de LINE-1 de forma endógena a diferencia de HeLa [452],

[469], a las mismas concentraciones que las anteriormente mencionadas. Ninguno de los tres RTis potentes y selectivos resultó tóxico a las concentraciones ensayadas en los ensayos de toxicidad (a corto y largo plazo), ni siquiera aquellas muy superiores a las necesarias para inhibir la retrotransposición, en células HeLa y PA-1. Para completar los controles se comprobó que estos compuestos no afectaban a la estabilidad de los plásmidos transfectados, como tampoco lo hacían de la expresión del retroelemento presente en ellos. Puesto que las copias de LINE-1 activas en el genoma humano son iguales en un 99.99% y que el centro activo de la RT está muy conservado [100], podemos especular que los tres RTis selectivos serán activos frente a cualquier LINE-1.

Desconocido es a día de hoy el papel de los retrotransposones en los organismos vivos y en concreto del único retrotransposón autónomo activo en humanos, el LINE-1. De hecho, resulta desconcertante que la expresión de LINE-1 en humanos esté inhibida en células somáticas (no tumorales) y, sin embargo, sea detectable la retrotransposición en células germinales, embrionarias o neuronales. Dada la importancia de estos tejidos por razones obvias, no parece lógico *a priori* que dichas células permitan la retrotransposición, al producir ésta inserciones que pueden resultar potencialmente letales para la célula. Esto puede ser motivo de especulación sobre si la retrotransposición simplemente está desregulada, por motivos que aún no alcanzamos a comprender, o si la retrotransposición es necesaria en estos tejidos para poder desarrollar su función biológica. La inhibición de la retrotransposición de LINE-1 mediante nucleósidos puede brindar una oportunidad para indagar en el papel de este retroelemento mediante una estrategia de pérdida de función. De hecho, haciendo uso de los datos de los estudios mencionados, han surgido otros en los que se han empleado dichos nucleósidos para estudiar cómo afecta la inhibición de la retrotransposición en los tejidos que lo expresan de forma endógena como en oocitos [474] y neuronas del hipocampo [475], pero también en procesos como el envejecimiento e inflamación [250], [251]. A menudo estos ensayos se realizan en ratones como modelos, los cuales contienen aún activos en su genoma tanto retrotransposones LTR como LINEs (no-LTR) [476], por lo que cuando se indaga en el papel de la

retrotransposición de LINE-1 mediante nucleósidos parece importante tener en cuenta lo anterior y emplear en la medida de posible **β-ddAC, emtricitabina** y **lamivudina** para inhibir selectivamente la retrotransposición de LINE-1 para poder diferenciar si un proceso biológico concreto se debe la inhibición de uno u otro tipo de retrotransposones. Además, estos compuestos carecen de toxicidad a concentraciones muy superiores a las inhibitorias. Sin embargo, cuando se quiera inhibir ambos procesos parece que el tenofovir resulta el más adecuado, al mostrar potente actividad frente a LINE-1 y LTR.

Por otro lado, numerosas son las patologías que cursan con una sobreexpresión de LINE-1, principalmente tumores de origen epitelial [232], [244] pero también en procesos autoinmunes, entre ellos el síndrome de Aicardi-Goutières [197], [477], [478]. En este último caso se está llevando a cabo un estudio clínico con nucleósidos inhibidores de la RT para determinar la utilidad terapéutica de estos compuestos en el tratamiento del síndrome de Aicardi-Goutières (NIH, ID: NCT02363452) [294]. En este estudio se está empleando una combinación de AZT, lamivudina y abacavir. Basándose en los datos de nuestro estudio, el uso de AZT podría evitarse por sus efectos tóxicos ampliamente reportados, así como el abacavir que, si bien no es tóxico, no está carente de efectos secundarios y su inhibición de la retrotransposición de LINE-1 es moderada a 25 μM (en cultivo celular). El uso de la lamivudina está justificado teniendo en cuenta nuestros ensayos, aunque si se quisiera combinar con otros RTis como en el estudio clínico de Aicardi-Goutières, los nucleósidos de elección serían los presentados en este estudio como inhibidores potentes y selectivos de la retrotransposición de LINE-1: **β-ddAC, emtricitabina** y **lamivudina**. La sobreexpresión de LINE-1 en tumores ha ido quedando patente en la última década, de la misma manera que el impacto que la retrotransposición tiene sobre el genoma del hospedador, sobre todo en un contexto de inestabilidad genómica como es el caso de células tumorales. La movilidad de retrotransposones durante el cáncer agrava la malignidad del tumor, así como su capacidad de metástasis. Parece haber además una relación entre el nivel de la hipometilación del promotor de LINE-1 (causa de la sobreexpresión) y una peor prognosis en los pacientes con cáncer. Con nuestro

trabajo esperamos abrir la puerta a estudios clínicos con pacientes con cáncer que presenten una alta expresión de LINE-1, en los que se usen RTis que inhiban la retrotransposición de este retroelemento como tratamiento coadyuvante. Por otra parte, algunos grupos de investigación han realizado ensayos con NRTIs y NNRTIs para tratar tumores en células en cultivo y en ratones y/o el efecto de estos compuestos sobre la diferenciación celular [289], [290], [479], [480], justificando la actividad antitumoral y los efectos sobre la diferenciación por la inhibición de la retrotransposición de LINE-1. Sin embargo, estos estudios carecen en general de controles adecuados apoyar dichas afirmaciones, así como de ensayos de toxicidad para corregir los resultados obtenidos. Aunque estos compuestos bien pueden haber ejercido su efecto mediante la inhibición de la movilidad de LINE-1, los efectos antitumorales también podrían deberse a la actividad intrínseca de algunos nucleósidos de comportarse como anticancerígenos (antimetabolitos). De forma similar, los efectos en la diferenciación pueden deberse a la interacción de estas moléculas con las enzimas celulares relacionadas con el metabolismo de ácidos nucleicos y no con la inhibición de la retrotransposición. Además, cuando se han usado ratones como modelo no se ha tenido en cuenta la falta de selectividad de los nucleósidos ensayados frente a LINEs y LTRs por lo que los efectos observados podrían deberse a la inhibición de uno, otro o ambos.

4. Conclusiones

4 CONCLUSIONES:

Conclusiones

1. La evaluación de un total de 33 análogos de nucleósidos en células HeLa, a través de un ensayo de retrotransposición que emplea un gen reportero, ha permitido la identificación de 3 compuestos como inhibidores selectivos de la retrotransposición de LINE-1 de mamíferos: β-ddAC, emtricitabina y lamivudina.

2. El compuesto β-ddAC ha revelado por primera vez ser un inhibidor de la retrotransposición de LINE-1 y ligeramente más potente que emtricitabina y lamivudina.

3. Todos los compuestos que han mostrado selectividad frente a la inhibición de LINE-1 de mamíferos tienen estructuralmente en común que poseen una base nitrogenada derivada de citosina.

4. Los inhibidores selectivos de la retrotransposición de LINE-1 β-ddAC, emtricitabina y lamivudina mostraron nula toxicidad tanto a corto (MTT) como a largo plazo (pU6ineo) en un amplio rango de concentraciones (5, 25, 50, 150 μM) y en dos líneas celulares: HeLA, con baja expresión endógena de LINE-1 y PA-1, con alta expresión endógena de LINE-1.

5. Los inhibidores selectivos de la retrotransposición de LINE-1 β-ddAC, emtricitabina y lamivudina mantuvieron su efecto con leves variaciones en su potencia al ser evaluados en distintas líneas celulares, HeLa y HEK293T, y empleando distintos sistemas reporteros de la retrotransposición: resistencia a neomicina, resistencia a blasticidina, luciferasa.

Conclusions:

1. The *in vitro* evaluation of 33 nucleoside analogues on HeLa cells, by means of a retrotransposition assay that uses a reporter gene, has allowed the characterization of 3 nucleosides that selectively inhibited the retrotransposition of mammalian LINE-1s: β-ddAC, emtricitabine y lamivudine.

2. The synthesized nucleoside analogue β-ddAC has revealed for the first time to be a LINE-1 retrotransposition inhibitor and, in addition, it is slightly more potent than emtricitabine y lamivudine.

3. Structurally, all compounds exhibiting selective inhibition of mammalian LINE-1 retrotransposition share that their nitrogenous base is derived from cytosine.

4. The selective inhibitors of LINE-1 retrotransposition β-ddAC, emtricitabine y lamivudine showed neither acute toxicity (MTT) nor long-term toxicity (pU6ineo) in a wide range of concentrations (5, 25, 50, 150 μM) and in two cell lines: HeLA, which express low levels of endogenous LINE-1 and PA-1, which express high levels of endogenous LINE-1.

5. The selective inhibitors of LINE-1 retrotransposition β-ddAC, emtricitabine y lamivudine exhibited similar effects with minor changes in potency when they were evaluated on different cell lines, HeLa and HEK293T, and using three reporter to measure the retrotransposition: neomycin resistance, blasticidina resistance and luciferase.

5. Materiales y Métodos

5 MATERIALS & METHODS

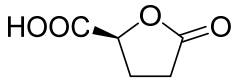
5.1 CHEMISTRY

5.1.1 General

All solvents and chemicals were used as purchased without further purification. Thin layer chromatography was performed on pre-coated silica gel 60 F_{254} plates (Merck), and silica gel 60 (230-400 mesh, Aldrich) was used for flash column chromatography (FCC). Preparative TLC was developed on pre-coated silica gel 60 F_{254}, using 0.5 mm glass plates (Merck). Optical rotatory power was determined employing a JASCO DIP-370 polarimeter. NMR spectra are provided in Supplemental Document 1, and were recorded using the following spectrometers: Varian Inova Unity 300 MHz, Varian Direct Drive 400 MHz, Varian Direct Drive 500 MHz and Varian Direct Drive 600 MHz. Chemical shifts (δ) are reported in parts per million relative to the residual peak of the deuterated solvent. High-resolution mass spectra were recorded on a Micromass LCT time-of-flight instrument using electrospray ionization (ESI). Low-resolution mass spectra (LRMS) were obtained operating in an electrospray ionization mode (ESI) coupled to high resolution liquid chromatography in a simple Quadrupole Agilent 6110 instrument, provided with a Zorbax Eclipse XDB-C18 4.6 x 150 mm column.

5.1.2 Synthesis of 2',3'-dideoxynucleosides

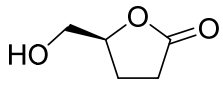
(*S*)-γ-Butyrolactone-γ-carboxylic acid (53). An aqueous solution of $NaNO_2$



(5.2 g, 75.4 mmol in 7.2 mL of water) and 12.5 mL of 5.6 N HCl were simultaneously added dropwise with pressure-equalizing dropping funnels to a suspension of L-glutamic acid (7.35 g, 50 mmol) in 25 mL of water over 2 h. While adding solutions, reaction mixture was kept at 0 °C and then allowed to reach room temperature and stirred overnight. The solvent was coevaporated with toluene under vacuum below 40 °C. The residue was dissolved in MeOH and stirred with 2 g of Dowex 50WX8 for 2 h in order to remove unreacted L-glutamic acid. Solids were filtered out, washed with MeOH and

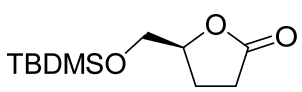
concentrated under vacuum to afford **53** as a crystalline solid, which was used in the next step without further purification.

(*S*)-γ-(Hydroxymethyl)-γ-butyrolactone (54). $BH_3{\cdot}SMe_2$ (9.2 mL, 92.24 mmol)



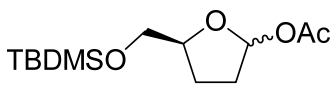
was added, at 1 mL/10 min approximately, under Ar atmosphere to a solution of **53** (8 g, 61.49 mmol) in anhydrous THF (40 mL). During the addition, reaction was cooled with an ice bath, then allowed to reach room temperature and stirred overnight. MeOH was carefully added to quench the reaction and the solvent was removed under vacuum. MeOH was added and evaporated twice and the residue was finally dried by coevaporation with toluene twice, obtaining **54** as a pale yellow syrup that was used without further purification.

(*S*)-γ-[[(*tert*-Butyldimethylsilyl)oxy]methyl]-γ-butyrolactone (55). Alcohol **54**



(9.7 g, 83.54 mmol) was dissolved in DCM (50 mL) followed by the addition of imidazole (7.5 g, 110.16 mmol). Reaction mixture was cooled with an ice bath and TBDMSCl (15.11 g, 100.24 mmol) was then added. The mixture was stirred overnight, letting it to slowly reach room temperature. The reaction was partitioned between water and DCM, and the organic layer was dried over Mg_2SO_4 and concentrated. Through FCC using hexane/diethyl ether (9:1) **55** was obtained (10.7 g, 48% from L-glutamic acid) as a colourless syrup. $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{28}}$ +11.8° (*c* 0.9, $CHCl_3$) (lit. $[\alpha]_D$ +11.11° (*c* 0.92, $CHCl_3$) [481]); **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.60 – 4.54 (m, 1H, H4), 3.85 (dd, $J_{5a,5b}$ = 11.3, $J_{5a,4}$ = 3.3 Hz, 1H, H5a), 3.68 (dd, $J_{5b,4}$ = 3.2 Hz, 1H, H5b), 2.65 – 2.55 (m, 1H, H2a), 2.50 – 2.41 (m, 1H, H2b), 2.31 – 2.22 (m, 1H, H3a), 2.22 – 2.11 (m, 1H, H3b), 0.88 [s, 9H, $SiC(CH_3)_3$], 0.06 [2s, 6H, $Si(CH_3)_2$] [481]. ^{13}C NMR (100 MHz, $CDCl_3$) δ 177.7 (C1), 80.2 (C4), 65.0 (C5), 28.7 (C2), 25.9 [$Si(CH_3)_2$], 23.7 (C3), 18.4 (Si*C*), -5.3 and -5.4 [$SiC(CH_3)_3$] [482].

1-*O*-Acetyl-5-(*tert*-butyldimethylsilyl)-2,3-D-dideoxy-ribofuranose (57). To a



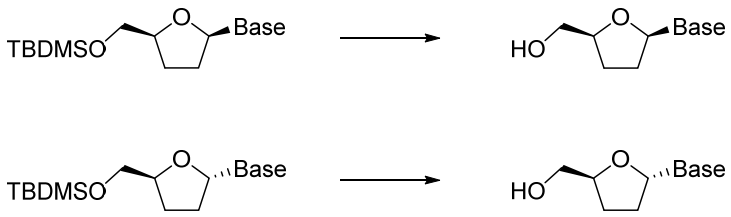
solution of **55** (500 mg, 2.17 mmol) in anhydrous DCM (10 mL) at -78 °C, 1 M solution of DIBAL-H in toluene (3.26 mL, 3.26 mmol) was added dropwise (10 min approximately) under Ar atmosphere. The reaction mixture was then stirred for an additional 45 min

at -78 °C, after which a solution of DMAP (265 mg, 2.17 mmol) in 4 mL of DCM, TEA (907 µL, 6.5 mmol) and Ac_2O (1026 µL, 10.85 mmol) were added. The reaction was stirred overnight allowing it to slowly reach room temperature and then a saturated aqueous solution of sodium potassium tartaric acid (4 mL) and a saturated aqueous solution of ammonium chloride (4 mL) were added and stirred for 30 min. The mixture was extracted with DCM, dried over Mg_2SO_4 and evaporated under vacuum. Purification through FCC with hexane/diethyl ether (95:5 + 1% TEA) led to the obtaining of the mixture of anomers **57** (400 mg, 67%) as a colourless syrup. **^{1}H NMR** (300 MHz, $CDCl_3$) δ 6.30 (d, *J* = 4.3 Hz, 1H), 6.26 (s, 1H), 4.35 – 4.26 (m, 1H), 4.21 – 4.11 (m, 1H), 3.73 (dd, *J* = 10.5, 5.4 Hz, 1H), 3.63 – 3.58 (m, 3H), 2.04 – 1.82 (m, 14H), 0.89 (2s, 18H), 0.05 (2s, 12H) [481].

5.1.3 General procedure for sugar-base coupling to 57

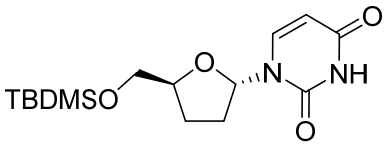
Coupling reactions between sugar moiety **57** and nitrogenous base were carried out as previously described [481]. A mixture of the nitrogenous base (0.73 mmol), HMDS (1.5 mL, 7.30 mmol), and $(NH_4)_2SO_4$ (cat.) was refluxed during 2 hours until a clear solution was obtained. The reaction mixture was cooled and evaporated in vacuum. The residue was dissolved in dry 1,2-dichloroethane (3 mL), under argon atmosphere, and a solution of acetate **1** (100 mg, 0.36 mmol) in dry 1,2-dichloroethane (2 mL) was added. The reaction mixture was cooled in an ice-bath, treated with $EtAlCl_2$ in toluene (1.8 M in toluene, 222 µL, 0.4 mmol) and allowed to stir for 16 hours at room temperature. The reaction was then quenched with saturated aqueous $NaHCO_3$ solution and partitioned between CH_2Cl_2 and water. The organic layer was dried over Mg_2SO_4, filtered and concentrated. The residue was initially purified by FCC to obtain a mixture of isomeric nucleosides that was next separated by preparative thin layer chromatography on a glass plate. The silica was scraped off and percolated through a pad of silica gel using DCM/MeOH (90:10).

5.1.4 General procedure for the desilylation of protected nucleosides

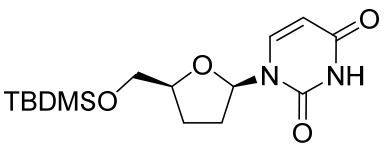


Each protected nucleoside (1.0 equiv) was dissolved in THF and treated with TBAF (1 M in THF, 1.5 equiv). After stirring during 15-30 minutes, the solvent evaporated under reduced pressure. The residue was dissolved in water and passed through a C8 SPE (Sulpelco, bed wt. 500 mg, volume 3 mL) eluting with water. Each tube was analysed with an LC-MS instrument in order to collect those tubes containing the unprotected nucleoside and lacking tetrabutylammonium salts.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-uracil (5'-*O*-silyl-protected α-ddU) has been previously synthesized [483]. TLC diethyl ether/hexane (9:1); preparative TLC diethyl ether/hexane (1:1). White solid (22 mg, 18%). **mp:** 126-128 °C; $[\alpha]_D^{29}$ -42.6° (*c* 1, CH_3Cl); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.15 (s, 1H, NH), 7.35 (d, $J_{6,5}$ = 8.1 Hz, 1H, H6), 6.05 (dd, $J_{1',2'a}$ = 6.2, $J_{1',2'b}$ = 3.1 Hz, 1H, H1'), 5.72 (d, 1H, H5), 4.40 (m, 1H, H4'), 3.68 (dd, $J_{5'a,5'b}$ = 10.9, $J_{5'a,4'}$ = 4.3 Hz, 1H, H5'a), 3.63 (dd, $J_{5'b,4'}$ = 4.1 Hz, 1H, H5'b), 2.51 (m, 1H, H2'a), 2.03 – 1.95 (m, 3H, H2'a, H3'a and H3'b), 0.90 [s, 9H, $SiC(CH_3)_3$], 0.07 [s, 6H, $Si(CH_3)_2$] [483]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 163.6 and 150.3 (C2 and C4), 139.4 (C6), 101.9 (C5), 88.1 (C1'), 82.2 (C4'), 65.4 (C5'), 32.9 (C2'), 26.0 [$SiC(CH_3)_3$], 25.8 (C3'), 18.4 (SiC), -5.21 and -5.27 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M + Na]^+$ calcd. for $C_{15}H_{26}N_2O_4NaSi$, 349.1560; found 349.1563.

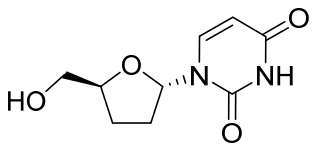


1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-uracil (5'-*O*-silyl-protected β-ddU) has been previously synthesized [483]. TLC diethyl ether/hexane (9:1); preparative TLC diethyl ether/hexane (1:1). Colourless syrup (28 mg, 23%). $[\alpha]_D^{28}$ +23.9° (*c* 1.2, CH_3Cl); **^{1}H NMR** (500 MHz,



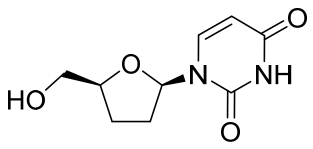
$CDCl_3$) δ 9.21 (s, 1H, NH), 8.08 (d, $J_{6,5}$ = 8.1 Hz, 1H, H6), 6.07 (dd, $J_{1',2'a}$ = 6.5, $J_{1',2'b}$ = 3.1 Hz, 1H, H1'), 5.65 (d, 1H, H5), 4.18 – 4.13 (m, 1H, H4'), 4.05 (dd, $J_{5'a,5'b}$ = 11.5, $J_{5'a,4'}$ = 2.4 Hz, 1H, H5'a), 3.70 (dd, $J_{5'b,4'}$ = 2.4 Hz, 1H, H5'b), 2.43 – 2.34 (m, 1H, H2'a), 2.10 – 2.04 (m, 1H, H2'b), 2.03 – 1.96 (m, 1H, H3'a), 1.92 – 1.86 (m, 1H, H3'b), 0.90 [s, 9H, $SiC(CH_3)_3$], 0.09 [2s, 6H, $Si(CH_3)_2$] [483]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 163.8 and 150.5 (C2 and C4), 140.7 (C6), 101.5 (C5), 86.4 (C1'), 82.2 (C4'), 63.7 (C5'), 33.6 (C2'), 26.0 [$SiC(CH_3)_3$], 24.4 (C3'), 18.6 (Si*C*), -5.33 and -5.43 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_{15}H_{27}N_2O_4Si$, 327.1740; found, 327.1746.

2',3'-Dideoxy-α-D-uridine (α-ddU) has been previously synthesized [484].



Obtained as a waxy solid (8 mg, quant.) from **5'-*O*-silyl-protected α-ddU** (12 mg, 0.037 mmol). TLC dichloromethane/methanol (9:1). **mp:** 96-98 °C (lit. waxy solid [485]); $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{28}}$ -28.9 (*c* 0.5, MeOH) (lit. $[\alpha]_D^{20}$ -13.1 (*c* 0.5, MeOH) [486]); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.99 (s, 1H, NH), 7.37 (d, $J_{6,5}$ = 8.1 Hz, 1H, H6), 6.10 (dd, $J_{1',2'a}$ = 6.2, $J_{1',2'b}$ = 4.5 Hz, 1H, H1'), 5.74 (d, 1H, H5), 4.50 – 4.42 (m, 1H, H4'), 3.76 (dd, $J_{5'a,5'b}$ = 11.9, $J_{5'a,4'}$ = 3.2 Hz, 1H, H5'a), 3.58 (dd, $J_{5'b,4'}$ = 5.7 Hz, 1H, H5'b), 2.60 – 2.47 (m, 1H, H2'a), 2.12 – 2.01 (m, 3H, H2'b, H3'a and OH), 1.99 – 1.88 (m, 1H, H3'b); **^{13}C NMR** (125 MHz, $CDCl_3$) δ 163.4 and 150.4 (C2 and C4), 139.5 (C6), 102.3 (C5), 87.7 (C1'), 82.1 (C4'), 64.8 (C5'), 32.9 (C2'), 25.9 (C3'); **HRMS** (*m*/*z*): $[M + Na]^+$ calcd. for $C_9H_{12}N_2O_4Na$, 235.0695; found 235.0694.

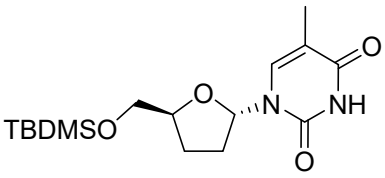
2',3'-Dideoxy-β-D-uridine (β-ddU) has been previously synthesized [487].



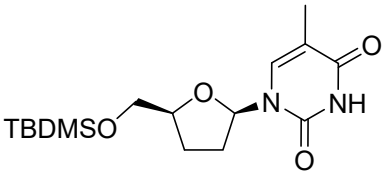
Obtained as a white solid (6 mg, quant.) from **5'-*O*-silyl-protected β-ddU** (9 mg, 0.027 mmol). TLC dichloromethane/methanol (9:1). **mp:** 122-124 °C (lit. 129-131 °C [486], lit. 121-122 °C [488]); $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{28}}$ +52.3 (*c* 0.5, MeOH) (lit. for its enantiomer $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{20}}$ -30.0 (*c* 0.2, MeOH) [486]); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.55 (s, 1H, NH), 7.79 (d, $J_{6,5}$ = 8.1 Hz, 1H, H6), 6.09 (dd, $J_{1',2'a}$ = 6.9, $J_{1',2'b}$ = 3.8 Hz, 1H, H1'), 5.71 (d, 1H, H5), 4.23 – 4.16 (m, 1H, H4'), 4.01 (dd, $J_{5'a,5'b}$ = 11.8, $J_{5'a,4'}$ = 2.7 Hz, 1H, H5'a), 3.74 (dd, $J_{5'b,4'}$ = 3.8 Hz, 1H, H5'b), 2.48 – 2.38 (m, 1H, H2'a), 2.20 – 2.07 (m, 2H, H2'b and OH), 2.06 – 1.95 (m, 2H, H3'a and H3'b); **^{13}C NMR** (125 MHz, $CDCl_3$) δ 163.3 and 150.3 (C2 and C4), 140.6 (C6), 102.1 (C5), 86.7 (C1'), 81.7 (C4'), 63.5 (C5'), 32.7

(C2'), 25.1 (C3'); **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_9H_{12}N_2O_4Na$, 235.0695; found 235.0683.

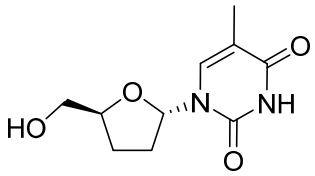
1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-thymine (5'-*O*-silyl-protected α-ddT). TLC diethyl ether/hexane (9:1); preparative TLC diethyl ether/hexane (1:1). Colourless syrup (16 mg, 13%). $[\alpha]_D^{28}$ -26.8 (*c* 0.7, CH_3Cl); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.70 (s, 1H, NH), 7.15 (s, 1H, H6), 6.07 (dd, $J_{1',2'a}$ = 6.3, $J_{1',2'b}$ = 3.8 Hz, 1H, H1'), 4.41 (m, 1H, H4'), 3.68 (dd, $J_{5'a,5'b}$ = 10.9, $J_{5'a,4'}$ = 4.3 Hz, 1H, H5'a), 3.63 (dd, $J_{5'b,4'}$ = 4.1 Hz, 1H, H5'b), 2.53 – 2.45 (m, 1H, H2'a), 2.06 – 1.95 (m, 3H, H2'b, H3'a and H3'b), 1.93 (s, 3H, Me), 0.90 [s, 9H, $SiC(CH_3)_3$], 0.07 [s, 6H, $Si(CH_3)_2$]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 163.9 and 150.2 (C2 and C4), 135.3 (C6), 110.4 (C5), 87.5 (C1'), 81.9 (C4'), 65.4 (C5'), 32.8 (C2'), 26.0 (C3'), 26.0 [$SiC(CH_3)_3$], 18.4 (Si*C*), 12.8 (CH_3), -5.20 and -5.26 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{16}H_{28}N_2O_4NaSi$, 363.1716; found 363.1717.



1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-thymine (5'-*O*-silyl-protected β-ddT) has been previously synthesized [489]. TLC diethyl ether/hexane (9:1); preparative TLC diethyl ether/hexane (1:1). White solid (28 mg, 22%). **mp:** 117-119 °C (lit. 125-128 °C [489]); $[\alpha]_D^{29}$ +11.4 (*c* 1, CH_3Cl) (lit. $[\alpha]_D^{23}$ +7 (*c* 1, $CHCl_3$) [490]; **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.31 (s, 1H, NH), 7.58 (d, $J_{6,Me}$ = 1.1 Hz, 1H, H6), 6.07 (dd, $J_{1',2'a}$ = 6.4, $J_{1',2'b}$ = 4.5 Hz, 1H, H1'), 4.19 – 4.11 (m, 1H, H4'), 3.99 (dd, $J_{5'a,5'b}$ = 11.4, $J_{5'a,4'}$ = 2.6 Hz, 1H, H5'a), 3.71 (dd, $J_{5'b,4'}$ = 2.9 Hz, 1H, H5'b), 2.43 – 2.29 (m, 1H, H2'a), 2.02 – 1.96 (m, 3H, H2'a, H3'a and H3'b), 1.92 (d, 3H, Me), 0.93 [s, 9H, $SiC(CH_3)_3$], 0.11 [s, 6H, $Si(CH_3)_2$] [490]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 164.1 and 150.5 (C2 and C4), 135.8 (C6), 110.3 (C5), 85.9 (C1'), 81.1 (C4'), 64.6 (C5'), 32.7 (C2'), 26.0 [$SiC(CH_3)_3$], 25.5 (C3'), 18.6 (Si*C*), 12.7 (CH_3), -5.17 and -5.20 [$Si(CH_3)_2$] [491]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{16}H_{28}N_2O_4NaSi$, 363.1716; found 363.1724.

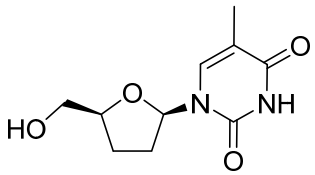


2',3'-Dideoxy-α-D-thymidine (α-ddT) has been previously synthesized [492].



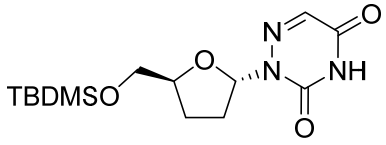
Obtained as a white solid (6 mg, quant.) from **5'-*O*-silyl-protected α-ddT** (9 mg, 0.026 mmol). TLC dichloromethane/methanol (9:1). **mp:** 102-104 °C (lit. 105-107 °C [486]); $[\alpha]_D^{29}$ -17.3 (*c* 0.2, MeOH) (lit. $[\alpha]_D^{20}$ -17.2 (*c* 0.1, MeOH) [486]); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.48 (s, 1H, NH), 7.15 (s, 1H, H6), 6.12 (t, $J_{1',2'a} = J_{1',2'b}$ = 5.7 Hz, 1H, H1'), 4.48 (m, 1H, H4'), 3.76 (dd, $J_{5'a,5'b}$ = 11.9, $J_{5'a,4'}$ = 3.2 Hz, 1H, H5'a), 3.58 (dd, $J_{5'b,4'}$ = 5.7 Hz, 1H, H5'b), 2.55 – 2.47 (m, 1H, H2'a), 2.14 – 1.99 (m, 2H, H2'b and H3'a), 1.98 – 1.88 (m, 1H, H3'b), 1.94 (s, 3H, Me) 1.63 (s, 1H, OH); **^{13}C NMR** (125 MHz, $CDCl_3$) δ 163.7 and 150.3 (C2 and C4), 135.2 (C6), 110.9 (C5), 87.2 (C1'), 81.8 (C4'), 64.9 (C5'), 32.7 (C2'), 26.1 (C3'), 12.8 (*C*H_3); **HRMS** (*m*/*z*): $[M + Na]^+$ calcd. for $C_{10}H_{14}N_2O_4Na$, 249.0851; found 249.0837.

2',3'-Dideoxy-β-D-thymidine (β-ddT) has been previously synthesized [493].



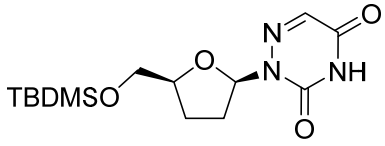
Obtained as a white solid (10 mg, quant.) from **5'-*O*-silyl-protected β-ddT** (15 mg, 0.026 mmol). TLC dichloromethane/methanol (9:1). **mp:** 150-152 °C (lit. 154-155 °C [486]); $[\alpha]_D^{28}$ +35.9 (*c* 0.6, MeOH) (lit. $[\alpha]_D^{20}$ -30.9 (*c* 0.4, MeOH) [486]); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.88 (s, 1H, NH), 7.52 (d, $J_{6,Me}$ = 1.1 Hz, 1H, H6), 6.10 (dd, $J_{1',2'a}$ = 7.0, $J_{1',2'b}$ = 4.0 Hz, 1H, H1'), 4.20 – 4.15 (m, 1H, H4'), 3.99 (dd, $J_{5'a,5'b}$ = 12.0, $J_{5a',4'}$ = 2.7 Hz, 1H, H5'a), 3.73 (dd, $J_{5'b,4'}$ = 4.0 Hz, 1H, H5'b), 2.44 – 2.35 (m, 1H, H2'a), 2.26 (s, 1H, OH), 2.12 – 1.98 (m, 3H, H2'b, H3'a and H3'b), 1.90 (d, 3H, CH_3) [491]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 164.0 and 150.5 (C2 and C4), 136.4 (C6), 110.7 (C5), 86.3 (C1'), 81.3 (C4'), 63.6 (C5'), 32.2 (C2'), 25.3 (C3'), 12.7 (*C*H_3) [491]; **HRMS** (*m*/*z*): $[M + Na]^+$ calcd. for $C_{10}H_{14}N_2O_4Na$, 249.0851; found 249.0866.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-6-azauracil (5'-*O*-silyl-protected α-ddAU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:2). Colourless syrup (28 mg, 23%). $[\alpha]_D^{28}$ +10.6 (*c* 1.1, $CHCl_3$); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.35 (s, 1H, NH), 7.43 (s, 1H, H5), 6.44 (dd, $J_{1',2'a}$ = 7.1, $J_{1',2'b}$ = 2.7 Hz, 1H, H1'), 4.45 – 4.33 (m, 1H, H4'), 3.65 (dd, $J_{5'a,5'b}$ = 10.8, $J_{5'a,4'}$ = 4.3 Hz, 1H, H5'a), 3.61 (dd, $J_{5'b,4'}$ = 4.0 Hz, 1H,

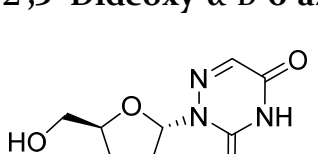


H5'b), 2.40 – 2.31 (m, 1H, H2'a), 2.30 – 2.18 (m, 2H, H2'a and H3'a), 1.97 – 1.86 (m, 1H, H3'b), 0.90 [s, 9H, SiC(CH_3)$_3$], 0.06 [s, 6H, Si(CH_3)$_2$]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 155.9 and 147.8 (C2 and C4), 135.4 (C5), 87.7 (C1'), 81.9 (C4'), 65.5 (C5'), 30.3 (C2'), 26.8 (C3'), 26.0 [SiC(CH_3)$_3$], 18.5 (Si*C*), -5.16 and -5.25 [Si(CH_3)$_2$]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{14}H_{25}N_3O_4NaSi$, 350.1512; found 350.1506.

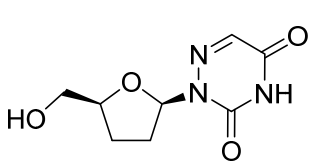
1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-6-azauracil (5'-*O*-silyl-protected β-ddAU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:2). Colourless syrup (36 mg, 30%). $[\alpha]_D^{30}$ -69.6 (*c* 0.5, $CHCl_3$); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.03 (s, 1H, NH), 7.42 (s, 1H, H5), 6.38 (dd, $J_{1',2'a}$ = 6.6, $J_{1',2'b}$ = 3.5 Hz, 1H, H1'), 4.17 – 4.11 (m, 1H, H4'), 3.72 (dd, $J_{5'a,5'b}$ = 10.6, $J_{5'a,4'}$ = 5.3 Hz, 1H, H5'a), 3.65 (dd, $J_{5'b,4'}$ = 5.5 Hz, 1H, H5'b), 2.32 – 2.23 (m, 2H, H2'a and H2'b), 2.07 – 1.99 (m, 2H, H3'a and H3'b), 0.88 [s, 9H, SiC(CH_3)$_3$], 0.04 [2s, 6H, Si(CH_3)$_2$]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 155.7 and 147.8 (C2 and C4), 135.3 (C5), 86.8 (C1'), 82.3 (C4'), 65.6 (C5'), 30.1 (C2'), 27.4 (C3'), 26.0 [SiC(CH_3)$_3$], 18.5 (Si*C*), -5.13 and -5.17 [Si(CH_3)$_2$]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{14}H_{25}N_3O_4NaSi$, 350.1512; found 350.1515.



2',3'-Dideoxy-α-D-6-azauridine (α-ddAU). Obtained as a colourless syrup (14 mg, quant.) from **5'-*O*-silyl-protected α-ddAU** (21 mg, 0.064 mmol). TLC dichloromethane/methanol (9:1). $[\alpha]_D^{28}$ +128.3 (*c* 1, MeOH); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.95 (s, 1H, NH), 7.43 (s, 1H, H5), 6.46 (t, $J_{1',2'a}$ = $J_{1',2'b}$ = 5.4 Hz, 1H, H1'), 4.46 – 4.39 (m, 1H, H4'), 3.74 (dd, $J_{5'a,5'b}$ = 12.0, $J_{5'a,4'}$ = 2.9 Hz, 1H, H5'a), 3.53 (dd, $J_{5'b,4'}$ = 5.4 Hz, 1H, H5'b), 2.58 (s, 1H, OH), 2.39 – 2.32 (m, 2H, H2'a and H2'b), 2.29 – 2.20 (m, 1H, H3'a), 1.91 – 1.82 (m, 1H, H3'b); **^{13}C NMR** (125 MHz, $CDCl_3$) δ 156.2 and 148.3 (C2 and C4), 135.7 (C5), 87.3 (C1'), 81.9 (C4'), 64.7 (C5'), 30.0 (C2'), 26.5 (C3'); **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_8H_{11}N_3O_4Na$, 236.0647; found 236.0658.



2',3'-Dideoxy-β-D-6-azauridine (β-ddAU) has been previously synthesized [494]. Obtained as a colourless syrup (5 mg, quant.) from **5'-*O*-silyl-protected β-ddAU** (8 mg, 0.024 mmol). TLC dichloromethane/methanol (9:1). $[\alpha]_D^{28}$ -117.3 (*c* 0.2,



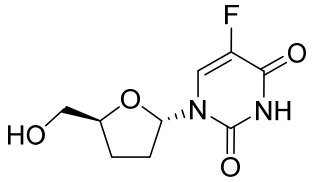
MeOH); **^{1}H NMR** (600 MHz, $CDCl_3$) δ 8.80 (s, 1H, NH), 7.46 (s, 1H, H5), 6.40 (dd, $J_{1',2'a}$ = 6.9, $J_{1',2'b}$ = 3.8 Hz, 1H, H1'), 4.29 – 4.23 (m, 1H, H4'), 3.85 (dd, $J_{5'a,5'b}$ = 12.0, $J_{5'b,4'}$ = 2.9 Hz, 1H, H5'a), 3.61 (dd, $J_{5'b,4'}$ = 4.9 Hz, 1H, H5'b), 2.38 – 2.27 (m, 2H, H2'a and H2'b), 2.20 – 2.11 (m, 1H, H3'a), 2.06 – 2.00 (m, 1H, H3'b), 1.60 (s, 1H, OH) [494]; **^{13}C NMR** (150 MHz, $CDCl_3$) δ 155.4 and 147.6 (C2 and C4), 135.8 (C5), 87.2 (C1'), 82.4 (C4'), 64.8 (C5'), 31.0 (C2'), 26.1 (C3') [494]; **HRMS** (*m*/*z*): [M - H]$^-$ calcd. for $C_8H_{10}N_3O_4$, 212.0671; found 212.0666.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-5-fluorouracil (5'-*O*-silyl-protected α-ddFU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:1). White solid (19 mg, 15%). **mp:** 106-108 °C (lit. 105-107 °C [495]); $[\alpha]_D^{29}$ -43.3 (*c* 1, $CHCl_3$); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.53 (s, 1H, NH), 7.42 (d, $J_{6,F}$ = 6.1 Hz, 1H, H6), 6.04 – 6.01 (m, 1H, H1'), 4.48 – 4.36 (m, 1H, H4'), 3.69 (dd, $J_{5'a,5'b}$ = 11.0, $J_{5'a,4'}$ = 4.2 Hz, 1H, H5'a), 3.63 (dd, $J_{5'b,4'}$ = 4.1 Hz, 1H, H5'b), 2.59 – 2.48 (m, 1H, H2'a), 2.04 – 1.93 (m, 3H, H2'a, H3'a and H3'b), 0.90 [s, 9H, $SiC(CH_3)_3$], 0.07 [s, 6H, $Si(CH_3)_2$] [495]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 157.2 (d, $J_{4,F}$ = 26.6 Hz, C4), 148.9 (C2), 140.5 (d, $J_{5,F}$ = 236.9 Hz, C5), 123.8 (d, $J_{6,F}$ = 33.7 Hz, C6), 88.2 (C1'), 82.3 (C4'), 65.3 (C5'), 32.9 (C2'), 26.0 [$SiC(CH_3)_3$], 25.7 (C3'), 18.4 (Si*C*), -5.23 and -5.29 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*)=: [M + Na]$^+$ calcd. for $C_{15}H_{25}N_2O_4NaFSi$, 367.1465; found 367.1450.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-5-fluorouracil (5'-*O*-silyl-protected β-ddFU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:1). White solid (35 mg, 28%). **mp:** 144-146 °C (lit. 133-135 °C [495]); $[\alpha]_D^{29}$ +21.2 (*c* 1.25, $CHCl_3$); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.00 (s, 1H, NH), 8.26 (d, $J_{6,F}$ = 6.4 Hz, 1H, H6), 6.07 – 6.02 (m, 1H, H1'), 4.20 – 4.14 (m, 1H, H4'), 4.10 (dd, $J_{5'a,5'b}$ = 11.7, $J_{5'a,4'}$ = 2.2 Hz, 1H, H5'a), 3.71 (dd, $J_{5'b,4'}$ = 2.1 Hz, 1H, H5'b), 2.45 – 2.36 (m, 1H, H2'a), 2.11 – 2.01 (m, 2H, H2'b and H3'a), 1.96 – 1.89 (m, 1H, H3'b), 0.93 [s, 9H, $SiC(CH_3)_3$], 0.12 [2s, 6H, $Si(CH_3)_2$] [495]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 157.1 (d, $J_{4,F}$ = 27.1 Hz, C4), 148.9 (C2), 140.3 (d, $J_{5,F}$ = 235.5 Hz, C5), 125.0 (d, $J_{6,F}$ = 34.3 Hz, C6), 86.7 (C1'), 82.3 (C4'), 63.9 (C5'),

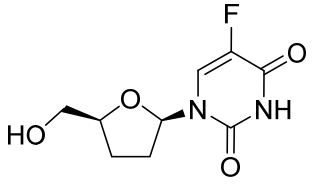
33.6 (C2'), 26.1 [SiC(CH_3)$_3$], 24.4 (C3'), 18.7 (Si*C*), -5.4[Si(CH_3)$_2$]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{15}H_{25}N_2O_4NaFSi$, 367.1465; found 367.1450.

2',3'-Dideoxy-α-D-5-fluorouridine (α-ddFU) has been previously synthesized



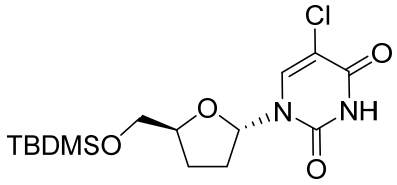
[496]. Obtained as a white solid (20 mg, quant.) from **5'-*O*-silyl-protected α-ddFU** (30 mg, 0.087 mmol). TLC dichloromethane/methanol (9:1). **mp:** 132-134 °C (lit. 141-143 °C [496]); $[\alpha]_D^{28}$ -41.7 (*c* 1.25, MeOH); **^{1}H NMR** (500 MHz, CD_3OD) δ 7.80 (d, $J_{6,F}$ = 6.6 Hz, 1H, H6), 6.06 – 6.02 (m, 1H, H1'), 4.51 – 4.45 (m, 1H, H4'), 3.64 (dd, $J_{5'a,5'b}$ = 11.9, $J_{5'a,4'}$ = 3.8 Hz, 1H, H5'a), 3.52 (dd, $J_{5'b,4'}$ = 5.1 Hz, 1H, H5'b), 2.51 – 2.42 (m, 1H, H, H2'a), 2.13 – 2.03 (m, 2H, H2'b and H3'a), 1.93 – 1.84 (m, 1H, H3'b); **^{13}C NMR** (125 MHz, CD_3OD) δ 159.6 (d, $J_{4,F}$ = 26.0 Hz, C4), 150.7 (C2), 141.8 (d, $J_{5,F}$ = 232.7 Hz, C5), 126.0 (d, $J_{6,F}$ = 34.3 Hz, C6), 89.1 (C1'), 83.4 (C4'), 65.1 (C5'), 33.2 (C2'), 26.7 (C3'); **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_9H_{11}N_2O_4FNa$, 253.0601; found 253.0604.

2',3'-Dideoxy-β-D-5-fluorouridine (β-ddFU) has been previously synthesized



[458]. Obtained as a white solid (10 mg, quant.) from **5'-*O*-silyl-protected β-ddFU** (15 mg, 0.043 mmol). TLC dichloromethane/methanol (9:1). **mp:** 123-125 °C (lit. 108-110 °C [495]); $[\alpha]_D^{28}$ +59.4 (*c* 0.62, MeOH) (lit. for its enantiomer $[\alpha]_D^{25}$ -53 (*c* 0.12, MeOH) [495]); **^{1}H NMR** (500 MHz, CD_3OD) δ 8.39 (d, $J_{6,F}$ = 7.0 Hz, 1H, H6), 6.03 – 5.99 (m, 1H, H1'), 4.17 – 4.11 (m, 1H, H4'), 3.91 (dd, $J_{5'a,5'b}$ = 12.3, $J_{5'a,4'}$ = 2.8 Hz, 1H, H5'a), 3.68 (dd, $J_{5'b,4'}$ = 3.3 Hz, 1H, H5'b), 2.44 – 2.34 (m, 1H, H2'a), 2.14 – 2.06 (m, 1H, H2'b), 2.03 – 1.92 (m, 2H, H3'a and H3'b); **^{13}C NMR** (125 MHz, CD_3OD) δ 159.7 (d, $J_{4,F}$ = 26.1 Hz, C4), 150.8 (C2), 141.5 (d, $J_{5,F}$ = 231.3 Hz, C5), 126.6 (d, $J_{6,F}$ = 35.0 Hz, C6), 87.9 (C1'), 83.7 (C4'), 63.3 (C5'), 33.7 (C2'), 25.4 (C3'); **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_9H_{11}N_2O_4FNa$, 253.0601; found 253.0625.

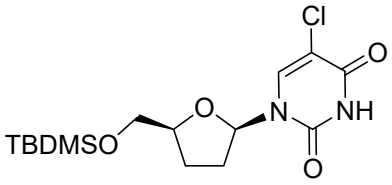
1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-5-



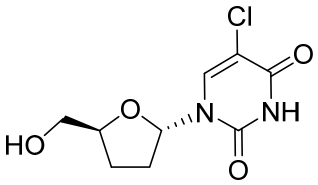
chlorouracil (5'-*O*-silyl-protected α-ddClU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:1). White solid (24 mg, 18%). **mp** =

157-159 °C; $[\alpha]_D^{28}$ -39.0 (*c* 0.1, $CHCl_3$); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.01 (s, 1H, NH), 7.56 (s, 1H, H6), 6.02 (dd, $J_{1',2'a}$ = 6.1, $J_{1',2'b}$ = 3.1 Hz, 1H, H1'), 4.49 – 4.42 (m, 1H, H4'), 3.71 (dd, $J_{5'a,5'b}$ = 10.9, $J_{5'a,4'}$ = 4.1 Hz, 1H, H5'a), 3.64 (dd, $J_{5'b,4'}$ = 4.0 Hz, 1H, H5'b), 2.60 – 2.50 (m, 1H, H2'a), 2.06 – 1.96 (m, 3H, H2'b, H3'a and H3'b), 0.90 [s, 9H, $SiC(CH_3)_3$], 0.08 [s, 6H, $Si(CH_3)_2$]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 159.1 and 149.3 (C2 and C4), 136.5 (C6), 108.7 (C5), 88.6 (C1'), 82.4(C4'), 65.3 (C5'), 33.1 (C2'), 26.0 [$SiC(CH_3)_3$], 25.6 (C3'), 18.4 (Si*C*), -5.21 and -5.27 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{15}H_{25}N_2O_4NaClSi$, 383.1170; found 383.1154.

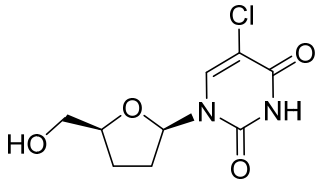
1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-5-chlorouracil (5'-*O*-silyl-protected β-ddClU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:1). White solid (44 mg, 33%). **mp** = 183-185 °C; $[\alpha]_D^{28}$ +2.45 (*c* 1.1, CH_3Cl); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.93 (s, 1H, NH), 8.15 (s, 1H, H6), 6.03 (dd, $J_{1',2'a}$ = 6.3, $J_{1',2'b}$ = 4.0 Hz, 1H, H1'), 4.22 – 4.16 (m, 1H, H4'), 4.06 (dd, $J_{5'a,5'b}$ = 11.6, $J_{5'a,4'}$ = 2.3 Hz, 1H, H5'a), 3.71 (dd, $J_{5'b,4'}$ = 2.5 Hz, 1H, H5'b), 2.46 – 2.36 (m, 1H, H2'a), 2.09 – 1.93 (m, 3H, H2'b, H3'a and H3'b), 0.94 [s, 9H, $SiC(CH_3)_3$], 0.14 [2s, 6H, $Si(CH_3)_2$]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 159.1 and 149.5 (C2 and C4), 137.4 (C6), 108.6 (C5), 87.0 (C1'), 82.0 (C4'), 64.3 (C5'), 33.4 (C2'), 26.1 [$SiC(CH_3)_3$], 24.9 (C3'), 18.7 (Si*C*), -5.15 and -5.16 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{15}H_{25}N_2O_4NaClSi$, 383.1170; found 383.1159.



2',3'-Dideoxy-α-D-5-chlorouridine (α-ddClU). Obtained as a white solid (13 mg, quantitative) from **5'-*O*-silyl-protected α-ddClU** (19 mg, 0.053 mmol). TLC (dichloromethane/methanol 9:1). **mp:** 165-167 °C; $[\alpha]_D^{28}$ -26.50 (*c* 0.7, MeOH); **^{1}H NMR** (500 MHz, CD_3OD) δ 7.87 (s, 1H, H6), 6.03 (dd, $J_{1',2'a}$ = 6.3, $J_{1',2'b}$ = 4.1 Hz, 1H, H1'), 4.53 – 4.47 (m, 1H, H4'), 3.66 (dd, $J_{5'a,5'b}$ = 11.9, $J_{5'a,4'}$ = 3.7 Hz, 1H, H5'a), 3.54 (dd, $J_{5'b,4'}$ = 5.2 Hz, 1H, H5'b), 2.54 – 2.43 (m, 1H, H2'a), 2.14 – 2.05 (m, 2H, H2'b and H3'a), 1.95 – 1.86 (m, 1H, H3'b); **^{13}C NMR** (125 MHz, CD_3OD) δ 161.7 and 151.3 (C2 and C4), 138.9 (C6), 109.1 (C5), 89.6 (C1'), 83.6 (C4'), 65.1 (C5'), 33.4 (C2'), 26.7 (C3'); **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_9H_{11}N_2O_4NaCl$, 269.0305; found 269.0328.



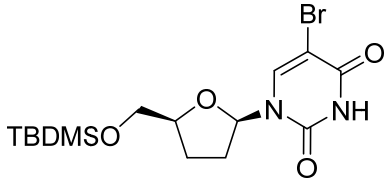
2',3'-Dideoxy-β-D-5-chlorouridine (β-ddClU) has been previously synthesized



[497]. Obtained as a white solid (13 mg, quantitative) from **5'-*O*-silyl-protected β-ddClU** (19 mg, 0.053 mmol). TLC (dichloromethane/methanol 9:1). **mp:** 159-161 °C (lit. 154-155 °C [497]); $[\alpha]_D^{28}$ +40.87 (*c* 0.55, MeOH); **^{1}H NMR** (400 MHz, CD_3OD) δ δ 8.56 (s, 1H, H6), 6.03 (dd, $J_{1',2'a}$ = 6.7, $J_{1',2'a}$ = 2.9 Hz, 1H, H1'), 4.23 – 4.14 (m, 1H, H4'), 3.95 (dd, $J_{5'a,5'b}$ = 12.3, $J_{5'a,4'}$ = 2.8 Hz, 1H, H5'a), 3.71 (dd, $J_{5'b,4'}$ = 3.2 Hz, 1H, H5'b), 2.49 – 2.36 (m, 1H, H2'a), 2.20 – 2.10 (m, 1H, H2'b), 2.05 – 1.94 (m, 2H, H3'a and H3'b); **^{13}C NMR** (100 MHz, CD_3OD) δ 161.8 and 151.3 (C2 and C4), 139.8 (C6), 108.6 (C5), 88.2 (C1'), 83.9 (C4'), 63.0 (C5'), 34.0 (C2'), 25.2 (C3'); **HRMS** (*m*/*z*): $[M + Na]^+$ calcd. for $C_9H_{11}N_2O_4NaCl$, 269.0305; found 269.0331.

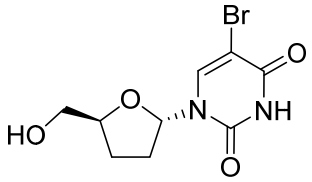
1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-5-bromouracil (5'-*O*-silyl-protected α-ddBrU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:1). White solid (24 mg, 16%). **mp:** 147-149 °C; $[\alpha]_D^{28}$ -20.9 (*c* 0.7, $CHCl_3$); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.84 (s, 1H, NH), 7.67 (s, 1H, H6), 6.02 (dd, $J_{1',2'a}$ = 6.1, $J_{1',2'b}$ = 3.0 Hz, 1H, H1'), 4.50 – 4.41 (m, 1H, H4'), 3.71 (dd, $J_{5'a,5'b}$ = 11.0, $J_{5'a,4'}$ = 4.1 Hz, 1H, H5'a), 3.64 (dd, $J_{5'b,4'}$ = 4.0 Hz, 1H, H5'b), 2.61 – 2.49 (m, 1H, H2'a), 2.06 – 1.96 (m, 3H, H2'b, H3'a and H3'b), 0.91 [s, 9H, $SiC(CH_3)_3$], 0.08 [s, 6H, $Si(CH_3)_2$]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 159.1 and 149.5 (C2 and C4), 139.1 (C6), 96.3 (C5), 88.7 (C1'), 82.4 (C4'), 65.3 (C5'), 33.2 (C2'), 26.0 [$SiC(CH_3)_3$], 25.6 (C3'), 18.4 (Si*C*), -5.21 and -5.26 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M - H]^-$ calcd. for $C_{15}H_{24}N_2O_4SiBr$, 403.0689; found 403.0679.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-5-bromouracil (5'-*O*-silyl-protected β-ddBrU). TLC diethyl ether/hexane (5:1); preparative TLC diethyl ether/hexane (1:1). White solid (32 mg, 21%). **mp:** 153-155 °C; $[\alpha]_D^{29}$ -3.5 (*c* 1.8, $CHCl_3$); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.44 (s, 1H, NH), 8.17 (s, 1H, H6), 6.02 (dd, $J_{1',2'a}$ = 6.4, $J_{1',2'b}$ = 4.3 Hz, 1H, H1'), 4.21 – 4.15 (m, 1H, H4'), 4.03 (dd, $J_{5'a,5'b}$ = 11.6, $J_{5'a,4'}$ = 2.3



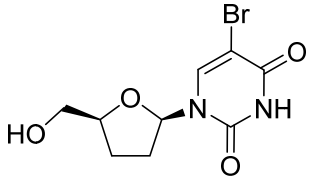
Hz, 1H, H5'a), 3.70 (dd, $J_{5'b,4'}$ = 2.7 Hz, 1H, H5'b), 2.44 – 2.36 (m, 1H, H2'a), 2.08 – 1.91 (m, 3H, H2'b, H3'a and H3'b), 0.93 [s, 9H, SiC(CH_3)$_3$], 0.13 [s, 6H, Si(CH_3)$_2$]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 159.4 and 149.9 (C2 and C4), 139.8 (C6), 96.4 (C5), 87.0 (C1'), 81.9 (C4'), 64.4 (C5'), 33.3 (C2'), 26.2 [SiC(CH_3)$_3$], 25.0 (C3'), 18.7 (Si*C*), -5.07 and -5.09 [Si(CH_3)$_2$]; **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{15}H_{25}N_2O_4NaSiBr$, 427.0665; found 427.0671.

2',3'-Dideoxy-α-D-5-bromouridine (α-ddBrU). Obtained as a colourless syrup



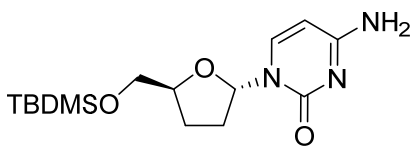
(13 mg, quant.) from **5'-*O*-silyl-protected α-ddBrU** (18 mg, 0.044 mmol). TLC dichloromethane/methanol (9:1). $[\alpha]_D^{28}$ -15.8 (*c* 1, MeOH); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 9.07 (s, 1H, NH), 7.67 (s, 1H, H6), 6.07 (dd, $J_{1',2'a}$ = 6.1, $J_{1',2'b}$ = 4.4 Hz, 1H, H1'), 4.54 – 4.48 (m, 1H, H4'), 3.79 (dd, $J_{5'a,5'b}$ = 12.0, $J_{5'a,4'}$ = 3.2 Hz, 1H, H5'a), 3.60 (dd, $J_{5'b,4'}$ = 5.6 Hz, 1H, H5'b), 2.62 – 2.51 (m, 1H, H2'a), 2.14 – 2.01 (m, 2H, H2'b and H3'a), 2.00 – 1.90 (m, 1H, H3'b), 1.70 (s, 1H, OH); **^{13}C NMR** (125 MHz, $CDCl_3$) δ 159.2 and 149.6 (C2 and C4), 139.1 (C6), 96.7 (C5), 88.2 (C1'), 82.3 (C4'), 64.7 (C5'), 33.1 (C2'), 25.7 (C3'); **HRMS** (*m*/*z*): [M + H]$^+$ calcd. for $C_9H_{11}N_2O_4Br$, 312.9800; found 312.9785.

2',3'-Dideoxy-β-D-5-bromouridine (β-ddBrU) has been previously synthesized



[498]. Obtained as a colourless solid (12 mg, quant.) from **5'-*O*-silyl-protected β-ddBrU** (17 mg, 0.044 mmol). TLC dichloromethane/methanol (9:1). **mp:** 182-184 °C (lit. 178-179 °C [498]); $[\alpha]_D^{27}$ +25.6 (*c* 0.7, MeOH); **^{1}H NMR** (600 MHz, CD_3OD) δ 8.64 (s, 1H, H6), 6.00 (dd, $J_{1',2'a}$ = 6.7, $J_{1',2'b}$ = 2.9 Hz, 1H, H1'), 4.19 – 4.14 (m, 1H, H4'), 3.93 (dd, $J_{5'a,5'b}$ = 12.3, $J_{5'a,4'}$ 2.8 Hz, 1H, H5'a), 3.68 (dd, $J_{5'b,4'}$ = 3.2 Hz, 1H, H5'b), 2.44 – 2.36 (m, 1H, H2'a), 2.15 – 2.09 (m, 1H, H2'b), 2.02 – 1.92 (m, 2H, H3'a and H3'b); **^{13}C NMR** (150 MHz, CD_3OD) δ 161.8 and 151.6 (C2 and C4), 142.4 (C6), 96.2 (C5), 88.3 (C1'), 84.0 (C4'), 63.0 (C5'), 34.0 (C2'), 25.2 (C3'); **HRMS** (*m*/*z*): [M - H]$^-$ calcd. for $C_9H_{10}N_2O_4Br$, 288.9824; found 288.9823.

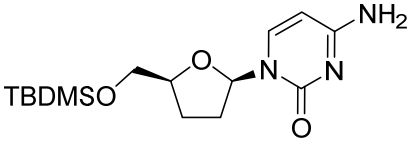
1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-cytosine



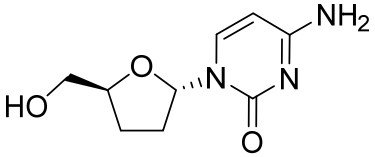
(5'-*O*-silyl-protected α-ddC) has been previously synthesized [481]. TLC dichloromethane/methanol (19:1); preparative TLC

chloroform/isopropanol (90:10). White solid (25 mg, 21%). **mp:** 186-188 °C (lit. 199-203 [481]); $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{28}}$ -50.0 (*c* 0.75, MeOH) (lit. $[\alpha]_D^{25}$ -53.9 (*c* 0.31, MeOH) [481]); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.43 (d, $J_{6,5}$ = 7.4 Hz, 1H, H6), 6.04 (dd, $J_{1',2'a}$ = 6.1, $J_{1',2'b}$ = 3.3 Hz, 1H, H1'), 5.74 (d, 1H, H5), 4.42 – 4.34 (m, 1H, H4'), 3.64 (d, $J_{5'a,4'}$ = $J_{5'b,4'}$ = 4.6 Hz, 2H, H5'a and H5'b), 2.56 – 2.47 (m, 1H, H2'a), 2.06 – 1.95 (m, 1H, H2'b), 1.94 – 1.87 (m, 2H, H3'a and H3'b), 0.89 [s, 9H, $SiC(CH_3)_3$], 0.06 [s, 6H, $Si(CH_3)_2$] [481]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 166.0 (C4), 156.1 (C2), 140.3 (C6), 93.8 (C5), 88.7 (C1'), 82.0 (C4'), 65.5 (C5'), 33.0 (C2'), 26.0 [$SiC(CH_3)_3$], 25.5 (C3'), 18.5 (Si*C*), -5.19 and -5.24 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_{15}H_{28}N_3O_3Si$, 326.1900; found 326.1890.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-cytosine (5'-*O*-silyl-protected β-ddC) has been previously synthesized [481]. TLC dichloromethane/methanol (19:1); preparative TLC chloroform/isopropanol (90:10). White solid (27 mg, 23%). **mp:** 198-200 °C (lit. 204-206 °C [481]); $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{28}}$ +41.1 (*c* 1, MeOH) (lit. $[\alpha]_D^{25}$ +45.8 (*c* 1.02, MeOH) [481]); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.10 (d, $J_{6,5}$ = 7.4 Hz, 1H, H6), 6.07 (dd, $J_{1',2'a}$ = 6.6, $J_{1',2'b}$ = 2.8 Hz, 1H, H1'), 5.63 (d, 1H, H5), 4.15 – 4.09 (m, 1H, H4'), 4.04 (dd, $J_{5'a,5'b}$ = 11.5, $J_{5'a,4'}$ = 2.6 Hz, 1H, H5'a), 3.71 (dd, $J_{5'b,4'}$ = 2.6 Hz, 1H, H5'b), 2.46 – 2.35 (m, 1H, H2'a), 2.10 – 2.02 (m, 1H, H2'b), 1.95 – 1.88 (m, 1H, H3'a), 1.87 – 1.79 (m, 1H, H3'b), 0.91 [s, 9H, $SiC(CH_3)_3$], 0.09 [2s, 6H, $Si(CH_3)_2$] [481]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 165.9 (C4), 156.1 (C2), 141.8 (C6), 93.3 (C5), 87.1 (C1'), 82.2 (C4'), 63.7 (C5'), 33.9 (C2'), 26.0 (C3'), 24.3 [$SiC(CH_3)_3$], 18.5 (Si*C*), -5.29 and -5.39 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_{15}H_{28}N_3O_3Si$, 326.1900; found 326.1891.

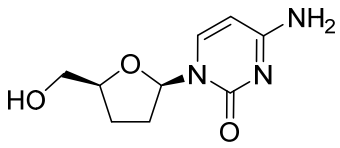


2',3'-Dideoxy-α-D-cytidine (α-ddC) has been previously synthesized [481]. Obtained as a white solid (9 mg, quantitative) from **5'-*O*-silyl-protected α-ddC** (14 mg, 0.043 mmol). TLC dichloromethane/methanol (9:1). **mp:** 168-170 °C (lit. 165-166 °C [486]); $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{27}}$ -74.5 (*c* 0.77, MeOH) (lit. $[\alpha]_D^{20}$ -83.4 (*c* 0.56, MeOH) [481]); **^{1}H NMR** (500 MHz, CD_3OD) δ 7.68 (d, $J_{6,5}$ = 7.5 Hz, 1H, H6), 6.05 (dd, $J_{1',2'a}$ = 6.2, $J_{1',2'b}$ = 3.7 Hz, 1H, H1'), 5.91 (d, 1H, H5), 4.49 – 4.44 (m, 1H, H4'),



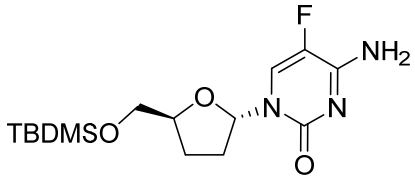
3.63 (dd, $J_{5'a,5'b}$ = 11.8, $J_{5'a,4'}$ = 4.0 Hz, 1H, H5'a), 3.54 (dd, $J_{5'b,4'}$ = 5.2 Hz, 1H, H5'b), 2.54 – 2.45 (m, 1H, H2'a), 2.07 – 1.96 (m, 2H, H2'b and H3'a), 1.93 – 1.84 (m, 1H, H3'b); **^{13}C NMR** (125 MHz, CD_3OD) δ 167.56 (C4), 158.05 (C2), 142.10 (C6), 95.57 (C5), 89.69 (C1'), 83.38 (C4'), 65.22 (C5'), 33.83 (C2'), 26.51 (C3'); **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_9H_{14}N_3O_3$, 212.1035; found 212.1028.

2',3'-Dideoxy-β-D-cytidine (β-ddC) has been previously synthesized [302].



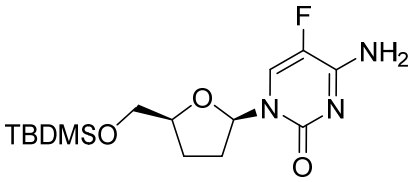
Obtained as a white solid (9 mg, quantitative) from **5'-*O*-silyl-protected β-ddC** (14 mg, 0.043 mmol). TLC dichloromethane/methanol (9:1). **mp:** 223-225 °C (lit. 215-217 °C [486]); $[\alpha]_D^{27}$ +82.1 (*c* 0.77, MeOH) (lit. $[\alpha]_D^{20}$ +102.5 (*c* 0.5, MeOH) [486]); **^{1}H NMR** (500 MHz, CD_3OD) δ 8.12 (d, $J_{6,5}$ = 7.5 Hz, 1H, H6), 6.02 (dd, $J_{1',2'a}$ = 6.7, $J_{1',2'b}$ = 3.2 Hz, 1H, H1'), 5.87 (d, 1H, H5), 4.18 – 4.11 (m, 1H, H4'), 3.87 (dd, $J_{5'a,5'b}$ = 12.2, $J_{5'a,4'}$ = 3.1 Hz, 1H, H5'a), 3.69 (dd, $J_{5'b,4'}$ = 4.0 Hz, 1H, H5'b), 2.46 – 2.37 (m, 1H, H2'a), 2.05 – 1.83 (m, 3H, H2'b, H3'a and H3'b) [499]; **^{13}C NMR** (125 MHz, CD_3OD) δ 167.6 (C4), 158.3 (C2), 142.8 (C6), 95.2 (C5), 88.4 (C1'), 83.6 (C4'), 63.7 (C5'), 34.1 (C2'), 25.7 (C3'); **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_9H_{14}N_3O_3$, 212.1035; found 212.1029.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-5-fluorocytosine (5'-*O*-silyl-protected α-ddFC). TLC dichloromethane/methanol

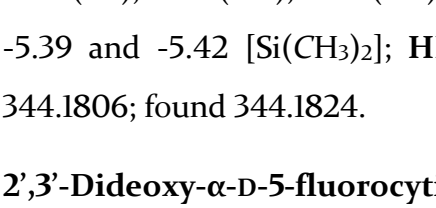


(19:1); preparative TLC acetone/dichloromethane (1:1). Colourless syrup (37 mg, 29%). $[\alpha]_D^{29}$ -59.0 (*c* 1, MeOH); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.05 (s, 1H, NHa), 7.45 (d, $J_{6,F}$ = 6.2 Hz, 1H), 6.02 – 5.94 (m, 1H, H1'), 5.54 (s, 1H, NHb), 4.45 – 4.35 (m, 1H, H4'), 3.69 – 3.60 (m, 2H, H5a' and H5'b), 2.59 – 2.48 (m, 1H, H2'a), 2.07 – 1.97 (m, 1H, H2'b), 1.97 – 1.87 (m, 2H, H3'a and H3'b), 0.89 [s, 9H, $SiC(CH_3)_3$], 0.06 [s, 6H, $Si(CH_3)_2$] [495]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 158.1 (d, $J_{4,F}$ = 13.5 Hz, C4), 154.0 (C2), 136.5 (d, $J_{5,F}$ = 240.9 Hz, C5), 124.7 (d, $J_{6,F}$ = 31.8 Hz, C6), 88.8 (C1'), 82.2 (C4'), 65.4 (C5'), 33.0 (C2'), 26.0 [$SiC(CH_3)_3$], 25.5 (C3'), 18.4 (SiC), -5.21 and -5.25 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_{15}H_{27}N_3O_3SiF$, 344.1806; found. 344.1823.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-5-fluorocytosine (5'-*O*-silyl-protected β-ddFC). TLC dichloromethane/methanol (19:1); preparative TLC acetone/dichloromethane (1:1). White solid (40 mg, 32%). **mp:** 180-182 °C [495]; $[\alpha]_D^{29}$ +51.6 (*c* 0.75, MeOH); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.28 (d, $J_{6,F}$ = 6.6 Hz, 1H, H6), 7.51 (s, 1H, NHa), 6.05 – 5.99 (m, 1H, H1'), 5.41 (s, 1H, NHb), 4.17 – 4.13 (m, 1H, H4'), 4.11 (dd, $J_{5'a,5'b}$ = 11.5, $J_{5'a,4'}$ = 2.3 Hz, 1H, H5'a), 3.72 (dd, $J_{5'b,4'}$ = 2.2 Hz, 1H, H5'b), 2.48 – 2.38 (m, 1H, H2'a), 2.15 – 2.06 (m, 1H, H2'b), 2.01 – 1.89 (m, 1H, H3'a), 1.88 – 1.78 (m, 1H, H3'b), 0.93 [s, 9H, $SiC(CH_3)_3$], 0.12 [2s, 6H, $Si(CH_3)_2$] [495]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 157.9 (d, $J_{4,F}$ = 13.7 Hz, C4), 154.1 (C2), 136.3 (d, $J_{5,F}$ = 239.0 Hz, C5), 126.3 (d, $J_{6,F}$ = 32.2 Hz, C6), 87.2 (C1'), 82.5 (C4'), 63.6 (C5'), 33.8 (C2'), 26.1 [$SiC(CH_3)_3$], 24.0 (C3'), 18.7 (Si*C*), -5.39 and -5.42 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_{15}H_{27}N_3O_3SiF$, 344.1806; found 344.1824.



2',3'-Dideoxy-α-D-5-fluorocytidine (α-ddFC). Obtained as a colourless syrup (12 mg, quantitative) from **5'-*O*-silyl-protected α-ddFC** (18 mg, 0.052 mmol). TLC dichloromethane/methanol (9:1). $[\alpha]_D^{28}$ -86.1 (*c* 0.62, MeOH); **^{1}H NMR** (500 MHz, CD_3OD) δ 7.79 (d, $J_{6,F}$ = 6.6 Hz, 1H, H6), 6.02 – 5.97 (m, 1H, H1'), 4.52 – 4.45 (m, 1H, H4'), 3.64 (dd, $J_{5'a,5'b}$ = 11.8, $J_{5'a,4'}$ = 3.9 Hz, 1H, H5'a), 3.53 (dd, $J_{5'b,4'}$ = 5.3 Hz, 1H, H5'b), 2.55 – 2.46 (m, 1H, H2'a), 2.08 – 1.96 (m, 2H, H2'b and H3'a), 1.92 – 1.83 (m, 1H, H3'b); **^{13}C NMR** (125 MHz, CD_3OD) δ 159.6 (d, $J_{4,F}$ = 13.9 Hz, C4), 156.5 (C2), 138.4 (d, $J_{5,F}$ = 242.0 Hz, C5), 126.2 (d, $J_{6,F}$ = 32.4 Hz, C6), 89.8 (C1'), 83.4 (C4'), 65.2 (C5'), 33.8 (C2'), 26.5 (C3'); **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_9H_{13}N_3O_3F$, 230.0941; found 230.0925.



2',3'-Dideoxy-β-D-5-fluorocytidine (β-ddFC) has been previously synthesized [500]. Obtained as a white foam (8 mg, quantitative) from **5'-*O*-silyl-protected β-ddFC** (12 mg, 0.035 mmol). TLC dichloromethane/methanol (9:1). **mp:** 157-159 °C (lit. 147-149 °C [495]); $[\alpha]_D^{28}$ +96.4 (*c* 0.5, MeOH) (lit. for its enantiomer $[\alpha]_D^{25}$ -108 (*c* 0.13, MeOH) [495]); **^{1}H NMR** (500 MHz, CD_3OD) δ 8.41 (d, $J_{6,F}$ = 7.0

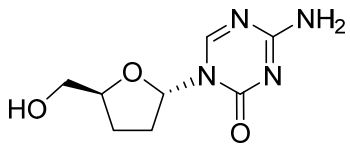
Hz, 1H, H6), 5.99 – 5.94 (m, 1H, H1'), 4.18 – 4.12 (m, 1H, H4'), 3.93 (dd, $J_{5'a,5'b}$ = 12.2, $J_{5'a,4'}$ = 2.8 Hz, 1H, H5'a), 3.70 (dd, $J_{5'b,4'}$ = 3.3 Hz, 1H, H5'b), 2.47 – 2.36 (m, 1H, H2'a), 2.09 – 2.01 (m, 1H, H2'b), 1.96 – 1.88 (m, 2H, H3'a and H3'b); **^{13}C NMR** (125 MHz, CD_3OD) δ 159.5 (d, $J_{4,F}$ = 14.1 Hz, C4), 156.5 (C2), 138.2 (d, $J_{5,F}$ = 240.7 Hz, C5), 127.2 (d, $J_{6,F}$ = 33.3 Hz, C6), 88.6 (C1'), 83.9 (C4'), 63.2 (C5'), 34.3 (C2'), 25.1 (C3'); **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_9H_{13}N_3O_3F$, 230.0941; found 230.0923.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-α-D-ribofuranosyl]-5-azacytosine (5'-*O*-silyl-protected α-ddAC) has been previously synthesized [501]. TLC dichloromethane/methanol (19:1); preparative TLC dichloromethane/methanol (97:3). White solid (26 mg, 22%). **mp** = 197-199 °C (lit. 191-193 °C [501]); $[\alpha]_D^{28}$ -50.7 (*c* 0.6, CH_3Cl); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.10 (s, 1H, H6), 6.85 (s, 1H, NHa), 6.01 (dd, $J_{1',2'a}$ = 6.2, $J_{1',2'b}$ = 2.9 Hz, 1H, H1'), 5.74 (s, 1H, NHb), , 4.48 – 4.39 (m, 1H, H4'), 3.69 (dd, $J_{5'a,5'b}$ = 10.9, $J_{5'a,4'}$ = 4.3 Hz, 1H, H5'a), 3.64 (dd, $J_{5'b,4'}$ = 4.2 Hz, 1H, H5'b), 2.63 – 2.52 (m, 1H, H2'a), 2.14 – 2.04 (m, 1H, H2'b), 2.00 – 1.92 (m, 2H, H3'a and H3'b), 0.90 [s, 9H, $SiC(CH_3)_3$], 0.07 [s, 6H, $Si(CH_3)_2$] [501]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 166.6 (C4), 154.8 (C6), 154.0 (C2), 89.1 (C1'), 82.5 (C4'), 65.3 (C5'), 33.1 (C2'), 26.0 [$SiC(CH_3)_3$], 25.3 (C3'), 18.4 (Si*C*), -5.20 and -5.26 [$Si(CH_3)_2$]; **HRMS** (*m*/*z*): $[M + H]^+$ calcd. for $C_{14}H_{27}N_4O_3Si$, 327.1852; found 327.1841.

1-[5'-*O*-(*tert*-Butyldimethylsilyl)-2',3'-dideoxy-β-D-ribofuranosyl]-5-azacytosine (5'-*O*-silyl-protected β-ddAC) has been previously synthesized [501]. TLC dichloromethane/methanol (19:1); preparative TLC dichloromethane/methanol (97:3). White solid (26 mg, 22%). **mp** = 183-185 °C (lit. 196-198 °C [501]); $[\alpha]_D^{28}$ +53.9 (*c* 0.25, CH_3Cl); **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.72 (s, 1H, H6), 6.64 (s, 1H, NHa), 6.02 (dd, $J_{1',2'a}$ = 6.6, $J_{1',2'b}$ = 2.4 Hz, 1H, H1'), 5.65 (s, 1H, NHb), 4.23 – 4.16 (m, 1H, H4'), 4.08 (dd, $J_{5'a,5'b}$ = 11.6, $J_{5'a,4'}$ = 2.5 Hz, 1H, H5'a), 3.72 (dd, $J_{5'a,4'}$ = 2.4 Hz, 1H, H5'b), 2.54 – 2.42 (m, 1H, H2'a), 2.20 – 2.13 (m, 1H, H2'b), 2.01 – 1.92 (m, 1H, H3'a), 1.91 – 1.84 (m, 1H, H3'b), 0.92 [s, 9H, $SiC(CH_3)_3$], 0.11 [2s, 6H, $Si(CH_3)_2$] [501]; **^{13}C NMR** (125 MHz, $CDCl_3$) δ 166.6 (C4), 156.2 (C6),

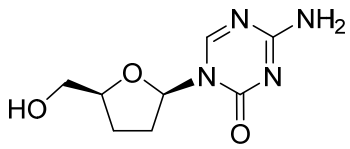
154.3 (C2), 87.5 (C1'), 82.9 (C4'), 63.4 (C5'), 33.9 (C2'), 26.1 [SiC(CH_3)$_3$], 23.9 (C3'), 18.6 (Si*C*), -5.28 and -5.39 [Si(CH_3)$_2$]; **HRMS** (*m*/*z*): [M + H]$^+$ calcd. for $C_{14}H_{27}N_4O_3Si$, 327.1852; found 327.1862.

2',3'-Dideoxy-α-D-5-azacytidine (α-ddAC) has been previously synthesized



[501]. Obtained as a white solid (8 mg, quantitative) from **5'-*O*-silyl-protected α-ddAC** (12 mg, 0.037 mmol). TLC dichloromethane/methanol (9:1). **mp:** 145-147 °C (lit. 132-134 °C [501]); $[\alpha]_D^{29}$ -52.8 (*c* 0.2, MeOH) (lit. for its enantioner $[\alpha]_D^{25}$ -59.7 (*c* 0.07, MeOH) [501]); **^{1}H NMR** (500 MHz, CD_3OD) δ 8.26 (s, 1H, H6), 5.99 (dd, $J_{1',2'a}$ = 6.3, $J_{1',2'b}$ = 3.5 Hz, 1H, H1'), 4.53 – 4.48 (m, 1H, H4'), 3.65 (dd, $J_{5'a,5'b}$ = 11.9, $J_{5'a,4'}$ = 3.9 Hz, 1H, H5'a), 3.54 (dd, $J_{5'b,4'}$ = 5.2 Hz, 1H, H5'b), 2.57 – 2.49 (m, 1H, H2'a), 2.16 – 2.03 (m, 2H, H2'b and H3'a), 1.95 – 1.88 (m, 1H, H3'b); **^{13}C NMR** (125 MHz, CD_3OD) δ 168.2 (C4), 156.62 (C6), 156.59 (C2), 90.1 (C1'), 83.7 (C4'), 65.1 (C5'), 33.7 (C2'), 26.4 (C3'); **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_8H_{12}N_4O_3Na$, 235.0807; found 235.0811.

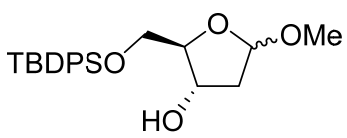
2',3'-Dideoxy-β-D-5-azacytidine (β-ddAC) has been previously synthesized



[501]. Obtained as a white solid (8 mg, quantitative) from **5'-*O*-silyl-protected β-ddAC** (12 mg, 0.037 mmol). TLC dichloromethane/methanol (9:1). **mp:** 240 °C (245 °C [501]); $[\alpha]_D^{29}$ +46.3 (*c* 0.1, MeOH) (lit. for its enantioner $[\alpha]_D^{25}$ -45 (*c* 0.1, MeOH) [501]); **^{1}H NMR** (300 MHz, D_2O) δ 8.63 (s, 1H, H6), 6.02 (dd, $J_{1',2'a}$ = 6.9, $J_{1',2'b}$ = 2.3 Hz, 1H, H1'), 4.36 – 4.25 (m, 1H, H4'), 3.93 (dd, $J_{5'a,5'b}$ = 12.7, $J_{5'a,4'}$ = 3.0 Hz, 1H, H5'a), 3.76 (dd, $J_{5'b,4'}$ = 4.9, 1H, H5'b), 2.59 – 2.43 (m, 1H, H2'a), 2.28 – 2.17 (m, 1H, H2'b), 2.11 – 1.99 (m, 1H, H3'a), 1.93 – 1.75 (m, 1H, H3'b); **^{13}C NMR** (150 MHz, CD_3OD) δ 168.0 (C4), 157.4 (C6), 156.6 (C2), 88.8 (C1'), 84.4 (C4'), 63.1 (C5'), 34.3 (C2'), 25.0 (C3'); **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_8H_{12}N_4O_3Na$, 235.0807; found 235.0809.

Synthesis of 3'-azido-2',3'-dideoxynucleosides

Methyl 5-*O*-*tert*-butyldiphenylsilyl-2-deoxy-α,β-D-erythro-pentofuranosides (58α, 58β). 2-deoxy-D-ribose (5 g, 37.2 mmol) was dissolved in MeOH (70 mL) and AcCl was then added (173 µL 2.4 mmol) and the reaction mixture was stirred for 1 h. $NaHCO_3$ (2 g) was added and stirred for an additional 5 min. The precipitate was filtered out and the solvent removed under vacuum. The resulting brown syrup was used without further purification. To a solution of the crude product (5.8 g) in DCM (80 mL), DMAP (956 mg, 7 mmol), TEA (6 mL) and TBDPSCl (11.16 mL, 43 mmol) were subsequently added in an ice bath. This mixture was allowed to stir overnight and partitioned between water and DCM. The aqueous layer was extracted twice with DCM and the combined organic layers were dried over Mg_2SO_4 and evaporated. The residue was purified by FCC using hexane-diethyl ether (9:1) to afford **58α** (5.7 g) and **58β** (3.0 g) as colourless syrups in a combined yield of 60%.



Methyl 5-*O*-*tert*-butyldiphenylsilyl-2-deoxy-α-D-erythro-pentofuranoside (58α). $[\alpha]_D^{28}$ +53.0° (*c* 1, CH_3Cl). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.70 – 7.63 (m, 4H, aryl), 7.46 – 7.34 (m, 6H, aryl), 5.11 (d, *J* = 4.5 Hz, 1H, H1), 4.30 (dd, *J* = 10.5, 6.1 Hz, 1H, H4), 4.19 – 4.14 (m, 1H, H3), 3.75 (dd, *J* = 11.0, 3.6 Hz, 1H, H5a), 3.61 (dd, *J* = 11.1, 4.9 Hz, 1H, H5b), 3.38 (s, 3H, OMe), 2.83 (d, *J* = 10.6 Hz, 1H, OH), 2.23 – 2.15 (m, 1H, H2), 2.02 (d, *J* = 13.7 Hz, 1H, H2), 1.05 [s, 9H, $SiC(CH_3)_3$]. **^{13}C NMR** (125 MHz, $CDCl_3$) δ 135.8, 135.7, 133.4, 133.3, 129.9, 129.9, 127.9 (aryl), 105.8 (C1), 88.0 (C4), 73.4 (C3), 64.6 (C5), 54.9 (OMe), 41.3 (C2), 27.0 [$SiC(CH_3)_3$], 19.4 (Si*C*) [434]. **HRMS** (*m*/*z*): $[M + Na]^+$ calcd. for $C_{22}H_{30}O_4NaSi$, 409.1811; found, 409.1793.

Methyl 5-*O*-*tert*-butyldiphenylsilyl-2-deoxy-β-D-erythro-pentofuranoside (58β). $[\alpha]_D^{28}$ -41.8° (*c* 1, CH_3Cl). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.71 – 7.65 (m, 4H, aryl), 7.48 – 7.35 (m, 6H, aryl), 5.05 (dd, *J* = 5.3, 2.1 Hz, 1H, H1), 4.55 – 4.47 (m, 1H, H4), 3.98 – 3.91 (m, 1H, H3), 3.82 (dd, *J* = 10.3, 5.1 Hz, 1H, H5a), 3.66 (dd, *J* = 10.2, 7.7 Hz, 1H, H5b), 3.27 (s, 3H, OMe), 2.24 – 2.16 (m, 1H, H2), 2.11 – 2.02 (m, 1H, H2), 1.81 (d, *J* = 4.1 Hz, 1H, OH), 1.07 [s, 9H, $SiC(CH_3)_3$]. **^{13}C NMR** (125 MHz, $CDCl_3$) δ 135.7, 133.4, 129.97, 129.95, 127.9 (aryl), 105.1 (C1), 85.8 (C4), 73.5 (C3), 65.6 (C5),

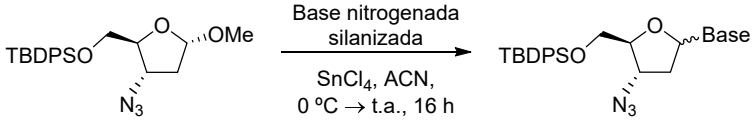
55.2 (OMe), 41.2 (C2), 27.0 [SiC(CH_3)$_3$], 19.4 (Si*C*) [434]. **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{22}H_{30}O_4NaSi$, 409.1811; found, 409.1816.

Methyl 5-*O*-*tert*-butyldiphenylsilyl-3-iodo-2,3-dideoxy-α-D-threo-pentofuranoside (59). Compound **58α** (5.6 g, 14.5 mmol) was dissolved in toluene (150 mL) and Ph_3P (7.9 g, 30.4 mmol), imidazole (4.4 g, 60.84 mmol) and iodine (7.7 g, 30.4 mmol) was then added. The mixture was stirred at 75 °C until the completion of the reaction. The suspension was filtered and the solvent evaporated. FCC using hexane-diethyl ether (5:1) yielded pure **59** (6.6 g, 92%) as a pale yellow syrup. $[\alpha]_D^{28}$ +41.9° (*c* 1, CH_3Cl). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.79 – 7.32 (m, 10H, aryl), 5.21 (dd, $J_{1,2a}$ = 5.5, $J_{1,2b}$ = 3.9 Hz, 1H, H1), 4.57 – 4.52 (m, 1H, H4), 3.96 (dd, $J_{5a,5b}$ = 10.5, $J_{5a,4}$ = 5.0 Hz, 1H, H5a), 3.72 (dd, $J_{5b,4}$ = 6.0 Hz, 1H, H5b), 3.44 (q, *J* = 5.1 Hz, 1H, H3), 3.39 (s, 3H, OCH_3), 2.81 (ddd, *J* = 14.8, 5.5, 3.3 Hz, 1H, H2a), 2.59 (ddd, *J* = 14.8, 6.5, 3.9 Hz, 1H, H2b), 1.07 [s, 9H, SiC(CH_3)$_3$]. **^{13}C NMR** (100 MHz, $CDCl_3$) δ 135.86, 135.77, 133.51, 133.45, 129.87 and 127.83 (aryl), 105.1 (C1), 80.1 (C4), 68.8 (C5), 55.8 (OCH_3), 46.3 (C2), 27.2 (C3), 27.0 [SiC(CH_3)$_3$], 19.4 (Si*C*) [434]. **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{22}H_{29}O_3NaSiI$, 519.0828; found, 519.0848.

Methyl 5-*O*-*tert*-butyldiphenylsilyl-3-azido-2,3-dideoxy-α-D-erythro-pentofuranoside (60). NaN_3 (3.0 g, 46.5 mmol) was suspended in a solution of **59** (6.6 g, 13.3 mmol) in DMF (60 mL). After stirring overnight at 80 °C, the solvent was evaporated in vacuum and the residue purified by FCC using hexane-diethyl ether (5:1) to afford pure **60** (3.0 g, 54%) as a pale yellow syrup. $[\alpha]_D^{27}$ +89.4° (*c* 1, CH_3Cl) (lit. $[\alpha]_D^{20}$ +84° (*c* 0.68, CH_3Cl) [502]). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.75 – 7.30 (m, 10H, aryl), 5.08 (d, *J* = 5.1 Hz, 1H, H1), 4.13 – 4.00 (m, 2H, H3 and H4), 3.82 – 3.68 (m, 2H, H5a and H5b), 3.38 (s, 3H, OCH_3), 2.41 – 2.29 (m, 1H, H2a), 2.03 (ddd, *J* = 14.0, 2.6, 1.2 Hz, 1H, H2b), 1.07 [s, 9H, SiC(CH_3)$_3$]. **^{13}C NMR** (100 MHz, $CDCl_3$) δ 135.79, 135.72, 133.3, 133.2, 130.0, 129.9, 127.94, 127.89 (aryl), 104.9 (C1), 83.5 (C4), 64.1 (C5), 60.9 (C3), 55.1 (OCH_3), 38.9 (C2), 27.0

[SiC(CH_3)$_3$], 19.4 (Si*C*) [502]. **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{22}H_{29}N_3O_3NaSi$, 434.1876; found, 434.1861.

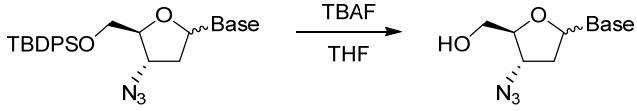
General procedure for sugar-base coupling to 60



60 **Nucleósidos protegidos**

Coupling reactions between sugar moiety **60** and nitrogenous base were carried out as previously described [439], [440]. Nitrogenous base (1.94 mmol) was dissolved, under Ar atmosphere, in anhydrous ACN (3 mL) and BSA (1.1 mL, 4.28 mmol) was added with a syringe. The mixture was refluxed for 1 h, after which the reaction was cooled to room temperature and a solution of **60** (100 mg, 0.24 mmol) in anhydrous ACN (2 mL) was then added. The reaction was cooled with an ice-bath, treated with $SnCl_4$ (1 M in DCM, 2.4 mL, 2.43 mmol) and allowed to stir overnight at room temperature. This mixture was diluted with DCM and washed with $NaHCO_3$ saturated aqueous solution, the organic phase was dried and evaporated. Nucleoside stereoisomers were purified by FCC and collected together since their separation was not possible.

General procedure for the desilylation of protected nucleosides



Nucleósidos protegidos **Nucleósidos**

Nucleoside stereoisomer mixtures were deprotected as previously described for 2',3'-dideoxynucleosides. The deprotected 3'-azido-2',3'-dideoxynucleoside estereoisomers mixtures were purified by preparative TLC using the following solvent systems:

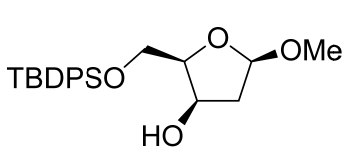
-For uracil and thymine: ethyl acetate-hexane (60:40).

-For azauracil: ethyl acetate-hexane (40:60).

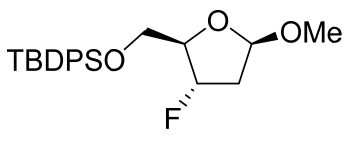
-For 5-fluorouracil, 5-chlorouracil and 5-bromouracil: ethyl acetate-hexane (30:70).

-For cytosine, 5-fluorocytosine and 5-azacytosine: dichloromethane-methanol (95:5).

Synthesis of 3'-fluoro-2',3'-dideoxynucleosides

Methyl 5-*O*-*tert*-butyldiphenylsilyl-2-deoxy-β-D-threo-pentofuranoside (62). 

Compound **58β** (2.4 g, 6.21 mmol) was dissolved in DCM (15 mL) and was then treated with a freshly prepared complex of CrO_3/Py/Ac_2O (1.9 g/3.1 mL/1.9 mL, in this order) in DCM (35 mL). After stirring for 20 min the reaction was finished and the mixture was slowly added to stirred ethyl acetate. The resulting suspension was passed through a path of silica gel washing with ethyl acetate and the solvent evaporated under reduced pressure below 40 °C. The residue was dissolved in EtOH (100 mL), cooled with an ice-bath previous addition of $NaBH_4$ (960 mg, 24.83 mmol) and stirred for 10-15 min. The reaction was then quenched by the dropwise addition of 80% acetic acid the solvent evaporated. Purification by FCC with hexane-diethyl ether (3:1) allowed obtaining **62** (1.7 g, 71%) as a white solid. $[\alpha]_D^{29}$ -54.6° (*c* 1, CH_3Cl) (lit. $[\alpha]_D^{20}$ -54° (*c* 0.31, CH_3Cl) [502]). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.78 – 7.30 (m, 10H, aryl), 5.05 (dd, $J_{1,2a}$ = 3.7, $J_{1,2b}$ = 1.5 Hz, 1H, H1), 4.39 – 4.30 (m, 1H, H4), 4.15 – 4.02 (m, 2H, H3 and H5), 3.87 (dd, $J_{5a,5b}$ = 10.3, $J_{5a,4}$ = 5.8 Hz, 1H, H5), 3.30 (s, 3H, OCH_3), 2.94 (d, *J* = 9.7 Hz, 1H, OH), 2.18 – 2.07 (m, 2H, H2a and H2b), 1.07 [s, 9H, $SiC(CH_3)_3$]. **^{13}C NMR** (100 MHz, $CDCl_3$) δ 135.8, 135.8, 133.6, 133.5, 129.8, 127.8 and 127.8 (aryl), 105.2 (C1), 84.7 (C3), 71.6 (C4), 63.8 (C5), 55.1 (OCH_3), 41.5 (C2), 27.0 [$SiC(CH_3)_3$], 19.3 (Si*C*) [502]. **HRMS** (*m*/*z*): [M + Na]$^+$ calcd. for $C_{22}H_{30}O_4NaSi$, 409.1811; found, 409.1805.

Methyl 5-*O*-*tert*-butyldiphenylsilyl-3-fluoro-2,3-dideoxy-β-D-erythro-pentofuranoside (63). 

To a solution of **62** (1.7 g, 4.4 mmol) in anhydrous toluene (20 mL) was added DAST (2.7 mL, 17.59 mmol). The reaction was stirred for 2 h, after which the mixture was slowly and extremely carefully passed through a path of silica gel washing with $CHCl_3$. After evaporation of the solvent, purification by FCC using hexane-diethyl ether (95:5) afforded **63** (425 mg, 25%) as a colourless syrup. $[\alpha]_D^{28}$ -42.8° (*c* 0.4, CH_3Cl) (lit. $[\alpha]_D^{20}$ -45° (*c* 0.25, CH_3Cl) [502]).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.70 – 7.35 (m, 10H, aryl), 5.34 – 5.31 (m, 1 ½H, 1 ½H3), 5.23 – 5.20 (m, 1H, 1 ½H, 1 ½H3), 5.17 (dd, *J* = 5.6, 3.8 Hz, 1H, H1), 4.33 – 4.23 (m, 1H, H4), 3.73 (ddd, $J_{5a,5b}$ = 10.6, 5.1, 2.1 Hz, 1H, H5a), 3.59 (dd, *J* = 8.1 Hz, 1H, H5b), 3.30 (s, 3H, OCH_3), 2.44 – 2.32 (m, 1H, H2a), 2.23 – 2.10 (m, 1H, H2b), 1.07 [s, 9H, $SiC(CH_3)_3$]. **^{13}C NMR** (125 MHz, $CDCl_3$) δ 135.7, 133.3, 129.97, 129.93, 127.9, 127.9, 105.8 (d, $J_{F,C1}$ = 1.9 Hz, C1), 94.75 (d, $J_{F,C3}$ = 177.8 Hz, C3), 84.76 (d, $J_{F,C4}$ = 22.9 Hz, C4), 63.91 (d, $J_{F,C5}$ = 9.9 Hz, C5), 55.71 (OCH_3), 39.57 (d, $J_{F,C2}$ = 21.5 Hz, C2), 26.94 [$SiC(CH_3)_3$], 19.36 (Si*C*). **HRMS** (*m*/*z*): $[M + Na]^+$ calcd. for $C_{22}H_{29}O_3FNaSi$, 411.1768; found, 411.1757.

5.2 MOLECULAR BIOLOGY

5.2.1 Chemicals Used in this Study

RTis used to treat HIV/AIDS (Abacavir, AZT, efavirenz, emtricitabine, etravirine, lamivudine, nevirapine, stavudine, and tenofovir) were obtained from Hospital Clinico San Cecilio, Granada, Spain. The name, provider and CAS number of commercially available nucleoside analogues used in this study are indicated in **Table 5**. All synthesized compounds were dissolved in water. Aliquots of all chemicals were prepared to avoid repeated freeze-thaw cycles.

Name	CAS	Provider	Solvent
Abacavir	136470-78-5	Hospital	Water
AZT	30516-87-1	Hospital	Water
Efavirenz	154598-52-4	Hospital	DMSO
Emtricitabine	143491-57-0	Hospital	Water
Etravirine	269055-15-4	Hospital	DMSO
Lamivudine	134678-17-4	Hospital	Water
Nevirapine	129618-40-2	Hospital	DMSO
Stavudine	3056-17-5	Hospital	Water
Tenofovir disoproxil	147127-20-6	Hospital	Water
2′,3′-Dideoxy-3′-fluorouridine (C1)	41107-56-6	Aldrich	Water
2',3'-Isopropylideneuridine (C2)	362-43-6	Aldrich	Water
2',3′-Dideoxy-3′-fluorothymidine (C4)	25526-93-6	Aldrich	Water
2′,3′-Dideoxy-5-iodouridine (C5)	105784-83-6	Aldrich	Water

2',3'-*O*-isopropylideneadenosine (C6)	362-75-4	Aldrich	Water
2'-Azido-2'-deoxyuridine (C7)	26929-65-7	Aldrich	Water
N4-Acetyl-2'-*O*-methylcytidine (C8)	113886-71-8	Carbosynth	Water
N4-Acetyl-2'-deoxy-2'-fluorocytidine (C9)	159414-97-8	Carbosynth	Water
5'-*O*-Acetyl-2',3'-*O*-isopropylideneadenosine (C10)	15888-38-7	Carbosynth	DMSO

Table 5. *HIV/AIDS RTis and commercially available nucleoside analogues used in this study.*

5.2.2 Cell Line Authentication

The cell lines used in this study were originally obtained from ATCC. Their identity was confirmed by STR analyses at least once a year at Lorgen, Granada, Spain. Furthermore, the absence of *Mycoplasma spp*. was confirmed every month using a PCR-based assay (Minerva).

5.2.3 Cell Line Conditions

All cell lines used in this study were grown at 37 °C, 5% CO_2 and atmospheric O_2, and passaged using Trypsin 0.05% (from Gibco, used for HeLa and PA-1 cells) or by pipetting up/down (for HEK293T cells).

-HeLa cell [116] were grown on Dulbecco's Modified Eagle Medium-High Glucose (DMEM, Invitrogen) supplemented with 1x penicillin/streptomycin/glutamine (Invitrogen) and 10% Fetal Bovine Serum (FBS, Hyclone).

-HEK293T cells were grown on DMEM-High Glucose (Invitrogen) supplemented with 1x penicillin/streptomycin/glutamine (Invitrogen) and 10% FBS (Hyclone).

-PA-1 cells [503] were cultured as described [469], using Minimum Essential Media (MEM, Invitrogen) supplemented with 1x penicillin/streptomycin/glutamine (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen), and 10% heat-inactivated FBS (Hyclone).

5.2.4 Expression Plasmids

Plasmids used in this study were propagated in Escherichia coli strain DH5α (ThermoFisher) and were purified using Plasmid Midi kits from Qiagen following manufacturer instructions. Plasmid DNAs were analysed by electrophoresis (0.7%

agarose-ethidium bromide gels) and we only used highly supercoiled preparations of plasmid DNA for transfections.

-JM101/L1.3 (Human L1.3) has been described previously [504]. It contains a full-length copy of the human L1.3 element (L1.3, accession number #L19088) tagged with the mneoI indicator cassette [116], [285]; it is cloned in vector pCEP4 (Life Technologies).

-pCEPL1SM (Mouse L1-T_F) has been described previously [473]. It contains a full-length mouse T_F LINE-1 element (L1Md-$L1_{Orf}$, [466]) where the coding sequence of the LINE-1 ORFs (L1-ORF1 and L1-ORF2) has been codon optimized, and is tagged with the *mneoI* indicator cassette [285]. It is cloned in vector pCEP4 (Life Technologies).

-pCEP-TGF21 (Mouse L1-G_F) has been described previously [471]. It contains a full-length mouse G_F LINE-1 element (L1Md-G_F21, accession number #AC021631.6, positions 62229-68991) tagged with the *mneoI* indicator cassette [285]. It is cloned in vector pCEP4 (Life Technologies).

-pCEP-A101 (Mouse L1-A) has been described previously [471]. It contains a full-length mouse A LINE-1 element (L1Md-A101, accession number #AY053455) tagged with the *mneoI* indicator cassette [285]. It is cloned in vector pCEP4 (Life Technologies).

-pCMV-MusD-6neo^{TNF} (Mouse MusD) has been described previously [454]. It contains a nearly full-length copy of a mouse MusD element (lacks the U3 sequence from the 5′LTR, accession number #AC124426, positions 9078–16,569 (+)) tagged with the neo^{TNF} indicator cassette [505] and is cloned in vector pCMVbeta (Clontech).

-pIAP-92L23neo^{TNF} (Mouse IAP) has been described previously [455]. It contains a full-length copy of a mouse IAP element (accession number #AC012382, positions 161,601–168,684, (+)) tagged with the neo^{TNF} indicator cassette [505] at IAP nucleotide position 5744, and is cloned in vector pGL3basic (Promega).

-pU6ineo has been described previously [191]. It contains the neomycin phosphotransferase (NEO) expression cassette from pEGFP-N1 (Clontech) cloned into a modified pBSKS-II(+) (Stratagene) that contains a U6 promoter in the multi-cloning site.

-JJ101/L1.3 (Human L1.3) has been described previously [465]. It contains a full-length copy of the human L1.3 element (L1.3, accession number #L19088) tagged with the *mblastI* indicator cassette [151], [287] and is cloned in pCEP4 (Life Technologies).

-JJL1SM (Mouse L1-T_F) has been described previously [467]. It contains a full-length mouse T_F LINE-1 element (L1Md-$L1_{Orf}$, [466]) where the coding sequence of the LINE-1 ORFs (L1-ORF1 and L1-ORF2) has been codon optimized, and is tagged with the *mblastI* indicator cassette [151], [287]. It is cloned in vector pCEP4 (Life Technologies).

-pXY014 (Human $L1_{RP}$) has been described previously [286]. It contains a full-length copy of the human $L1_{RP}$ element (accession number #AF148856.1, [468]) tagged with the *mflucI* indicator cassette [286] and is cloned in a modified pCEP4 (Life Technologies) that contains a Renilla firefly expression cassette.

-pXY015 has been described previously [286]. It is derived from plasmid pXY014 but the cloned $L1_{RP}$ element contains two missense mutations in the RNA binding domain of L1-ORF1p (RR261/62AA). This plasmid was used as a negative control of the luciferase-based retrotransposition assays.

5.2.5 Retrotransposition Assays

All retrotransposition and clonability assays conducted in this study were conducted at least in duplicate and several independent times (>3). Clonability and retrotransposition assays using *mneoI* or neo^{TNF} tagged retrotransposons on HeLa cells were carried out as previously described [150], [177], [506]. HeLa cells were plated in 6-well plates (Corning) at the indicated number ($2x10^4$ cells for plasmids JM101/L1.3, pCEP-TG_F21, pCEP-A101, and pCMV-MusD-$6neo^{TNF}$; $1x10^4$ cells for plasmids pCEPL1SM, pIAP-92L23neo^{TNF}, and pU6ineo). Eighteen hours after

plating, DNA transfections were carried out using FuGene 6 transfection reagent (Promega) and Opti-MEM (Life Technologies) following the protocol provided by the manufacturer (for a well of a 6-well plate: 3 µL of FuGene, 97 µL of Opti-MEM and 1 µg of plasmid DNA). The day after transfection, media was replaced with fresh media containing the indicated amount of each RTi (0, 5, and 25 µM unless otherwise indicated). Neomycin selection was started 72 h post-transfection using 400 µg/mL G418 (Life Technologies) and the indicated amount of each RTi. Selection media containing the indicated RTi was replaced every other day and selection was continued for 11 additional days. HeLa cells were then washed with 1x PBS (Gibco), fixed with 2% paraformaldehyde/0.4% glutaraldehyde, and stained with 0.1% (w/v) crystal violet solution as described [116], [506], to visualize and count foci representing successful retrotransposition events.

Retrotransposition assays using *mblastI* tagged LINE-1 vectors (JJ101/L1.3 and JJL1SM) on HeLa cells were carried out as previously described [150], [177], [507]. $2x10^4$ (JJ101/L1.3) or $1x10^4$ (JJL1SM) HeLa cells were plated per well of a 6-well tissue culture plate and transfected 18h later using 1 µg of each plasmid and 3 µL of FuGene 6 as described above. The day after transfection, media was replaced with fresh media containing the indicated amount of each RTi (0, 5, and 25 µM unless otherwise indicated), and cells cultured for 5 additional days changing the media every other day. Blasticidin S selection (5 µg/mL, Life Technologies) in the presence of the indicated RTi was started 120 h post-transfection, and selection was continued for 9 additional days. HeLa cells were then washed with 1x PBS (Gibco), fixed with 2% paraformaldehyde/0.4% glutaraldehyde, and stained with 0.1% (w/v) crystal violet solution as described [116], [506], to visualize and count foci representing successful retrotransposition events.

Retrotransposition assays using *mflucI* tagged LINE-1 constructs (pX014 and pX015 as a negative control) were carried out in HEK293T cells as previously described [150]. $1x10^5$ cells per well were plated on 24-well tissue culture plates (Corning), and cells transfected 16-18 hours later using Lipofectamine 2000 (Invitrogen) and Opti-MEM (Gibco), and following the protocol provided by the manufacturer. Prior to transfection, the culture media was replaced with antibiotic-free culture

media, and 200 ng of each LINE-1 construct and 1 µL of Lipofectamine 2000 were used per well (of a 24-well tissue culture plate). 6 h after transfection, media was replaced with fresh complete media containing the indicated amount of each RTi (0, 2,5, 5, 10, 15, and 20 µM unless otherwise indicated). Puromycin selection (1 µg/mL, Sigma) was started 24 h post-transfection, using culture media containing the indicated amount of each RTi. Selection was continued for 3 additional days, and then Firefly and Renilla luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega) following manufacturer's instructions and a GloMax-Multi Detection System (Promega). Luciferase values for RTi untreated wells were designated as 1.

5.2.6 MTT Assays

Assays were conducted in triplicate using $1x10^3$ cells/well (HeLa or PA-1 cells) plated on 96-well tissue culture plates (Corning). 16-18 h after plating, culture media was replaced with fresh media containing the indicated amount of each RTi (0, 5, and 25 µM unless otherwise indicated), and cells cultured for 72 h. MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Life Technologies) was dissolved in 1x PBS (5 mg/mL), and 10 µL/well were added to 96-well tissue culture plates. Plates were incubated during 3 hours at 37 °C, media removed, and 100 µL DMSO were then added to each well, followed by an incubation at 37 °C during 15 minutes. After the incubation, the absorbance was measured using a GloMax-Multi Detection System (Promega) with a test wavelength of 560 nm and a reference wavelength of 750 nm to obtain sample signal (OD_{570}-OD_{630}). Blank wells for background containing only media were included as controls, as well as blank wells containing each RTi and culture media (to discard the possible contribution of RTis to absorbance values). Data is represented as the mean values with standard deviation, where untreated cells were arbitrarily designated as 1 for comparisons.

5.2.7 qPCR and RT-qPCR Control Assays

To determine whether RTi treatments affect the stability of LINE-1 constructs in cultured cells, we transfected a human L1 overexpression plasmid (JM101/L1.3) in

HeLa cells, and we then quantified the amount of plasmid DNA after RTi treatment. $8x10^4$ HeLa cells were plated per well of a 6-well tissue culture plate in triplicate; after 16-18 hours, cells were transfected with 1 µg of plasmid JM101/L1.3 using 3 µL of FuGene 6 (see above). After 24 h, cells were feed with fresh media containing 25 µM of the indicated RTi (β-ddAC, Emtricitabine, Lamivudine, Tenofovir, C5, and AZT) or vehicle, and cells cultured for 48h. After 48h, genomic DNA from transfected cells was extracted using phenol:chloroform:isoamyl alcohol (25:24:1) as described [150]. Next, 50 ng of each extracted DNA were used in qPCR reactions, using a StepOnePlus Real-Time PCR System (Applied Biosystems), GoTaq qPCR Master Mix Kit (Promega) and the following program: 1x (95 ºC, 10 min); 40x (95 ºC, 15 s; 60 ºC, 60 s). To quantify the relative amount of L1 plasmids after RTi treatments, we amplified the unique EBNA-1 region included in plasmid JM101/L1.3 and a portion of the human GAPDH gene to normalize for copy number differences, and using pair of primers whose amplification efficiencies do not differ more than 5%. To do that, we first generated a standard curve by serially diluting genomic DNA or plasmid JM101/L1.3, starting with 1 µg and making five five-fold dilutions. We then quantified copy numbers from three technical replicates of each reaction using the StepOnePlus Software v2.3. We used the EBNA-1/GAPDH ratio to calculate differences in copy number. Untransfected controls were used to discard plasmid DNA contaminations.

Similarly, and to analyse if RTi treatments affect the expression of L1 RNAs from transfected plasmids, we transfected a human L1 overexpression plasmid (JM101/L1.3) in HeLa cells, and we then quantified the amount of expressed L1 RNAs after RTi treatment. To quantify expressed L1 RNAs, we used: i) a NEO primer pair designed to only amplify spliced *mneoI*-tagged L1 transcripts; ii) a EBNA-1 primer pair designed to detect expressed EBNA-1 RNAs from transfected plasmids; and iii) GAPDH to normalize for expression differences. Briefly, HeLa cells were plated, transfected and treated with RTis as described above, and total RNA was isolated using TRIzol (Invitrogen). Next, 1 µg of total RNA from each treatment condition (β-ddAC, Emtricitabine, Lamivudine, Tenofovir, C5, AZT, or vehicle) was treated twice with 10 units of RNase-free DNaseI (Invitrogen), to

completely remove the transfected plasmid DNA from extracted RNAs. cDNAs were then synthesized using a High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) following the instructions provided by the manufacturer. Diluted cDNAs (1/5 and 1/10, and in triplicate) were then analyzed using a StepOne Real-Time PCR System (Applied Biosystems), GoTaq qPCR MasterMix (Promega) and 0.15 μM of each primer (Sigma). We included an internal control (no RT added) in all subsequent qPCR reactions. The qPCR cycling conditions were: 1x (95 ºC, 10 min); 40x (95 ºC, 15 s; 60 ºC, 60 s); a melting curve was recorded to confirm the identity of amplified products. To calculate differences in L1 RNA expression from transfected plasmids after RTi treatments, we used the NEO/EBNA-1 ratio and the comparative C_T ($\Delta\Delta C_T$) method [508].

5.2.8 Western Blot Analyses

PA-1 cells were plated ($2x10^5$ cells/well of a 6-well tissue culture plate) and treated with 25 μM of the indicated RTi during 96 h. After the treatment, cells were harvested using a cell scrapper and duplicate Whole Cell Extracts (WCE) prepared using RIPA buffer (Sigma) supplemented with 1x Complete Mini EDTA-free Protease Inhibitor cocktail (Roche), 0.1% Phosphatase Inhibitor 1&2 (Sigma), 1 mM (PMSF) (Sigma) and 0.25% β-mercaptoethanol (Sigma), by incubating cells during 10 min on ice. Cellular debris was removed by centrifugation (1,000 g for 5 min at 4 °C) and total protein concentration was determined using the Micro BCA Protein Assay Kit (Thermo) following standard procedures. Equal amounts of protein lysates were run on 10% SDS-PAGE gels and transferred to nitrocellulose membranes (Bio-Rad). Membranes were blocked in 5% milk/1x TBST (TBS + 0.2% Tween-20 (v/v)) and incubated with primary antibodies diluted in 5% milk/1xTBST overnight at 4 °C. Membranes were then washed 3 times with 1xTBST, incubated with secondary antibodies for 1 h at RT, washed again and developed. A chemiluminescent detection system (Clarity Western ECL Substrate, Bio-Rad; ImageQuant LAS 4000, GE Healthcare) was used, following manufacturer's recommendations. The following antibodies were used for immunoblotting (at indicated dilutions): mouse anti-L1Hs-ORF1p (1:1000, Merck Millipore); mouse anti-p53 (1:300; Santa Cruz Biotechnology); mouse anti–β-actin (1:20,000; Sigma);

horse anti-mouse HRP-linked secondary antibody (1:20,000, Cell Signaling Technology).

5.2.9 Quantification and statistical analysis

We used GraphPad Prism (version 6; GraphPad Software, Inc.) for statistical analyses. Data from multiple independent experiments is reported as retrotransposition mean ± SD, and we arbitrarily designed untreated cells as 100% Retrotransposition. Student t-test was used to calculate statistical significance and * indicates $p < 0.05$; **, $p < 0.01$, ***, $p < 0.001$, and ****, $p < 0.0001$.

CHAPTER 2: GENERATING HUMAN PLURIPOTENT CELL MODELS DEVOID OF LINE-1 RETROTRANSPOSON ACTIVITY USING CRISPR/CAS9 SYSTEM

The next part of this Dissertation was researched during a 3-month stay at the MRC Human Genetics Unit (MRC lnstitute of Genetics & Molecular Medicine, Western General Hospital, University of Edinburgh, Edinburgh, Scotland, UK) under the supervision of Dr. Thomas Widmann.

ABSTRACT

The use of human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) is a promising strategy that will open the door to a new era in regenerative medicine because of their capacity to be differentiated into any cell type. However, some obstacles must be solved prior the use of such cells in clinic, like the recently reported LINE-1 retrotransposition during hESCs and hiPSCs culture. LINE-1s comprise 17% of the human genome, although most L1 copies are inactive. However, an average human genome contains 80-100 active L1s, which move by a copy and paste mechanism that involves the reverse transcription of an intermediate L1 RNA. In addition, during this process other mRNAs are retrotranscribed (Alu, SVA and others) so hESCs and hiPSCs culture involves the impact produced by transposable elements mobilization. In fact, 11 *de novo* insertions have been described during hESCs and hiPSCs culture, one of them impacting the *CAPDS2* gene leading to protein synthesis failure. Similarly, transposable element mobilization prevents the use of pig epithelial tissues in xenotransplants, but recently, the use of a CRISPR/Cas9 system has been proved to efficiently inactivate all copies of pig endogenous retroviruses (PERVs). In this work, our aim is to design a CRISPR/Cas9 system that specifically targets currently active LINE-1 copies in human genome. To do this, up to 6 sgRNAs were designed to specifically target L1-Ta subfamily and were then tested on PA-1 cells, which naturally express endogenous LINE-1, to analyse their ability inactivating LINE-1 elements. Although our results are not conclusive, data seem to indicate that the designed CRISPR/Cas9 systems may inactivate LINE-1 copies and could potentially be used in the future on hESCs and hiPSCs, thus bringing closer the use of these cells in regenerative medicine.

1. INTRODUCTION

1.1. Human Embryonic Stem Cells and Human Induced Pluripotent Stem Cells

Human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) have the capacity to be derived into virtually any cell type. This has obviously opened the door to regenerative therapies that could be potentially used to treat or completely remove a wide range of diseases.

The first ESCs line obtained was mouse embryonic stem cells (mESCs) [509], [510] whereas hESC line were derived from the inner cell mass of preimplantation blastocysts almost two decades later [511]. This cell types have been very useful for research purposes for situations difficult or impossible to be studied *in vivo*. This includes the study of cell differentiation, disease modelling, treatment and prevention of birth defects, tissue transplantation and testing of drugs on hESC-derived cell lines when obtaining them by other sources is not achievable [512], [513]. The ethical issues around hESCs has limited the research and applications of this cell types so that mESCs have been used for certain experiments such as animal models. ESCs can be genetically modified to lead to a specific disease in the adult animal after embryo reimplantation [514]. Other techniques different from fertilization have been developed for obtaining hESCs like ESCs derived from parthenogenetic embryos (pESCs) [515], [516] or the transplantation of somatic cell nuclei into oocyte cytoplasm which reprograms that nuclei to pluripotency leading to the so called nuclear transfer (NT)-ESCs [517] (technique used to clone Dolly [518]). These two types of ESCs were principally obtained in order to avoid autoimmune response when regenerative medicine want to be applied. In this sense, hiPSCs share the same features and potential uses than hESCs and represent a more feasible and efficient strategy to avoid immune rejection and possess less ethical restrictions since they are not derived from the fertilization process but from somatic tissues. hiPSCs were firstly obtained by reprogramming fibroblasts through the introduction of specific transcription factors using retroviruses [519] but other transcription factors have been described

for the same purpose [520]–[522] among other strategies different from transcription factor integration such as excision of reprogramming factors [523]–[525], episomal vectors [526] or transfection of proteins [527] or mRNA [528]. The culture of hiPSCs derived from patients with a genetic disease, allows the generation of disease models to study the pathogenesis and also to test therapeutic strategies [529]–[531].

Despite the promising future of hESCs and hiPSCs, some concerns have to be solved previous the medical application of such therapeutic strategy, reason why they are subjected to a wide battery of quality controls. When cultured, hESCs and hiPSCs are prone to gain chromosomes [532], to accumulate point mutations that frequently target oncogenic pathways [533] and copy number variants [534] and are epigenetically variable [535] among others (reviewed in [529]). Remarkably, LINE-1 retrotransposition is activated during cell reprogramming caused by the promoter hypomethylation profile changes, as well as hESCs [167]. In addition, engineered LINE-1 retrotransposed more efficiently in hiPSCs than in their parental cell line. These experiments evidenced the impact that LINE-1 produces in the genome of hESCs and hiPSCs when cultured with the associated structural and functional effects described in Chapter 1. In fact, later was described 11 insertions of retroelements (L1, *Alu* and SVA) during pluripotent stem cells (PSCs: hESCs and hiPSCs) culture, of which 6 had took place in protein-coding genes expressed in PSCs and 7 LINE-1 insertions, being 4 of them full length. Furthermore, a *de novo* LINE-1 insertion was found in an intron of the *CAPDS2* gene causing protein synthesis failure [214]. Taking into account that there are ~100 active copies of the LINE-1 element, the risk for PSCs of suffering a deleterious insertion is importantly high. (For more information about LINE-1 see Chapter 1). Similarly, porcine endogenous retroviruses (PERVs) represent the main barrier for the use of pig epithelial tissues for xenotransplantation because of the possibility to transmit these PERVs to the human genome. Recently, CRISPR/Cas9 technology was applied for the successful removal (in one clone) of the 62 copies of PERVs present in a porcine kidney epithelial cell line (PK15), targeting the *pol* sequence corresponding to

the reverse transcriptase domain [536]. They demonstrated that no PERV transmission from PERVs-inactivated pig PK15 cells to human HEK293T cells was detected, but more importantly, they showed that targeting a high number of genomic sites was possible with CRISPR/Cas9, since the previous reported simultaneous edition had been 6 [537]. The same approach could be used to inactivate LINE-1 copies.

Many are the invaluable possibilities offered by hESCs and hiPSCs, but even greater are these if we combine the pluripotent capacity of hESCs and hiPSCs with the genome editing provided by the CRISPR/Cas9 technology to solve some problematic issues around the use of these cell types (and others, like pig epithelial tissues) in regenerative medicine.

1.2. Genome Editing: CRISPR/Cas9

First attempts to modify the genome were based on double-strand breaks (DSBs) generated by means of site-specific nucleases (SSNs), after which the cell is prone to repair it through homology directed repair (HDR) using an exogenously provided repair template [538], [539]. The low efficiency of the previous method led to the development of designed zinc-finger nucleases (ZFNs) that recognize a unique genomic site where these proteins generate a DBS, which is repaired by cell machinery through HDR [540]. Adding the specific template, researchers successfully corrected an X-linked severe combined immune deficiency (SCID) mutation in the IL2Rγ gene in a yield of 18% [541]. This method was the first to earned the name of "genome editing" and resulted in a clinical trial to disrupt CCR5 in T cells to treat persons infected with HIV [542]. Similar yields can be obtained with the later discovered transcription activator-like effector nucleases (TALENs), which are comprised of a non-specific FokI nuclease domain fused to a DNA-binding domain that is edited to recognize a specific DNA sequence [543], [544]. ZFNs and TALENs have been used in PSCs with many purposes [523], [545]–[548].

However, the real revolution came with the discovery by the Spanish researcher Francisco Mojica of clustered regularly interspaced short palindromic repeats (CRISPR). Found in prokaryotes, CRISPR were originated from retrovirus

infections or from the acquisition of exogenous plasmids that were integrated into the genome. CRISPR loci structure includes direct repeats (23-47 bp) separated by short sequences (21-72 bp) termed protospacers (sequences acquired from exogenous DNA) and are frequently adjacent to Cas (CRISPR-associated) genes. There are up to 45 families of Cas genes in prokaryotes with various functional domains including polymerases, nucleases, helicases and polynucleotide-binding proteins required for CRISPR functionality. Since prokaryotes are highly exposed to foreign DNA (via transduction, conjugation and transformation), they have developed different systems to protect their genome integrity. CRISPR acts as an acquired immunity, targeting exogenous DNA in a sequence-specific manner because of their complementarity since they target "the same sequences" they come from [549]–[552]. Its mechanism has not been fully elucidated, but it is thought that some Cas proteins mediate the acquisition of new spacers while others are required for the interference against invasive DNA [553]–[555]. The Cas9 protein soon emerged as a special one. It was detected that in a subset of CRISPR systems (CRISPR type-2), Cas9 was able to associate with an engineered single guide RNA (sgRNA) complementary to a desired genomic location where Cas9 exhibited it nuclease activity producing a DSB. The system was highly specific, property conferred by the 20 nucleotides engineered sgRNA [556]. Soon after this discovery, full described methods for the application of CRISPR/Cas9 technology in genome editing were published, together with the option to repair the DSB with a desired template [537], [557]–[559]. The potential of this genome editing strategy encouraged many research groups to its application on PSCs and, as a result, a large number of publication regarding this matter have appeared, from human disease modelling disrupting key genes to repairing point mutations that lead to a disease (review in [560], [561]).

Many modifications have been made on CRISPR/Cas9 system aiming to improve its versatility and efficiency (reviewed in [562]). Some of the variants allow targeting of mRNA (CRISPR/C2c2) [563] or promoter methylation of desired gene (CRISPRi) [564], [565] as well as Cas9 proteins that either cleave one strand or both strands. Here, the mechanism of the method used in the present work will be

briefly described [558], [559]. Prokaryotic CRISPR system can be applied to mammalian cells through three basic elements: the CRISPR-associated nuclease Cas9 (SpCas9), the CRISPR RNA (crRNA) that determines the target specificity and an auxiliary trans-activating crRNA (tracrRNA) partially complementary to the crRNA. Upon transcription of pre-crRNA and tracrRNA and hybridization, the dsRNA form a complex with Cas9 protein. Host RNase III processes this complex into discrete units that, guided by the complementary crRNA, specifically recognize a 20-bp genomic target (contiguous to a protospacer adjacent motif or PAM) and then, endonuclease activity of Cas9 cleaves the sequence for which the crRNA has been designed, generating a DSB. In practice, tracrRNA and crRNA base pairing is exogenously prepared to form a chimeric sgRNA. The DSBs generated by CRISPR can be repaired by the host cell through two main pathways. When a template is provided to the system, cells tend to repair DNA damage using the HDR mechanism, what provides a way to edit the genome, like repairing a disease-causing mutation or simulating a naturally occurred mutation to study the caused disease. However, this pathway occurs almost exclusively in dividing cells and shows a variable efficiency across cell type and state and depending on the genomic target and the exogenous template. The other repair mechanism, nonhomologous end joining (NHEJ), simply religates the DSB leaving indel mutations. This strategy is the preferred one when a gene want to be disrupted because it entails a simpler protocol since an indel in an exon usually lead to a frameshift mutation.

1.3. LINE families

Three LINE-1 families can be found in human genomes, namely L1, L2 and L3, that have successively replaced each other over the last 100 million years so the only active family nowadays is LINE-1 family (or L1HS for human specific) [3], [36]. The analysis of the LINE-1 5'UTR sequences led to the subclassification of this LINE family into 17 classes (L1PA1 - L1PA17) that resemble each other in their ORF sequences but significantly differ in their 5'UTR. The younger L1PA1 (L1HS), also known as L1-Ta (for transcribed active) subfamily encompasses all active (full-length) LINE-1 elements. Deeper studies led to the establishment of L1-Ta subsets based on sequence differences: Ta-1 (Ta-1d and Ta-n1d), Ta-0 and non-Ta. These

studies showed that ORF1 and ORF2 are conserved while 5'UTRs are not [566], [567].

When using CRISPR/Cas9 to target LINE-1, it should be kept in mind the sequence variations between LINE-1 subfamilies in order to design sgRNAs capable of discerning active from inactive elements to avoid off-target DSBs that could triggers apoptosis.

2. OBJECTIVES

The mutagenic effects of LINE-1 mobilization during hESCs and hiPSCs culture will prevent their use in regenerative medicine. Since around 80-100 copies of L1 elements are still active within the human genome, this risk must to be solved prior therapeutic use of PSCs. According to the inactivation of PERVs abovementioned, it is possible to target multiple genomic locus with CRISPR/Cas9 technology. Based on that, our aim is to inactivate active LINE-1 elements within human genome.

1. Design of a highly specific CRISPR/Cas9 system to target active LINE-1 elements.

    1.1. Design of single guide RNAs that specifically target active LINE-1 copies.

    1.2. Transfection of the CRISPR/Cas9 system designed and analysis of indels on PA-1 cells.

2. Application of the validated system on hESCs and hiPSCs.

3. RESULTS AND DISCUSSION

3.1. Design of a highly specific CRISPR/Cas9 system to target active LINE-1 elements

3.1.1. Design of single guide RNAs that specifically target active LINE-1 copies

For the design of the sgRNAs that target LINE-1 sequence, an online tool provided by Zhang lab. (http://crispr.mit.edu:8079/, no longer available but others have appeared). Given that ORF1 and ORF2 are more conserved, their consensus sequences (from L1.3, accession number #L19088) were input into the online tool and a series of sgRNAs were returned. This tool only recognized sequences adjacent to the required PAM sequences (NGG) so the initial list of sgRNAs must be filtered taking into account the L1PA1-L1PA16 sequence alignment simultaneously with the sequence alignment of currently active LINE-1 elements (L1PA1) [100], [566] together with the score (quality) and potential number of off-targets provided by the online tool. Despite the fact that CRISPR/Cas9 system is highly specific, sgRNAs allow certain degree of mismatches with target DNA [558] so filtration process was carried out carefully to select proper guides. Regarding this, selected sgRNAs should be as complementary as possible to conserved ORF regions within L1PA1 subfamily but, at the same time, they should be as non-complementary as possible to LINE-1 elements distinct from L1PA1, decreasing this way possible off-targets thus, less DSBs and in consequence, less DNA damage.

After this process a total of 6 guides, containing the 20 nucleotide guide and the PAM stretch (NGG), were selected for further research:

-3 targeted around position 150 of the ORF1 sequence (overlapped):

-1568: GATCAAATTACTCTGAGCTACGG
-1570: ATCAAATTACTCTGAGCTACGGG
-1588: AAATTACTCTGAGCTACGGGAGG

-1 targeted around position 280 of the ORF1 sequence

-1593: TTCACGTAGTTCTCGAGCCTTGG

-1 targeted the endonuclease coding sequence

-2567: GGTCAATTTTGGAATAGGTGTGG

-1 targeted the reverse transcriptase coding sequence.

2280: CAATCATGTCGTCTGCAAACAGG

As mentioned in Chapter 1, endonulease and reverse transcriptase are strictly required for LINE-1 retrotransposition, therefore, potential indels generated by sgRNAs **2567** and **2280** may inactivate L1PA1 elements. On the other hand, sgRNAs **1568**, **1570** and **1588** could lead to frameshift mutations that also inactivate L1PA1 copies. To test the selected sgRNAs, they must be hybridized and cloned into expression plasmid pX458 (pSpCas9(BB)-2A-GFP) [559], which mainly contains the ampicillin resistance gene for bacteria selection, enhanced green fluorescent protein for transfection selection, the Cas9 gene and a cloning site. The plasmid was digested with *BbsI* and purified in an agarose gel to cut the band corresponding to the digested plasmid, which was subsequently purified with Qiagen gel extraction kit. sgRNA top and bottom oligonucleotides were annealed and ligated into the digested pX458 plasmid and the product was propagated in transformed bacteria. Colony PCR using U6 forward primer and bottom sgRNA oligonucleotide, was used to check the presence of the desired plasmid within colonies (**Figure 65**). Finally, plasmids were purified with Qiagen miniprep kit and sequenced with U6-Fwd primer to definitely confirm the successful ligation.

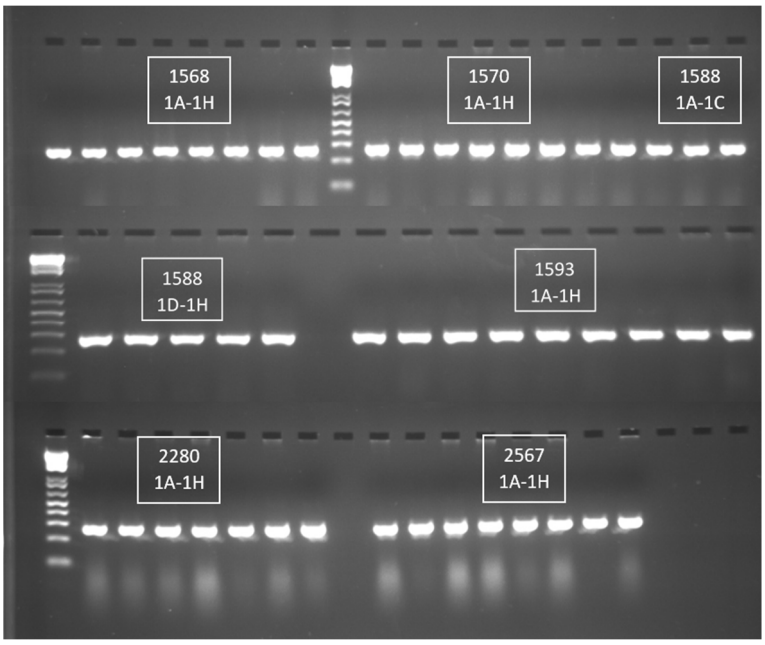


Figure 65. *Agarose gel electrophoresis of colony PCR products. Eight colonies per plasmid ligation were picked (1A-1H).*

Prior transfection of CRISPR plasmids on PA-1 cells, a method to analyse later the impact made on genomic LINE-1 elements was required. Blue-white screening (in bacteria) provides a useful method to determine the no mutations/indels ratio and is based on the frameshift mutation (or not) of the LacZ gene, contained in the pBlueScript II SK (+) plasmid (pBSSKII(+)), upon PCR product ligation. Thereby, if the PCR product, once ligated, does not alter the frameshift of LacZ gene, colonies will be able to express β-galactosidase that transform X-gal presents in the culture media into a blue compound, but if the PCR product leads to a frameshift mutation, colonies will not be able to express such enzyme and therefore will remain white. For this purpose, different set of primers flanking the sgRNA targets that amplify 300 bp, were designed so that PCR products lacking any mutation could be cloned into the LacZ gene without producing a frameshift, while PCR products containing indels (excepting multiples of 3 bp) would do. As controls, plasmid LM101/L1.3 DNA (see Chapter 1) and PA-1 genomic DNA were used to test that designed primer pairs were able to amplify a consensus LINE-1, what was positively confirmed (**Figure 66**).

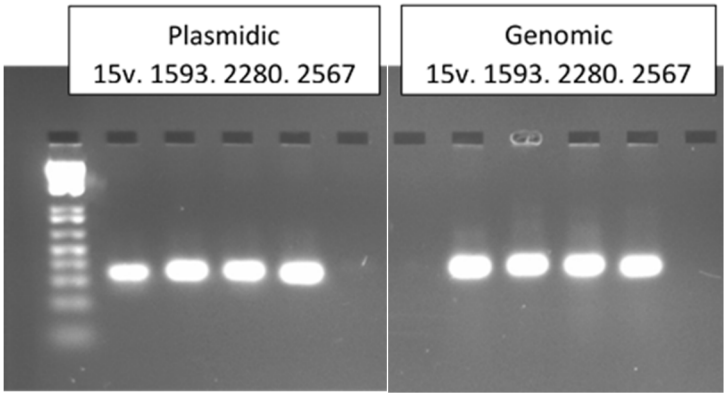


Figure 66. *Agarose gel electrophoresis of plasmid JM101/L1.3 and PA-1 genomic DNA PCR products. 15v includes the three overlapped targeted sequences. 100 and 300 indicate the size of the PCR products. Ladder is included in the left lane.*

PCR products were purified, digested and cloned into pBSSKII(+) digested plasmid. Bacteria transformation and plating on X-gal-containing dishes showed that all colonies were blue. To check if PCR products of plasmid DNA had been correctly inserted, three colonies per dish were selected for plasmid purification, which were then digested with the same restriction enzymes used for PCR product

digestions to confirm PCR ligation (**Figura 67**). Positive samples were sent to sequencing with M13 Fwd primer.

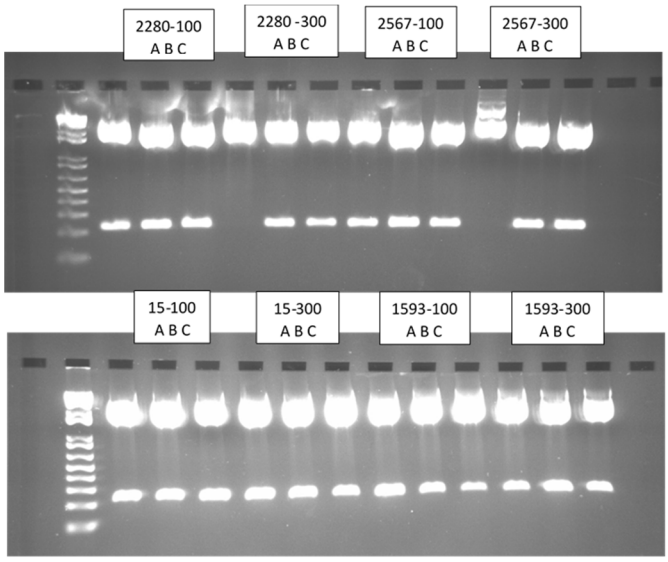


***Figure 67**. Electrophoresis in agarose gel of digested pBSSKII(+) plasmids containing cloned PCR products.*

At the same time, and to analogously confirm the insertion of PCR products from plasmid and genomic DNA, colony PCRs of two colonies per well were carried out (**Figure 68**).

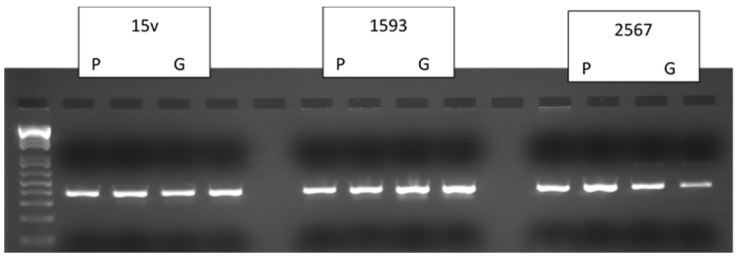


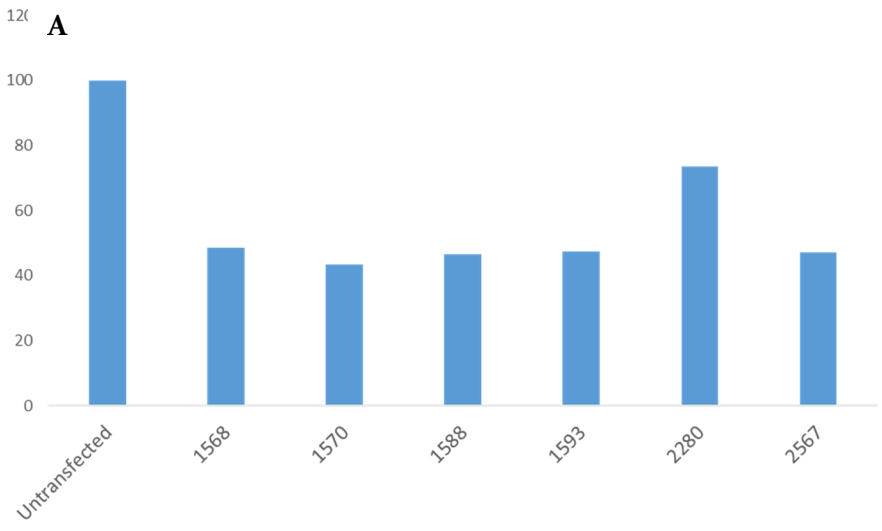
***Figure 68**. Agarose gel electrophoresis of colony PCRs of ligated plasmid (P) and genomic (G) PCR products.*

3.1.2. Transfection of the designed CRISPR/Cas9 system and analysis of indels on PA-1 cells

Since PA-1 cells naturally express LINE-1, we decided to use this cell line to analyse the efficiency of the designed sgRNAs to inactivate LINE-1 elements. To do this, PA-1 cells were seeded and transfected with the corresponding pX458 plasmids, each one containing the different sgRNAs, using FuGene6 as described in Chapter 1. After 48-72 h, cells were sorted and transfected population was

selected for further growing, after which samples for genomic DNA, mRNA and protein analysis were taken.

Western Blot analysis were performed to determine if the sgRNAs were able to produce any significant change in the levels of ORF1p (**Figure 69B**). Quantification of the Western Blot showed a general reduction of the ORF1p levels in PA-1 cells treated with CRISPR/Cas9 system when compared with untransfected PA-1 cells (**Figure 69A**). Noteworthily, all sgRNAs whose target was different from the RT domain led to a reduction of approximately the 50% of ORF1p. The only sgRNA that targeted the RT domain just decreased this value in a 30%. However, it must make clear that these results are based on just one sample so they have to be validated with more biological replicates and using other tests. Unfortunately, the analysis of genomic indels generated by the sgRNAs designed, as well as LINE-1 transcript quantification, were not possible due to the end of the short-term stay.



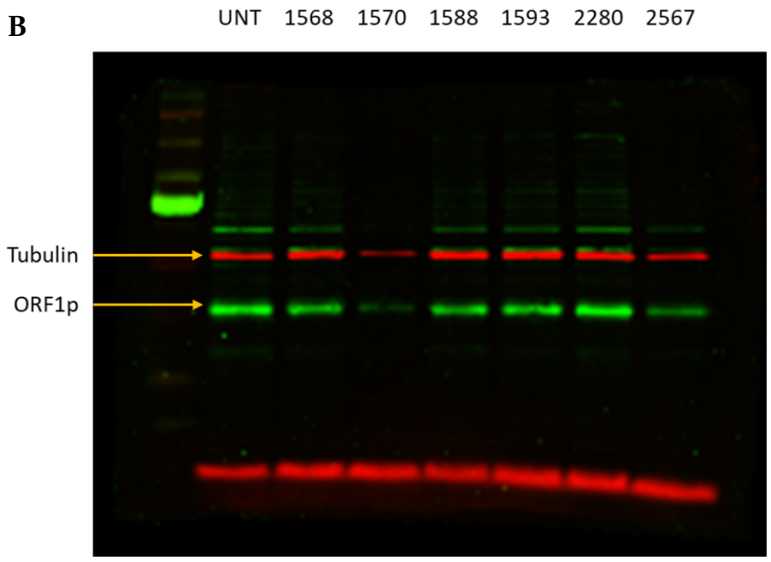


***Figure 69**. Western Blot of PA-1 protein extract treated with CRISPR/Cas9 system. Quantification of the LINE-1 ORF1p. Data were normalized with α-tubulin.*

The mobilization of human retrotransposons during culture of hESCs and hiPSCs supposes a great challenge that must be solved previously to the use of these cells in regenerative medicine. The use of a CRISPR/Cas9 system has proved its potential inactivating a total of 62 PERVs present in pig genome [536]. In a similar approach, we took advantage of this strategy with the aim of inactivating active LINE-1 elements in human genome. We designed 6 sgRNAs that specifically target this subfamily of retrotransposons and tested them in PA-1 cells, which endogenously express LINE-1. Although the study of such sgRNAs was not completed and many controls remain to be done, Western Blot analysis of the extracts of transfected PA-1 cells showed some promising potential in inactivating LINE-1 elements. If this tendency would be confirmed, a second round of transfection with CRISPR/Cas9 systems could lead to further inactivation. In addition to the quantification of LINE-1 expression and the analysis of targeted genomic sequences, the system have to be tested on hESCs and hiPSCs. Furthermore, off-target effects and specificity of sgRNAs evaluation are still required. With this research and future work, we hope to lay the foundation on this matter.

4. CONCLUSIONS

1. Six sgRNAs were designed to specifically target active LINE-1 elements and were cloned into CRISPR/Cas9 expression vector.

2. The designed CRISPR/Cas9 systems seem to be effective decreasing LINE-1 ORF1p levels.

5. MATERIALS AND METHODS

5.1. Cell Line Authentication and Cell Line Conditions

It has been described in Chapter 1.

5.2. Expression Plasmids

Plasmids used in this study were gifted by Martin Reijns (MRC Human Genetics Unit, Edinburgh, UK) and purified using Plasmid Midi kits from Qiagen following manufacturer instructions. Plasmid DNAs were analysed by electrophoresis (0.7% agarose-ethidium bromide gels) and we only used highly supercoiled preparations of plasmid DNA for transfections.

-**pBSSKII(+)** (Stratagene, cat. number #212205) is a standard cloning vector and was propagated in Escherichia coli strain DH5α (ThermoFisher).

-**pX458** has been described previously [559] and contains the Cas9 gene. It was propagated in Stbl3 strain (ThermoFisher).

5.3. sgRNAs hybridization and cloning into pX458 plasmid

It has been previously described [559]. Briefly, plasmid pX458 was digested with *BbsI* restriction enzyme following manufacturer's protocol. Digested plasmid was resolved in a 1% agarose-SYBR safe gel, excised and purified (Qiagen Gel Extraction Kit).

For sgRNA oligonucleotides hybridization, 20 µL of top and 20 µL of bottom sgRNA strands (100 µM, final concentracion 50 µM, **Table 6**) were mixed and incubated at 95 °C for 5 minutes, after which reactions were slowly cooled down, diluted 1:50 and ligated into the digested pX458 plasmid. Reactions without inserts were used as controls.

sgRNA	Sequences ordered	
1568	Top strand:	5'-caccggGATCAAATTACTCTGAGCTA
	Bottom strand:	5'-aaacTAGCTCAGAGTAATTTGATCcc
1570	Top strand:	5'-caccggATCAAATTACTCTGAGCTAC
	Bottom strand:	5'-aaacGTAGCTCAGAGTAATTTGATcc
1588	Top strand:	5'-caccggAAATTACTCTGAGCTACGGG
	Bottom strand:	5'-aaacCCCGTAGCTCAGAGTAATTTcc

1593	Top strand:	5'-caccggTTCACGTAGTTCTCGAGCCT
	Bottom strand:	5'-aaacAGGCTCGAGAACTACGTGAAcc
2280	Top strand:	5'-caccggCAATCATGTCGTCTGCAAAC
	Bottom strand:	5'-aaacGTTTGCAGACGACATGATTGcc
2567	Top strand:	5'-caccggGGTCAATTTTGGAATAGGTG
	Bottom strand:	5'-aaacCACCTATTCCAAAATTGACCcc

Table 6. *sgRNAs used in this study.*

5.4. CRISPR/Cas9 systems evaluation on PA-1 cells

To determine the efficiency of the designed CRISPR/Cas9 systems, $1x10^5$ PA-1 cells were plated per well of a 6-well tissue culture plate in duplicate. After 16-18 hours, cells were transfected using FuGene6 with pX458 plasmids containing the different guides as described in Chapter 1. 48-72 h after transfection cells were trypsinized and sorted. Selected population were seeded again in 10-cm dishes in order to expand them. Cells were scrapped off to be analysed (genomic DNA, LINE-1 mRNA and proteins).

5.5. PCR analysis

All PCRs were carried out using Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher) following provided protocol and 10 ng of plasmid DNA or 100 ng of extracted genomic DNA (as described in Chapter 1) from transfected cells were used as templates in PCRs. DNA-free water was included as a negative control. PCR products were purified with Qiagen PCR purification kit and analysed in a 2% agarose-ethidium bromide gel.

When analysis of targeted sequences, either genomic or plasmidic, were carried out, primers pairs in **Table 7** were used and PCR conditions for amplification were as follows: 1× (98°C, 10 s); 35× (98 °C, 1 s; 58 °C, 5 s; 72 °C, 5 s); 1× (72°C, 1 min).

Primer	Sequences ordered	
15V	Fwd primer:	5'-aGGATCCgatggggaaaaaacagaacagaaaaac
	Rv primer:	5'-aGGTACCaggcttctgcattcttcacgtagttc
1593	Fwd primer:	5'-aGGATCCaggcttcagacgatcaaattactc
	Rv primer:	5'-aGGTACCacatagtcccatatttcttggagg
2280	Fwd primer:	5'-aGGATCCagcattccctttgaaaactggcac
	Rv primer:	5'-aCTCGAGcccatgatttggctctctgtttg
2567	Fwd primer:	5'-aGGATCCagacagatcaacgagacagaaagtc
	Rv primer:	5'-aCTCGAGagcaggttgttcagtttccatg

Table 7. *Primers used in this study to analyse CRISPR/Cas9 systems designed effects.*

When sgRNA ligation wanted to be confirmed, colony PCRs were carried out using sgRNA bottom strands and U6-Fwd primer (GAGGGCCTATTTCCCATGATTCC) to amplify desired regions as follows: 1× (95°C, 5 min); 40× (98 °C, 1 s; 58 °C, 55 s; 72 °C, 5 s); 1× (72°C, 1 min).

When PCR products ligation wanted to be confirmed, colony PCRs were carried out using Fwd primers from **Table 2** and M13-Fwd primer (GTAAAACGACGGCCAG) as abovementioned.

5.6. Analysis of targeted sequences

PCR products from amplification of targeted sequences were digested as follows: for products corresponding to sequences targeted with 1568, 1570, 1588 and 1593 CRISPR/Cas9 systems, *BamHI*/*KpnI* restriction enzymes (from NEB) were used, while for products corresponding to sequences targeted with 2280 and 2567 CRISPR/Cas9 systems, *BamHI*/*XhoI* restriction enzymes (from NEB) were used. Digested PCR products were purified with Qiagen PCR Purification Kit. Plasmid pBSSKII(+) was separately digested with both mixtures of enzymes, resolved in a 1% agarose-SYBR safe gel, excised and purified with Qiagen Gel Extraction Kit. Inserts were ligated into purified digested pBSSKII(+). Bacteria of the *E. coli* DH5α strain were transformed with these plasmids, seeded over agar plates in which 100 µL of X-Gal (stock solution 20 mg/mL) and 40 µL IPTG (stock solution 100 mM) were spread. After incubating overnight at 37 °C, blue/white colonies were counted.

5.7. Western Blot Analysis

Whole cell extract of transfected and sorted PA-1 cells were prepared as previously described in Chapter 1. For immunoblotting, the following antibodies were used (at indicated dilutions): rabbit polyclonal anti-LIHs-ORF1p (provided by Dr. Oliver Weichenrieder, Max-Planck, Germany, 1:5000), mouse monoclonal anti-α-tubulin (1:1000, Santa Cruz); goat anti-rabbit and anti-mouse fluorescent secondary antibodies (LI-COR) were used at a 1:20,000 dilution (goat anti-rabbit IRDye 800CW; goat anti-mouse IRDye 680LT; goat anti-mouse IRDye 800CW). Signals

were quantified using the freely available software LI-COR Image Studio Lite (LI-COR).

6.Bibliografía

6 BIBLIOGRAFÍA:

[1] Britten, R. J. and Kohne, D. E., 'Repeated Sequences in DNA', *Science*, 161, (3841), 529–540, 1968.

[2] Britten, R. J. and Davidson, E. H., 'Repetitive and Non-Repetitive DNA Sequences and a Speculation on the Origins of Evolutionary Novelty', *The Quarterly Review of Biology*, 46, (2), 111–138, 1971.

[3] Lander, E. S. *et al.*, 'Initial sequencing and analysis of the human genome', *Nature*, 409, (6822), 860–921, 2001.

[4] Koning, A. P. J. de, Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D., 'Repetitive Elements May Comprise Over Two-Thirds of the Human Genome', *PLOS Genetics*, 7, (12), e1002384, 2011.

[5] Lupski, J. R. and Weinstock, G. M., 'Short, interspersed repetitive DNA sequences in prokaryotic genomes.', *J. Bacteriol.*, 174, (14), 4525–4529, 1992.

[6] Charlesworth, B., Sniegowski, P., and Stephan, W., 'The evolutionary dynamics of repetitive DNA in eukaryotes', *Nature*, 371, (6494), 215–220, 1994.

[7] Gregory, T. R., 'Synergy between sequence and size in Large-scale genomics', *Nature Reviews Genetics*, 6, (9), 699–708, 2005.

[8] Cordaux, R. and Batzer, M. A., 'The impact of retrotransposons on human genome evolution', *Nat Rev Genet*, 10, (10), 691–703, 2009.

[9] Ohno, S., 'So much "junk" DNA in our genome', *Brookhaven Symp. Biol.*, 23, 366–370, 1972.

[10] Doolittle, W. F. and Sapienza, C., 'Selfish genes, the phenotype paradigm and genome evolution', *Nature*, 284, (5757), 601–603, 1980.

[11] Orgel, L. E. and Crick, F. H. C., 'Selfish DNA: the ultimate parasite', *Nature*, 284, (5757), 604–607, 1980.

[12] Makaowski, W., 'Genomic scrap yard: how genomes utilize all that junk', *Gene*, 259, (1), 61–67, 2000.

[13] Shapiro, J. A. and Sternberg, R. von, 'Why repetitive DNA is essential to genome function', *Biological Reviews*, 80, (2), 227–250, 2005.

[14] Rearden, A., Magnet, A., Kudo, S., and Fukuda, M., 'Glycophorin B and glycophorin E genes arose from the glycophorin A ancestral gene via two duplications during primate evolution.', *J. Biol. Chem.*, 268, (3), 2260–2267, 1993.

[15] Gatchel, J. R. and Zoghbi, H. Y., 'Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles', *Nature Reviews Genetics*, 6, (10), 743–755, 2005.

[16] Castel, A. L., Cleary, J. D., and Pearson, C. E., 'Repeat instability as the basis for human diseases and as a potential target for therapy', *Nature Reviews Molecular Cell Biology*, 11, (3), 165–170, 2010.

[17] McClintock, B., 'The origin and behavior of mutable loci in maize', *Proc. Natl. Acad. Sci. U.S.A.*, 36, (6), 344–355, 1950.

[18] McClintock, B., 'Controlling Elements and the Gene', *Cold Spring Harbor Symposia on Quantitative Biology*, 21, (0), 197–216, 1956.

[19] Kapitonov, V. V. and Jurka, J., 'A universal classification of eukaryotic transposable elements implemented in Repbase', *Nat Rev Genet*, 9, (5), 411–412, 2008.

[20] Wicker, T. *et al.*, 'A unified classification system for eukaryotic transposable elements', *Nature Reviews Genetics*, 8, (12), 973–982, 2007.

[21] Kleckner, N., 'Regulation of Transposition in Bacteria', *Annu. Rev. Cell. Biol.*, 6, (1), 297–327, 1990.

[22] Feschotte, C., Jiang, N., and Wessler, S. R., 'Plant transposable elements: where genetics meets genomics', *Nature Reviews Genetics*, 3, (5), 329–341, 2002.

[23] Emmons, S. W. and Yesner, L., 'High-frequency excision of transposable element Tc1 in the nematode caenorhabditis elegans is limited to somatic cells', *Cell*, 36, (3), 599–605, 1984.
[24] Jacobson, J. W., Medhora, M. M., and Hartl, D. L., 'Molecular structure of a somatically unstable transposable element in Drosophila', *PNAS*, 83, (22), 8684–8688, 1986.
[25] Mitra, R. *et al.*, 'Functional characterization of piggyBat from the bat Myotis lucifugus unveils an active mammalian DNA transposon', *PNAS*, 110, (1), 234–239, 2013.
[26] Ray, D. A. *et al.*, 'Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus', *Genome Res.*, 18, (5), 717–728, 2008.
[27] Kapitonov, V. V. and Jurka, J., 'RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons', *PLOS Biology*, 3, (6), e181, 2005.
[28] Agrawal, A., Eastman, Q. M., and Schatz, D. G., 'Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system', *Nature*, 394, (6695), 744–751, 1998.
[29] Robertson, H. M. and Lampe, D. J., 'Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera.', *Mol Biol Evol*, 12, (5), 850–862, 1995.
[30] Casse, N., Bui, Q. T., Nicolas, V., Renault, S., Bigot, Y., and Laulier, M., 'Species sympatry and horizontal transfers of Mariner transposons in marine crustacean genomes', *Molecular Phylogenetics and Evolution*, 40, (2), 609–619, 2006.
[31] Lampe, D. J., Witherspoon, D. J., Soto-Adames, F. N., and Robertson, H. M., 'Recent Horizontal Transfer of Mellifera Subfamily Mariner Transposons into Insect Lineages Representing Four Different Orders Shows that Selection Acts Only During Horizontal Transfer', *Mol Biol Evol*, 20, (4), 554–562, 2003.
[32] Houck, M. A., Clark, J. B., Peterson, K. R., and Kidwell, M. G., 'Possible horizontal transfer of Drosophila genes by the mite Proctolaelaps regalis', *Science*, 253, (5024), 1125–1128, 1991.
[33] Hartl, D. L., Lohe, A. R., and Lozovskaya, E. R., 'Modern thoughts on an ancyent marinere: function, evolution, regulation.', *Annu Rev Genet*, 31, 337–358, 1997.
[34] Haber, J. E., 'Partners and pathways: repairing a double-strand break', *Trends in Genetics*, 16, (6), 259–264, 2000.
[35] Munoz-Lopez, M. and Garcia-Perez, J., 'DNA Transposons: Nature and Applications in Genomics', *Current Genomics*, 11, (2), 115–128, 2010.
[36] Smit, A. F., 'The origin of interspersed repeats in the human genome', *Current Opinion in Genetics & Development*, 6, (6), 743–748, 1996.
[37] Smit, A. F. and Riggs, A. D., 'Tiggers and DNA transposon fossils in the human genome', *PNAS*, 93, (4), 1443–1448, 1996.
[38] Lambowitz, A. M., Gellert, M., Chandler, M., Craig, N. L., Sandmeyer, S. B., and Rice, P. A., Eds., *Mobile DNA III*. American Society of Microbiology, 2015.
[39] Bénit, L., Lallemand, J.-B., Casella, J.-F., Philippe, H., and Heidmann, T., 'ERV-L Elements: a Family of Endogenous Retrovirus-Like Elements Active throughout the Evolution of Mammals', *J. Virol.*, 73, (4), 3301–3308, 1999.
[40] Tristem, M., 'Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database', *J. Virol.*, 74, (8), 3715–3730, 2000.
[41] Soriano, P. and Jaenisch, R., 'Retroviruses as probes for mammalian development: Allocation of cells to the somatic and germ cell lineages', *Cell*, 46, (1), 19–29, 1986.
[42] Malik, H. S. and Eickbush, T. H., 'Phylogenetic Analysis of Ribonuclease H Domains Suggests a Late, Chimeric Origin of LTR Retrotransposable Elements and Retroviruses', *Genome Res.*, 11, (7), 1187–1197, 2001.

[43] Garfinkel, D. J., Boeke, J. D., and Fink, G. R., 'Ty element transposition: Reverse transcriptase and virus-like particles', *Cell*, 42, (2), 507–517, 1985.
[44] Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R., 'Ty elements transpose through an RNA intermediate', *Cell*, 40, (3), 491–500, 1985.
[45] Moyes, D., Griffiths, D. J., and Venables, P. J., 'Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease', *Trends in Genetics*, 23, (7), 326–333, 2007.
[46] Marchi, E., Kanapin, A., Byott, M., Magiorkinis, G., and Belshaw, R., 'Neanderthal and Denisovan retroviruses in modern humans', *Current Biology*, 23, (22), R994–R995, 2013.
[47] Wildschutte, J. H., Williams, Z. H., Montesion, M., Subramanian, R. P., Kidd, J. M., and Coffin, J. M., 'Discovery of unfixed endogenous retrovirus insertions in diverse human populations', *PNAS*, 113, (16), E2326–E2334, 2016.
[48] Dewannieux, M. *et al.*, 'Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements', *Genome Res.*, 16, (12), 1548–1556, 2006.
[49] Lee, Y. N. and Bieniasz, P. D., 'Reconstitution of an Infectious Human Endogenous Retrovirus', *PLOS Pathogens*, 3, (1), e10, 2007.
[50] Zhang, Y., Maksakova, I. A., Gagnier, L., Lagemaat, L. N. van de, and Mager, D. L., 'Genome-Wide Assessments Reveal Extremely High Levels of Polymorphism of Two Active Families of Mouse Endogenous Retroviral Elements', *PLOS Genetics*, 4, (2), e1000007, 2008.
[51] Mager, D. L. and Stoye, J. P., 'Mammalian Endogenous Retroviruses', *Microbiology Spectrum*, 3, (1), 2015.
[52] Bannert, N. and Kurth, R., 'The Evolutionary Dynamics of Human Endogenous Retroviral Families', *Annual Review of Genomics and Human Genetics*, 7, (1), 149–173, 2006.
[53] Stocking, C. and Kozak, C. A., 'Murine endogenous retroviruses', *Cell Mol Life Sci*, 65, (21), 3383–3398, 2008.
[54] Stoye, J. P., 'Studies of endogenous retroviruses reveal a continuing evolutionary saga', *Nature Reviews Microbiology*, 10, (6), 395–406, 2012.
[55] Grow, E. J. *et al.*, 'Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells', *Nature*, 522, (7555), 221–225, 2015.
[56] Mi, S. *et al.*, 'Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis', *Nature*, 403, (6771), 785–789, 2000.
[57] Mallet, F. *et al.*, 'The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology', *PNAS*, 101, (6), 1731–1736, 2004.
[58] Chuong, E. B., Elde, N. C., and Feschotte, C., 'Regulatory evolution of innate immunity through co-option of endogenous retroviruses', *Science*, 351, (6277), 1083–1087, 2016.
[59] Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H., 'Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition', *Cell*, 72, (4), 595–605, 1993.
[60] Cost, G. J., Feng, Q., Jacquier, A., and Boeke, J. D., 'Human L1 element target-primed reverse transcription in vitro', *The EMBO Journal*, 21, (21), 5899–5910, 2002.
[61] Houck, C. M., Rinehart, F. P., and Schmid, C. W., 'A ubiquitous family of repeated DNA sequences in the human genome', *Journal of Molecular Biology*, 132, (3), 289–306, 1979.
[62] Deininger, P. L. and Daniels, G. R., 'The recent evolution of mammalian repetitive DNA elements', *Trends in Genetics*, 2, 76–80, 1986.

[63] Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J., and Schmitz, J., 'Evolutionary history of 7SL RNA-derived SINEs in Supraprimates', *Trends in Genetics*, 23, (4), 158–161, 2007.
[64] Weiner, A. M., 'An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome', *Cell*, 22, (1), 209–218, 1980.
[65] Ullu, E. and Tschudi, C., 'Alu sequences are processed 7SL RNA genes', *Nature*, 312, (5990), 171–172, 1984.
[66] Quentin, Y., 'Fusion of a free left Alu monomer and a free right Alu monometer at the origin of the Alu family in the primate genomes', *Nucleic Acids Res*, 20, (3), 487–493, 1992.
[67] Quentin, Y., 'Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements', *Nucleic Acids Res*, 20, (13), 3397–3401, 1992.
[68] Dewannieux, M., Esnault, C., and Heidmann, T., 'LINE-mediated retrotransposition of marked Alu sequences', *Nature Genetics*, 35, (1), 41–48, 2003.
[69] Wallace, N., Wagstaff, B. J., Deininger, P. L., and Roy-Engel, A. M., 'LINE-1 ORF1 protein enhances Alu SINE retrotransposition', *Gene*, 419, (1), 1–6, 2008.
[70] Medstrand, P., Lagemaat, L. N. van de, and Mager, D. L., 'Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes', *Genome Res.*, 12, (10), 1483–1495, 2002.
[71] Paolella, G., Lucero, M. A., Murphy, M. H., and Baralle, F. E., 'The Alu family repeat promoter has a tRNA-like bipartite structure', *EMBO J.*, 2, (5), 691–696, 1983.
[72] Chu, W. M., Liu, W. M., and Schmid, C. W., 'RNA polymerase III promoter and terminator elements affect Alu RNA expression', *Nucleic Acids Res*, 23, (10), 1750–1757, 1995.
[73] Batzer, M. A. and Deininger, P. L., 'Alu repeats and human genomic diversity', *Nature Reviews Genetics*, 3, (5), 370–379, 2002.
[74] Batzer, M. A. *et al.*, 'Dispersion and Insertion Polymorphism in Two Small Subfamilies of Recently Amplified HumanAluRepeats', *Journal of Molecular Biology*, 247, (3), 418–427, 1995.
[75] Deininger, P. L., Batzer, M. A., Hutchison, C. A., and Edgell, M. H., 'Master genes in mammalian repetitive DNA amplification', *Trends in Genetics*, 8, (9), 307–311, 1992.
[76] Shen, M. R., Batzer, M. A., and Deininger, P. L., 'Evolution of the master Alu gene(s)', *J Mol Evol*, 33, (4), 311–320, 1991.
[77] Hormozdiari, F. *et al.*, 'Alu repeat discovery and characterization within human genomes', *Genome Res.*, 21, (6), 840–849, 2011.
[78] Carroll, M. L. *et al.*, 'Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity1IEdited by J. Karn', *Journal of Molecular Biology*, 311, (1), 17–40, 2001.
[79] Ade, C., Roy-Engel, A. M., and Deininger, P. L., 'Alu elements: an intrinsic source of human genome instability', *Current Opinion in Virology*, 3, (6), 639–645, 2013.
[80] Oler, A. J., Traina-Dorge, S., Derbes, R. S., Canella, D., Cairns, B. R., and Roy-Engel, A. M., 'Alu expression in human cell lines and their retrotranspositional potential', *Mobile DNA*, 3, 11, 2012.
[81] Wang, H. *et al.*, 'SVA Elements: A Hominid-specific Retroposon Family', *Journal of Molecular Biology*, 354, (4), 994–1007, 2005.
[82] Hancks, D. C. and Kazazian, H. H., 'SVA retrotransposons: Evolution and genetic instability', *Seminars in Cancer Biology*, 20, (4), 234–245, 2010.
[83] Raiz, J. *et al.*, 'The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery', *Nucl. Acids Res.*, 40, (4), 1666–1683, 2012.

[84] Hancks, D. C., Goodier, J. L., Mandal, P. K., Cheung, L. E., and Kazazian, H. H., 'Retrotransposition of marked SVA elements by human L1s in cultured cells', *Hum Mol Genet*, 20, (17), 3386–3400, 2011.
[85] Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S., and Devine, S. E., 'Natural Genetic Variation Caused by Transposable Elements in Humans', *Genetics*, 168, (2), 933–951, 2004.
[86] Esnault, C., Maestre, J., and Heidmann, T., 'Human LINE retrotransposons generate processed pseudogenes', *Nature Genetics*, 24, (4), 363–367, 2000.
[87] Zhang, Z., Harrison, P. M., Liu, Y., and Gerstein, M., 'Millions of Years of Evolution Preserved: A Comprehensive Catalog of the Processed Pseudogenes in the Human Genome', *Genome Res.*, 13, (12), 2541–2558, 2003.
[88] Torrents, D., Suyama, M., Zdobnov, E., and Bork, P., 'A Genome-Wide Survey of Human Pseudogenes', *Genome Res.*, 13, (12), 2559–2567, 2003.
[89] Zhang, Z., Harrison, P., and Gerstein, M., 'Identification and Analysis of Over 2000 Ribosomal Protein Pseudogenes in the Human Genome', *Genome Res.*, 12, (10), 1466–1482, 2002.
[90] Ewing, A. D. *et al.*, 'Retrotransposition of gene transcripts leads to structural variation in mammalian genomes', *Genome Biology*, 14, R22, 2013.
[91] Schrider, D. R. *et al.*, 'Gene Copy-Number Polymorphism Caused by Retrotransposition in Humans', *PLOS Genetics*, 9, (1), e1003242, 2013.
[92] Richardson, S. R., Salvador-Palomeque, C., and Faulkner, G. J., 'Diversity through duplication: Whole-genome sequencing reveals novel gene retrocopies in the human population', *BioEssays*, 36, (5), 475–481, 2014.
[93] Sayah, D. M., Sokolskaja, E., Berthoux, L., and Luban, J., 'Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1', *Nature*, 430, (6999), 569–573, 2004.
[94] Pyle, A. M., 'Group II Intron Self-Splicing', *Annu. Rev. Biophys.*, 45, (1), 183–205, 2016.
[95] Zimmerly, S. and Semper, C., 'Evolution of group II introns', *Mobile DNA*, 6, 7, 2015.
[96] Lambowitz, A. M. and Belfort, M., 'Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution', *Microbiology Spectrum*, 3, (1), 2015.
[97] Doolittle, W. F., 'The trouble with (group II) introns', *PNAS*, 111, (18), 6536–6537, 2014.
[98] Lambowitz, A. M. and Zimmerly, S., 'Group II Introns: Mobile Ribozymes that Invade DNA', *Cold Spring Harb Perspect Biol*, 3, (8), a003616, 2011.
[99] Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M., and Nienhuis, A. W., 'A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene', *Nucleic Acids Res*, 8, (24), 6113–6128, 1980.
[100] Brouha, B. *et al.*, 'Hot L1s account for the bulk of retrotransposition in the human population', *PNAS*, 100, (9), 5280–5285, 2003.
[101] Swergold, G. D., 'Identification, characterization, and cell specificity of a human LINE-1 promoter.', *Molecular and Cellular Biology*, 10, (12), 6718–6729, 1990.
[102] Speek, M., 'Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes', *Molecular and Cellular Biology*, 21, (6), 1973–1985, 2001.
[103] Yang, N. and Jr, H. H. K., 'L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells', *Nature Structural & Molecular Biology*, 13, (9), 763–771, 2006.
[104] Macia, A. *et al.*, 'Epigenetic Control of Retrotransposon Expression in Human Embryonic Stem Cells', *Mol. Cell. Biol.*, 31, (2), 300–316, 2011.
[105] Denli, A. M. *et al.*, 'Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity', *Cell*, 163, (3), 583–593, 2015.

[106] Becker, K. G., Swergold, G., Ozato, K., and Thayer, R. E., 'Binding of the ubiquitous nuclear transcription factor YY1 to a *cis* regulatory sequence in the human LINE-1 transposable element', *Human Molecular Genetics*, 2, (10), 1697–1702, 1993.
[107] Athanikar, J. N., Badge, R. M., and Moran, J. V., 'A YY1-binding site is required for accurate human LINE-1 transcription initiation', *Nucleic Acids Res*, 32, (13), 3846–3855, 2004.
[108] Yang, N., Zhang, L., Zhang, Y., Jr, K., and H, H., 'An important role for RUNX3 in human L1 transcription and retrotransposition', *Nucleic Acids Res*, 31, (16), 4929–4940, 2003.
[109] Tchénio, T., Casella, J.-F., and Heidmann, T., 'Members of the SRY family regulate the human LINE retrotransposons', *Nucleic Acids Res*, 28, (2), 411–415, 2000.
[110] Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., and Gage, F. H., 'Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition', *Nature*, 435, (7044), 903–910, 2005.
[111] Yu, F., Zingler, N., Schumann, G., and Strätling, W. H., 'Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription', *Nucleic Acids Res*, 29, (21), 4493–4501, 2001.
[112] Holmes, S. E., Singer, M. F., and Swergold, G. D., 'Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element.', *J. Biol. Chem.*, 267, (28), 19765–19768, 1992.
[113] Taylor, M. S. *et al.*, 'Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition', *Cell*, 155, (5), 1034–1048, 2013.
[114] Martin, S. L. *et al.*, 'A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity', *Nucleic Acids Res*, 36, (18), 5845–5854, 2008.
[115] Martin, S. L. *et al.*, 'LINE-1 Retrotransposition Requires the Nucleic Acid Chaperone Activity of the ORF1 Protein', *Journal of Molecular Biology*, 348, (3), 549–561, 2005.
[116] Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., and Kazazian Jr., H. H., 'High Frequency Retrotransposition in Cultured Mammalian Cells', *Cell*, 87, (5), 917–927, 1996.
[117] Khazina, E. and Weichenrieder, O., 'Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame', *PNAS*, 106, (3), 731–736, 2009.
[118] Khazina, E., Truffault, V., Büttner, R., Schmidt, S., Coles, M., and Weichenrieder, O., 'Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition', *Nat Struct Mol Biol*, 18, (9), 1006–1014, 2011.
[119] Martin, S. L., Branciforte, D., Keller, D., and Bain, D. L., 'Trimeric structure for an essential protein in L1 retrotransposition', *PNAS*, 100, (24), 13815–13820, 2003.
[120] Kolosha, V. O. and Martin, S. L., 'High-affinity, Non-sequence-specific RNA Binding by the Open Reading Frame 1 (ORF1) Protein from Long Interspersed Nuclear Element 1 (LINE-1)', *J. Biol. Chem.*, 278, (10), 8112–8117, 2003.
[121] Martin, S. L., 'Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells.', *Molecular and Cellular Biology*, 11, (9), 4804–4807, 1991.
[122] Hohjoh, H. and Singer, M. F., 'Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA.', *The EMBO Journal*, 15, (3), 630–639, 1996.
[123] Kolosha, V. O. and Martin, S. L., 'In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition', *PNAS*, 94, (19), 10155–10160, 1997.
[124] Doucet, A. J. *et al.*, 'Characterization of LINE-1 Ribonucleoprotein Particles', *PLOS Genetics*, 6, (10), e1001150, 2010.

[125] Kulpa, D. A. and Moran, J. V., 'Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition', *Hum. Mol. Genet.*, 14, (21), 3237–3248, 2005.
[126] Feng, Q., Moran, J. V., Kazazian Jr., H. H., and Boeke, J. D., 'Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition', *Cell*, 87, (5), 905–916, 1996.
[127] Mathias, S. L., Scott, A. F., Kazazian, H. H., Boeke, J. D., and Gabriel, A., 'Reverse transcriptase encoded by a human transposable element', *Science*, 254, (5039), 1808–1810, 1991.
[128] Weichenrieder, O., Repanas, K., and Perrakis, A., 'Crystal Structure of the Targeting Endonuclease of the Human LINE-1 Retrotransposon', *Structure*, 12, (6), 975–986, 2004.
[129] Jurka, J., 'Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons', *PNAS*, 94, (5), 1872–1877, 1997.
[130] Cost, G. J. and Boeke, J. D., 'Targeting of Human Retrotransposon Integration Is Directed by the Specificity of the L1 Endonuclease for Regions of Unusual DNA Structure†', *Biochemistry*, 37, (51), 18081–18093, 1998.
[131] Xiong, Y. and Eickbush, T. H., 'Origin and evolution of retroelements based upon their reverse transcriptase sequences.', *The EMBO Journal*, 9, (10), 3353–3362, 1990.
[132] Piskareva, O., Denmukhametova, S., and Schmatchenko, V., 'Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells', *Protein Expression and Purification*, 28, (1), 125–130, 2003.
[133] Herschhorn, A. and Hizi, A., 'Retroviral reverse transcriptases', *Cell. Mol. Life Sci.*, 67, (16), 2717–2747, 2010.
[134] Menéndez-Arias, L., Sebastián-Martín, A., and Álvarez, M., 'Viral reverse transcriptases', *Virus Research*, 234, 153–176, 2017.
[135] Stamos, J. L., Lentzsch, A. M., and Lambowitz, A. M., 'Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications', *Molecular Cell*, 68, (5), 926-939.e4, 2017.
[136] Bolton, E. C., Mildvan, A. S., and Boeke, J. D., 'Inhibition of Reverse Transcription In Vivo by Elevated Manganese Ion Concentration', *Molecular Cell*, 9, (4), 879–889, 2002.
[137] Kirshenboim, N., Hayouka, Z., Friedler, A., and Hizi, A., 'Expression and characterization of a novel reverse transcriptase of the LTR retrotransposon Tf1', *Virology*, 366, (2), 263–276, 2007.
[138] Vashishtha, A. K., Wang, J., and Konigsberg, W. H., 'Different Divalent Cations Alter the Kinetics and Fidelity of DNA Polymerases', *J. Biol. Chem.*, 291, (40), 20869–20875, 2016.
[139] Kulpa, D. A. and Moran, J. V., '*Cis*-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles', *Nature Structural & Molecular Biology*, 13, (7), 655–660, 2006.
[140] Fanning, T. and Singer, M., 'The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins', *Nucleic Acids Res*, 15, (5), 2251–2260, 1987.
[141] Piskareva, O., Ernst, C., Higgins, N., and Schmatchenko, V., 'The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding', *FEBS Open Bio*, 3, (1), 433–437, 2013.
[142] Usdin, K. and Furano, A. V., 'The structure of the guanine-rich polypurine:polypyrimidine sequence at the right end of the rat L1 (LINE) element.', *J. Biol. Chem.*, 264, (26), 15681–15687, 1989.
[143] Moran, J. V., DeBerardinis, R. J., and Kazazian, H. H., 'Exon Shuffling by L1 Retrotransposition', *Science*, 283, (5407), 1530–1534, 1999.

[144] Lindtner, S., Felber, B. K., and Kjems, J., 'An element in the 3' untranslated region of human LINE-1 retrotransposon mRNA binds NXF1(TAP) and can function as a nuclear export element.', *RNA*, 8, (3), 345–356, 2002.
[145] Alisch, R. S., Garcia-Perez, J. L., Muotri, A. R., Gage, F. H., and Moran, J. V., 'Unconventional translation of mammalian LINE-1 retrotransposons', *Genes Dev.*, 20, (2), 210–224, 2006.
[146] Kubo, S. *et al.*, 'L1 retrotransposition in nondividing and primary human somatic cells', *PNAS*, 103, (21), 8036–8041, 2006.
[147] Xie, Y., Mates, L., Ivics, Z., Izsvák, Z., Martin, S. L., and An, W., 'Cell division promotes efficient retrotransposition in a stable L1 reporter cell line', *Mobile DNA*, 4, 10, 2013.
[148] Monot, C. *et al.*, 'The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts', *PLOS Genetics*, 9, (5), e1003499, 2013.
[149] Viollet, S., Monot, C., and Cristofari, G., 'L1 retrotransposition', *Mob Genet Elements*, 4, 2014.
[150] Benitez-Guijarro, M. *et al.*, 'RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition', *The EMBO Journal*, e98506, 2018.
[151] Morrish, T. A. *et al.*, 'DNA repair mediated by endonuclease-independent LINE-1 retrotransposition', *Nature Genetics*, 31, (2), 159–165, 2002.
[152] Sen, S. K., Huang, C. T., Han, K., and Batzer, M. A., 'Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome', *Nucleic Acids Res*, 35, (11), 3741–3751, 2007.
[153] Goodier, J. L., 'Restricting retrotransposons: a review', *Mobile DNA*, 7, 16, 2016.
[154] Ariumi, Y., 'Guardian of the Human Genome: Host Defense Mechanisms against LINE-1 Retrotransposition', *Front. Chem.*, 4, 2016.
[155] Friedli, M. and Trono, D., 'The Developmental Control of Transposable Elements and the Evolution of Higher Species', *Annu. Rev. Cell Dev. Biol.*, 31, (1), 429–451, 2015.
[156] Mita, P. and Boeke, J. D., 'How retrotransposons shape genome regulation', *Current Opinion in Genetics & Development*, 37, 90–100, 2016.
[157] Ostertag, E. M. and Kazazian, H. H., 'Twin Priming: A Proposed Mechanism for the Creation of Inversions in L1 Retrotransposition', *Genome Res.*, 11, (12), 2059–2065, 2001.
[158] Boissinot, S., Entezam, A., and Furano, A. V., 'Selection Against Deleterious LINE-1-Containing Loci in the Human Lineage', *Mol Biol Evol*, 18, (6), 926–935, 2001.
[159] Grimaldi, G., Skowronski, J., and Singer, M. F., 'Defining the beginning and end of KpnI family segments.', *The EMBO Journal*, 3, (8), 1753–1759, 1984.
[160] Belancio, V. P., Roy-Engel, A. M., and Deininger, P., 'The impact of multiple splice sites in human L1 elements', *Gene*, 411, (1), 38–45, 2008.
[161] Perepelitsa-Belancio, V. and Deininger, P., 'RNA truncation by premature polyadenylation attenuates human mobile element activity', *Nature Genetics*, 35, (4), 363–366, 2003.
[162] Suzuki, M. M. and Bird, A., 'DNA methylation landscapes: provocative insights from epigenomics', *Nature Reviews Genetics*, 9, (6), 465–476, 2008.
[163] Cedar, H. and Bergman, Y., 'Programming of DNA Methylation Patterns', *Annu. Rev. Biochem.*, 81, (1), 97–117, 2012.
[164] Hata, K. and Sakaki, Y., 'Identification of critical CpG sites for repression of L1 transcription by DNA methylation', *Gene*, 189, (2), 227–234, 1997.
[165] Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P., and Warren, W. D., 'Asymmetric Methylation in the Hypermethylated CpG Promoter Region of the Human L1 Retrotransposon', *J. Biol. Chem.*, 272, (12), 7810–7816, 1997.
[166] Thayer, R. E., Singer, M. F., and Fanning, T. G., 'Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1 -encoded protein', *Gene*, 133, (2), 273–277, 1993.

[167] Wissing, S. *et al.*, 'Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility', *Hum. Mol. Genet.*, 21, (1), 208–218, 2012.
[168] Kato, Y. *et al.*, 'Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse', *Hum Mol Genet*, 16, (19), 2272–2280, 2007.
[169] Bourc'his, D. and Bestor, T. H., 'Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L', *Nature*, 431, (7004), 96–99, 2004.
[170] Bulut-Karslioglu, A. *et al.*, 'Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells', *Molecular Cell*, 55, (2), 277–290, 2014.
[171] Heras, S. R., Macias, S., Cáceres, J. F., and Garcia-Perez, J. L., 'Control of mammalian retrotransposons by cellular RNA processing activities', *Mobile Genetic Elements*, 4, (2), e28439, 2014.
[172] Siomi, M. C., Sato, K., Pezic, D., and Aravin, A. A., 'PIWI-interacting small RNAs: the vanguard of genome defence', *Nature Reviews Molecular Cell Biology*, 12, (4), 246–258, 2011.
[173] Watanabe, T. *et al.*, 'Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes', *Nature*, 453, (7194), 539–543, 2008.
[174] Gregory, R. I. *et al.*, 'The Microprocessor complex mediates the genesis of microRNAs', *Nature*, 432, (7014), 235–240, 2004.
[175] Kim, V. N., Han, J., and Siomi, M. C., 'Biogenesis of small RNAs in animals', *Nature Reviews Molecular Cell Biology*, 10, (2), 126–139, 2009.
[176] Hamdorf, M. *et al.*, 'miR-128 represses L1 retrotransposition by binding directly to L1 RNA', *Nature Structural & Molecular Biology*, 22, (10), 824–831, 2015.
[177] Heras, S. R. *et al.*, 'The Microprocessor controls the activity of mammalian retrotransposons', *Nat Struct Mol Biol*, 20, (10), 1173–1181, 2013.
[178] Zhang, A., Dong, B., Doucet, A. J., Moldovan, J. B., Moran, J. V., and Silverman, R. H., 'RNase L restricts the mobility of engineered retrotransposons in cultured human cells', *Nucleic Acids Research*, 42, (6), 3803–3820, 2014.
[179] Gregersen, L. H. *et al.*, 'MOV10 Is a 5′ to 3′ RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3′ UTRs', *Molecular Cell*, 54, (4), 573–585, 2014.
[180] Skariah, G. *et al.*, 'Mov10 suppresses retroelements and regulates neuronal development and function in the developing brain', *BMC Biology*, 15, (1), 54, 2017.
[181] Goodier, J. L., Cheung, L. E., and Jr, H. H. K., 'MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells', *PLOS Genetics*, 8, (10), e1002941, 2012.
[182] Li, X. *et al.*, 'The MOV10 Helicase Inhibits LINE-1 Mobility', *J. Biol. Chem.*, 288, (29), 21148–21160, 2013.
[183] Choi, J., Hwang, S.-Y., and Ahn, K., 'Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition', *Nucleic Acids Research*, 2018.
[184] Moldovan, J. B. and Moran, J. V., 'The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition', *PLoS Genet*, 11, (5), e1005121, 2015.
[185] Goodier, J. L., Pereira, G. C., Cheung, L. E., Rose, R. J., and Jr, H. H. K., 'The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition', *PLOS Genetics*, 11, (5), e1005252, 2015.
[186] Kinomoto, M. *et al.*, 'All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition', *Nucleic Acids Res*, 35, (9), 2955–2964, 2007.
[187] MacDuff, D. A., Demorest, Z. L., and Harris, R. S., 'AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity', *Nucleic Acids Res*, 37, (6), 1854–1867, 2009.

[188] Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S., 'AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification', *Nature*, 418, (6893), 99–104, 2002.
[189] Harris, R. S., Petersen-Mahrt, S. K., and Neuberger, M. S., 'RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators', *Molecular Cell*, 10, (5), 1247–1253, 2002.
[190] Ikeda, T. *et al.*, 'Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons', *Nucleic Acids Res*, 39, (13), 5538–5554, 2011.
[191] Richardson, S. R., Narvaiza, I., Planegger, R. A., Weitzman, M. D., and Moran, J. V., 'APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition', *eLife Sciences*, 3, e02008, 2014.
[192] Stenglein, M. D. and Harris, R. S., 'APOBEC3B and APOBEC3F Inhibit L1 Retrotransposition by a DNA Deamination-independent Mechanism', *J. Biol. Chem.*, 281, (25), 16837–16841, 2006.
[193] Feng, Y., Goubran, M. H., Follack, T. B., and Chelico, L., 'Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H', *Scientific Reports*, 7, (1), 10881, 2017.
[194] Horn, A. V. *et al.*, 'Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity', *Nucleic Acids Res*, 42, (1), 396–416, 2014.
[195] Goldstone, D. C. *et al.*, 'HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase', *Nature*, 480, (7377), 379–382, 2011.
[196] Lahouassa, H. *et al.*, 'SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates', *Nature Immunology*, 13, (3), 223–228, 2012.
[197] Zhao, K. *et al.*, 'Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1', *Cell Reports*, 4, (6), 1108–1115, 2013.
[198] White, T. E. *et al.*, 'Effects of human SAMHD1 polymorphisms on HIV-1 susceptibility', *Virology*, 460–461, 34–44, 2014.
[199] Upton, K. R. *et al.*, 'Ubiquitous L1 Mosaicism in Hippocampal Neurons', *Cell*, 161, (2), 228–239, 2015.
[200] Hu, S. *et al.*, 'SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation', *PLOS Genetics*, 11, (7), e1005367, 2015.
[201] Stetson, D. B., Ko, J. S., Heidmann, T., and Medzhitov, R., 'Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity', *Cell*, 134, (4), 587–598, 2008.
[202] Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E., 'Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man', *Nature*, 332, (6160), 164–166, 1988.
[203] Trelogan, S. A. and Martin, S. L., 'Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis.', *PNAS*, 92, (5), 1520–1524, 1995.
[204] Georgiou, I. *et al.*, 'Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes', *Hum Mol Genet*, 18, (7), 1221–1228, 2009.
[205] Lucifero, D., La Salle, S., Bourc'his, D., Martel, J., Bestor, T. H., and Trasler, J. M., 'Coordinate regulation of DNA methyltransferase expression during oogenesis', *BMC Developmental Biology*, 7, (1), 36, 2007.
[206] Branciforte, D. and Martin, S. L., 'Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition.', *Molecular and Cellular Biology*, 14, (4), 2584–2592, 1994.

[207] Freeman, P., Macfarlane, C., Collier, P., Jeffreys, A. J., and Badge, R. M., 'L1 hybridization enrichment: a method for directly accessing de novo L1 insertions in the human germline', *Human Mutation*, 32, (8), 978–988, 2011.
[208] Ostertag, E. M. *et al.*, 'A mouse model of human L1 retrotransposition', *Nature Genetics*, 32, (4), 655–660, 2002.
[209] Packer, A. I., Manova, K., and Bachvarova, R. F., 'A Discrete LINE-1 Transcript in Mouse Blastocysts', *Developmental Biology*, 157, (1), 281–283, 1993.
[210] Walter, M., Teissandier, A., Pérez-Palacios, R., and Bourc'his, D., 'An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells', *eLife*, 5, e11418, 2016.
[211] Garcia-Perez, J. L. *et al.*, 'LINE-1 retrotransposition in human embryonic stem cells', *Hum. Mol. Genet.*, 16, (13), 1569–1577, 2007.
[212] Hurk, J. A. J. M. van den *et al.*, 'L1 retrotransposition can occur early in human embryonic development', *Hum. Mol. Genet.*, 16, (13), 1587–1592, 2007.
[213] Friedli, M. *et al.*, 'Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency', *Genome Res.*, 24, (8), 1251–1259, 2014.
[214] Klawitter, S. *et al.*, 'Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells', *Nature Communications*, 7, 10286, 2016.
[215] Lu, X. *et al.*, 'The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity', *Nature Structural & Molecular Biology*, 21, (4), 423–425, 2014.
[216] Wang, J. *et al.*, 'Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells', *Nature*, 516, (7531), 405–409, 2014.
[217] Kano, H. *et al.*, 'L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism', *Genes Dev.*, 23, (11), 1303–1312, 2009.
[218] Coufal, N. G. *et al.*, 'L1 retrotransposition in human neural progenitor cells', *Nature*, 460, (7259), 1127–1131, 2009.
[219] Evrony, G. D. *et al.*, 'Single-Neuron Sequencing Analysis of L1 Retrotransposition and Somatic Mutation in the Human Brain', *Cell*, 151, (3), 483–496, 2012.
[220] Evrony, G. D. *et al.*, 'Cell Lineage Analysis in Human Brain Using Endogenous Retroelements', *Neuron*, 85, (1), 49–59, 2015.
[221] Baillie, J. K. *et al.*, 'Somatic retrotransposition alters the genetic landscape of the human brain', *Nature*, 479, (7374), 534–537, 2011.
[222] Philippe, C. *et al.*, 'Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci', *eLife*, 5, e13926, 2016.
[223] Clayton, E. A., Wang, L., Rishishwar, L., Wang, J., McDonald, J. F., and Jordan, I. K., 'Patterns of Transposable Element Expression and Insertion in Cancer', *Front. Mol. Biosci.*, 3, 2016.
[224] Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R., and Deininger, P., 'Somatic expression of LINE-1 elements in human tissues', *Nucleic Acids Res*, 38, (12), 3909–3922, 2010.
[225] Scott, E. C., Gardner, E. J., Masood, A., Chuang, N. T., Vertino, P. M., and Devine, S. E., 'A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer', *Genome Res.*, 2016.
[226] Miki, Y. *et al.*, 'Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer', *Cancer Res*, 52, (3), 643–645, 1992.
[227] Alves, G., Tatro, A., and Fanning, T., 'Differential methylation of human LINE-1 retrotransposons in malignant cells', *Gene*, 176, (1), 39–44, 1996.
[228] Kazazian Jr, H. H., 'An estimated frequency of endogenous insertional mutations in humans', *Nature Genetics*, 22, (2), 130, 1999.

[229] Xing, J. *et al.*, 'Mobile elements create structural variation: Analysis of a complete human genome', *Genome Res.*, 19, (9), 1516–1526, 2009.
[230] Cordaux, R., Hedges, D. J., Herke, S. W., and Batzer, M. A., 'Estimating the retrotransposition rate of human Alu elements', *Gene*, 373, 134–137, 2006.
[231] Callinan, P. A. and Batzer, M. A., 'Retrotransposable Elements and Human Disease', *Genome and Disease*, 1, 104–115, 2006.
[232] Burns, K. H., 'Transposable elements in cancer', *Nature Reviews Cancer*, 17, (7), 415–424, 2017.
[233] Hancks, D. C. and Kazazian, H. H., 'Roles for retrotransposon insertions in human disease', *Mobile DNA*, 7, 9, 2016.
[234] Garcia-Perez, J. L., Widmann, T. J., and Adams, I. R., 'The impact of transposable elements on mammalian development', *Development*, 143, (22), 4101–4114, 2016.
[235] Richardson, S. R., Doucet, A. J., Kopera, H. C., Moldovan, J. B., Garcia-Perez, J. L., and Moran, J. V., 'The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes', *Microbiology Spectrum*, 3, (2), 2015.
[236] Kaer, K. and Speek, M., 'Retroelements in human disease', *Gene*, 518, (2), 231–241, 2013.
[237] Beck, C. R., Garcia-Perez, J. L., Badge, R. M., and Moran, J. V., 'LINE-1 Elements in Structural Variation and Disease', *Annual Review of Genomics and Human Genetics*, 12, (1), 187–215, 2011.
[238] Goodier, J. L. and Kazazian Jr., H. H., 'Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites', *Cell*, 135, (1), 23–35, 2008.
[239] Miné, M. *et al.*, 'A large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element', *Human Mutation*, 28, (2), 137–142, 2007.
[240] Takasu, M. *et al.*, 'Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon', *Tissue Antigens*, 70, (2), 144–150, 2007.
[241] Gasior, S. L., Wakeman, T. P., Xu, B., and Deininger, P. L., 'The Human LINE-1 Retrotransposon Creates DNA Double-strand Breaks', *Journal of Molecular Biology*, 357, (5), 1383–1393, 2006.
[242] Flemr, M. *et al.*, 'A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes', *Cell*, 155, (4), 807–816, 2013.
[243] Xing, J., Wang, H., Belancio, V. P., Cordaux, R., Deininger, P. L., and Batzer, M. A., 'Emergence of primate genes by retrotransposon-mediated sequence transduction', *PNAS*, 103, (47), 17608–17613, 2006.
[244] Scott, E. C. and Devine, S. E., 'The Role of Somatic L1 Retrotransposition in Human Cancers', *Viruses*, 9, (6), 131, 2017.
[245] Hancks, D. C. and Kazazian Jr., H. H., 'Active human retrotransposons: variation and disease', *Current Opinion in Genetics & Development*, 22, (3), 191–203, 2012.
[246] Shukla, R. *et al.*, 'Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma', *Cell*, 153, (1), 101–111, 2013.
[247] Helman, E., Lawrence, M. S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M., 'Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing', *Genome Res.*, 24, (7), 1053–1063, 2014.
[248] Belancio, V. P., Hedges, D. J., and Deininger, P., 'Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health', *Genome Res.*, 18, (3), 343–358, 2008.
[249] Van Meter, M. *et al.*, 'SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age', *Nature Communications*, 5, 5011, 2014.
[250] Cecco, M. D. *et al.*, 'L1 drives IFN in senescent cells and promotes age-associated inflammation', *Nature*, 566, (7742), 73, 2019.

[251] Simon, M. *et al.*, 'LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation', *Cell Metabolism*, 0, (0), 2019.
[252] Tubio, J. M. C. *et al.*, 'Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes', *Science*, 345, (6196), 1251343, 2014.
[253] Iskow, R. C. *et al.*, 'Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons', *Cell*, 141, (7), 1253–1261, 2010.
[254] Solyom, S. *et al.*, 'Extensive somatic L1 retrotransposition in colorectal tumors', *Genome Res.*, 22, (12), 2328–2338, 2012.
[255] Lee, E. *et al.*, 'Landscape of Somatic Retrotransposition in Human Cancers', *Science*, 337, (6097), 967–971, 2012.
[256] Pitkänen, E. *et al.*, 'Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer', *Oncotarget*, 5, (3), 853–859, 2014.
[257] Paterson, A. L. *et al.*, 'Mobile element insertions are frequent in oesophageal adenocarcinomas and can mislead paired-end sequencing analysis', *BMC Genomics*, 16, (1), 473, 2015.
[258] Rodić, N. *et al.*, 'Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma', *Nature Medicine*, 21, (9), 1060–1064, 2015.
[259] Ewing, A. D. *et al.*, 'Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution', *Genome Res.*, 25, (10), 1536–1545, 2015.
[260] Doucet-O'Hare, T. T., Sharma, R., Rodić, N., Anders, R. A., Burns, K. H., and Kazazian, H. H., 'Somatically Acquired LINE-1 Insertions in Normal Esophagus Undergo Clonal Expansion in Esophageal Squamous Cell Carcinoma', *Human Mutation*, 37, (9), 942–954, 2016.
[261] Wolff, E. M. *et al.*, 'Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer', *PLOS Genetics*, 6, (4), e1000917, 2010.
[262] Daskalos, A. *et al.*, 'Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer', *International Journal of Cancer*, 124, (1), 81–87, 2009.
[263] Saito, K., Kawakami, K., Matsumoto, I., Oda, M., Watanabe, G., and Minamoto, T., 'Long Interspersed Nuclear Element 1 Hypomethylation Is a Marker of Poor Prognosis in Stage IA Non–Small Cell Lung Cancer', *Clin Cancer Res*, 16, (8), 2418–2426, 2010.
[264] Ikeda, K. *et al.*, 'Long Interspersed Nucleotide Element 1 Hypomethylation Is Associated With Poor Prognosis of Lung Adenocarcinoma', *The Annals of Thoracic Surgery*, 96, (5), 1790–1794, 2013.
[265] Iwagami, S. *et al.*, 'LINE-1 Hypomethylation Is Associated With a Poor Prognosis Among Patients With Curatively Resected Esophageal Squamous Cell Carcinoma':, *Annals of Surgery*, 257, (3), 449–455, 2013.
[266] Kawano, H. *et al.*, 'Chromosomal Instability Associated with Global DNA Hypomethylation is Associated with the Initiation and Progression of Esophageal Squamous Cell Carcinoma', *Ann Surg Oncol*, 21, (4), 696–702, 2014.
[267] Ogino, S. *et al.*, 'A Cohort Study of Tumoral LINE-1 Hypomethylation and Prognosis in Colon Cancer', *J Natl Cancer Inst*, 100, (23), 1734–1738, 2008.
[268] Inamura, K. *et al.*, 'Tumor LINE-1 Methylation Level and Microsatellite Instability in Relation to Colorectal Cancer Prognosis', *J Natl Cancer Inst*, 106, (9), 2014.
[269] Mima, K. *et al.*, 'Tumor LINE-1 methylation level and colorectal cancer location in relation to patient survival', *Oncotarget*, 7, (34), 55098–55109, 2016.
[270] van Hoesel, A. Q. *et al.*, 'Hypomethylation of LINE-1 in primary tumor has poor prognosis in young breast cancer patients: a retrospective cohort study', *Breast Cancer Res Treat*, 134, (3), 1103–1114, 2012.

[271] Harris, C. R. *et al.*, 'Association of Nuclear Localization of a Long Interspersed Nuclear Element-1 Protein in Breast Tumors with Poor Prognostic Outcomes', *Genes & Cancer*, 1, (2), 115–124, 2010.
[272] Harada, K. *et al.*, 'LINE-1 Methylation Level and Patient Prognosis in a Database of 208 Hepatocellular Carcinomas', *Ann Surg Oncol*, 22, (4), 1280–1287, 2015.
[273] Zhu, C. *et al.*, 'Hypomethylation of Long Interspersed Nuclear Element-1 (LINE-1) is Associated with Poor Prognosis via Activation of c-MET in Hepatocellular Carcinoma', *Ann Surg Oncol*, 21, (4), 729–735, 2014.
[274] Gao, X. *et al.*, 'Hypomethylation of long interspersed nuclear element-1 promoter is associated with poor outcomes for curative resected hepatocellular carcinoma', *Liver International*, 34, (1), 136–146, 2014.
[275] Jeong, S. *et al.*, 'Tumoral LINE-1 hypomethylation is associated with poor survival of patients with intrahepatic cholangiocarcinoma', *BMC Cancer*, 17, (1), 588, 2017.
[276] Furlan, C. *et al.*, 'Prognostic significance of LINE-1 hypomethylation in oropharyngeal squamous cell carcinoma', *Clinical Epigenetics*, 9, 58, 2017.
[277] Pattamadilok, J. *et al.*, 'LINE-1 hypomethylation level as a potential prognostic factor for epithelial ovarian cancer', *International Journal of Gynecological Cancer*, 18, (4), 711–717, 2008.
[278] Chen, L., Dahlstrom, J. E., Chandra, A., Board, P., and Rangasamy, D., 'Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer', *Breast Cancer Res Treat*, 136, (1), 129–142, 2012.
[279] Baba, Y. *et al.*, 'LINE-1 Hypomethylation, DNA Copy Number Alterations, and CDK6 Amplification in Esophageal Squamous Cell Carcinoma', *Clin Cancer Res*, 20, (5), 1114–1124, 2014.
[280] Hur, K. *et al.*, 'Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis.', *Gut*, 63, (4), 635–646, 2014.
[281] Li, M. Y. *et al.*, 'Long interspersed nucleotide acid element-1 ORF-1 protein promotes proliferation and invasion of human colorectal cancer LoVo cells through enhancing ETS-1 activity', *Genetics and Molecular Research*, 13, (3), 6981–6994, 2014.
[282] Yang, Q. *et al.*, 'LINE-1 ORF-1p functions as a novel HGF/ETS-1 signaling pathway co-activator and promotes the growth of MDA-MB-231 cell', *Cellular Signalling*, 25, (12), 2652–2660, 2013.
[283] Balaj, L. *et al.*, 'Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences', *Nature Communications*, 2, 180, 2011.
[284] Cruickshanks, H. A. *et al.*, 'Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer', *Nucleic Acids Res*, 41, (14), 6857–6869, 2013.
[285] Freeman, J. D., Goodchild, N. L., and Mager, D. L., 'A modified indicator gene for selection of retrotransposition events in mammalian cells', *BioTechniques*, 17, (1), 46, 48–49, 52, 1994.
[286] Xie, Y., Rosser, J. M., Thompson, T. L., Boeke, J. D., and An, W., 'Characterization of L1 retrotransposition with high-throughput dual-luciferase assays', *Nucl. Acids Res.*, 39, (3), e16–e16, 2011.
[287] Goodier, J. L., Zhang, L., Vetter, M. R., and Kazazian, H. H., 'LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced Silencing Complex', *Mol. Cell. Biol.*, 27, (18), 6469–6483, 2007.
[288] Ostertag, E. M. *et al.*, 'Determination of L1 retrotransposition kinetics in cultured cells', *Nucleic Acids Res*, 28, (6), 1418–1423, 2000.

[289] Mangiacasale, R. *et al.*, 'Exposure of normal and transformed cells to nevirapine, a reverse transcriptase inhibitor, reduces cell growth and promotes differentiation', *Oncogene*, 22, (18), 2750, 2003.
[290] Sciamanna, I. *et al.*, 'Inhibition of endogenous reverse transcriptase antagonizes human tumor growth', *Oncogene*, 24, (24), 3923–3931, 2005.
[291] Jones, R. B., Garrison, K. E., Wong, J. C., Duan, E. H., Nixon, D. F., and Ostrowski, M. A., 'Nucleoside Analogue Reverse Transcriptase Inhibitors Differentially Inhibit Human LINE-1 Retrotransposition', *PLoS ONE*, 3, (2), 2008.
[292] Carlini, F. *et al.*, 'The Reverse Transcription Inhibitor Abacavir Shows Anticancer Activity in Prostate Cancer Cell Lines', *PLOS ONE*, 5, (12), e14221, 2010.
[293] Dai, L., Huang, Q., and Boeke, J. D., 'Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition', *BMC Biochemistry*, 12, 18, 2011.
[294] Rice, G. I. *et al.*, 'Reverse-Transcriptase Inhibitors in the Aicardi–Goutières Syndrome', *New England Journal of Medicine*, 379, (23), 2275–2277, 2018.
[295] Faria, N. R. *et al.*, 'The early spread and epidemic ignition of HIV-1 in human populations', *Science*, 346, (6205), 56–61, 2014.
[296] Gallo, R. C. *et al.*, 'Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS)', *Science*, 220, (4599), 865–867, 1983.
[297] Barre-Sinoussi, F. *et al.*, 'Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)', *Science*, 220, (4599), 868–871, 1983.
[298] Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C., 'Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS', *Science*, 224, (4648), 497–500, 1984.
[299] Gallo, R. C. *et al.*, 'Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS', *Science*, 224, (4648), 500–503, 1984.
[300] Sarngadharan, M. G., Popovic, M., Bruch, L., Schupbach, J., and Gallo, R. C., 'Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS', *Science*, 224, (4648), 506–508, 1984.
[301] Schupbach, J., Popovic, M., Gilden, R. V., Gonda, M. A., Sarngadharan, M. G., and Gallo, R. C., 'Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS', *Science*, 224, (4648), 503–505, 1984.
[302] Horwitz, J. P., Chua, J., Noel, M., and Donatti, J. T., 'Nucleosides. XI. 2',3'-Dideoxycytidine', *J. Org. Chem.*, 32, (3), 817–818, 1967.
[303] 'WHO | HIV/AIDS', *WHO*. [Online]. Available: http://www.who.int/gho/hiv/en/. [Accessed: 17-Jan-2019].
[304] Clercq, E. D. and Li, G., 'Approved Antiviral Drugs over the Past 50 Years', *Clinical Microbiology Reviews*, 29, (3), 695–747, 2016.
[305] Tantillo, C. *et al.*, 'Locations of Anti-AIDS Drug Binding Sites and Resistance Mutations in the Three-dimensional Structure of HIV-1 Reverse Transcriptase: Implications for Mechanisms of Drug Inhibition and Resistance', *Journal of Molecular Biology*, 243, (3), 369–387, 1994.
[306] Ding, J. *et al.*, 'Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors', *Nature Structural & Molecular Biology*, 2, (5), 407–415, 1995.
[307] Das, K. *et al.*, 'Crystal Structures of 8-Cl and 9-Cl TIBO Complexed with Wild-type HIV-1 RT and 8-Cl TIBO Complexed with the Tyr181Cys HIV-1 RT Drug-resistant Mutant', *Journal of Molecular Biology*, 264, (5), 1085–1100, 1996.

[308] Tuaillon, E. *et al.*, 'Phenotypic Susceptibility to Nonnucleoside Inhibitors of Virion-Associated Reverse Transcriptase From Different HIV Types and Groups', *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 37, (5), 1543, 2004.
[309] Ren, J., Bird, L. E., Chamberlain, P. P., Stewart-Jones, G. B., Stuart, D. I., and Stammers, D. K., 'Structure of HIV-2 reverse transcriptase at 2.35-Å resolution and the mechanism of resistance to non-nucleoside inhibitors', *PNAS*, 99, (22), 14410–14415, 2002.
[310] Cano-Soldado, P. and Pastor-Anglada, M., 'Transporters that translocate nucleosides and structural similar drugs: structural requirements for substrate recognition', *Medicinal Research Reviews*, 32, (2), 428–457, 2012.
[311] Jordheim, L. P., Durantel, D., Zoulim, F., and Dumontet, C., 'Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases', *Nature Reviews Drug Discovery*, 12, (6), 447–464, 2013.
[312] Balzarini, J., Herdewijn, P., and Clercq, E. D., '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.*, 264, (11), 6127–6133, 1989.
[313] Larder, B. A., Purifoy, D. J. M., Powell, K. L., and Darby, G., 'Site-specific mutagenesis of AIDS virus reverse transcriptase', *Nature*, 327, (6124), 716, 1987.
[314] Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C., 'Structure of a Covalently Trapped Catalytic Complex of HIV-1 Reverse Transcriptase: Implications for Drug Resistance', *Science*, 282, (5394), 1669–1675, 1998.
[315] Harris, D., Kaushik, N., Pandey, P. K., Yadav, P. N. S., and Pandey, V. N., 'Functional Analysis of Amino Acid Residues Constituting the dNTP Binding Pocket of HIV-1 Reverse Transcriptase', *J. Biol. Chem.*, 273, (50), 33624–33634, 1998.
[316] Sarafianos, S. G. *et al.*, 'Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA', *The EMBO Journal*, 21, (23), 6614–6624, 2002.
[317] Das, K., Martinez, S. E., Bauman, J. D., and Arnold, E., 'HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism', *Nature Structural & Molecular Biology*, 19, (2), 253–259, 2012.
[318] Furman, P. A. *et al.*, 'Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase', *PNAS*, 83, (21), 8333–8337, 1986.
[319] Chu, C. K., Schinazi, R. F., Ahn, M. K., Ullas, G. V., and Gu, Z. P., 'Structure-activity relationships of pyrimidine nucleosides as antiviral agents for human immunodeficiency virus type 1 in peripheral blood mononuclear cells', *J. Med. Chem.*, 32, (3), 612–617, 1989.
[320] Fowler, B. J. *et al.*, 'Nucleoside reverse transcriptase inhibitors possess intrinsic anti-inflammatory activity', *Science*, 346, (6212), 1000–1003, 2014.
[321] Beck-Engeser, G. B., Eilat, D., and Wabl, M., 'An autoimmune disease prevented by anti-retroviral drugs', *Retrovirology*, 8, (1), 91, 2011.
[322] Shelton, J., Lu, X., Hollenbaugh, J. A., Cho, J. H., Amblard, F., and Schinazi, R. F., 'Metabolism, Biochemical Actions, and Chemical Synthesis of Anticancer Nucleosides, Nucleotides, and Base Analogs', *Chem. Rev.*, 116, (23), 14379–14455, 2016.
[323] Galmarini, C. M., Mackey, J. R., and Dumontet, C., 'Nucleoside analogues and nucleobases in cancer treatment', *The Lancet Oncology*, 3, (7), 415–424, 2002.
[324] Ward, D. C., 'Chapter 25. Relationship between Nucleoside Conformation and Biological Activity', in *Annual Reports in Medicinal Chemistry*, 5, C. K. Cain, Ed. Academic Press, 1970, 272–284.

[325] Roey, P. V., Salerno, J. M., Chu, C. K., and Schinazi, R. F., 'Correlation between preferred sugar ring conformation and activity of nucleoside analogues against human immunodeficiency virus', *PNAS*, 86, (11), 3929–3933, 1989.
[326] Roey, P. V., Taylor, E. W., Chu, C. K., and Schinazi, R. F., 'Correlation of Molecular Conformation and Activity of Reverse Transcriptase Inhibitorsa', *Annals of the New York Academy of Sciences*, 616, (1), 29–40, 1990.
[327] Dijkstra, S., Benevides, J. M., and Thomas, G. J., 'Solution conformations of nucleoside analogues exhibiting antiviral activity against human immunodeficiency virus', *Journal of Molecular Structure*, 242, 283–301, 1991.
[328] Mansour, T. S. *et al.*, 'Structure-Activity Relationship of Pyrimidine Heterosubstituted Nucleoside Analogues', *Nucleosides and Nucleotides*, 16, (7–9), 993–1001, 1997.
[329] Bowater, R. P. and Gates, A. J., 'Nucleotides: Structure and Properties', in *eLS*, John Wiley & Sons Ltd, Ed. Chichester, UK: John Wiley & Sons, Ltd, 2015, 1–9.
[330] McNaught, A. D., Wilkinson, A., and International Union of Pure and Applied Chemistry, Eds., *Compendium of chemical terminology: IUPAC recommendations*, 2nd ed. Oxford [England] ; Malden, MA, USA: Blackwell Science, 1997.
[331] Favre, H. A. and Powell, W. H., *Nomenclature of Organic Chemistry: IUPAC Recommendations and Preferred Names 2013.* Cambridge: Royal Society of Chemistry, 2013.
[332] Bonnett, R., 'The Chemistry of the Vitamin B12 Group.', *Chem. Rev.*, 63, (6), 573–605, 1963.
[333] Migawa, M. T. *et al.*, 'Design, Synthesis, and Antiviral Activity of α-Nucleosides: d- and l-Isomers of Lyxofuranosyl- and (5-Deoxylyxofuranosyl)benzimidazoles', *J. Med. Chem.*, 41, (8), 1242–1251, 1998.
[334] Post, M. L., Birnbaum, G. I., Huber, C. P., and Shugar, D., 'α-Nucleosides in biological systems: Crystal structure and conformation of α-cytidine', *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 479, (2), 133–142, 1977.
[335] Sundaralingam, M., 'Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides', *Biopolymers*, 7, (6), 821–860, 1969.
[336] Cadet, J., Taïeb, C., Remin, M., Niemczura, W. P., and Hruska, F. E., 'Conformational studies of α- and β-pyrimidine 2′-deoxyribonucleosides in the syn and anti conformation', *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 608, (2), 435–445, 1980.
[337] Sundaralingam, M., 'Stereochemistry of nucleic acids and their constituents. XVIII. Conformational analysis of .alpha. nucleosides by x-ray crystallography', *J. Am. Chem. Soc.*, 93, (24), 6644–6647, 1971.
[338] Viswamitra, M. A., Reddy, B. S., Lin, G. H. Y., and Sundaraligam, M., 'Stereochemistry of nucleic acids and their constituents. XVII. Crystal and molecular structure of deoxycytidine 5'-phosphate monohydrate. Possible puckering for the furanoside ring in B-deoxyribonucleic acid', *Journal of the American Chemical Society*, 93, (18), 4565–4573, 1971.
[339] Sutor, D. J., '204. Evidence for the existence of C–H···O hydrogen bonds in crystals', *J. Chem. Soc.*, 0, (0), 1105–1110, 1963.
[340] Chekhlov, A. N., 'Crystallographic evidence for intramolecular C–H...O interactions in principal nucleosides. Refinement of the crystal structure of thymidine', *J Struct Chem*, 36, (1), 155–161, 1995.
[341] Green, E. A., Rosenstein, R. D., Shiono, R., Abraham, D. J., Trus, B. L., and Marsh, R. E., 'The crystal structure of uridine', *Acta Crystallographica Section B*, 31, (1), 102–107, 1975.

[342] Saenger, W., 'Molecular and crystal structure of an arabino nucleoside. X-ray analysis of 1-.beta.-D-arabinofuranosyl-4-thiouracil monohydrate', *J. Am. Chem. Soc.*, 94, (2), 621–626, 1972.
[343] Seeman, N. C., Sussman, J. L., Berman, H. M., and Kim, S. H., 'Nucleic Acid Conformation: Crystal Structure of a Naturally Occurring Dinucleoside Phosphate (UpA)', *Nature New Biology*, 233, (37), 90–92, 1971.
[344] Palafox, M. A., 'Molecular structure differences between the antiviral Nucleoside Analogue 5-iodo-2′-deoxyuridine and the natural nucleoside 2′-deoxythymidine using MP2 and DFT methods: conformational analysis, crystal simulations, DNA pairs and possible behaviour', *Journal of Biomolecular Structure and Dynamics*, 32, (5), 831–851, 2014.
[345] Palafox, M. A., 'Structure and conformational analysis of the anti-HIV reverse transcriptase inhibitor AZT using MP2 and DFT methods. Differences with the natural nucleoside thymidine. Simulation of the 1st phosphorylation step with ATP', *Phys. Chem. Chem. Phys.*, 16, (45), 24763–24783, 2014.
[346] G. Ponomareva, A., P. Yurenko, Y., O. Zhurakivsky, R., van Mourik, T., and M. Hovorun, D., 'Conformational Landscape of the Nucleoside Reverse Transcriptase Inhibitor d4T: a Comprehensive Quantum-Chemical Approach', *Current Physical Chemistry*, 3, (1), 83–92, 2013.
[347] Ahmed, M., Yu, A., and Wang, F., 'DFT Study on the Conformational and Vibrational Properties of 3'-Deoxycytidine and Its Analogues', *International Journal of Chemistry*, 5, (2), 2013.
[348] G. Ponomareva, A., P. Yurenko, Y., O. Zhurakivsky, R., Mourik, T. van, and M. Hovorun, D., 'Complete conformational space of the potential HIV-1 reverse transcriptase inhibitors d4U and d4C. A quantum chemical study', *Physical Chemistry Chemical Physics*, 14, (19), 6787–6795, 2012.
[349] Yurenko, Y. P., Zhurakivsky, R. O., Samijlenko, S. P., and Hovorun, D. M., 'Intramolecular CH...O Hydrogen Bonds in the AI and BI DNA-like Conformers of Canonical Nucleosides and their Watson-Crick Pairs. Quantum Chemical and AIM Analysis', *Journal of Biomolecular Structure and Dynamics*, 29, (1), 51–65, 2011.
[350] Yurenko, Y. P., Zhurakivsky, R. O., Ghomi, M., Samijlenko, S. P., and Hovorun, D. M., 'Ab Initio Comprehensive Conformational Analysis of 2'-Deoxyuridine, the Biologically Significant DNA Minor Nucleoside, and Reconstruction of Its Low-Temperature Matrix Infrared Spectrum', *J. Phys. Chem. B*, 112, (4), 1240–1250, 2008.
[351] Yurenko, Y. P., Zhurakivsky, R. O., Samijlenko, S. P., Ghomi, M., and Hovorun, D. M., 'The whole of intramolecular H-bonding in the isolated DNA nucleoside thymidine. AIM electron density topological study', *Chemical Physics Letters*, 447, (1), 140–146, 2007.
[352] Yurenko, Y. P., Zhurakivsky, R. O., Ghomi, M., Samijlenko, S. P., and Hovorun, D. M., 'Comprehensive Conformational Analysis of the Nucleoside Analogue 2'-β-Deoxy-6-azacytidine by DFT and MP2 Calculations', *J. Phys. Chem. B*, 111, (22), 6263–6271, 2007.
[353] Hocquet, A., 'Intramolecular hydrogen bonding in 2′-deoxyribonucleosides: an AIM topological study of the electronic density', *Physical Chemistry Chemical Physics*, 3, (15), 3192–3199, 2001.
[354] Fidanza, N. G., Suvire, F. D., Sosa, G. L., Lobayan, R. M., Enriz, R. D., and Peruchena, N. M., 'A search for C–H···O type hydrogen bonds in Lamivudine (3TC). An exploratory conformational and electronic analysis', *Journal of Molecular Structure: THEOCHEM*, 543, (1), 185–193, 2001.
[355] Hocquet, A., Leulliot, N., and Ghomi, M., 'Ground-State Properties of Nucleic Acid Constituents Studied by Density Functional Calculations. 3. Role of Sugar Puckering

and Base Orientation on the Energetics and Geometry of 2'-Deoxyribonucleosides and Ribonucleosides', *J. Phys. Chem. B*, 104, (18), 4560–4568, 2000.
[356] Hocquet, A. and Ghomi, M., 'The peculiar role of cytosine in nucleoside conformational behaviour: Hydrogen bond donor capacity of nucleic bases', *Phys. Chem. Chem. Phys.*, 2, (23), 5351–5353, 2000.
[357] Miljkovic, M., 'Stereoelectronic Effects in Nucleosides and Nucleotides', in *Electrostatic and Stereoelectronic Effects in Carbohydrate Chemistry*, M. Miljkovic, Ed. Boston, MA: Springer US, 2014, 181–196.
[358] Vargason, J. M., Henderson, K., and Ho, P. S., 'A crystallographic map of the transition from B-DNA to A-DNA', *PNAS*, 98, (13), 7265–7270, 2001.
[359] Li, D. *et al.*, 'Tautomerism provides a molecular explanation for the mutagenic properties of the anti-HIV nucleoside 5-aza-5,6-dihydro-2′-deoxycytidine', *PNAS*, 111, (32), E3252–E3259, 2014.
[360] Ichikawa, E. and Kato, K., 'Sugar-Modified Nucleosides in Past 10 Years, A Review', *Current Medicinal Chemistry*, 8, (4), 385–423, 2001.
[361] Singh, S., Bhattarai, D., Veeraswamy, G., Choi, Y., and Lee, K., 'Nucleosides with Modified Sugar Ring: Synthesis and Biological Activities', *Current Organic Chemistry*, 20, (8), 856–897, 2016.
[362] Romeo, G., Chiacchio, U., Corsaro, A., and Merino, P., 'Chemical Synthesis of Heterocyclic–Sugar Nucleoside Analogues', *Chem. Rev.*, 110, (6), 3337–3370, 2010.
[363] De Clercq, E., 'C-Nucleosides To Be Revisited', *J. Med. Chem.*, 59, (6), 2301–2311, 2016.
[364] Ferrero, M., Fernández, S., and Gotor, V., 'Biocatalytic Methodologies for Selective Modified Nucleosides', in *Chemical Synthesis of Nucleoside Analogues*, P. Merino, Ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013, 1–48.
[365] Ferrero, M. and Gotor, V., 'Biocatalytic Selective Modifications of Conventional Nucleosides, Carbocyclic Nucleosides, and C-Nucleosides', *Chem. Rev.*, 100, (12), 4319–4348, 2000.
[366] Huang, B., Chen, B., and Hui, Y., 'First Synthesis of 4′-Acetamido-2′-deoxythymidine', *Synthesis*, 1993, (08), 769–771, 1993.
[367] Reist, E. J., Gueffroy, D. E., Blackford, R. W., and Goodman, L., 'Pyrrolidine Sugars. Synthesis of 4'-Acetamidoadenosine and Other Derivatives of 4-Amino-4-deoxy-D-ribosel', *J. Org. Chem.*, 31, (12), 4025–4030, 1966.
[368] Yokoyama, M. and Momotake, A., 'Synthesis and Biological Activity of Azanucleosides', *Synthesis*, 1999, (09), 1541–1554, 1999.
[369] Reist, E. J., Gueffroy, D. E., and Goodman, Leon., 'Synthesis of 4-Thio-D- and -L-ribofuranose and the Corresponding Adenine Nucleosides', *J. Am. Chem. Soc.*, 86, (24), 5658–5663, 1964.
[370] Dyson, M. R., Coe, P. L., and Walker, R. T., 'The synthesis and antiviral properties of E-5-(2-bromovinyl)-4′-thio-2′-deoxyuridine', *J. Chem. Soc., Chem. Commun.*, 0, (10), 741–742, 1991.
[371] Austin, A. T. and Howard, J., '642. The steric course of lactonisation following the deamination of glutamic acid, glutamine, and γ-aminovaleric acid', *J. Chem. Soc.*, 0, (0), 3278–3284, 1961.
[372] Secrist, J. A., Riggs, R. M., Tiwari, K. N., and Montgomery, J. A., 'Synthesis and anti-HIV activity of 4'-thio-2',3'-dideoxynucleosides', *J. Med. Chem.*, 35, (3), 533–538, 1992.
[373] Evans, C. A. *et al.*, 'Divergent asymmetric syntheses of dioxolane nucleoside analogues', *Tetrahedron: Asymmetry*, 4, (11), 2319–2322, 1993.
[374] Belleau, B. R., Evans, C. A., Tse, H. L. A., Jin, H., Dixit, D. M., and Mansour, T. S., 'Oxidative degradation of L-ascorbic acid acetals to 2′,3′-dideoxy-3′-oxaribofuranosides. Synthesis of enantiomerically pure 2′,3′-dideoxy-3′-oxacytidine stereoisomers as potential antiviral agents', *Tetrahedron Letters*, 33, (46), 6949–6952, 1992.

[375] Kim, H. O. *et al.*, 'L-.beta.-(2S,4S)- and L-.alpha.-(2S,4R)-dioxolanyl nucleosides as potential anti-HIV agents: asymmetric synthesis and structure-activity relationships', *J. Med. Chem.*, 36, (5), 519–528, 1993.
[376] Kim, H. O. *et al.*, '1,3-Dioxolanylpurine nucleosides (2R,4R) and (2R,4S) with selective anti-HIV-1 activity in human lymphocytes', *J. Med. Chem.*, 36, (1), 30–37, 1993.
[377] Nguyen-Ba, N. *et al.*, 'Synthesis and anti-HIV activity of 1,3-dithiolane nucleosides', *Chem. Commun.*, 0, (13), 1245–1246, 1999.
[378] Caputo, R., Guaragna, A., Palumbo, G., and Pedatella, S., 'Asymmetric Synthesis of 1,3-Dithiolane Nucleoside Analogues', *European Journal of Organic Chemistry*, 2003, (2), 346–350, 2003.
[379] Jin, H. *et al.*, 'Diastereoselective Synthesis of the Potent Antiviral Agent (-)-2'-Deoxy-3'-thiacytidine and Its Enantiomer', *J. Org. Chem.*, 60, (8), 2621–2623, 1995.
[380] Fernández, F., García, G., and Rodríguez, J. E., 'A Useful Synthesis of Chiral Glyoxylates', *Synthetic Communications*, 20, (18), 2837–2847, 1990.
[381] Goodyear, M. D., Hill, M. L., West, J. P., and Whitehead, A. J., 'Practical enantioselective synthesis of lamivudine (3TC™) via a dynamic kinetic resolution', *Tetrahedron Letters*, 46, (49), 8535–8538, 2005.
[382] Milton, J., Brand, S., Jones, M. F., and Rayner, C. M., 'Enzymatic Resolution of α-acetoxysulfides: A new approach to the synthesis of homochiral S,O-Acetals', *Tetrahedron: Asymmetry*, 6, (8), 1903–1906, 1995.
[383] Milton, J., Brand, S., Jones, M. F., and Rayner, C. M., 'Enantioselective enzymatic synthesis of the anti-viral agent lamivudine (3TC™)', *Tetrahedron Letters*, 36, (38), 6961–6964, 1995.
[384] Jeong, L. S. *et al.*, 'Asymmetric synthesis and biological evaluation of .beta.-L-(2R,5S)- and .alpha.-L-(2R,5R)-1,3-oxathiolane-pyrimidine and -purine nucleosides as potential anti-HIV agents', *J. Med. Chem.*, 36, (2), 181–195, 1993.
[385] Evans, M. E. and Parrish, F. W., 'A simple synthesis of L-gulose', *Carbohydrate Research*, 28, (2), 359–364, 1973.
[386] Dinkelaar, J. *et al.*, 'Stereoselective Synthesis of L-Guluronic Acid Alginates', *Chem. Eur. J.*, 14, (30), 9400–9411, 2008.
[387] Chu, C. K. *et al.*, 'Enantiomeric synthesis of (+)-BCH-189 [(+)-(2S,5R)-1-[2-(hydroxymethyl)-1,3-oxathiolan-5-yl]cytosine] from D-mannose and its anti-HIV activity', *J. Org. Chem.*, 56, (23), 6503–6505, 1991.
[388] Jeong, L. S. *et al.*, 'Structure-activity relationships of .beta.-D-(2S,5R)- and .alpha.-D-(2S,5S)-1,3-oxathiolanyl nucleosides as potential anti-HIV agents', *J. Med. Chem.*, 36, (18), 2627–2638, 1993.
[389] Crimmins, M. T., 'New developments in the enantioselective synthesis of cyclopentyl carbocyclic nucleosides', *Tetrahedron*, 54, (32), 9229–9272, 1998.
[390] Leclerc, E., 'Chemical Synthesis of Carbocyclic Analogues of Nucleosides', in *Chemical Synthesis of Nucleoside Analogues*, John Wiley & Sons, Ltd, 2013, 535–604.
[391] Moon, H. R., Choi, W. J., Kim, H. O., and Jeong, L. S., 'Improved and alternative synthesis of d- and l-cyclopentenone derivatives, the versatile intermediates for the synthesis of carbocyclic nucleosides', *Tetrahedron: Asymmetry*, 13, (11), 1189–1193, 2002.
[392] Wang, P., Agrofoglio, L. A., Gary Newton, M., and Chu, C. K., 'Asymmetric synthesis of L-cyclopentyl carbocyclic nucleosides', *Tetrahedron Letters*, 38, (24), 4207–4210, 1997.
[393] Guo, H.-M., Wu, S., Niu, H.-Y., Song, G., and Qu, G.-R., 'Chemical Synthesis of Acyclic Nucleosides', in *Chemical Synthesis of Nucleoside Analogues*, P. Merino, Ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013, 103–162.

[394] Xie, M.-S., Niu, H.-Y., Qu, G.-R., and Guo, H.-M., 'The development for the synthesis of chiral acyclic nucleosides and their phosphonates', *Tetrahedron Letters*, 55, (52), 7156–7166, 2014.
[395] Brown Ripin, D. H. *et al.*, 'Process Improvements for the Manufacture of Tenofovir Disoproxil Fumarate at Commercial Scale', *Org. Process Res. Dev.*, 14, (5), 1194–1201, 2010.
[396] Dueholm, K. L. and Pedersen, E. B., '2,3-Dideoxy-furanoses in Convergent Syntheses of 2′,3′-Dideoxy Nucleosides', *Synthesis*, 1992, (01/02), 1–22, 1992.
[397] Häfele, B. and Jäger, V., 'A Short Synthesis of (S)-5-Hydroxy-2-penten-4-olide from D-Mannitol via 2,3-O-Isopropylidene-D-glyceraldehyde', *Liebigs Annalen der Chemie*, 1987, (1), 85–87, 1987.
[398] Motawia, M. S. and Pedersen, E. B., 'A new route to 2′,3′-dideoxycytidine', *Liebigs Annalen der Chemie*, 1990, (6), 599–602, 1990.
[399] Clercq, E. D., Aerschot, A. V., Herdewijn, P., Baba, M., Pauwels, R., and Balzarini, J., 'Anti-Hiv-1 Activity of 2′,3′-Dideoxinucleoside Analogues : Structure-Activity Relationship', *Nucleosides and Nucleotides*, 8, (5–6), 659–671, 1989.
[400] Sari, O., Roy, V., and Agrofoglio, L. A., 'Nucleosides Modified at the Base Moiety', in *Chemical Synthesis of Nucleoside Analogues*, John Wiley & Sons, Ltd, 2013, 49–101.
[401] Okello, M. and Nair, V., 'Methodologies for the Synthesis of Isomeric Nucleosides and Nucleotides of Antiviral Significance', in *Chemical Synthesis of Nucleoside Analogues*, John Wiley & Sons, Ltd, 2013, 317–343.
[402] Wróblewski, A. E., Głowacka, I. E., and Piotrowska, D. G., '1′-Homonucleosides and their structural analogues: A review', *European Journal of Medicinal Chemistry*, 118, 121–142, 2016.
[403] Hilbert, G. E. and Johnson, T. B., 'Researches On Pyrimidines. CXVII. A Method For The Synthesis Of Nucleosides', *J. Am. Chem. Soc.*, 52, (11), 4489–4494, 1930.
[404] Niedballa, U. and Vorbrüggen, H., 'A General Synthesis of Pyrimidine Nucleosides', *Angewandte Chemie International Edition in English*, 9, (6), 461–462, 1970.
[405] Vorbrüggen, H. and Ruh-Pohlenz, C., 'Synthesis Of Nucleosides', in *Organic Reactions*, John Wiley & Sons, Inc., 2004.
[406] Wilson, L. J., Hager, M. W., El-Kattan, Y. A., and Liotta, D. C., 'Nitrogen Glycosylation Reactions Involving Pyrimidine and Purine Nucleoside Bases with Furanoside Sugars', *Synthesis*, 1995, (12), 1465–1479, 1995.
[407] van Tilburg, E. W. *et al.*, '5'-O-Alkyl Ethers of N,2-Substituted Adenosine Derivatives: Partial Agonists for the Adenosine A1 and A3 Receptors', *J. Med. Chem.*, 44, (18), 2966–2975, 2001.
[408] 'Vorbrüggen Glycosylation: (Vorbrüggen Reaction, Vorbrüggen Glycosidation, Vorbrüggen Coupling)', in *Comprehensive Organic Name Reactions and Reagents*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010.
[409] Mikhailopulo, I. A. *et al.*, 'Synthesis and antiviral and cytostatic properties of 3'-deoxy-3'-fluoro- and 2'-azido-3'-fluoro-2',3'-dideoxy-D-ribofuranosides of natural heterocyclic bases', *J. Med. Chem.*, 34, (7), 2195–2202, 1991.
[410] Downey, A. M., Richter, C., Pohl, R., Mahrwald, R., and Hocek, M., 'Direct One-Pot Synthesis of Nucleosides from Unprotected or 5-O-Monoprotected d-Ribose', *Org. Lett.*, 17, (18), 4604–4607, 2015.
[411] Wang, Z., Prudhomme, D. R., Buck, J. R., Park, M., and Rizzo, C. J., 'Stereocontrolled Syntheses of Deoxyribonucleosides via Photoinduced Electron-Transfer Deoxygenation of Benzoyl-Protected Ribo- and Arabinonucleosides', *J. Org. Chem.*, 65, (19), 5969–5985, 2000.
[412] Vorbrüggen, H. and Höfle, G., 'Nucleoside Syntheses, XXIII1) On the Mechanism of Nucleoside Synthesis', *Chem. Ber.*, 114, (4), 1256–1268, 1981.

[413] Liang, C., Ju, W., Ding, S., Sun, H., and Mao, G., 'Effective Synthesis of Nucleosides Utilizing O-Acetyl-Glycosyl Chlorides as Glycosyl Donors in the Absence of Catalyst: Mechanism Revision and Application to Silyl-Hilbert-Johnson Reaction', *Molecules*, 22, (1), 84, 2017.
[414] Vorbrüggen, H., 'Adventures in Silicon-Organic Chemistry', *Acc. Chem. Res.*, 28, (12), 509–520, 1995.
[415] Paquette, L. A., Seekamp, C. K., Kahane, A. L., Hilmey, D. G., and Gallucci, J., 'Stereochemical Features of Lewis Acid-Promoted Glycosidations Involving 4'-Spiroannulated DNA Building Blocks', *J. Org. Chem.*, 69, (22), 7442–7447, 2004.
[416] Okabe, M., Sun, R. C., Tam, S. Y. K., Todaro, L. J., and Coffen, D. L., 'Synthesis of the dideoxynucleosides "ddC" and "CNT" from glutamic acid, ribonolactone, and pyrimidine bases', *J. Org. Chem.*, 53, (20), 4780–4786, 1988.
[417] Johnstone, R. A. W., '1.11 - Reduction of Carboxylic Acids to Aldehydes by Metal Hydrides', in *Comprehensive Organic Synthesis*, B. M. Trost and I. Fleming, Eds. Oxford: Pergamon, 1991, 259–281.
[418] Kim, E. E., Onyango, E. O., Pace, J. R., Abbot, T. M., Fu, L., and Gribble, G. W., 'Three-component reductive alkylation of 2-hydroxy-1,4-naphthoquinones with lactols', *Tetrahedron Letters*, 57, (8), 864–867, 2016.
[419] Wang, X.-L., Huang, W.-F., Lei, X.-S., Wei, B.-G., and Lin, G.-Q., 'A facile synthesis of 1,4-dideoxy-1,4-imino-l-ribitol (LRB) and (–)-8a-epi-swainsonine from d-glutamic acid', *Tetrahedron*, 67, (26), 4919–4923, 2011.
[420] Choi, W. B., Wilson, L. J., Yeola, S., Liotta, D. C., and Schinazi, R. F., 'In situ complexation directs the stereochemistry of N-glycosylation in the synthesis of thialanyl and dioxolanyl nucleoside analogs', *J. Am. Chem. Soc.*, 113, (24), 9377–9379, 1991.
[421] Gil, A., Lorente, A., Albericio, F., and Álvarez, M., 'Stereoselective Allylstannane Addition for a Convergent Synthesis of a Complex Molecule', *Org. Lett.*, 17, (24), 6246–6249, 2015.
[422] Sujino, K. and Sugimura, H., 'A Useful Method for the Synthesis of β-2′,3′-Dideoxynucleosides via a Thioglycoside', *Synlett*, 1992, (07), 553–555, 1992.
[423] 'Protection for the Hydroxyl Group, Including 1,2- and 1,3-Diols', in *Greene's Protective Groups in Organic Synthesis*, John Wiley & Sons, Ltd, 2014, 17–471.
[424] Barral, K., Balzarini, J., Neyts, J., De Clercq, E., Hider, R. C., and Camplo, M., 'Synthesis and Antiviral Evaluation of Cyclic and Acyclic 2-Methyl-3-hydroxy-4-pyridinone Nucleoside Derivatives', *J. Med. Chem.*, 49, (1), 43–50, 2006.
[425] Cirillo, P. F. and Panek, J. S., 'Diastereoselectivity in nucleophilic addition reactions to (.alpha.,.beta.-dialkoxyacyl) silanes: an operationally useful route to optically active 1,2,3-syn-triols', *J. Org. Chem.*, 55, (25), 6071–6073, 1990.
[426] Kim, Y. A. *et al.*, 'Structure–Activity Relationships of 7-Deaza-6-benzylthioinosine Analogues as Ligands of Toxoplasma gondii Adenosine Kinase', *J. Med. Chem.*, 51, (13), 3934–3945, 2008.
[427] Smith III, A. B. *et al.*, 'The Spongistatins: Architecturally Complex Natural Products—Part Two: Synthesis of the C(29–51) Subunit, Fragment Assembly, and Final Elaboration to (+)-Spongistatin 2', *Angew. Chem.*, 113, (1), 202–205, 2001.
[428] Bhatt, U., Christmann, M., Quitschalle, M., Claus, E., and Kalesse, M., 'The First Total Synthesis of (+)-Ratjadone', *J. Org. Chem.*, 66, (5), 1885–1893, 2001.
[429] Brennan, N. K., Guo, X., and Paquette, L. A., 'Second-Generation, Highly Abbreviated Route for Elaboration of the Oxetane D-Ring in a Fully Functionalized Taxane', *J. Org. Chem.*, 70, (2), 732–734, 2005.
[430] Nicolaou, K. C. *et al.*, 'Synthesis and biological evaluation of 2′,4′- and 3′,4′-bridged nucleoside analogues', *Bioorganic & Medicinal Chemistry*, 19, (18), 5648–5669, 2011.

[431] Neuner, S., Santner, T., Kreutz, C., and Micura, R., 'The "Speedy" Synthesis of Atom-Specific 15N Imino/Amido-Labeled RNA', *Chem. Eur. J.*, 21, (33), 11634–11643, 2015.
[432] Hogrefe, R. I., McCaffrey, A. P., Borozdina, L. U., McCampbell, E. S., and Vaghefi, M. M., 'Effect of excess water on the desilylation of oligoribonucleotides using tetrabutylammonium fluoride', *Nucl. Acids Res.*, 21, (20), 4739–4741, 1993.
[433] Kaburagi, Y. and Kishi, Y., 'Operationally Simple and Efficient Workup Procedure for TBAF-Mediated Desilylation: Application to Halichondrin Synthesis', *Org. Lett.*, 9, (4), 723–726, 2007.
[434] Hansen, P. *et al.*, 'A Simple Synthetic Route to Silylated Methyl 3-Azido-2,3-dideoxy-alpha,beta-D-erythro-pentofuranoside.', *Acta Chemica Scandinavica*, 44, 522–523, 1990.
[435] Quintiliani, M., Balzarini, J., and McGuigan, C., 'Design, synthesis, and biological evaluation of C1–phosphonamidate analogues of 2-deoxy-d-ribose-1-phosphate', *Tetrahedron*, 69, (43), 9111–9119, 2013.
[436] Hanessian, S. *et al.*, 'Total Synthesis of Pactamycin', *Angew. Chem. Int. Ed.*, 50, (15), 3497–3500, 2011.
[437] Nowak, I. and Robins, M. J., 'Synthesis of 3'-Deoxynucleosides with 2-Oxabicyclo[3.1.0]hexane Sugar Moieties: Addition of Difluorocarbene to a 3',4'-Unsaturated Uridine Derivative and 1,2-Dihydrofurans Derived from d- and l-Xylose1', *J. Org. Chem.*, 72, (9), 3319–3325, 2007.
[438] Li, N.-S. and Piccirilli, J. A., 'Efficient synthesis of 2′-C-α-aminomethyl-2′-deoxynucleosides', *Chem. Commun.*, 48, (70), 8754–8756, 2012.
[439] Paquette, L. A., Seekamp, C. K., and Kahane, A. L., 'Conformational Restriction of Nucleosides by Spirocyclic Annulation at C4' Including Synthesis of the Complementary Dideoxy and Didehydrodideoxy Analogues', *J. Org. Chem.*, 68, (22), 8614–8624, 2003.
[440] Crich, D. and Hao, X., 'Asymmetric Synthesis of C4'α-Carboxylated 2'-Deoxynucleosides. Preparation of Oxetanone Derivatives and Influence of Solvent on the Stereochemistry of Base Introduction', *J. Org. Chem.*, 64, (11), 4016–4024, 1999.
[441] Liu, P., Sharon, A., and Chu, C. K., 'Fluorinated nucleosides: Synthesis and biological implication', *Journal of Fluorine Chemistry*, 129, (9), 743–766, 2008.
[442] Herdewijn, P. *et al.*, '3'-Substituted 2',3'-dideoxynucleoside analogs as potential anti-HIV (HTLV-III/LAV) agents', *J. Med. Chem.*, 30, (8), 1270–1278, 1987.
[443] Hartmann, H., Vogt, M. W., Durno, A. G., Hirsch, M. S., Hunsmann, G., and Eckstein, F., 'Enhanced In Vitro Inhibition of HIV-1 Replication by 3′-Fluoro-3′-deoxythymidine Compared to Several Other Nucleoside Analogs', *AIDS Research and Human Retroviruses*, 4, (6), 457–466, 1988.
[444] Chidgeavadze, Z. G. *et al.*, '3'-Fluoro-2',3'-dideoxyribonucleoside 5'-triphosphates: terminators of DNA synthesis', *FEBS Letters*, 183, (2), 275–278, 1985.
[445] Motawia, M. S. and Pedersen, E. B., 'A short route to 3′-deoxy-3′-fluorothymidine', *Liebigs Annalen der Chemie*, 1990, (11), 1137–1139, 1990.
[446] Garegg, P. J. and Samuelsson, B., 'Oxidation of primary and secondary alcohols in partially protected sugars with the chromium trioxide-pyridine complex in the presence of acetic anhydride', *Carbohydrate Research*, 67, (1), 267–270, 1978.
[447] Hansske, F. and Robins, M. J., 'Nucleic acid related compounds. 43. A convenient procedure for the synthesis of 2′ and 3′-ketonucleosides', *Tetrahedron Letters*, 24, (15), 1589–1592, 1983.
[448] Izquierdo, I., Plaza, M. T., Rodríguez, M., Tamayo, J. A., and Martos, A., 'Lipase mediated resolution of 1,3-butanediol derivatives: chiral building blocks for pheromone enantiosynthesis. Part 3†', *Tetrahedron: Asymmetry*, 12, (2), 293–300, 2001.

[449] Thiverny, M., Demory, E., Baptiste, B., Philouze, C., Chavant, P. Y., and Blandin, V., 'Inexpensive, multigram-scale preparation of an enantiopure cyclic nitrone via resolution at the hydroxylamine stage', *Tetrahedron: Asymmetry*, 22, (12), 1266–1273, 2011.
[450] Kohout, M., Kählig, H., Wolrab, D., Roller, A., and Lindner, W., 'Novel Chiral Selector Based on Mefloquine – A Comparative NMR Study to Elucidate Intermolecular Interactions with Acidic Chiral Selectands', *Chirality*, 24, (11), 936–943, 2012.
[451] Heidmann, O. and Heidmann, T., 'Retrotransposition of a mouse IAP sequence tagged with an indicator gene', *Cell*, 64, (1), 159–170, 1991.
[452] Leibold, D. M., Swergold, G. D., Singer, M. F., Thayer, R. E., Dombroski, B. A., and Fanning, T. G., 'Translation of LINE-1 DNA elements in vitro and in human cells.', *PNAS*, 87, (18), 6990–6994, 1990.
[453] Dombroski, B. A., Scott, A. F., and Kazazian, H. H., 'Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element', *PNAS*, 90, (14), 6513–6517, 1993.
[454] Ribet, D., Dewannieux, M., and Heidmann, T., 'An active murine transposon family pair: Retrotransposition of "master" MusD copies and ETn trans-mobilization', *Genome Res.*, 14, (11), 2261–2267, 2004.
[455] Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G., and Heidmann, T., 'Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells', *Nat Genet*, 36, (5), 534–539, 2004.
[456] Mosmann, T., 'Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays', *Journal of Immunological Methods*, 65, (1–2), 55–63, 1983.
[457] Georgopapadakou, N., 'Discontinued drugs in 2005: anti-infectives', *Expert Opinion on Investigational Drugs*, 16, (1), 1–10, 2007.
[458] Khwaja, T. A. and Heidelberger, Charles., 'Fluorinated Pyrimidines. XXIX. Syntheses of 2',3'-Dehydro-5-fluoro-2'-deoxyuridine and 2',3'-Dideoxy-5-fluorouridine', *J. Med. Chem.*, 10, (6), 1066–1070, 1967.
[459] Heidelberger, C. *et al.*, 'Fluorinated Pyrimidines, A New Class of Tumour-Inhibitory Compounds', *Nature*, 179, (4561), 663, 1957.
[460] Feng, J. Y., Johnson, A. A., Johnson, K. A., and Anderson, K. S., 'Insights into the Molecular Mechanism of Mitochondrial Toxicity by AIDS Drugs', *J. Biol. Chem.*, 276, (26), 23832–23837, 2001.
[461] Newman, E. M. and Santi, D. V., 'Metabolism and mechanism of action of 5-fluorodeoxycytidine', *PNAS*, 79, (21), 6419–6423, 1982.
[462] Zhao, C., Liu, F., and Pyle, A. M., 'An ultra-processive, accurate reverse transcriptase encoded by a metazoan group II intron', *RNA*, rna.063479.117, 2017.
[463] Nowak, E. *et al.*, 'Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry', *Nature Structural & Molecular Biology*, 21, (4), 389–396, 2014.
[464] Izumi, M., Miyazawa, H., Kamakura, T., Yamaguchi, I., Endo, T., and Hanaoka, F., 'Blasticidin S-resistance gene (bsr): A novel selectable marker for mammalian cells', *Experimental Cell Research*, 197, (2), 229–233, 1991.
[465] Kopera, H. C., Moldovan, J. B., Morrish, T. A., Garcia-Perez, J. L., and Moran, J. V., 'Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase', *PNAS*, 108, (51), 20345–20350, 2011.
[466] Takahara, T. *et al.*, 'Dysfunction of the Orleans Reeler Gene Arising from Exon Skipping Due to Transposition of a Full-Length Copy of an Active L1 Sequence into the Skipped Exon', *Hum Mol Genet*, 5, (7), 989–993, 1996.
[467] MacLennan, M. *et al.*, 'Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells', *eLife*, 6, 2017.

[468] Kimberland, M. L., Divoky, V., Prchal, J., Schwahn, U., Berger, W., and Kazazian, H. H., 'Full-Length Human L1 Insertions Retain the Capacity for High Frequency Retrotransposition in Cultured Cells', *Hum Mol Genet*, 8, (8), 1557–1560, 1999.
[469] Garcia-Perez, J. L. *et al.*, 'Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells', *Nature*, 466, (7307), 769–773, 2010.
[470] Naas, T. P. *et al.*, 'An actively retrotransposing, novel subfamily of mouse L1 elements', *The EMBO Journal*, 17, (2), 590–597, 1998.
[471] Goodier, J. L., Ostertag, E. M., Du, K., and Kazazian, H. H., 'A Novel Active L1 Retrotransposon Subfamily in the Mouse', *Genome Res.*, 11, (10), 1677–1685, 2001.
[472] DeBerardinis, R. J., Goodier, J. L., Ostertag, E. M., and Kazazian, H. H., 'Rapid amplification of a retrotransposon subfamily is evolving the mouse genome', *Nat Genet*, 20, (3), 288–290, 1998.
[473] Han, J. S. and Boeke, J. D., 'A highly active synthetic mammalian retrotransposon', *Nature*, 429, (6989), nature02535, 2004.
[474] Malki, S., van der Heijden, G. W., O'Donnell, K. A., Martin, S. L., and Bortvin, A., 'A Role for Retrotransposon LINE-1 in Fetal Oocyte Attrition in Mice', *Developmental Cell*, 29, (5), 521–533, 2014.
[475] Bachiller, S., del-Pozo-Martín, Y., and Carrión, Á. M., 'L1 retrotransposition alters the hippocampal genomic landscape enabling memory formation', *Brain, Behavior, and Immunity*, 64, 65–70, 2017.
[476] Mouse Genome Sequencing Consortium, 'Initial sequencing and comparative analysis of the mouse genome', *Nature*, 420, (6915), 520, 2002.
[477] Volkman, H. E. and Stetson, D. B., 'The enemy within: endogenous retroelements and autoimmune disease', *Nature Immunology*, 15, (5), 415–422, 2014.
[478] Crow, Y. J. and Manel, N., 'Aicardi–Goutières syndrome and the type I interferonopathies', *Nature Reviews Immunology*, 15, (7), 429–440, 2015.
[479] Landriscina, M., Spadafora, C., Cignarelli, M., and Barone, C., 'Anti-Tumor Activity of Non-Nucleosidic Reverse Transcriptase Inhibitors', *Current Pharmaceutical Design*, 13, (7), 737–747, 2007.
[480] Patnala, R. *et al.*, 'Inhibition of LINE-1 retrotransposon-encoded reverse transcriptase modulates the expression of cell differentiation genes in breast cancer cells', *Breast Cancer Res Treat*, 143, (2), 239–253, 2014.
[481] Okabe, M., Sun, R. C., Tam, S. Y. K., Todaro, L. J., and Coffen, D. L., 'Synthesis of the dideoxynucleosides "ddC" and "CNT" from glutamic acid, ribonolactone, and pyrimidine bases', *J. Org. Chem.*, 53, (20), 4780–4786, 1988.
[482] Chakraborty, T. K., Samanta, R., and Kumar, P. K., 'A radical mediated approach to the stereoselective formal total synthesis of (+)-Sch 642305', *Tetrahedron*, 65, (34), 6925–6931, 2009.
[483] Mitsudo, K., Kawaguchi, T., Miyahara, S., Matsuda, W., Kuroboshi, M., and Tanaka, H., 'Electrooxidative Glycosylation through C–S Bond Cleavage of 1-Arylthio-2,3-dideoxyglycosides. Synthesis of 2',3'-Dideoxynucleosides', *Org. Lett.*, 7, (21), 4649–4652, 2005.
[484] Kawakami, H. *et al.*, 'Stereoselectivities in the Coupling Reaction between Silylated Pyrimidine Bases and 1-Halo-2,3-dideoxyribose', *HETEROCYCLES*, 31, (11), 2041, 1990.
[485] Draanen, N. A. V. and Koszalka, G. W., 'Synthesis and Biological Evaluation of Pyrimidine and Purine α-L-2′,3′-Dideoxy Nucleosides', *Nucleosides and Nucleotides*, 13, (8), 1679–1693, 1994.
[486] Rassu, G., Zanardi, F., Battistini, L., Gaetani, E., and Casiraghi, G., 'Expeditious Syntheses of Sugar-Modified Nucleosides and Collections Thereof Exploiting Furan-, Pyrrole-, and Thiophene-Based Siloxy Dienes', *J. Med. Chem.*, 40, (2), 168–180, 1997.
[487] Pfitzner, K. E. and Moffatt, J. G., 'The Synthesis and Hydrolysis of 2',3'-Dideoxyuridine', *J. Org. Chem.*, 29, (6), 1508–1511, 1964.

[488] Beach, J. W. *et al.*, 'A highly stereoselective synthesis of anti-HIV 2',3'-dideoxy- and 2',3'-didehydro-2',3'-dideoxynucleosides', *J. Org. Chem.*, 57, (14), 3887–3894, 1992.
[489] Camarasa, M.-J., Heras, F. G. D. las, and Pérez-Pérez, M. J., 'Aldol Reaction of Nucleoside 5′-Carboxaldehydes with Acetone. Synthesis of 5′-C-Chain Extended Thymidine Derivatives', *Nucleosides and Nucleotides*, 9, (4), 533–546, 1990.
[490] Takahashi, Y., Honda, Y., and Tsuchiya, T., 'Synthesis of (3′R)-3′-deoxy-3′-C-nitromethylthymidine', *Carbohydrate Research*, 270, (1), 77–83, 1995.
[491] Audat, S. A. S., Trzasko Love, C., Al-Oudat, B. A. S., and Bryant-Friedrich, A. C., 'Synthesis of C3′ Modified Nucleosides for Selective Generation of the C3′-Deoxy-3′-thymidinyl Radical: A Proposed Intermediate in LEE Induced DNA Damage', *J. Org. Chem.*, 77, (8), 3829–3837, 2012.
[492] Agyei-Aye, K. and Baker, D. C., 'Synthesis and evaluation of a series of 1-(3-alkyl-2,3-dideoxy-α,β-d-erythro-pentofuranosyl)thymines', *Carbohydrate Research*, 183, (2), 261–275, 1988.
[493] Michelson, A. M. and Todd, A. R., 'Deoxyribonucleosides and related compounds. Part V. cycloThymidines and other thymidine derivatives. The configuration at the glycosidic centre in thymidine', *J. Chem. Soc.*, 0, (0), 816–823, 1955.
[494] Rosowsky, A. and Pai, N. N., 'Synthesis of the 2′-,3′-Didehydro-2′-,3′-dideoxy and 2′-,3′-Dideoxy Derivatives of 6-Azauridine and a New Route to 2′-,3′-Didehydro-2′-,3′-dideoxy-5-chlorouridine', *Nucleosides and Nucleotides*, 10, (4), 837–851, 1991.
[495] Lin, T.-S., Luo, M.-Z., Liu, M.-C., Pai, S. B., Dutschman, G. E., and Cheng, Y.-C., 'Synthesis and Biological Evaluation of 2',3'-Dideoxy-L-pyrimidine Nucleosides as Potential Antiviral Agents against Human Immunodeficiency Virus (HIV) and Hepatitis B Virus (HBV)', *J. Med. Chem.*, 37, (6), 798–803, 1994.
[496] Zhuk, R. A., Berzinya, A. É., Silinya, V. N., Liepin'sh, É. É., and Giller, S. A., 'Analogs of pyrimidine mono- and polynucleotides', *Chem Heterocycl Compd*, 15, (8), 926–929, 1979.
[497] Van Aerschot, A. *et al.*, 'Synthesis and anti-HIV evaluation of 2',3'-dideoxyribo-5-chloropyrimidine analogs: reduced toxicity of 5-chlorinated 2',3'-dideoxynucleosides', *J. Med. Chem.*, 33, (6), 1833–1839, 1990.
[498] Furukawa, Y., Yoshioka, Y., Imai, K., and Honjo, M., 'Studies on the Synthesis of Pyrimidine Deoxynucleosides. I. Synthesis of 2', 3'-Dideoxyuridine and 1-(3-Ethylthio-3-deoxy-β-D-xylofuranosyl) uracil', *Chemical & Pharmaceutical Bulletin*, 18, (3), 554–560, 1970.
[499] Sivets, G. G., Klennitskaya, T. V., Zhernosek, E. V., and Mikhailopulo, I. A., 'Synthesis of Peracylated Derivatives of l-Ribofuranose from d-Ribose and Their Use for the Preparation of β-l-Ribonucleosides', *Synthesis*, 2002, (02), 0253–0259, 2002.
[500] Kim, C. H., Marquez, V. E., Broder, S., Mitsuya, H., and Driscoll, J. S., 'Potential anti-AIDS drugs. 2',3'-Dideoxycytidine analogs', *J. Med. Chem.*, 30, (5), 862–866, 1987.
[501] Lin, T.-S., Luo, M.-Z., and Liu, M.-C., 'Synthesis of several pyrimidine l-nucleoside analogues as potential antiviral agents', *Tetrahedron*, 51, (4), 1055–1068, 1995.
[502] Fleet, G. W. J., Son, J. C., and Derome, A. E., 'Methyl 5-0-tert-butyldiphenylsilyl-2-deoxy-α β -d-threo-pentofuranoside as a divergent intermediate for the synthesis of 3'-substituted-2',3'-dideoxynucleosides: synthesis of 3'-azido-3'-deoxythymidine, 3'-deoxy-3'-fluorothymidine and 3'-cyano-3'-deoxythymidine.', *Tetrahedron*, 44, (2), 625–636, 1988.
[503] Zeuthen, J. *et al.*, 'Characterization of a human ovarian teratocarcinoma-derived cell line', *International Journal of Cancer*, 25, (1), 19–32, 1980.
[504] Sassaman, D. M. *et al.*, 'Many human L1 elements are capable of retrotransposition', *Nat Genet*, 16, (1), 37–43, 1997.

[505] Esnault, C., Casella, J.-F., and Heidmann, T., 'A Tetrahymena thermophila ribozyme-based indicator gene to detect transposition of marked retroelements in mammalian cells', *Nucleic Acids Res*, 30, (11), e49–e49, 2002.
[506] Wei, W., Morrish, T. A., Alisch, R. S., and Moran, J. V., 'A Transient Assay Reveals That Cultured Human Cells Can Accommodate Multiple LINE-1 Retrotransposition Events', *Analytical Biochemistry*, 284, (2), 435–438, 2000.
[507] Morrish, T. A., Garcia-Perez, J. L., Stamato, T. D., Taccioli, G. E., Sekiguchi, J., and Moran, J. V., 'Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres', *Nature*, 446, (7132), 208–212, 2007.
[508] Livak, K. J. and Schmittgen, T. D., 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2–ΔΔCT Method', *Methods*, 25, (4), 402–408, 2001.
[509] Evans, M. J. and Kaufman, M. H., 'Establishment in culture of pluripotential cells from mouse embryos', *Nature*, 292, (5819), 154, 1981.
[510] Martin, G. R., 'Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells', *PNAS*, 78, (12), 7634–7638, 1981.
[511] Thomson, J. A. *et al.*, 'Embryonic Stem Cell Lines Derived from Human Blastocysts', *Science*, 282, (5391), 1145–1147, 1998.
[512] Murry, C. E. and Keller, G., 'Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development', *Cell*, 132, (4), 661–680, 2008.
[513] Clevers, H., 'Modeling Development and Disease with Organoids', *Cell*, 165, (7), 1586–1597, 2016.
[514] Kuehn, M. R., Bradley, A., Robertson, E. J., and Evans, M. J., 'A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice', *Nature*, 326, (6110), 295, 1987.
[515] Kim, K. *et al.*, 'Recombination Signatures Distinguish Embryonic Stem Cells Derived by Parthenogenesis and Somatic Cell Nuclear Transfer', *Cell Stem Cell*, 1, (3), 346–352, 2007.
[516] Mai, Q. *et al.*, 'Derivation of human embryonic stem cell lines from parthenogenetic blastocysts', *Cell Research*, 17, (12), 1008–1019, 2007.
[517] Tachibana, M. *et al.*, 'Human Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer', *Cell*, 153, (6), 1228–1238, 2013.
[518] Campbell, K. H. S., McWhir, J., Ritchie, W. A., and Wilmut, I., 'Sheep cloned by nuclear transfer from a cultured cell line', *Nature*, 380, (6569), 64, 1996.
[519] Takahashi, K. *et al.*, 'Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors', *Cell*, 131, (5), 861–872, 2007.
[520] Yu, J. *et al.*, 'Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells', *Science*, 318, (5858), 1917–1920, 2007.
[521] Park, I.-H. *et al.*, 'Reprogramming of human somatic cells to pluripotency with defined factors', *Nature*, 451, (7175), 141–146, 2008.
[522] Buganim, Y. *et al.*, 'Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase', *Cell*, 150, (6), 1209–1222, 2012.
[523] Soldner, F. *et al.*, 'Parkinson's Disease Patient-Derived Induced Pluripotent Stem Cells Free of Viral Reprogramming Factors', *Cell*, 136, (5), 964–977, 2009.
[524] Woltjen, K. *et al.*, '*piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells', *Nature*, 458, (7239), 766–770, 2009.
[525] Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K., 'Virus-free induction of pluripotency and subsequent excision of reprogramming factors', *Nature*, 458, (7239), 771–775, 2009.

[526] Yu, J. *et al.*, 'Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences', *Science*, 324, (5928), 797–801, 2009.
[527] Kim, D. *et al.*, 'Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins', *Cell Stem Cell*, 4, (6), 472–476, 2009.
[528] Warren, L. *et al.*, 'Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA', *Cell Stem Cell*, 7, (5), 618–630, 2010.
[529] Robinton, D. A. and Daley, G. Q., 'The promise of induced pluripotent stem cells in research and therapy', *Nature*, 481, (7381), 295–305, 2012.
[530] Avior, Y., Sagi, I., and Benvenisty, N., 'Pluripotent stem cells in disease modelling and drug discovery', *Nature Reviews Molecular Cell Biology*, 17, (3), 170–182, 2016.
[531] Sterneckert, J. L., Reinhardt, P., and Schöler, H. R., 'Investigating human disease using stem cell models', *Nature Reviews Genetics*, 15, (9), 625–639, 2014.
[532] Ben-David, U., Mayshar, Y., and Benvenisty, N., 'Large-Scale Analysis Reveals Acquisition of Lineage-Specific Chromosomal Aberrations in Human Adult Stem Cells', *Cell Stem Cell*, 9, (2), 97–102, 2011.
[533] Gore, A. *et al.*, 'Somatic coding mutations in human induced pluripotent stem cells', *Nature*, 471, (7336), 63–67, 2011.
[534] Hussein, S. M. *et al.*, 'Copy number variation and selection during reprogramming to pluripotency', *Nature*, 471, (7336), 58–62, 2011.
[535] Lister, R. *et al.*, 'Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells', *Nature*, 471, (7336), 68–73, 2011.
[536] Yang, L. *et al.*, 'Genome-wide inactivation of porcine endogenous retroviruses (PERVs)', *Science*, 350, (6264), 1101–1104, 2015.
[537] Wang, H. *et al.*, 'One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering', *Cell*, 153, (4), 910–918, 2013.
[538] Rouet, P., Smih, F., and Jasin, M., 'Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells', *PNAS*, 91, (13), 6064–6068, 1994.
[539] Rouet, P., Smih, F., and Jasin, M., 'Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.', *Molecular and Cellular Biology*, 14, (12), 8096–8106, 1994.
[540] Kim, Y. G., Cha, J., and Chandrasegaran, S., 'Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain', *PNAS*, 93, (3), 1156–1160, 1996.
[541] Urnov, F. D. *et al.*, 'Highly efficient endogenous human gene correction using designed zinc-finger nucleases', *Nature*, 435, (7042), 646, 2005.
[542] Tebas, P. *et al.*, 'Gene Editing of *CCR5* in Autologous CD4 T Cells of Persons Infected with HIV', *New England Journal of Medicine*, 370, (10), 901–910, 2014.
[543] Moscou, M. J. and Bogdanove, A. J., 'A Simple Cipher Governs DNA Recognition by TAL Effectors', *Science*, 326, (5959), 1501–1501, 2009.
[544] Boch, J. *et al.*, 'Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors', *Science*, 326, (5959), 1509–1512, 2009.
[545] Soldner, F. *et al.*, 'Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations', *Cell*, 146, (2), 318–331, 2011.
[546] DeKelver, R. C. *et al.*, 'Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome', *Genome Res.*, 20, (8), 1133–1142, 2010.
[547] Zou, J. *et al.*, 'Gene Targeting of a Disease-Related Gene in Human Induced Pluripotent Stem and Embryonic Stem Cells', *Cell Stem Cell*, 5, (1), 97–110, 2009.
[548] Lombardo, A. *et al.*, 'Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery', *Nature Biotechnology*, 25, (11), 1298–1306, 2007.

[549] Mojica, F. J. M., Juez, G., and Rodriguez-Valera, F., 'Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites', *Molecular Microbiology*, 9, (3), 613–621, 1993.
[550] Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E., 'Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements', *J Mol Evol*, 60, (2), 174–182, 2005.
[551] Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim., and Schouls, Leo. M., 'Identification of genes that are associated with DNA repeats in prokaryotes', *Molecular Microbiology*, 43, (6), 1565–1575, 2002.
[552] Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E., 'A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes', *PLOS Computational Biology*, 1, (6), e60, 2005.
[553] Brouns, S. J. J. *et al.*, 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes', *Science*, 321, (5891), 960–964, 2008.
[554] Beloglazova, N. *et al.*, 'A Novel Family of Sequence-specific Endoribonucleases Associated with the Clustered Regularly Interspaced Short Palindromic Repeats', *J. Biol. Chem.*, 283, (29), 20361–20371, 2008.
[555] Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., and Doudna, J. A., 'Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense', *Structure*, 17, (6), 904–912, 2009.
[556] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E., 'A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337, (6096), 816–821, 2012.
[557] Cong, L. *et al.*, 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science*, 339, (6121), 819–823, 2013.
[558] Hsu, P. D. *et al.*, 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nat Biotech*, 31, (9), 827–832, 2013.
[559] Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F., 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*, 8, (11), 2281–2308, 2013.
[560] Barrangou, R. and Horvath, P., 'A decade of discovery: CRISPR functions and applications', *Nature Microbiology*, 2, nmicrobiol201792, 2017.
[561] Moreno, A. M. and Mali, P., 'Therapeutic genome engineering via CRISPR-Cas systems', *WIREs Syst Biol Med*, n/a-n/a, 2017.
[562] Koonin, E. V., Makarova, K. S., and Zhang, F., 'Diversity, classification and evolution of CRISPR-Cas systems', *Current Opinion in Microbiology*, 37, 67–78, 2017.
[563] Abudayyeh, O. O. *et al.*, 'C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector', *Science*, 353, (6299), aaf5573, 2016.
[564] Gilbert, L. A. *et al.*, 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes', *Cell*, 154, (2), 442–451, 2013.
[565] Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S., 'CRISPR interference (CRISPRi) for sequence-specific control of gene expression', *Nature Protocols*, 8, (11), 2180–2196, 2013.
[566] Khan, H., Smit, A., and Boissinot, S., 'Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates', *Genome Res.*, 16, (1), 78–87, 2006.
[567] Boissinot, S., Chevret, P., and Furano, A. V., 'L1 (LINE-1) Retrotransposon Evolution and Amplification in Recent Human History', *Mol Biol Evol*, 17, (6), 915–928, 2000.

7. Abreviaturas

7 ABREVIATURAS

$(n\text{-}Bu)_3SnH$: tributyltin hydride

Ac_2O: acetic anhydride

AcCl: acetyl chloride

ACN: acetonitrile

AcOH: acetic acid

AcSH: thioacetic acid

Ag_2CO_3: silver carbonate

AGS: Aicardi-Goutières syndrome

AIBN: 2,2′-azobis(2-methylpropionitrile)

AID: activation-induced cytidine deaminase

APC: adenomatous polyposis coli

APOBEC: apolipoprotein B mRNA editing complex

ASP: antisense promoter

$BaCO_3$: barium carbonate

$BH_3{\cdot}SMe_2$: borane dimethyl sulfide complex

$BnEt_3NCl$: benzyltriethylammonium chloride

BnSH: benzyl mercaptan

BrBn: benzyl bromide

$BrCH_2CH(OEt)_2$: bromoacetaldehyde diethyl acetal

BSA: bis(trimethylsilyl)acetamide

Bu_4NI: tetrabutylammonium iodide

Bu_4NOH: tetrabutylammonium hydroxide

BzCl: benzoil chloride

CC: coiled-coil

$CdCO_3$; cadmium carbonate

cDNA: complementary DNA

CH_3PPh_3Br: methyltriphenylphosphonium bromide

CMV: citomegalovirus

CPFS: cleavage polyadenylation specific factor

CrO_3: chromium(VI) oxide

CTD: carboxyl-terminal domain

DAST: (diethylamino)sulfur trifluoride

DCE: 1,2-dichloroethane

DCM: dichloromethane

DEAD: diethyl azodicarboxylate

DIBAL-H: diisobutylaluminum hydride

DMAP: 4-(dimethylamino)pyridine

DMF: dimethylformamide

DMP: 2,2-dimethoxipropane

DMSO: dimethylsulfoxide

DSB: double strand break

dsRNA: double-stranded RNA

EN: endonuclease

Env: envelope

ERV: endogenous retrovirus

ESC: embryonic stem cell

$EtAlCl_2$: ethylaluminum dichloride

EtOH: ethanol

FBS: fetal bovine serum

FCC: flash column chromatography

Gag: group-specific antigen gen

GFP: green fluorescent protein

H_2O_2: hydrogen peroxide

H_2SO_4: sulphuric acid

HBr: hydrogen bromide

HCl: hydrogen chloride

HERV: human endogenous retrovirus

HF: hydrogen fluoride

$HgCl_2$: mercuric chloride

HMDS: hexamethyldisilazane

HR: homologous recombination

*i*PrOH: isopropanol

iPSC: induced pluripotent stem cell

K_2CO_3: potassium carbonate

KSC(S)OEt: potassium ethyl xanthogenate

LA: lewis acid

LC-MS: liquid chromatography-mass spectra

$LiAlH_4$: lithium aluminum hydride

$LiBH_4$: lithium borohydride

LINE-1 or L1: long interspersed element-1

LTR: long terminal repeats

m-CPBA: 3-chloroperbenzoic acid

Me_2SO_4: dimethyl sulfate

MeCOCl: acetyl chloride

MeOH: methanol

MeONa: sodium methoxide

miRNA: microRNA

MOV10: moloney leukemia virus 10

MsCl: methanesulfonyl chloride

MTT: methylthiazolyldiphenyltetrazolium bromide

$NaBH_4$: sodium borohydride

NaH: sodium hydride

NaI: sodium iodide

$NaIO_4$: sodium periodate

NaN_3: sodium azide

$NaNO_2$: sodium nitrite

NaOAc: sodium acetate

NaOCl: sodium hypochlorite

NBS: *N*-bromosuccinimide

NH_4OH: ammonium hydroxide

NHEJ: non-homologous end joining

NMR: nuclear magnetic resonance

NNRTI: non-nucleoside reverse-transcriptase inhibitor

NOMM: 4-methylmorpholine *N*-oxide

NPC: neural progenitor cell

NRTI: nucleoside reverse-transcriptase inhibitor

ORF: open reading frame

ORF1p: ORF1 protein

ORF2p: ORF2 protein

OsO_4: osmium tetroxide

$Pb(OAc)_4$: lead tetraacetate

PDC: pyridinium dichromate

Ph_3PCHCO_2Et: triphenylcarbethoxymethylenephosphorane

$PhCO_2H$: benzoic acid

PhCOSK: potassium thiobenzoate

PhOC(S)Cl: *O*-phenyl chlorothionoformate

PhSH: thiophenol

piRISC: piRNA-induced silencing complex

piRNA: PIWI-interacting RNA

Pol: polymerase

PP: processed pseudogene

PPh_3: triphenylphosphine

Prt: protease

PTEN: phosphatase and tensin homolog

Py: pyridine

qPCR: quantitative polymerase chain reaction

RC-L1: retrocompetent LINE-1

REP: reporter gene

RNAi: RNA interference

RNP: ribonucleoprotein particle

RRM: RNA recognition motif

RT: reverse transcriptase

RTi: reverse-transcriptase inhibitor

RT-qPCR: reverse transcription-quantitative polymerase chain reaction

$RuCl_3$: ruthenium(III) chloride

SA: splice acceptor

SD: splice dono

siRNA: small interfering RNA

$SnCl_2$: tin(II) chloride

SnCl4: tin(IV) chloride

$SOCl_2$: thionyl chloride

ssRNA: small silencing RNA

ST18: suppression of tumorigenicity 18

TBAF: tetrabutylammonium fluoride

TBDMSCl: *tert*-butyldimethylsilyl chloride

TBDPSCl: *tert*-butyldiphenylsilyl chloride

t-BuOK: potassium *tert*-butoxide

*t*BuOOH: *tert*-butyl hydroperoxide

TE: transposable element

TEA: trimethylamine

Tf_2O: trifluoromethanesulfonic anhydride

THF: tetrahydrofuran

$Ti(OiPr)_4$: titanium(IV) isopropoxide

TIR: terminal inverted repeat

TLC: thin layer chromatografy

TMSOTf: trimethylsilyl trifluoromethanesulfonate

TPRT: target-primed reverse transcription

TrCl: trityl chloride

TsCl: *p*-toluenesulfonyl chloride

TSD: target site duplication

TSG: tumor suppressor gene

TsOH: *p*-toluenesulfonic acid

UTR: untranslated region

VHS: virus herpes simple

VIH: virus de inmunodeficiencia humana

VNTR: variable number target repeats

ZAP: zinc-finger antiviral protein

$ZnCl_2$: zinc chloride

8. Anexo

8 ANEXO

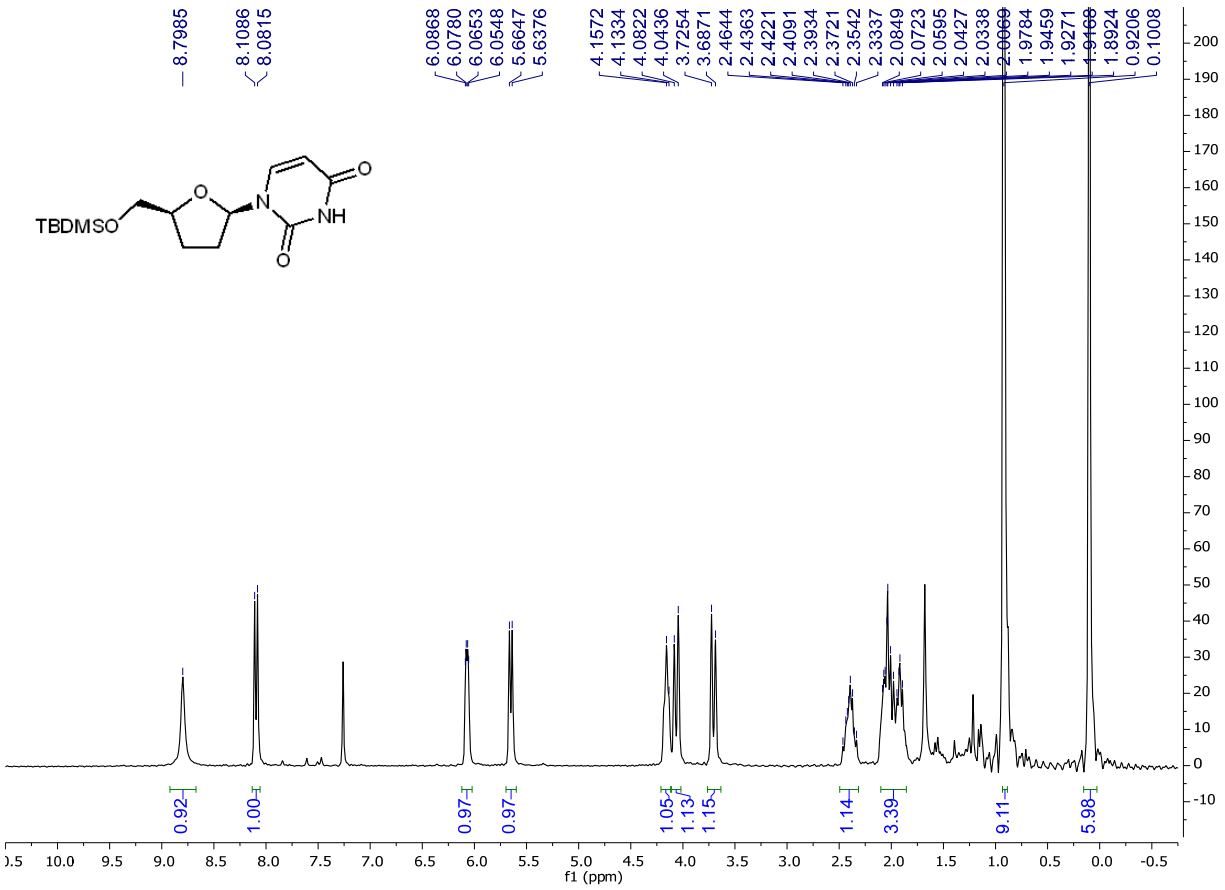


Figure S1. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddU** (300 MHz, $CDCl_3$).

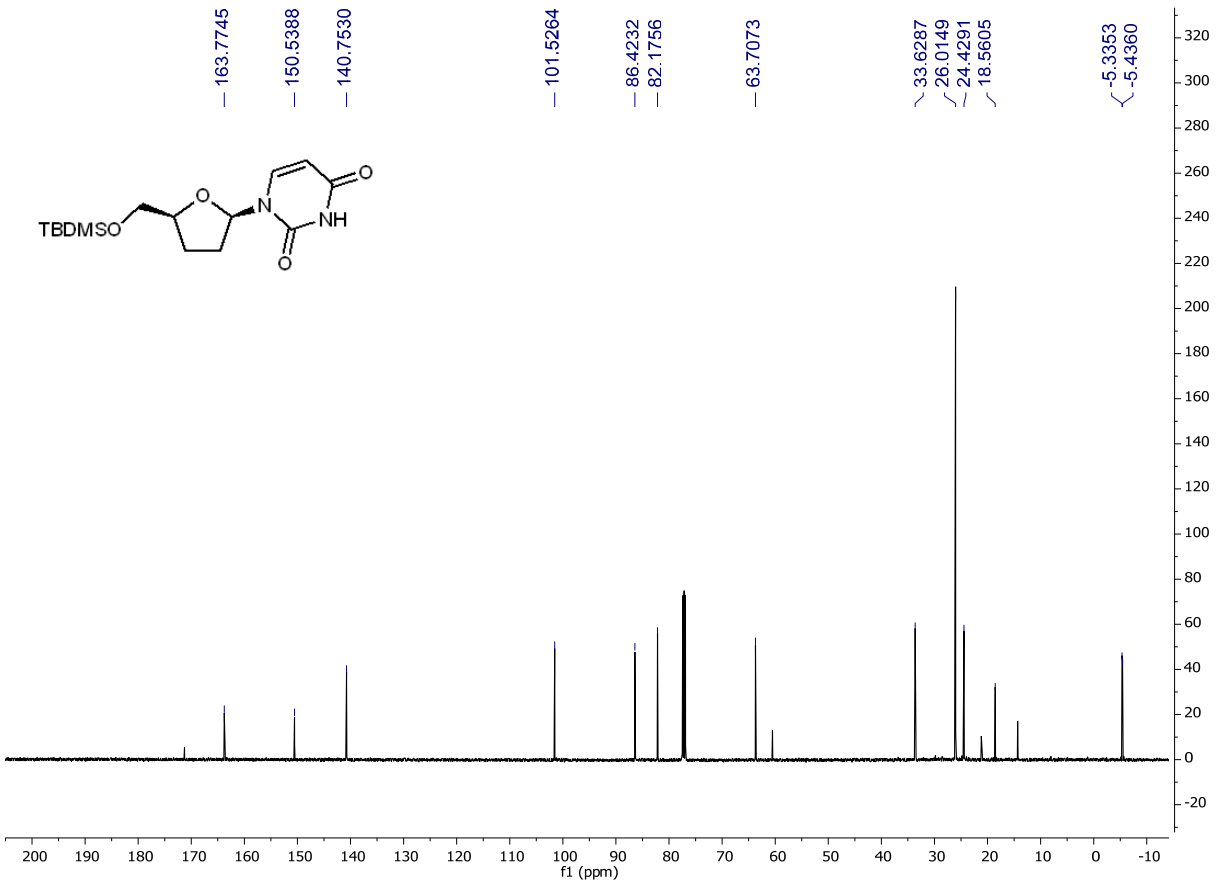


Figure S2. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddU** (125 MHz, $CDCl_3$).

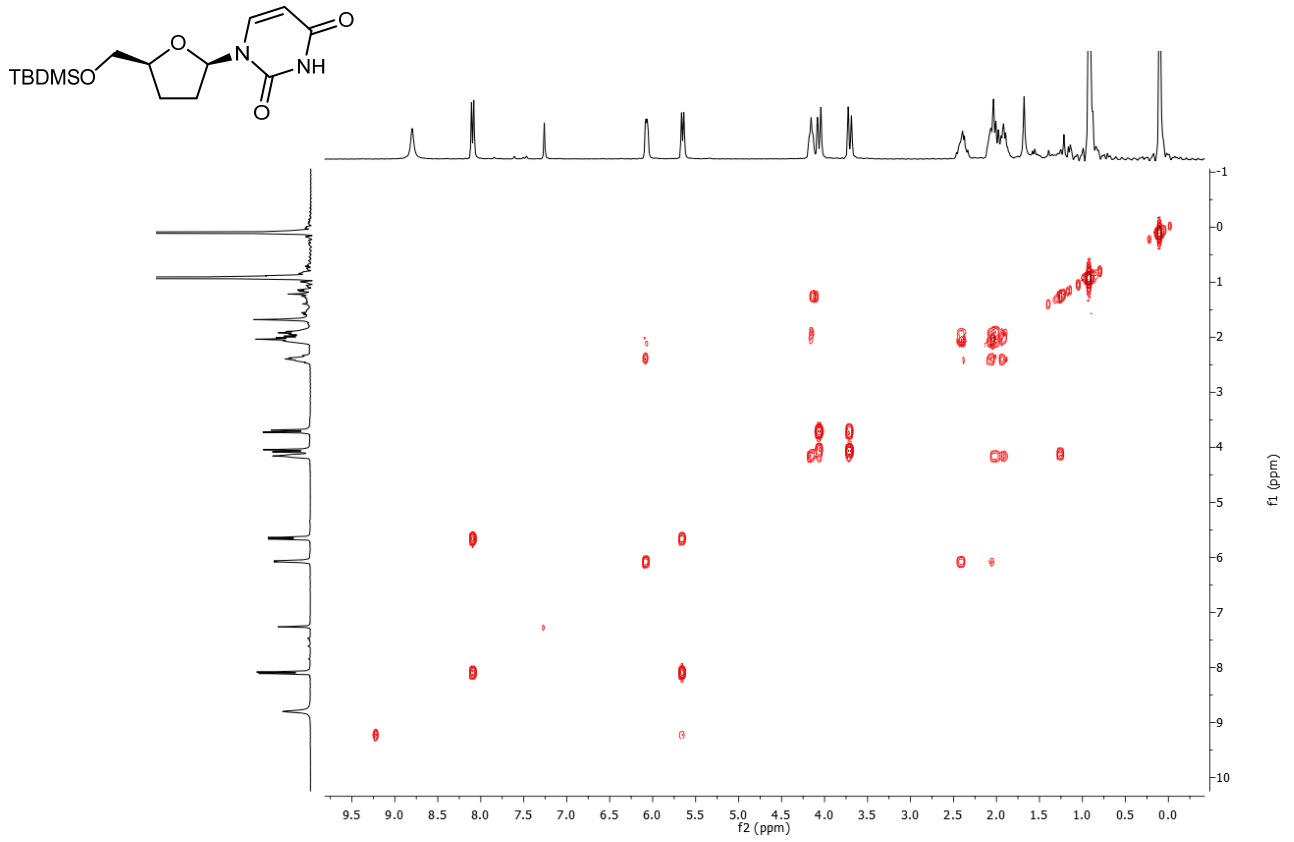


Figure S3. COSY spectrum of **5'-*O*-silyl-protected β-ddU**.

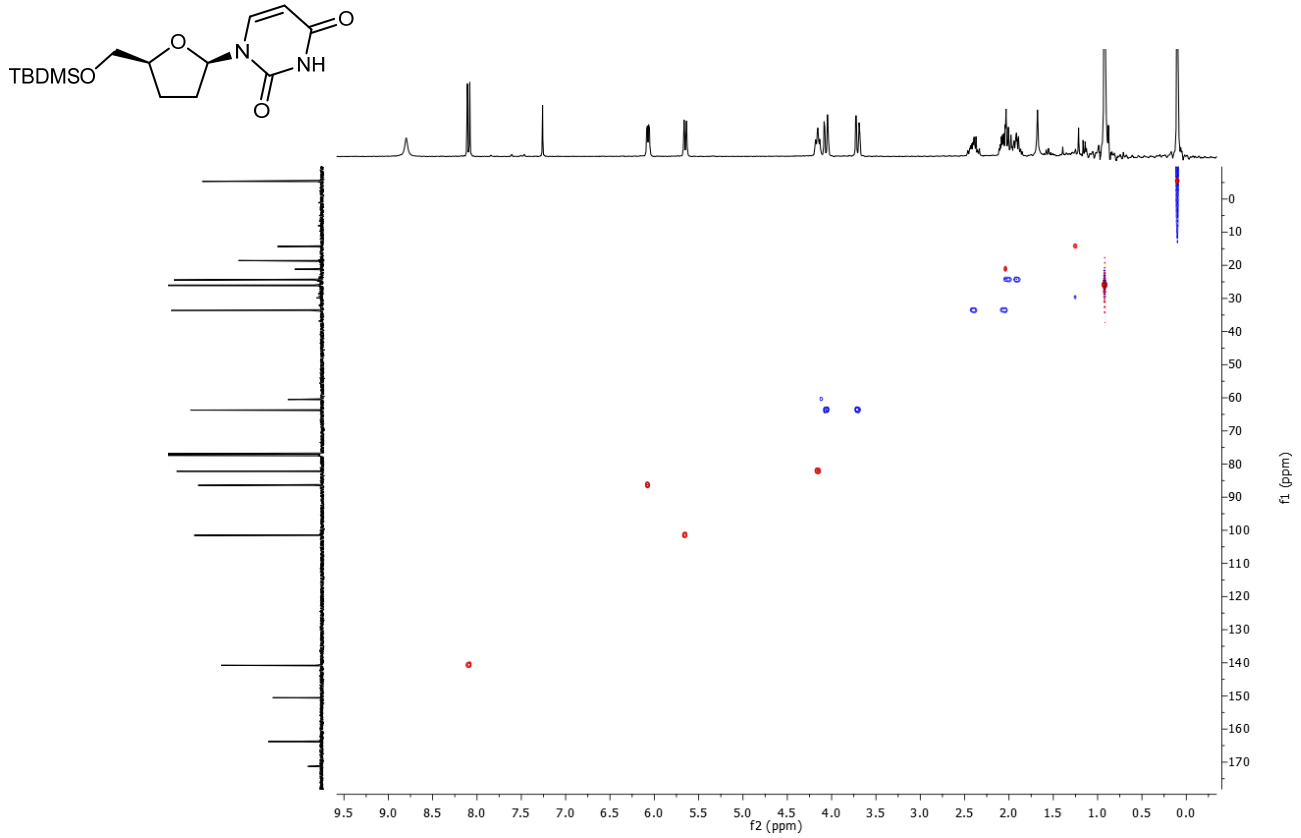


Figure S4. HSQC spectrum of **5'-*O*-silyl-protected β-ddU**.

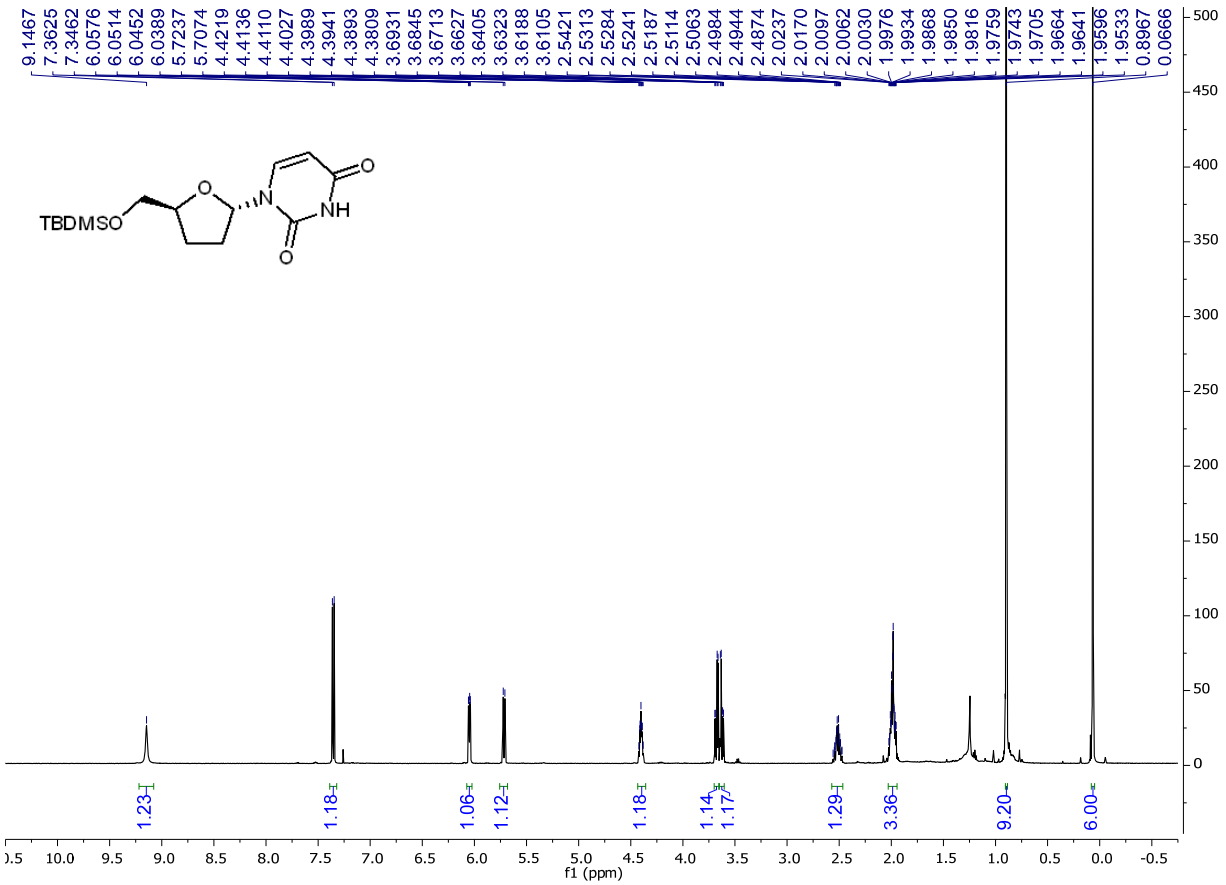


Figure S5. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddU** (500 MHz, $CDCl_3$).

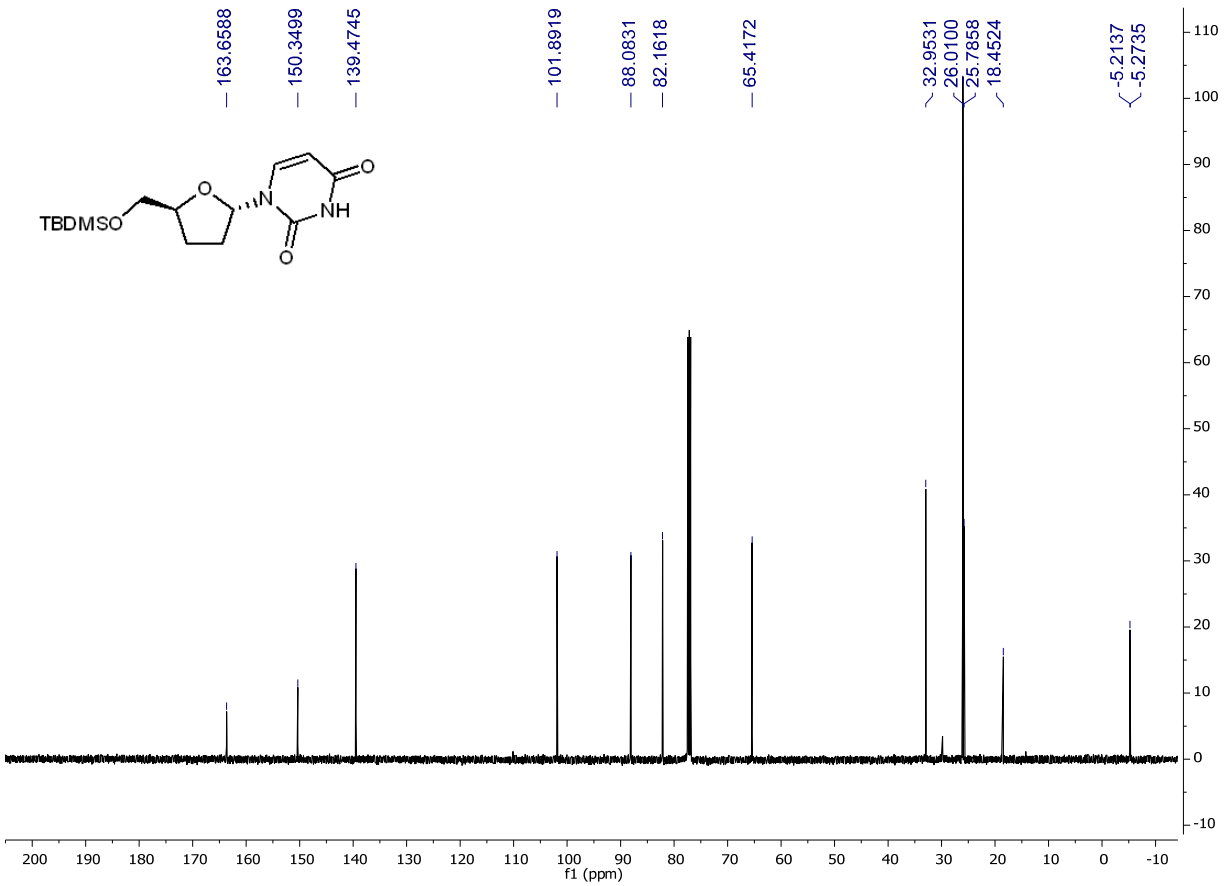


Figure S6. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddU** (125 MHz, $CDCl_3$).

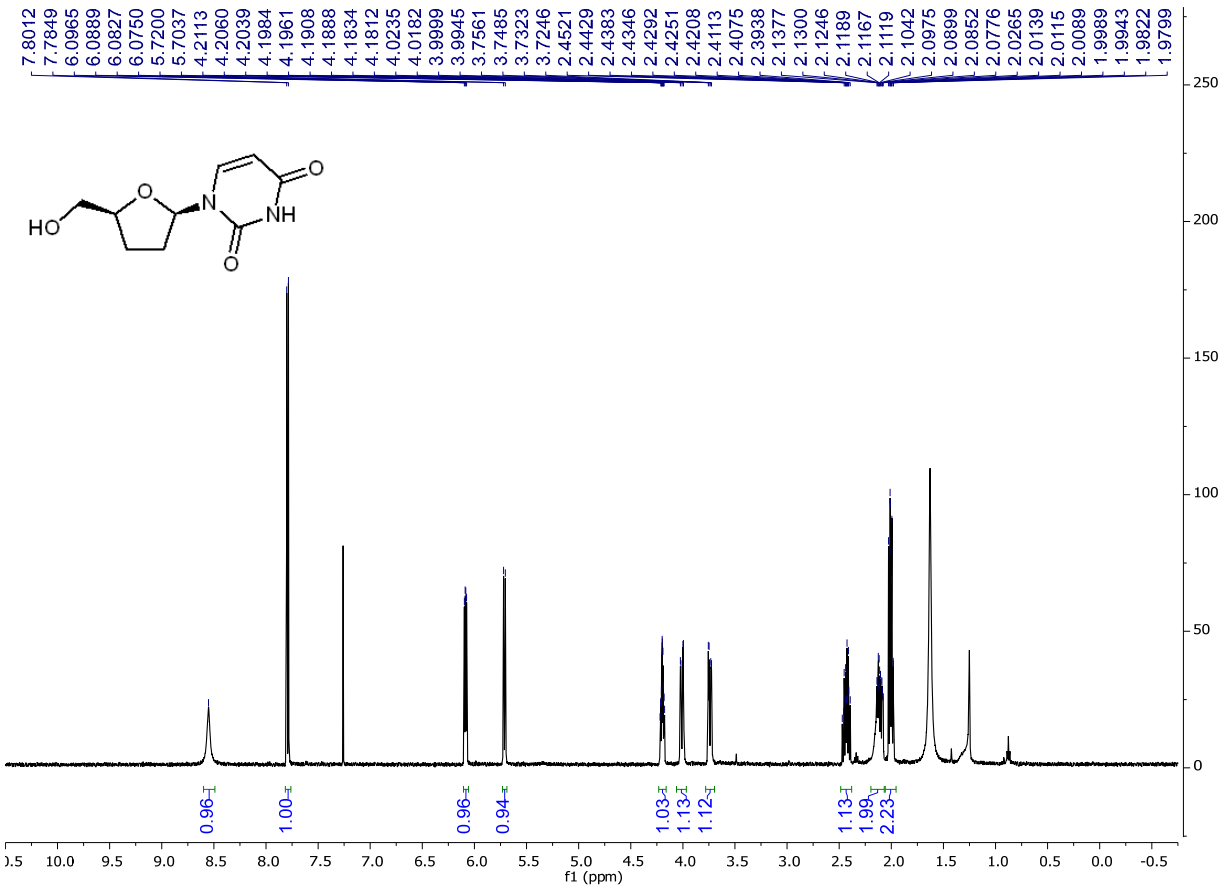


Figure S7. ^{1}H spectrum of **β-ddU** (500 MHz, $CDCl_3$).

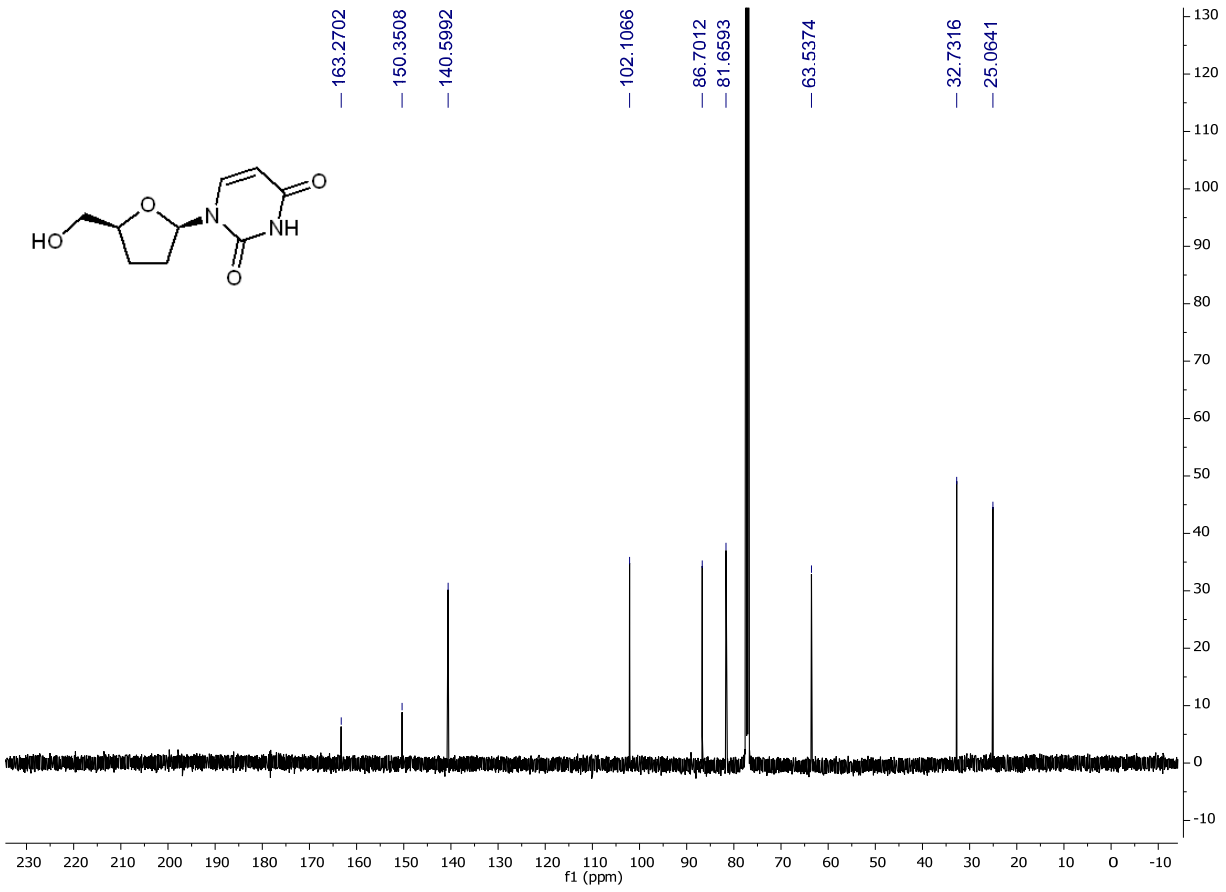


Figure S8. ^{13}C spectrum of **β-ddU** (125 MHz, $CDCl_3$).

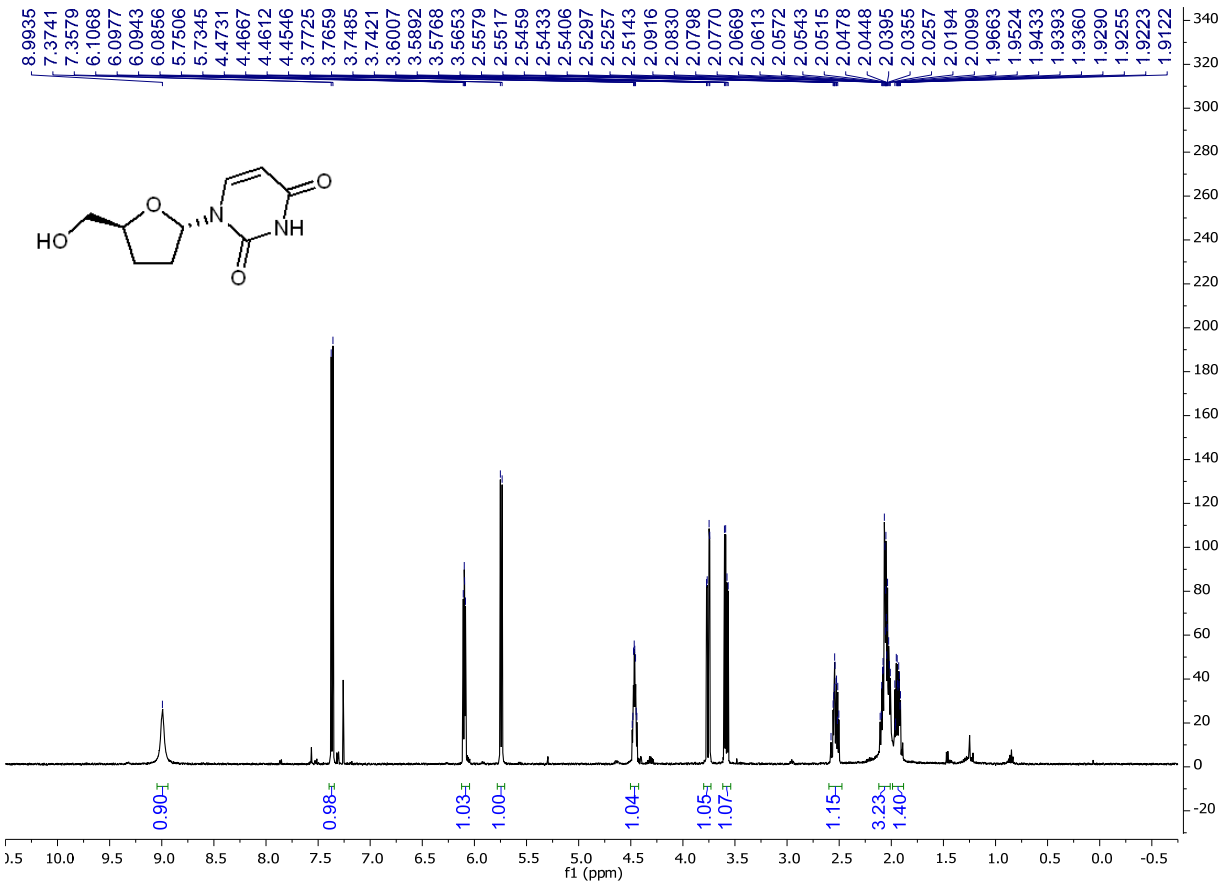


Figure S9. ^{1}H spectrum of **α-ddU** (500 MHz, $CDCl_3$).

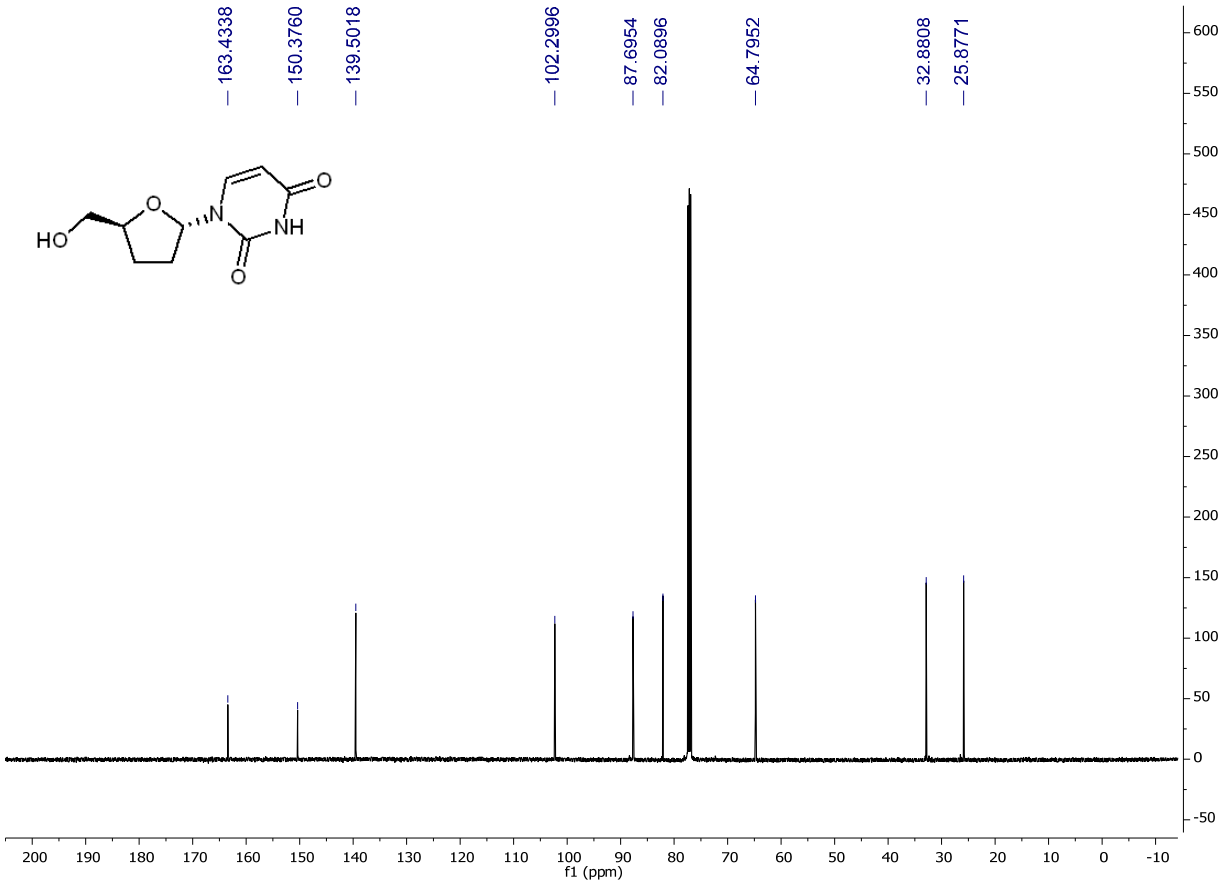


Figure S10. ^{13}C spectrum of **α-ddU** (125 MHz, $CDCl_3$).

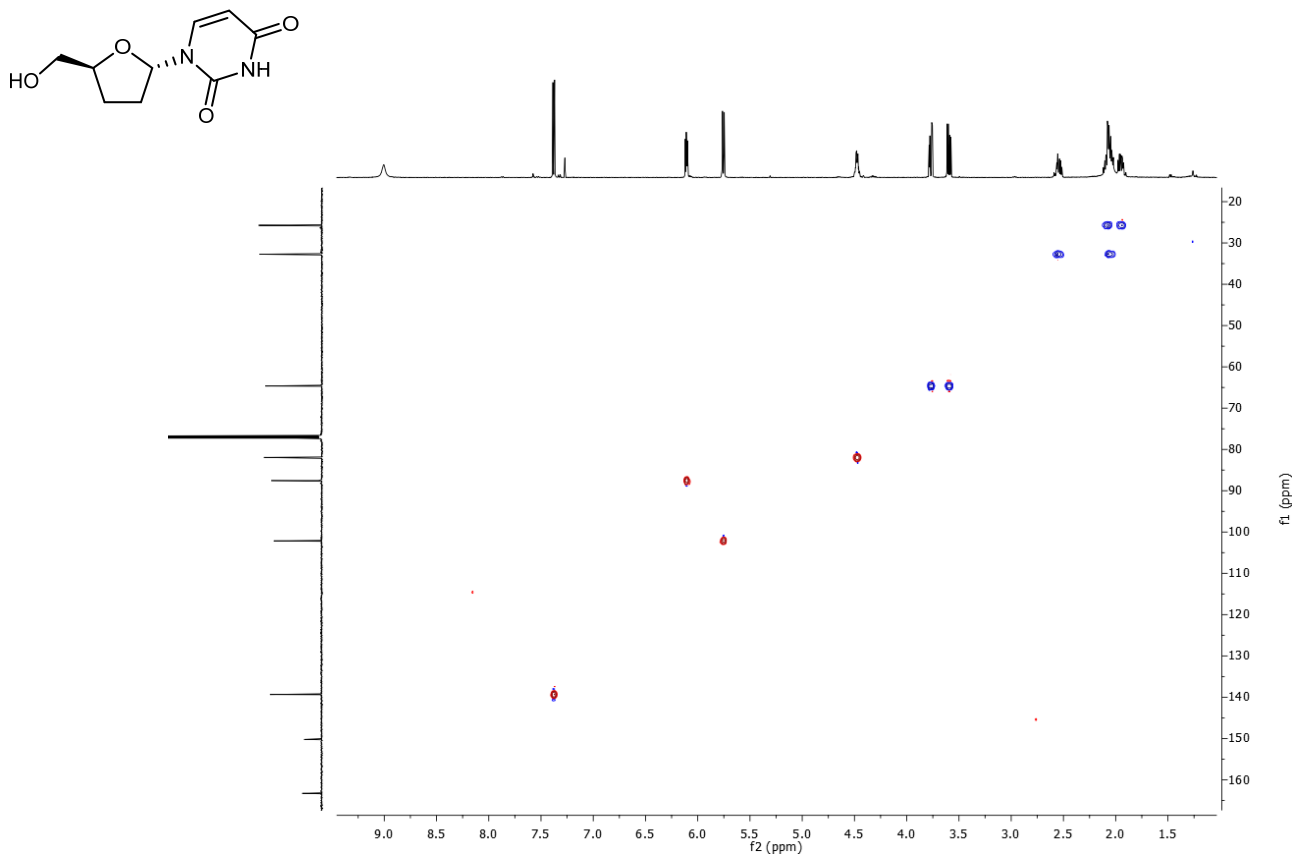


Figure S11. HSQC spectrum of **α-ddU**.

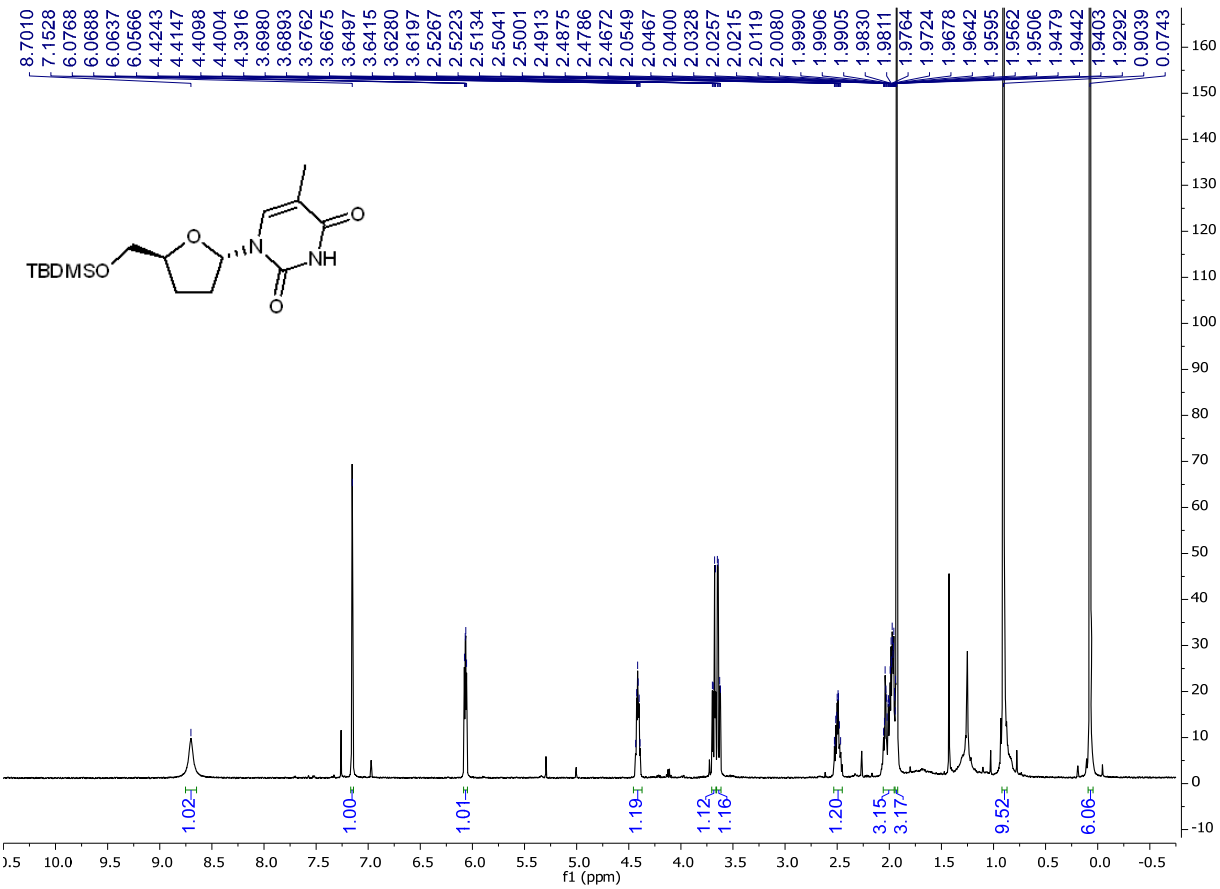


Figure S12. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddT** (500 MHz, $CDCl_3$).

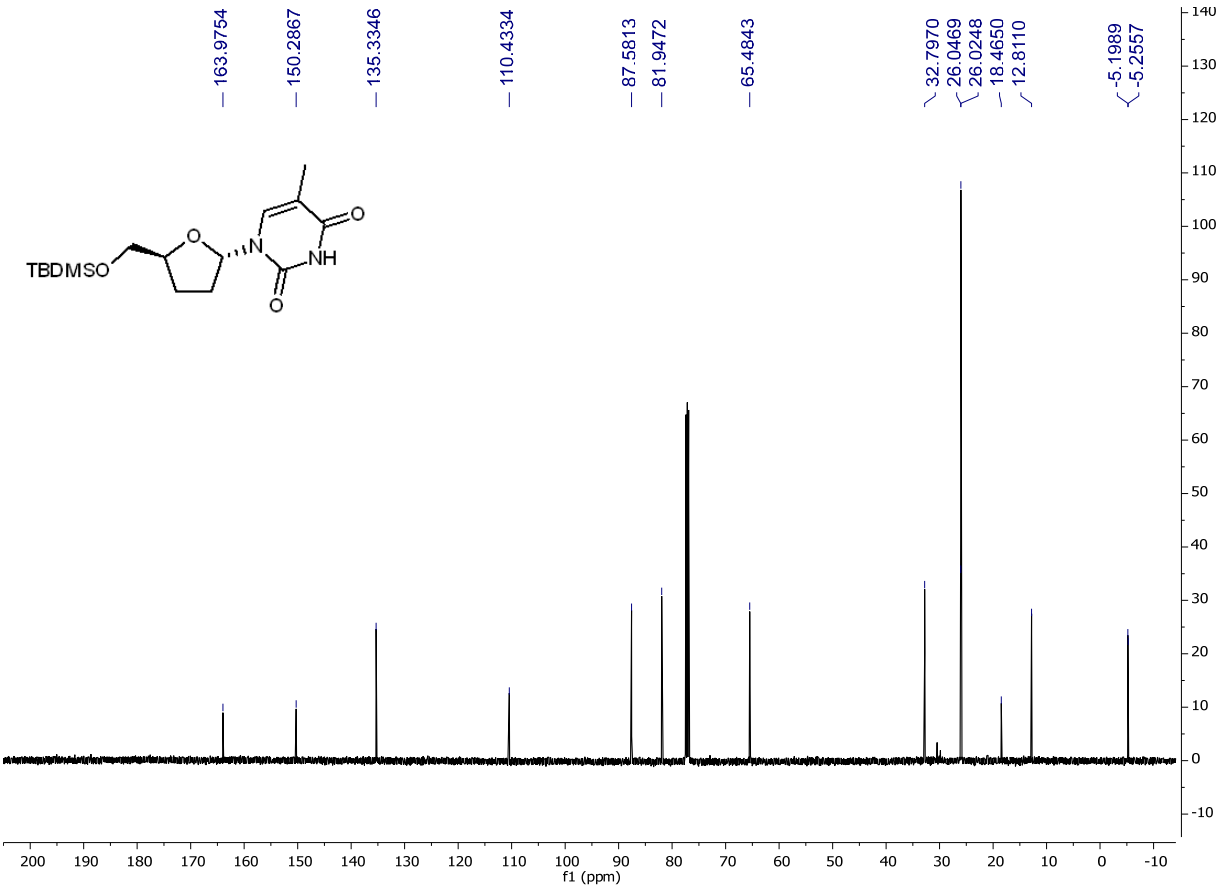


Figure S13. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddT** (125 MHz, $CDCl_3$).

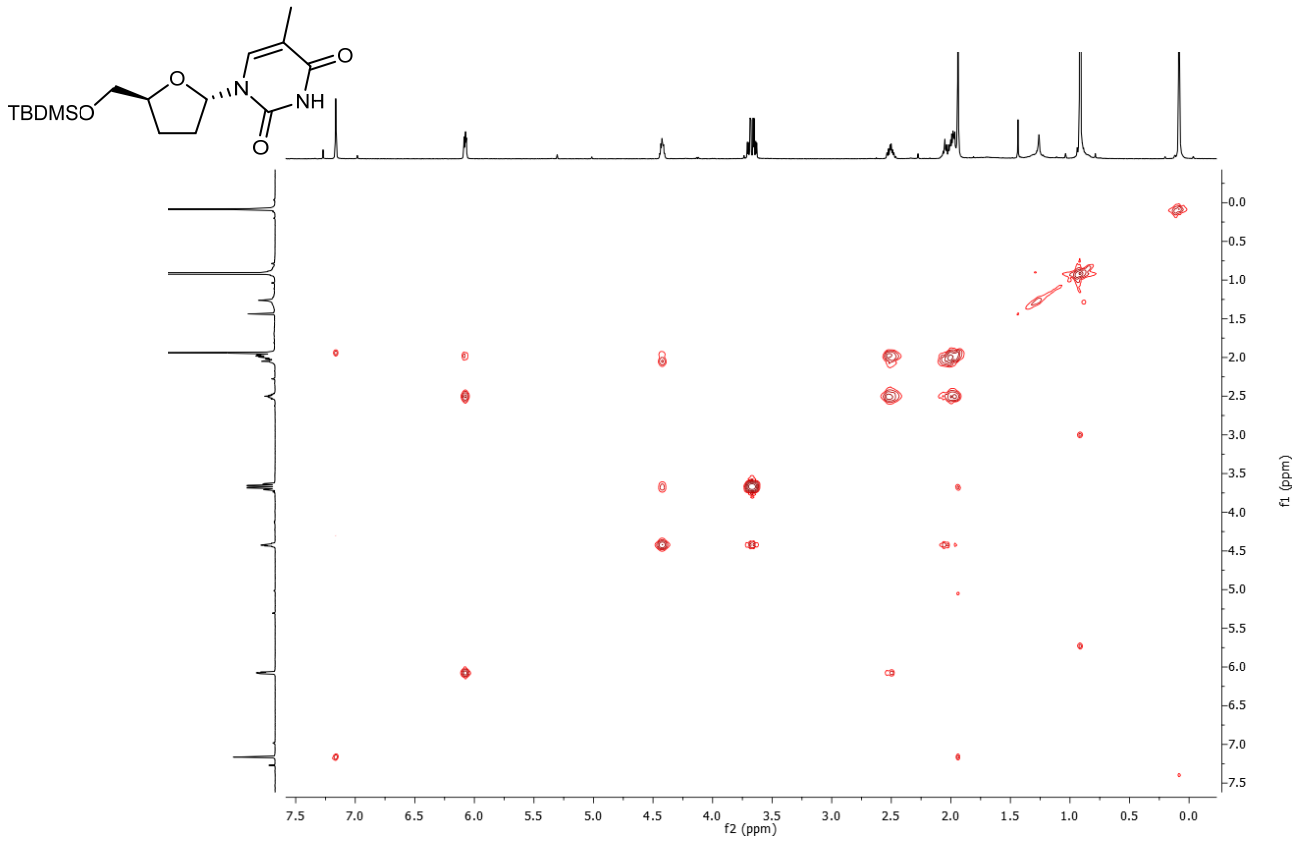


Figure S14. COSY spectrum of **5'-*O*-silyl-protected α-ddT**.

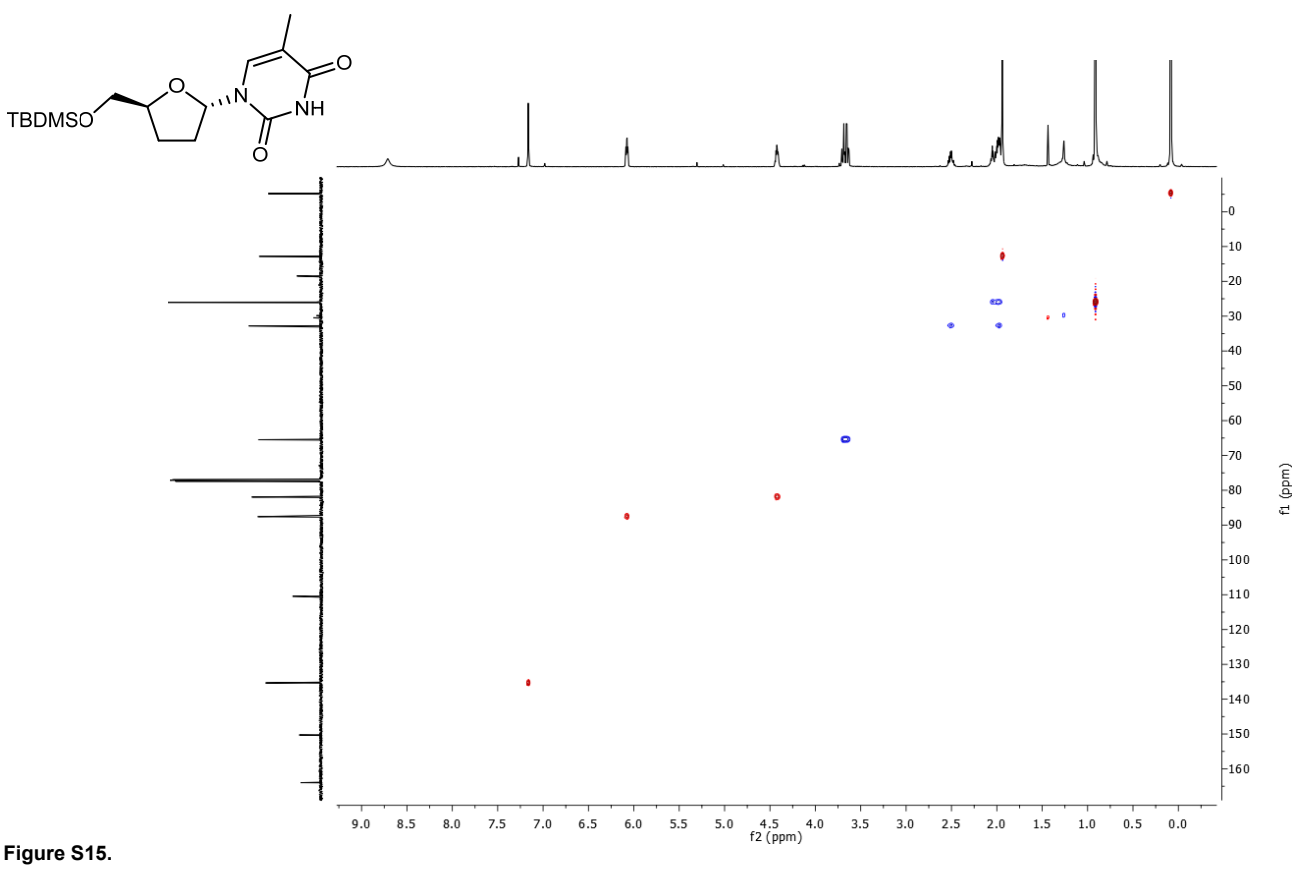


Figure S15.
HSQC spectrum of **5'-*O*-silyl-protected α-ddT**.

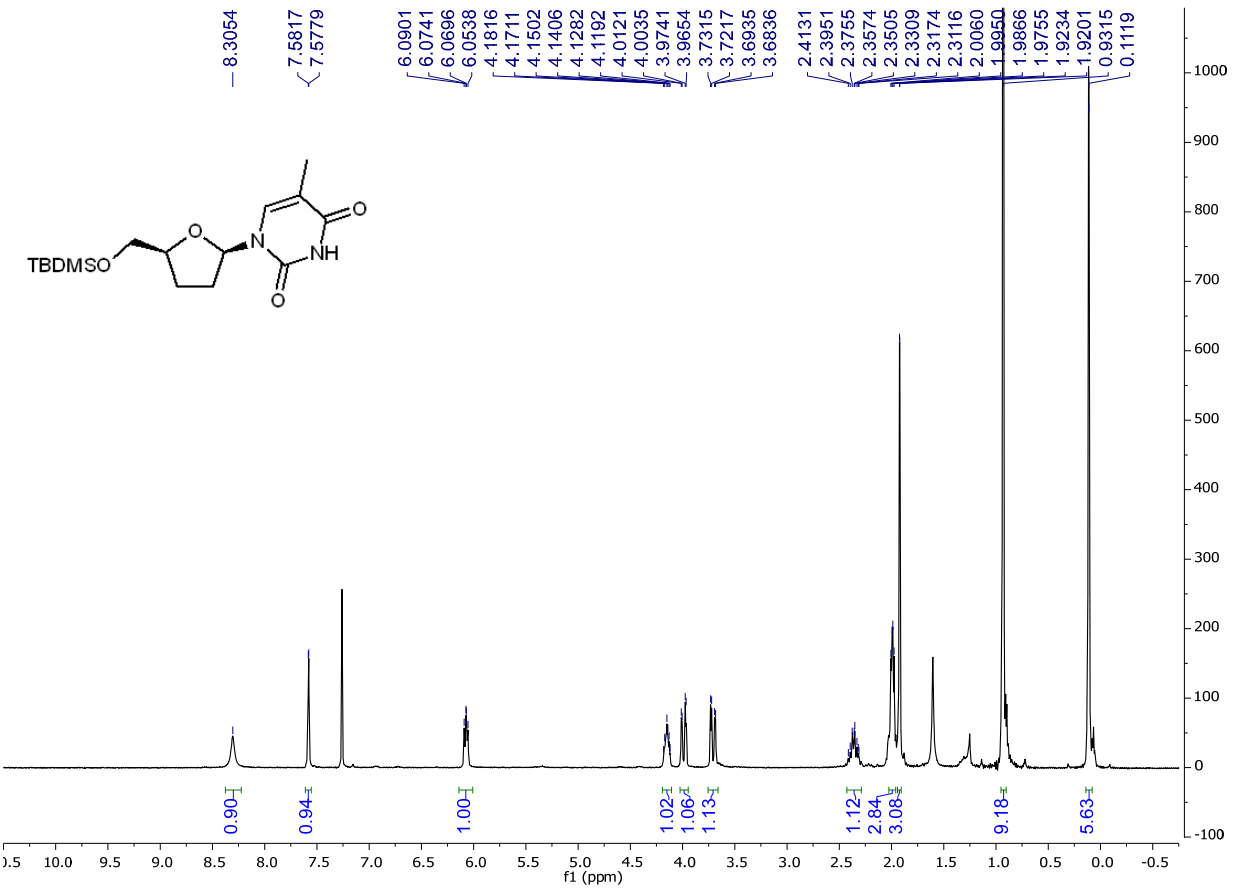


Figure S16. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddT** (300 MHz, $CDCl_3$).

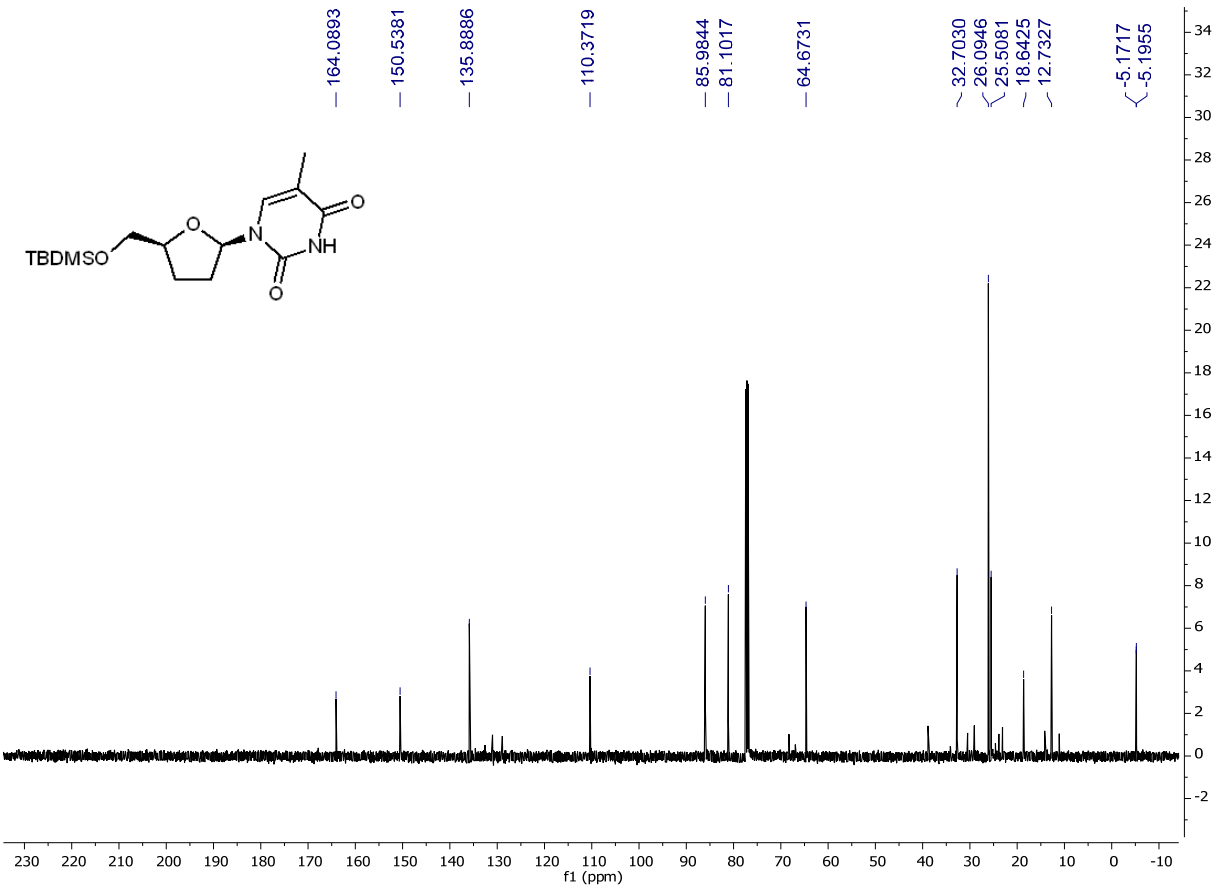


Figure S17. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddT** (125 MHz, $CDCl_3$).

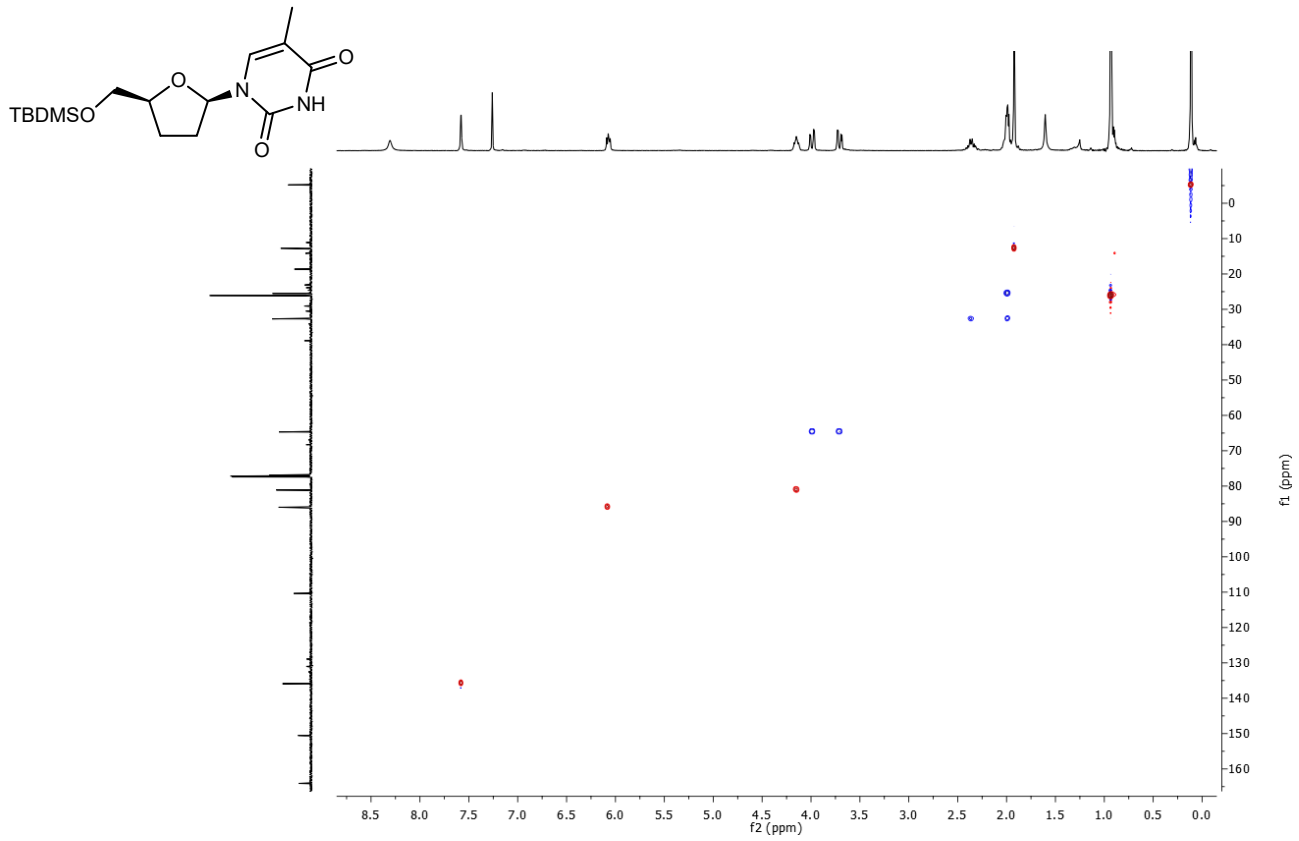


Figure S18. HSQC spectrum of **5'-*O*-silyl-protected β-ddT**.

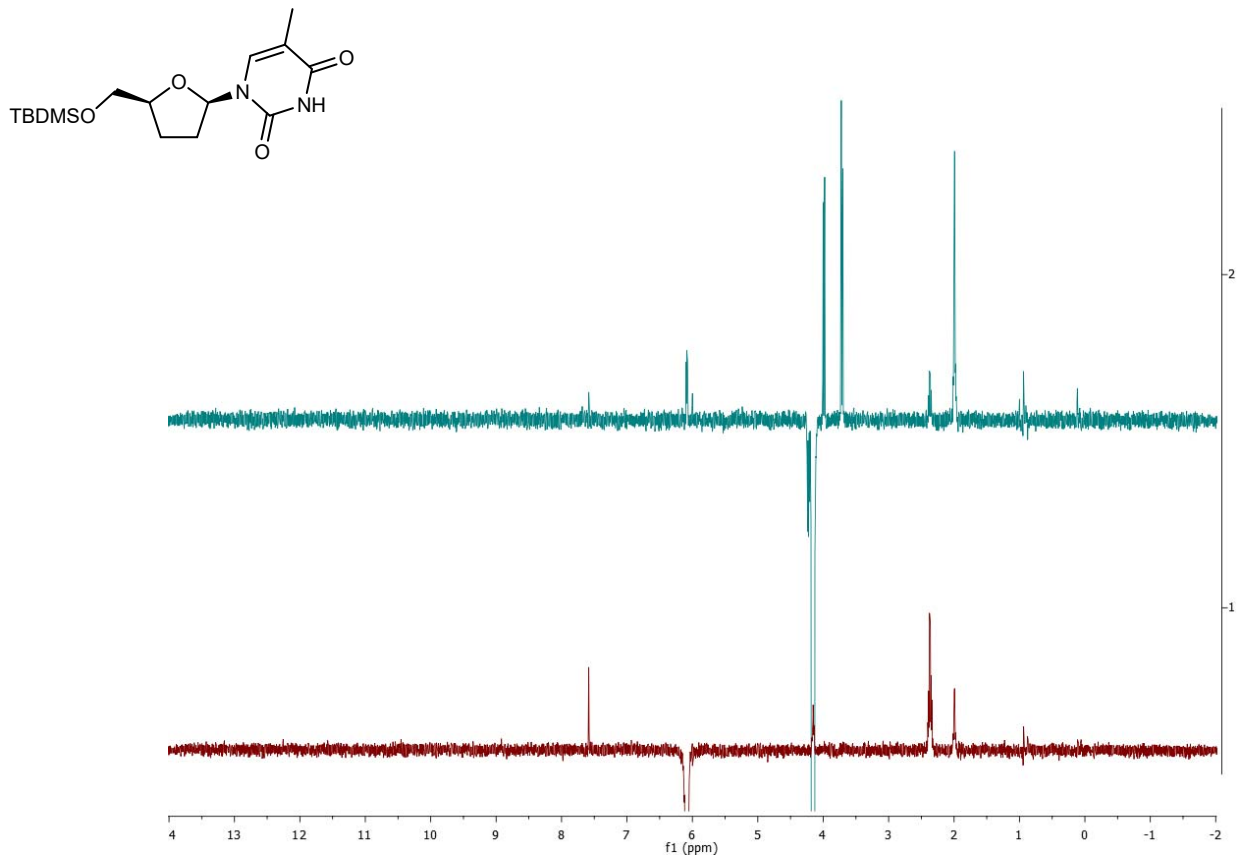


Figure S19. 1D-NOESY spectra of **5'-*O*-silyl-protected β-ddT.**

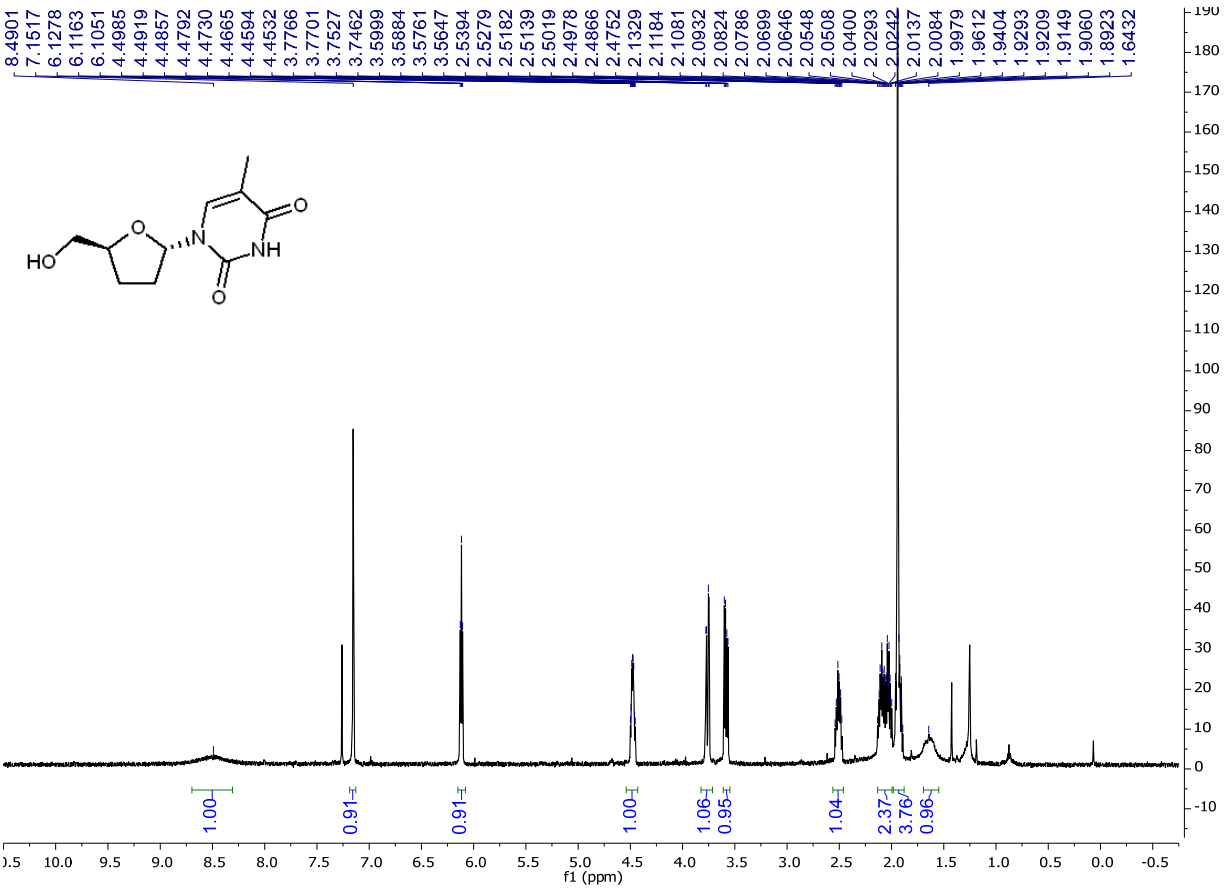


Figure S20. ^{1}H spectrum of **α-ddT** (500 MHz, $CDCl_3$).

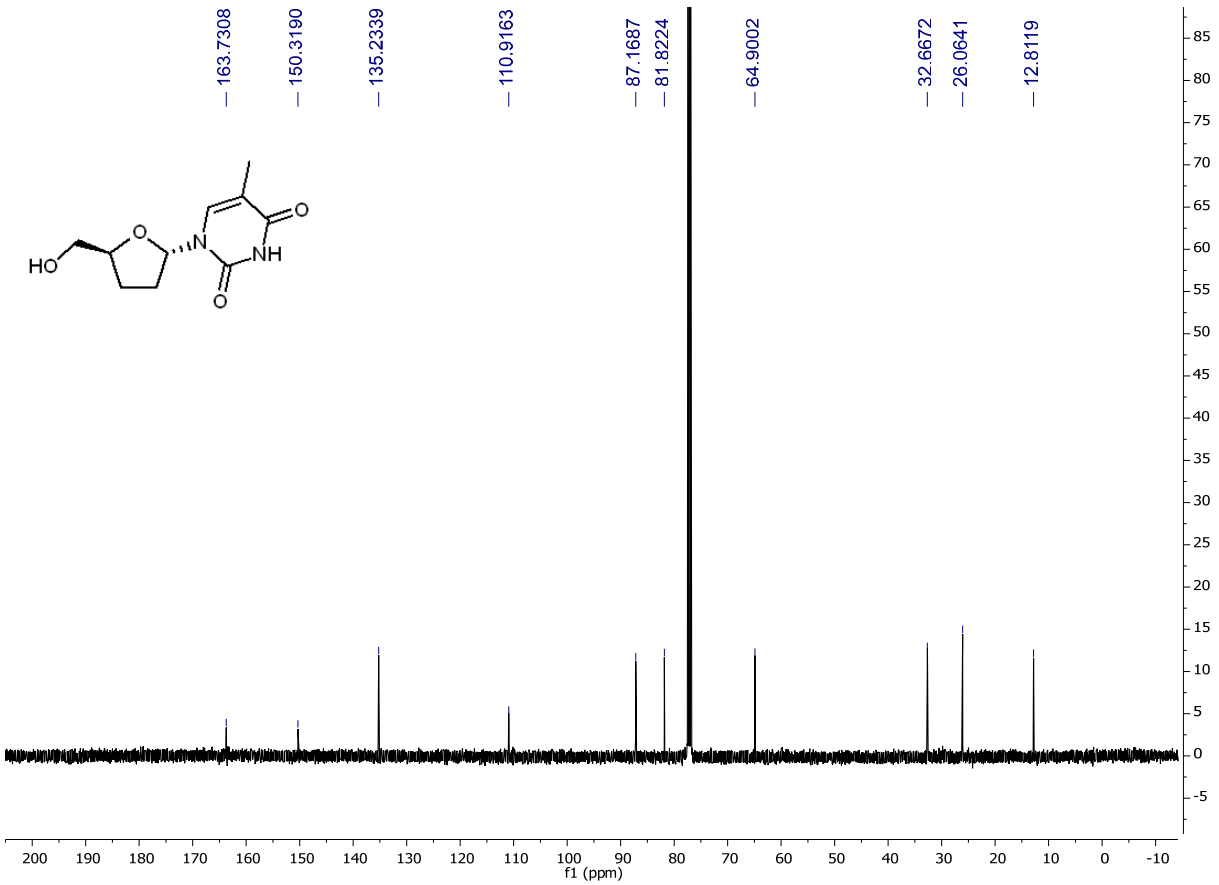


Figure S21. ^{13}C spectrum of **α-ddT** (125 MHz, $CDCl_3$).

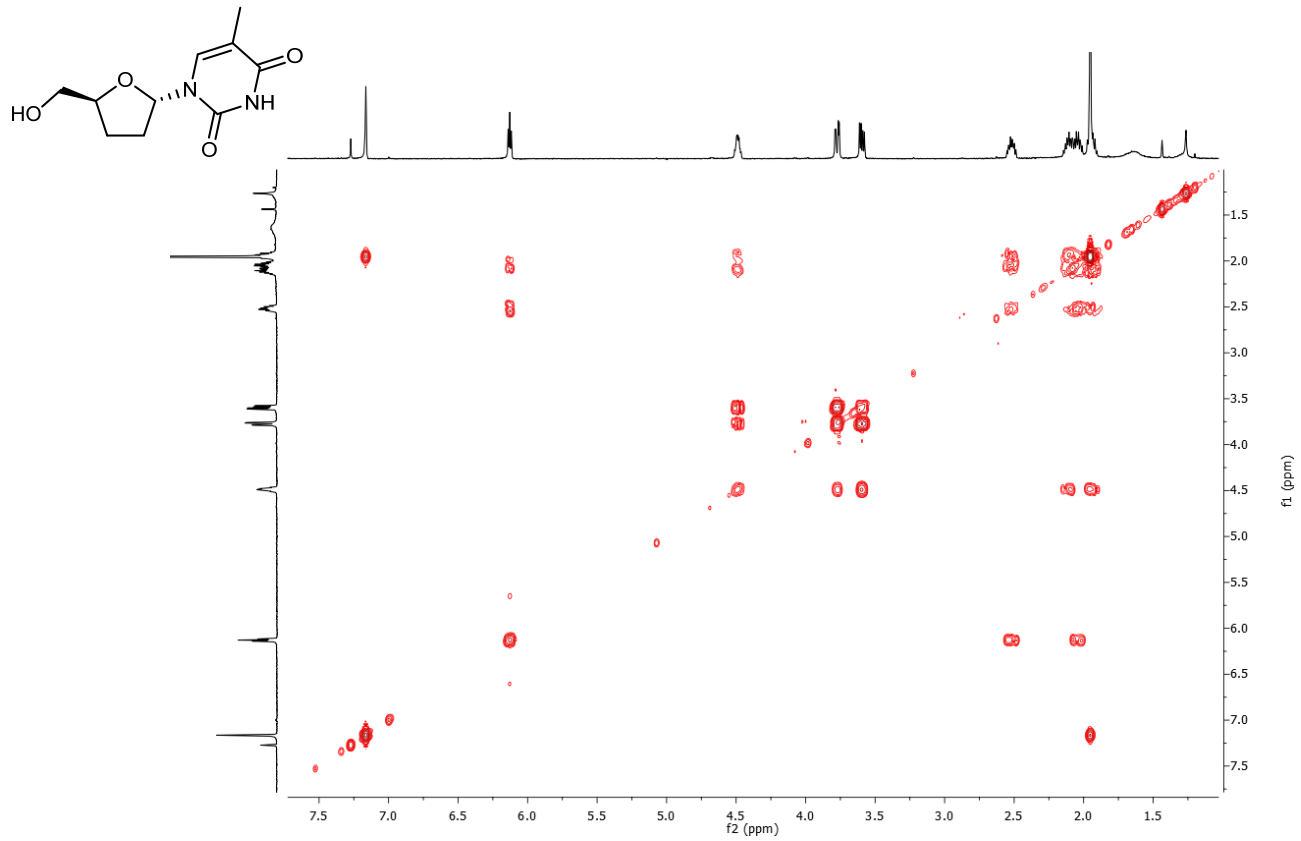


Figure S22. COSY spectrum of **α-ddT.**

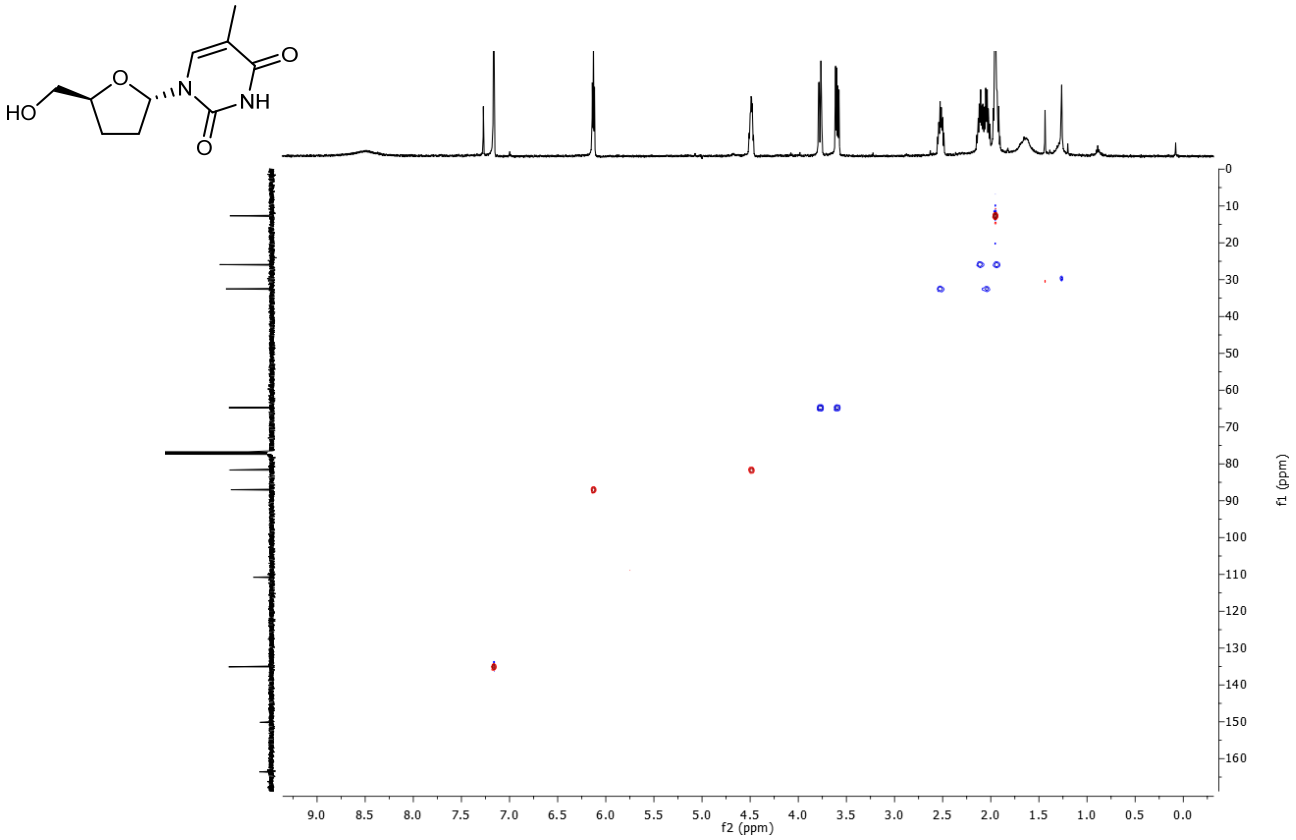


Figure S23. HSQC spectrum of **α-ddT.**

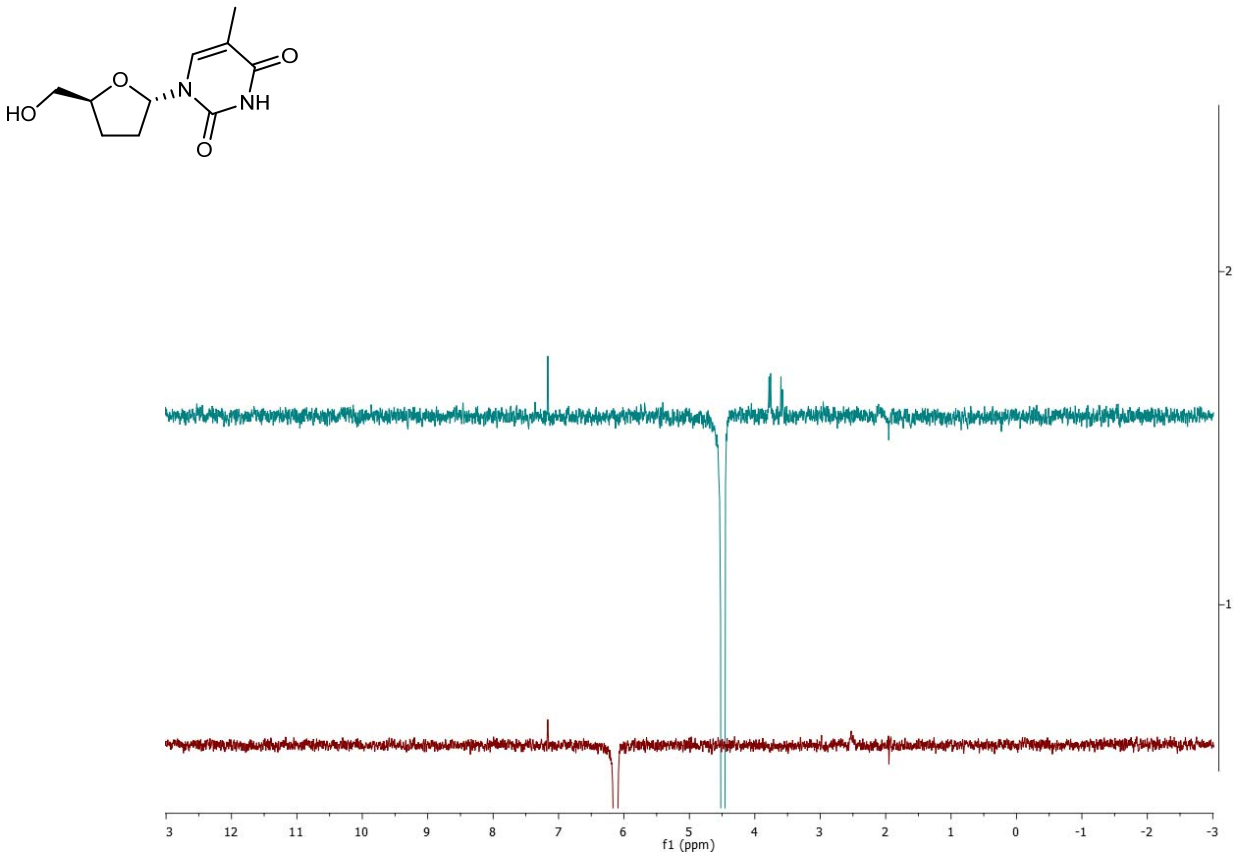


Figure S24. 1D-NOESY spectra of **α-ddT.**

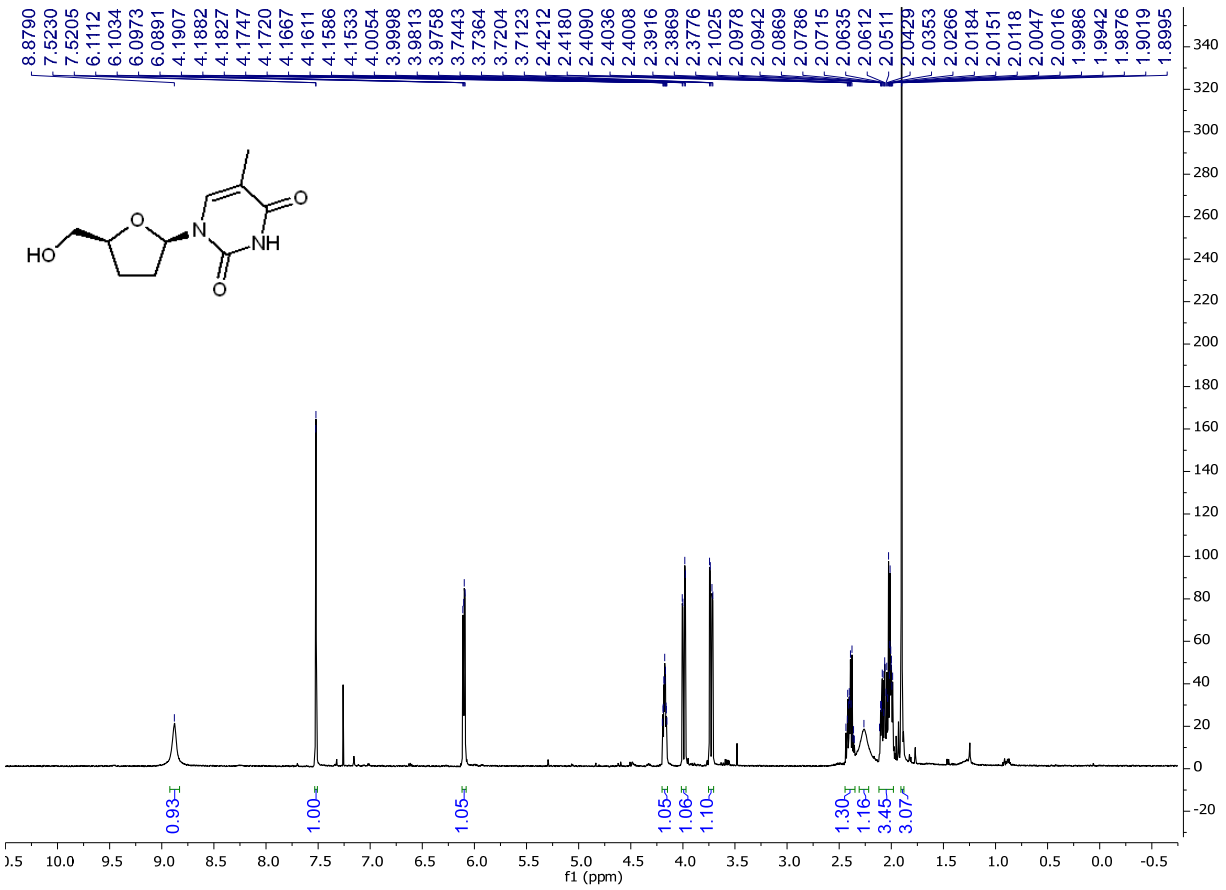


Figure S25. ^{1}H spectrum of **β-ddT** (500 MHz, $CDCl_3$).

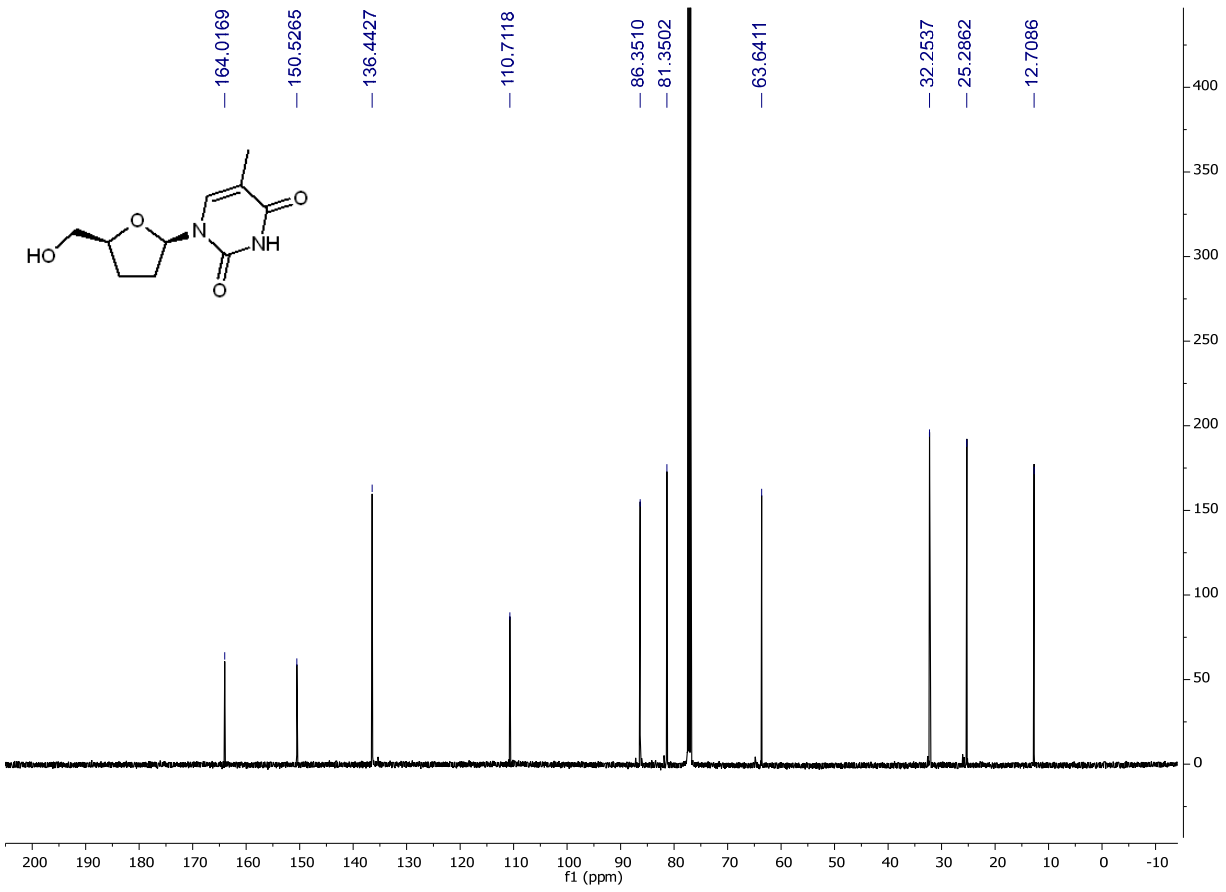


Figure S26. ^{13}C spectrum of **β-ddT** (125 MHz, $CDCl_3$).

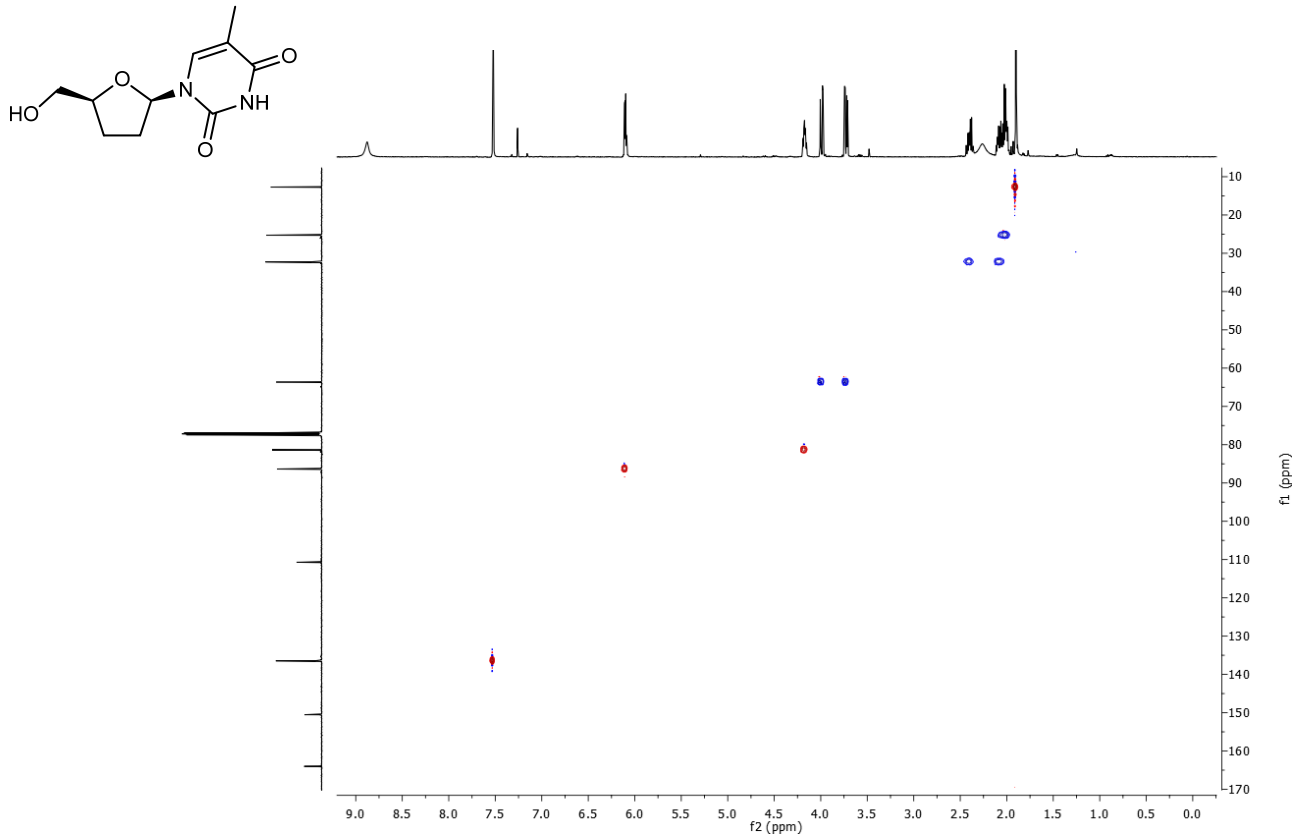


Figure S27. HSQC spectrum of **β-ddT.**

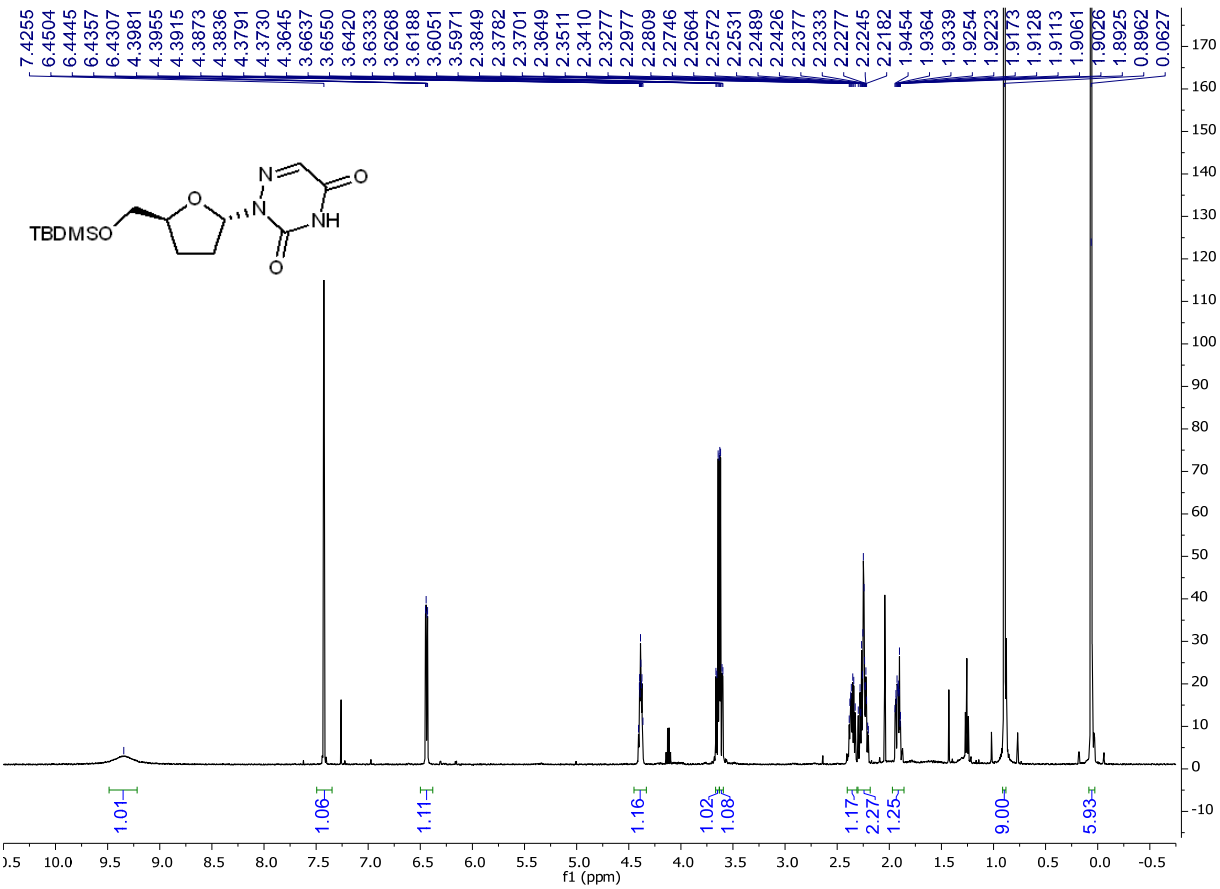


Figure S28. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddAU** (500 MHz, $CDCl_3$)**.**

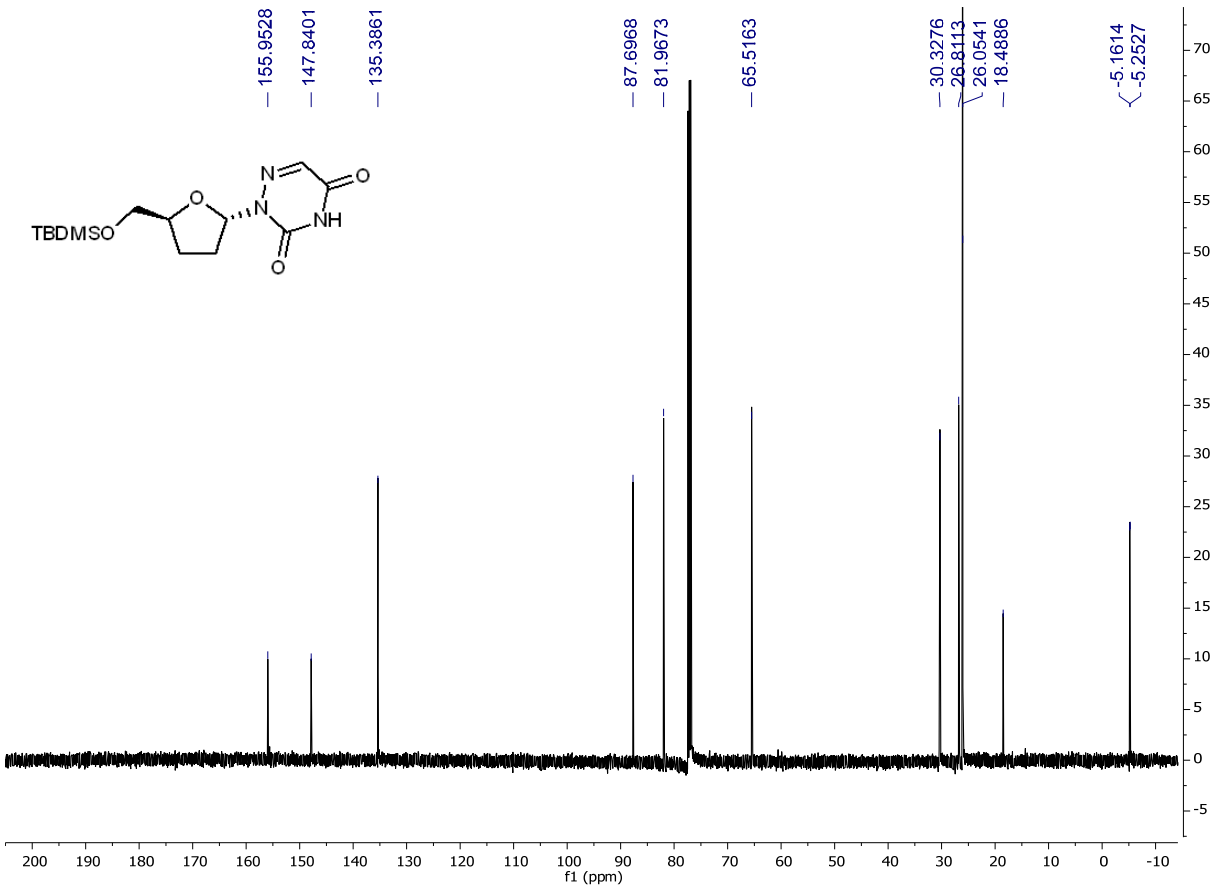


Figure S29. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddAU** (125 MHz, $CDCl_3$).

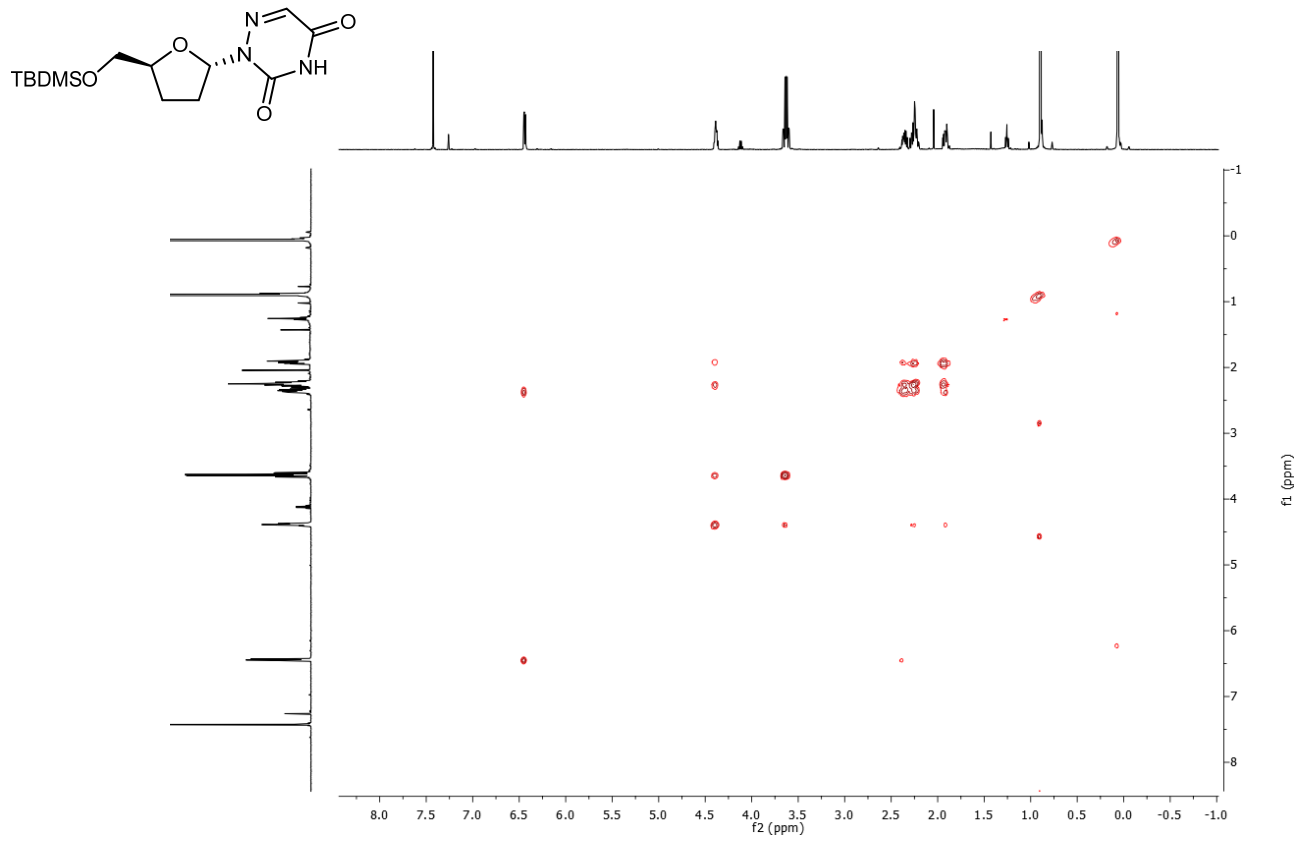


Figure S30. COSY spectrum of **5'-*O*-silyl-protected α-ddAU.**

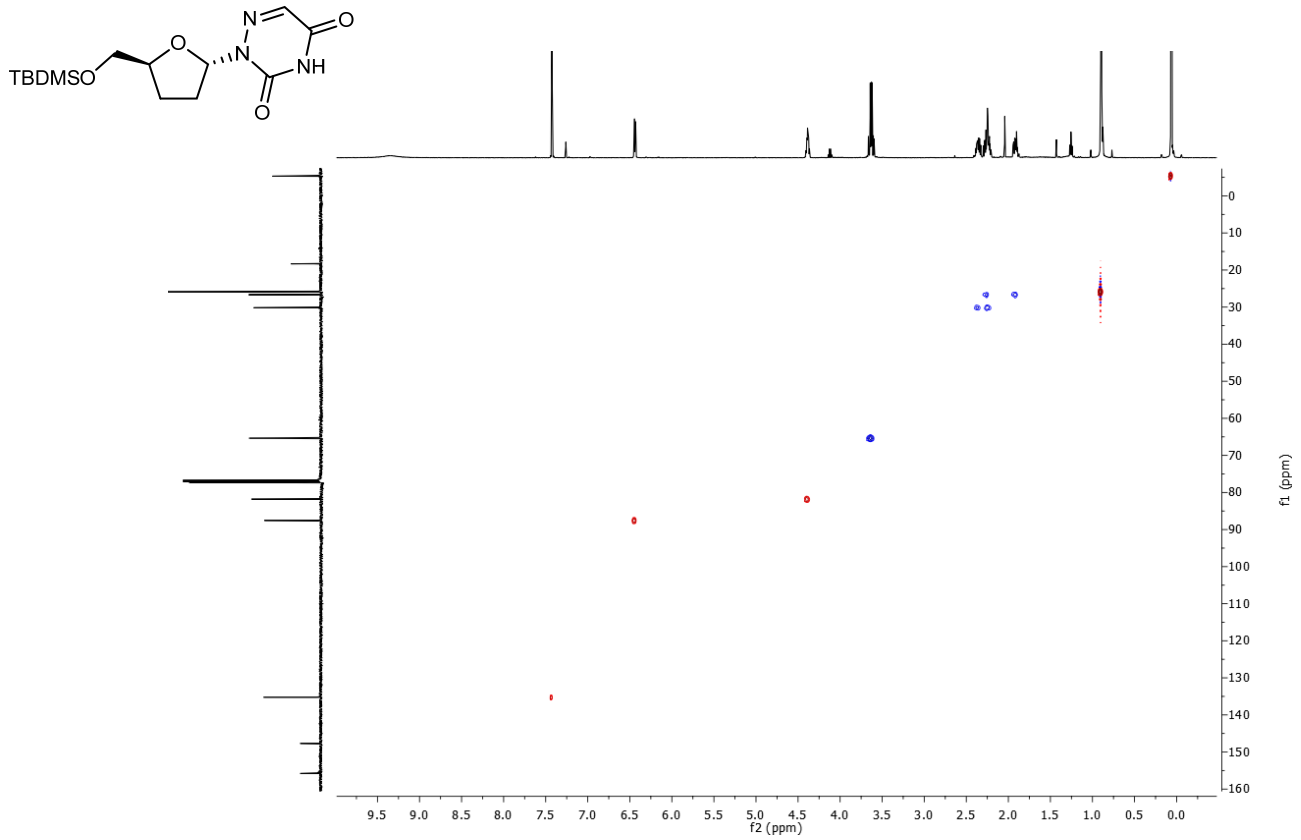


Figure S31. HSQC spectrum of **5'-*O*-silyl-protected α-ddAU.**

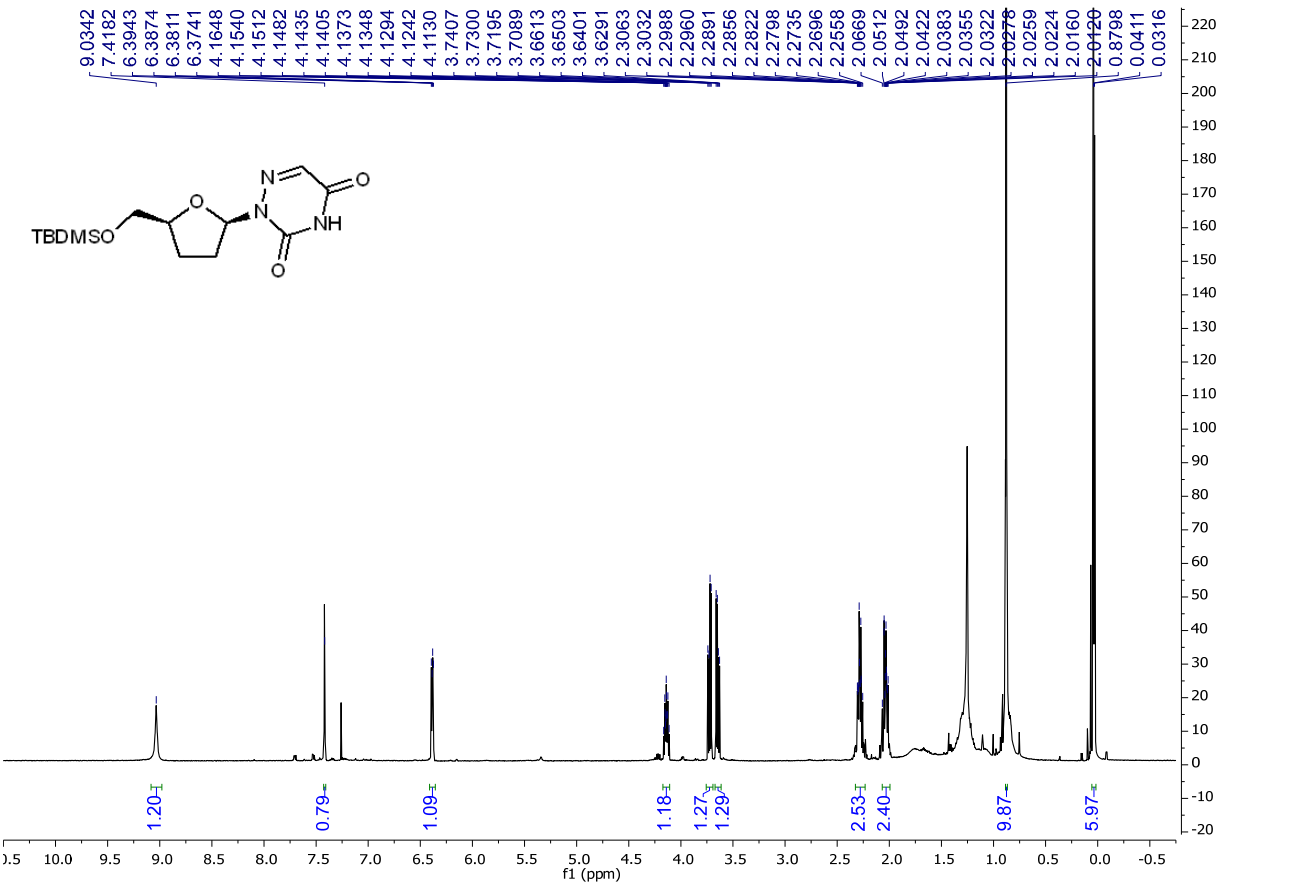


Figure S32. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddAU** (500 MHz, $CDCl_3$).

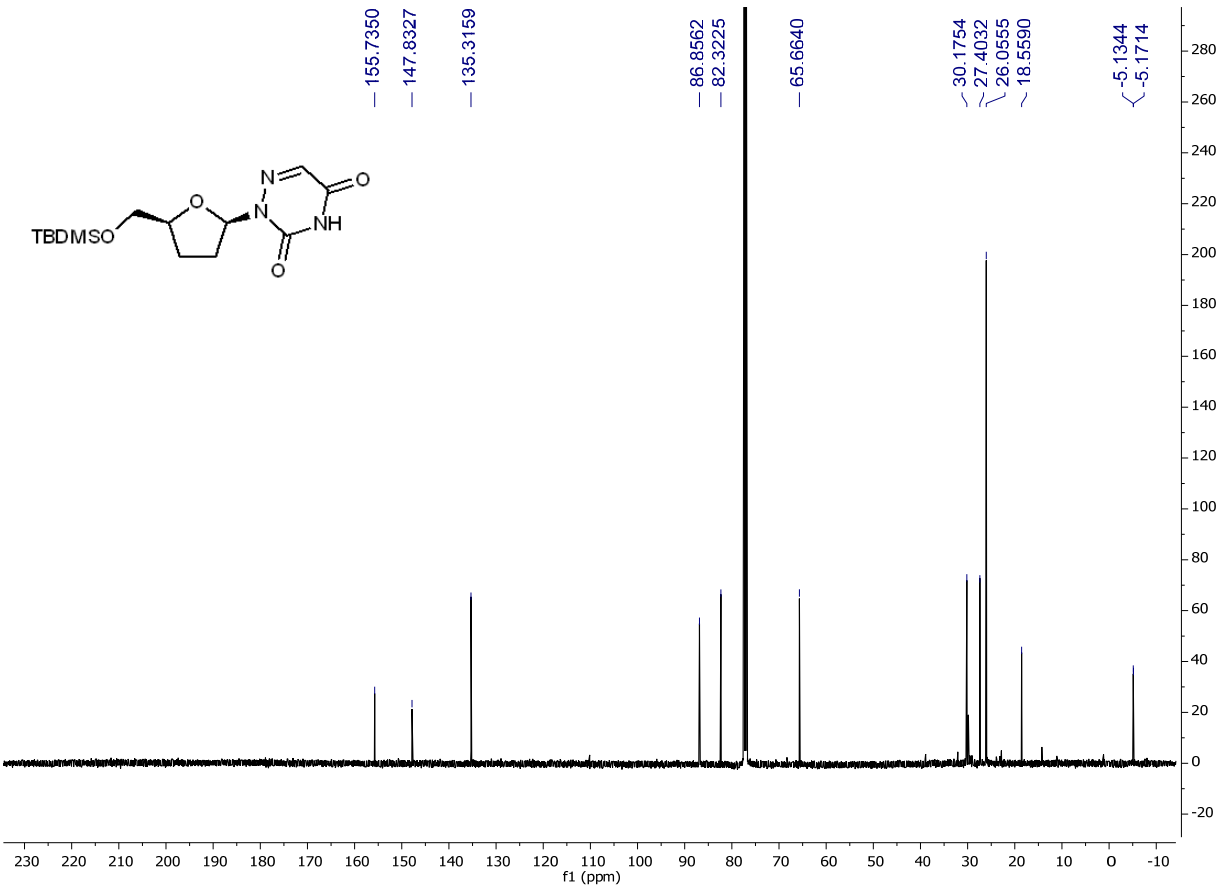


Figure S33. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddAU** (125 MHz, $CDCl_3$)**.**

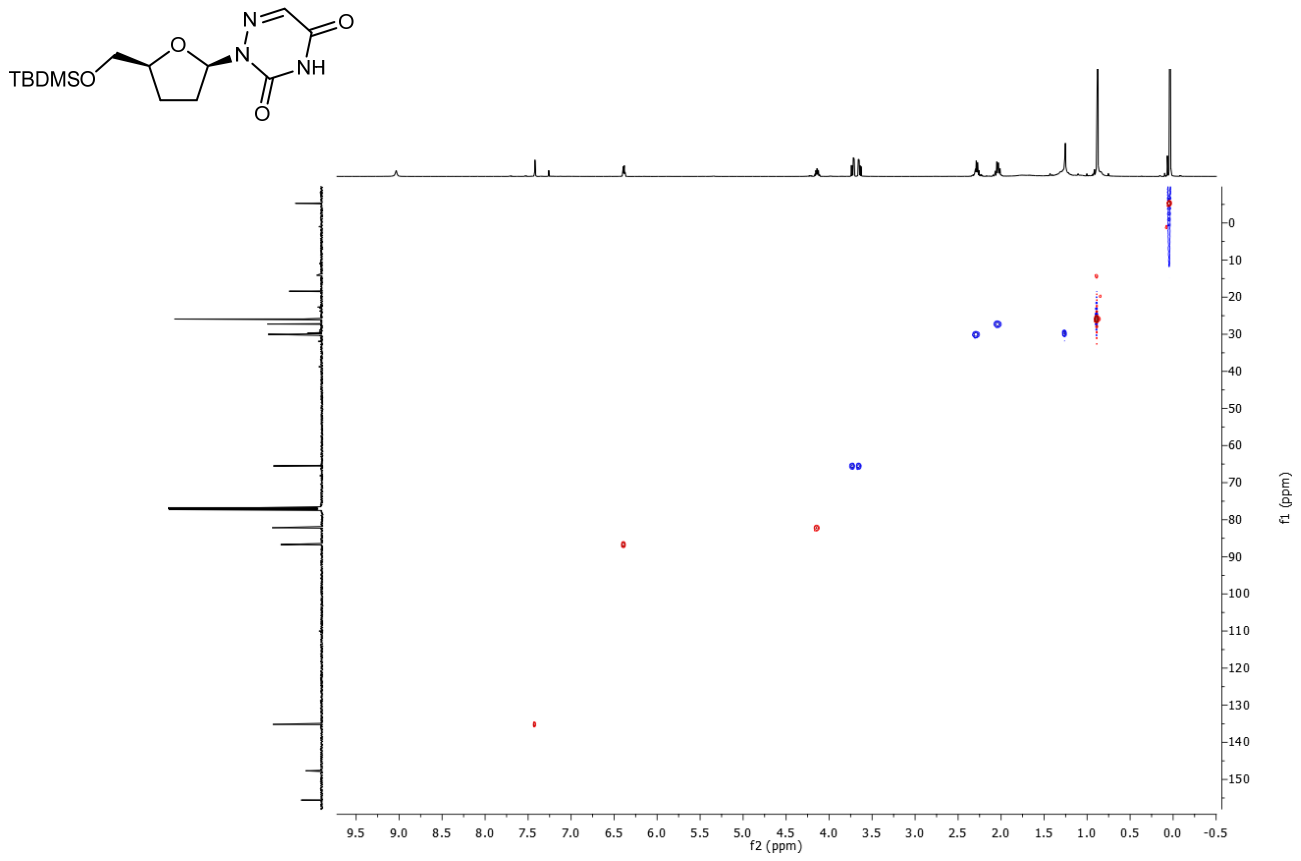


Figure S34. HSQC spectrum of **5'-*O*-silyl-protected β-ddAU.**

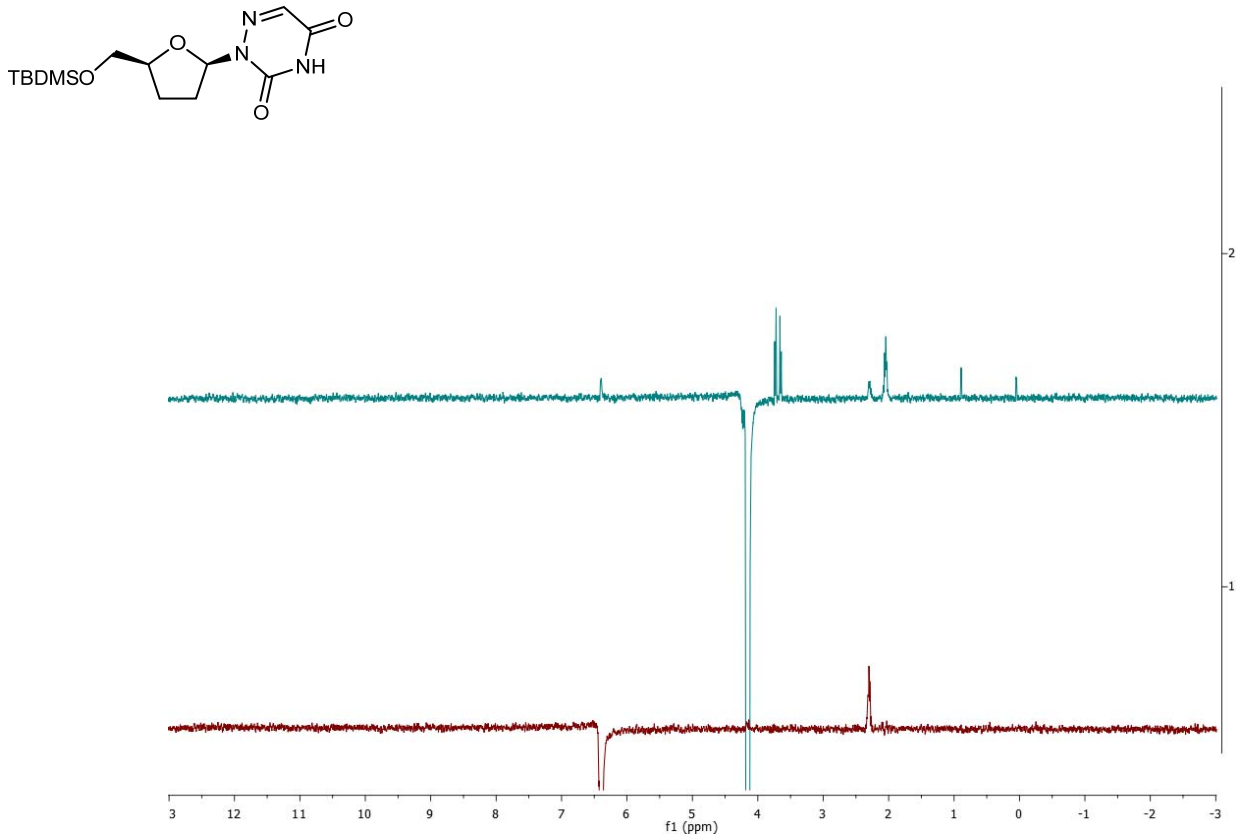


Figure S35. 1D-NOESY spectra of **5'-*O*-silyl-protected β-ddAU.**

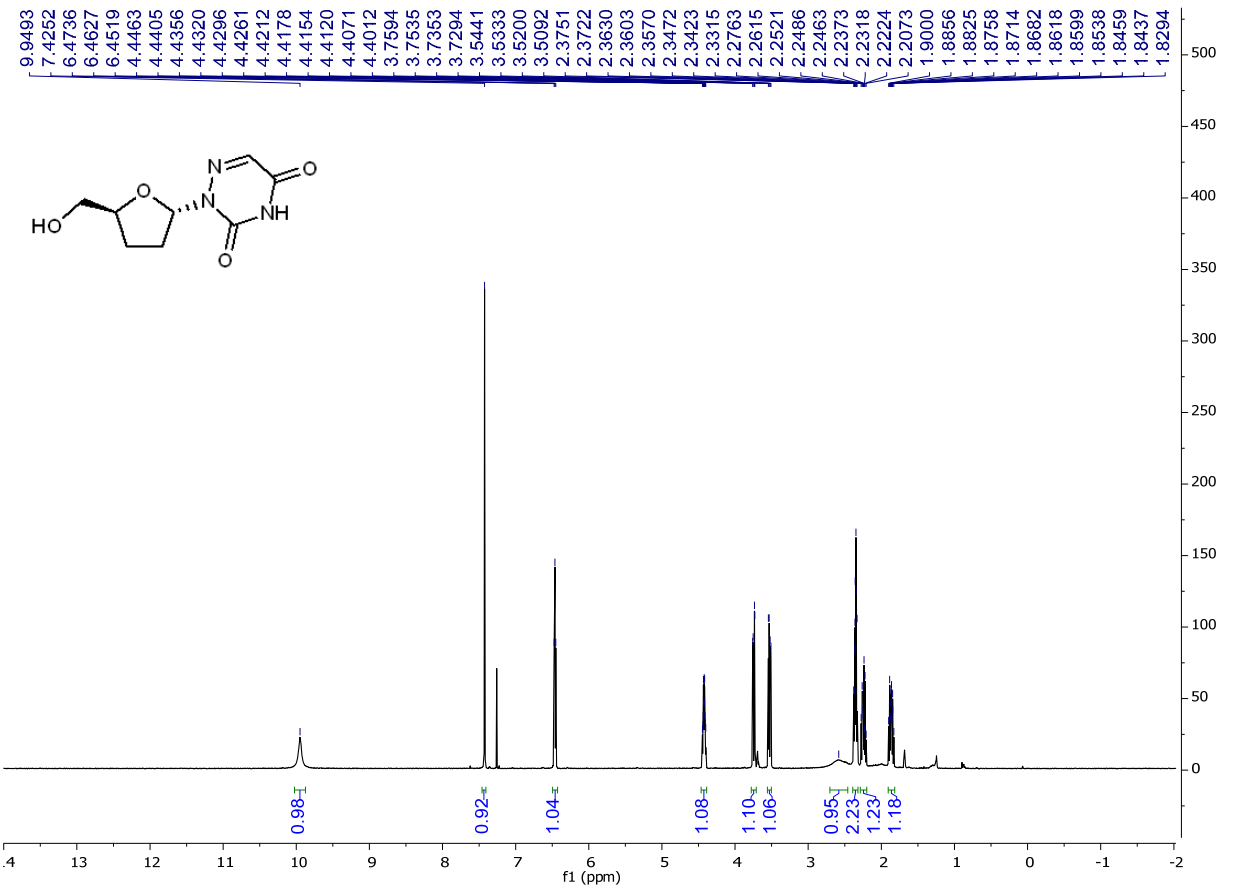


Figure S36. ^{1}H spectrum of **α-ddAU** (500 MHz, $CDCl_3$).

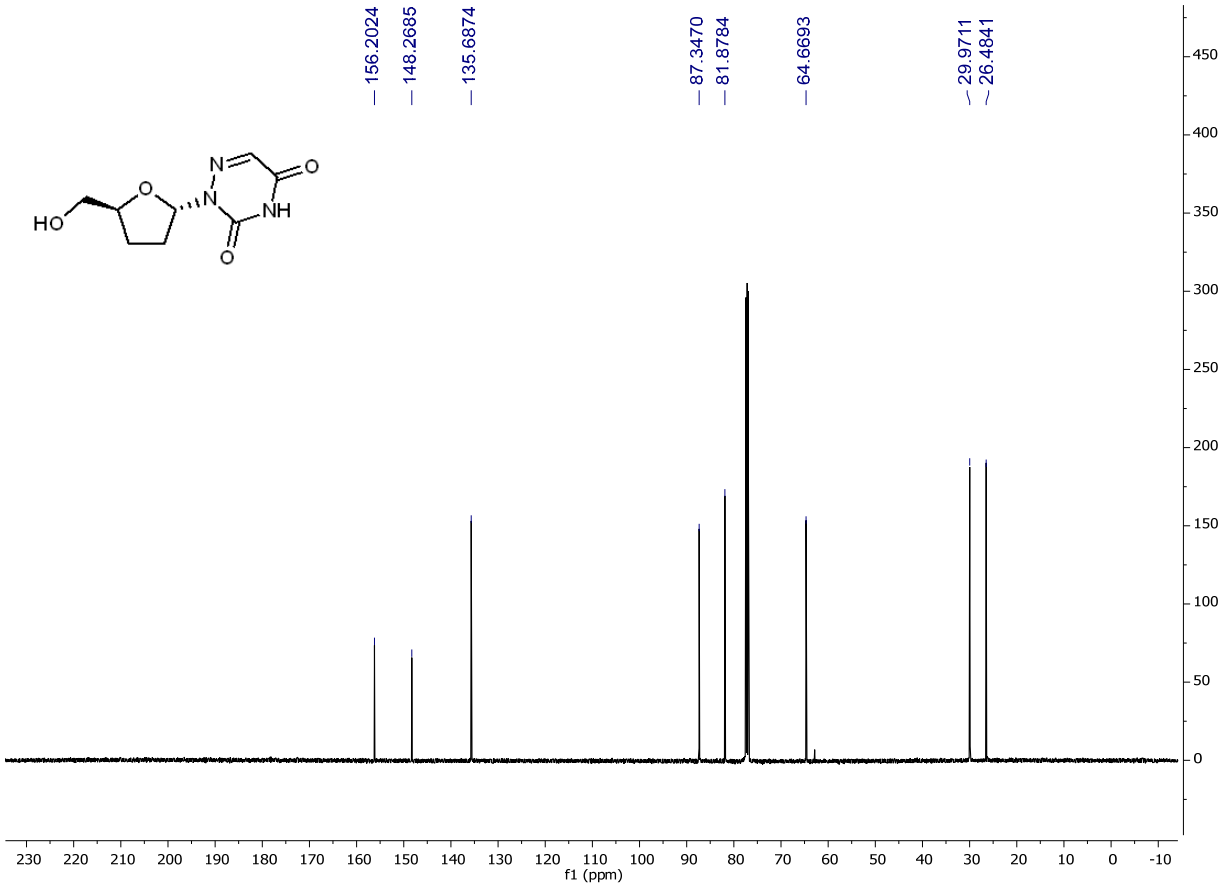


Figure S37. ^{13}C spectrum of **α-ddAU** (125 MHz, $CDCl_3$).

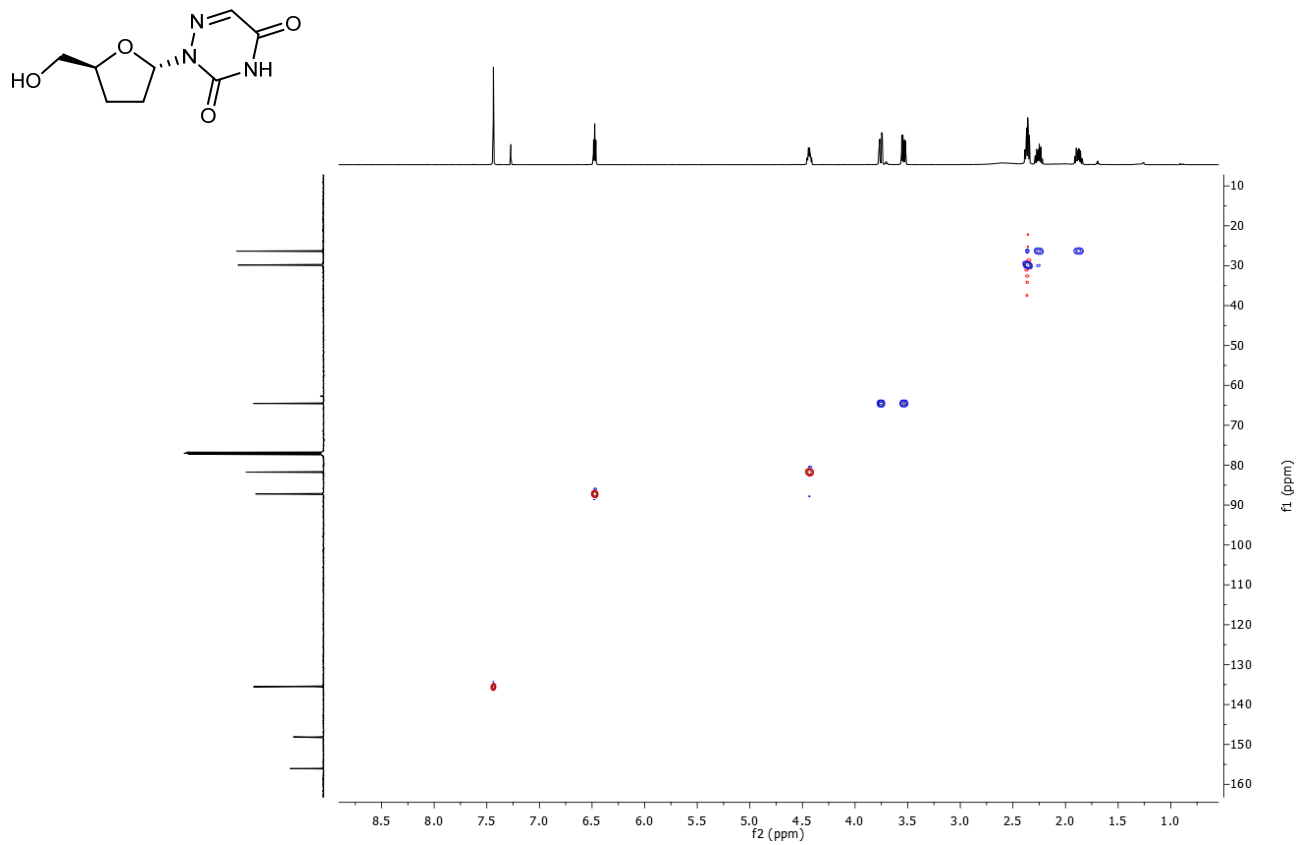


Figure S38. HSQC spectrum of **α-ddAU.**

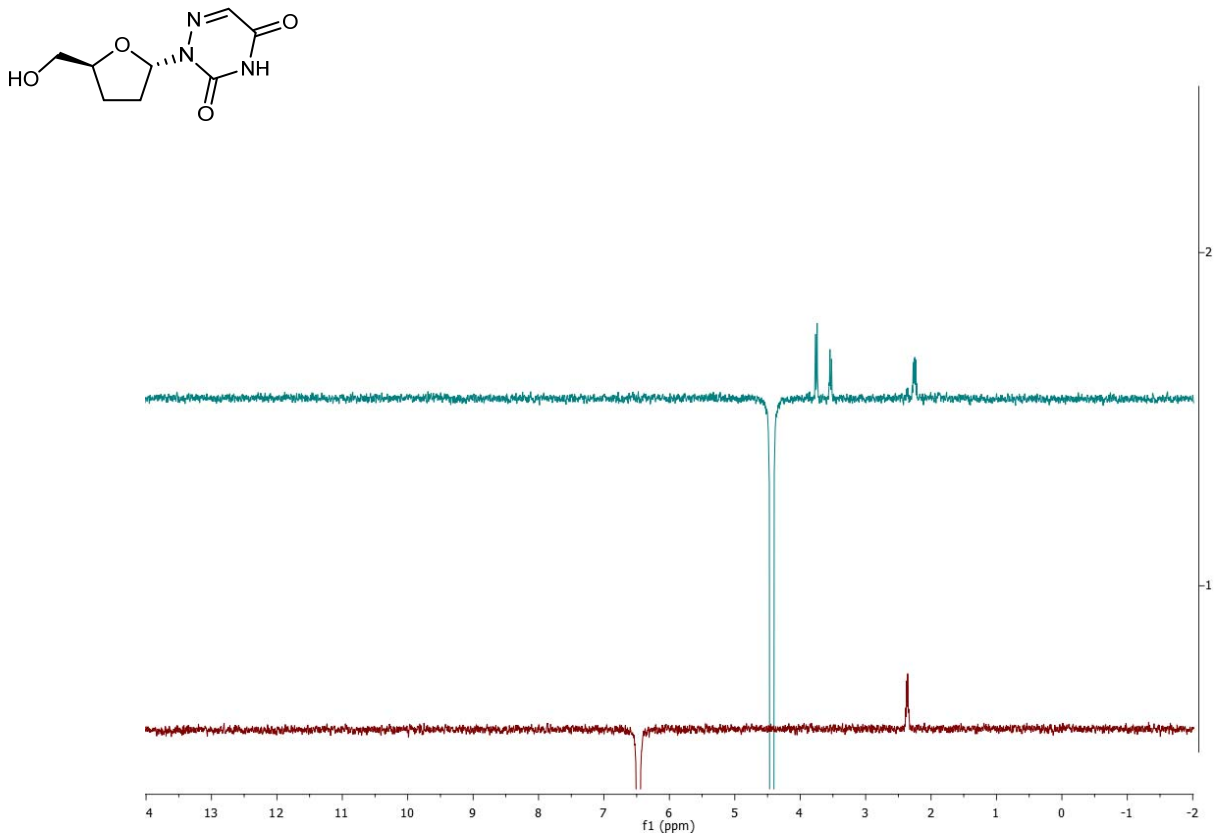


Figure S39. 1D-NOESY spectra of **α-ddAU.**

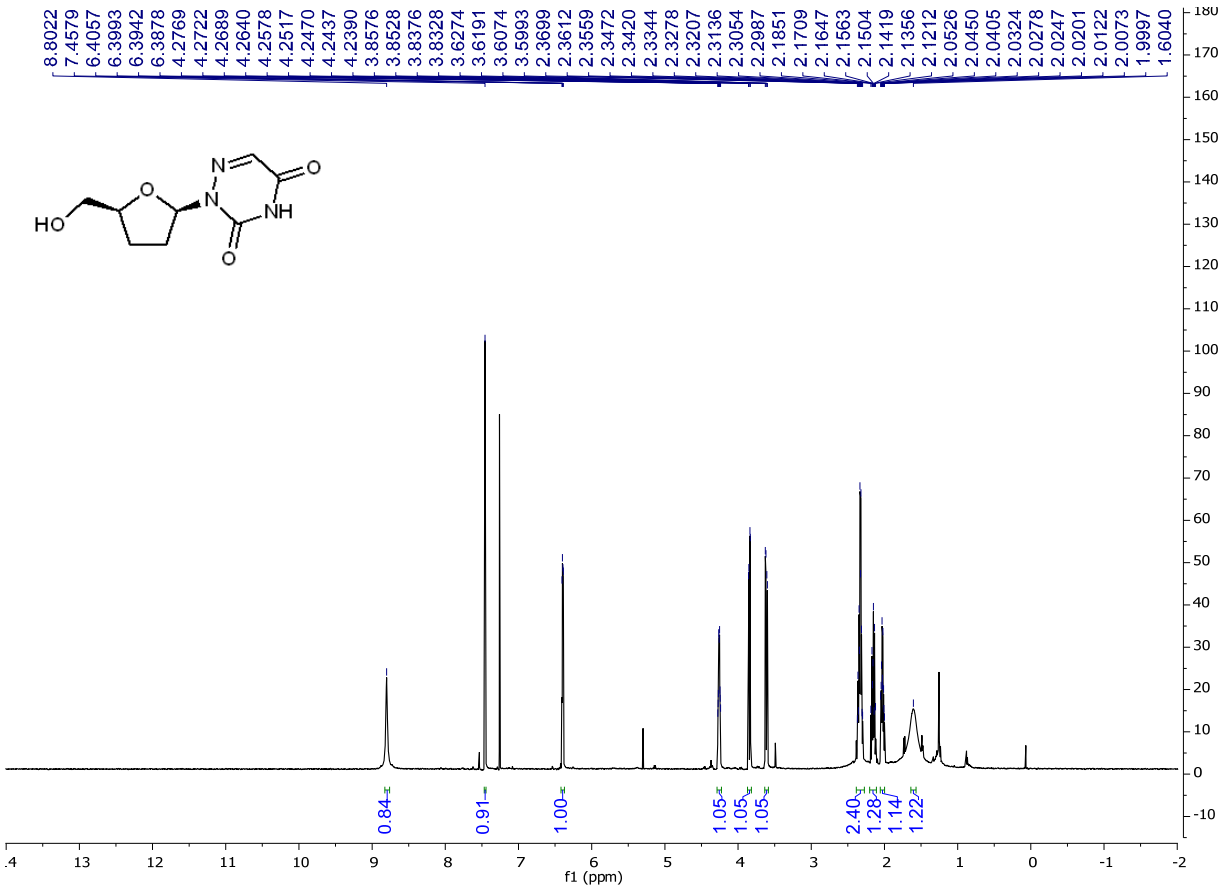


Figure S40. ^{1}H spectrum of **β-ddAU** (600 MHz, $CDCl_3$).

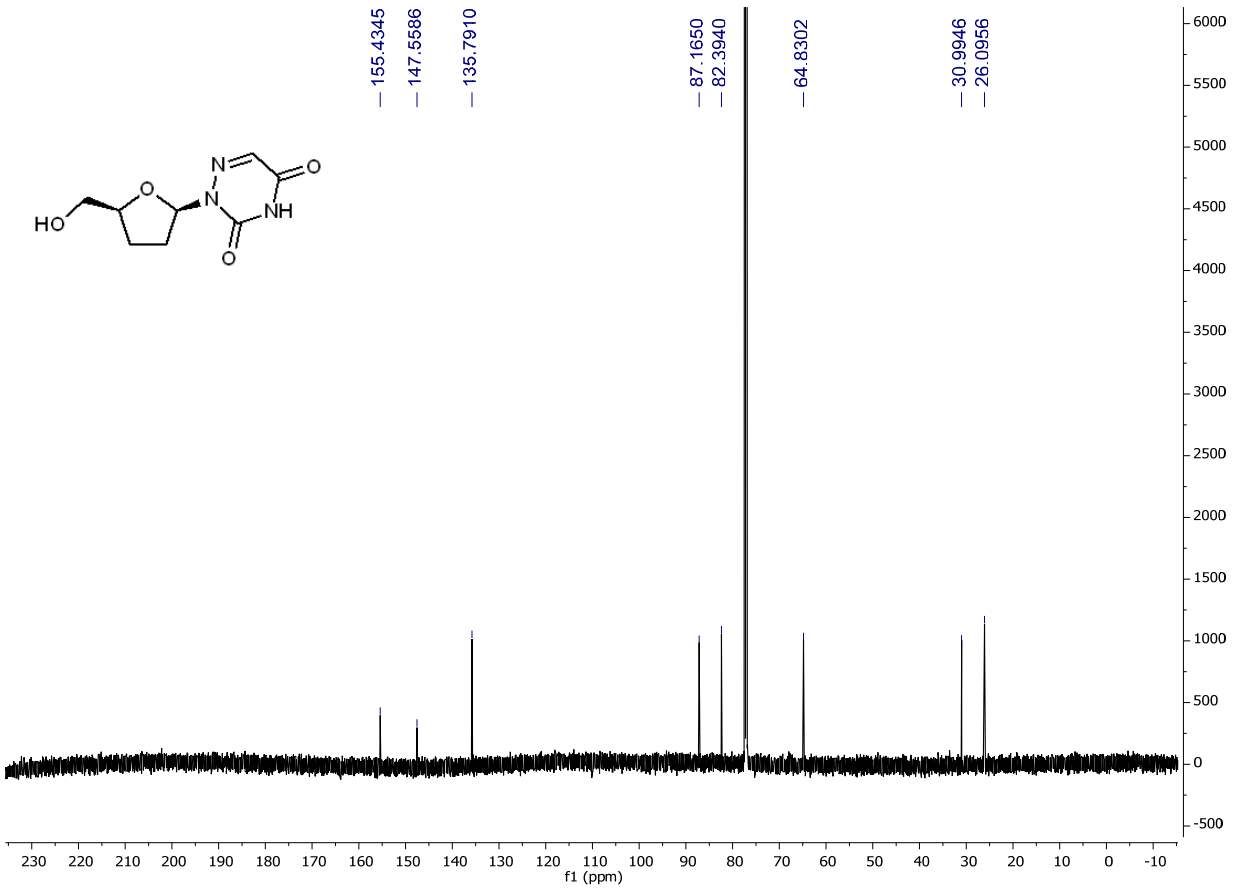


Figure S41. ^{13}C spectrum of **β-ddAU** (150 MHz, $CDCl_3$)**.**

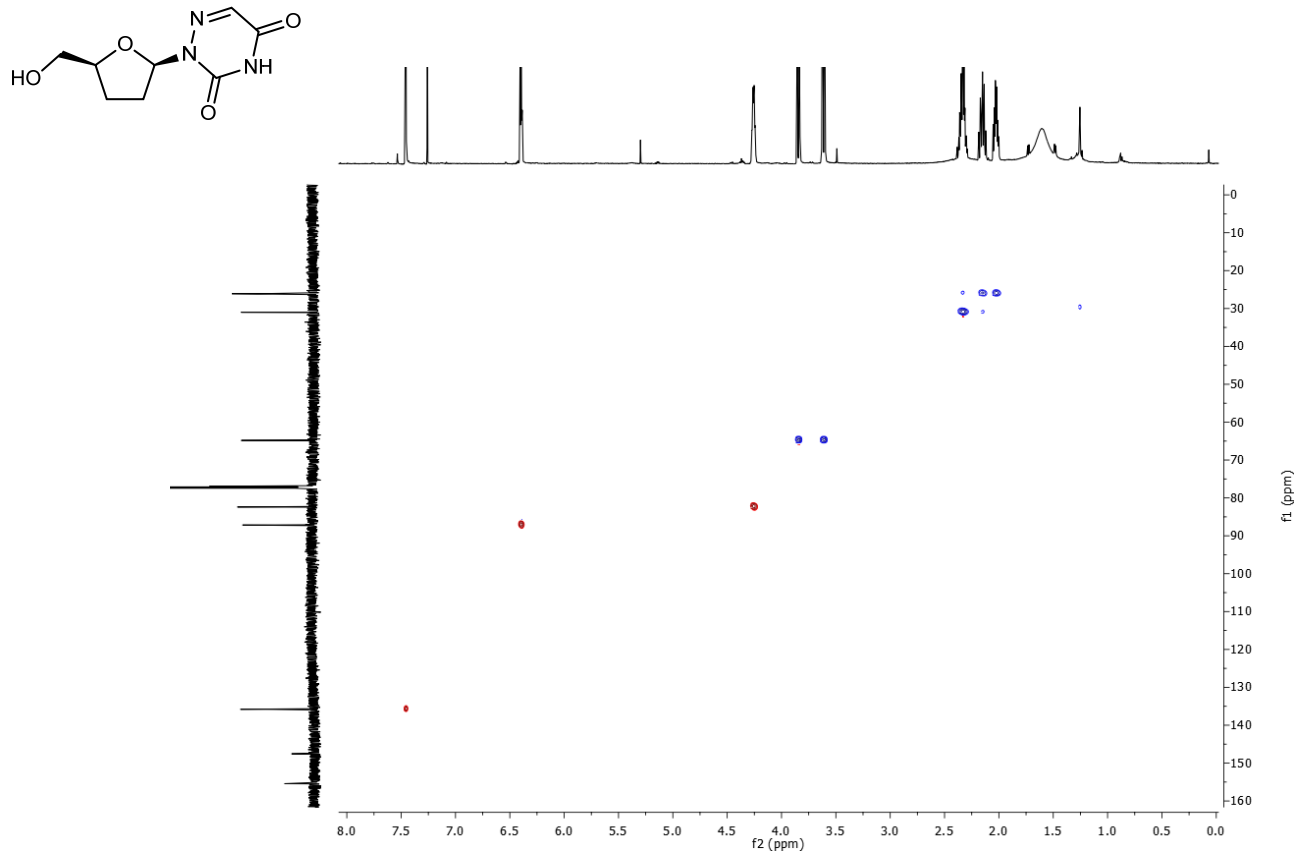


Figure S42. HSQC spectrum of **β-ddAU.**

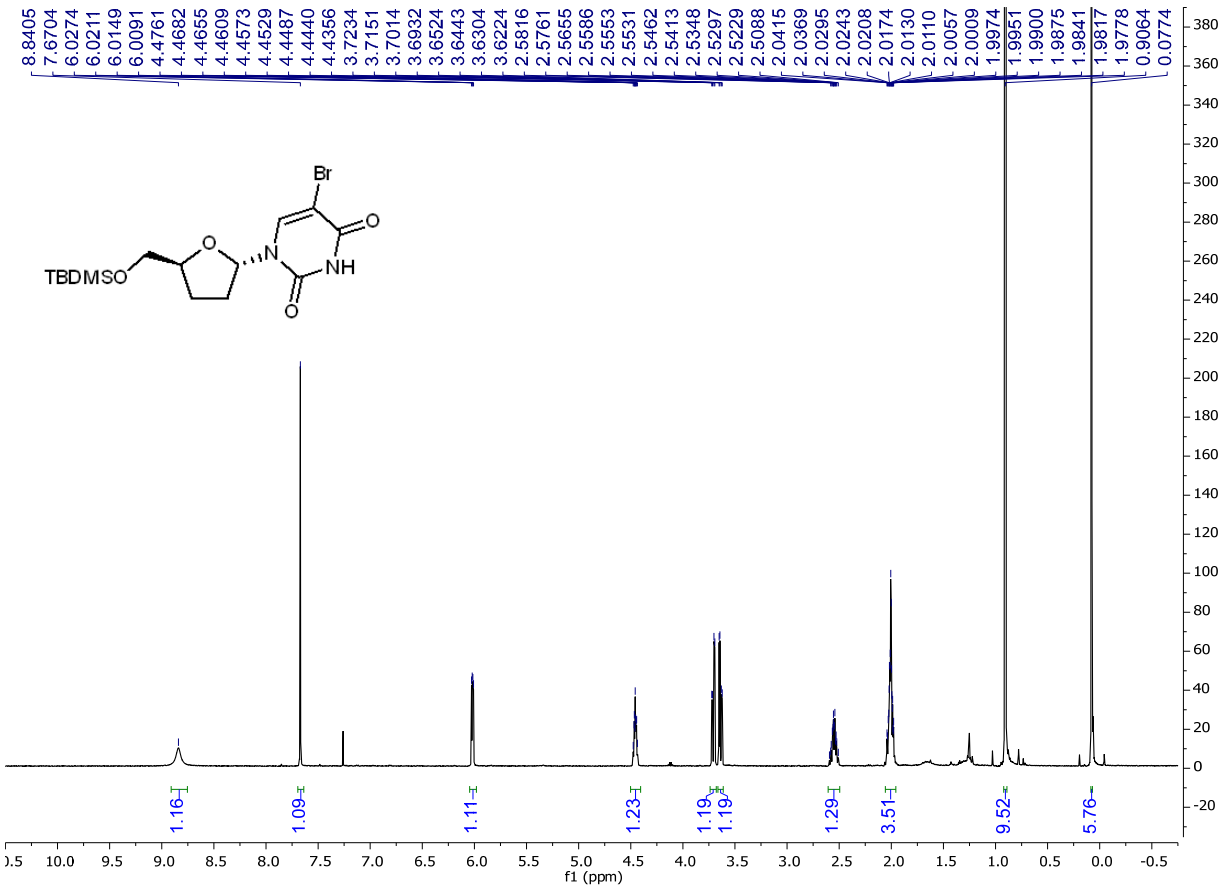


Figure S43. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddBrU** (500 MHz, $CDCl_3$).

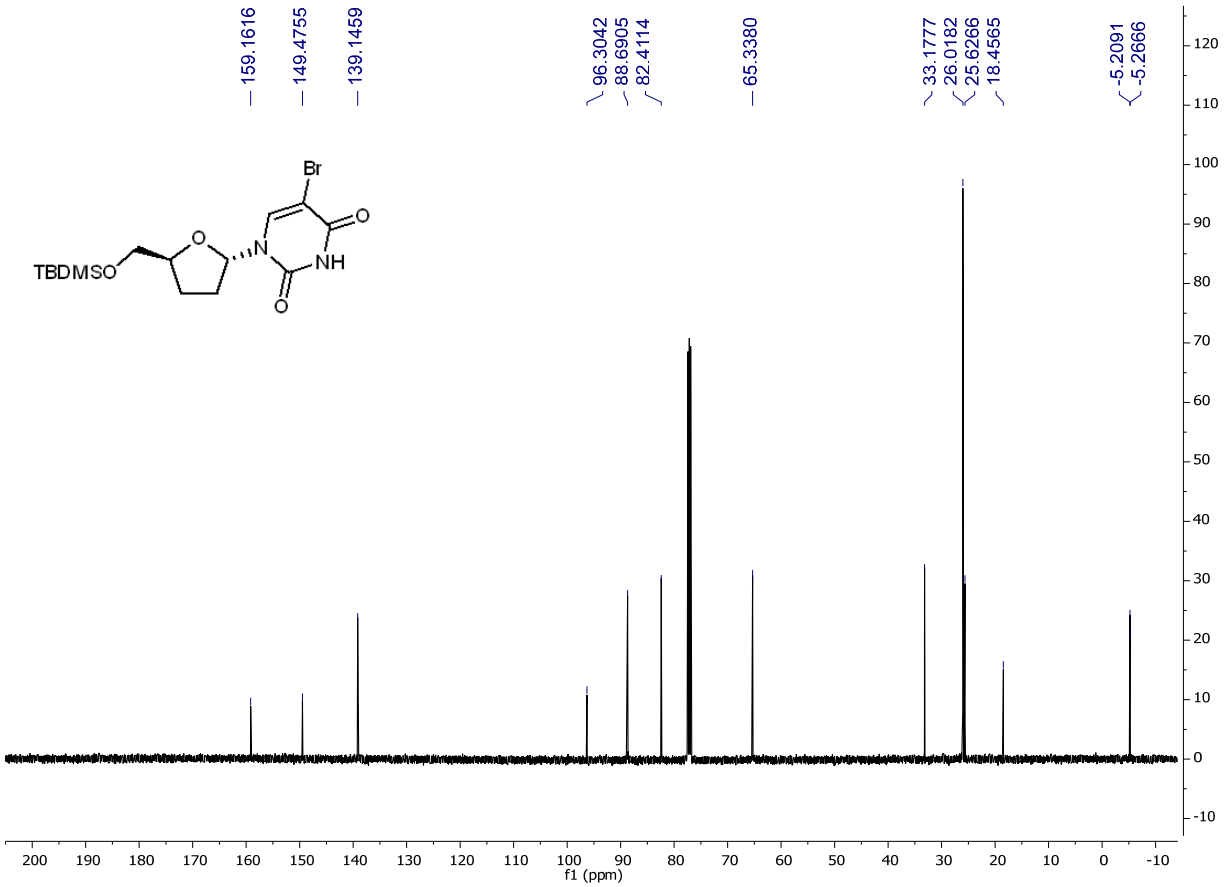


Figure S44. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddBrU** (125 MHz, $CDCl_3$).

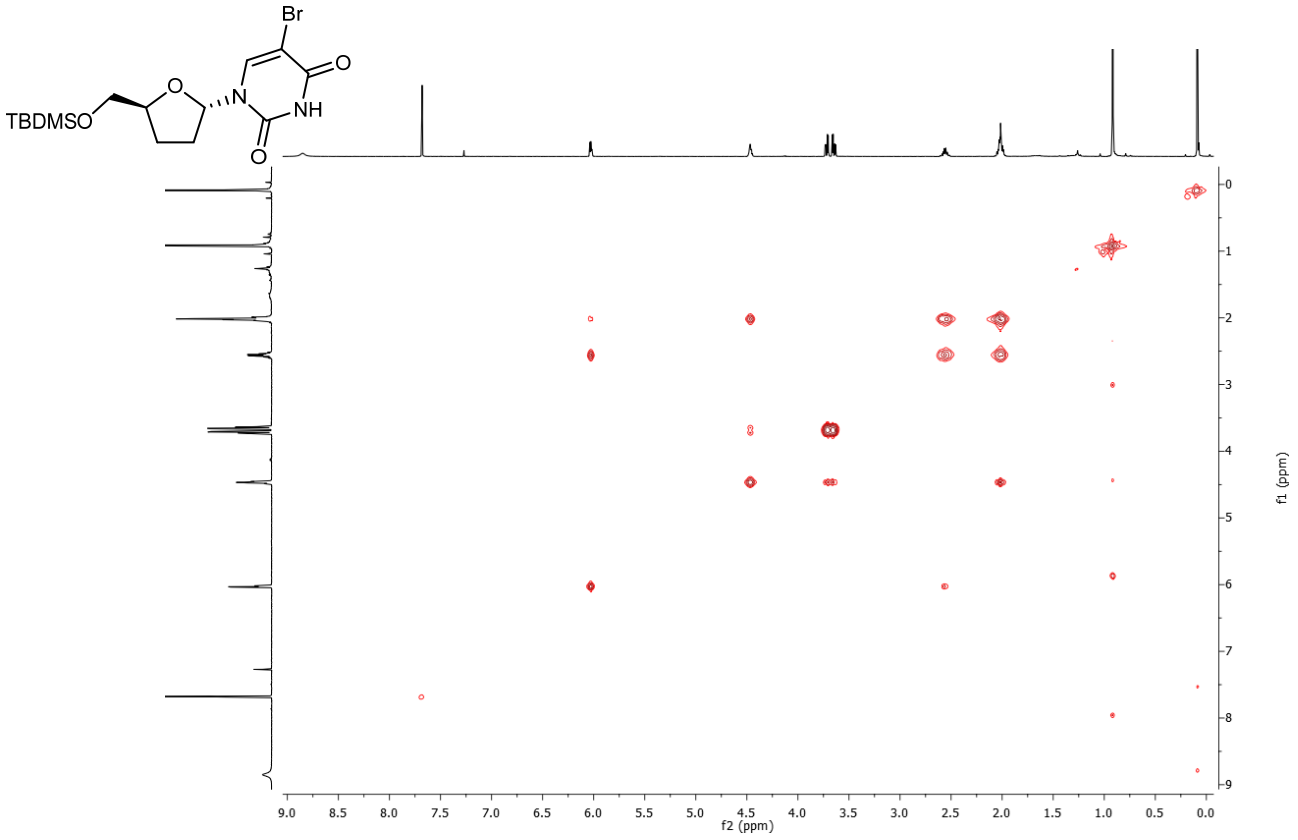


Figure S45. COSY spectrum of **5'-*O*-silyl-protected α-ddBrU**.

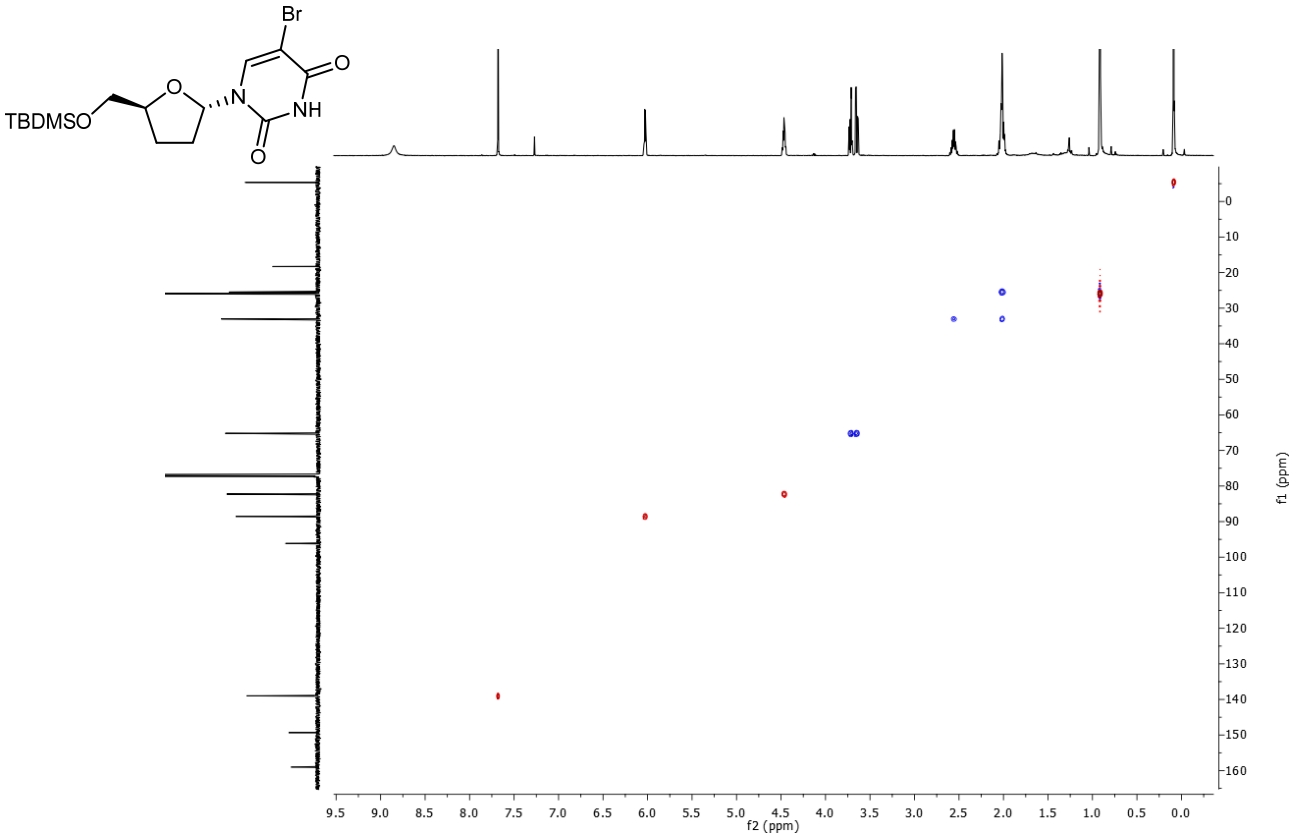


Figure S46. HSQC spectrum of **5'-*O*-silyl-protected α-ddBrU**.

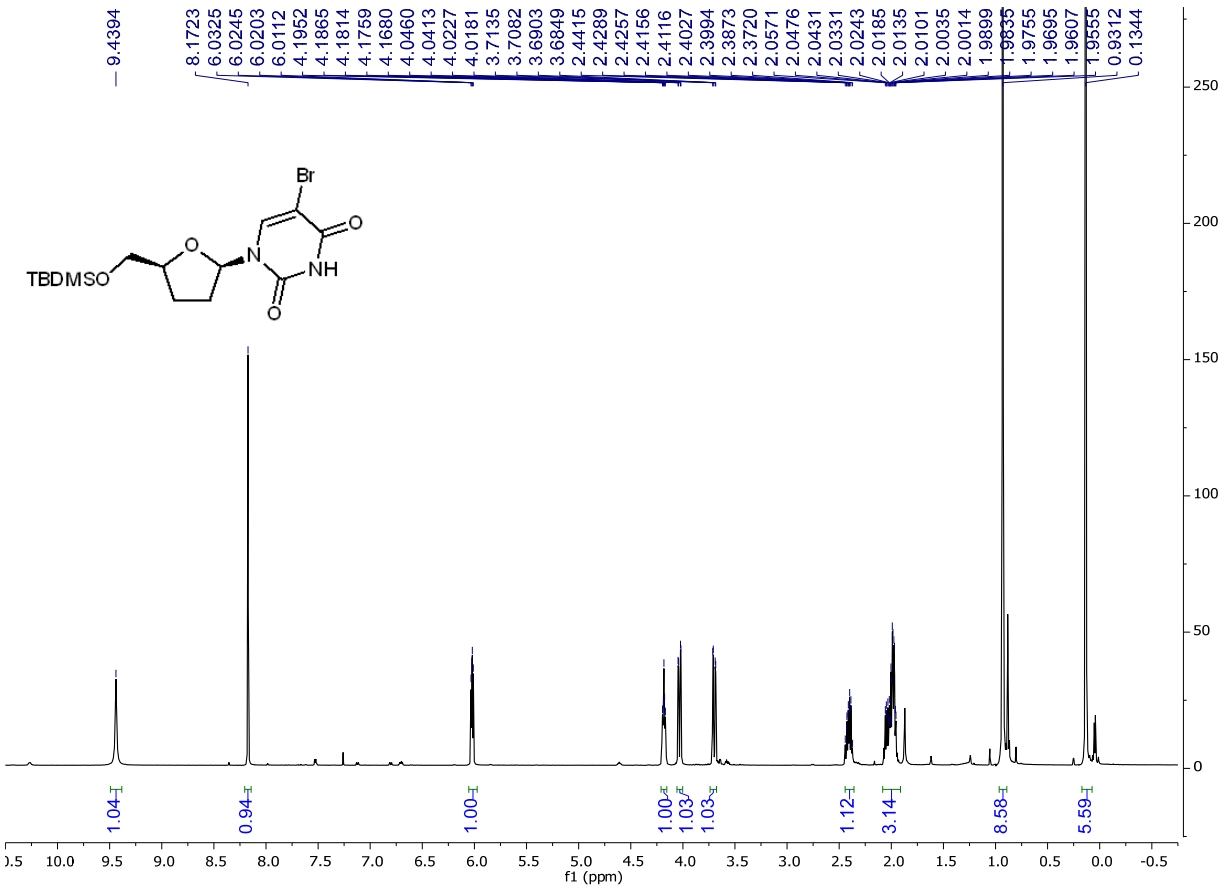


Figure S47. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddBrU** (500 MHz, $CDCl_3$).

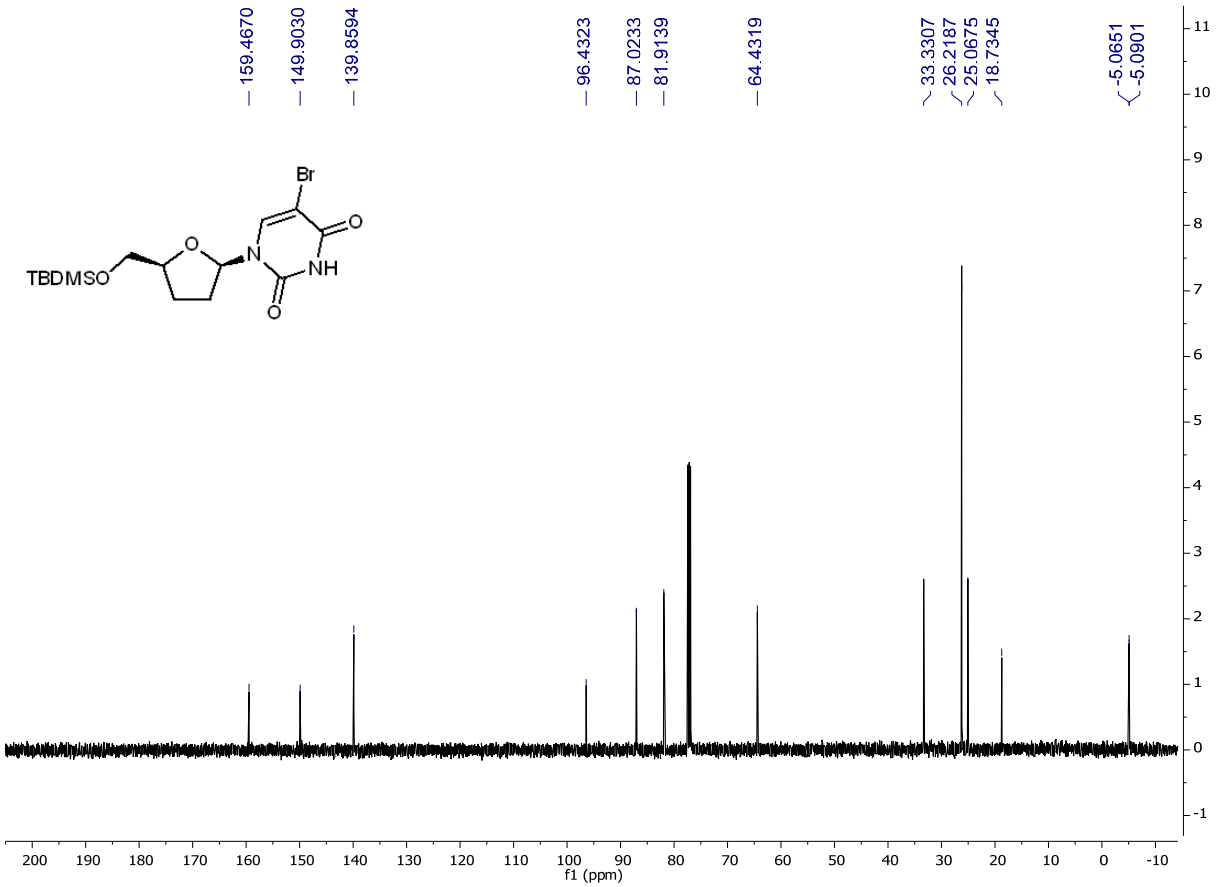


Figure S48. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddBrU** (125 MHz, $CDCl_3$).

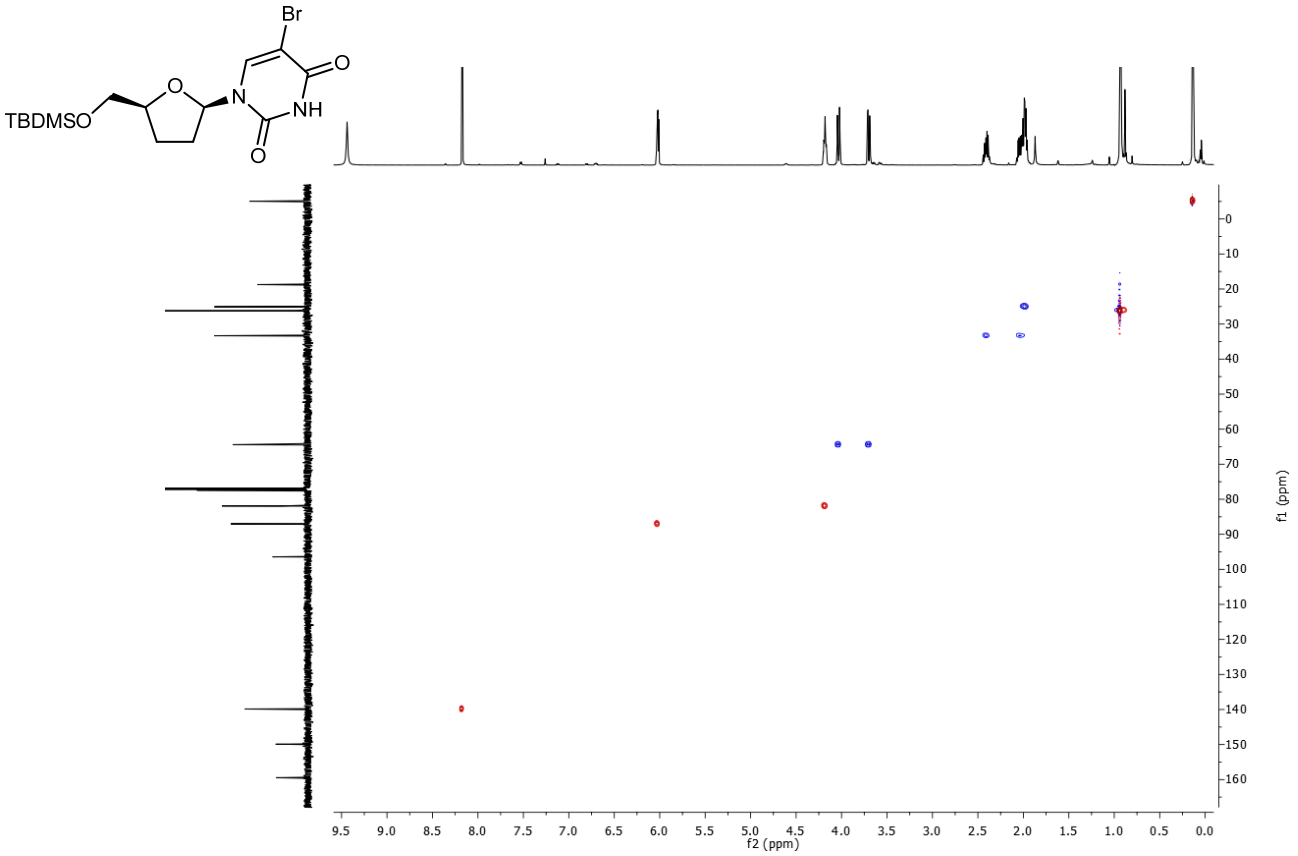


Figure S49. HSQC spectrum of **5'-*O*-silyl-protected β-ddBrU**.

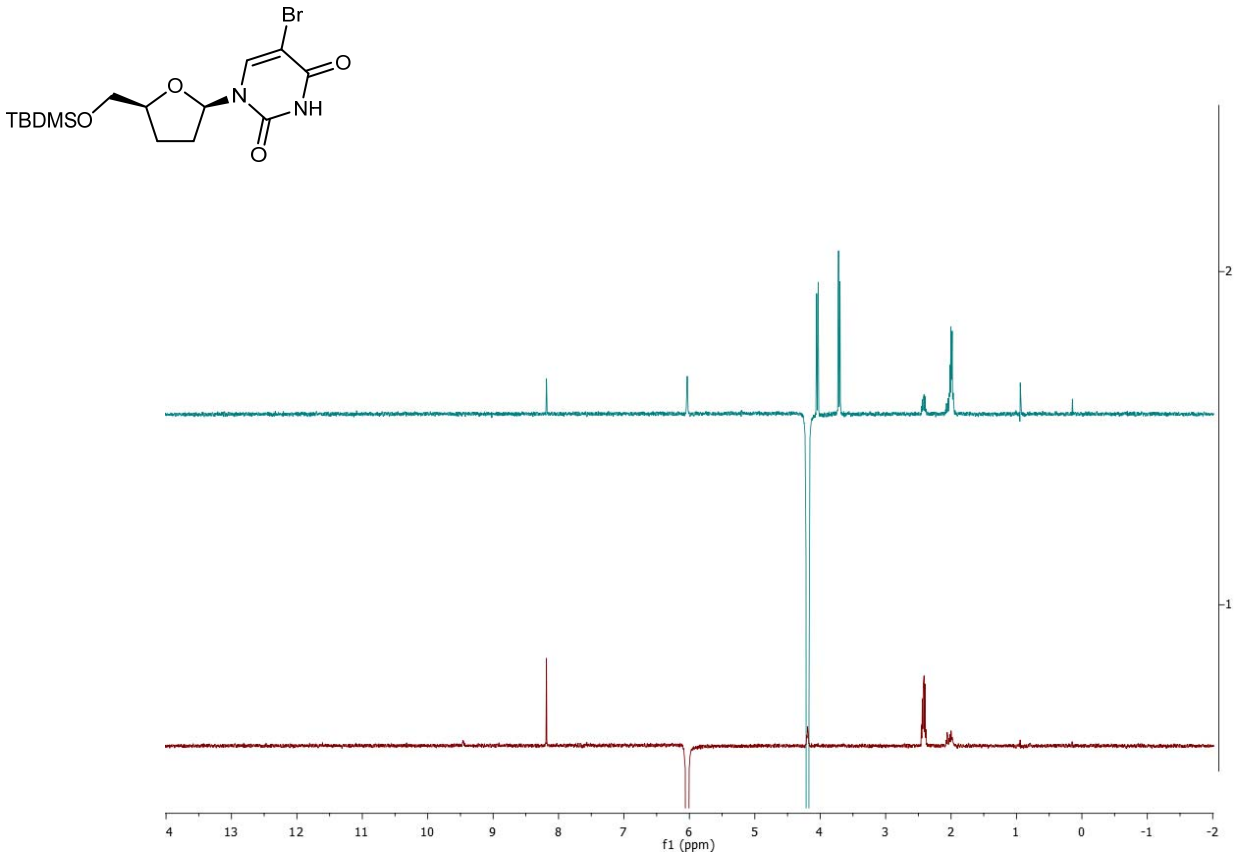


Figure S50. 1D-NOESY spectra of **5'-*O*-silyl-protected β-ddBrU**.

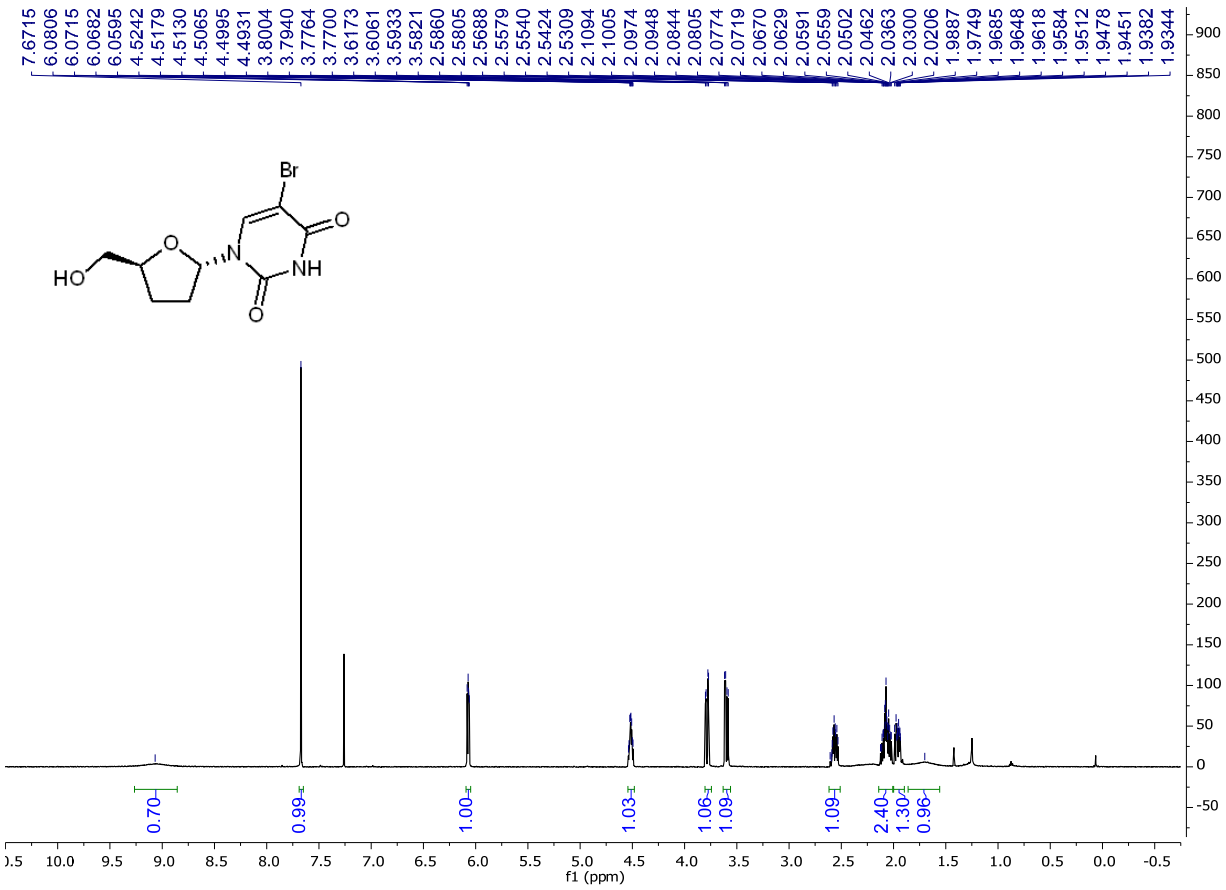


Figure S51. ^{1}H spectrum of **α-ddBrU** (500 MHz, $CDCl_3$).

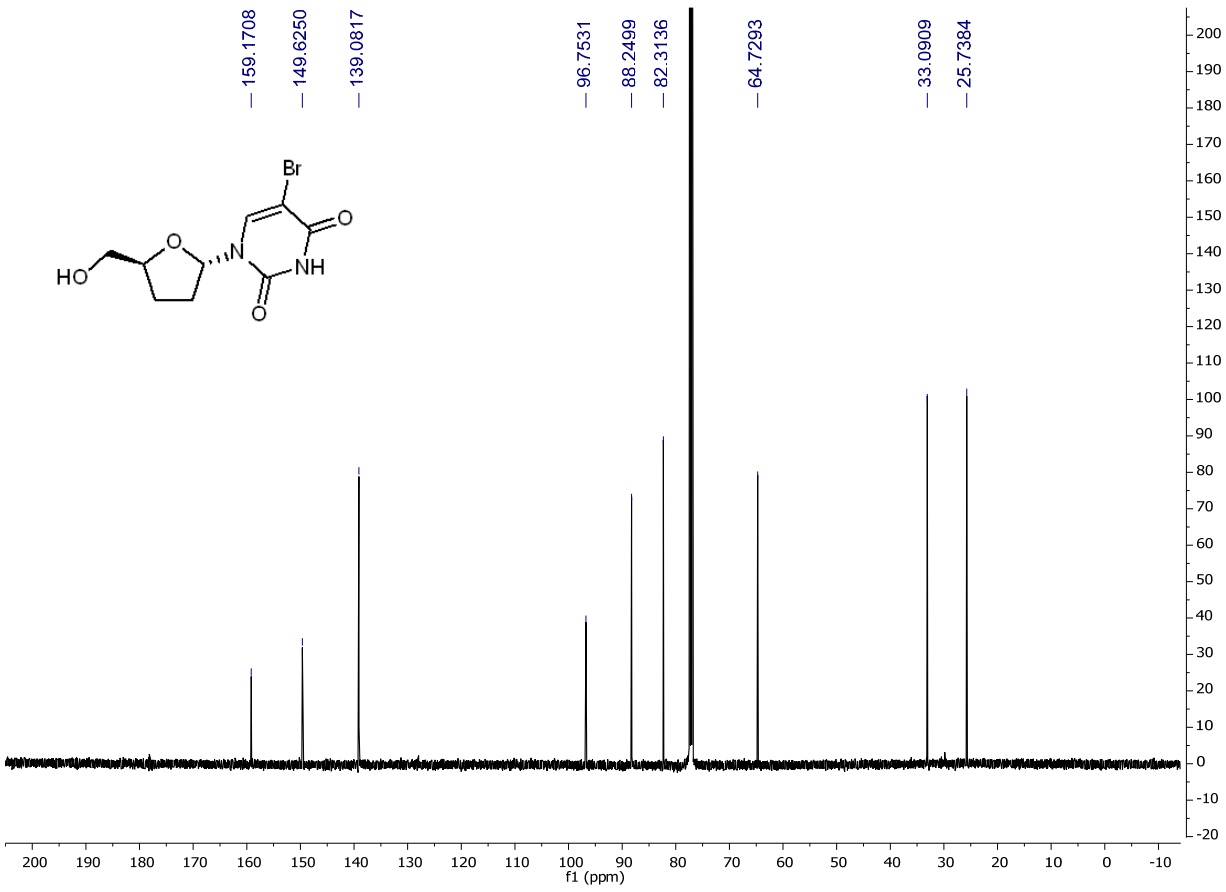


Figure S52. ^{13}C spectrum of **α-ddBrU** (125 MHz, $CDCl_3$).

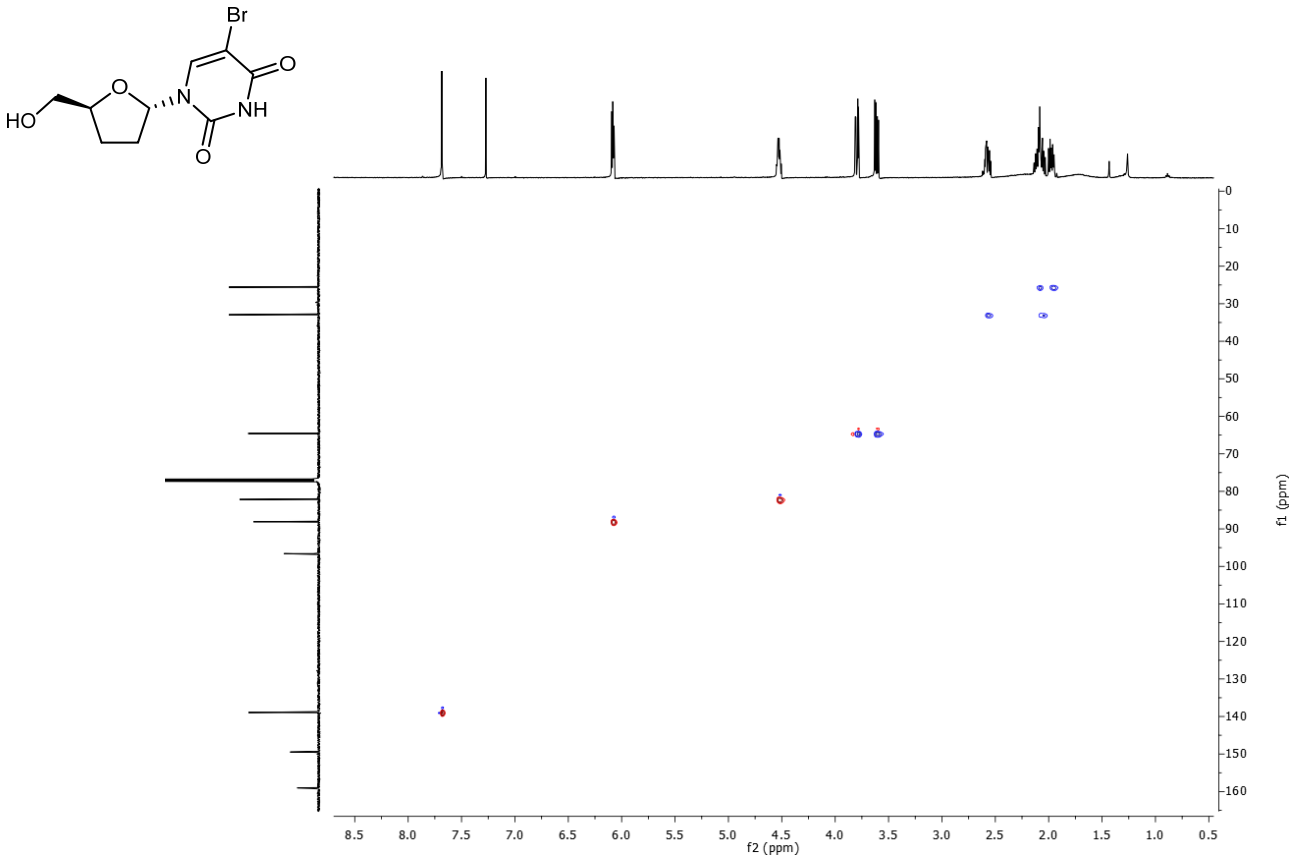


Figure S53. HSQC spectrum of **α-ddBrU**.

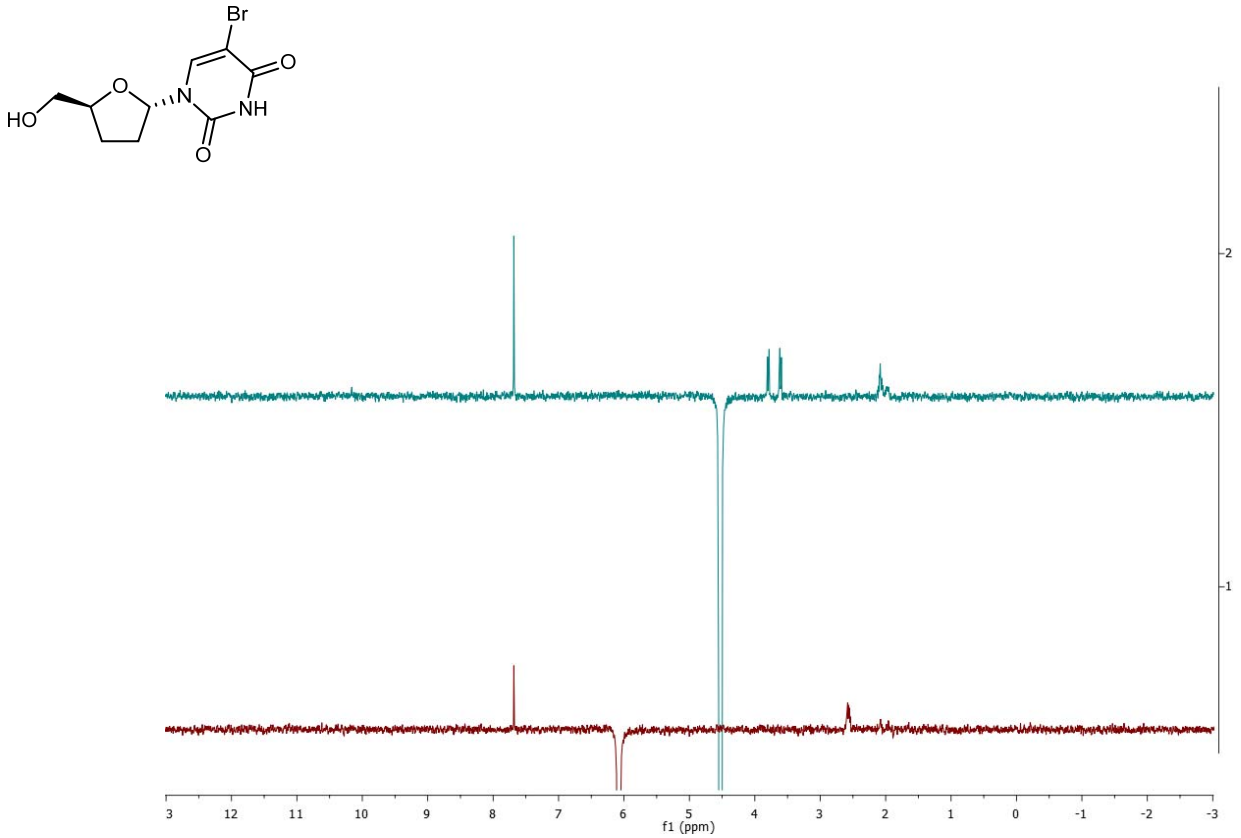


Figure S54. 1D-NOESY spectra of **α-ddBrU**.

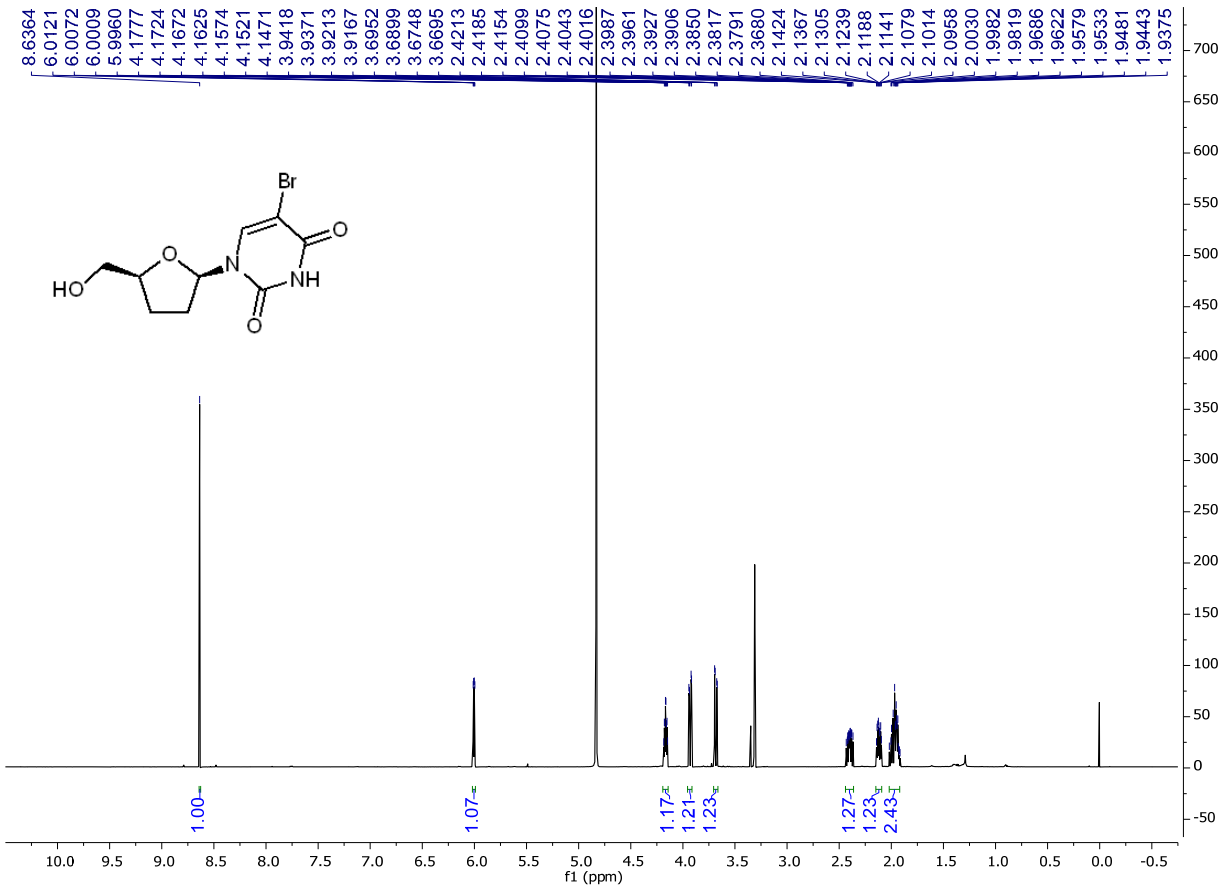


Figure S55. ^{1}H spectrum of **β-ddBrU** (600 MHz, CD_3OD).

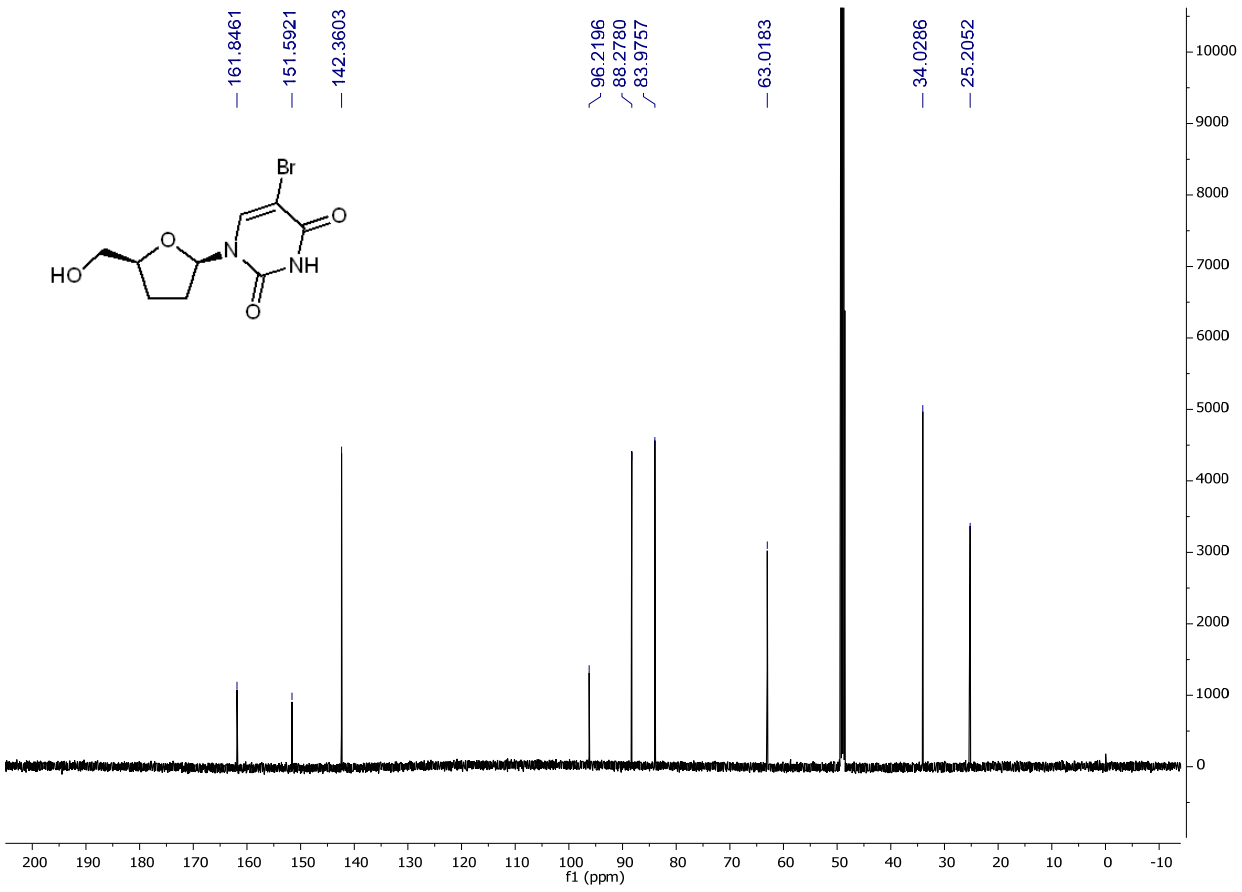


Figure S56. ^{13}C spectrum of **β-ddBrU** (150 MHz, CD_3OD).

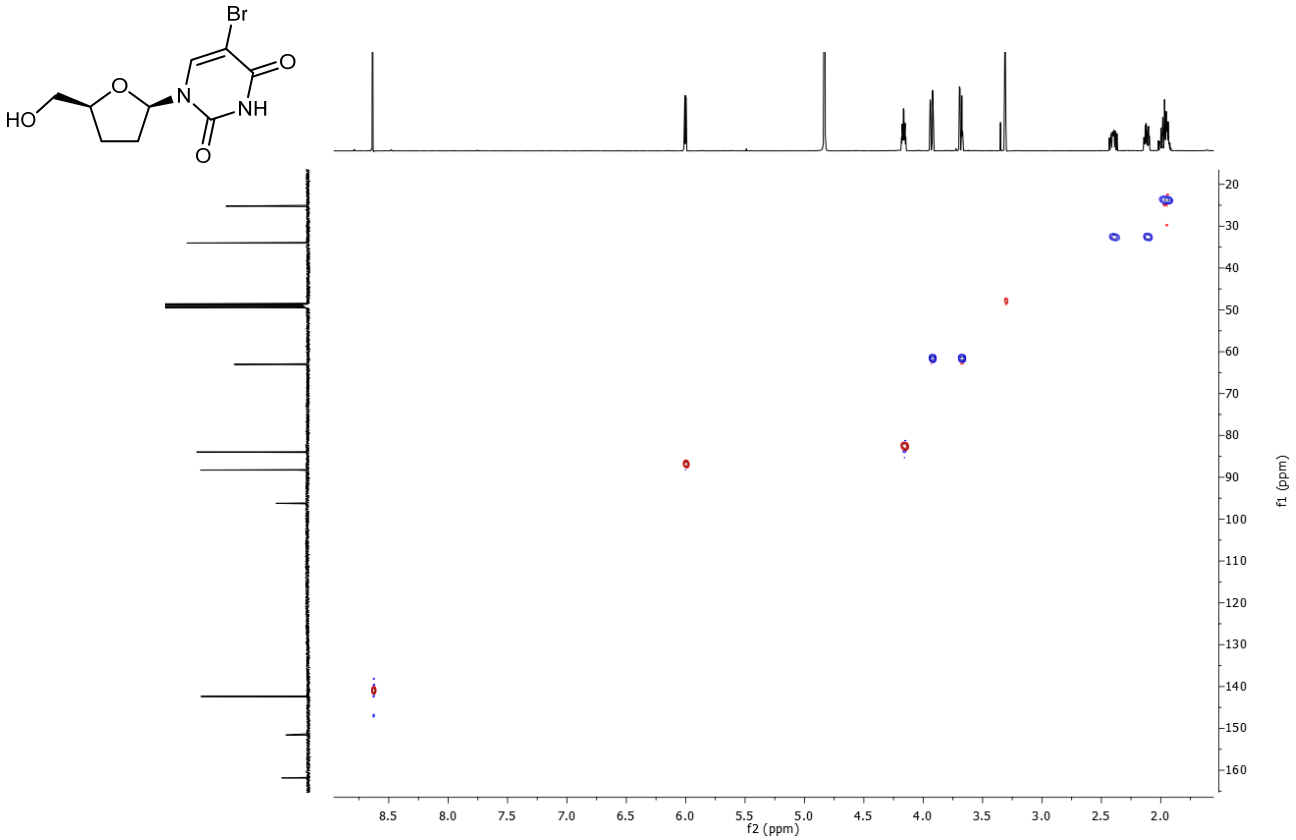


Figure S57. HSQC spectrum of **β-ddBrU.**

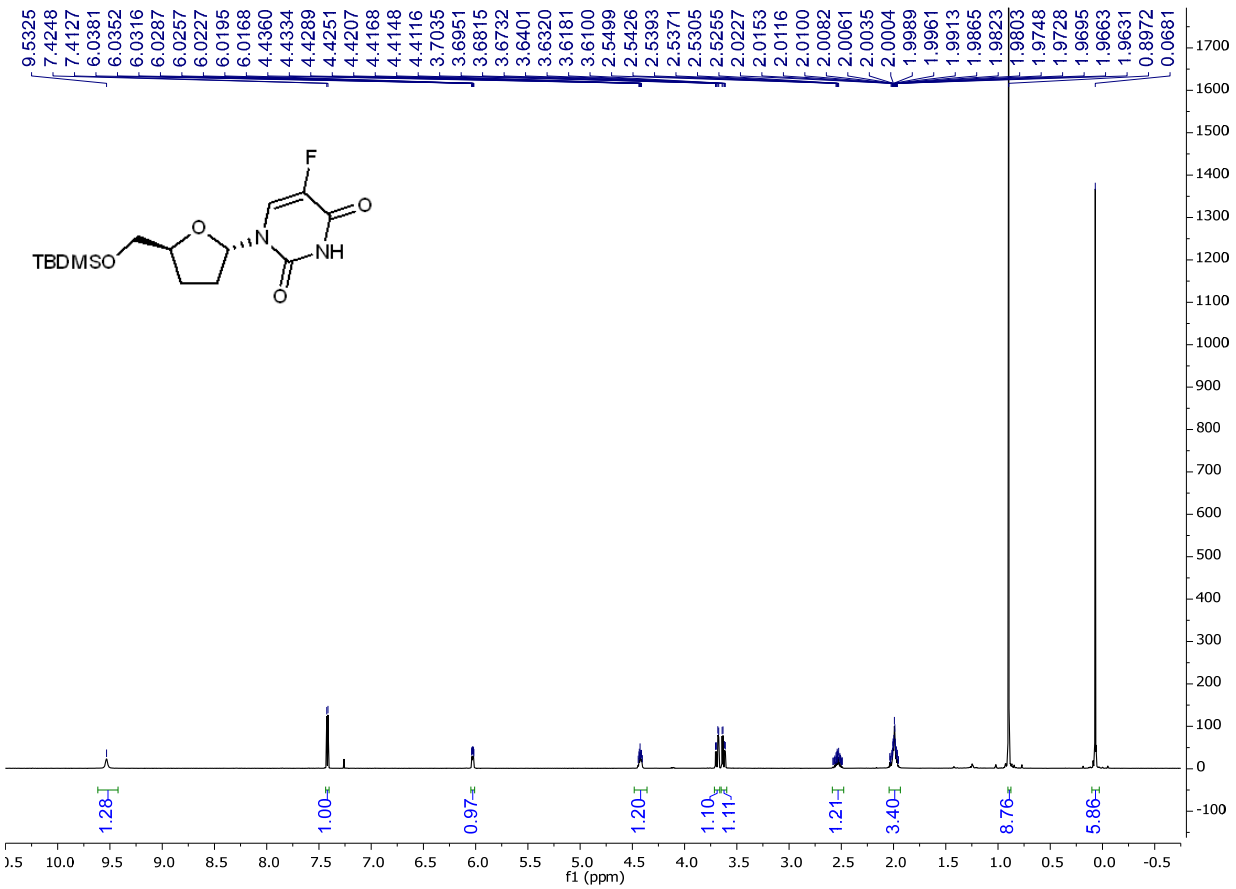


Figure S58. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddFU** (500 MHz, $CDCl_3$).

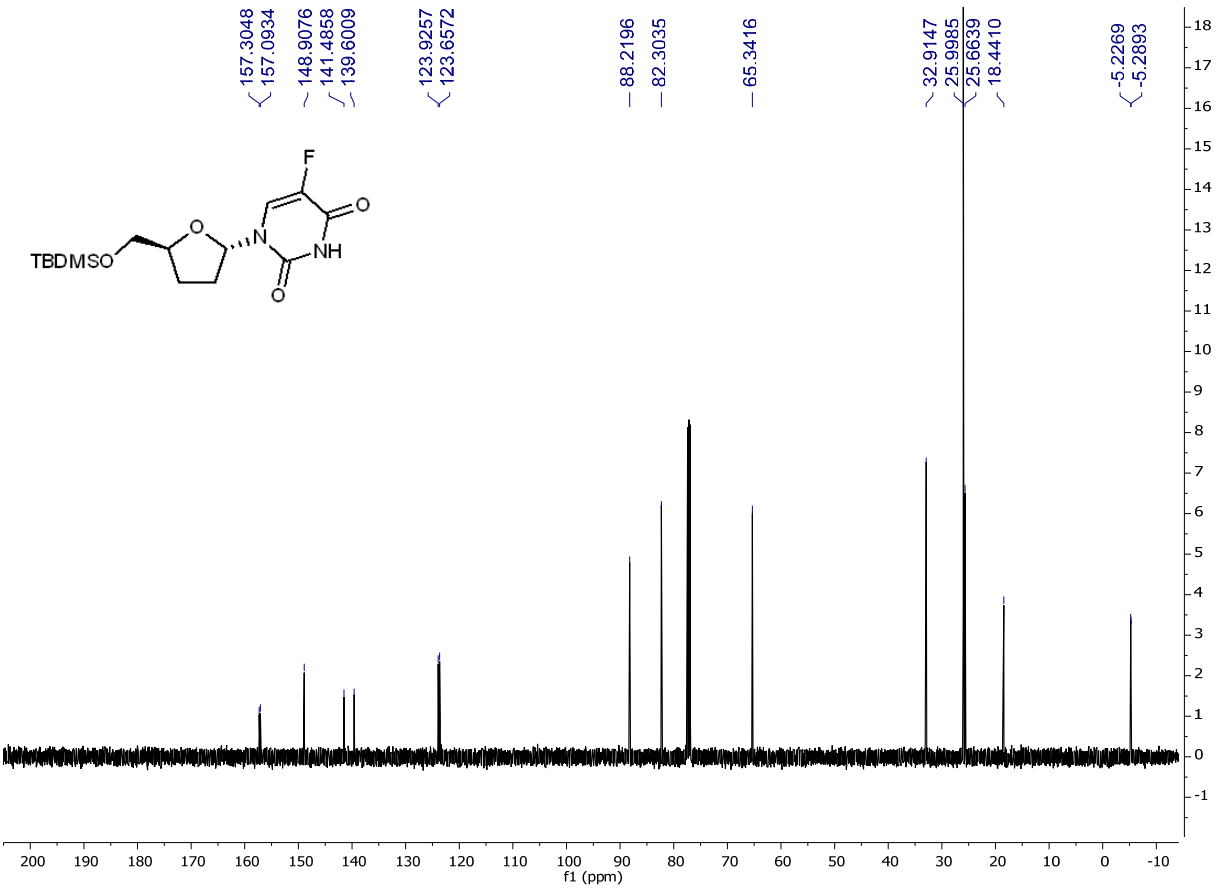


Figure S59. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddFU** (125 MHz, $CDCl_3$).

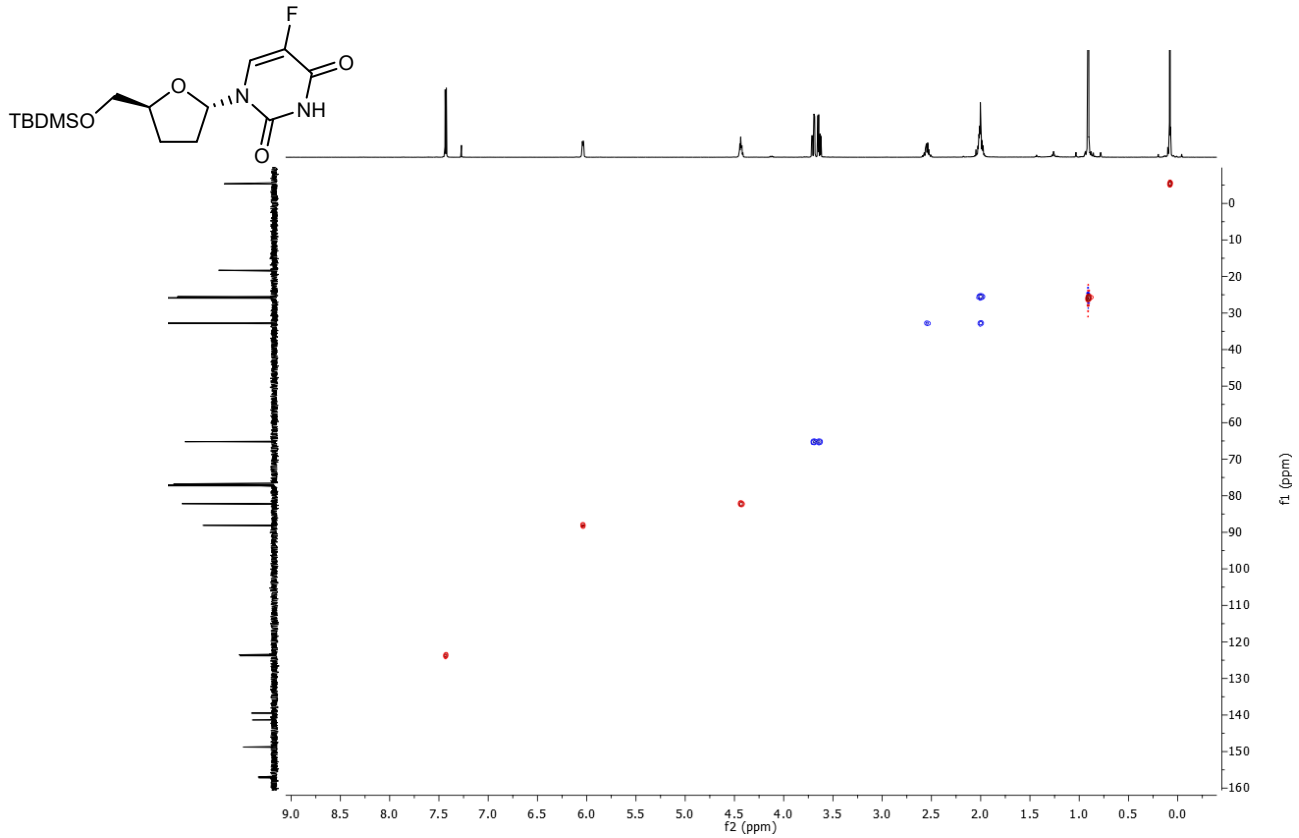


Figure S60. HSQC spectrum of **5'-*O*-silyl-protected α-ddFU**.

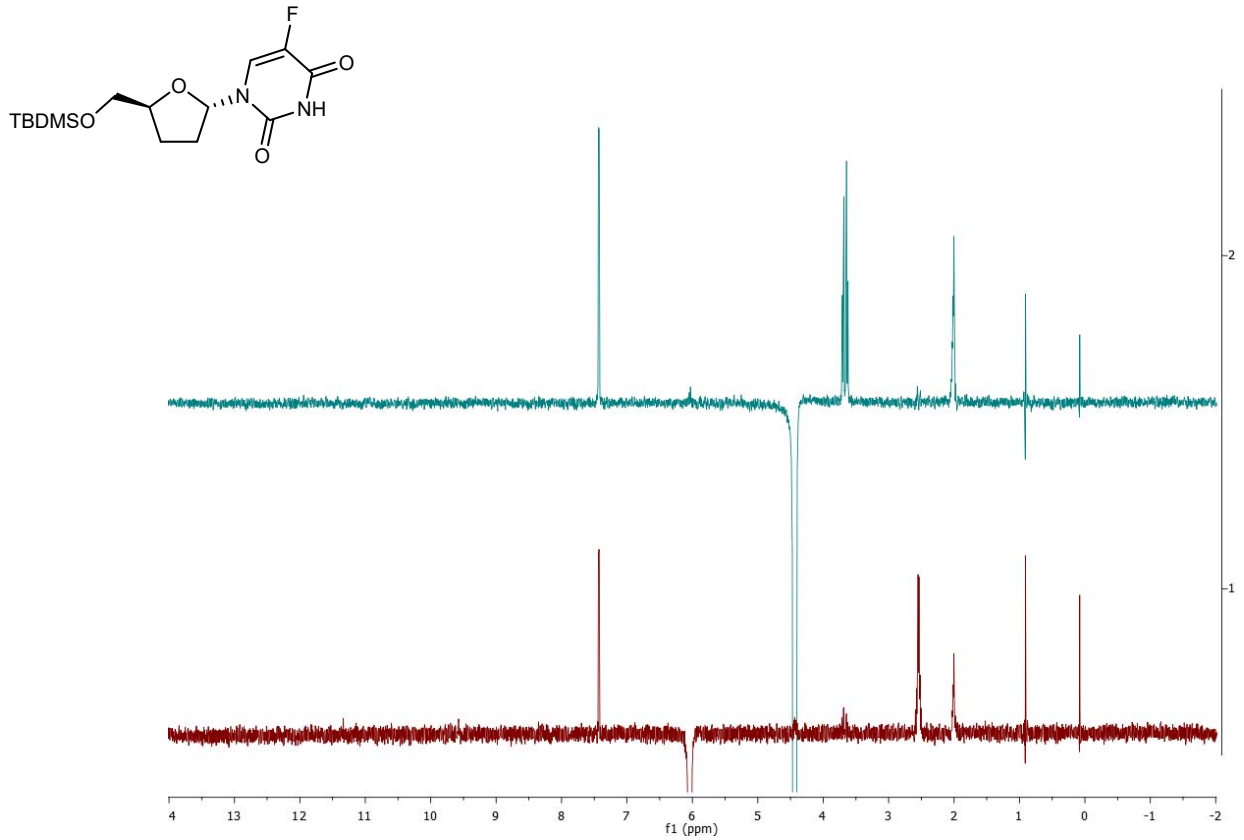


Figure S61. 1D-NOESY spectra of **5'-*O*-silyl-protected α-ddFU**.

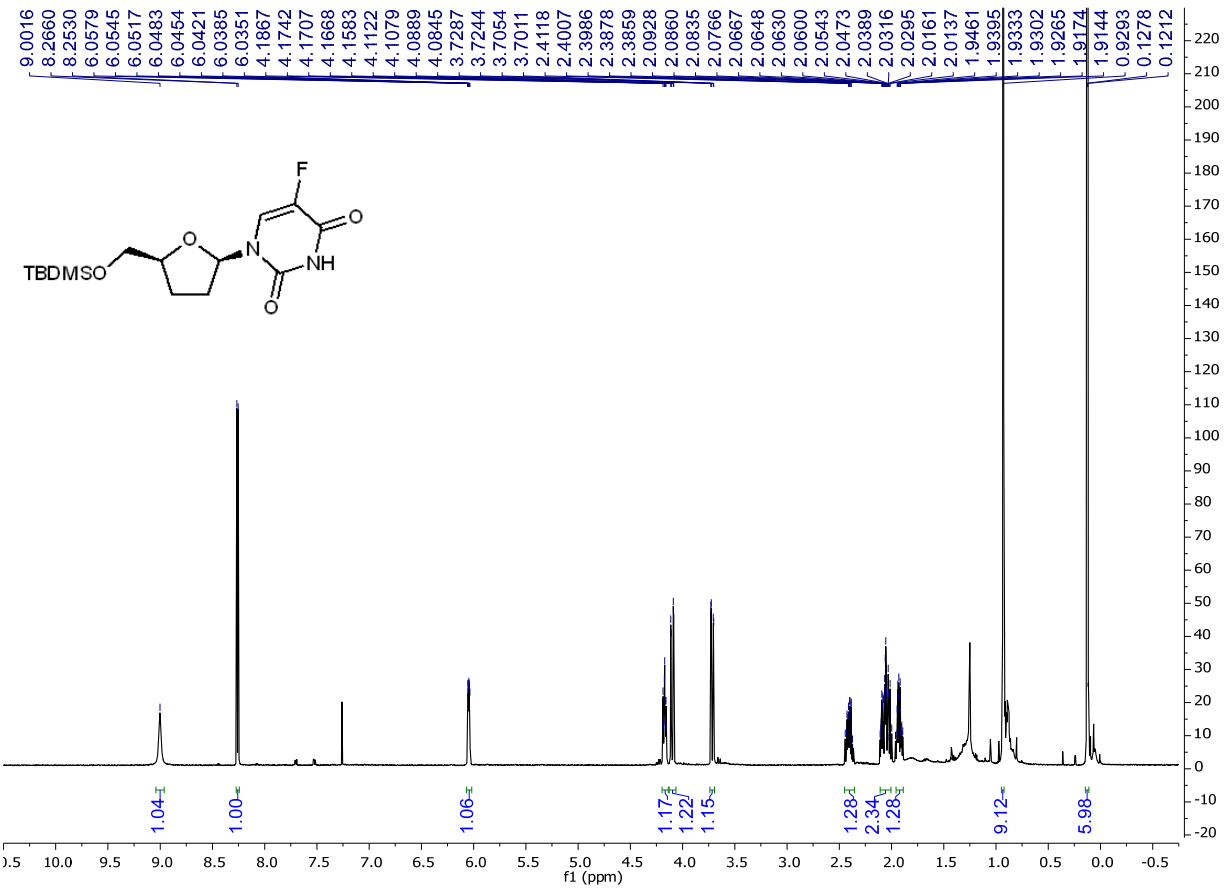


Figure S62. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddFU** (500 MHz, $CDCl_3$).

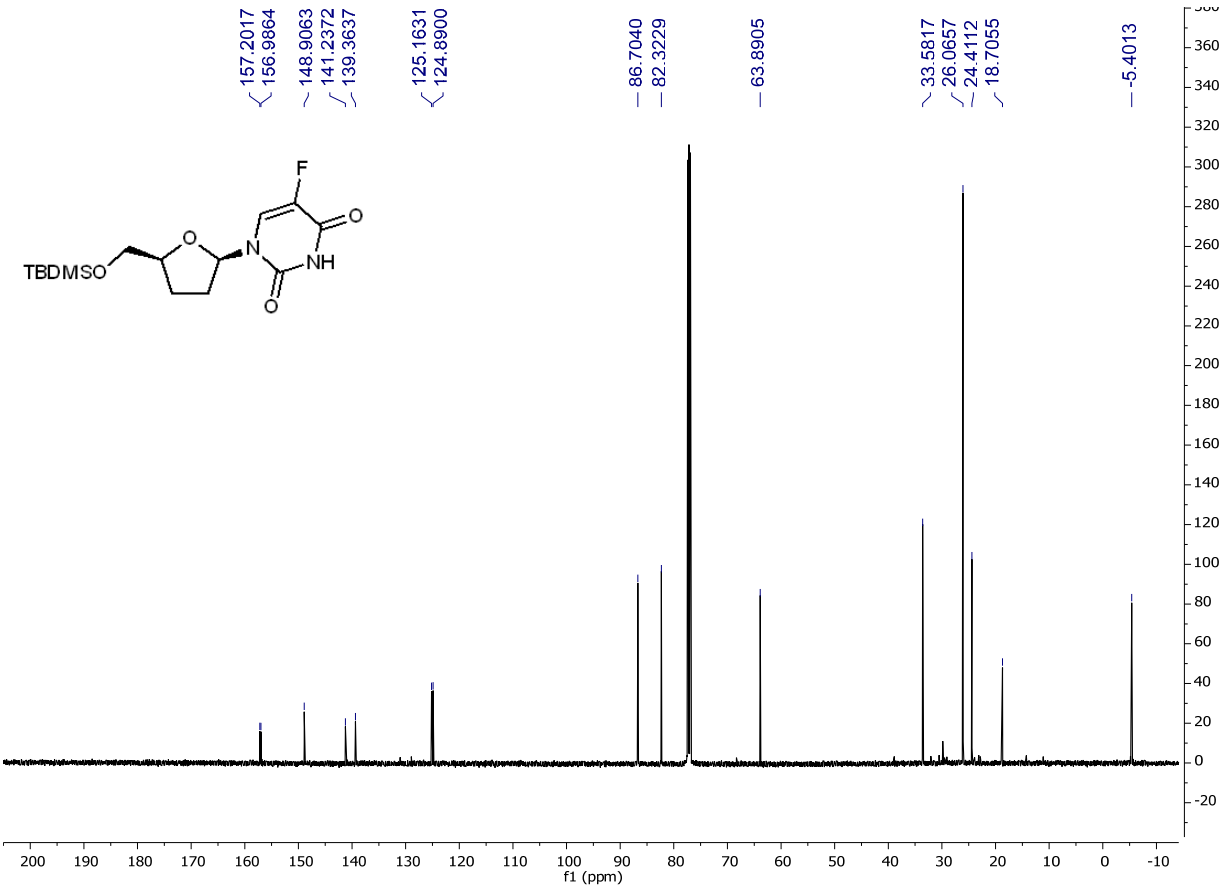


Figure S63. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddFU** (125 MHz, $CDCl_3$).

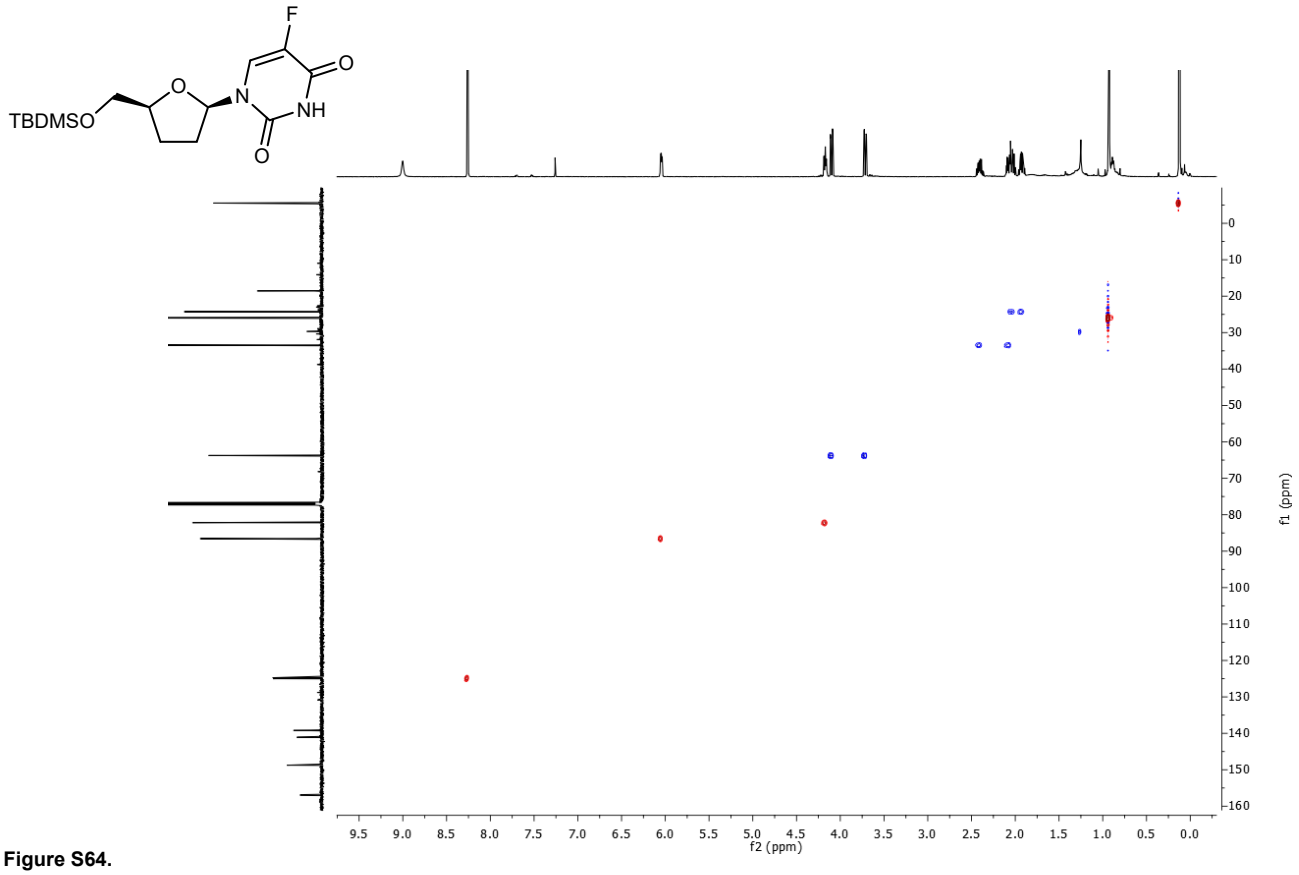


Figure S64.
HSQC spectrum of **5'-*O*-silyl-protected β-ddFU**.

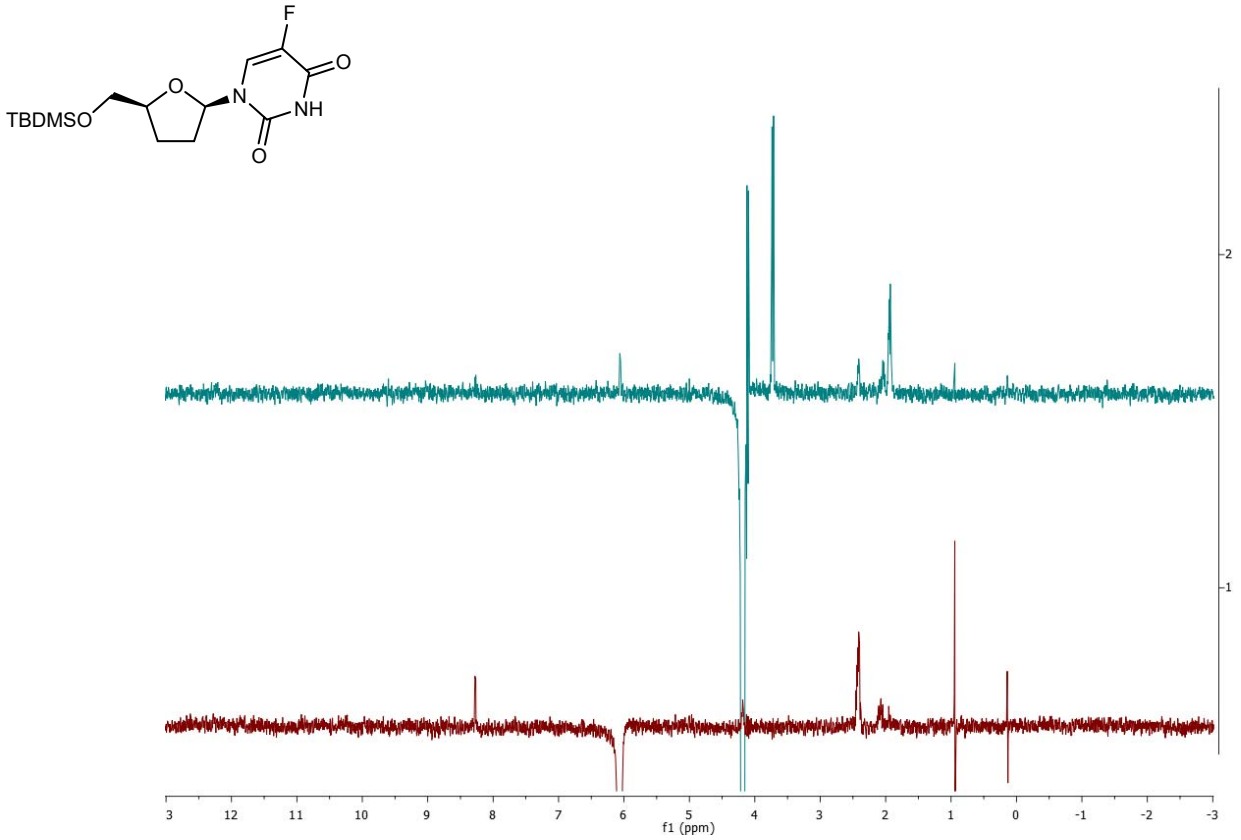


Figure S65. 1D-NOESY spectra of **5'-*O*-silyl-protected β-ddFU**.

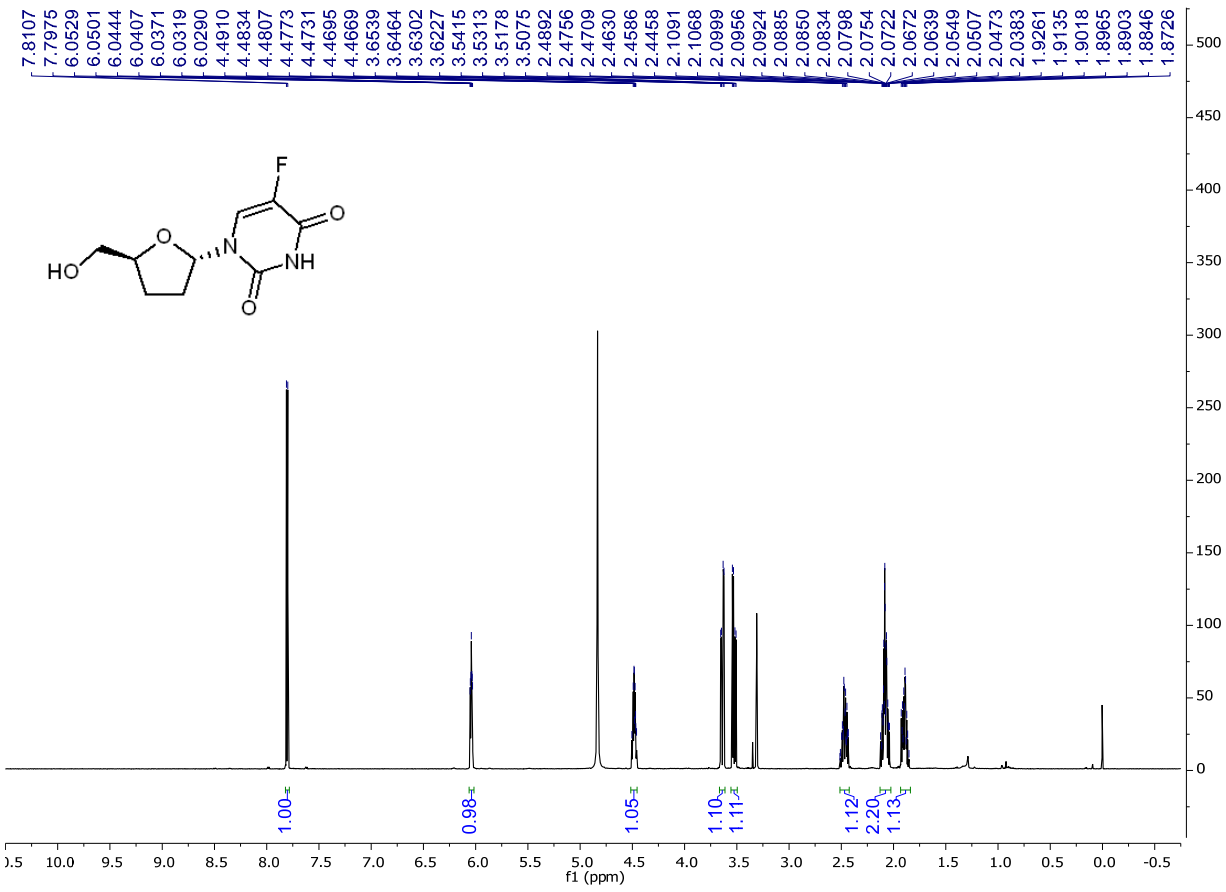


Figure S66. ^{1}H spectrum of **α-ddFU** (500 MHz, CD_3OD).

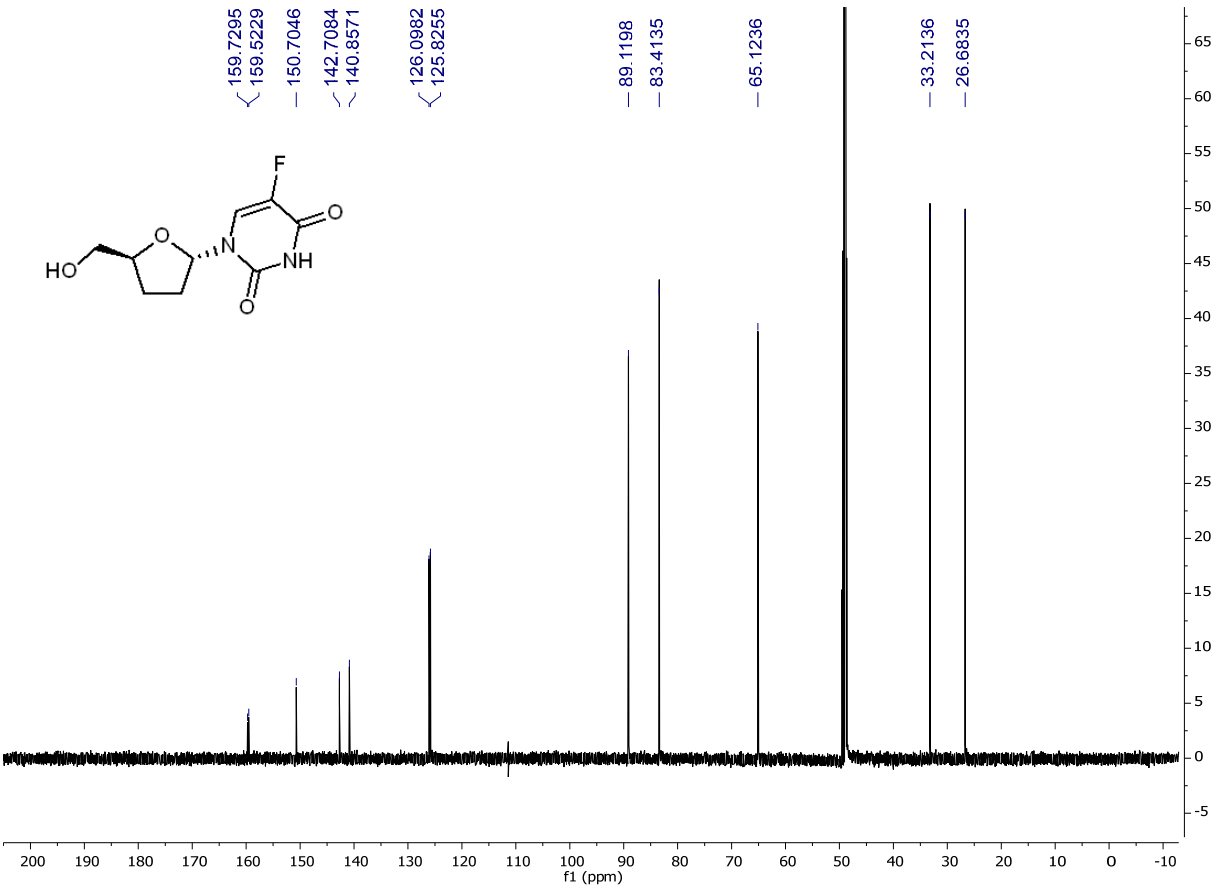


Figure S67. ^{13}C spectrum of **α-ddFU** (125 MHz, CD_3OD).

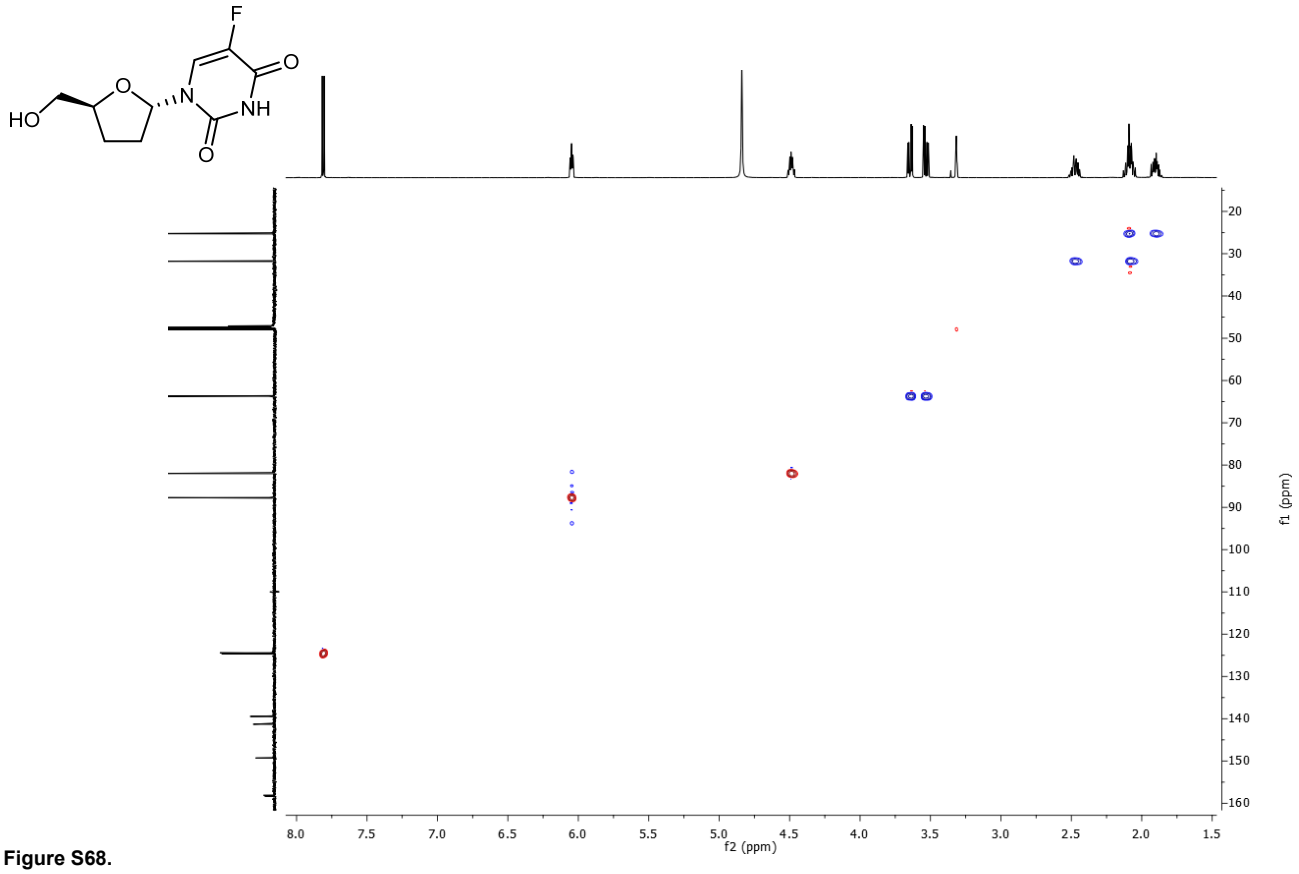


Figure S68.
HSQC spectrum of **α-ddFU**.

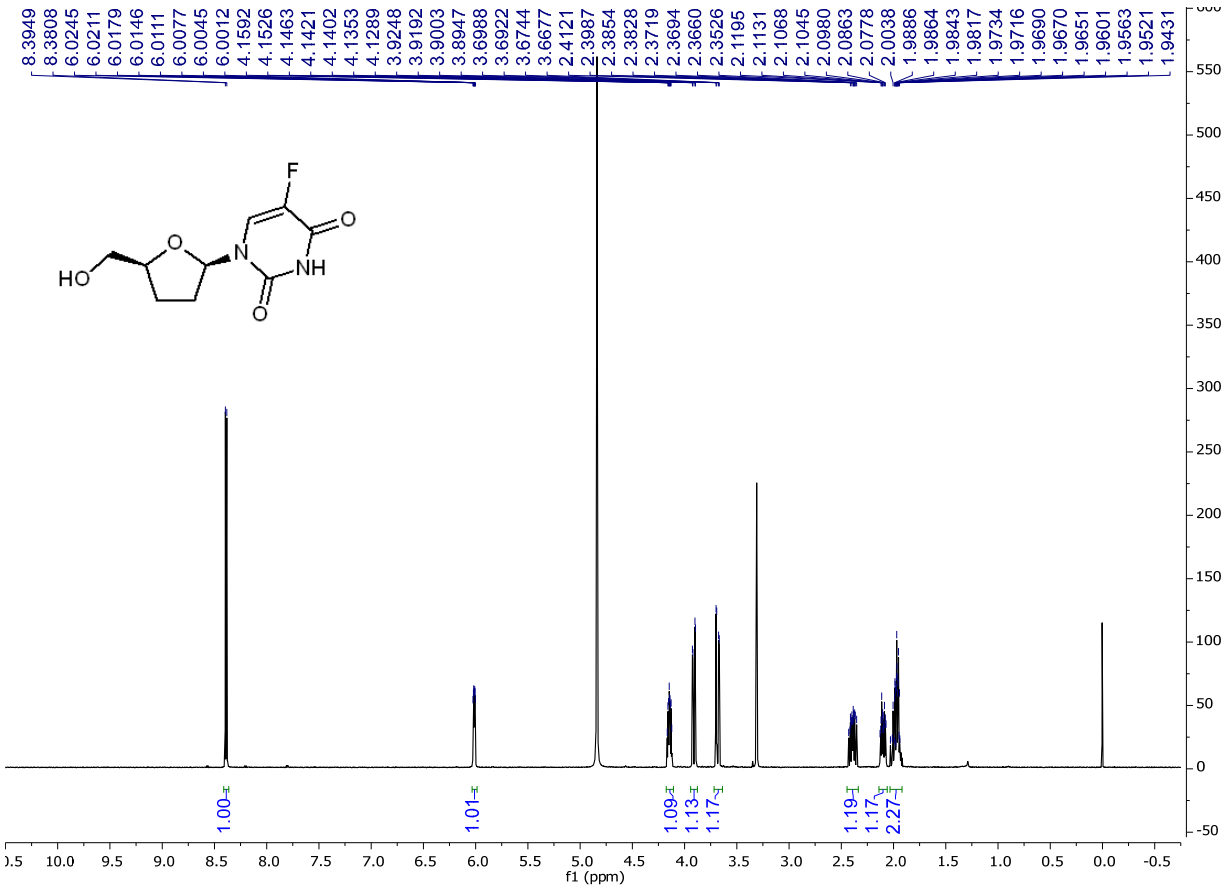


Figure S69. ^{1}H spectrum of **β-ddFU** (500 MHz, CD_3OD).

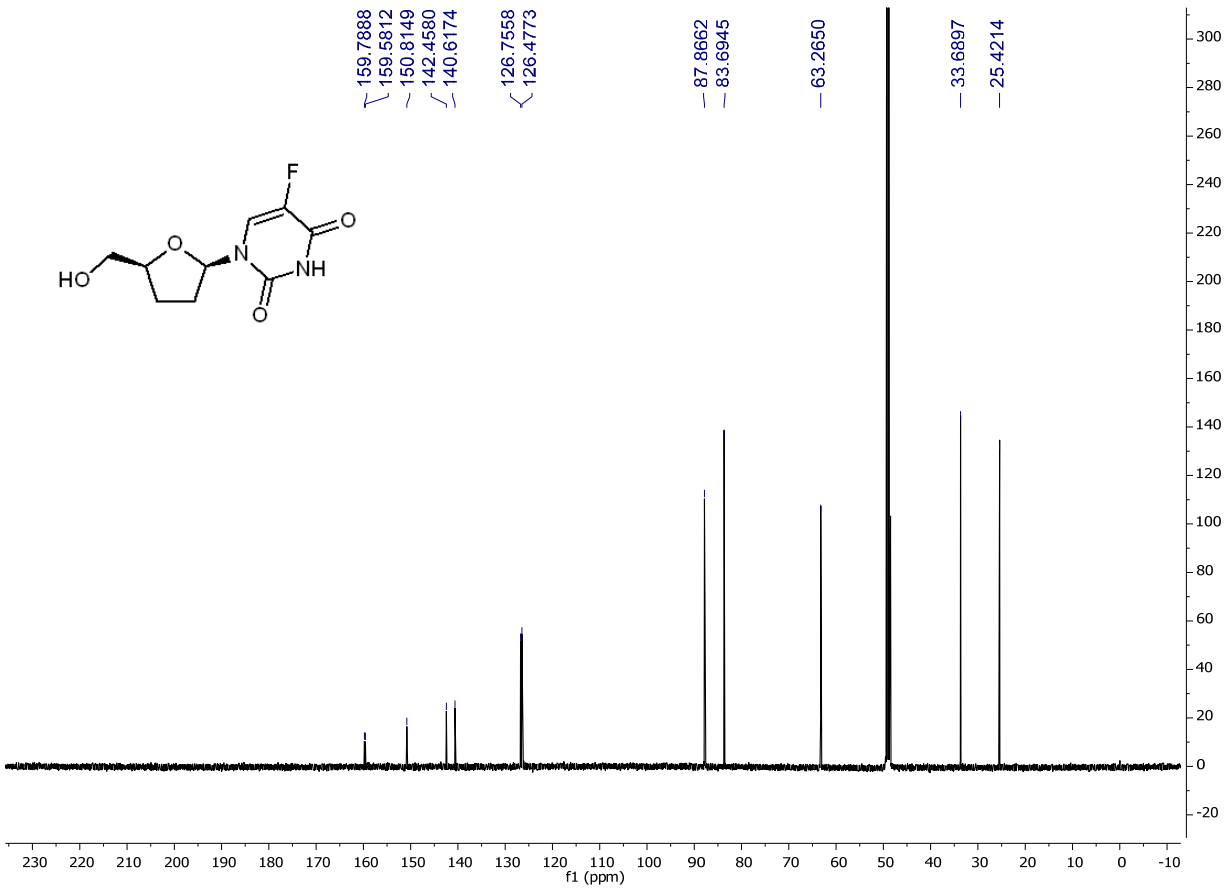


Figure S70. ^{13}C spectrum of **β-ddFU** (125 MHz, CD_3OD).

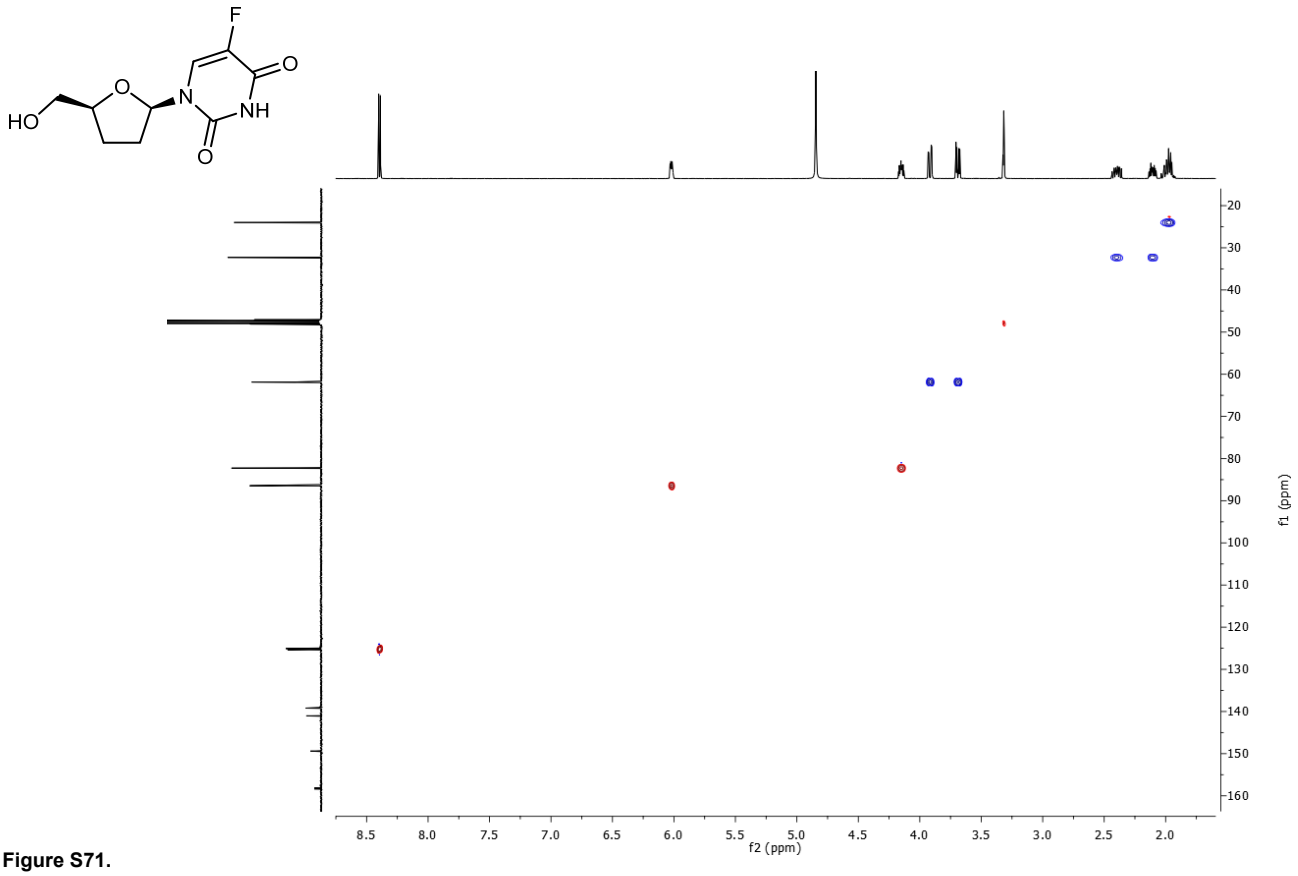


Figure S71.
HSQC spectrum of **β-ddFU**.

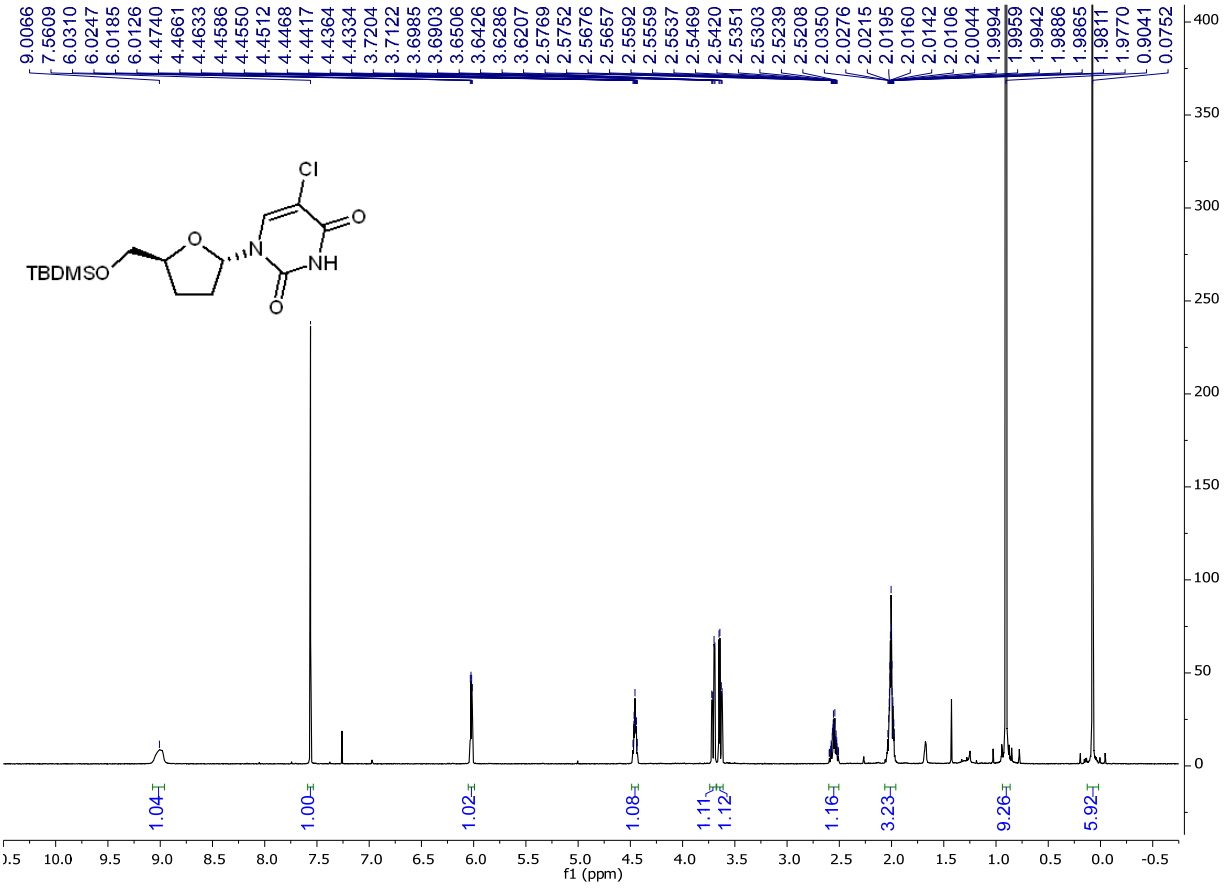


Figure S72. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddClU** (500 MHz, $CDCl_3$).

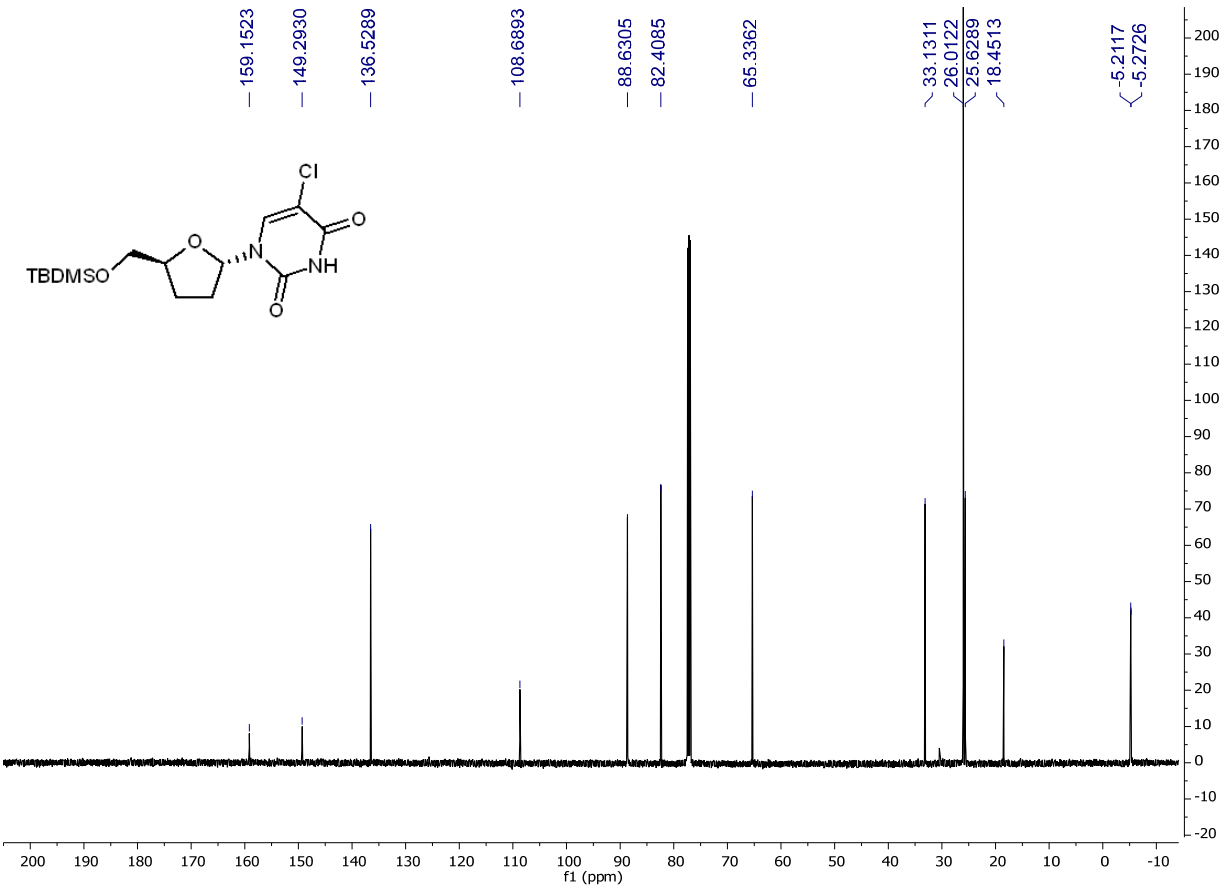


Figure S73. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddClU** (125 MHz, $CDCl_3$).

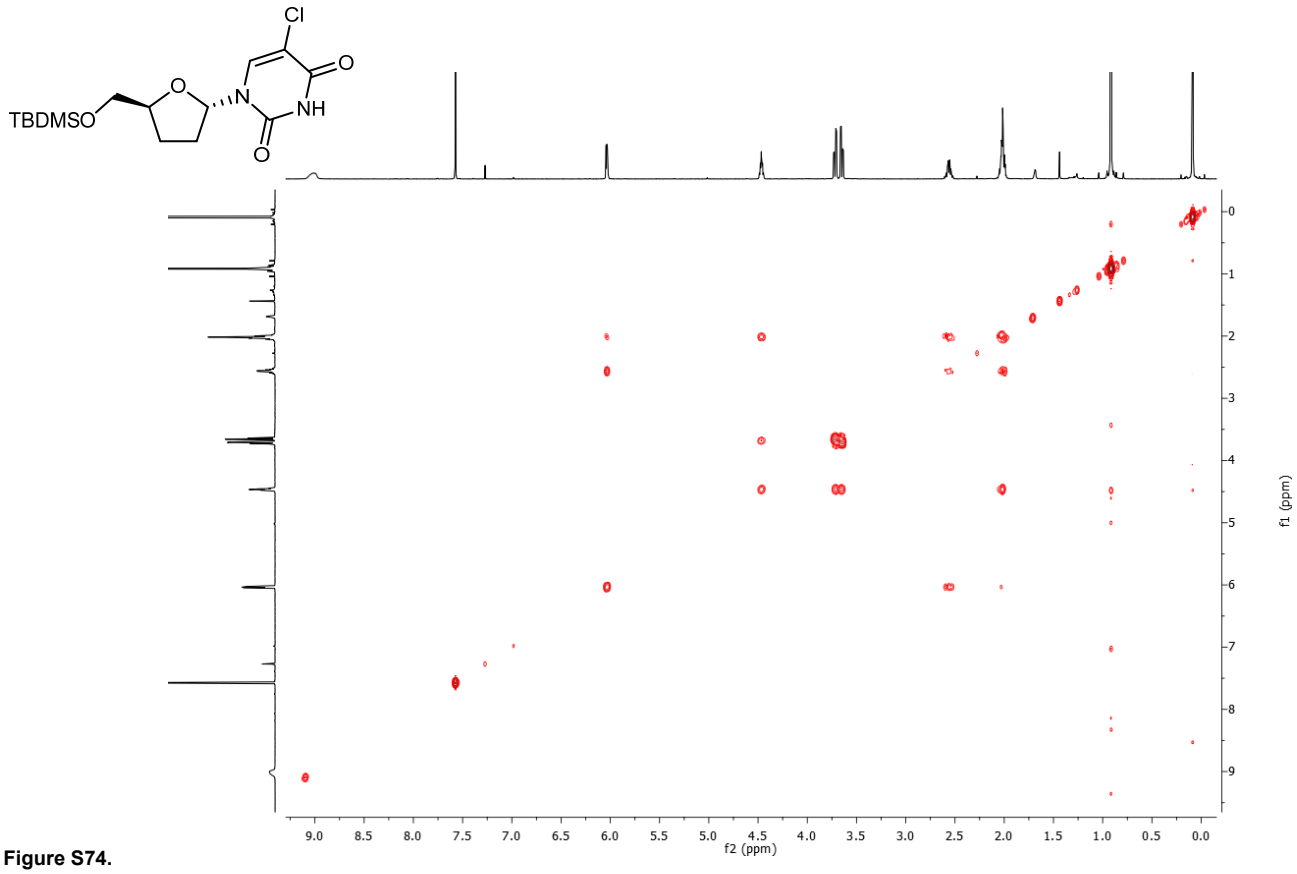


Figure S74.
COSY spectrum of **5'-*O*-silyl-protected α-ddClU**.

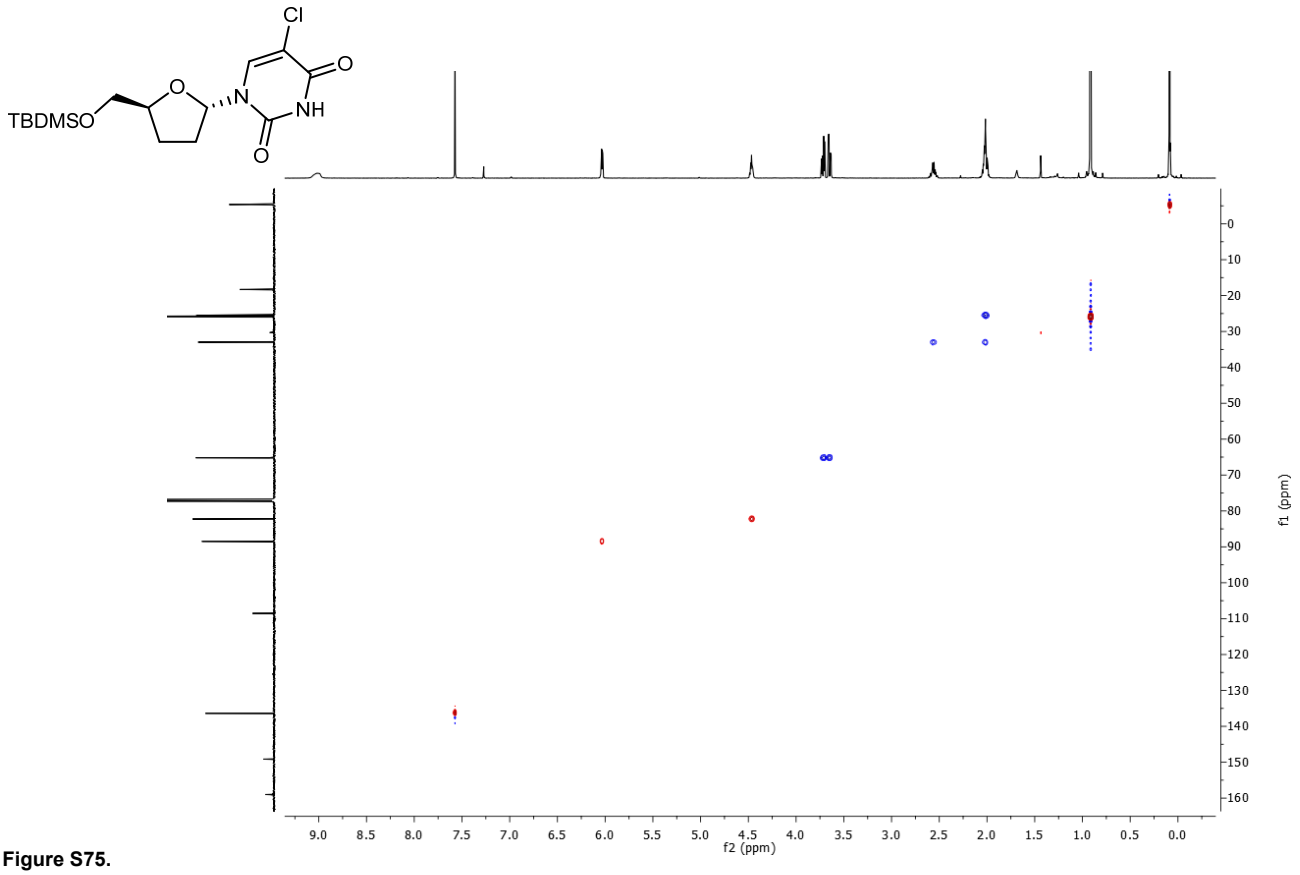


Figure S75.
HSQC spectrum of **5'-*O*-silyl-protected α-ddClU**.

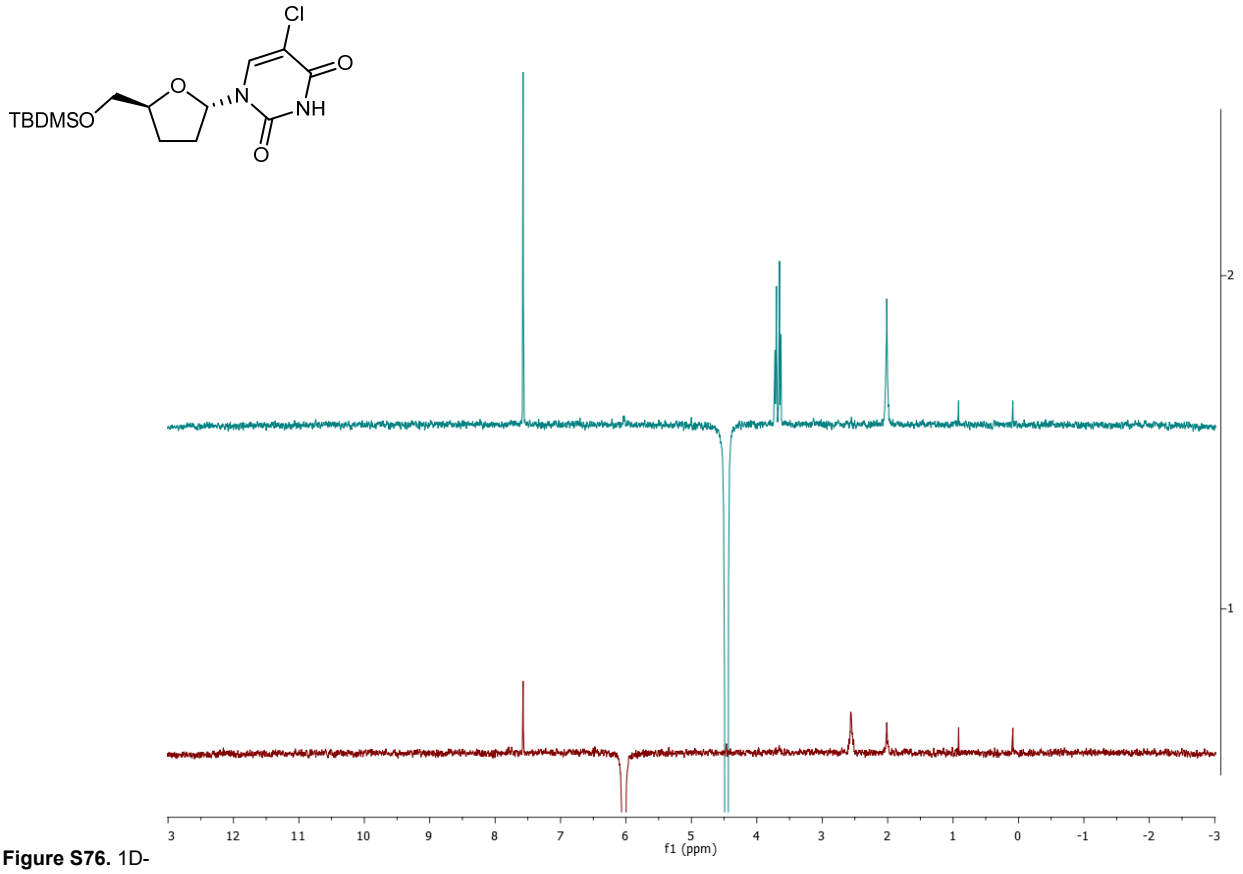


Figure S76. 1D-
NOESY spectra of **5'-*O*-silyl-protected α-ddClU**.

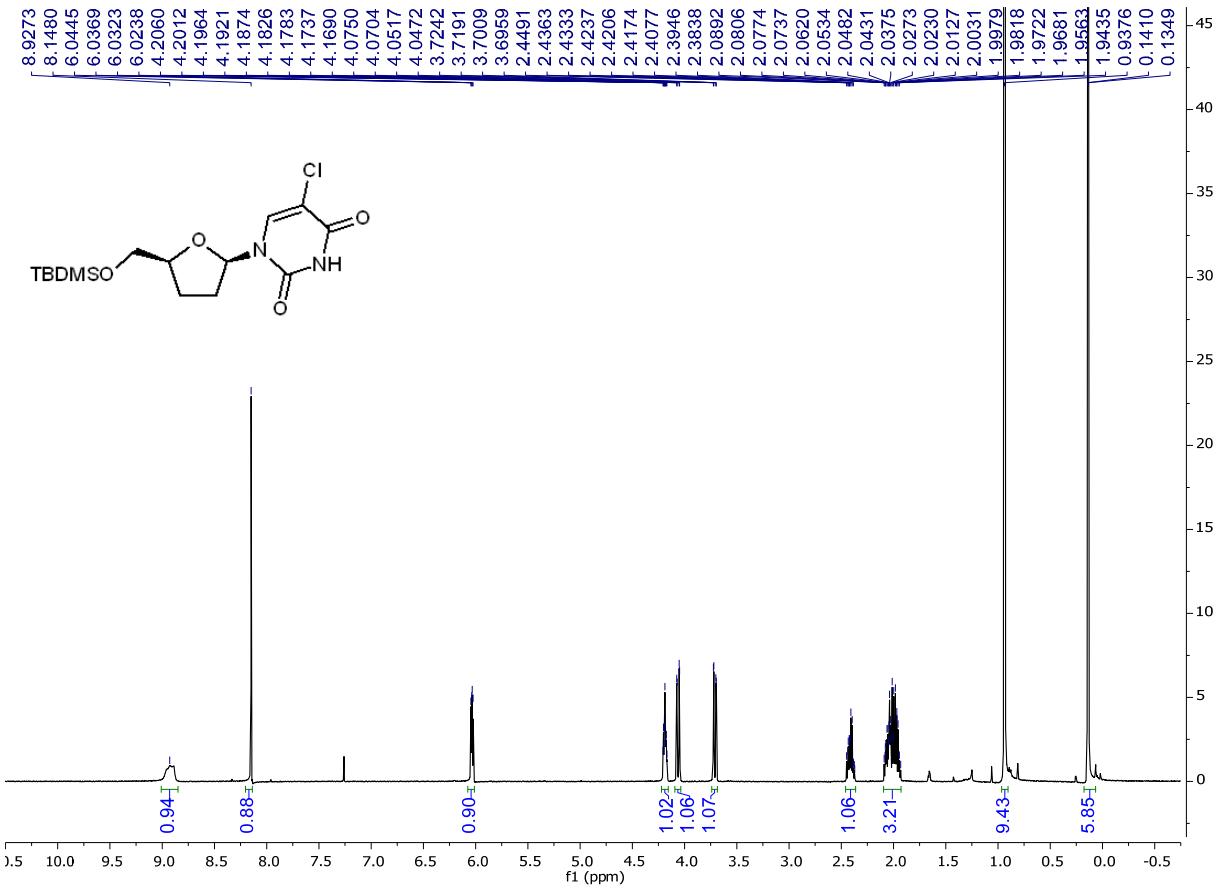


Figure S77. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddClU** (500 MHz, $CDCl_3$).

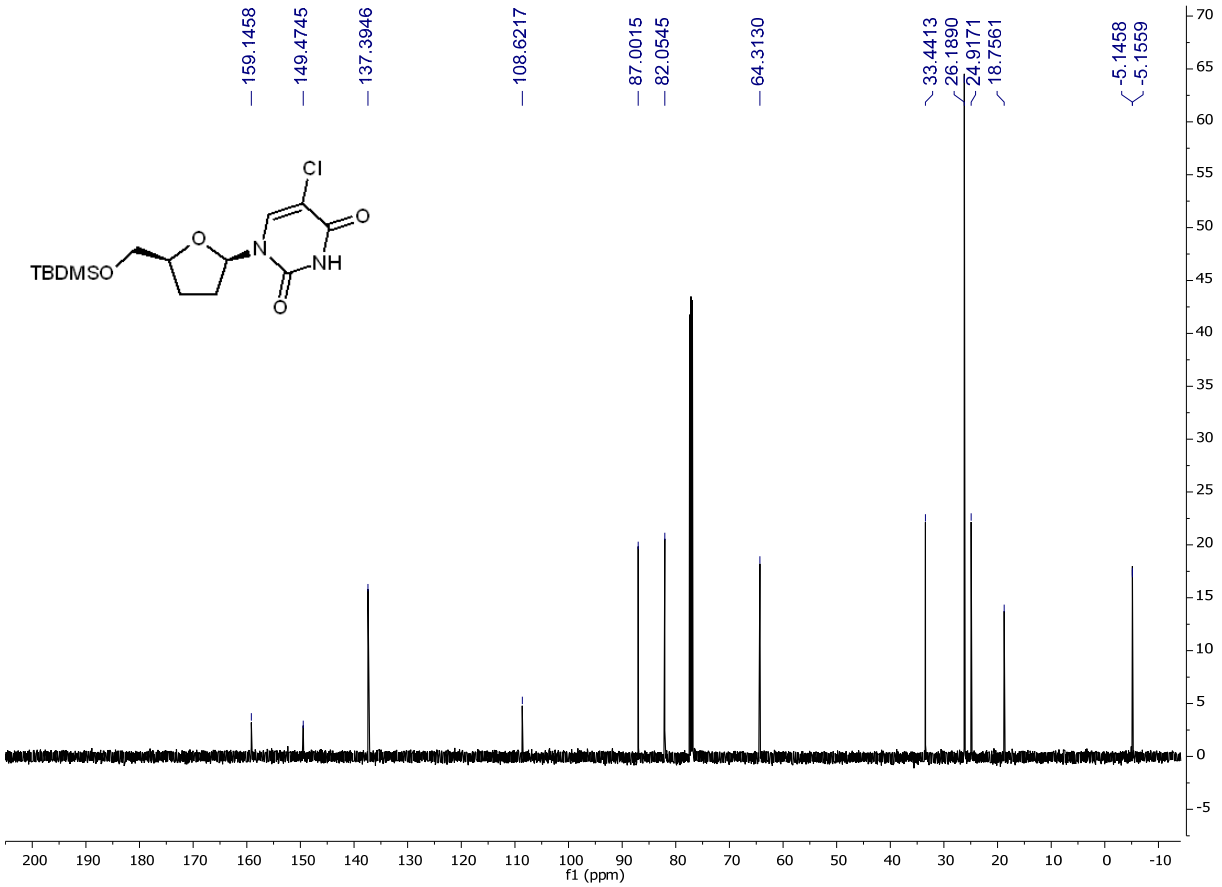


Figure S78. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddClU** (125 MHz, $CDCl_3$).

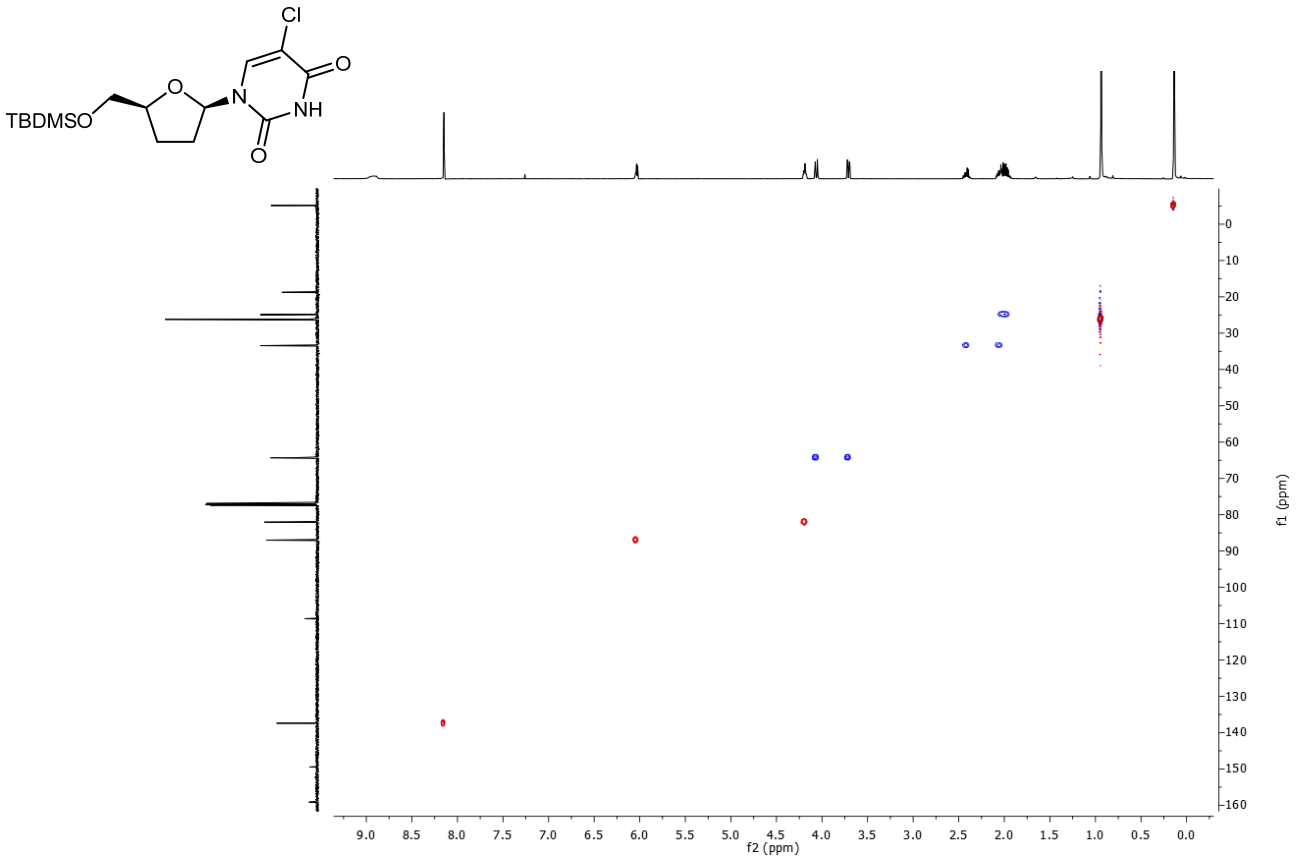


Figure S79. HSQC spectrum of **5'-*O*-silyl-protected β-ddClU**.

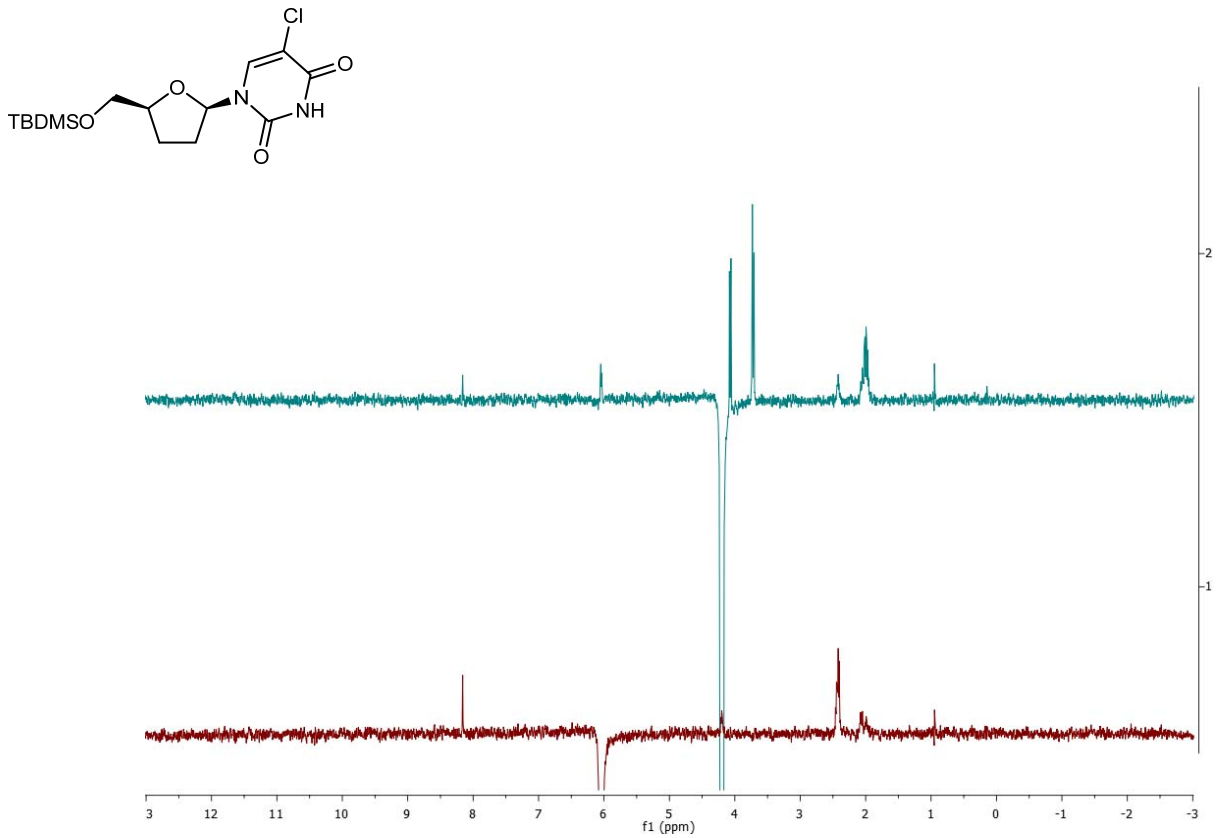


Figure S80. 1D-NOESY spectra of **5'-*O*-silyl-protected β-ddClU**.

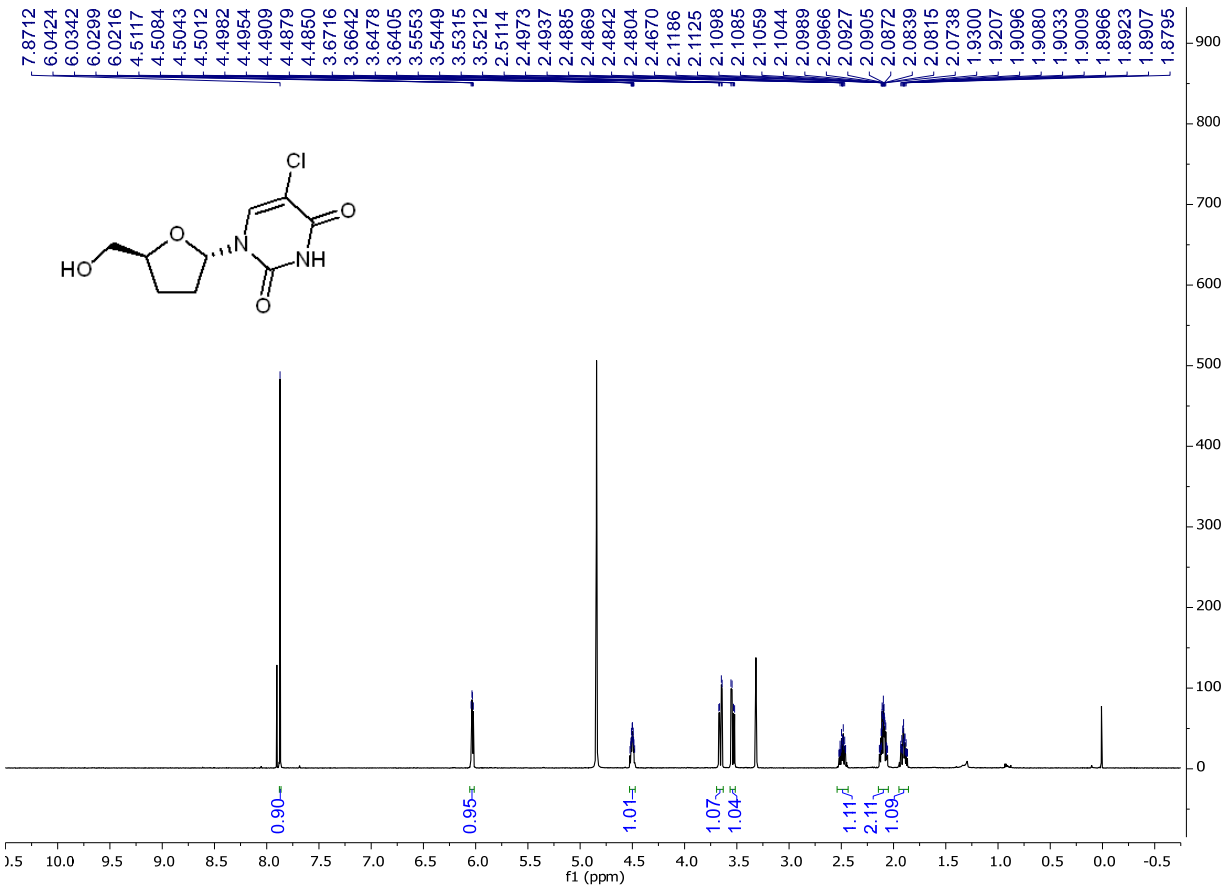


Figure S81. ^{1}H spectrum of **α-ddClU** (500 MHz, CD_3OD).

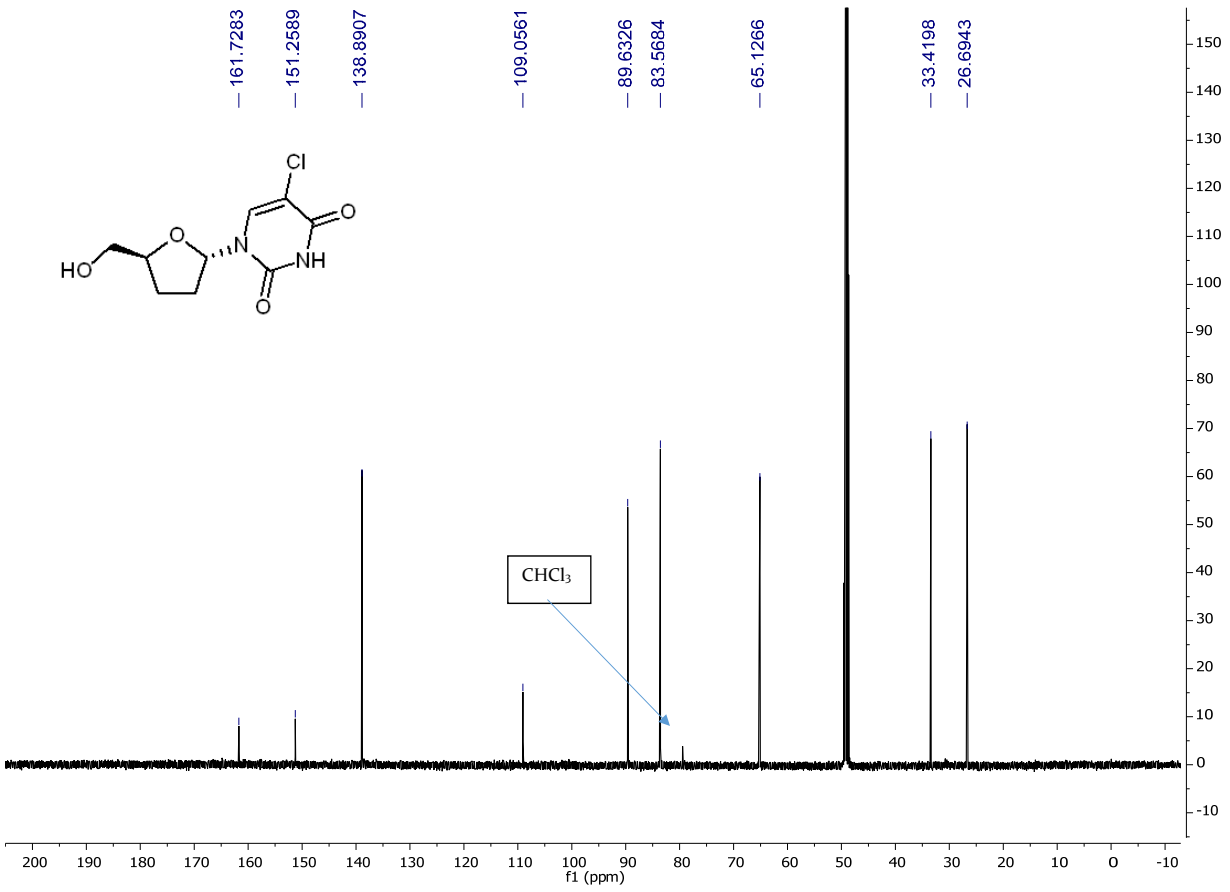


Figure S82. ^{13}C spectrum of **α-ddClU** (125 MHz, CD_3OD).

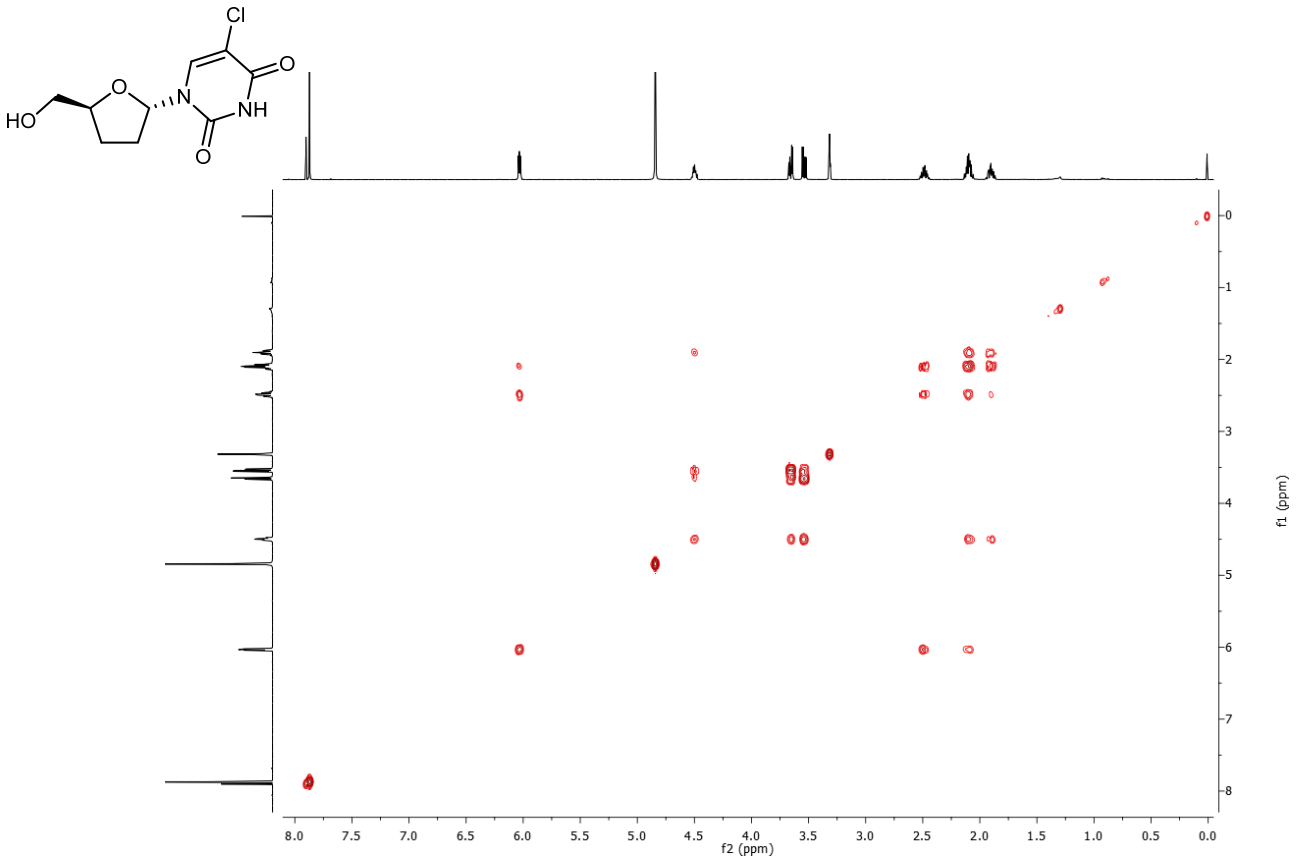


Figure S83. COSY spectrum of **α-ddClU**.

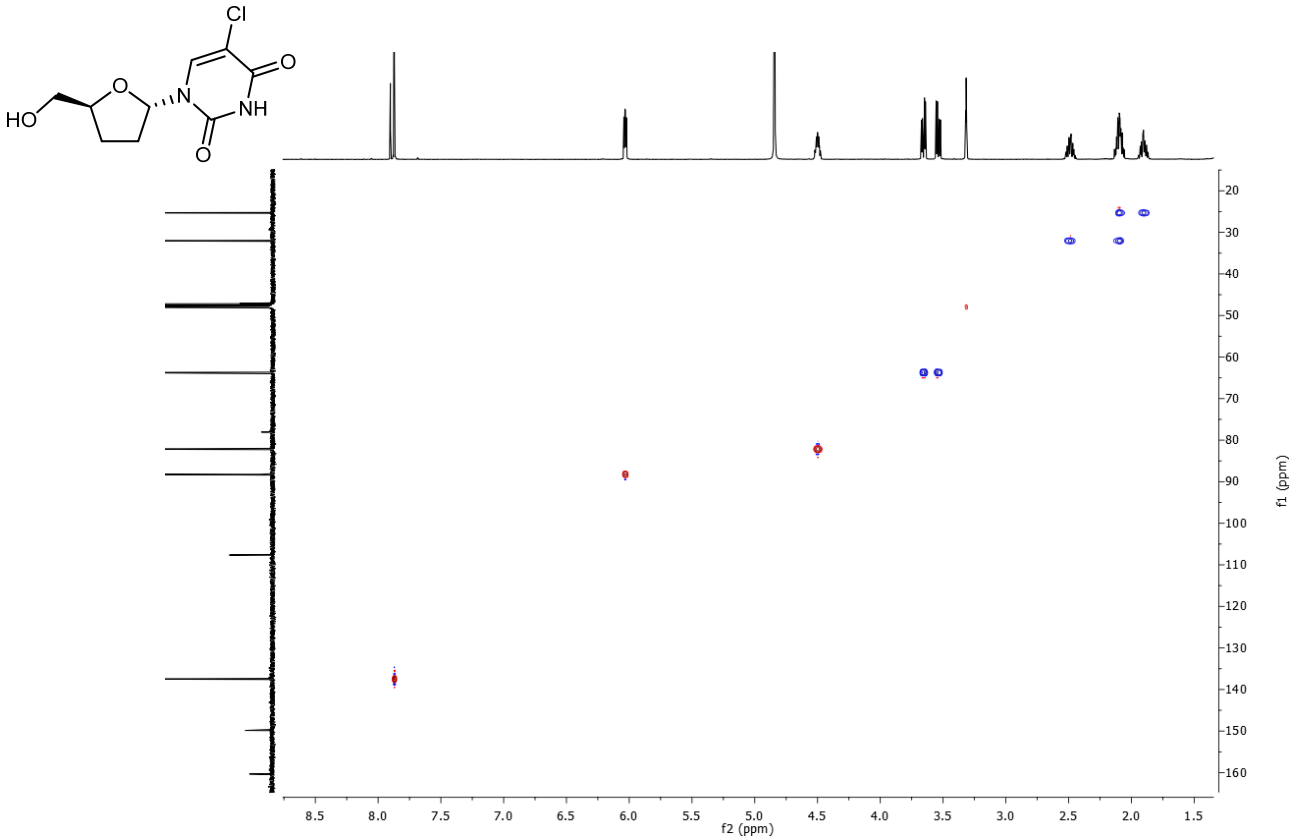


Figure S84. HSQC spectrum of **α-ddClU**.

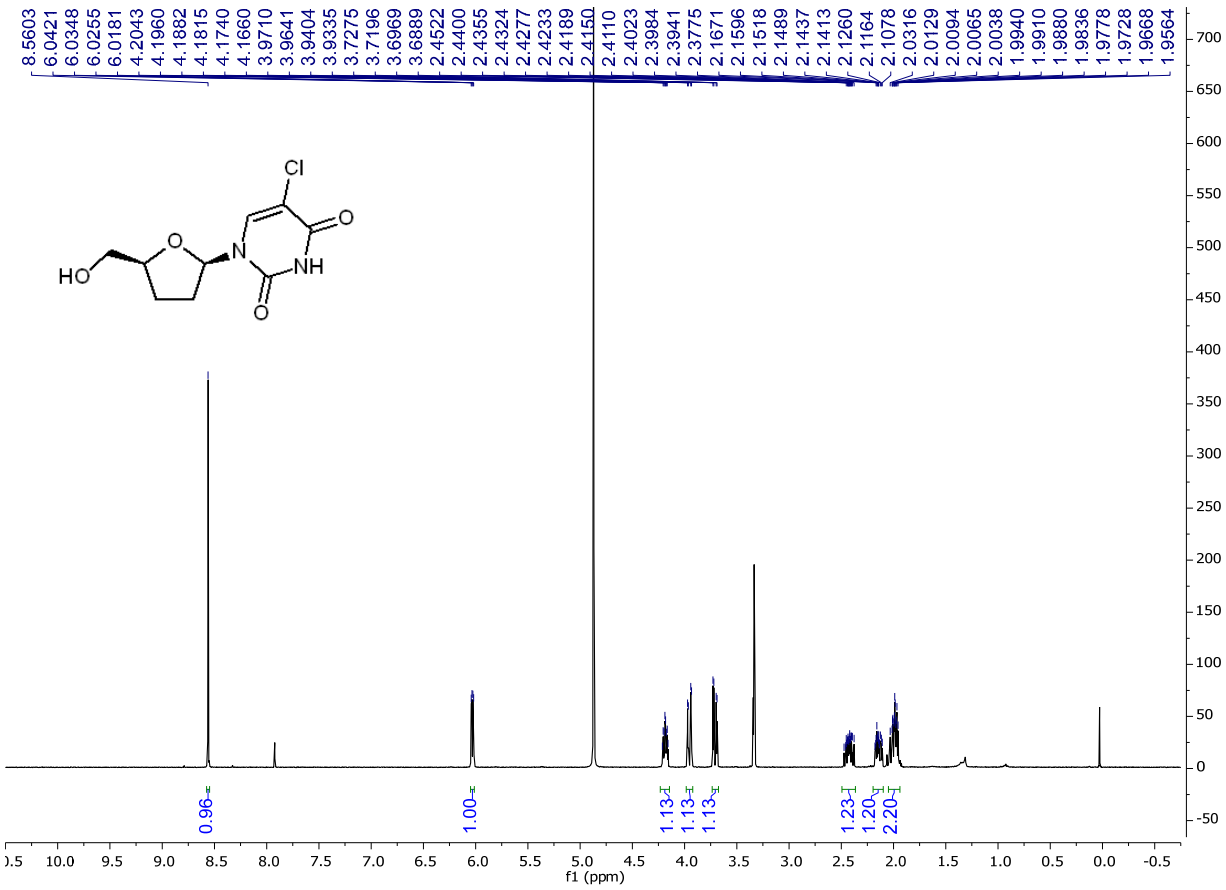


Figure S85. ^{1}H spectrum of **β-ddClU** (400 MHz, CD_3OD).

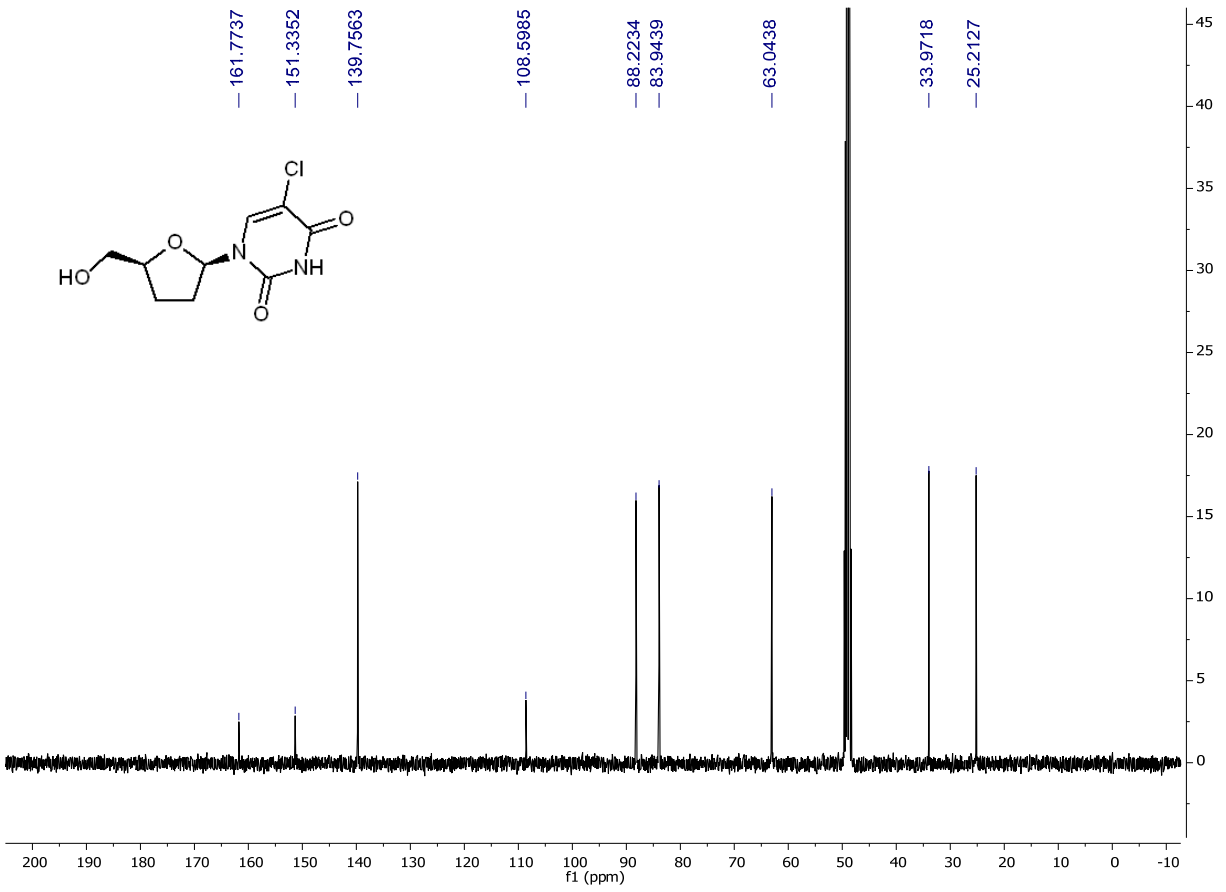


Figure S86. ^{13}C spectrum of **β-ddClU** (100 MHz, CD_3OD).

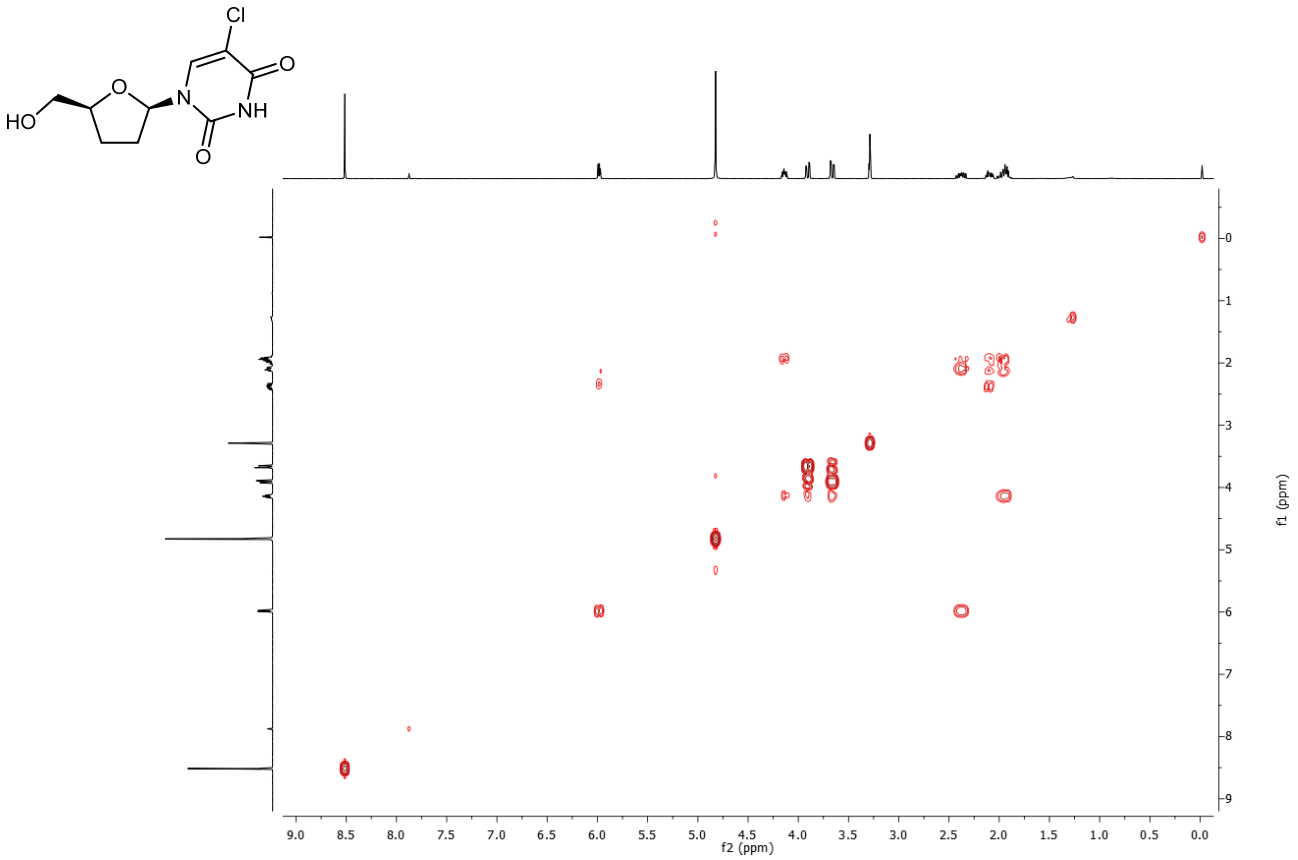


Figure S87. COSY spectrum of **β-ddClU**.

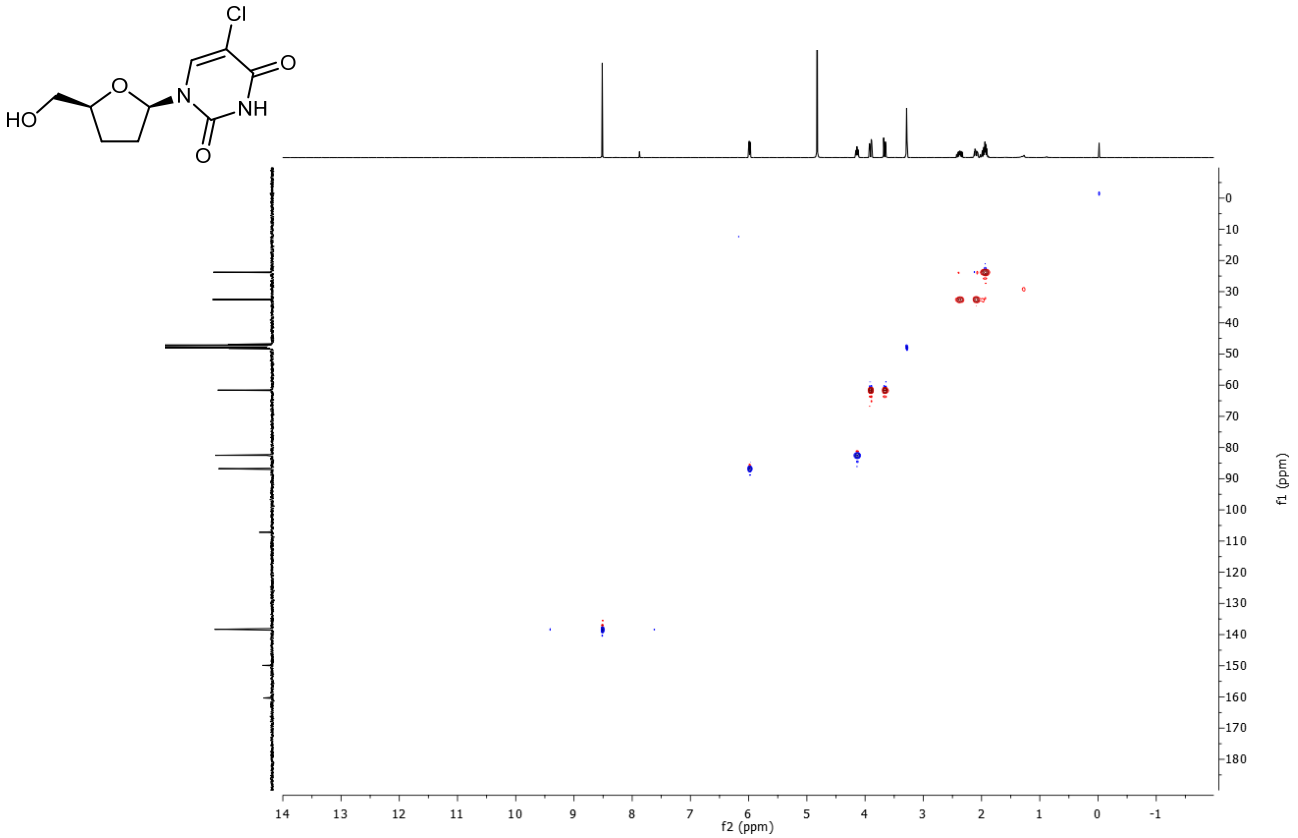


Figure S88. HSQC spectrum of **β-ddClU**.

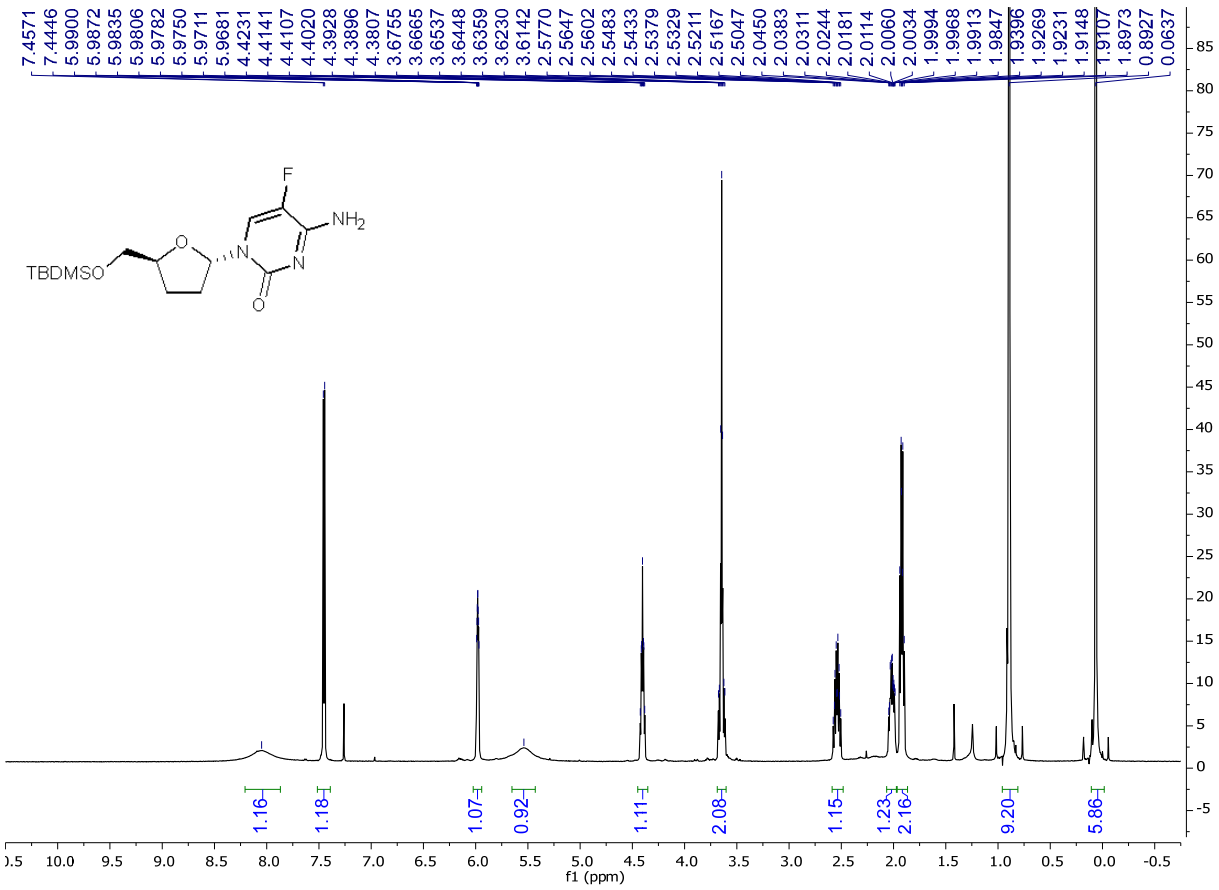


Figure S89. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddFC** (500 MHz, $CDCl_3$).

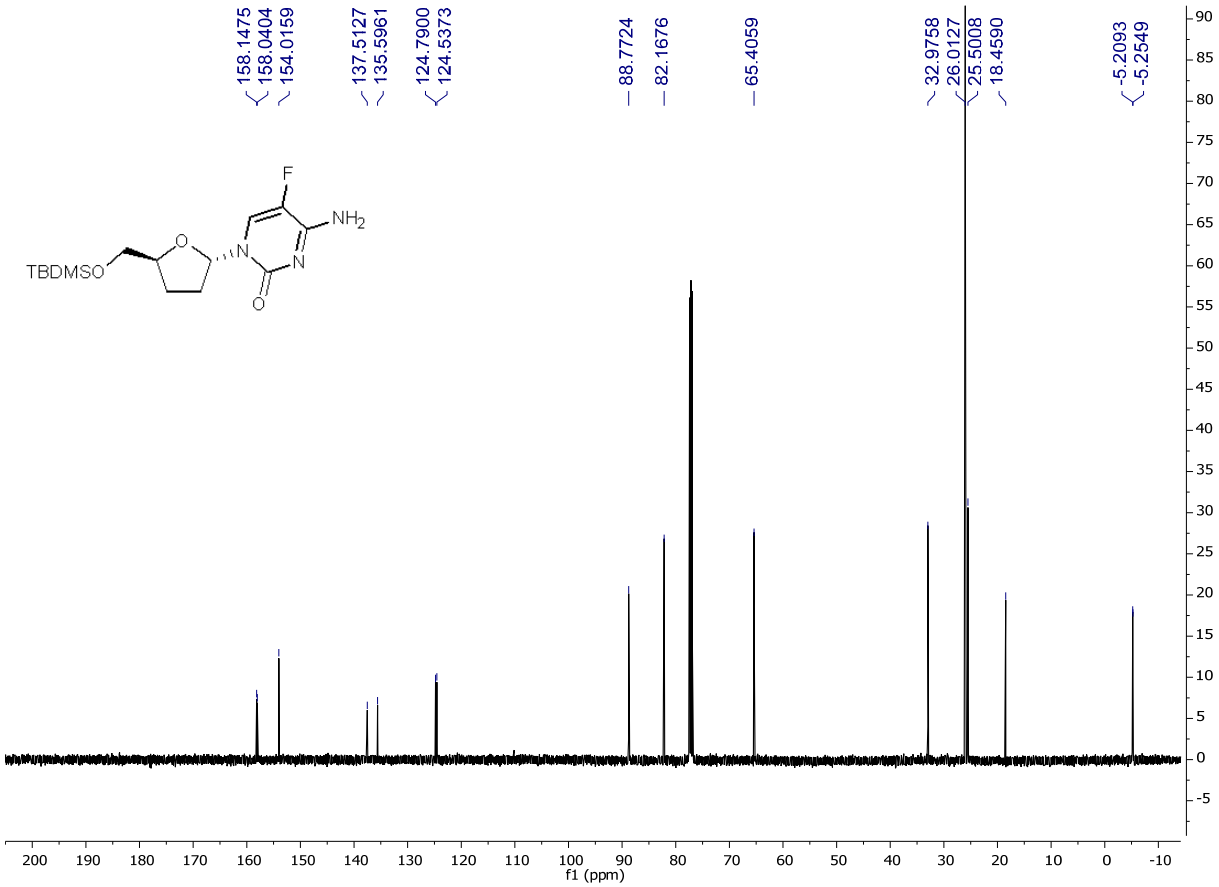


Figure S90. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddFC** (125 MHz, $CDCl_3$).

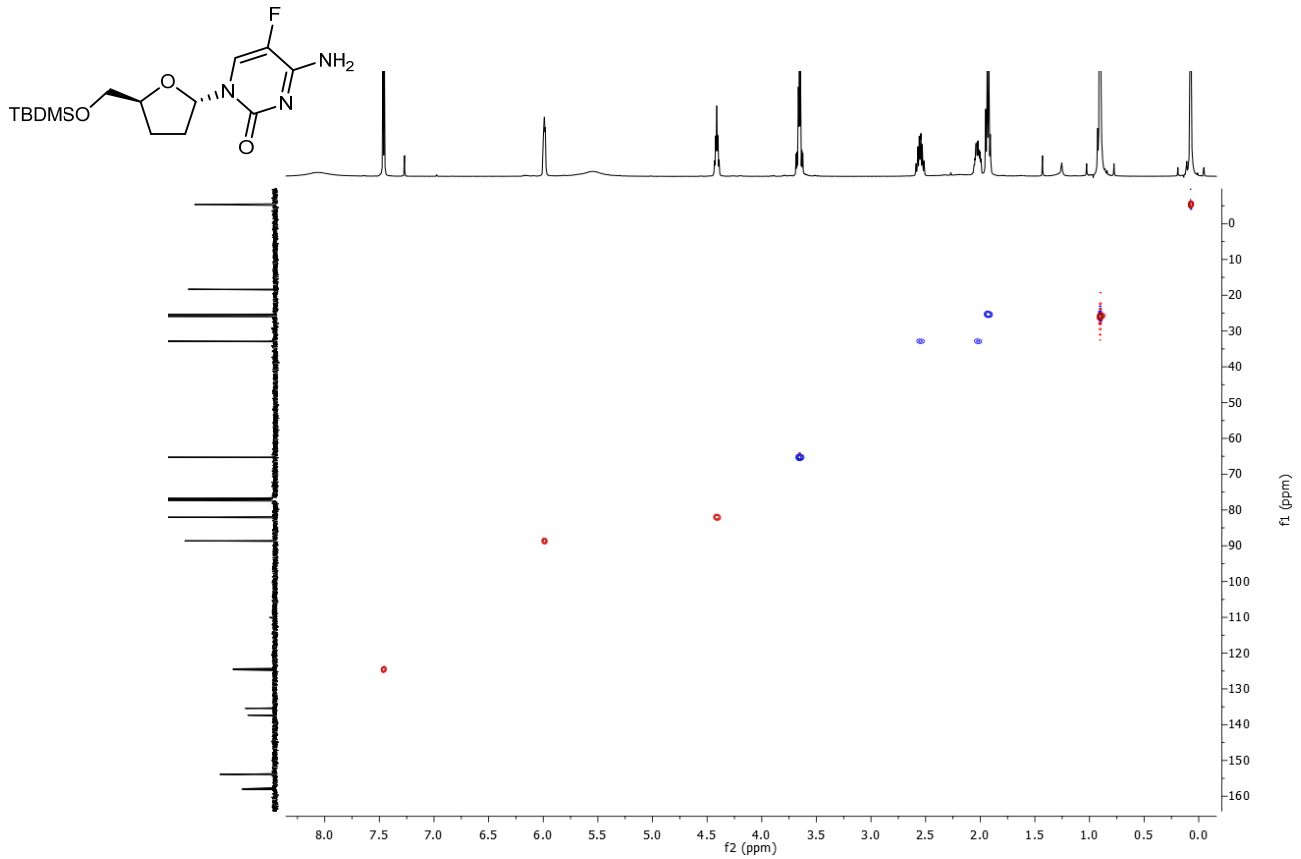


Figure S91. HSQC spectrum of **5'-*O*-silyl-protected α-ddFC**.

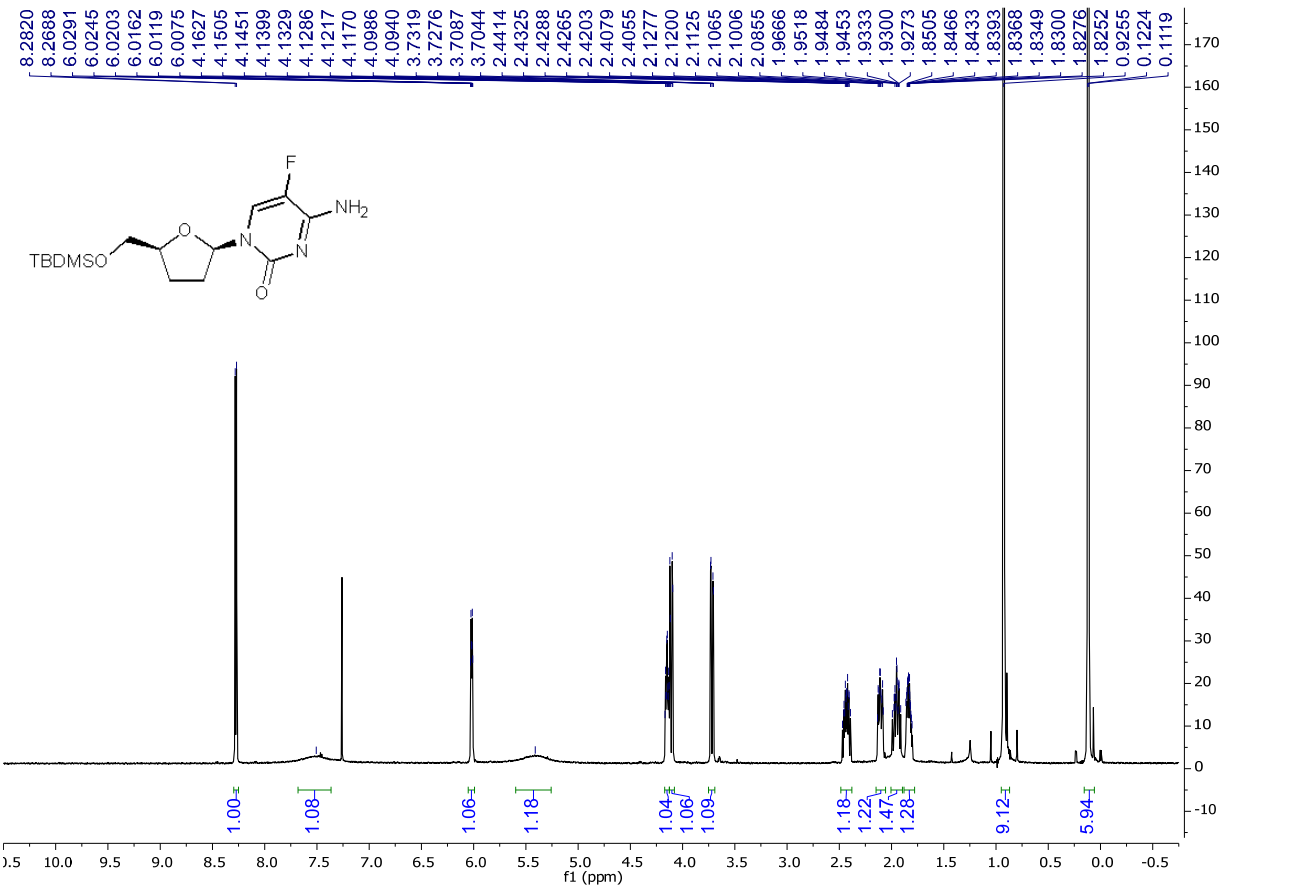


Figure S92. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddFC** (500 MHz, $CDCl_3$).

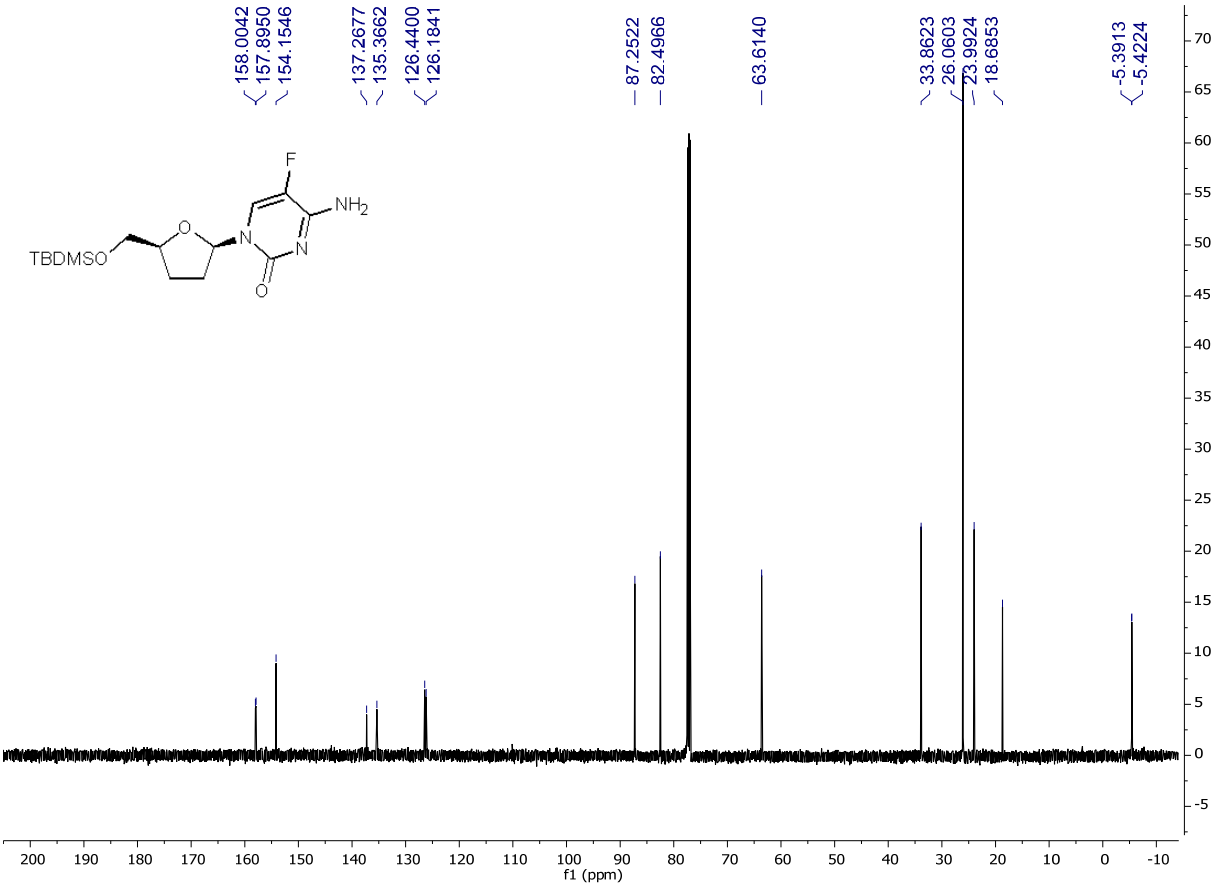


Figure S93. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddFC** (125 MHz, $CDCl_3$).

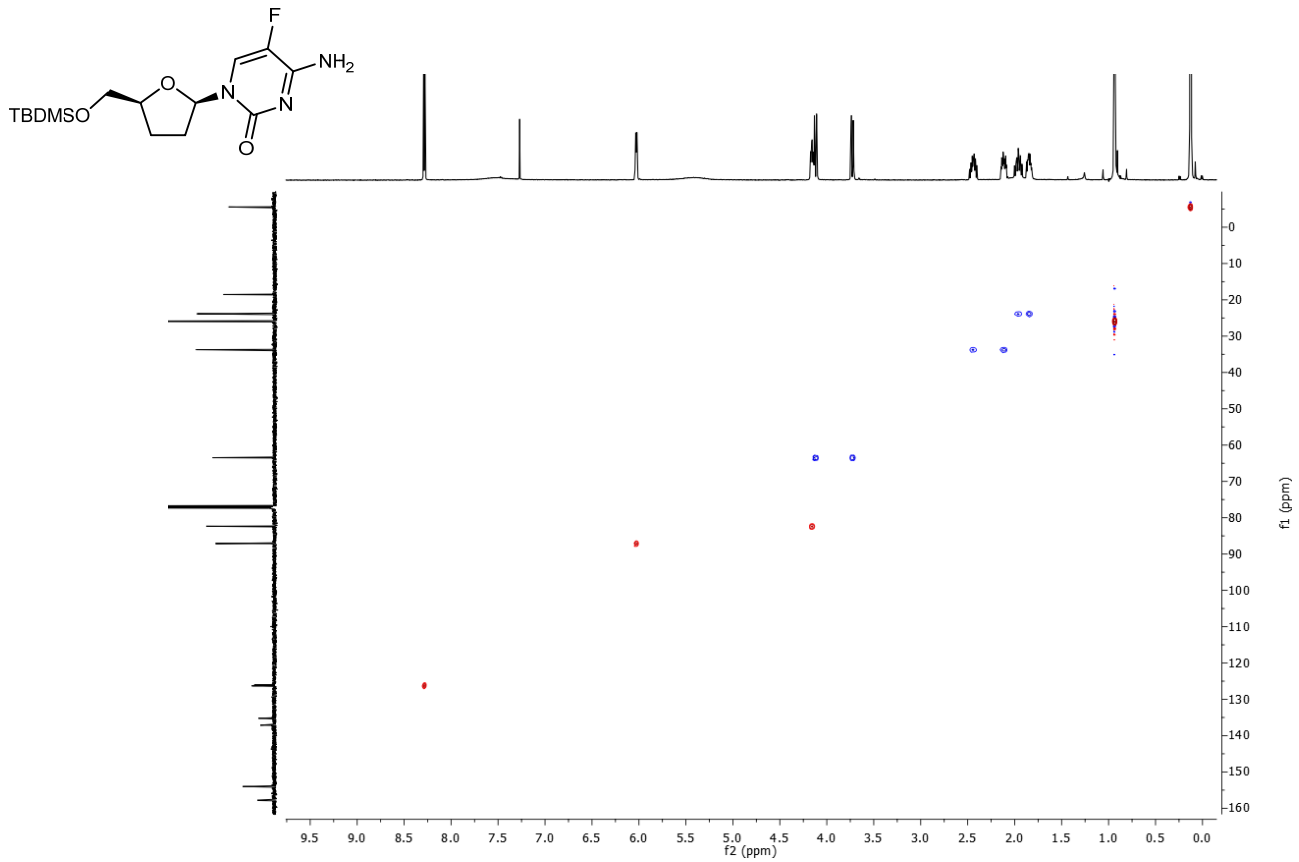


Figure S94. HSQC spectrum of **5'-*O*-silyl-protected β-ddFC**.

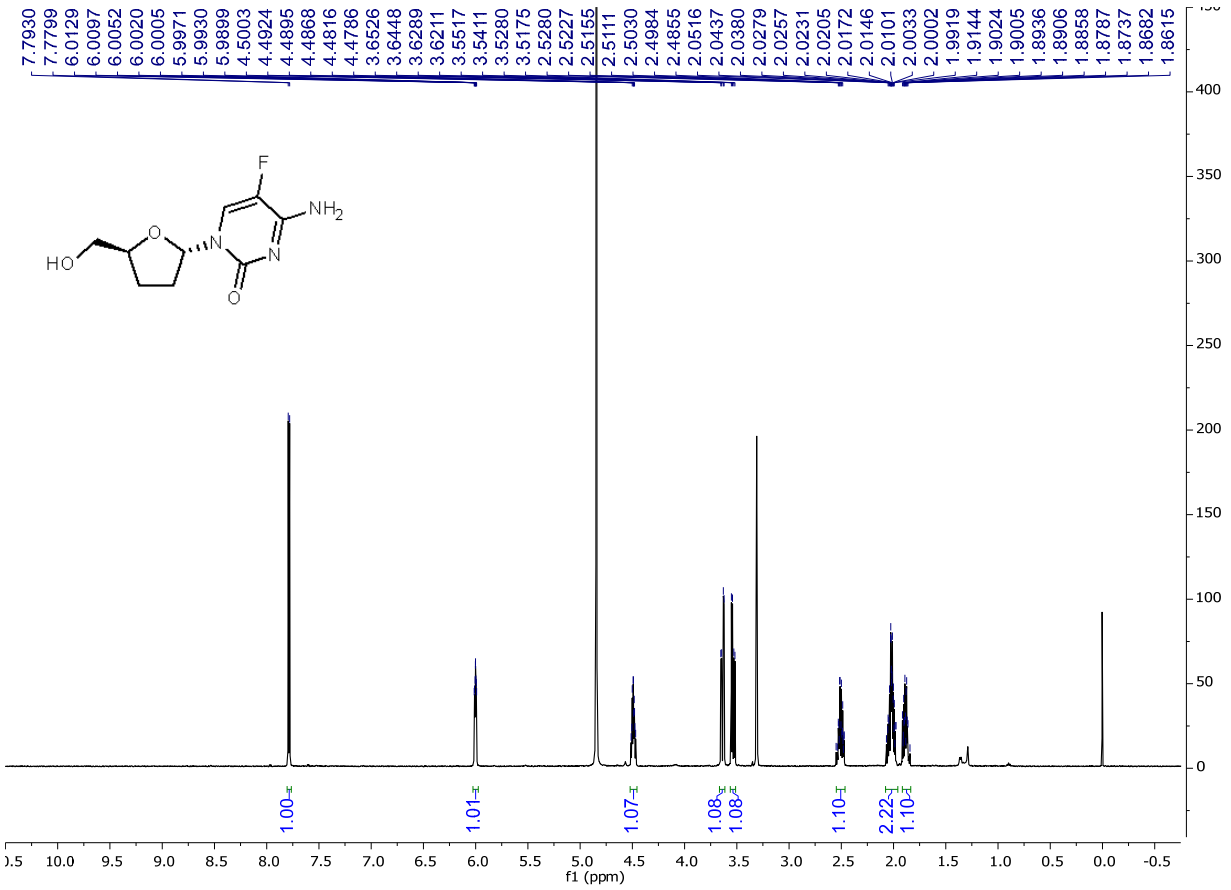


Figure S95. ^{1}H spectrum of **α-ddFC** (500 MHz, CD_3OD).

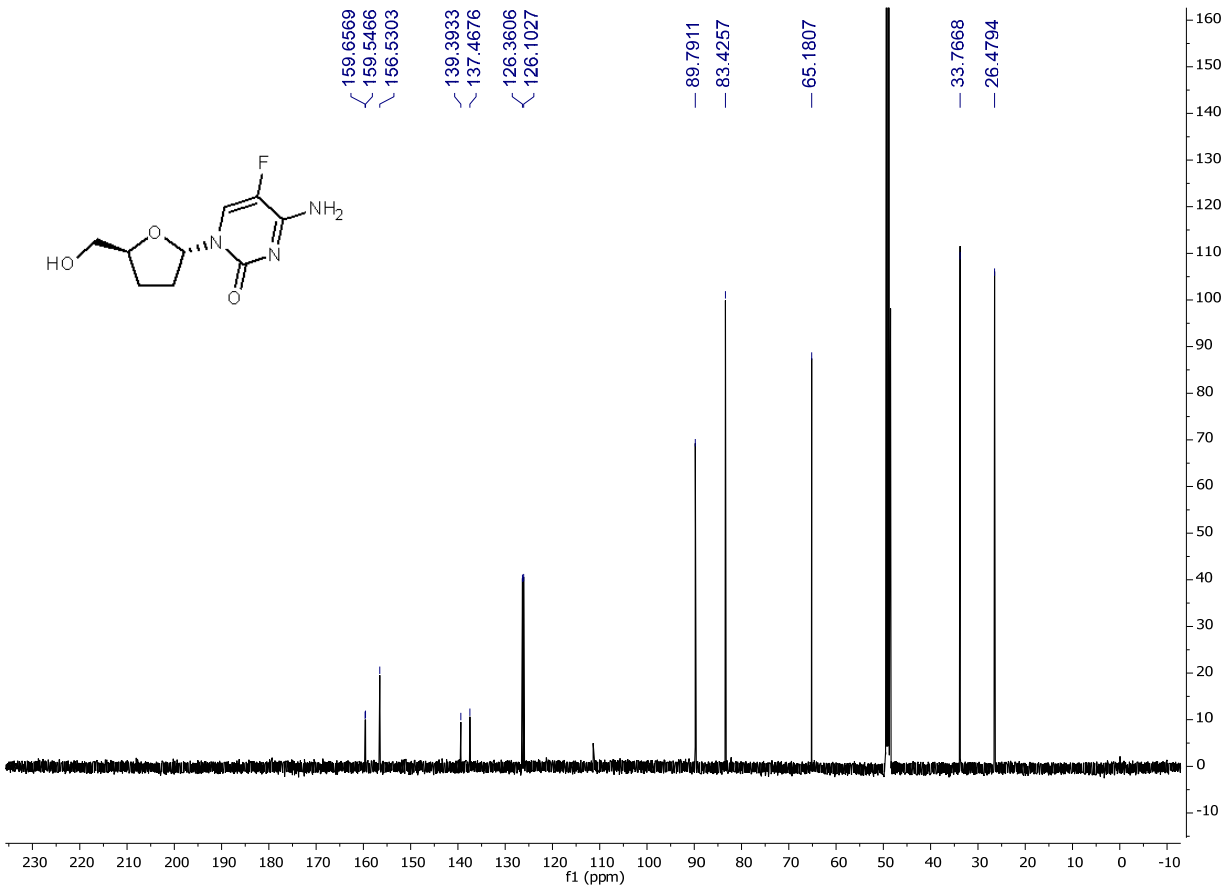


Figure S96. ^{13}C spectrum of **α-ddFC** (125 MHz, CD_3OD).

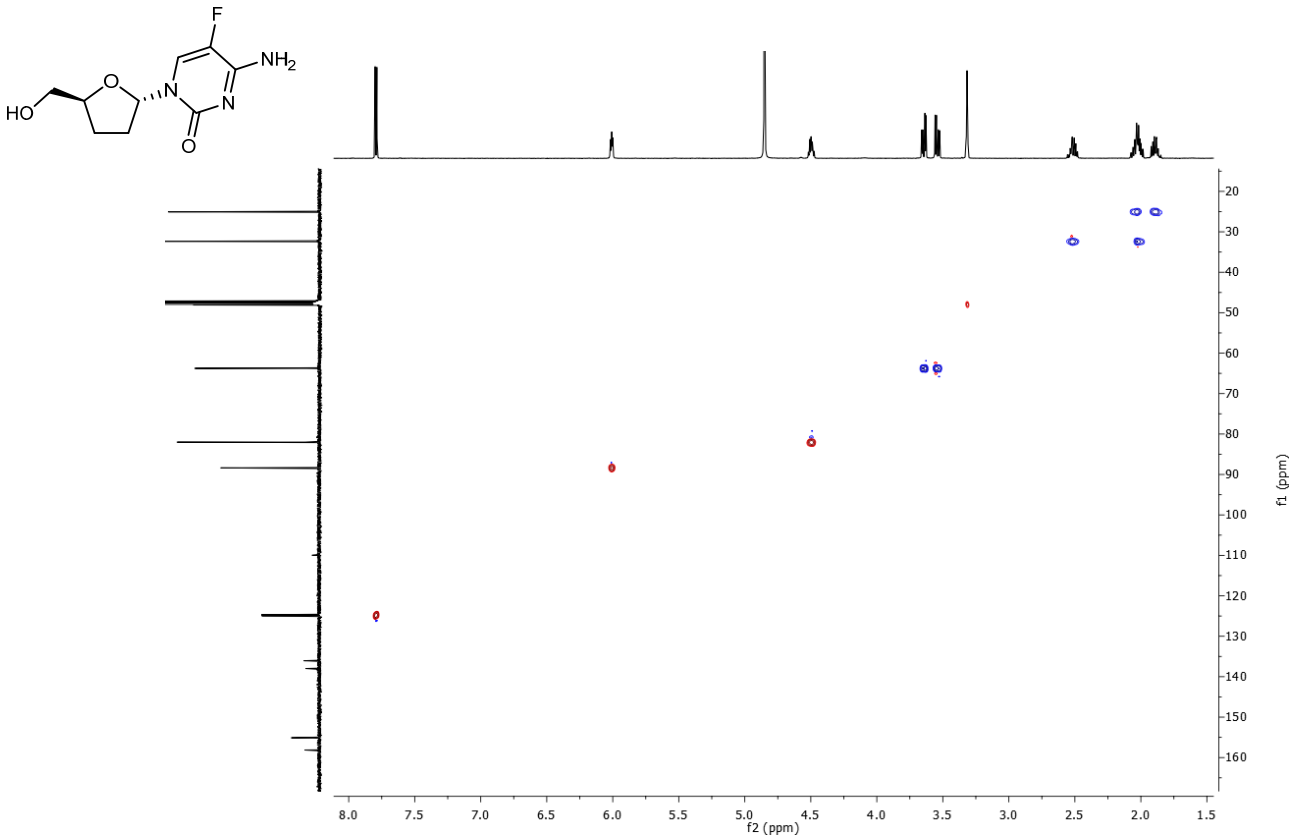


Figure S97. HSQC spectrum of **α-ddFC**.

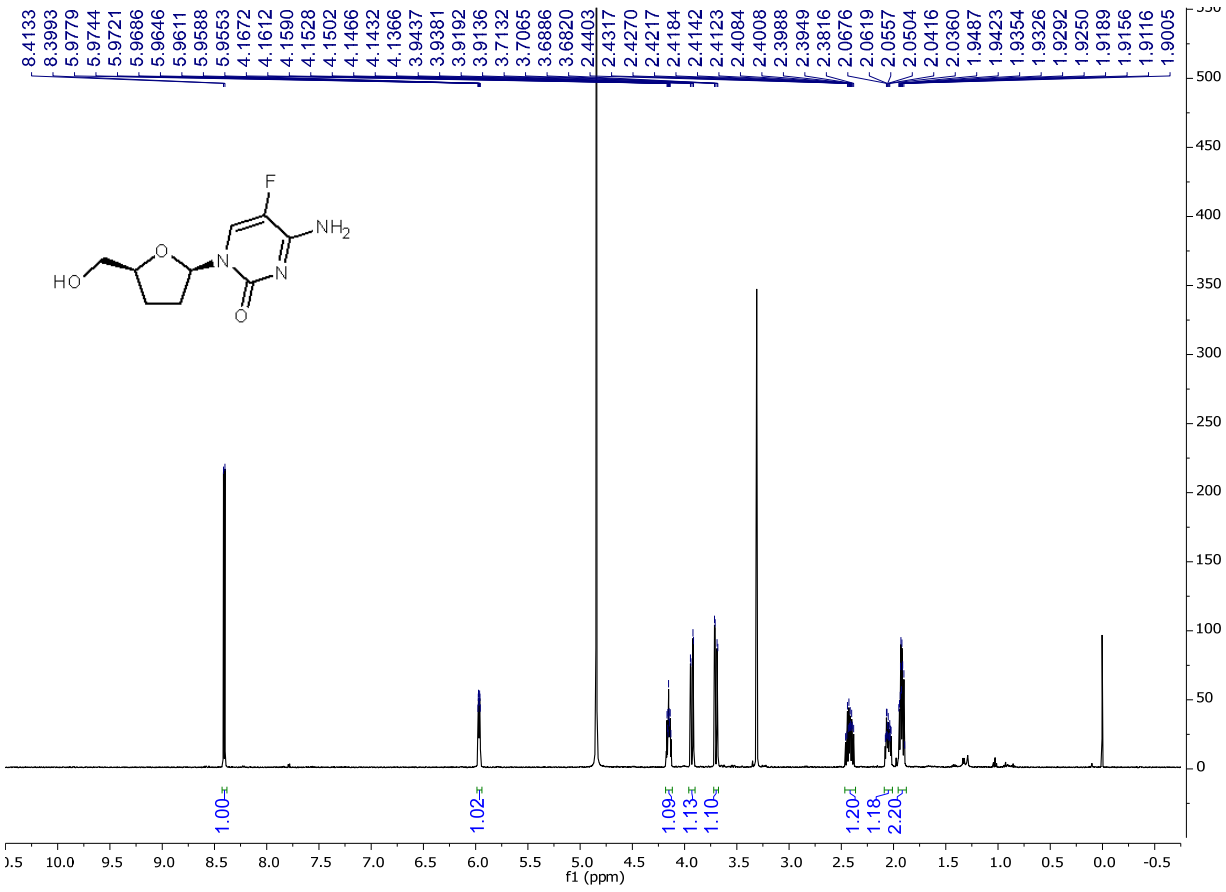


Figure S98. ^{1}H spectrum of **β-ddFC** (500 MHz, CD_3OD).

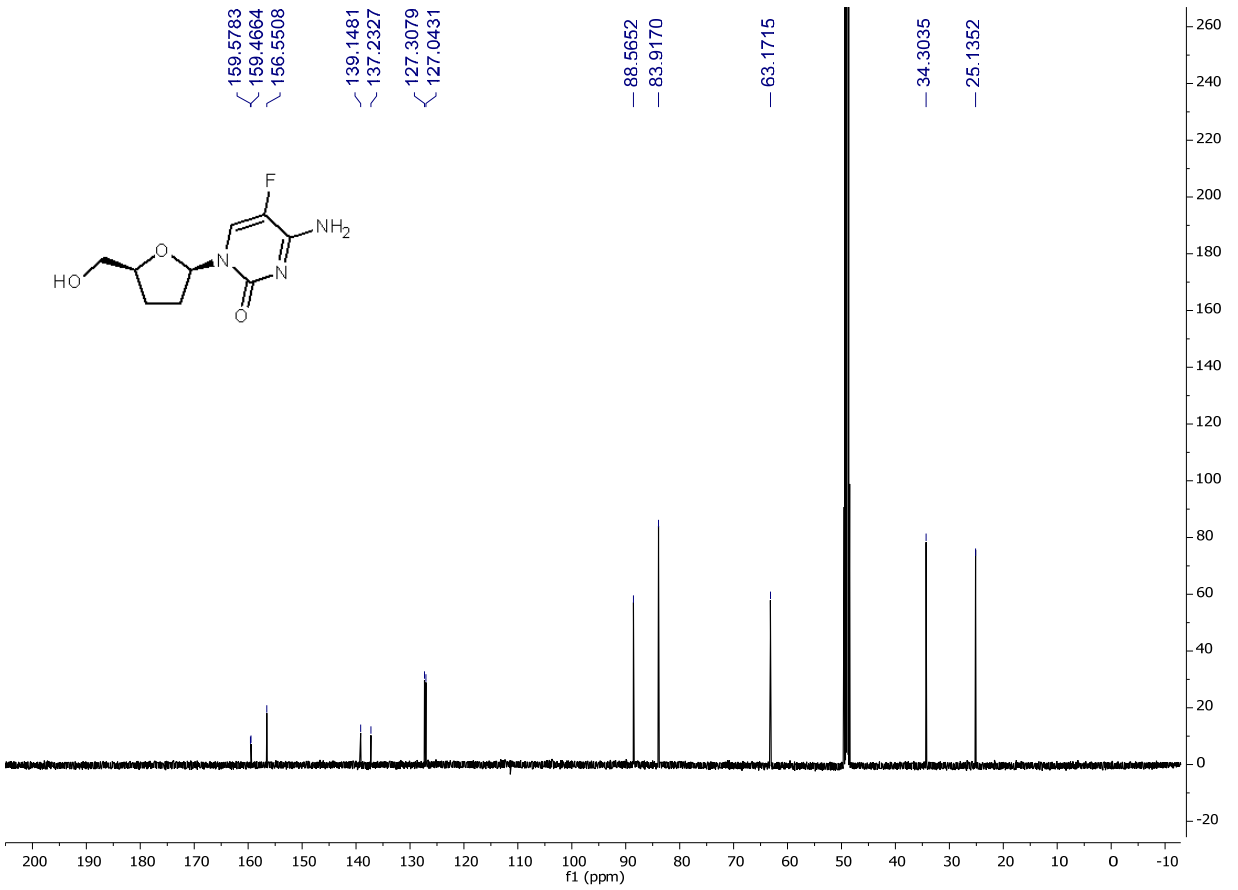


Figure S99. ^{13}C spectrum of **β-ddFC** (125 MHz, CD_3OD).

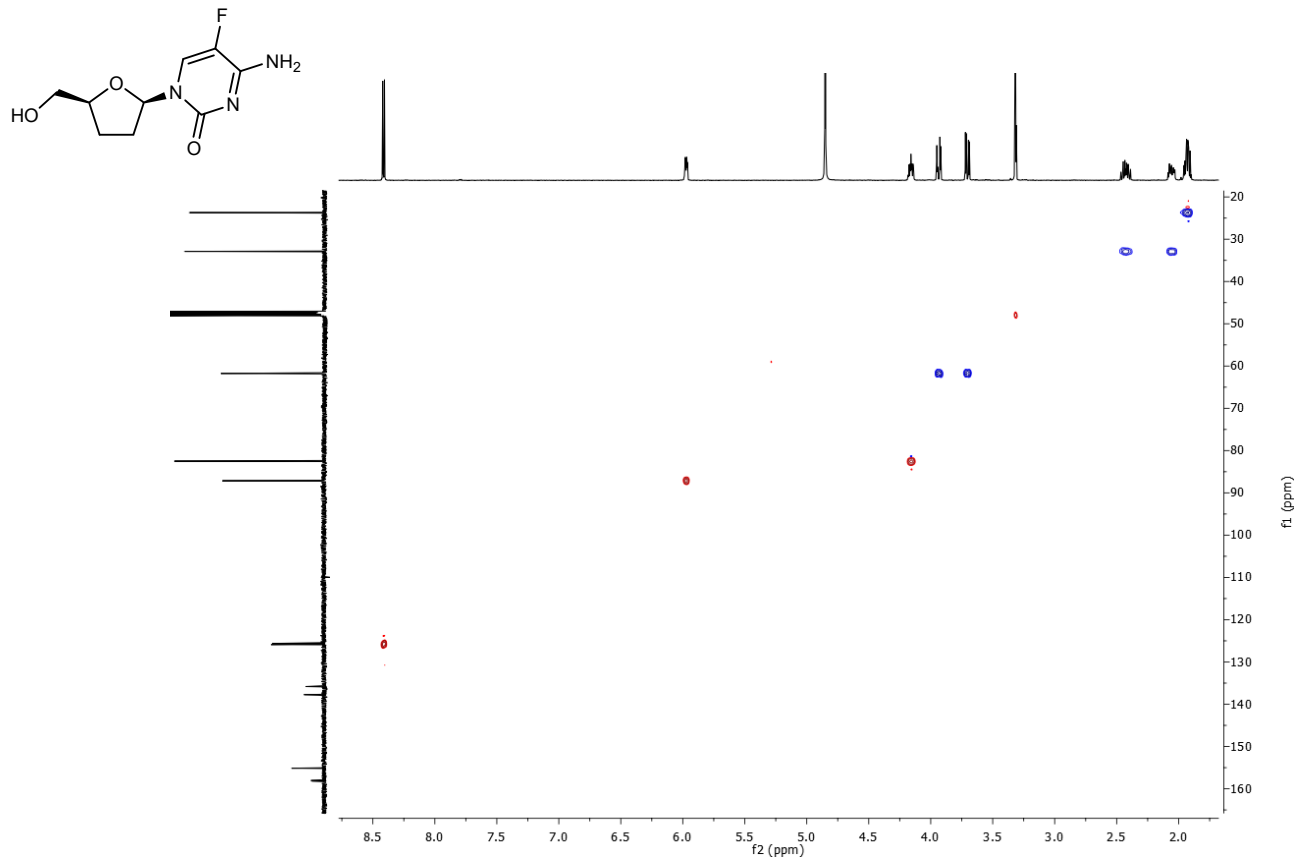


Figure S100. HSQC spectrum of **β-ddFC**.

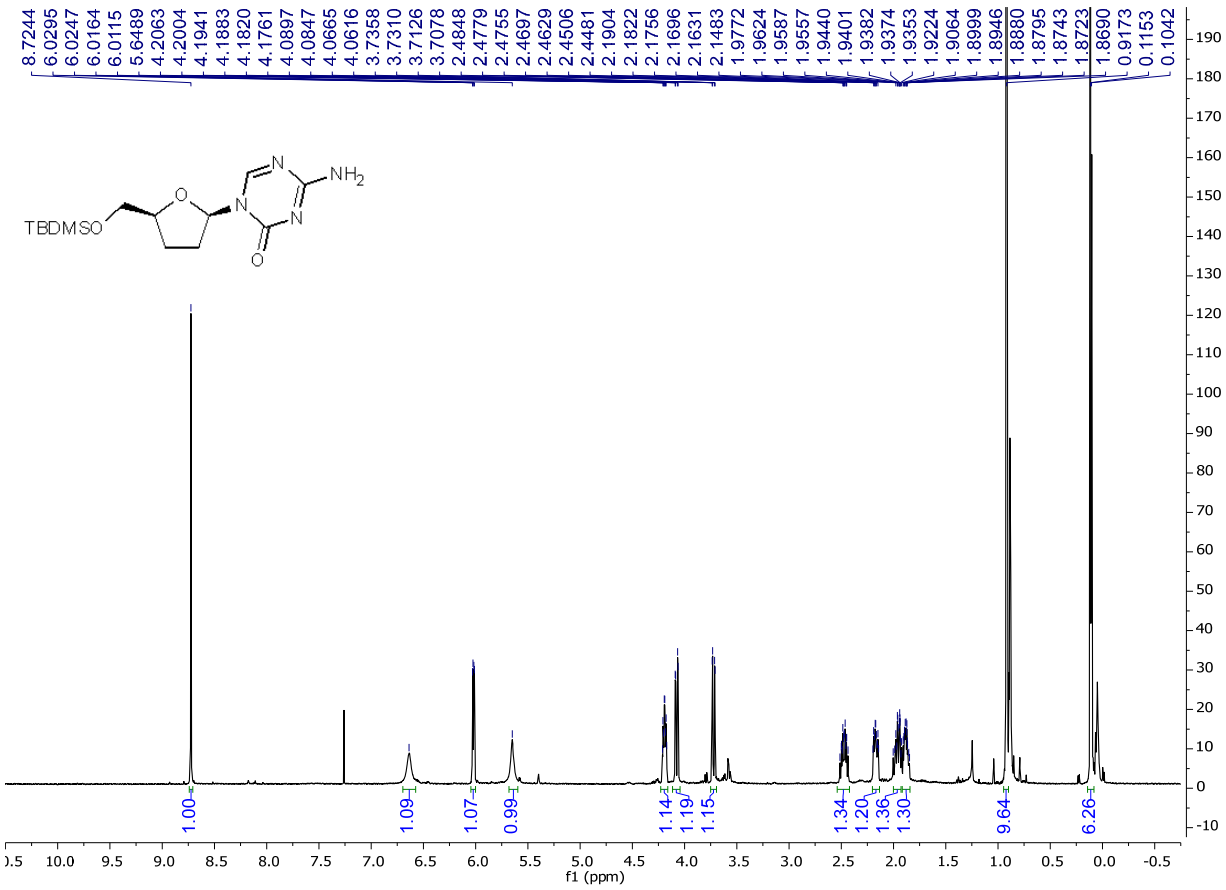


Figure S101. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddAC** (500 MHz, $CDCl_3$).

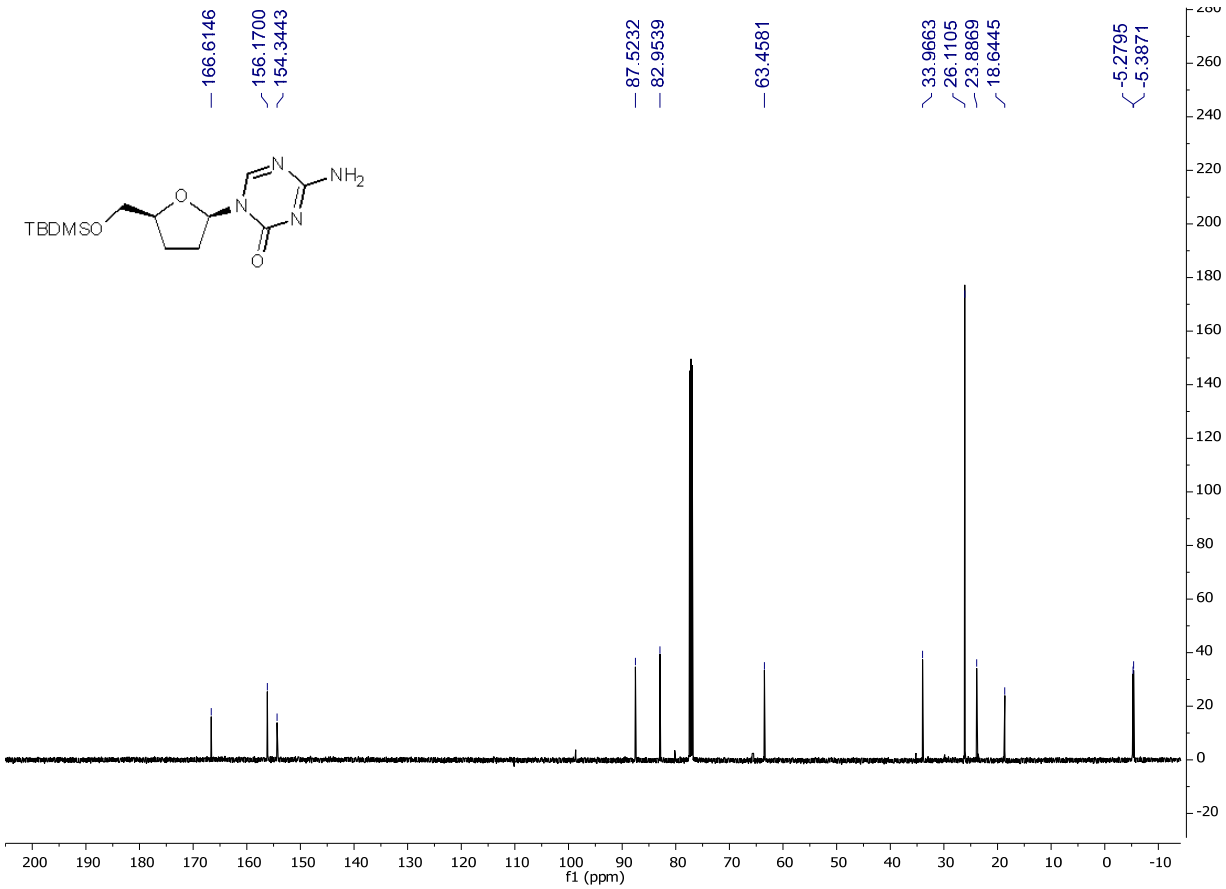


Figure S102. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddAC** (125 MHz, $CDCl_3$).

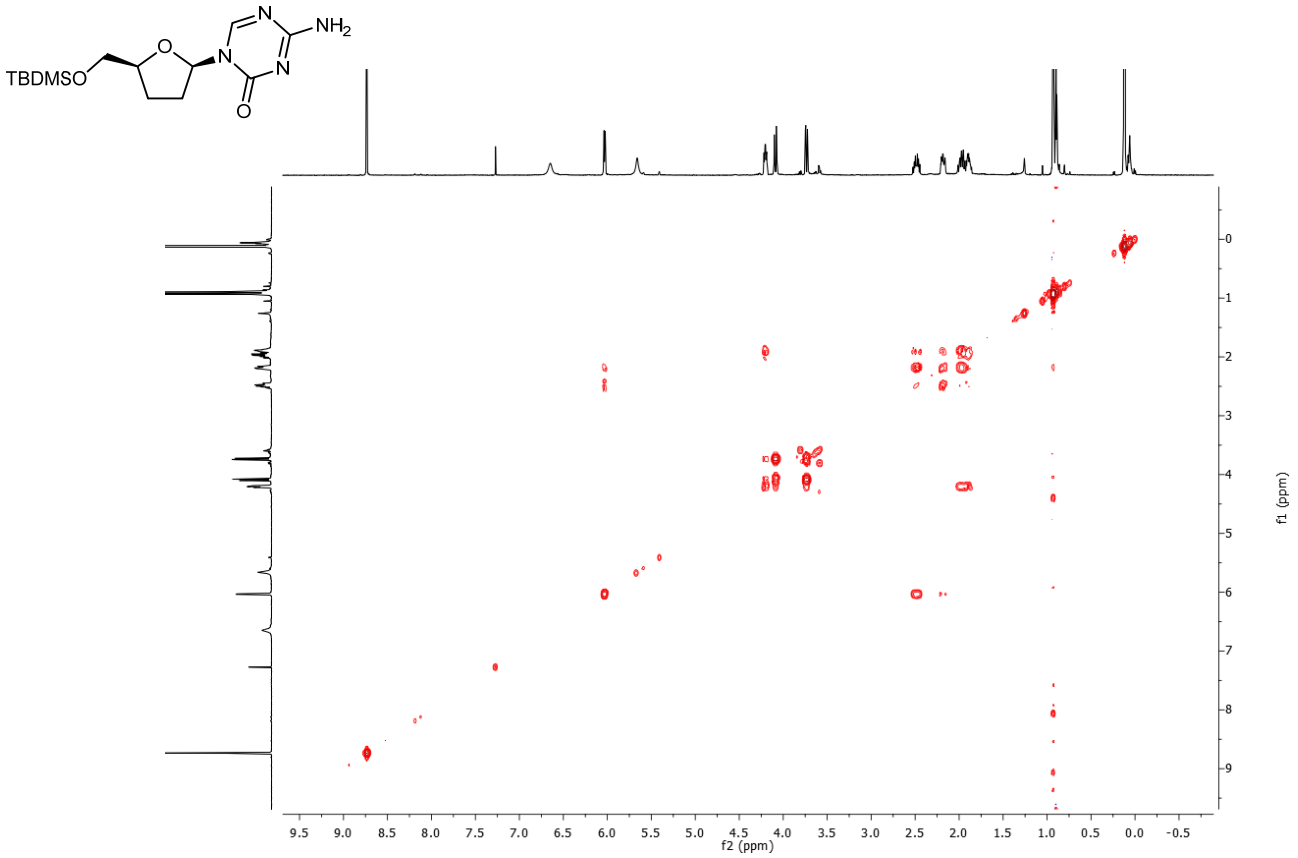


Figure S103. COSY spectrum of **5'-*O*-silyl-protected β-ddAC**.

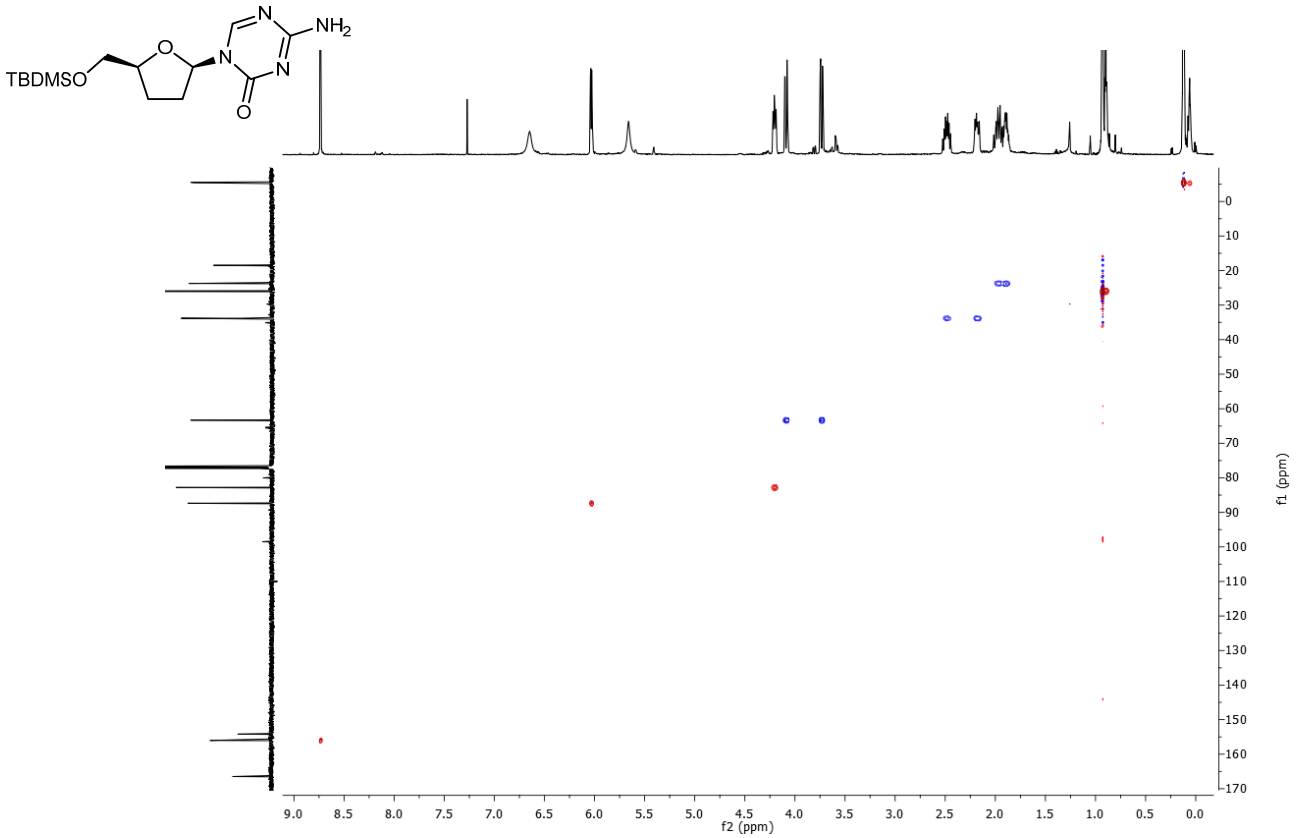


Figure S104. HSQC spectrum of **5'-*O*-silyl-protected β-ddAC**.

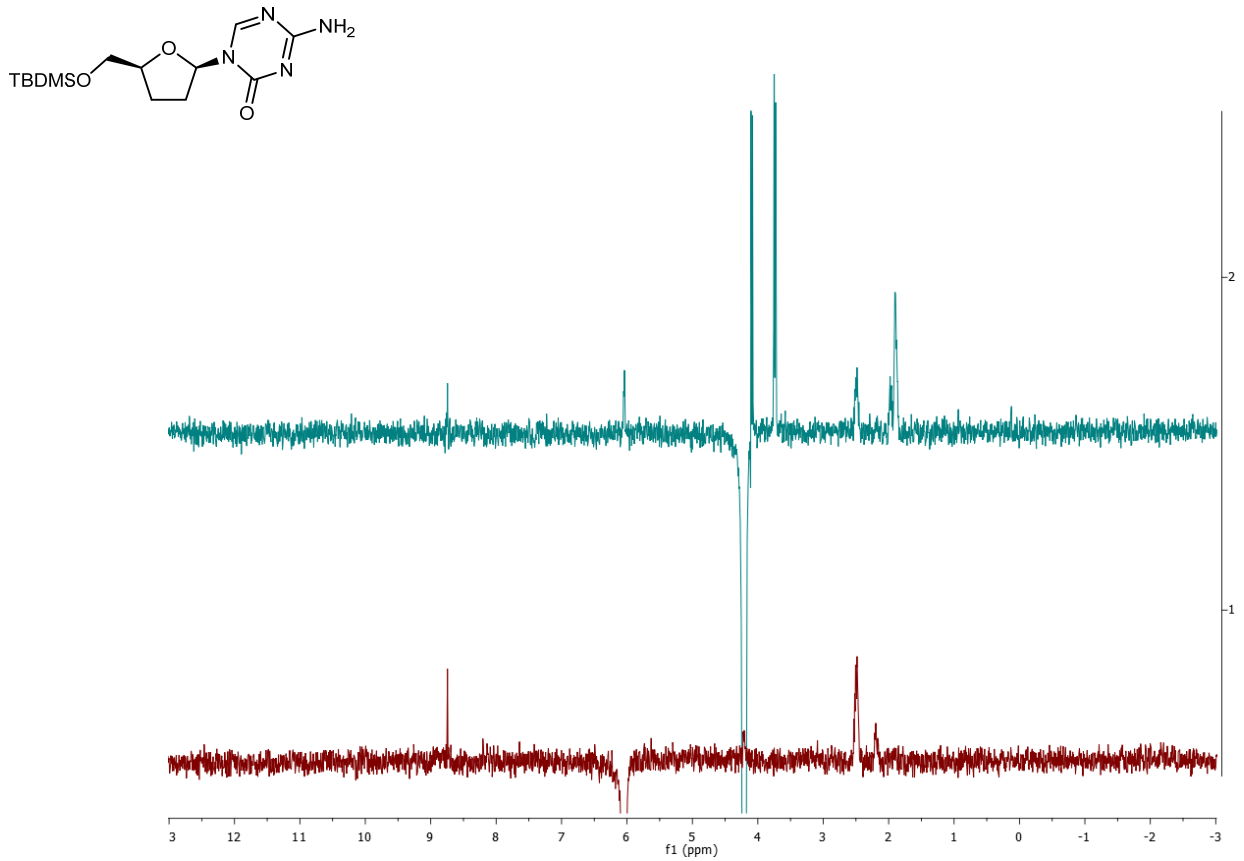


Figure S105. 1D-NOESY spectra of **5'-*O*-silyl-protected β-ddAC**.

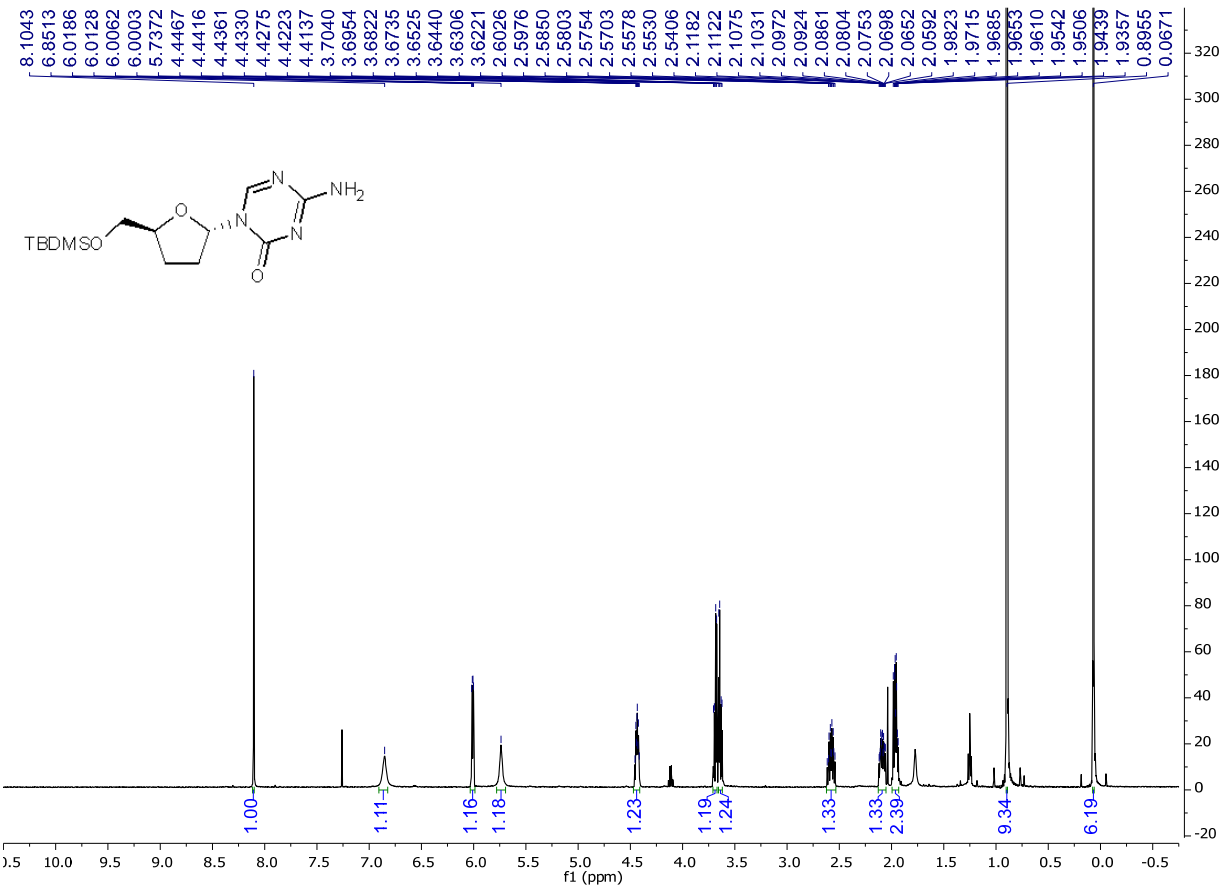


Figure S106. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddAC** (500 MHz, $CDCl_3$).

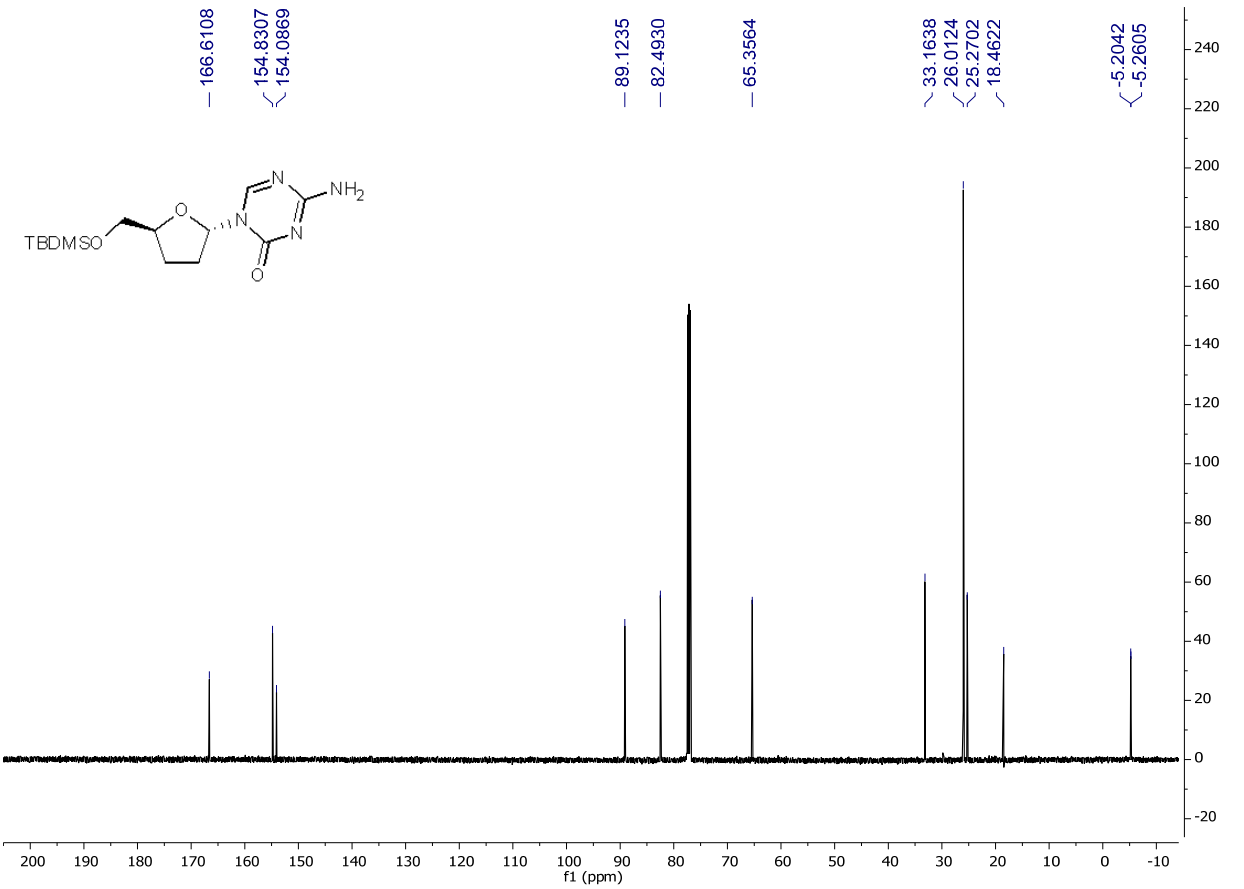


Figure S107. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddAC** (125 MHz, $CDCl_3$).

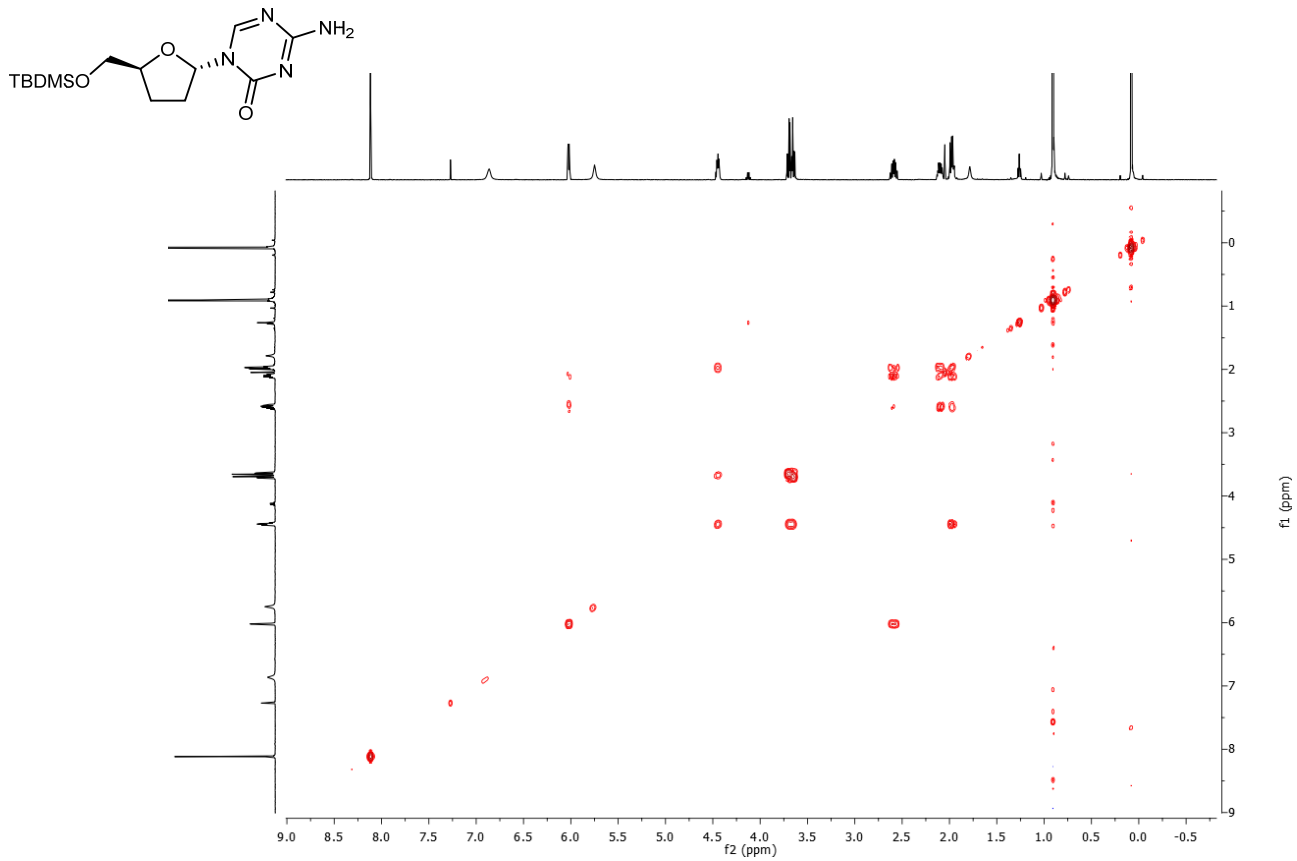


Figure S108. COSY spectrum of **5'-*O*-silyl-protected α-ddAC**.

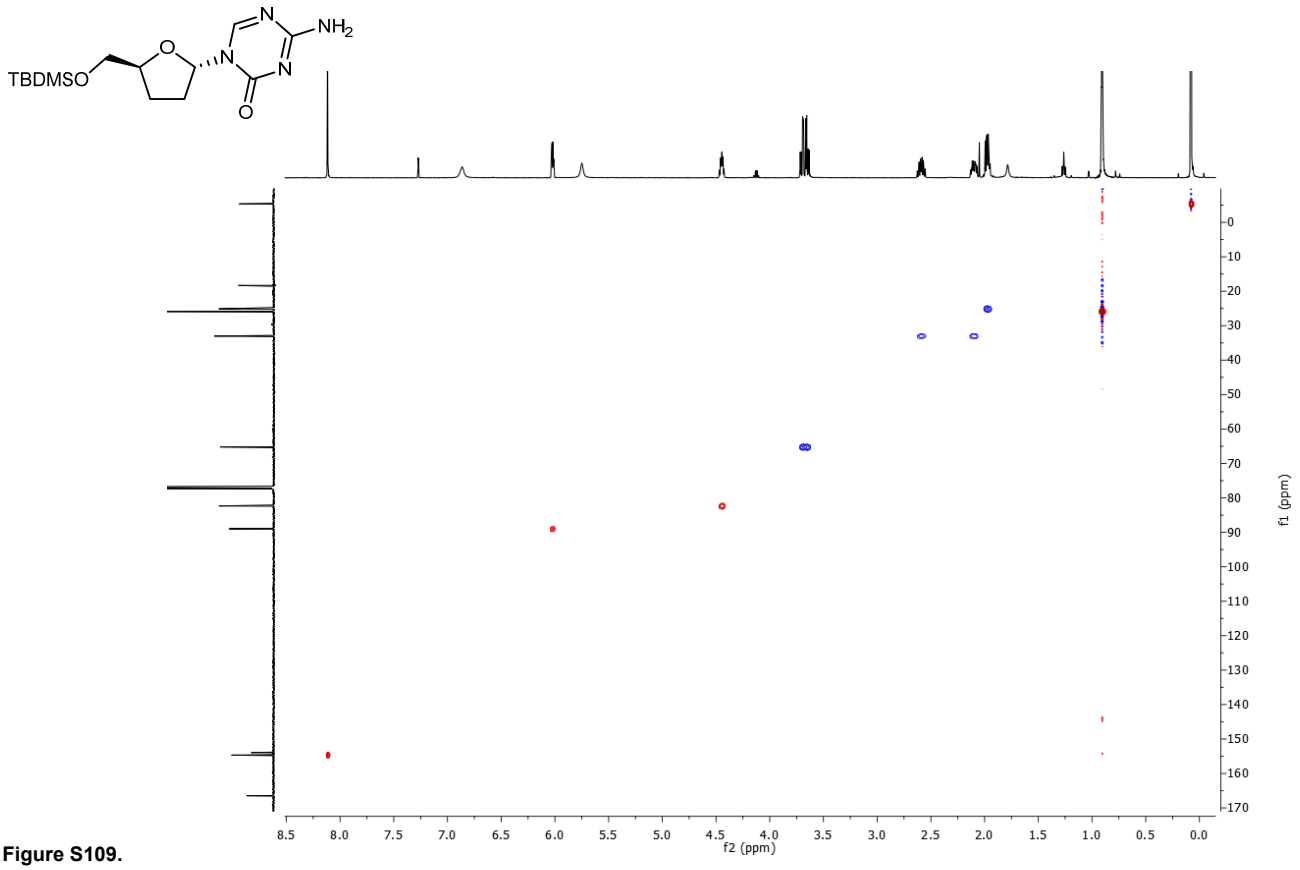


Figure S109.
HSQC spectrum of **5'-*O*-silyl-protected α-ddAC**.

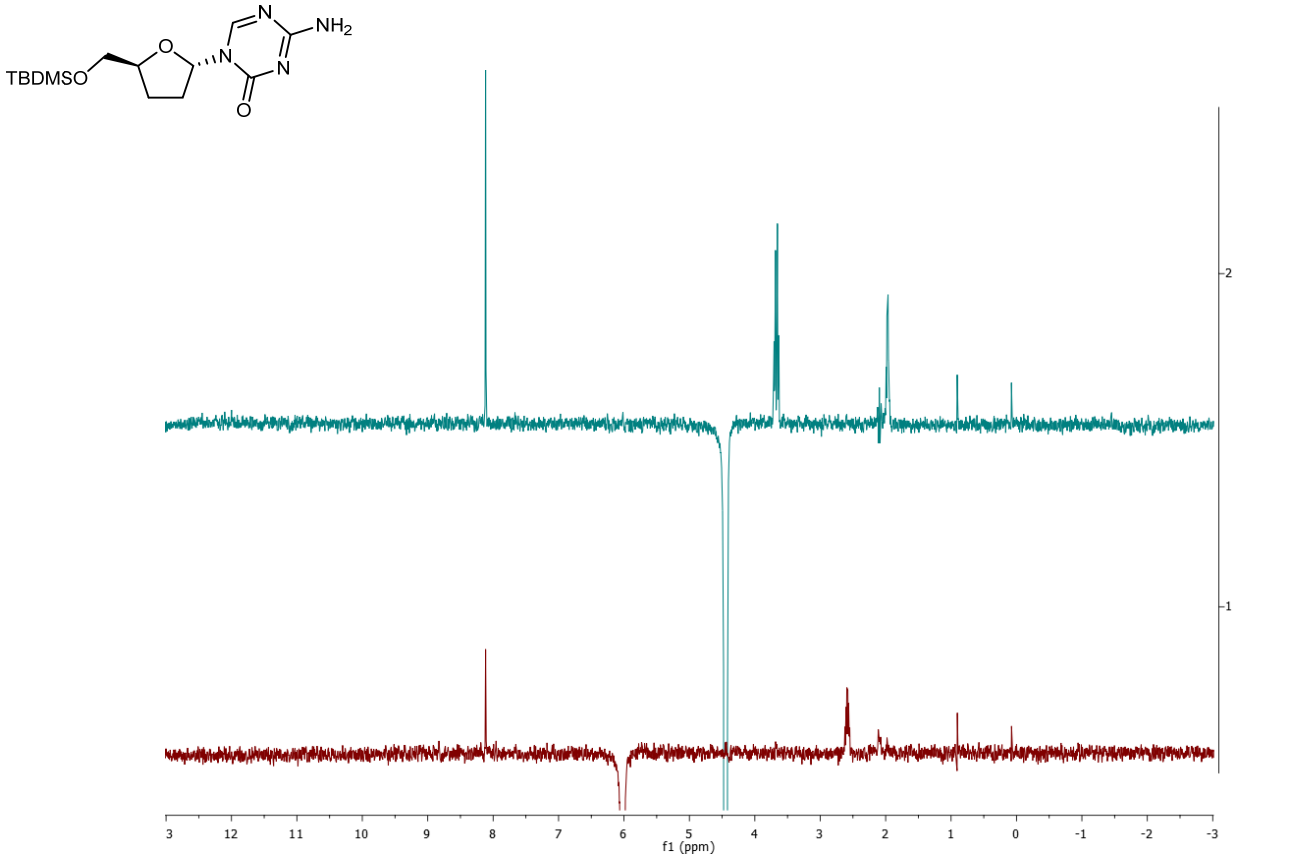


Figure S110. 1D-NOESY spectra of **5'-*O*-silyl-protected α-ddAC**.

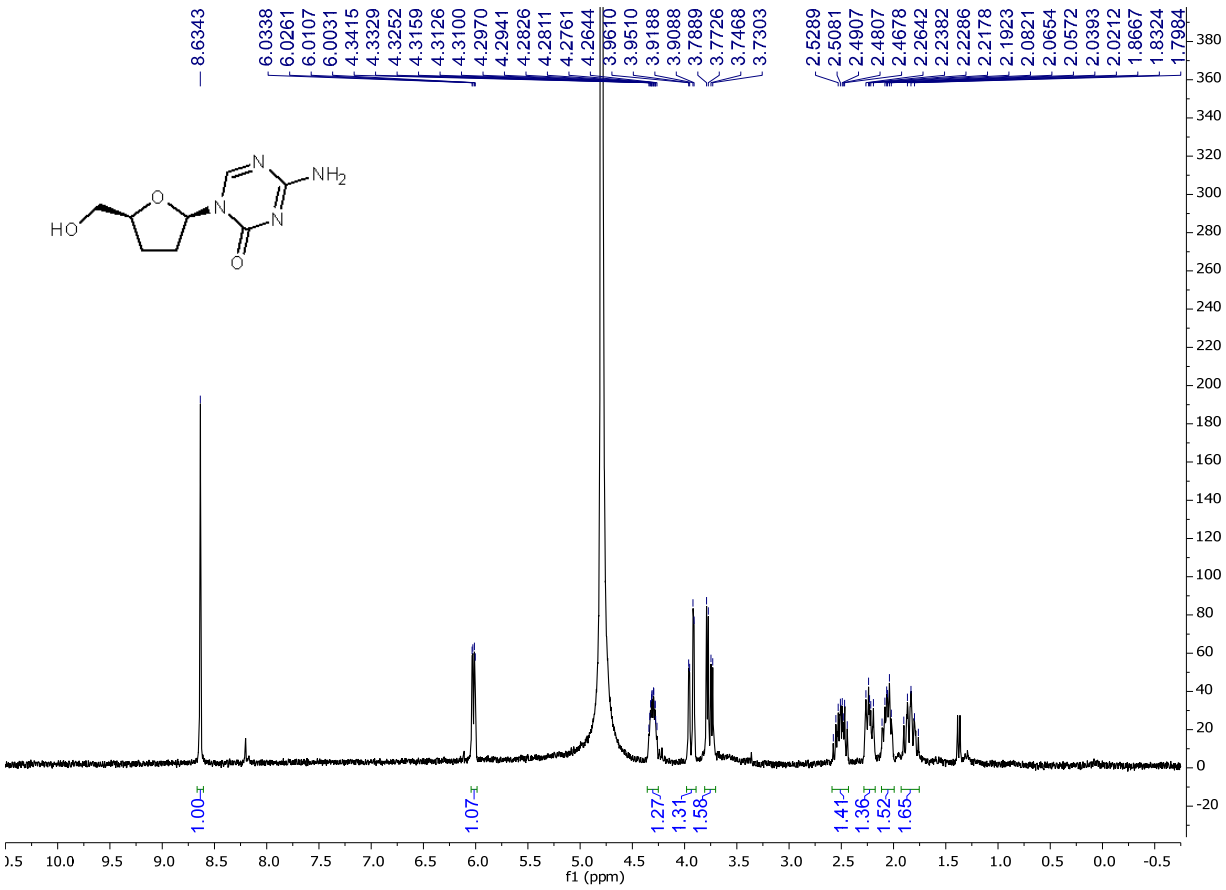


Figure S111. ^{1}H spectrum of **β-ddAC** (300 MHz, D_2O).

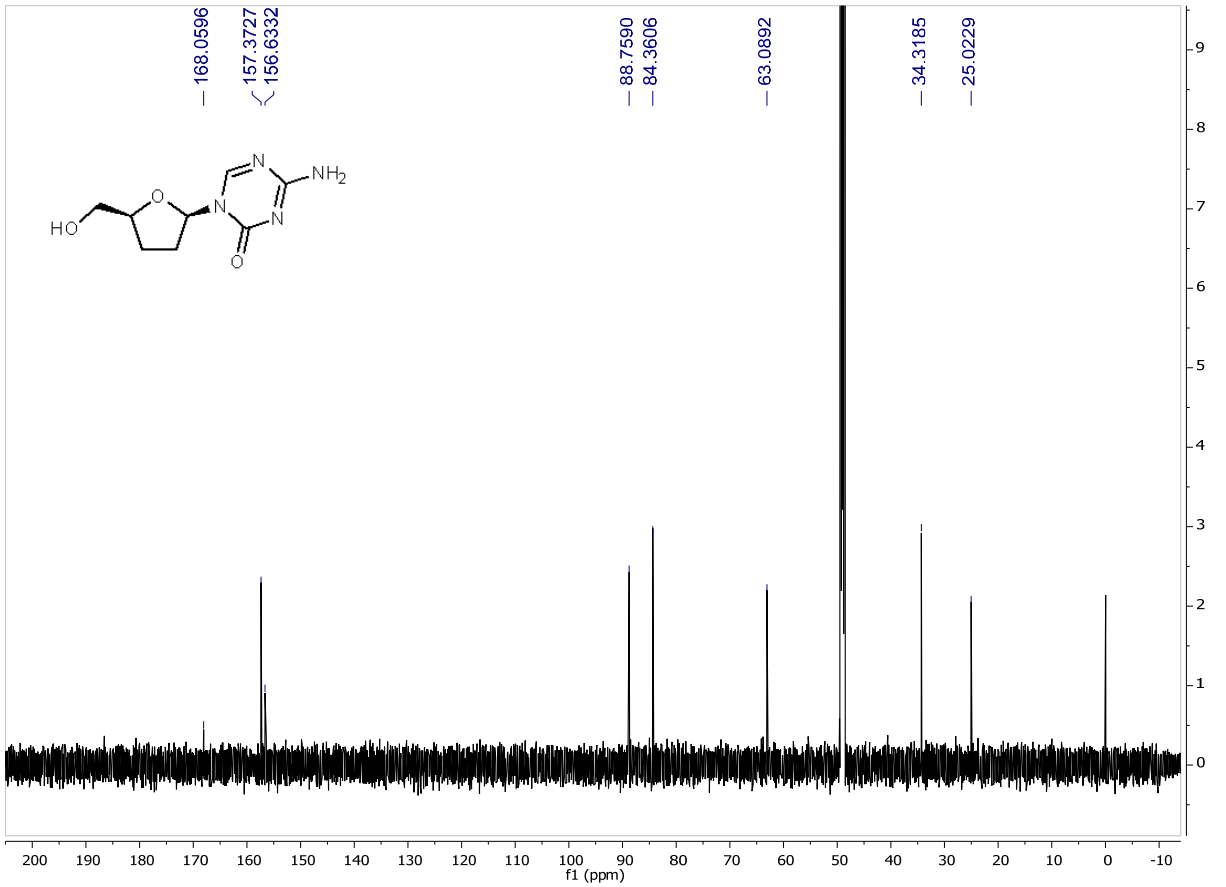


Figure S112. ^{13}C spectrum of **β-ddAC** (150 MHz, CD_3OD).

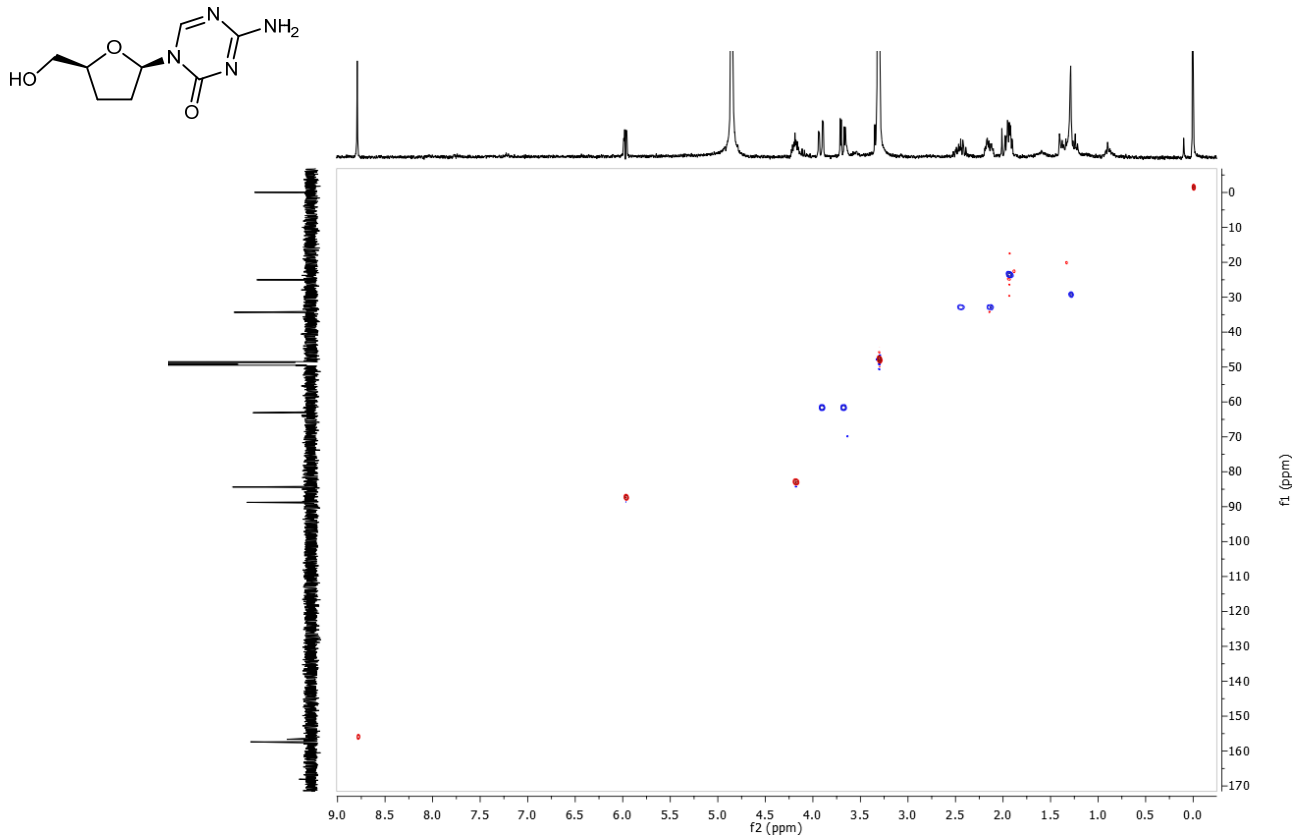


Figure S113. HSQC spectrum of **β-ddAC**.

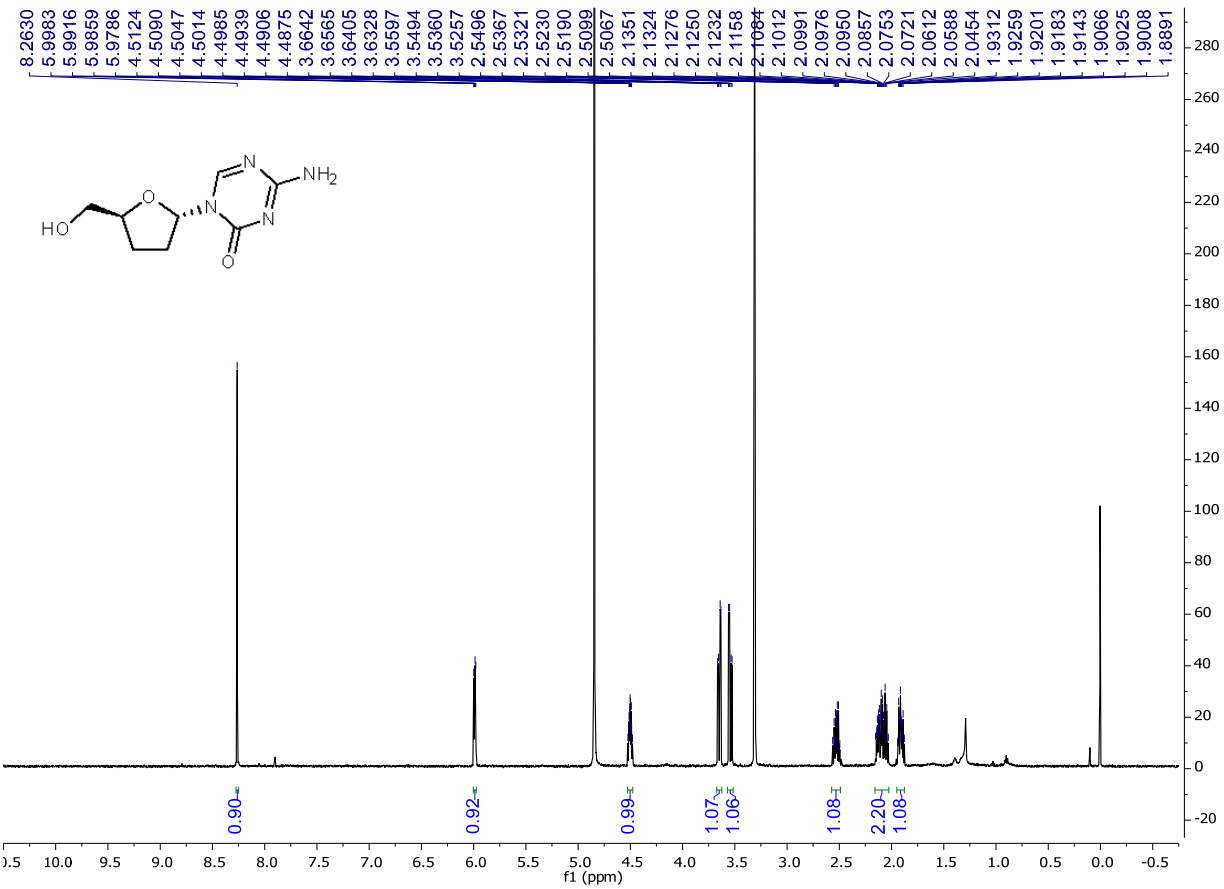


Figure S114. ^{1}H spectrum of **α-ddAC** (500 MHz, CD_3OD).

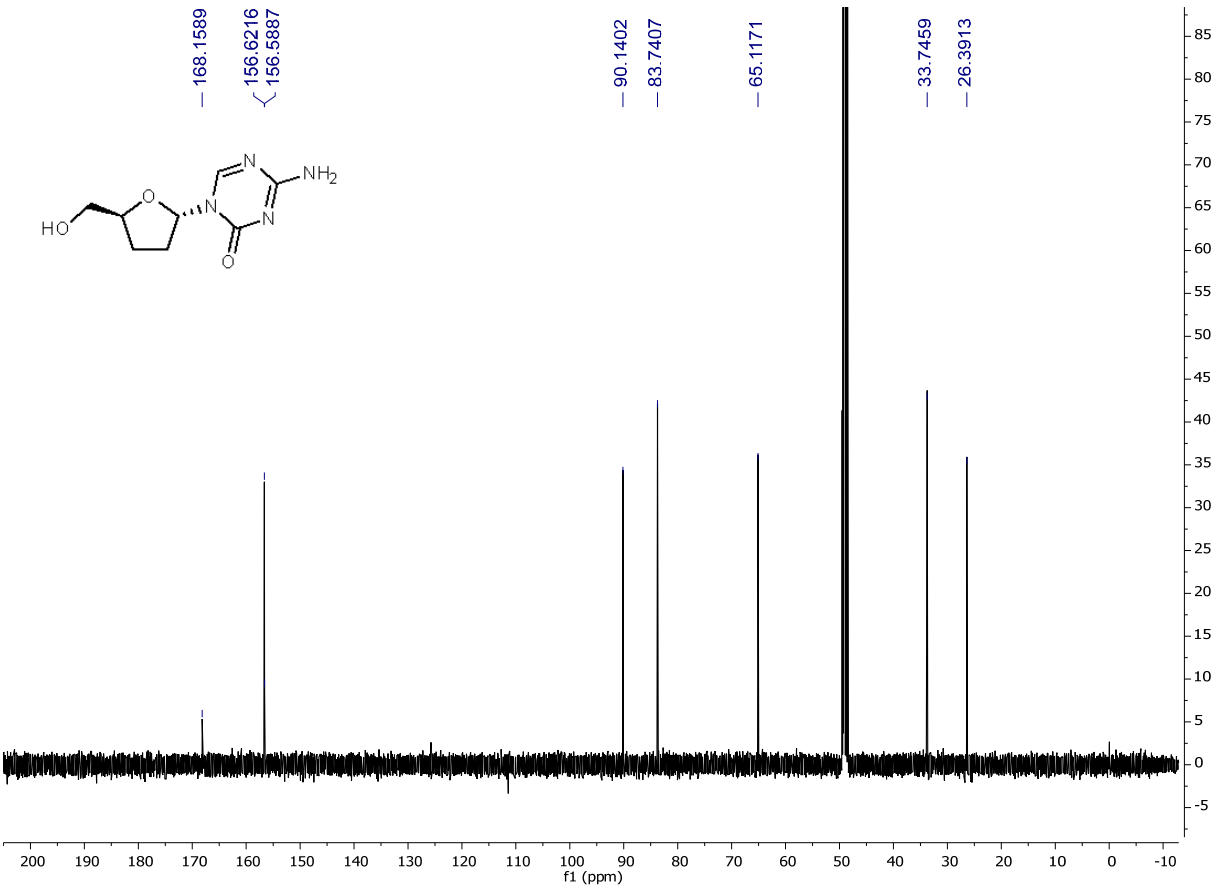


Figure S115. ^{13}C spectrum of **α-ddAC** (125 MHz, CD_3OD).

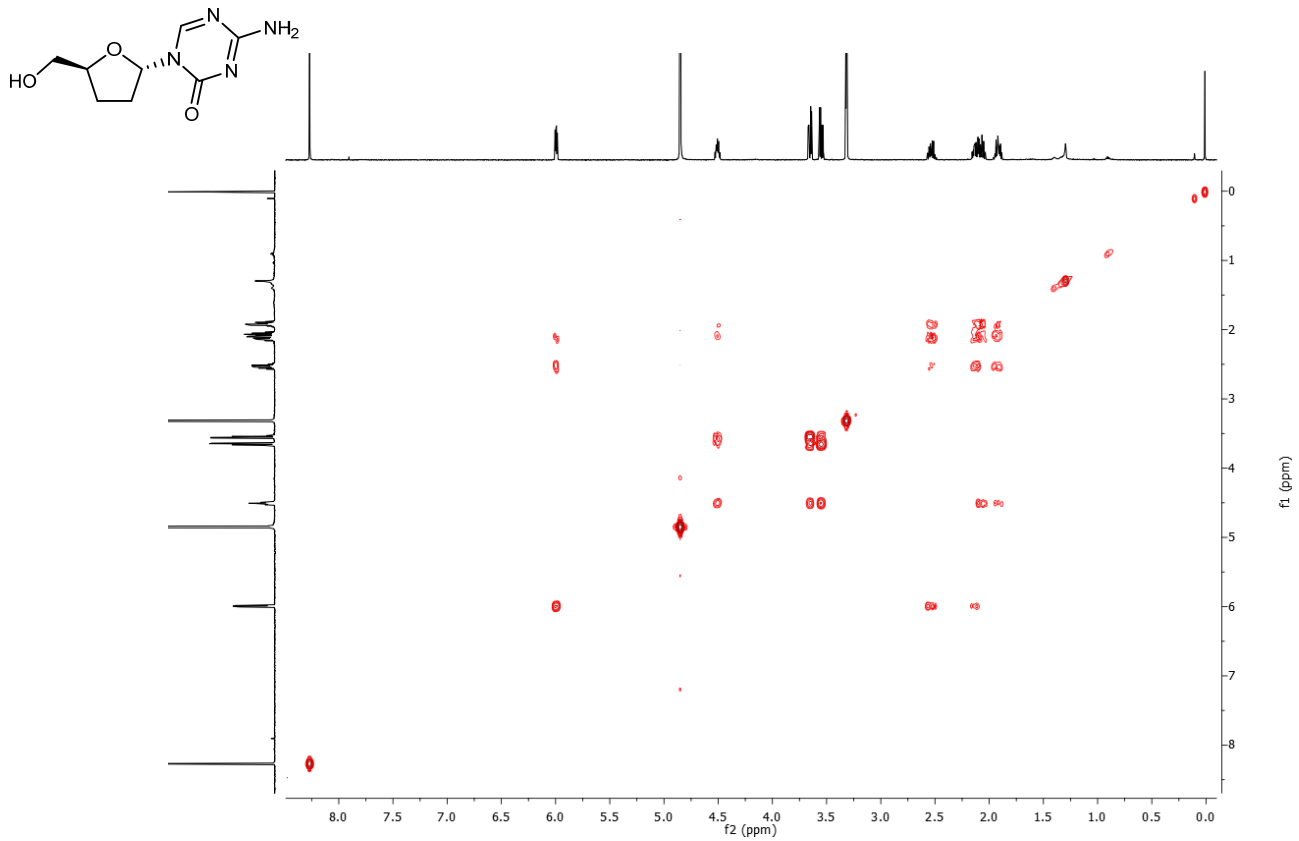


Figure S116. COSY spectrum of **α-ddAC**.

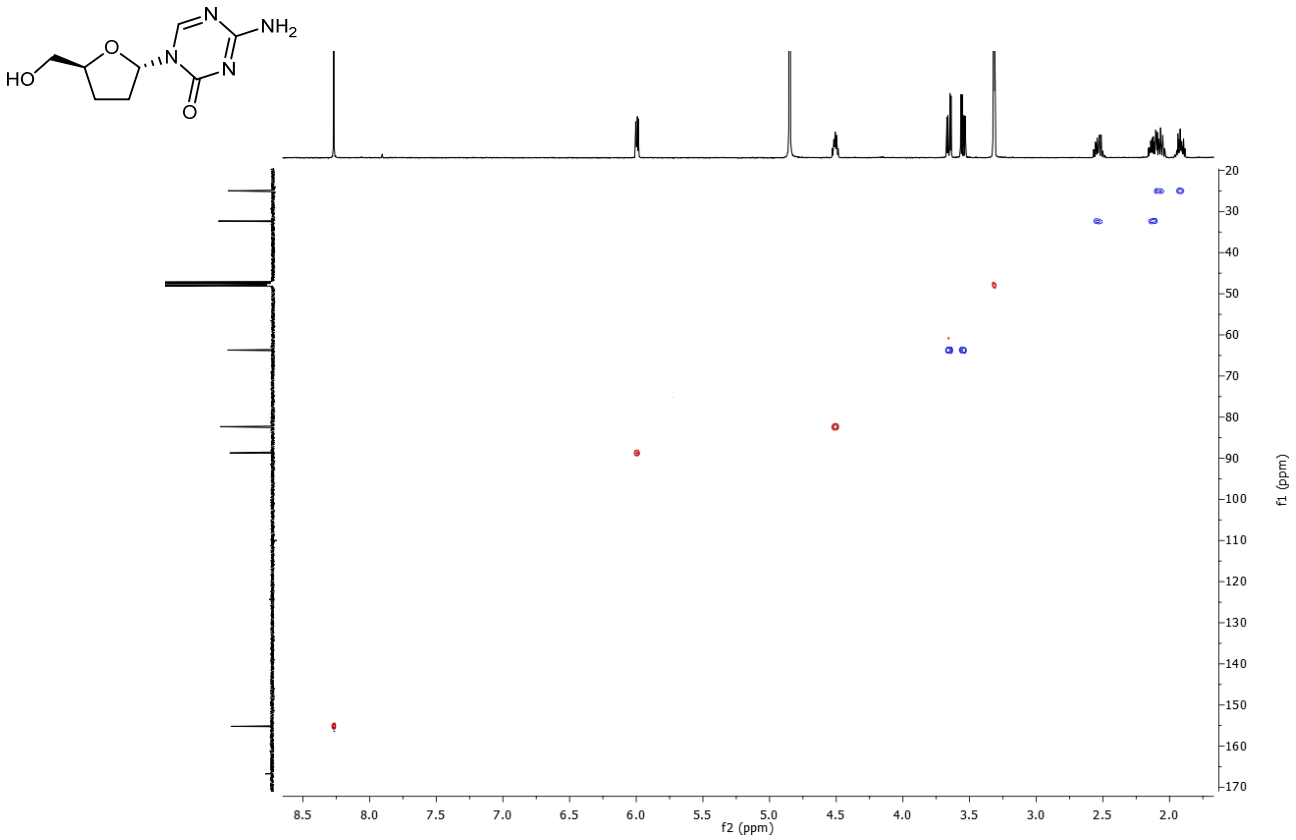


Figure S117. HSQC spectrum of **α-ddAC**.

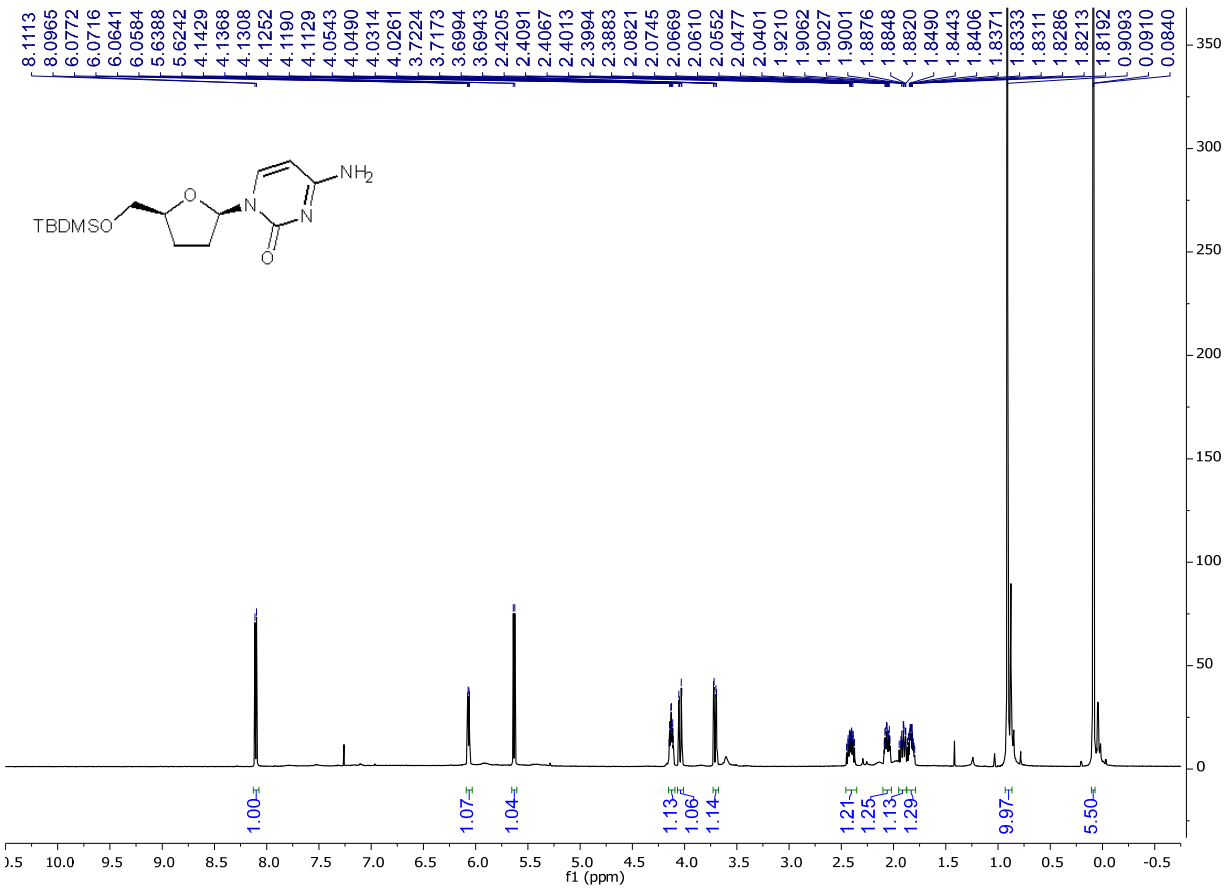


Figure S118. ^{1}H spectrum of **5'-*O*-silyl-protected β-ddC** (500 MHz, $CDCl_3$).

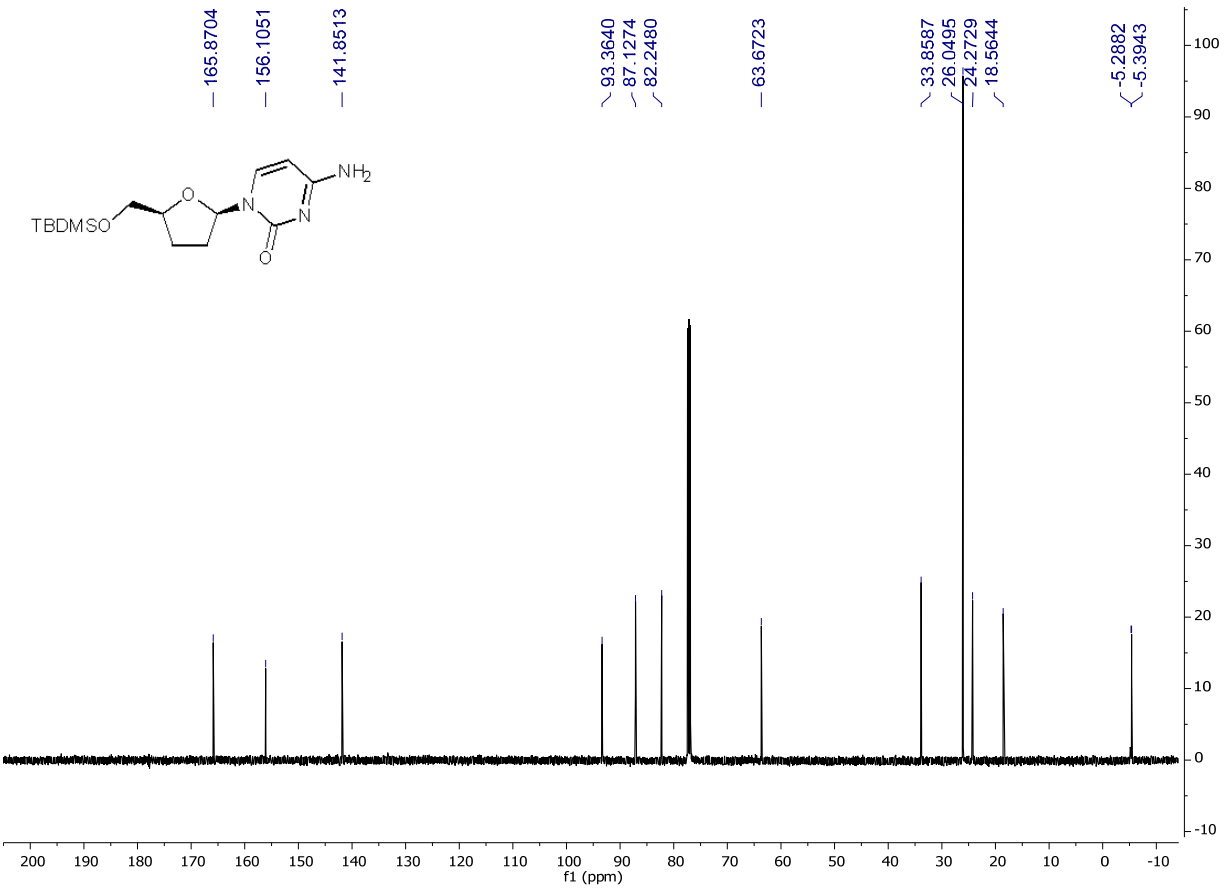


Figure S119. ^{13}C spectrum of **5'-*O*-silyl-protected β-ddC** (125 MHz, $CDCl_3$).

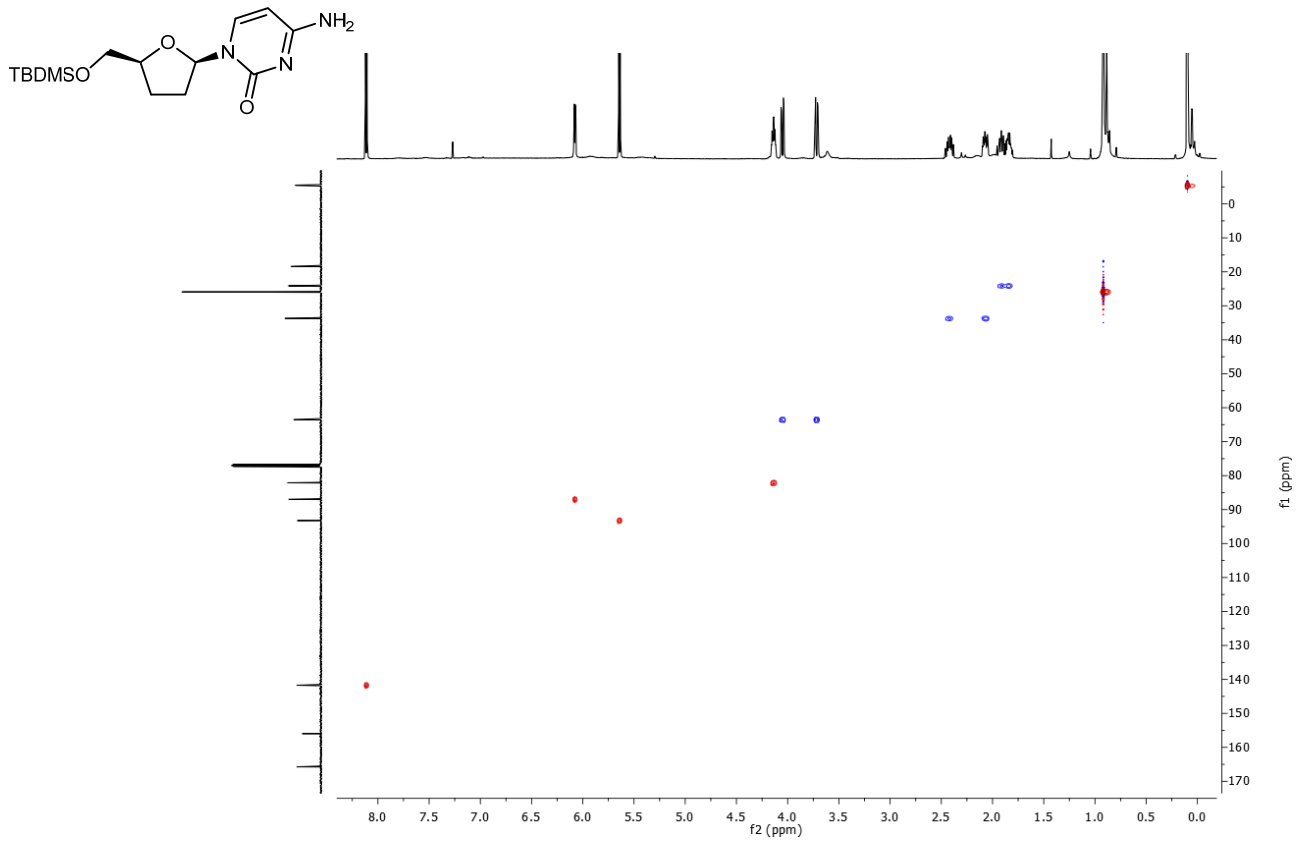


Figure S120. HSQC spectrum of **5'-*O*-silyl-protected β-ddC**.

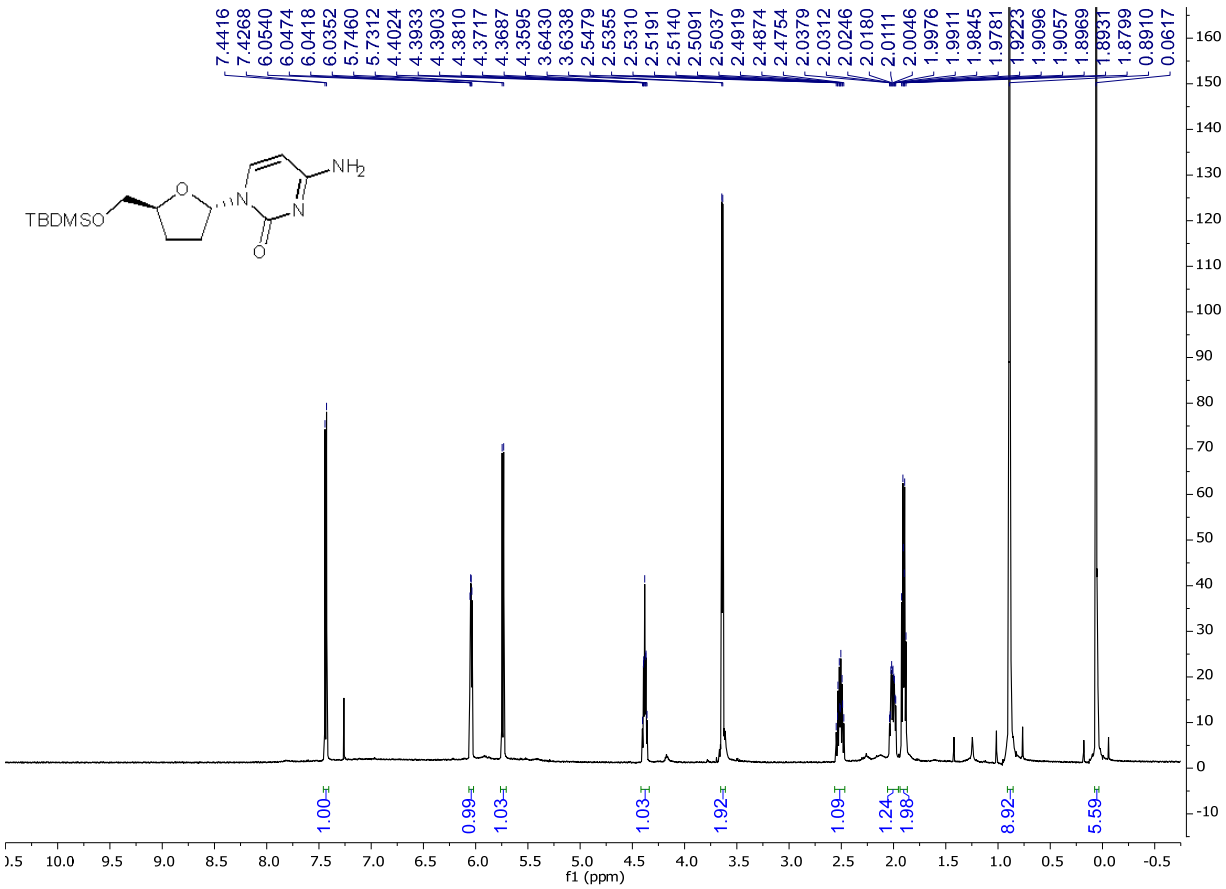


Figure S121. ^{1}H spectrum of **5'-*O*-silyl-protected α-ddC** (500 MHz, $CDCl_3$).

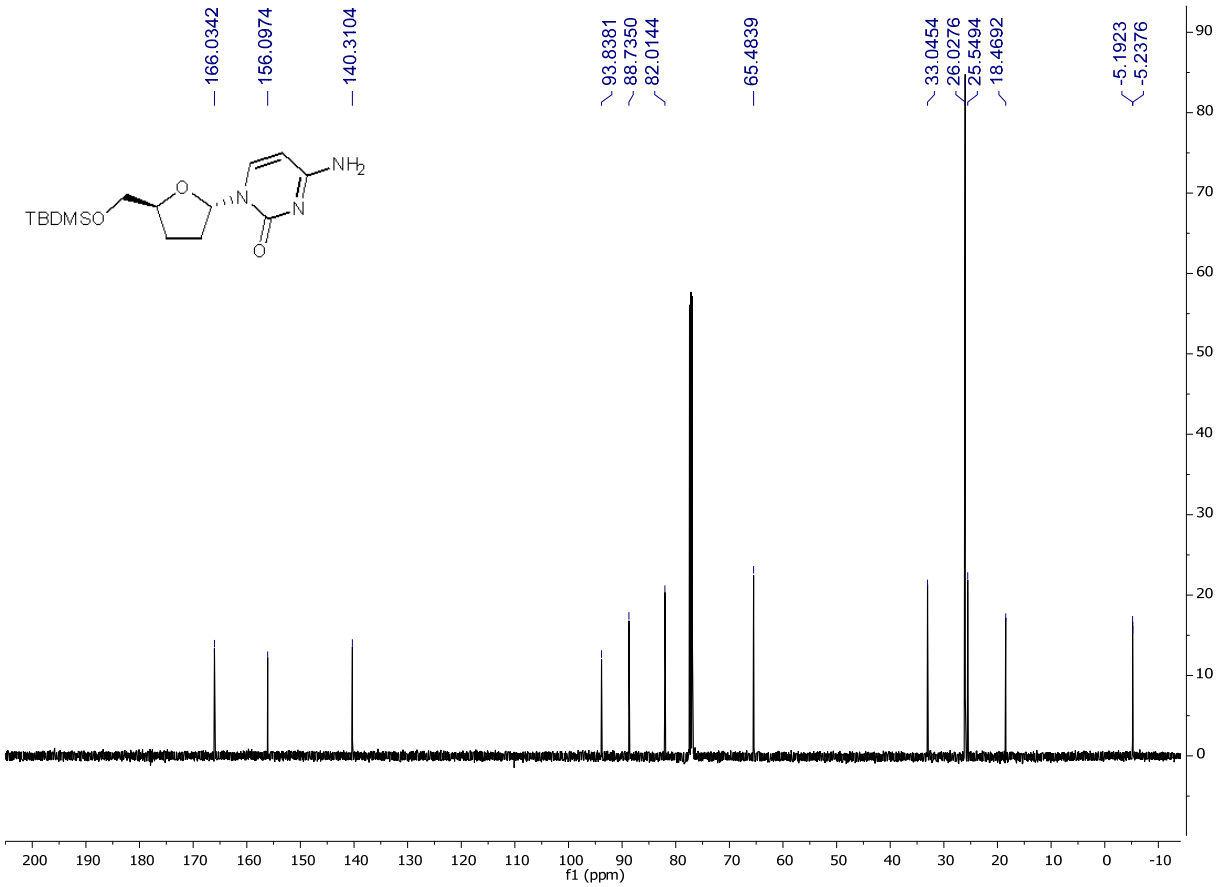


Figure S122. ^{13}C spectrum of **5'-*O*-silyl-protected α-ddC** (125 MHz, $CDCl_3$).

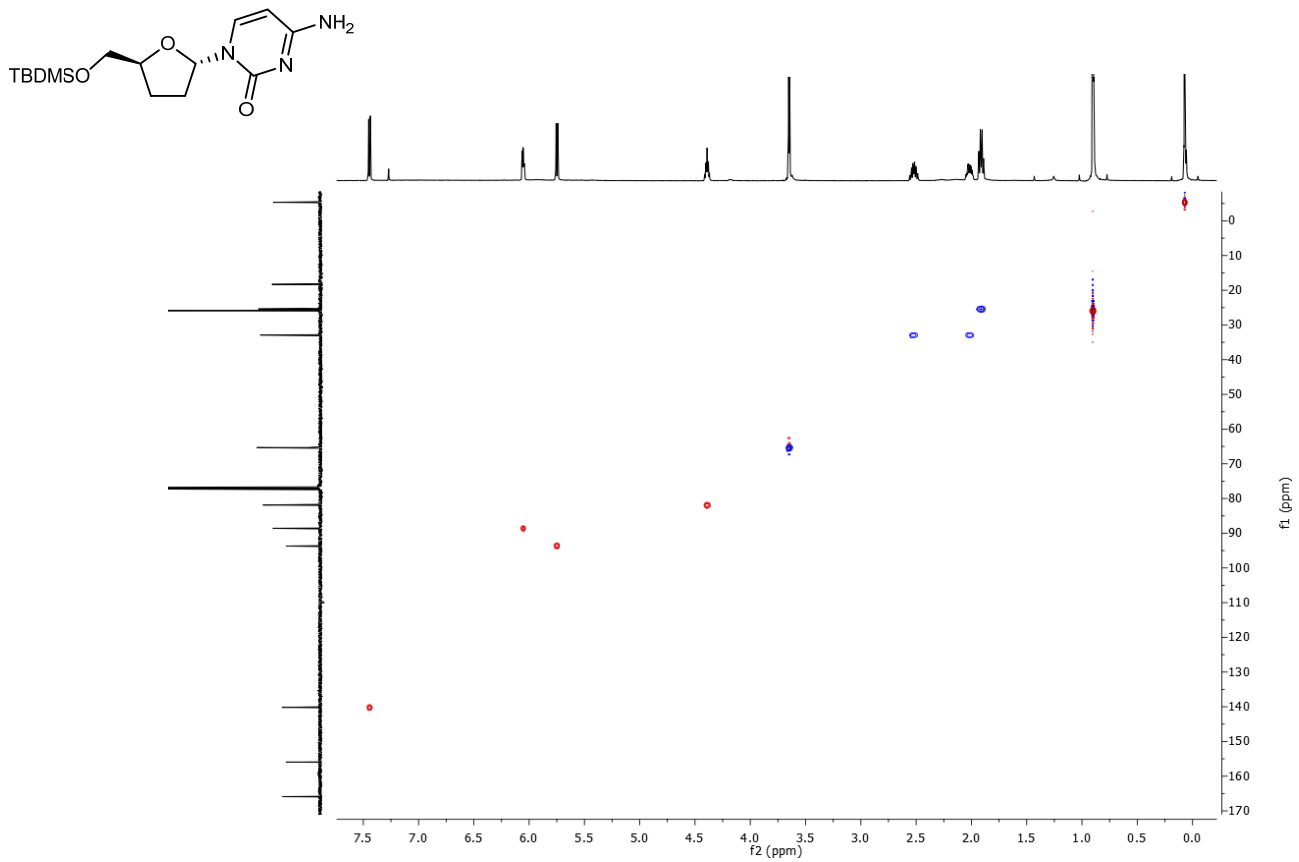


Figure S123. HSQC spectrum of **5'-*O*-silyl-protected α-ddC**.

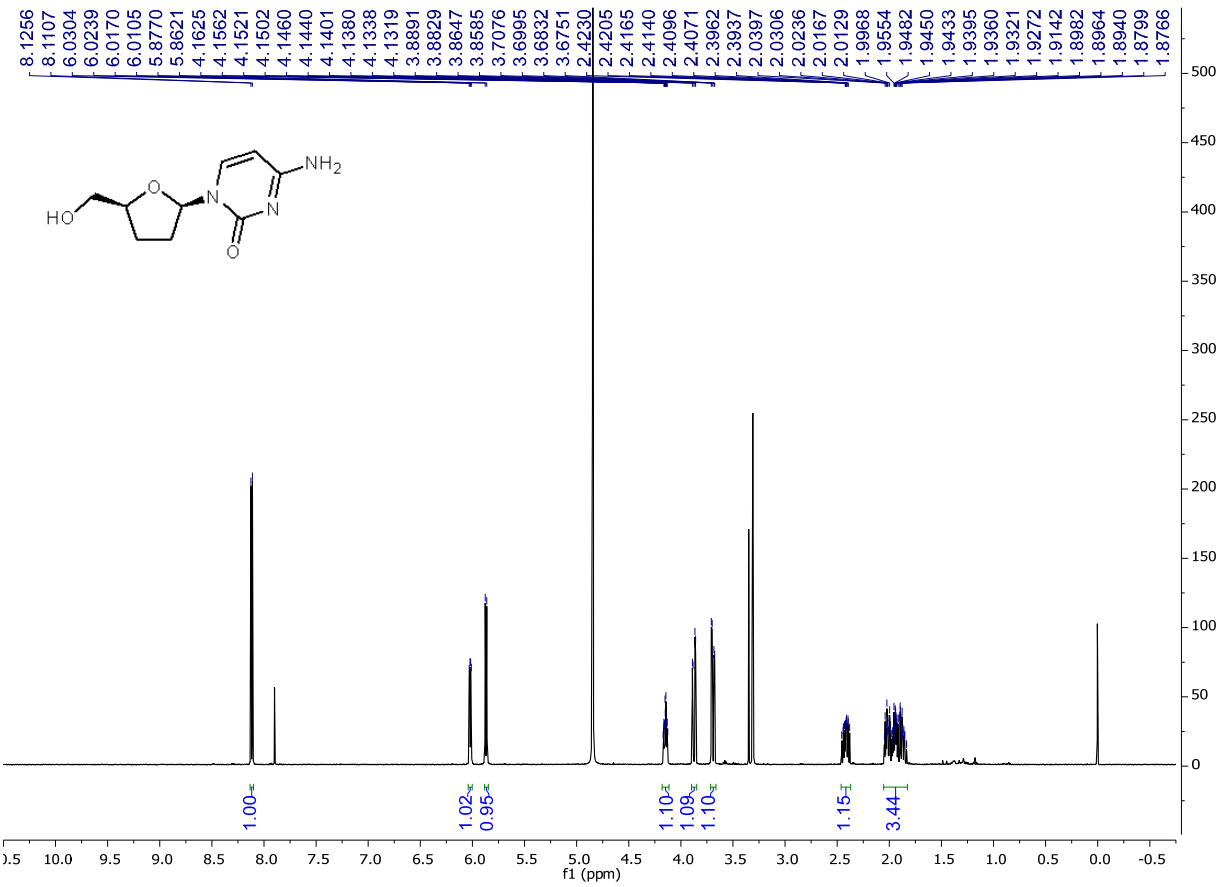


Figure S124. ^{1}H spectrum of **β-ddC** (500 MHz, CD_3OD).

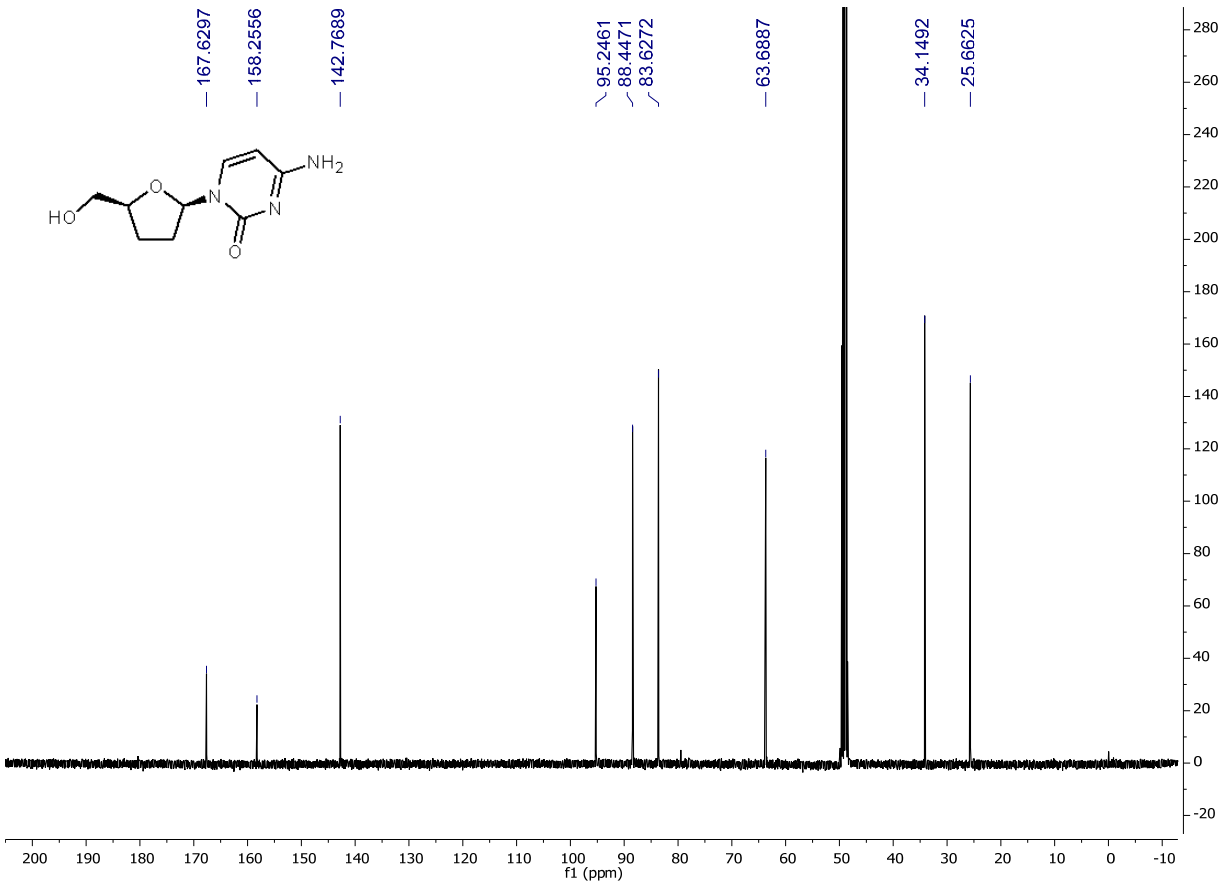


Figure S125. ^{13}C spectrum of **β-ddC** (125 MHz, CD_3OD).

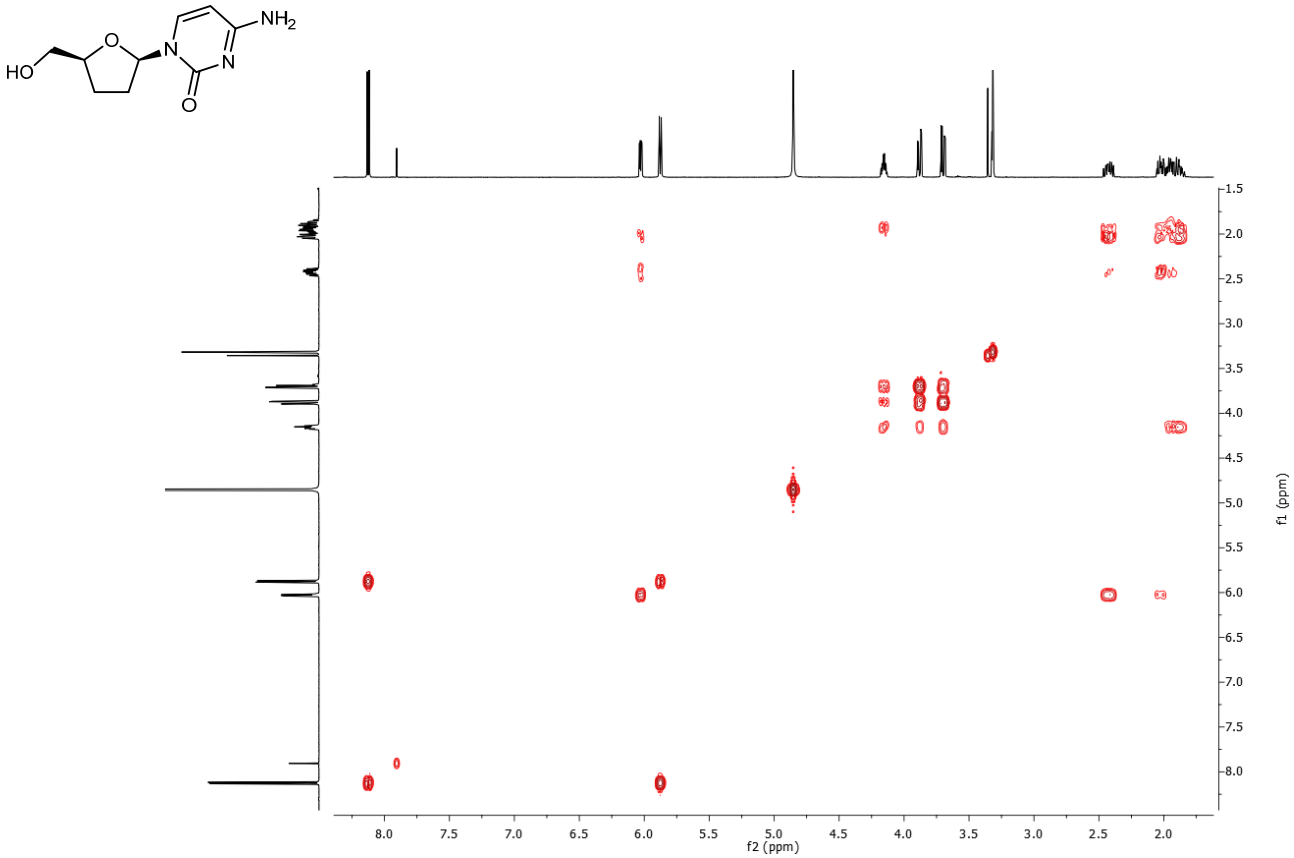


Figure S126. COSY spectrum of **β-ddC**.

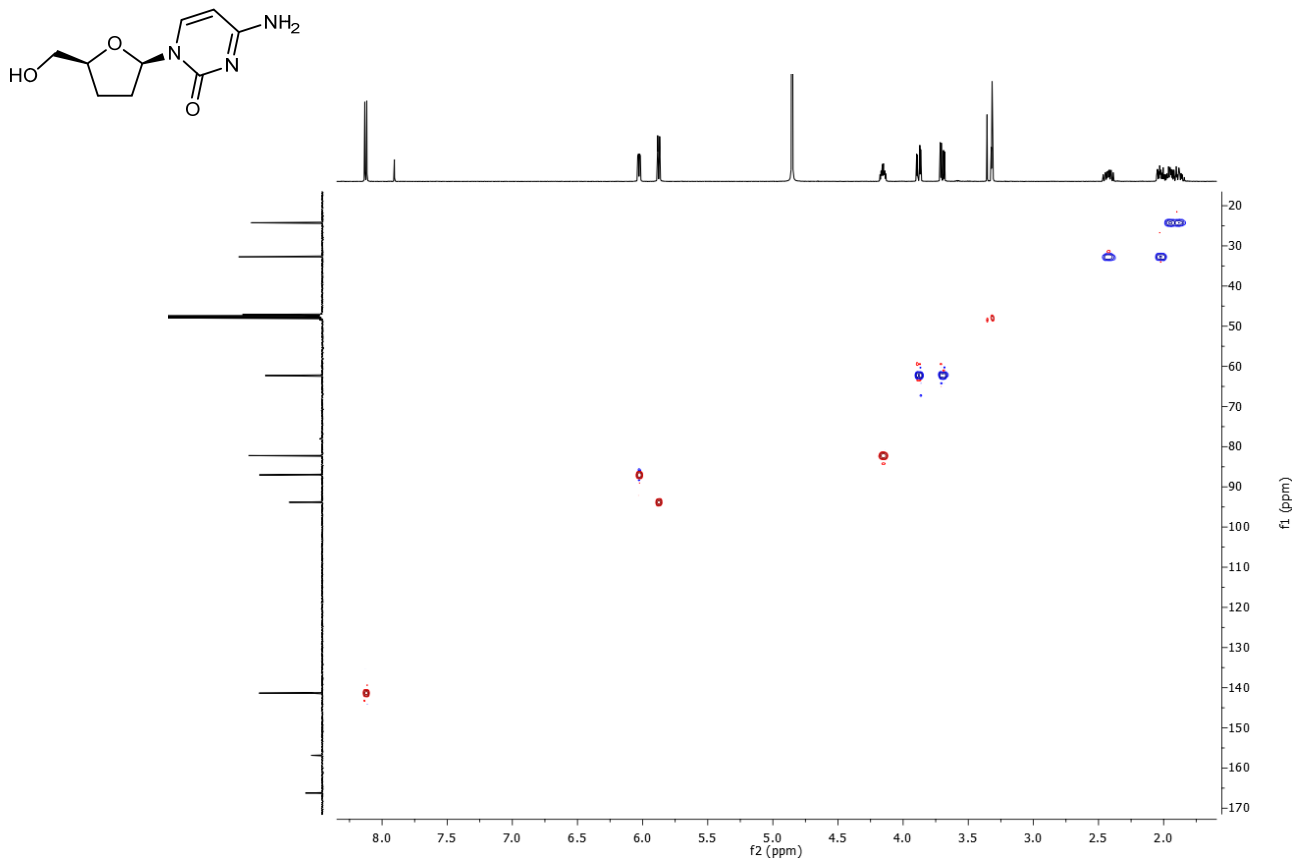


Figure S127. HSQC spectrum of **β-ddC**.

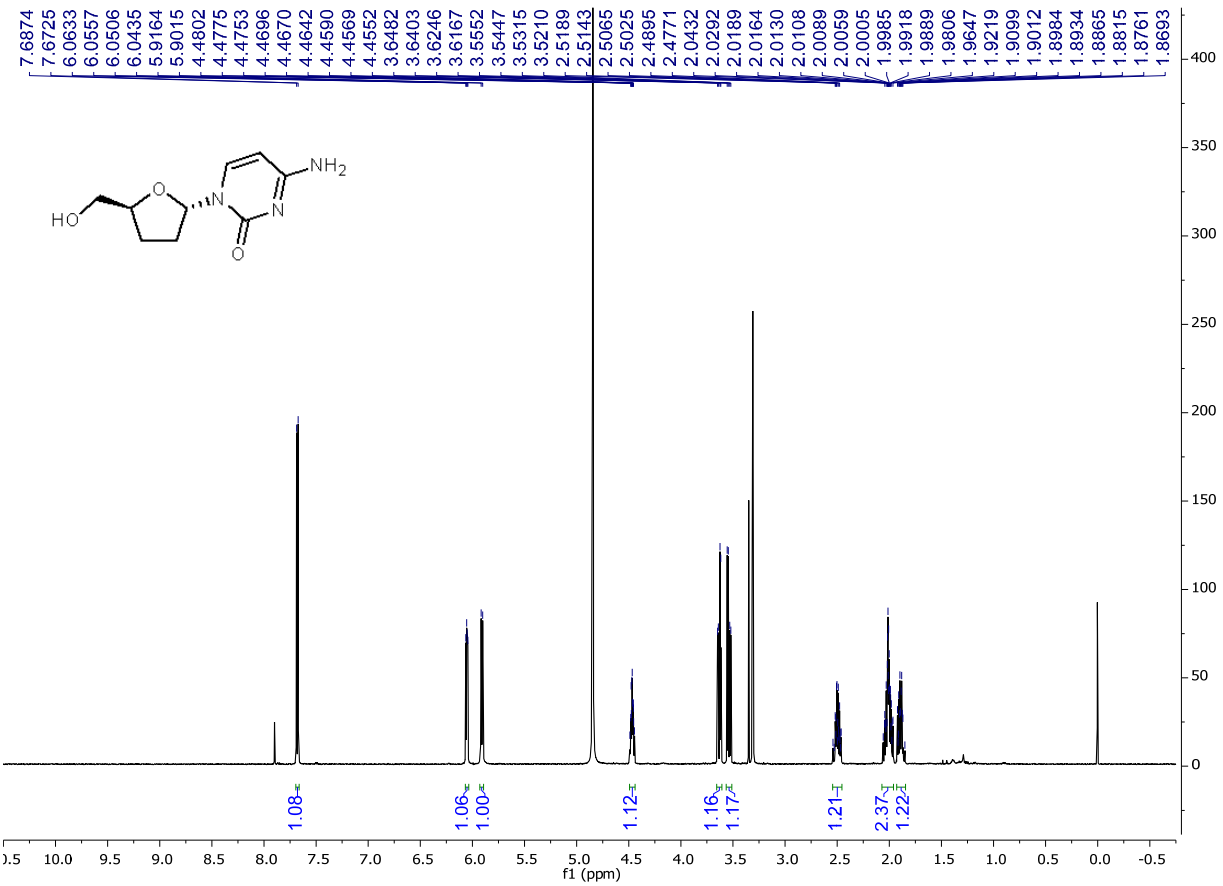


Figure S128. ^{1}H spectrum of **α-ddC** (500 MHz, CD_3OD).

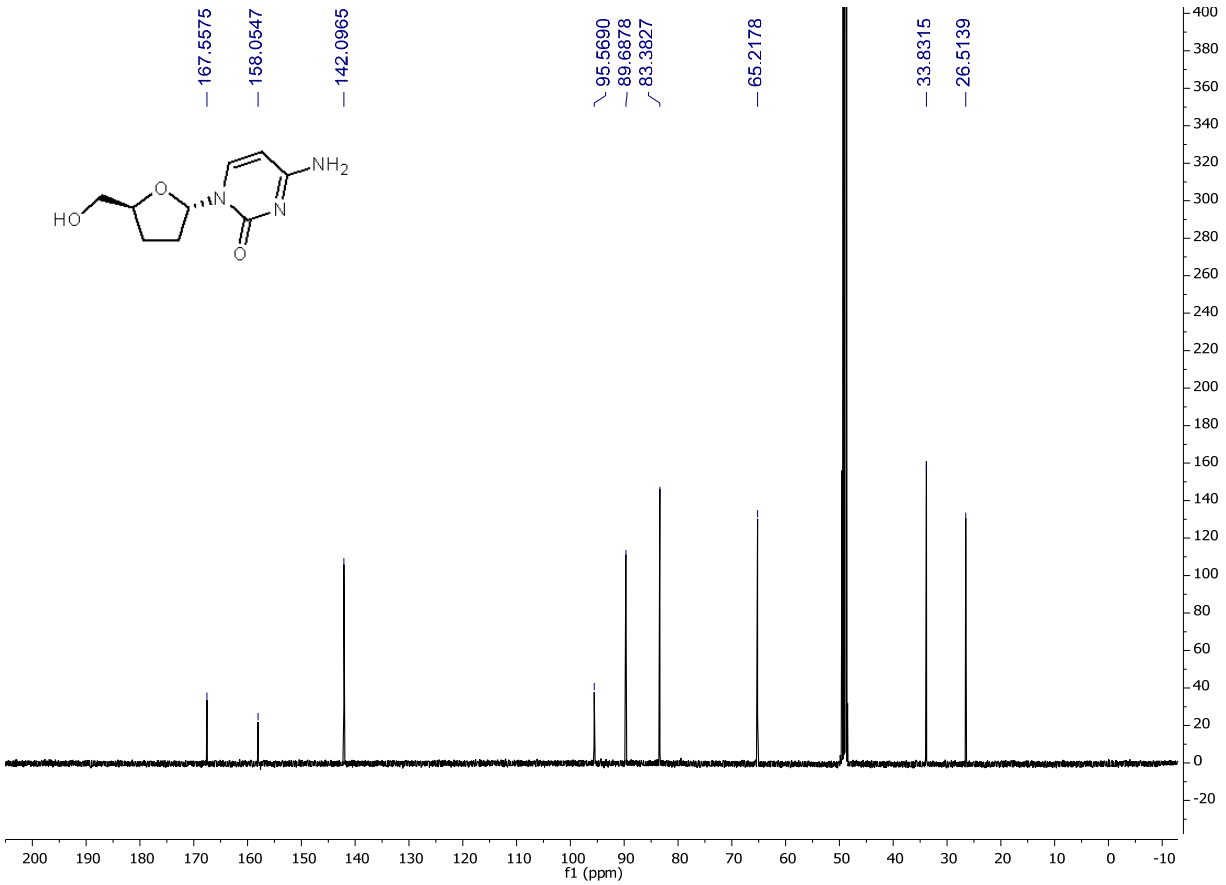


Figure S129. ^{13}C spectrum of **α-ddC** (125 MHz, CD_3OD).

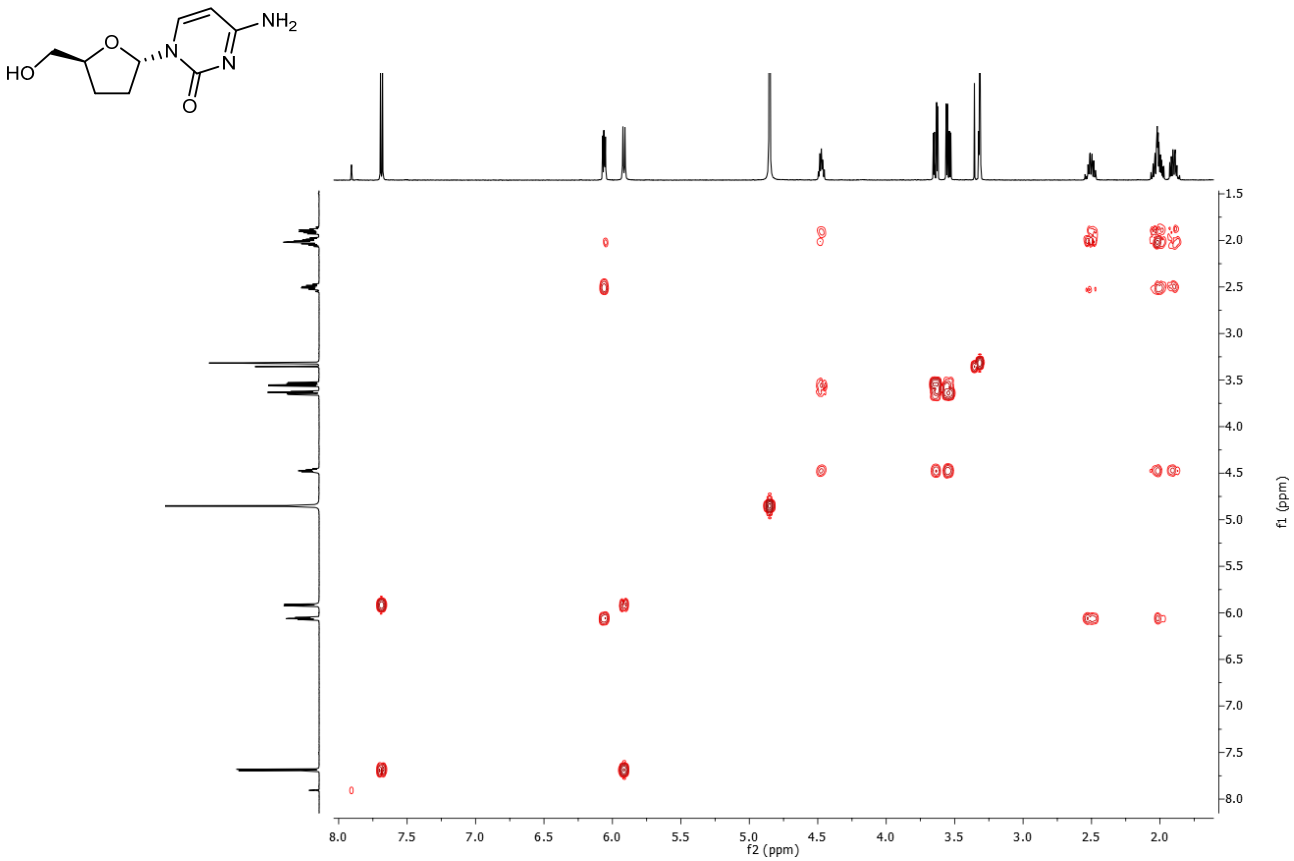


Figure S130. COSY spectrum of **α-ddC**.

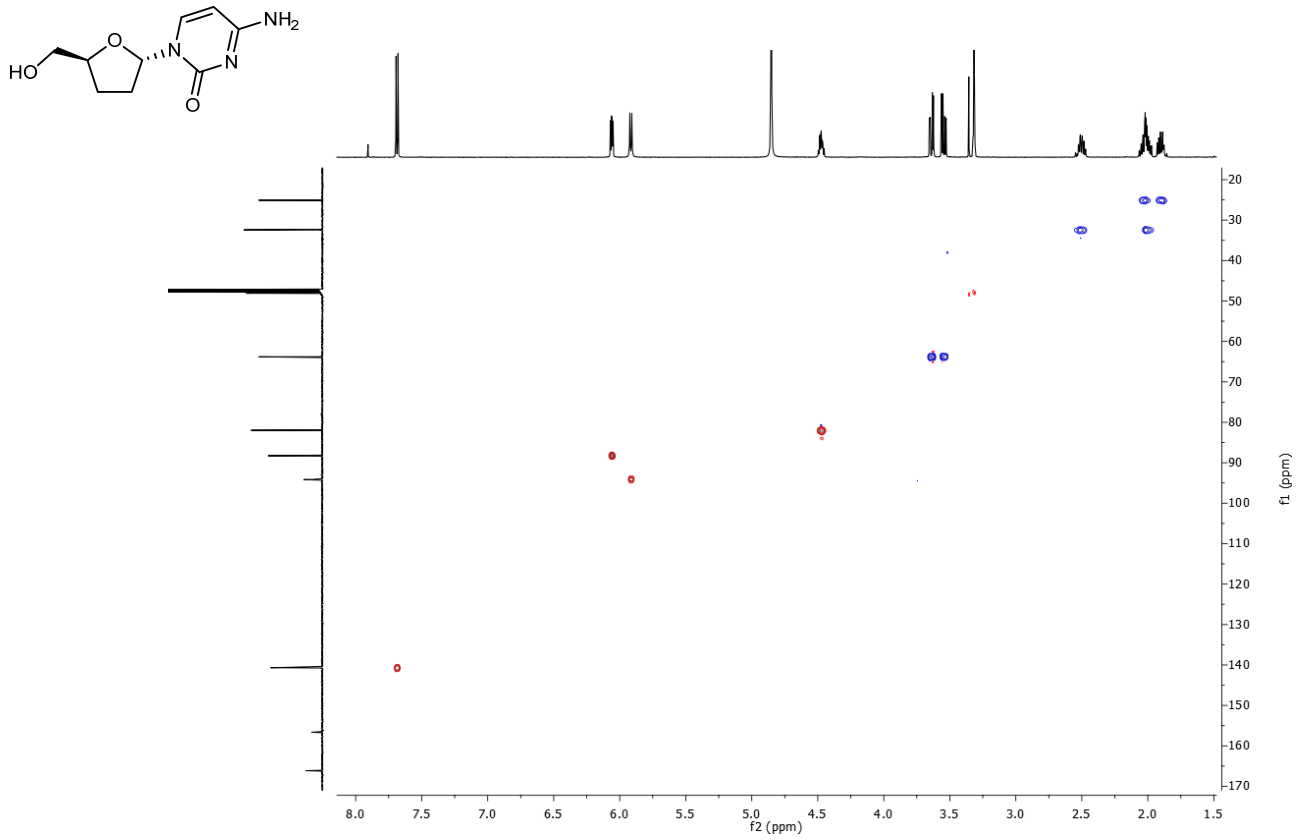


Figure S131. HSQC spectrum of **α-ddC**.

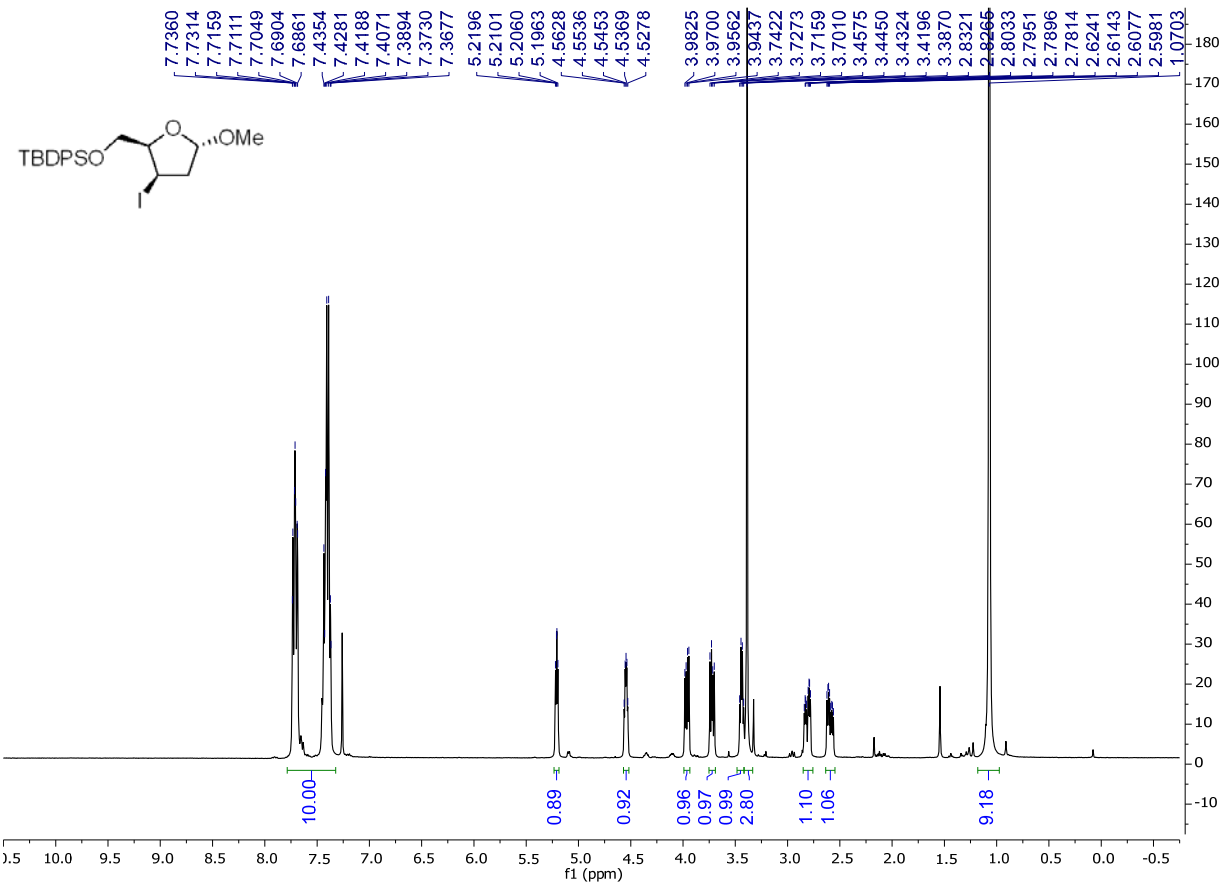


Figure S132. ^{1}H spectrum of **59** (400 MHz, $CDCl_3$).

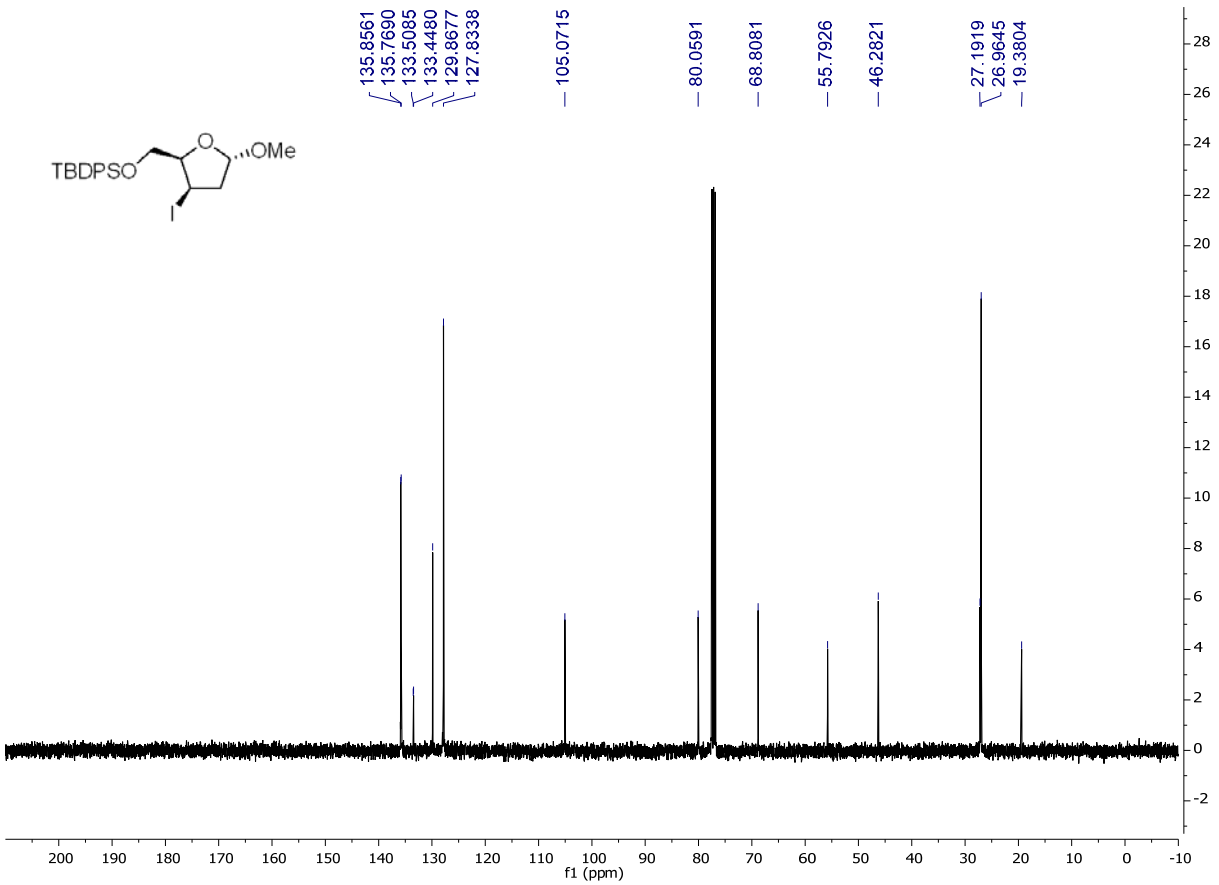


Figure S133. ^{13}C spectrum of **59** (100 MHz, $CDCl_3$).

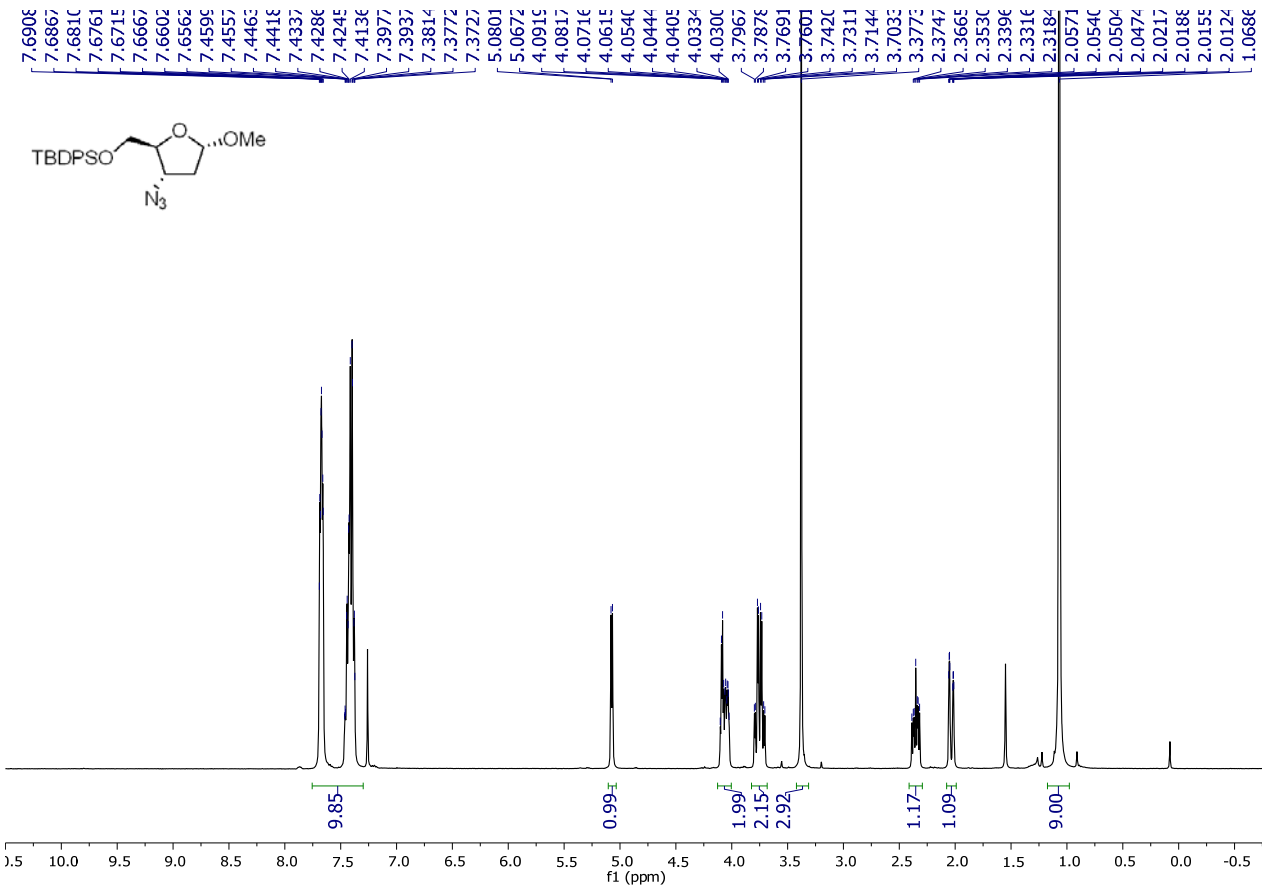


Figure S134. ^{1}H spectrum of **60** (400 MHz, $CDCl_3$).

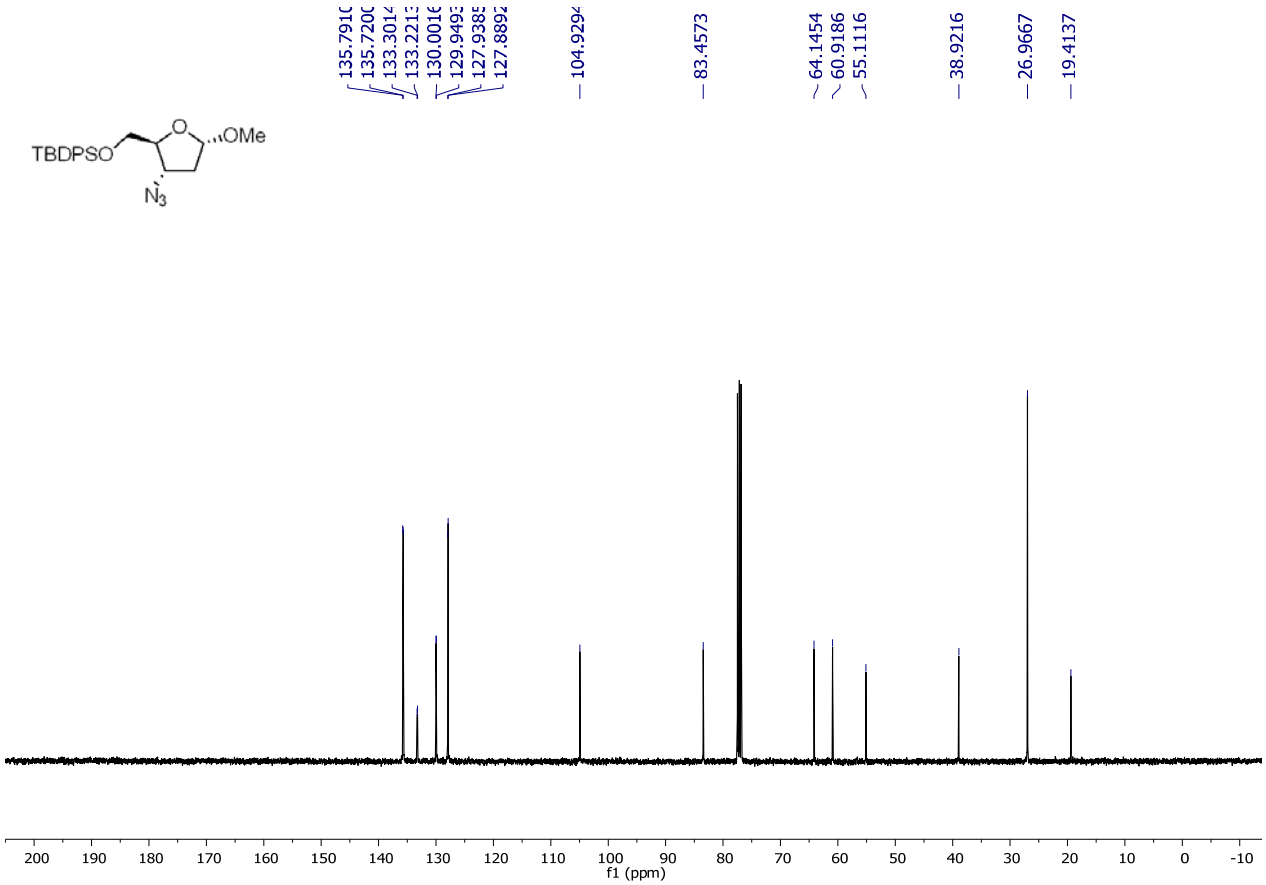


Figure S135. ^{13}C spectrum of **60** (400 MHz, $CDCl_3$).

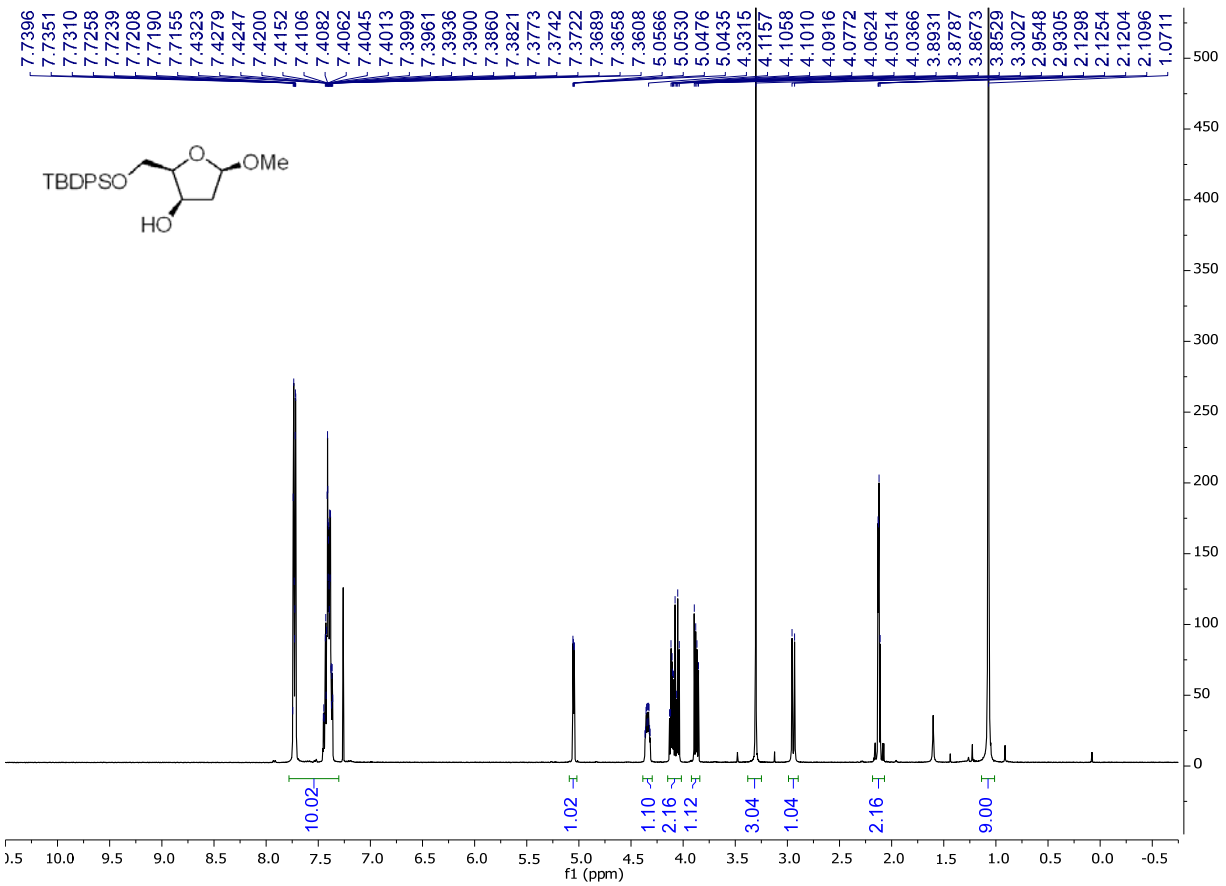


Figure S136. ^{1}H spectrum of **62** (400 MHz, $CDCl_3$).

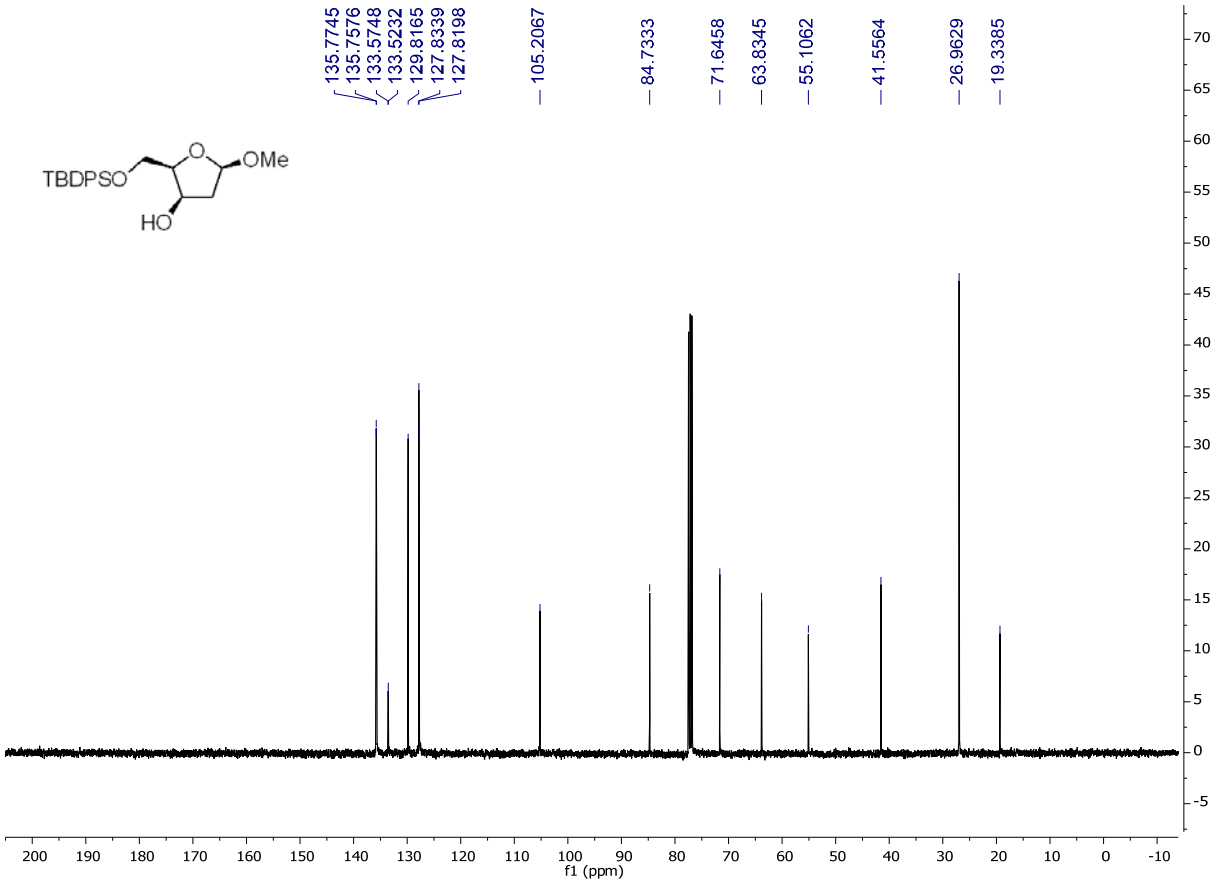


Figure S137. ^{13}C spectrum of **62** (400 MHz, $CDCl_3$).

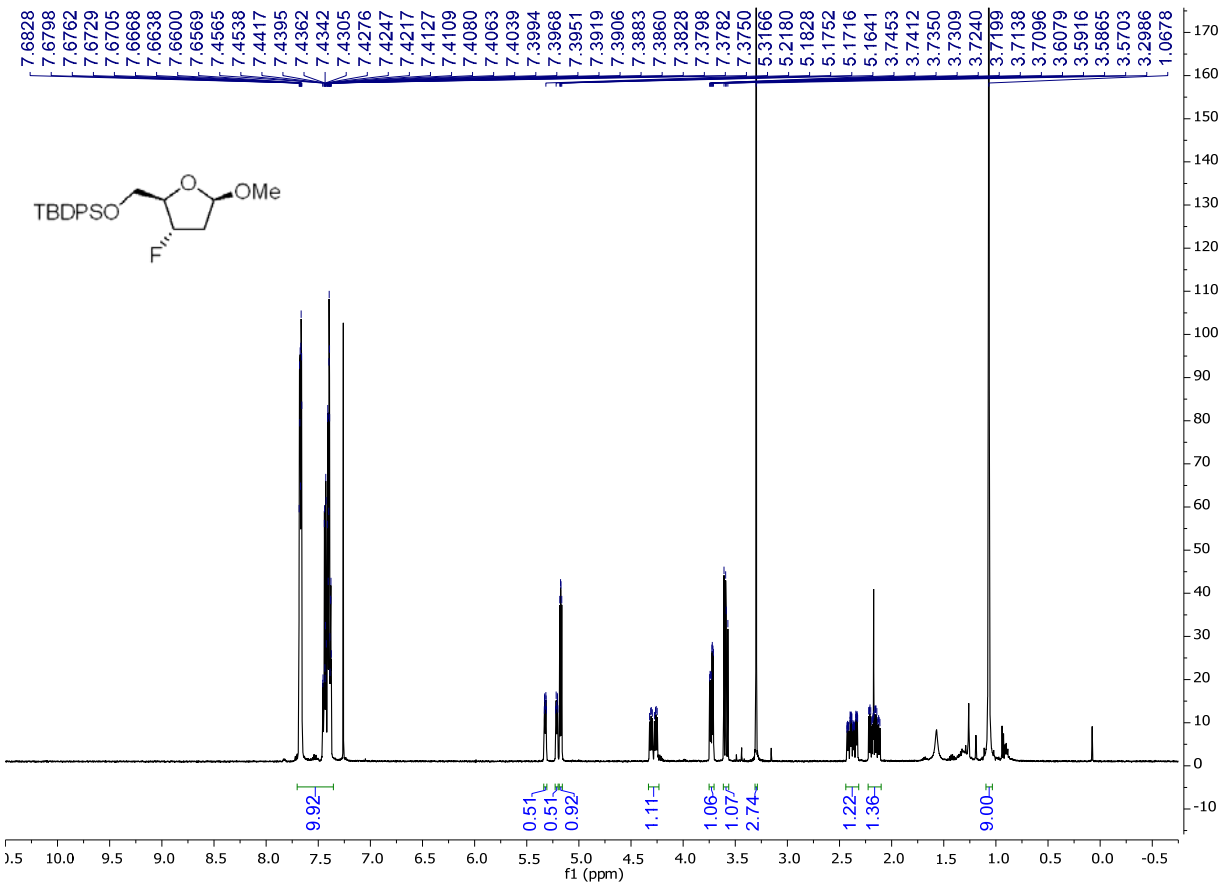


Figure S138. ^{1}H spectrum of **63** (500 MHz, $CDCl_3$).

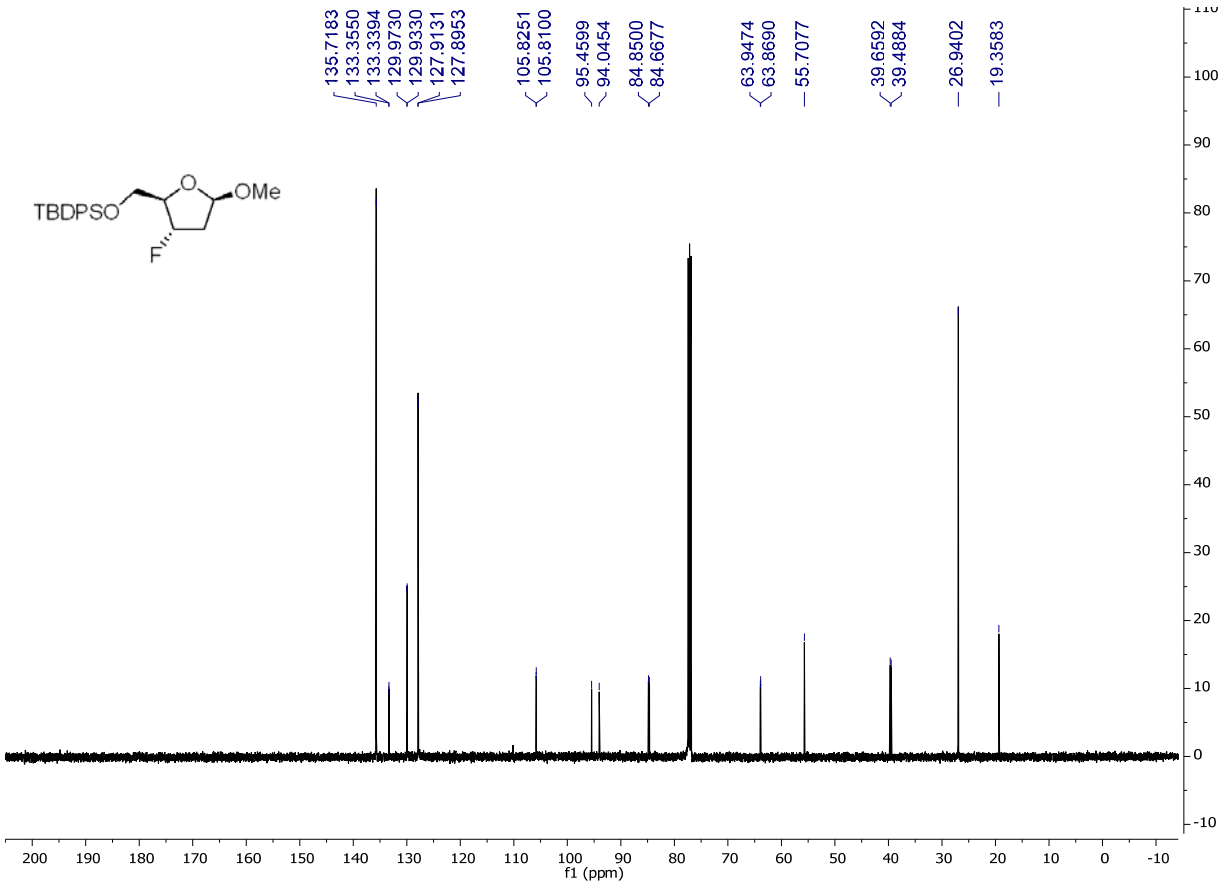


Figure S139. ^{13}C spectrum of **63** (125 MHz, $CDCl_3$).