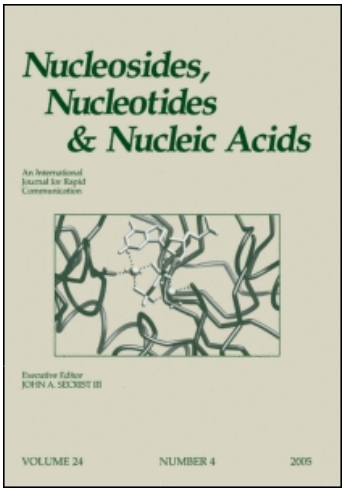


Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

A proposal for a convenient notation for *P*-chiral nucleotide analogues. Part 2. Dinucleoside monophosphate analogues

Michal Sobkowski[a]; Jacek Stawinski[ab]; Adam Kraszewski[a]
[a] Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland [b] Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

Online publication date: 22 December 2010

Nucleosides, Nucleotides, and Nucleic Acids, 25:1363–1375, 2006

ISSN: 1525-7770 print / 1532-2335 online
DOI: 10.1080/15257770600918862

Taylor & Francis
Taylor & Francis Group

A PROPOSAL FOR A CONVENIENT NOTATION FOR *P*-CHIRAL NUCLEOTIDE ANALOGUES. PART 2. DINUCLEOSIDE MONOPHOSPHATE ANALOGUES

Michal Sobkowski □ *Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland*

Jacek Stawinski □ *Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland and Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden*

Adam Kraszewski □ *Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland*

□ *A configuration of ligands around a phosphorus atom in P-chiral dinucleoside monophosphate analogues can be described using D_P/L_P stereochemical notation, which allows immediate correlation between the notation of configuration and the actual spatial arrangement of the phosphorus ligands. The area of applications of this new stereochemical nomenclature covers dinucleoside units bridged by virtually any type of tri- and tetra-coordinated phosphorus moieties, that is, phosphorothioates, phosphoramidates, phosphoramidites, boranephosphates, methanephosphonates, H-phosphonates, and many others.*

Keywords Stereochemical notation; D_P/L_P; *P*-chiral; Nucleotide analogues

INTRODUCTION

Application of R_P/S_P descriptors for absolute configuration of *P*-chiral dinucleoside monophosphate analogues is a reliable and commonly used implementation of the CIP convention.[1,2] However, since this system is based entirely on formal priority rules, stereochemically related compounds

Received 15 August 2005; accepted 12 June 2006.

The authors would like to express their gratitude to many scientists worldwide for their helpful remarks and suggestions during preparation of this paper. The financial support from the State Committee for Scientific Research, Republic of Poland (Grant No. 4 T09A 100 23) and the Swedish Research Council is gratefully acknowledged.

Address correspondence to Michal Sobkowski, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland. E-mail: msob@ibch.poznan.pl or Jacek Stawinski, Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106-91 Stockholm, Sweden. E-mail: js@organ.su.se

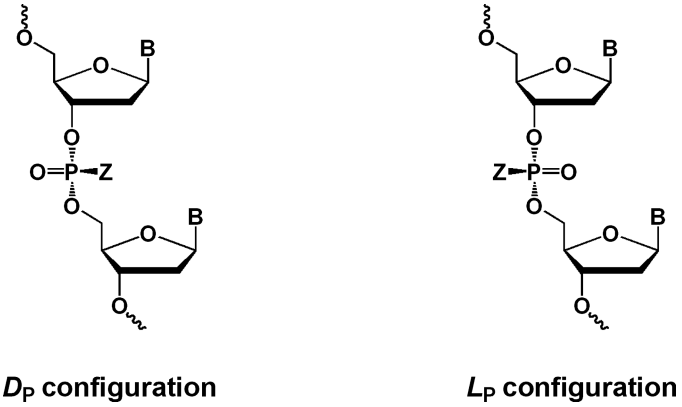


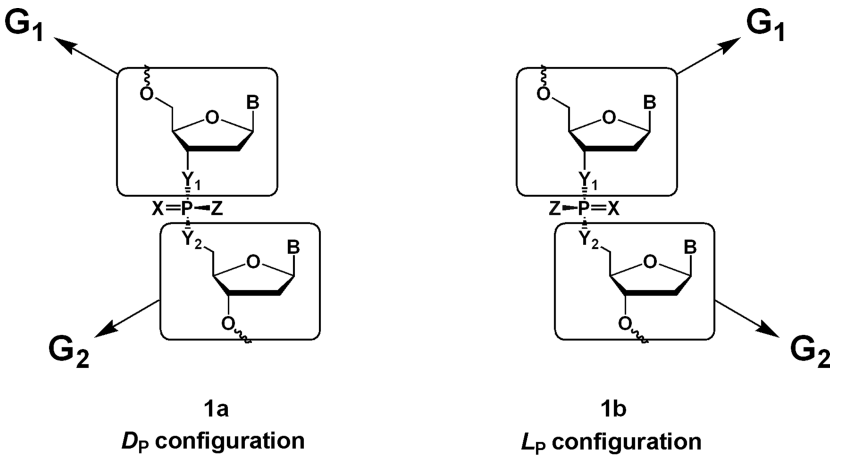
FIGURE 1 Representative structures for D_P and L_P configurations.

with the same spatial distribution of substituents often may have opposite configurations. This has been recognized as a longstanding problem in nucleic acids chemistry, but no alternative stereochemical notation useful for bioorganic phosphorus compounds has been proposed. Although in 1996 Lebedev and Wickstrom, while discussing stereochemistry at the phosphorus center in oligonucleotide analogues, used descriptive terms "pseudoaxial/pseudoequatorial,"[3] this was never developed into a comprehensive stereochemical notation.

To solve, at least partly, this problem recently we proposed a new system for describing stereochemistry of tetracoordinated phosphorus in nucleotide analogues using D_P/L_P descriptors,[4] which correlates configuration at the phosphorus center with spatial layout of the substituents. According to this convention, if a dinucleoside monophosphate fragment is drawn as in Figure 1 (a Fischer-like projection of the phosphorus center, with a nucleosid-3′-yl as an upper unit and a nucleosid-5′-yl as a lower one), species with a single-bonded ligand **Z** on the left side of the structure are defined as having L_P configuration at the phosphorus center, while those with a **Z** ligand on the right side, as having D_P configuration.

Such a convention offers clear advantages during analysis of structural aspects of nucleic acids analogues and mechanistic problems associated with chemical transformations during internucleotide bond formation.

The proposed D_P/L_P nomenclature is not intended as a replacement for the absolute R_P/S_P notation in the field of nucleic acids, but rather as its complementation. Although it works accurately for a wide range of *P*-chiral nucleotide analogues, its fundamentals are less rigid than those of the CIP convention. Thus, to avoid confusion or possible ambiguities, we suggest to use D_P/L_P descriptors together with R_P/S_P ones. For example, name D_P-(S_P)-methanephosphonate would stand for a S_P-methanephosphonate having D_P configuration, and D_P-(R_P)-phosphorothioate, for a R_P-phosphorothioate having D_P configuration.



Z: -H, -SR, -NR_2, -Me, -SiR_3, -BH_3^-, -I, -Cl, -F, -OR, *etc.*
Y_1, Y_2 : O, S, NR, CR_2, *etc.*
X : O, S, Se, Te, NR, CR_2, *etc.* or a free electron pair
G_1: nucleoside-3'-yl, nucleoside-2'-yl, alkyl, aryl, acyl, *etc.*
G_2: nucleoside-5'-yl, alkyl, aryl, acyl, *etc.*

FIGURE 2 Structures for definition of the extended D_P/L_P system.

DISCUSSION

The initial framework of D_P/L_P notation referred to dinucleoside monophosphate analogues with one modification (the **Z** group) at the phosphorus atom.[4] However, the system can be extended to nucleotide derivatives with one or 2 nucleoside units, and having atoms other than oxygen in bridging or nonbridging positions at the phosphorus center (Figure 2).

In general structures for compounds with D_P and L_P configurations depicted in Figure 2, the ligands at the phosphorus center are designated as $\mathbf{G_1}$, $\mathbf{G_2}$, **X**, and **Z**. Ligands $\mathbf{G_1}$ and $\mathbf{G_2}$ have fixed positions as the upper and lower units, respectively, while ligands **X** and **Z** can take either left- and right-side positions, depending on absolute configuration at the phosphorus center. For DNA or RNA analogues, $\mathbf{G_1}$ stands for a nucleosid-3′-yl (or a nucleosid-2′-yl) unit while $\mathbf{G_2}$ group for a nucleosid-5′-yl, and $\mathbf{Y_1}$ and $\mathbf{Y_2}$ are usually integral parts of these moieties. Ligands forming a double bond to the phosphorus (e.g., oxygen or sulfur in the P=O and P=S groups, respectively) are always designated as **X** while single-bonded ligands, as **Z**.

To ensure an unequivocal assignment of D_P/L_P configurations of various nucleotide analogues, we have defined a set of simple rules:

Rule 1: For dinucleoside phosphates presented in a form similar to the Fischer projection as shown in Figure 2, that is, with a nucleosid-3′-yl standing for $\mathbf{G_1}$ and a nucleosid-5′-yl standing for $\mathbf{G_2}$ moiety, D_P configuration is defined as the one having a single-bonded ligand **Z** to the right, and the P=X group to left. For compounds with L_P configuration, **Z** is to the left, and the P=X group to the right. **X** can be any atom double-bonded to the phosphorus or a free electron pair. $\mathbf{Y_1}$ and $\mathbf{Y_2}$ are atoms, or group of atoms, that are integral parts of $\mathbf{G_1}$ and $\mathbf{G_2}$ units.

Rule 2: For nucleotide analogues having carbohydrate residues other than ribo- or deoxyribofuranose, the assignment of ligands as $\mathbf{G_1}$ and $\mathbf{G_2}$ is done primarily with respect to their resemblance to the natural nucleosidic residues. If such analogy is not obvious, the following rules should be used:

(a) A carbohydrate residue having a phosphorus center bound to a carbon atom of higher order is assigned as $\mathbf{G_1}$.

(b) If carbon atoms to which the phosphorus center is attached are of the same order in both carbohydrate residues, the residue of higher CIP priority is assigned as $\mathbf{G_1}$.

Rule 3: If both **X** and **Z** can form a double bond with phosphorus, the following order of priority should be used for the assignment of ligand **X** in the P=X bond:

(a) P=O > P=S > P=Se > P=Te > P=N.

(b) In other instances, the double bond should be set to an atom of lower CIP priority.

Configurations with **X** and **Z** assigned oppositely to that specified Rule 3a or 3b, should be referred to as pseudo-D_p or pseudo-L_p.

Below, the assignment of D_P/L_P configuration to dinucleoside monophosphate analogues is discussed in detail. Applications of D_P/L_P descriptors to nucleoside phosphate monoester analogues and to non-nucleosidic compounds are presented in the accompanying article in this issue.[5]

1. Dinucleoside Monophosphate Analogues Containing P=O Bond (Rule 1)

For analogues of dinucleoside monophosphates (or compounds bearing a dinucleoside monophosphate as a fragment of a more complex framework) with 2 nucleosidic ligands already designated as $\mathbf{G_1}$ and $\mathbf{G_2}$, D_P/L_P configuration depends on the positions of the **Z** and **X** groups only (Figure 3).

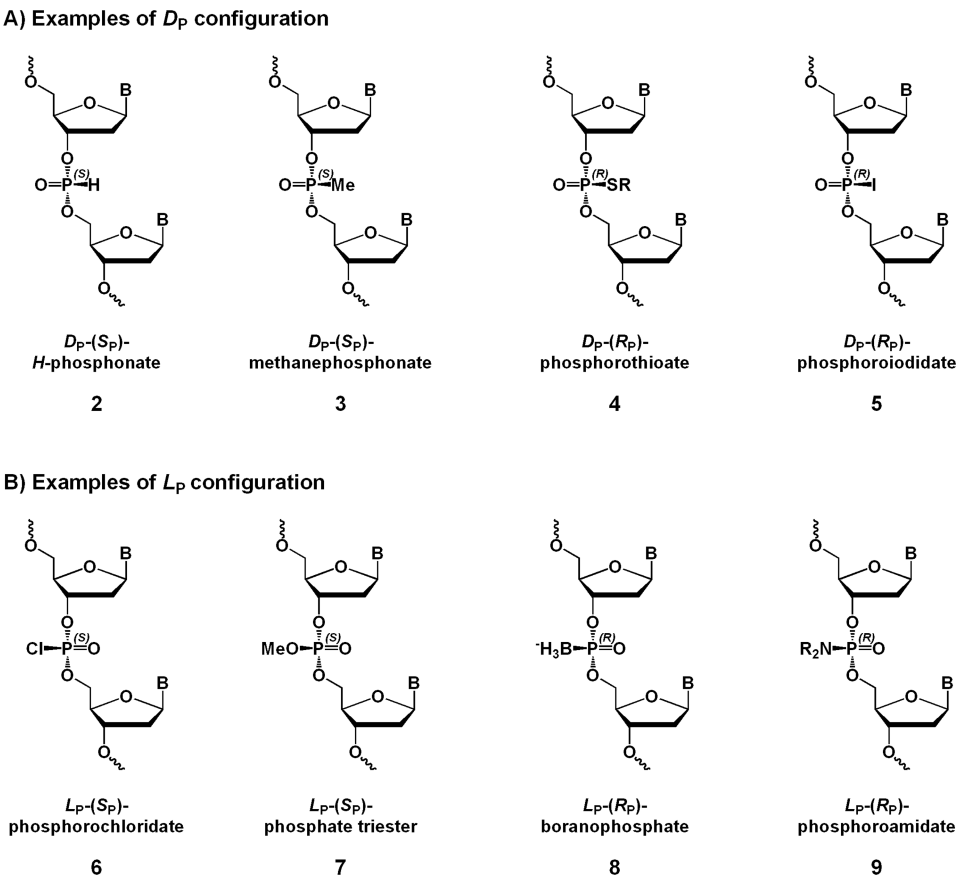


FIGURE 3 D_P/L_P notation for dinucleoside monophosphate analogues with a P=O bond.

For all compounds with D_P configuration, group **Z** is pointing to the right, and for compounds with L_P configuration, to the left. For this class of compounds, a correlation between D_P/L_P and R_P/S_P configurations can be found easily with a rule of thumb stating that configuration $D_P = R_P$ (and $L_P = S_P$) if atomic mass of the **Z** atom > 16, while configuration $D_P = S_P$ (and $L_P = R_P$), if atomic mass of the **Z** atom < 16.*

As discussed in Part 1 of this series,[4] the D_P/L_P notation facilitates analysis of enzymatic reactions of nucleotide and nucleic acids analogues as—in contrast to the R_P/S_P notation—this system emphasizes the importance of structural motifs for substrate recognition by enzymes. For example, D_P-(R_P)-dinucleoside phosphorothioates and D_P-(S_P)-methyl nucleoside-5′-yl phosphorothioates are both substrates for SVPD (D_P

*This rule of thumb can be extended according to the following scheme: m[**Z**] > m[**X**] $\Rightarrow D_P = R_P$ and $L_P = S_P$; m[**Z**] < m[**X**] $\Rightarrow D_P = S_P$ and $L_P = R_P$; where m[**Z**] = atomic mass of atom **Z**; m[**X**] = atomic mass of atom **X**. The rule may fail if **Y** atom is not an oxygen (e.g., **30** or **32**) or if $\mathbf{G_2}$ has higher CIP priority than $\mathbf{G_1}$ (e.g., **26**).

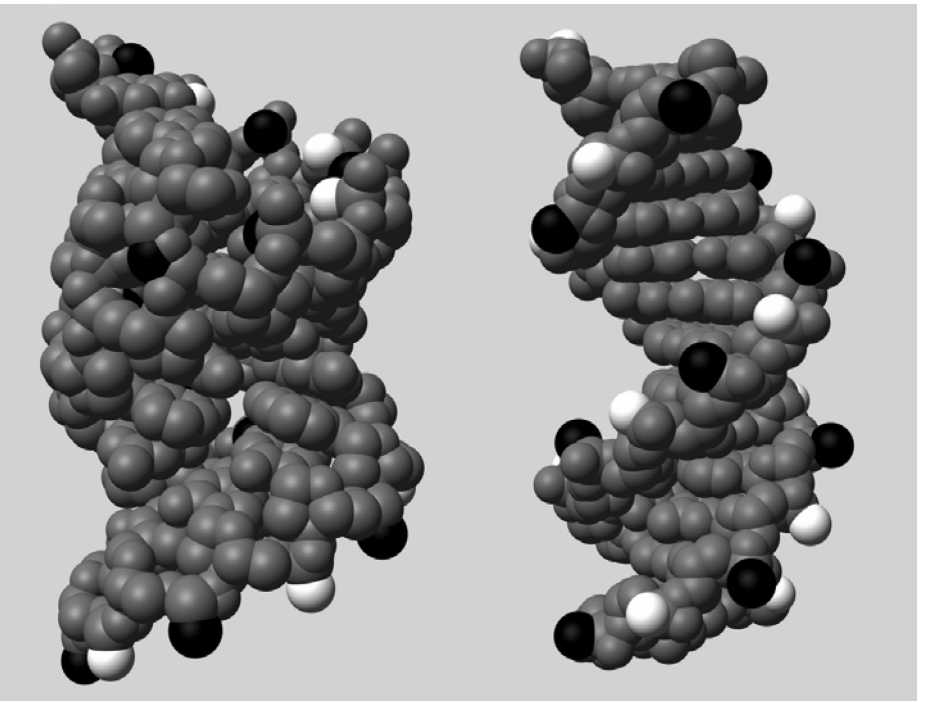
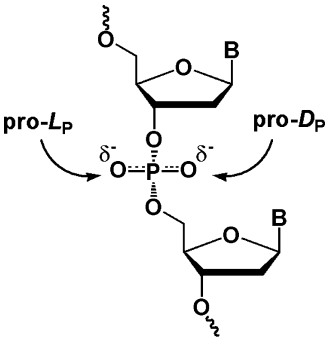


FIGURE 4 (a) Chimeric A-DNA and (b) B-DNA duplex forms containing *P*-chiral internucleotide linkages. R_P-phosphorothioate and S_P-phosphoramidate residues all have D_P configuration, while S_P-phosphorothioates and R_P-phosphoramidates have L_P configuration. For both duplexes, one strand contains D_P residues exclusively, while the complementary one, L_P residues only (• = sulfur, ○ = nitrogen). The DNA structures were generated by HyperChem 7.0 program and were converted into graphics images using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081). (http://www.cgl.ucsf.edu/chimera)

configurations) although they have opposite absolute configurations (R_P and S_P, respectively).

P-chiral internucleotide linkages can be denoted as D_P/L_P analogously as it is practiced for the R_P/S_P convention. However, in contrast to the absolute R_P/S_P configuration, the new notation unequivocally correlates configuration at the phosphorus center with a spatial position of the attached ligands in nucleic acid framework. For example, in the case of a double-stranded B-DNA form, a **Z** substituent of D_P diastereomers always is located in the major groove, while for L_P counterparts it points away from the duplex (Figure 4b). In an A-DNA form, **Z** is hidden deeply in the major groove for D_P diastereomers and located on its edge for analogues with L_P configuration (Figure 4a).

For description of nucleic acid-protein interactions, 2 nonbridging oxygens at the phosphorus center are often designated as pro-R_P/pro-S_P.[6–8] These descriptors, unfortunately, are rather imprecise, as they always require definition of an analogue to which they refer. The D_P/L_P convention



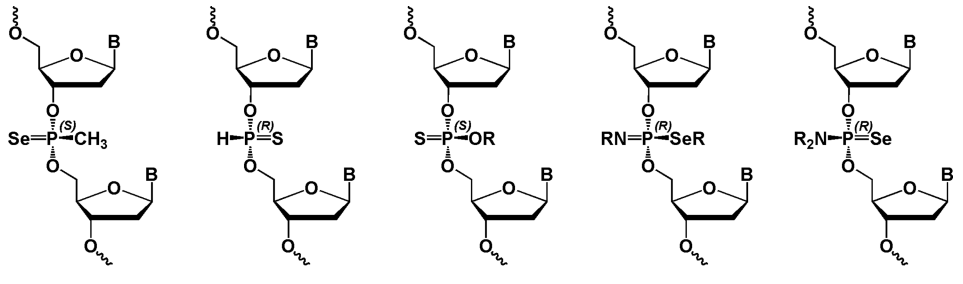
10

FIGURE 5 Pro-D_P/pro-L_P notation for prochiral internucleotidic phosphate.

alleviates this problem and gives the pro-D_P and pro-L_P descriptors an absolute meaning, namely, that replacement of pro-D_P oxygen by any heteroatom or carbon will result in an analogue having D_P configuration, and vice versa (Figure 5).

2. Dinucleoside Monophosphate Analogues Without a P=O Bond (Rule 1)

There are many phosphate analogues in which the phosphoryl group P=O has been replaced by a P=X functionality, where **X** stands, for example, for S, Se, N, C, etc. (Rule 1). A set of exemplary nonionic compounds is shown in Figure 6. Since bond orders for ligands attached to the phosphorus centers in these compounds are well defined, the assignment of D_P/L_P configuration is usually rather straightforward using Rule 1. For not fully esterified derivatives, a more careful analysis is required and this will be discussed separately (see Sections 5 and 6).



D_P-(S_P)-methane-phosphonoselenoate | L_P-(R_P)-*H*-phosphonothioate | D_P-(S_P)-phosphothioate | D_P-(R_P)-phospho-selenoimidate | L_P-(R_P)-phospho-selenoamidate

11 **12** **13** **14** **15**

FIGURE 6 D_P/L_P notation applied to dinucleoside monophosphate analogues without a P=O bond.

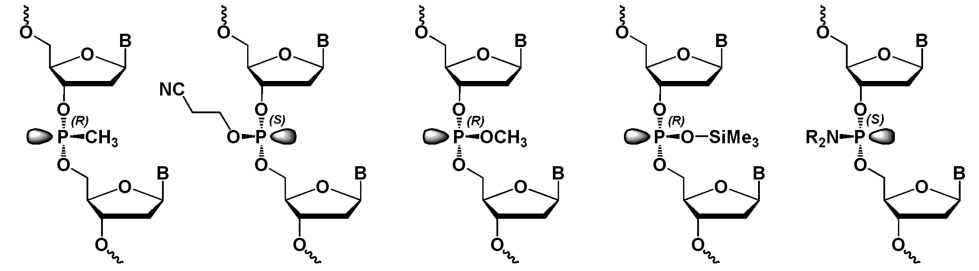


FIGURE 7 D_P/L_P notation for tervalent phosphorus dinucleoside analogues.

3. Dinucleoside Phosphite Analogues (Rule 1)

Although phosphite triesters and their analogues (phosphinates, phosphoramidites, etc.) contain only single-bonded ligands, they can be accommodated within the D_P/L_P notation system, by treating a lone electron pair as ligand **X**, and a nonnucleosidic moiety, as ligand **Z** (Rule 1). Examples of application of the D_P/L_P notation to tervalent phosphorus compounds are shown in Figure 7.

4. Dinucleoside Phosphate Analogues Containing Modified Ribose or Other Sugar Moieties (Rule 2)

The nucleotide analogues discussed so far contained 2′-deoxyribose as a sugar residue. However, D_P/L_P notation also can be applied to systems containing various furanoses (e.g., modified ribose, arabinose, or xylose), hexoses, or acyclic nucleoside derivatives. Moreover, inversion of stereochemistry at the 1′ position or lack of an aglycone does not affect the definition of the D_P/L_P notation, allowing application of this system also to α-nucleotides, fragments containing abasic units, or carbohydrate phosphates (Figure 8).

For dinucleoside monophosphates bearing sugar units other than ribose, the important aspect is a proper assignment of the ligands to $\mathbf{G_1}$ and $\mathbf{G_2}$ positions. If the attachment sites of a phosphate group to carbohydrate residues differ in carbon orders (1°, 2°, or 3°), $\mathbf{G_1}$ position (an upper unit) should be assigned to a group of higher order and $\mathbf{G_2}$ position (a lower unit) to the ligand attached through a carbon of lower order (Rule 2a). This is an extension of the conventional manner in which native dinucleotide units are usually presented and can be applied, that is, to analogues in which a phosphate group spans 1° and 2° carbon atoms (cf. structures **21–24**) or 2° and 3° ones (e.g. **25**).

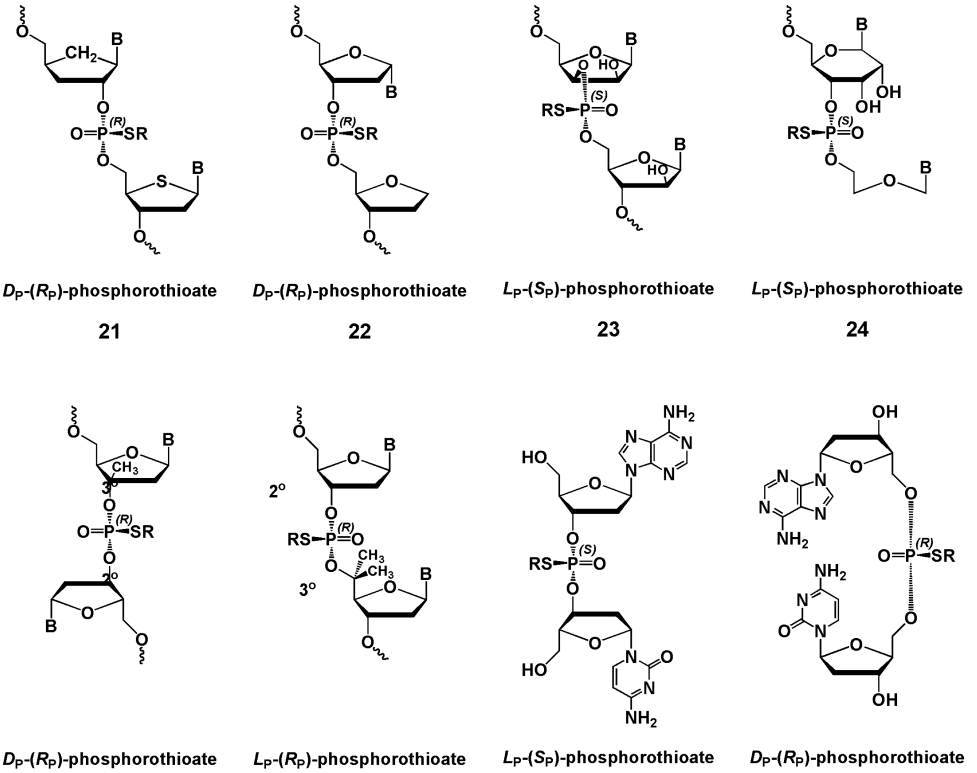


FIGURE 8 D_P/L_P notation for nucleotide analogues containing various sugar moieties and/or internucleotide linkages other than 3′–5′.

When 2 nucleoside units in a dinucleotide analogue bear obvious resemblance to a ribose-3′-yl or a ribose-5′-yl moiety, then these should be designated as $\mathbf{G_1}$ or $\mathbf{G_2}$ ligands, irrespective of the order of the carbon atoms at the attachment positions. For example, triester **26** contains 5,5′-dimethylribose, and although the attachment point is 3° carbon, this unit should be designed as $\mathbf{G_2}$ ligand, due to its obvious resemblance to a ribose-5′-yl moiety.

In the case of phosphate analogues containing carbohydrates esterified via carbons of the same order (e.g., 3′-3′ or 5′-5′ esters), position $\mathbf{G_1}$ should be assigned to the one with higher CIP priority, according to Rule 2b (e.g., phosphorothioates **27** and **28**).

The D_P/L_P system also can be applied to nucleotide analogues containing any atom in place of oxygen in a bridging position of the phosphorus center (3′ and/or 5′) (Figure 9). In contrast to the R_P/S_P system, such modifications do not affect D_P/L_P configuration of possible isomers (e.g., compare structures **29** and **30**).

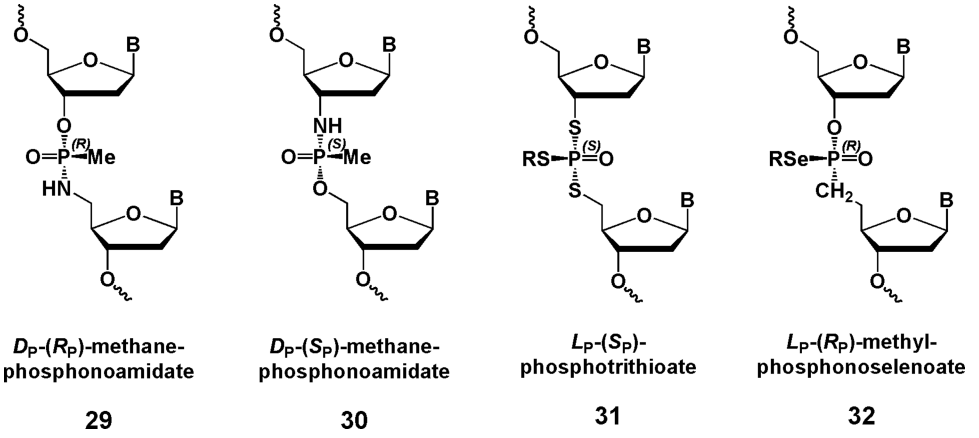


FIGURE 9 D_P/L_P notation for 3′ and 5′ modified nucleoside esters.

5. Dinucleoside Monophosphate Analogues Containing Ambident Anions (Rule 3a)

Partially esterified *P*-chiral derivatives are drawn often in their anionic forms. In such compounds, the negative charge is delocalized and results in fractional bond orders between phosphorus and both non-bridging atoms (e.g. phosphorothioate **33**, Figure 10).

Compounds with the charge distributed over the phosphorus center, in principle, are illegible to D_P/L_P notation, since none of the ligands can be assigned as **X** nor **Z**. For such derivatives, one mesomeric structure with the negative charge localized on **Z** atom should be chosen for the purpose of assignment of D_P/L_P configuration. Rule 3 provides guidance as to which of the 2 heteroatoms at phosphorus should be treated as ligand **X** (double-bonded to phosphorus), and which should bear a formal negative charge. The priority order for a P=X residue was defined as follows:

$$P{=}O > P{=}S > P{=}Se > P{=}Te > P{=}N$$

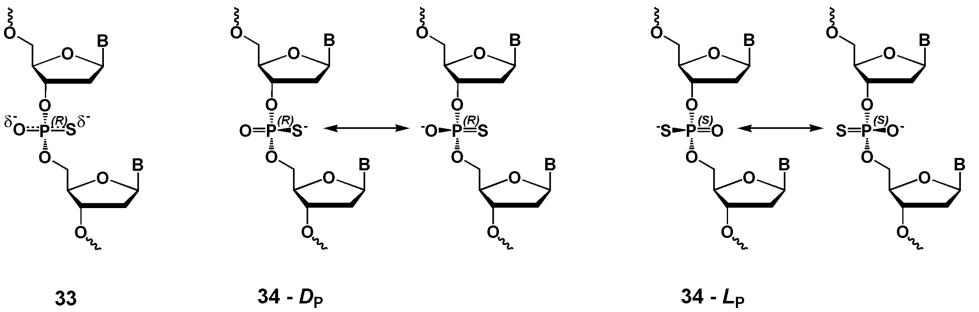


FIGURE 10 Mesomeric forms of phosphorothioate diester analogues.

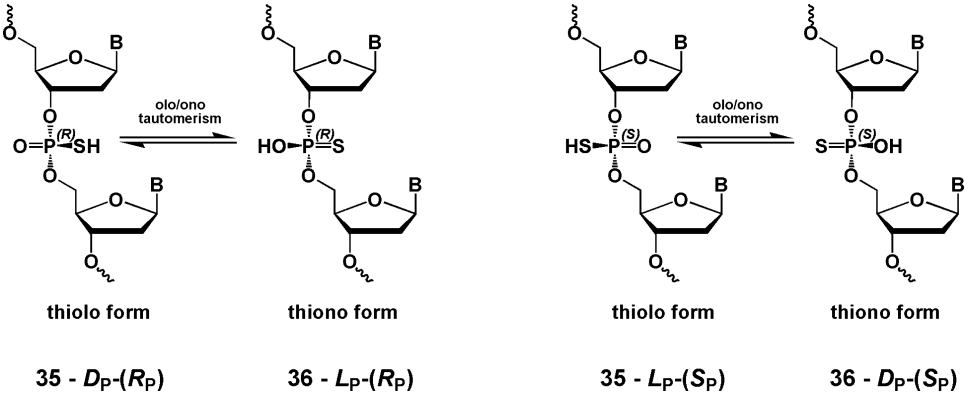


FIGURE 11 Tautomerism of phosphorothioate diesters.

Such arbitrarily chosen order of priority is congruent with the most common way of presenting ambident anions in literature. Thus, phosphorothioates should be considered as phosphorothiolates (O=P–S^-, e.g., as **34-*D*$_P$** or **34-*L*$_P$**), phosphoroselenoates as phosphoroselenolates (O=P–Se^-), phosphoroselenothioates as phosphoroselenothionates (S=P–Se^-), etc.

If, for any reason, it would be necessary to discuss separately 2 mesomeric forms of such ambident esters, one may use descriptors "pseudo-L_P" and "pseudo-D_P" to indicate that the formulas do not follow Rule 3.

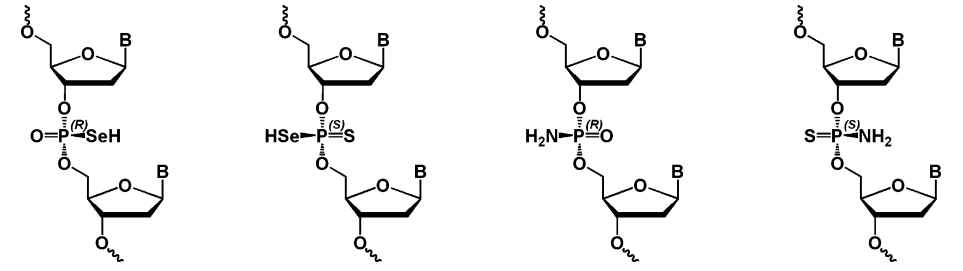
The Ono/Olo Tautomerism

When ambident phosphodiester analogues are protonated and form neutral species, their single and double bonds clearly are distinguishable, allowing direct assignment of their D_P/L_P configuration. As charged diesters have two mesomeric forms, the neutral species may exist in two tautomeric ono/olo forms. Such tautomers should be treated as separate compounds (e.g., **35** and **36**, Figure 11) and they will have opposite D_P/L_P configurations. Examples of these types of compounds include phosphorothioates, phosphoroselenothioates or other types of chiral ambident phosphodiesters.

If the preferred tautomeric form is unknown, Rule 3 should be applied for setting appropriate heteroatoms as ligands **Z** and **X**. For example, for tautomeric pairs shown in Figure 11, it would be tautomer **35**-D_P-(R_P) and **35**-L_P-(S_P) that should be used. Further examples are shown in Figure 12.

6. Dinucleoside Monophosphate Isotopomers (Rule 3b)

While Rule 3a should cover the majority of cases met in the nucleotide chemistry, nevertheless, there are compounds beyond its scope. Thus, to complete the system, a Rule 3b was formulated: if the assignment of **X** and



D_P-(R_P)-phosphoroselenolate — **37**

L_P-(S_P)-phosphoroselenothionate — **38**

L_P-(R_P)-phosphoramidate — **39**

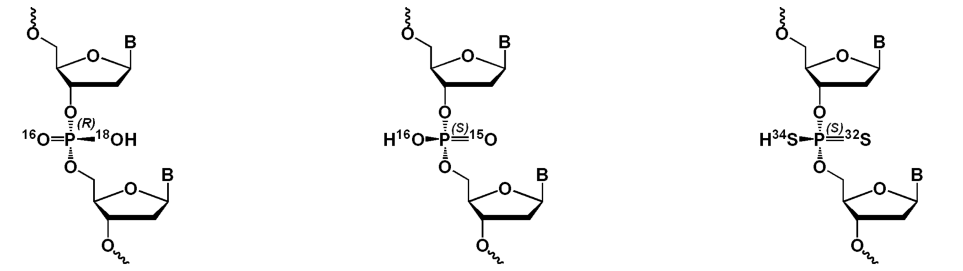
D_P-(S_P)-phosphoramidothioate — **40**

FIGURE 12 D_P/L_P notation of phosphodiester analogues.

Z positions cannot be made on the basis of the previous rules, the double bond should be assigned to the atom of lower CIP priority. The main area of application for Rule 3b is probably the assignment of D_P/L_P descriptors for isotopomers (Figure 13). For example, for dinucleoside monophosphate **41** having oxygen isotopes, ^{16}O and ^{18}O, at the nonbridging positions of the phosphorus center, oxygen ^{16}O should be chosen as a double-bonded ligand (ligand **X**) and ^{18}O, as ligand **Z**.

SUMMARY

The D_P/L_P stereochemical notation has been extended to encompass dinucleoside monophosphate analogues with single or multiple modifications at the phosphorus center, and for derivatives bearing various sugar moieties. For the purpose of configurational assignment of ionic *P*-chiral analogues with charge delocalization across the phosphorus center (chiral ambident anions), one mesomeric form should be chosen.



D_P-(R_P)-^{18}O-phosphate — **41**

L_P-(S_P)-^{15}O-phosphate — **42**

L_P-(S_P)-^{34}S-phosphodithioate — **43**

FIGURE 13 D_P/L_P notation for isotopomers.

REFERENCES

1. Cahn, R.S.; Ingold, C.; Prelog, V. Specification of molecular chirality. *Angew. Chem. Int. Ed.* **1966**, 5, 385–415.
2. Prelog, V.; Helmchen, G. Basic principles of the CIP-system and proposals for a revision. *Angew. Chem. Int. Ed.* **2005**, 21, 567–583.
3. Lebedev, A.V.; Wickstrom, E. The chirality problem in *P*-substituted oligonucleotides. *Perspect. Drug Discov. Design.* **1996**, 4, 17–40.
4. Sobkowski, M.; Stawinski, J.; Kraszewski, A. A proposal for a new stereochemical notation for *P*-chiral nucleotide analogues and related compounds. *Nucleosides Nucleotides Nucleic Acids* **2005**, 25, 1301–1309.
5. Sobkowski, M.; Stawinski, J.; Kraszewski, A. A proposal for a convenient notation for *P*-chiral nucleotide analogues. Part 3. Compounds with one nucleoside residue and non-nucleosidic derivatives. *Nucleosides Nucleotides Nucleic Acids* **2006**, 12, 1377–1389.
6. Kanaori, K.; Sakamoto, S.; Yoshida, H.; Guga, P.; Stec, W.; Tajima, K.; Makino, K. Effect of phosphorothioate chirality on i-motif structure and stability. *Biochemistry* **2004**, 43, 5672–5679.
7. Warashina, M.; Takagi, Y.; Stec, W.J.; Taira, K. Differences among mechanisms of ribozyme-catalyzed reactions. *Curr. Opinion Biotech.* **2000**, 11, 354–362.
8. Deval, J.; Alvarez, K.; Selmi, B.; Bermond, M.; Boretto, J.; Guerreiro, C.; Mulard, L.; Canard, B. Mechanistic insights into the suppression of drug resistance by human immunodeficiency virus type 1 reverse transcriptase using α-boranophosphate nucleoside analogs. *J. Biol. Chem.* **2005**, 280, 3838–3846.