PROSIT: Pseudo-Rotational Online Service and Interactive Tool, Applied to a Conformational Survey of Nucleosides and Nucleotides

Guangyu Sun, Johannes H. Voigt,† Igor V. Filippov,‡ Victor E. Marquez,* and Marc C. Nicklaus*

Laboratory of Medicinal Chemistry, Center for Cancer Research, National Cancer Institute, NIH, DHHS, NCI-Frederick, 376 Boyles St., Frederick, Maryland 21702

Received April 5, 2004

A Pseudo-Rotational Online Service and Interactive Tool (PROSIT) designed to perform complete pseudorotational analysis of nucleosides and nucleotides is described. This service is freely available at http://cactus.nci.nih.gov/prosit/. Files containing nucleosides/nucleotides or DNA/RNA segments, isolated or bound to other molecules (e.g., a protein) can be uploaded to be processed by PROSIT. The service outputs the pseudorotational phase angle *P*, puckering amplitude ν_{max}, and other related information for each nucleoside/nucleotide detected. The service was implemented using the chemoinformatics toolkit CACTVS. PROSIT was used for a survey of nucleosides contained in the Cambridge Structural Database and nucleotides in high-resolution crystal structures from the Nucleic Acid Database. Special cases discussed include nucleosides having constrained sugar moieties with extreme puckering amplitudes, and several specific DNA/RNA helices and protein-bound DNA oligonucleotides (Dickerson-Drew dodecamer, RNA/DNA hybrid viral polypurine tract, Z-DNA enantiomers, B-DNA containing (L)-α-threofuranosyl nucleotides, TATA-box binding protein/TATA-box complex, and DNA (cytosine C5)-methyltransferase complexed with an oligodeoxyribonucleotide containing transition state analogue 5,6-dihydro-5-azacytosine). When the puckering amplitude decreases to a small value, the sugar becomes increasingly planar, thus reducing the significance of the phase angle *P*. We introduce the term "central conformation" to describe this part of the pseudorotational hyperspace in contrast to the conventional north and south conformations.

INTRODUCTION

The ability of the furanose ring in nucleosides(tides) to adopt multiple conformations, often with comparatively little difference in energy, represents a formidable challenge to our understanding of the relationship between conformation and biological activity. A picture emerging from recent studies is that the majority of nucleoside(tide) target enzymes, whether anabolic or catabolic, appear to have strict conformational requirements for substrate binding, accepting the furanose ring only in a specific well-defined shape.[1−4] Four conformational parameters[5] are required to define the shape of a nucleoside(tide): (A) the glycosyl torsion angle χ, which determines the syn or anti disposition of the base relative to the sugar moiety; (B) the torsion angle γ, which determines the orientation of the 5′-OH (or 5′-phosphate); (C) the puckering of the furanose ring described by the phase angle of pseudorotation *P* (0°−360°); and (D) the degree of deviation from planarity of the furanose ring indicated by the maximum out-of-plane pucker ν_{max}. These last two parameters were first defined by Altona and Sundaralingam[6] when they introduced the concept of pseudorotation to describe the puckering motion of the five-member furanose ring in nucleosides(tides). The value of *P* in combination with the puckering amplitude ν_{max} can adequately describe the sugar pucker because only two torsion angles are independent variables even though there are five endocyclic torsion angles in a furanose ring. The pseudorotational wheel results when *P* and ν_{max} are plotted in a polar graph, as shown in Figure 1. By convention, a phase angle of $P = 0°$ corresponds to an absolute north conformation possessing a symmetrical twist form $^{3}_{2}$T, whereas its south antipode, $^{2}_{3}$T, is represented by $P = 180°$. For the majority of nucleosides, the value of *P* normally falls in a tight range in the vicinity of either one of these north and south extremes. While in the solid state only one form predominates, in solution the two conformations are in a rapid dynamic equilibrium dictated by the balance of stereoelectronic, gauche, and anomeric effects, which are in turn influenced by the electronegativity, ionization state, steric bulk, and relative stereochemistry of all the substituents on the furanose ring.

When the furanose ring occupies similar domains of *P* ($^{3}_{2}$T or $^{2}_{3}$T) in polymeric structures (e.g., RNA and DNA), this leads to two different categories of polynucleotide conformations known as A- and B-type families. The A-type conformation is generically described as having the furanose rings close to a 3′-endo environment ($P = 18°$) whereas in the B-type the conformations the furanose rings are in an environment close to a 2′-endo conformation ($P = 162°$). Changes between A- and B-type conformations can occur in response to environmental factors, such as degree of hydration and metal ion coordination, as well as interactions with small molecules and protein binding. Such an inherent malleability of DNA and RNA is inextricably linked to their important biological roles.

* Corresponding author fax (301)846-6033; e-mail: VEM, marquezv@mail.nih.gov; MCN, mn1@helix.nih.gov.

† Current address: Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033.

‡ Funded in part by DHHS #NO1-CO-12400 through Basic Research Program, SAIC-Frederick, Inc.

10.1021/ci049881+
Published on Web 07/13/2004

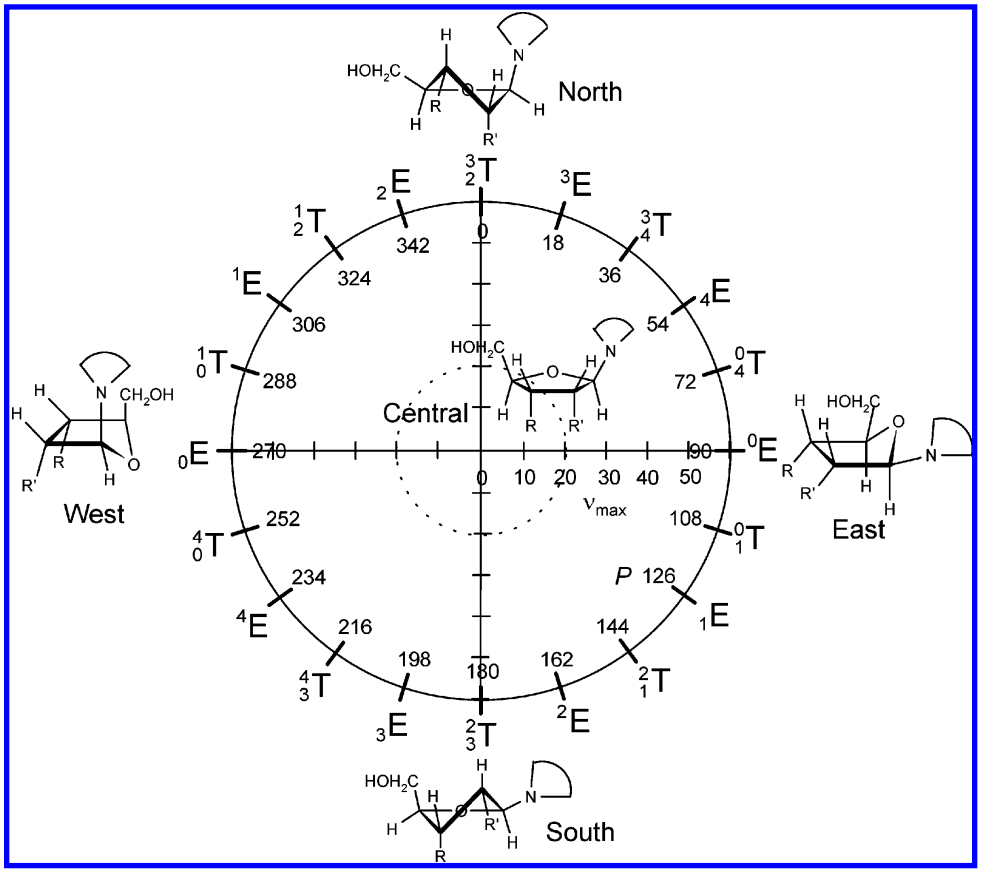


Figure 1. The pseudorotation wheel for pentofuranosyl nucleosides showing the characteristic north, south, east, west, and central conformations. T denotes twist and E denotes envelope. The units of the *P* and ν_{max} are degrees. The radius of the wheel corresponds to ν_{max} and the small concentric circle encompasses a central region characterized by a flatter sugar moiety ($\nu_{max} < 20°$).

The need to rapidly and accurately access information about the state of the furanose moiety in nucleosides(tides) individually, embedded in a polynucleotide DNA/RNA chain isolated, or in complex with other molecule(s) (most commonly proteins) is thus encountered in various contexts when dealing with such molecules: One may need to gain a quick overview of the structural details of a crystal structure; one may want to know what the sugar pucker would be in a proposed novel molecule, for example, to be synthesized in the context of drug design; or one may want to analyze what the effect of incorporation of an unnatural nucleotide would be on the overall structure of a DNA tract. This need, often felt in our own modeling work in these fields, and to our knowledge unmet by any of the major molecular modeling packages in an automated, rapid, and convenient manner, prompted us to develop the Pseudo-Rotational Online Service and Interactive Tool (PROSIT). PROSIT is an online tool developed to rapidly calculate Altona's pseudorotational parameters[6] and is available at http://cactus.nci.nih.gov/prosit/. Files containing 3D coordinates of a single compound, multiple compounds, or complexes can be submitted through the web browser at the above URL. The tool automatically identifies all the nucleoside(tide) molecules amenable for the calculations, even when present as part of a large macromolecular complex. This allows the user to submit directly a file from the Protein Data Bank (PDB) without the need to identify or mark the DNA, RNA, small oligonucleotide, or individual nucleoside(tide) molecule(s) contained in the file. When PROSIT is applied to an individual nucleoside(tide), the tool is complementary to the program PSEUROT developed by Altona and co-workers[7−9] to calculate pseudorotational parameters based on experimental NMR coupling constants. PROSIT is one of several online services offered by the Computer-Aided Drug Design Group of the Laboratory of Medicinal Chemistry, National Cancer Institute at Frederick, National Institutes of Health (NIH). These services include, among others, the Enhanced NCI Database Browser Release 2,[10] the Self-Organized Map (SOM) of Compounds Tested in the NCI anti-HIV Screen, the Cancer Screening Data 3Dminer, and the Online SMILES Translator and File Format Converter. All these services can be accessed at http://cactus.nci.nih.gov/. Particularly worth mentioning in this context is the Pseudorotation Visualization tool (http://cactus.nci.nih.gov/pseudo-visual/), which allows the user to modify *P* and ν_{max} interactively and observe the resulting conformational changes in a 3D model of a five-membered ring.

While for the analysis of "normal" DNA and RNA structures it may seem sufficient to be able to detect and process the five standard nucleotides, in other contexts, particularly in drug development, analogues of the standard nucleosides(tides) with "unnatural" sugar and/or nucleobase moieties play an important role. Therefore, great care was taken to make the pattern used in the substructure search routine as general as possible, while preserving the requirement that the structure(s) recognized should still be similar to what most chemists in the field would consider to be a nucleoside(tide) analogue. Obviously, the exact location of this boundary is a matter of definition. It will therefore always be possible to construct a structure that someone might see

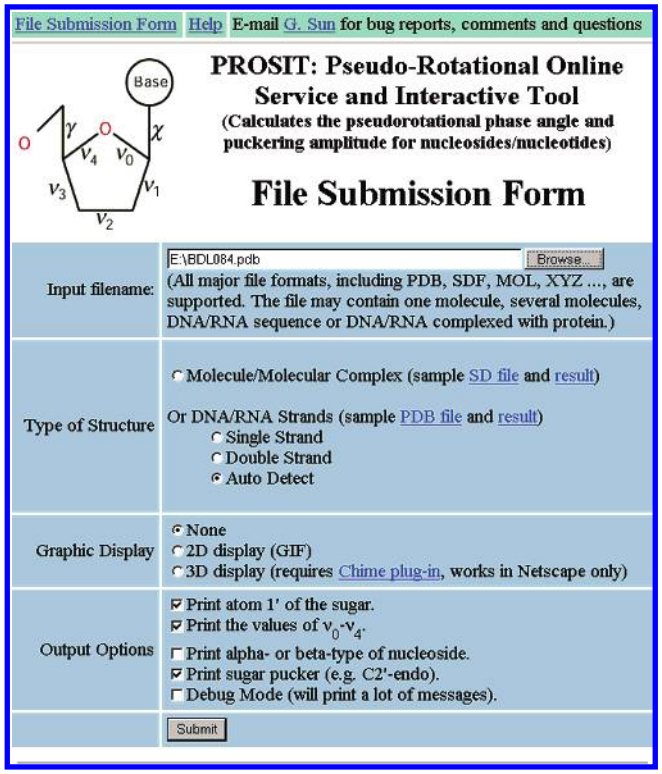


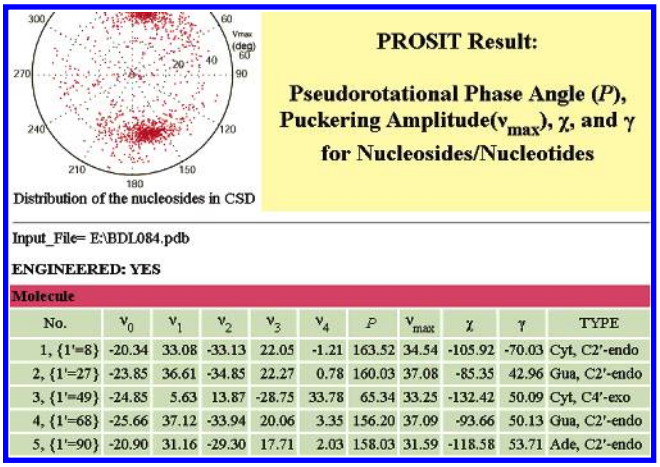
Figure 2. Typical screen display of the File Submission Form of PROSIT available at http://cactus.nci.nih.gov/prosit/.

as a nucleoside analogue but that the tool will not recognize as such. Examples of what the tool currently recognizes include analogues with carbocyclic "sugar" moieties, sugars containing double bonds, sugars or pseudosugars with practically any replacement of hydrogen by a functional group, and α-nucleosides. Examples of structures not currently recognized include DNA abasic sites and molecules containing seven-membered rings in the nucleobase moiety.

The present manuscript first describes the user interface and the underlying algorithm of the online tool. We then present a survey in which PROSIT was used to analyze all the nucleosides(tides) contained in high-resolution structures in the Cambridge Structural Database (CSD)[11] and in the Nucleic Acid Database (NDB),[12] which both serves to exemplify the usefulness of the tool for convenient and rapid analysis of structures and delivers an interesting overview of actually occurring molecules. Finally, selected molecules with extreme sugar conformations and several DNA or RNA fragments are presented as examples to illustrate the scope and robustness of the online service.

USER INTERFACE

To calculate the pseudorotational parameters using PROSIT, the 3D coordinates of the compound or complex are required. The user needs to upload the chosen file containing the 3D coordinates through the File Submission Form at http://cactus.nci.nih.gov/prosit/. A typical screen sample of the File Submission Form is shown in Figure 2. The file can be in various chemical representation formats such as PDB, SDF, MOL, or XYZ. Removal of molecules that are not of interest is not necessary since the nucleosides(tides) are automatically identified. The user can specify whether the file contains a molecule or molecular complex, or DNA/RNA strands. For DNA/RNA strands, the user can select a



No.	ν0	ν1	ν2	ν3	ν4	P	νmax	χ	γ	TYPE
1, {1'=8}	-20.34	33.08	-33.13	22.05	-1.21	163.52	34.54	-105.92	-70.03	Cyt, C2'-endo
2, {1'=27}	-23.85	36.61	-34.85	22.27	0.78	160.03	37.08	-85.35	42.96	Gua, C2'-endo
3, {1'=49}	-24.85	5.63	13.87	-28.75	33.78	65.34	33.25	-132.42	50.09	Cyt, C4'-exo
4, {1'=68}	-25.66	37.12	-33.94	20.06	3.35	156.20	37.09	-93.66	50.13	Gua, C2'-endo
5, {1'=90}	-20.90	31.16	-29.30	17.71	2.03	158.03	31.59	-118.58	53.71	Ade, C2'-endo

Figure 3. Screenshot of a typical result page of PROSIT.

"Single Strand" option that causes the results of the identified nucleotides to be printed following the order in which they are found. The "Double Strand" option pairs the first nucleotide with the last, the second with the second to last, etc. to denote the existence of a double-helix structure when the number of nucleosides is even. The user should be aware that a double-stranded DNA/RNA with an overhang of an even number of nucleotides would lead to a mismatch in the assignment of the base pairs by the tool with the "Double Strand" option. If the number of identified nucleotides is odd, the nucleotides will simply be listed in the order they were found. The "Auto Detect" option determines the sequence of each strand and prints the results for each strand sequentially from the 5′-end to the 3′-end. Additional display options can be selected in the "File Submission Form", including 2D or 3D graphic displays. The values of the torsion angles $\nu_0-\nu_4$, the anomeric position ($C1'$), stereochemistry of the sugar, and sugar pucker characteristics can be requested to be included as part of the final output. The Debug Mode prints large amounts of messages for debugging purposes.

Once the Submit button on the File Submission Form is clicked, the chosen file is transferred to the web server together with the selection of display options. The pseudorotational phase angle (P), puckering amplitude (ν_{max}), torsion angle χ, and torsion angle γ for the nucleosides(tides) are calculated, tabulated in HTML format, and displayed in the web browser client. Figure 3 shows a typical output of the first five nucleotides belonging to one strand of a DNA duplex. Any 2D or 3D display of the compound would follow (not shown in Figure 3). If the uploaded file contains more than one entry detected to be processable, each entry is treated and displayed sequentially on the same results page. After the file is processed, the definition of the parameters appears, followed by a summary that lists the number of compounds in the file, number of nucleosides with purine/pyrimidine bases, and the number of α-/β-anomers. At the end of the results page a line indicates the normal termination of the script, and the time of the calculation is posted.

COMPUTATIONAL METHODS

PROSIT consists of two parts: (A) The File Submission Form and the Help page, both written in standard HTML; and (B) a CGI script written in the Tcl scripting language that performs the actual identification of nucleosides(tides),

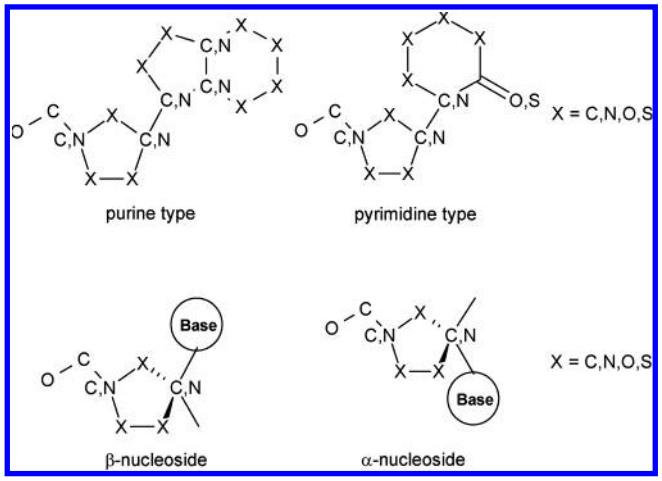


Figure 4. Connection patterns of purine- and pyrimidine-type nucleosides used in substructure searches to identify nucleosides(tides) in PROSIT. PROSIT is able to distinguish between α- and *β*-nucleosides.

calculation of pseudorotational parameters and presentation of the results. PROSIT's CGI script uses the Chemical Algorithms Construction, Threading, and Verification System (CACTVS) v.3.258.[13] CACTVS is a chemoinformatics extension of Tcl capable of handling chemistry-related information. The current release of free academic version of CACTVS can be obtained from http://www.xemistry.com/academic/.

At the beginning of the script, the uploaded file is written to a temporary location. Each compound in the file is, in turn, read in (automatically performed by the CACTVS command "molfile loop"), and is checked for the existence of 3D coordinates. Those compounds with 3D coordinates are subjected to a substructure search ("match ss" CACTVS command) in order to identify all possible sugar moieties. The atoms at the *C*1′ position of the sugar moieties are then used as anchors in the search for nucleosides. The patterns shown in Figure 4 are coded in the script as SMARTS strings, the substructure specification dialect of the SMILES language,[14] and used as the search criteria for nucleosides with purine and pyrimidine bases. The resulting list of atoms is used to define the torsion angles needed to calculate the pseudorotational parameters. To accommodate the diverse variety of synthetic nucleosides, we use a fairly loose definition of what constitutes a nucleoside in the substructure search. It is the user's responsibility to verify that the structure(s) matched by the search patterns do indeed correspond to nucleosides(tides) or analogues. The standard deoxyribose and ribose nucleosides(tides) are recognized at this step and labeled on the results page as Ade, Gua, Cyt, and Thy for 2′-deoxyriboses and ade, gua, cyt, and ura for riboses. The α/*β* stereochemistry of the sugars (Figure 4), defined by the stereo configuration of the *C*1′ atom, is determined within the context of the CACTVS substructure search functionality. The stereo type is denoted accordingly as "Alpha" or "Beta" on the results page. In cases where the stereo configuration of the *C*1′ atom is not clear, it is labeled as "unknown". Though quite unlikely to happen with "normal" structures submitted with reasonable 3D coordinates, this could occur, for example, if the nucleoside analogue contains a double bond to the *C*1′ position in the "sugar" ring or if the *C*1′ position has a nitrogen instead of a carbon atom—both cases that would be matched by the SMARTS pattern currently defined as acceptable substructures.

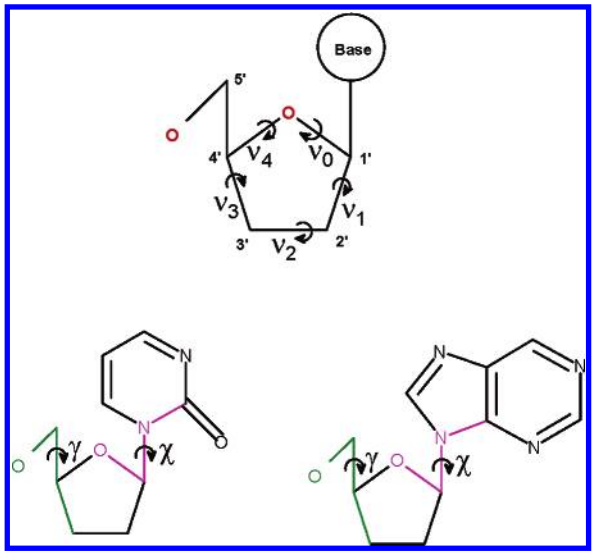


Figure 5. Depiction of torsion angles present in nucleosides (ν_0–ν_4, χ and γ).

The torsion angles ν_0–ν_4, χ, and γ are calculated from the 3D coordinates, following Saenger's notation[5] as shown in Figure 5. The pseudorotational phase angle (*P*) and the maximum puckering amplitude (ν_{max}) are calculated from eqs 1–3. Since the arctan function produces angles in the range of −90° to 90°, 180° is added to arctan(*p*) when ν_2 is negative to obtain the *P* values in the range of 90°−270° (south); 360° is added to arctan(*p*) in the remaining cases where *p* is negative to obtain *P* values in the range of 270°−360° (north).

$$p = \frac{(\upsilon_4 - \upsilon_0) - (\upsilon_3 - \upsilon_1)}{2{*}\upsilon_2{*}(\sin 36^\circ + \sin 72^\circ)} \tag{1}$$

$$P = \begin{Bmatrix} \arctan p + 180^\circ \cdots\cdots\cdots \textit{if } \upsilon_2 < 0, \textit{else} \\ \arctan p + 360^\circ \cdots\cdots\cdots \textit{if } p < 0, \textit{else} \\ \arctan p \end{Bmatrix} \tag{2}$$

$$\upsilon_{max} = abs\left(\frac{\upsilon_2}{\cos P}\right) \tag{3}$$

When the "Auto Detect" option is selected for DNA/RNA strands, the number of strands and the sequence of each strand are automatically determined by searching for the phosphorus linkage between the nucleotides. First, the 5′ and 3′ phosphorus atoms are separately identified for each nucleotide. The sequence of a strand is then determined when a phosphorus atom is shared between one nucleotide at the 3′ position and another at the 5′ position. The results are listed sequentially in the order of 5′ → 3′ for each strand.

If requested by the user, the optional 2D graphic display is generated using the GIF structure-drawing feature of the CACTVS toolkit. For 3D display, the Chime plug-in from MDL Information Systems, Inc. is required. The Chime plug-in can be obtained free of charge from http://www.mdlchime.com/.

The script for PROSIT is structured in such a manner that later modifications to the script for the purpose of adding new types of nucleosides(tides) to the list of recognized motifs requires only the simple addition of the SMARTS

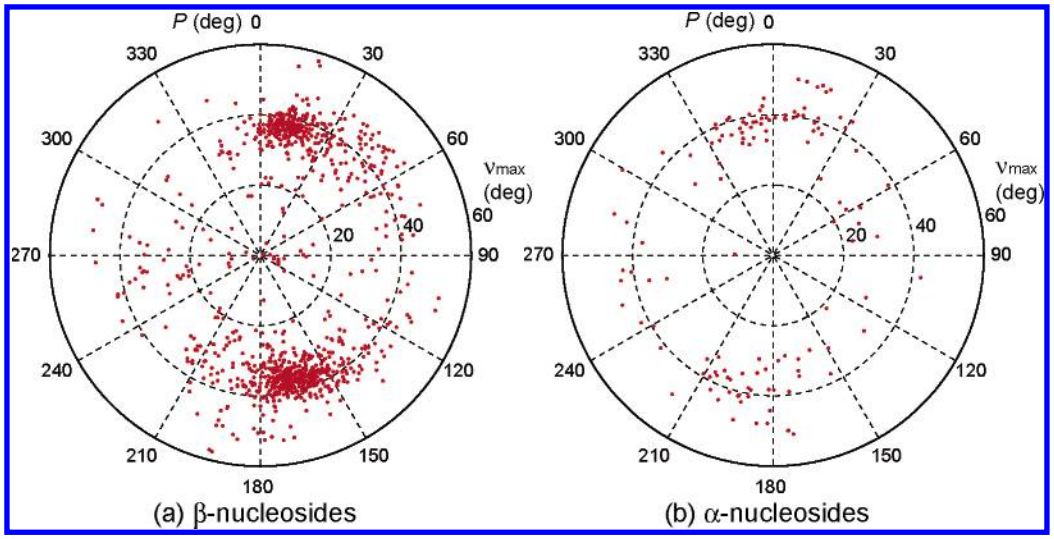


Figure 6. Distribution of the pseudorotational phase angle P and puckering amplitude ν_{max} for β-nucleosides (a) and α-nucleosides (b) in the Cambridge Structural Database.

strings of the new sugar and nucleoside structures. This capability is demonstrated below with an example of a nontraditional DNA structure containing (L)-α-threofuranosyl nucleotides.

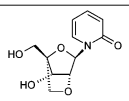
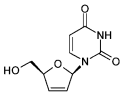
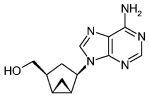
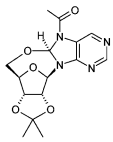
RESULTS AND DISCUSSION

We have used PROSIT to calculate the pseudorotational parameters for all the nucleosides found in the Cambridge Structural Database, as well as nucleotides in either RNA or DNA strands that crystallize alone or bound to proteins found in the Nucleic Acid Database. This section presents the results of these studies. This both illustrates the convenience and power of the tool for analyzing even quite large datasets and allowed us to generate an overview of structural trends of interest in the fields of (oligo-) nucleotides, nucleosides, and analogues thereof.

Nucleosides in the Cambridge Structural Database. Pattern searches in the CSD[11] using the ConQuest software[15] returned more than 980 entries having at least one nucleoside with either a purine or pyrimidine base. Among them, 871 entries that have 3D coordinates and an R-factor of less than 0.075 were selected for the calculation of pseudorotational parameters. Since many asymmetric units in the CSD entries contain more than one molecule, a total of 1161 nucleosides were identified. The distribution of P and ν_{max} is plotted in Figure 6 showing the hyperspace of geometries accessible to north- and south-type pseudorotamers for β- and α-nucleosides. Most β-nucleosides (Figure 6a) have either a north-type conformation with P within the 0°−30° range, or a south-type conformation with P values between 150°−180°. The north/south hyperspace for the fewer α-nucleosides appears slightly shifted toward the western hemisphere (Figure 6b). The distributions obtained from these calculations are nearly identical to previously reported figures in Deleeuw et al.[16] and Latha et al.[17] The ν_{max} values, corresponding to the radii of the concentric circles in Figure 6, fall mainly in the range of 30°−45°.

While the majority of nucleosides have a puckering amplitude (ν_{max}) below 45°, a number of compounds have ν_{max} values as large as 60°. A group of compounds with a large ν_{max} value are the synthetic 2′-*O*, 4′-*C*-methylene bridged nucleotides. One such a compound (Table 1, entry 1, EBOZAW) has a locked north conformation with P = 16.67° and ν_{max} = 57.59°.[18] We performed quantum-mechanical calculations in the gas phase for this compound with Gaussian 03[19] using Density Functional Theory (DFT) at the B3LYP/6-31G* level of theory.[20,21] The results (Figure S1 and Table S1 in Supporting Information) show that the X-ray crystal structure is indeed very close to the optimal structure. This confirms that the unusually high ν_{max} value of the molecule is not the result of crystal packing forces and that the extreme C3′-endo pucker observed for the sugar reflects the strain of the [2.2.1] bicyclic moiety.

Table 1. Several Compounds with Extreme Sugar Conformation from the Cambridge Structural Database

CSD entry	Structure	P (deg)	ν_{max} (deg)	Sugar Conformation	Ref.
EBOZAW		16.67	57.59	North	18
WAXLEM		97.15 65.35	6.77 0.91	Central Central	22
Not Available		92.38	29.84	East	27
JEHNIT		276.02	34.50	West	28

At the other end of the scale, there are compounds that have small puckering amplitudes. When the puckering amplitude is very small, the five-member ring has essentially become planar and the meaning of the phase angle P is therefore less significant. As a complement to the north/south terminology, we introduce here the concept of a "central conformation" to describe nucleosides with small ν_{max} values of less than 20° (Figure 1). Compounds having a central conformation include 2′,3′-didehydro-2′,3′-dideoxynucleosides[22] with a C=C bond between C2′ and C3′. One member of this class of compounds, 2′,3′-didehydro-2′,3′-dideoxyuridine (Table 1, entry 2, WAXLEM), has ν_{max} values of 0.91°

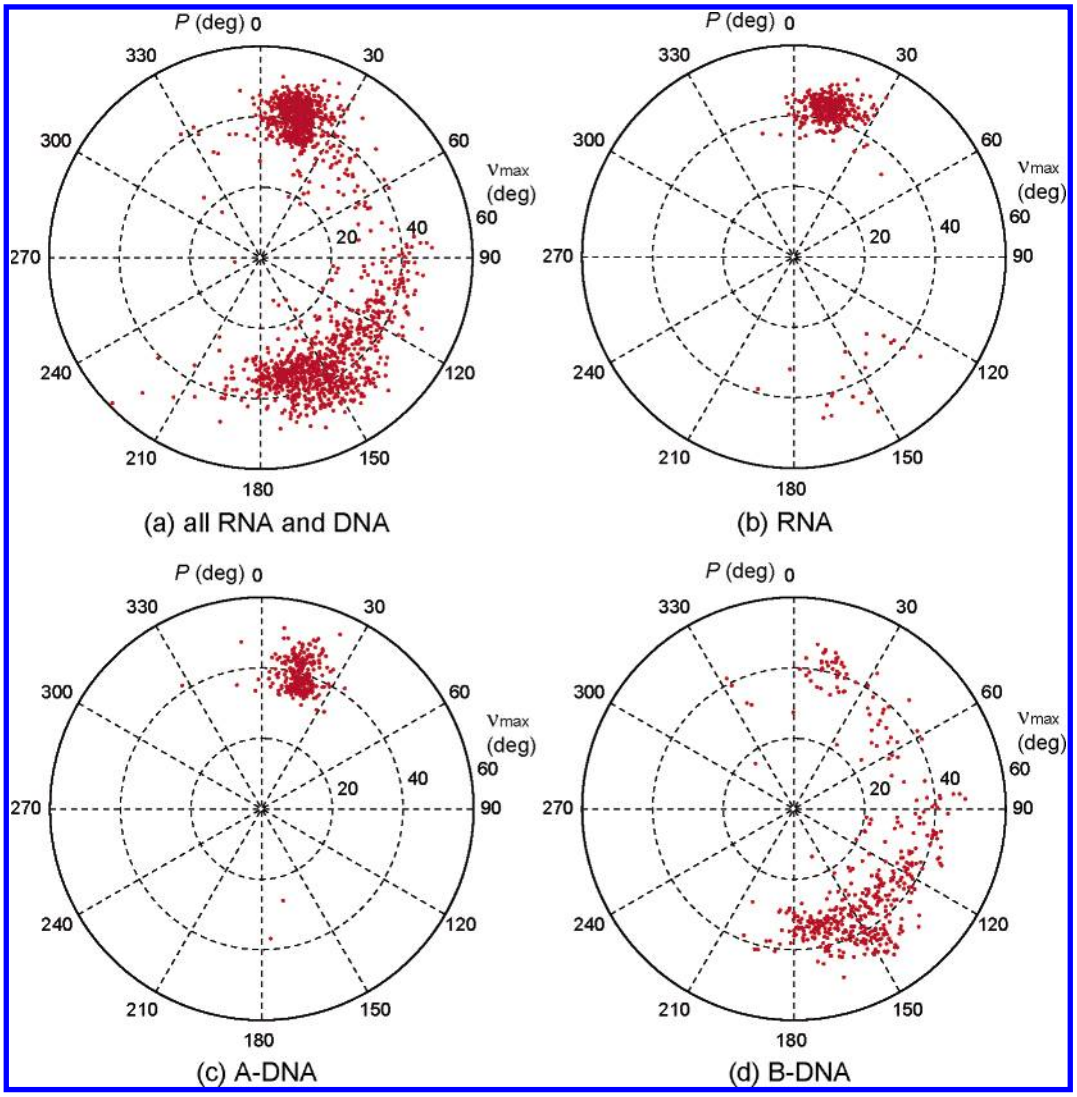


Figure 7. Distribution of the pseudorotational phase angle P and puckering amplitude ν_{max} of nucleotides of all (a) RNA and DNA combined, (b) RNA, (c) A-DNA, and (d) B-DNA in selected high-resolution crystal structures from the Nucleic Acid Database (NDB).

and 6.77°, respectively, for the two molecules in the asymmetric unit. Nucleosides having a central conformation also include the anti-HIV active compounds D4T[23] and N-MCD4T.[24] On the other hand, the structurally similar anti-HIV active compound, carbovir, has a ν_{max} value of 28° and P around 90° in the single-crystal structure,[25] which may seem to conflict with structures observed for D4T and N-MCD4T. However, a recent ab initio study[26] has shown that carbovir has a flat potential energy surface when the tip of the five-membered ring (4′-O) is moved "up" or "down" between two envelope conformations, while at the same time retaining a planar conformation around the 2′,3′ C═C bond. This indicates that the molecule can access the east, central, and west sugar conformations easily. Such a flexibility in carbovir has been speculated to be essential for its anti-HIV activity.[26]

Few compounds are found in the east or west conformations. These relatively rare cases result mainly when the molecule has a constrained sugar structure. Examples of these are an east methanocarba nucleoside with a fused *endo* cyclopropane ring at the 2′,3′ position (Table 1, entry 3)[27] and the west 7,8-dihydro-8,5′-cyclonebularine analogue with a two-atom tether connecting C8 and C4′ (Table 1, entry 4).[28]

DNA/RNA Duplexes. Single-crystal structures of DNA segments obtained from X-ray diffraction studies have provided a wealth of structural information about the nature of DNA duplexes.[29] PROSIT can rapidly and automatically provide the pseudorotational parameters associated with the individual sugar moieties of DNA and other oligomeric structures. However, an adequate assessment of the importance of the conformational polymorphism of the sugar moieties in these structures can be questioned if the uncertainties of the refined 3D coordinates are such that they entail considerable uncertainty in the ν_0 through ν_4 values used as input for the pseudorotational calculations. Therefore, only high-resolution structures were selected to calculate pseudorotational parameters with PROSIT in the following analysis.

In the Nucleic Acid Database,[12] we found 72 entries with 10 or more residues per strand that were obtained with 1.55 Å resolution or better. Among these 72 entries (see Table 2S in Supporting Information), twelve are RNA structures and nine are DNA or RNA structures bound to proteins; the rest are DNA structures. The combined results of the pseudorotational parameters of the 1629 nucleotides identified in these structures are shown in Figure 7a with the exception of two nucleotides having ν_{max} values larger than

60° (which thus fall outside the graph area). The majority of the nucleotides in these crystal structures of DNA or RNA are found in the usual north and south hemispheres. On the north side of the pseudorotation wheel, the sugar conformations concentrate heavily within the $0° \leq P \leq 30°$ and $30° \leq \nu_{max} \leq 50°$ area. RNA (Figure 7b) and A-form DNA (Figure 7c) are the main constituents of the north cluster. On the south side, a wider range of sugar conformations seems to be accessible in oligonucleotides compared to what is seen with isolated single nucleosides (compare Figures 6a and 7a). The normal range of $140° \leq P \leq 180°$ and $30° \leq \nu_{max} \leq 50°$ is still more populated than other areas, but a significant number of nucleotides have an east sugar conformation with *P* values of ca. 90° (O4′-endo). Figure 7d shows that B-form DNA represents the largest source of sugars with a south conformation. The central part and the entire west side of the pseudorotation wheel are sparsely populated. This is not unexpected since nucleotides having a west or central sugar conformation are normally chemically modified, and such nonstandard nucleotides have not appeared in many of the deposited high-resolution crystal structures of DNA or RNA helices.

In the following section, we present and discuss several interesting individual DNA/RNA structures as useful applications of PROSIT.

B-DNA Dodecamers. The Dickerson-Drew DNA dodecamer d(CGCGAATTCGCG) was the first right-handed B-form DNA whose X-ray crystal structure was solved.[29,30] The most recent high-resolution crystal structures of this molecule (NDB BDL084)[31] and two modified sequences[32] with one or two 2′-deoxy-2′-fluoroarabinofuranosyl thymine (S) residues per strand (ST dodecamer, NDB BDLB84 and SS dodecamer, NDB BDLB85) were evaluated using PROSIT with the Auto Detect option "on" to let the tool determine the sequence. The calculated phase angle *P* and puckering amplitude ν_{max} (Figure 8) were essentially the same as the values reported earlier.[31] The B-form helical structure of the Dickerson-Drew dodecamer is the result of the Z-incompatible AATT segment counteracting the Z-forming tendencies of the tetrameric CGCG ends.[29] The Z-forming tendencies of the CGCG segments, however, are still visible in the slight zigzag modulation at both ends of the helix, whereas the central part of the helix is much smoother. The majority of the sugars in the Dickerson-Drew dodecamer have south sugar conformations with a C2′-endo pucker. Some deviations from the low-energy sugar conformations are seen in the central part of the helix where T7, T8, and A18 have the C1′-exo pucker with *P* around 126°, and T19 has a ${}^{2}_{1}\mathrm{T}$ conformation with $P = 143.02°$.

The overall structures of the modified SS and ST dodecamers resemble that of the native dodecamer, but noticeable structural differences were found for the modified residues themselves.[32] The 2′-deoxy-2′-fluoroarabinofuranosyl thymines incorporated in the oligonucleotides assume the east conformation, although the individual nucleosides have been shown to favor the south conformation.[33] All six S residues (four in the SS dodecamer and two in the ST dodecamer) have *P* values within 95°−104°, clearly indicating the presence of an O4′-endo pucker. The east conformations of the S residues seem to follow the trend of the T residues in the native dodecamer insofar as they deviate from the low-energy south conformation toward the higher-energy east conformation.

```
BDL084
                    5' 3'
  (163.52, 34.54)  1 C--G 24  (13.75, 34.32)
  (160.03, 37.08)    G--C     (17.53, 36.16)
   (65.34, 33.25)    C--G    (151.20, 40.28)
  (156.20, 37.09)    G--C     (93.33, 40.61)
  (158.03, 31.59)    A--T    (151.41, 37.05)
  (152.63, 30.64)    A--T    (143.02, 34.24)
  (125.24, 34.76)    T--A    (127.51, 32.43)
  (126.89, 36.72)    T--A    (173.29, 31.36)
  (157.94, 32.21)    C--G    (168.80, 30.79)
  (145.61, 43.62)    G--C     (42.00, 37.79)
  (163.82, 36.64)    C--G    (129.57, 33.12)
  (207.36, 25.36) 12 G--C 13 (163.51, 30.67)
                    3' 5'

BDLB85
                    5' 3'
  (169.36, 35.52)  1 C--G 24   (8.76, 37.58)
  (159.34, 33.31)    G--C     (16.74, 34.53)
   (39.20, 35.26)    C--G    (149.22, 39.91)
  (168.79, 32.85)    G--C    (127.54, 39.81)
  (152.79, 35.29)  5 A--T 20 (120.11, 40.02)
  (131.32, 35.99)    A--S     (96.30, 46.24)
   (95.34, 41.85)    S--A    (135.25, 33.05)
  (121.95, 36.83)    T--A    (163.38, 32.73)
  (159.78, 34.54)    C--G    (163.25, 33.02)
  (149.83, 43.39) 10 G--C 15  (35.92, 37.19)
  (160.98, 34.74)    C--G    (136.02, 30.77)
   (88.24, 42.09) 12 G--C 13 (151.12, 45.05)
                    3' 5'

BDLB84
                    5' 3'
  (169.21, 34.61)  1 C--G 24  (17.56, 36.49)
  (158.41, 32.63)    G--C     (20.81, 32.25)
   (38.41, 37.09)    C--G    (154.92, 41.16)
  (163.81, 34.16)    G--C    (110.96, 33.47)
  (160.16, 30.94)  5 A--S 20  (95.72, 44.59)
  (144.89, 29.50)    A--S    (102.40, 34.91)
  (103.09, 38.97)    S--A    (148.75, 33.04)
   (98.36, 40.62)    S--A    (170.60, 31.83)
  (153.77, 34.91)    C--G    (172.60, 32.96)
  (147.85, 43.68) 10 G--C 15  (31.94, 35.40)
  (163.40, 33.97)    C--G     (41.14, 26.77)
   (73.91, 35.65) 12 G--C 13 (164.33, 38.19)
                    3' 5'
```

Figure 8. Pseudorotational phase angle *P* and puckering amplitude ν_{max} for a conventional Dickerson-Drew dodecamer (BDL084) and two modified dodecamers containing 2′-deoxy-2′-fluoroarabinofuranosyl thymine (S) residues at positions 7 and 19 (BDLB85) and at positions 7, 8, 19, and 20 (BDLB84).

RNA/DNA Hybrid Viral Polypurine Tract. During the retroviral conversion of single-stranded RNA into double-stranded DNA in human immunodeficiency virus (HIV), the viral polypurine tract (PPT) is the primer in the synthesis of the second strand.[34] These purine tracts are resistant to cleavage during first-strand synthesis and need to be cleaved by reverse transcriptase (RT) for the second strand synthesis. At the RNase H active site, RT makes contacts primarily along the backbone of the RNA of which the sugar moieties are an integral part. Recently, a high-resolution (1.10 Å) X-ray structure has been obtained for the first 10 base-pairs of the RNA/DNA hybrid PPT, r(caaagaaaag)/d(CTTTTC-TTTG).[35] It was found that in the RNA strand the adenine-2 ribose possesses C2′-endo conformation instead of the usual C3′-endo conformation of RNA. Using PROSIT to confirm this finding we calculated the pseudorotational parameters for the sugar moieties in the high-resolution X-ray structures

```
UH0005
                    5' 3'
  (4.79, 39.79)   1 c--G 20 (178.32, 42.41)
(173.72, 48.75)     a--T     (24.80, 33.75)
  (9.31, 42.04)     a--T     (13.36, 38.60)
 (12.63, 37.47)     a--T     (19.02, 39.20)
  (0.24, 43.95)     g--C     (16.90, 39.66)
 (12.58, 44.79)     a--T     (12.83, 43.83)
 (16.59, 42.77)     a--T     (12.95, 41.09)
  (1.32, 37.77)     a--T     (11.76, 41.10)
  (6.77, 43.28)     a--T     (12.90, 44.06)
 (10.52, 41.43) 10 g--C 11    (6.54, 39.18)
                    3' 5'

UH0006
                     5' 3'
   (1.93, 42.16)   1 c--G 20 (152.88, 40.45)
 (171.34, 48.66)     a--T     (83.84, 21.80)
   (0.05, 40.85)     a--T     (15.92, 39.09)
  (14.35, 38.37)     a--T     (12.53, 39.25)
 (356.95, 41.47)     g--C     (16.19, 40.16)
  (10.35, 45.17)     a--T     (10.28, 44.63)
  (16.97, 44.77)     a--T     (18.04, 39.96)
   (7.33, 37.88)     a--T     (15.09, 41.21)
   (2.49, 42.99)     a--T     (13.86, 40.86)
   (8.89, 40.82) 10 g--C 11    (9.43, 40.07)
                     3' 5'
```

Figure 9. Pseudorotational phase angle *P* and puckering amplitude ν_{max} for the RNA/DNA hybrid viral polypurine tracts in the NDB structures UH0005 and UH0006.

of the crystals grown with Mg (UH0005) and Ca (UH0006), respectively. The results are shown in Figure 9.

Most sugar moieties in both structures have north conformations with *P* values ranging from −4° to 20° and ν_{max} in the range from 37° to 46°. However, two nucleotides showed south characteristic. One is the adenine-2 ribose sugar at the 5′-end of the RNA strand, with a C2′-endo conformation ($P = 173.72°$; $\nu_{max} = 48.75°$ in UH0005 and $P = 171.34°$; $\nu_{max} = 48.66°$ in UH0006). This switch from C3′-endo to C2′-endo was speculated to be important for the rejection of the PPT by RNase H as it is not caused by crystal packing contacts.[35] The other south sugar is found in a nucleotide on the DNA strand: Guanine-20 has its 2′-deoxyribose moiety clearly in the south conformation according to the calculated *P* values of 178.32° and 152.88° in UH0005 and UH0006, respectively.

Z-DNA Enantiomers. In some special circumstances D-2′-deoxyribonucleotides can form DNA duplexes that adopt a left-handed Z-DNA conformation[36] rather than the much more common right-handed A- or B-DNA conformations. In contrast, chemically synthesized L-2′-deoxynucleotides form more readily an unusual right-handed Z-DNA. The X-ray crystal structure of the oligonucleotide d(CGCGCG)[37] contains both a left-handed Z-DNA composed of D-2′-deoxyribose units and a right-handed Z-DNA composed of L-2′-deoxyribose units. We have calculated the pseudorotational parameters for these D- and L-DNA duplexes, as shown in Figure 10. In the zigzag-shaped Z-DNA, each zigzag step has two nucleotides, C and G, respectively. The first (C) of each zigzag step in the D-DNA duplex has south conformation with *P* in the range of 150°−170°, while the second nucleotide (G) has north conformation with *P* between 18° and 40°. In the L-DNA duplex, all the torsion angles are inverted, and therefore the sugar conformations are also inverted. The first nucleotide (C) of each zigzag step has north conformation with *P* in 330°−350° range

```
ZDF040
D-DNA
                    5′ 3′
 (170.34, 30.47) 1 C--G 12 (151.96, 30.51)
  (34.04, 33.03)   G--C    (157.37, 38.84)
 (165.98, 42.11)   C--G     (19.90, 32.40)
  (36.78, 32.74)   G--C    (153.06, 44.59)
 (155.28, 36.81)   C--G     (23.74, 30.98)
  (18.04, 40.86) 6 G--C  7 (149.80, 37.50)
                    3′ 5′

L-DNA
                    5′ 3′
 (350.34, 30.47) 1 C--G 12 (331.96, 30.51)
 (214.04, 33.03)   G--C    (337.37, 38.84)
 (345.98, 42.11)   C--G    (199.90, 32.40)
 (216.78, 32.74)   G--C    (333.06, 44.59)
 (335.28, 36.81)   C--G    (203.74, 30.98)
 (198.04, 40.86) 6 G--C  7 (329.80, 37.50)
                    3′ 5′
```

Figure 10. Pseudorotational phase angle *P* and puckering amplitude ν_{max} for Z D-DNA and L-DNA helices in the NDB structure ZDF040.

```
BD0060
                    5' 3'
(168.79, 35.50)   1 C--G 24  (11.55, 40.39)
(141.84, 38.20)     G--C     (21.09, 39.87)
 (63.45, 32.94)     C--G    (147.71, 45.67)
(162.57, 32.52)     G--C    (100.11, 40.28)
(148.30, 33.47)     A--T    (154.73, 31.80)
(146.05, 41.45)     A--T*    (46.61, 44.89)
 (57.78, 41.35)     T*-A    (128.67, 42.20)
(155.71, 33.83)     T--A    (171.84, 36.76)
(150.31, 33.57)     C--G    (172.15, 35.32)
(143.29, 49.01)     G--C     (42.83, 39.43)
(156.79, 40.67)     C--G    (110.30, 34.79)
(109.97, 27.47) 12 G--C 13 (165.78, 30.09)
                    3' 5'
```

Figure 11. Pseudorotational phase angle *P* and puckering amplitude ν_{max} for the B-DNA helix containing two (L)-α-threofuranosyl nucleotides in the NDB structure BD0060. T* = (L)-α-threofuranosyl thymine.

while the second (G) has south conformation with *P* between 198° and 220°.

B-DNA Containing (L)-α-Threofuranosyl Nucleotides. (L)-α-Threofuranosyl nucleotides (TNA) containing vicinally connected (3′ → 2′) phosphodiester linkages have been shown to form base pairing in an antiparallel strand orientation and cross-pair with DNA and RNA.[38] Here we look at the structure of the B-form DNA duplex [d(CGCGAA)T*d(TCGCG)]$_2$ that has a single (L)-α-threofuranosyl thymine (T*) per strand.[39] Since the TNAs are not common nucleotides, PROSIT was modified to recognize these structures. The modification included the addition of the SMARTS strings to identify the sugar, the nucleoside, the stereochemistry of the sugar, and the 5′ and 3′ phosphorus atoms of the nucleotide. Our results, shown in Figure 11, are similar to the previously reported values of *P* and ν_{max},[39] with the largest difference being 4° for *P* and 2° for ν_{max}. The authors of the original work did not identify the method they used for the calculation of *P* and ν_{max}. Little difference in the global structure was seen between the structure of a DNA duplex containing a single TNA residue per strand and that of the reference structure consisting of normal residues. Considerable geometric changes, however, are seen at the sites of the TNA incorporation. Both threoses have C4′-exo pucker, as shown by the *P* values: 57.78° for T*7 and 46.61° for T*19. This particular sugar configuration partly compensates

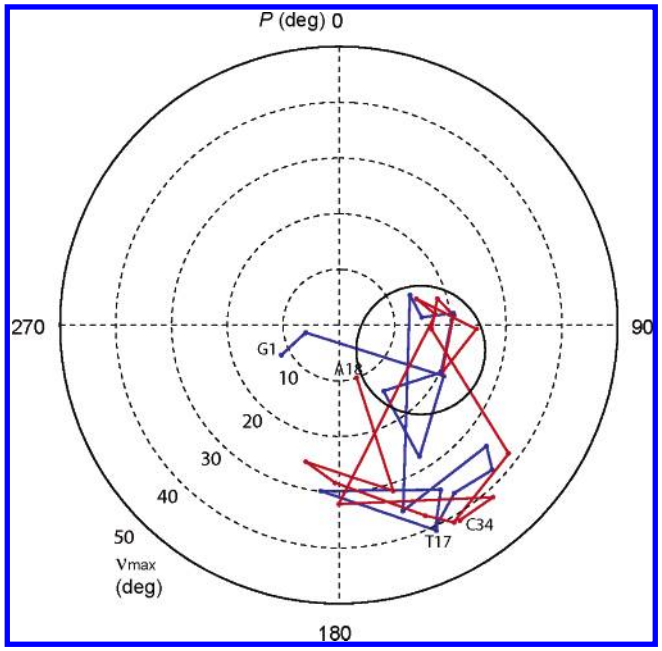


Figure 12. Pseudorotational plot of the sugar conformations of the DNA duplex bound to the TATA-box binding protein in the NDB structure PDR031. The complementary strands are colored in blue and red.

for the shortened repetitive unit arising from the five backbone bonds in TNA compared to the six backbone bonds in DNA.

Besides the above examples, we also analyzed the sugar conformations in other DNA arrangements, including a parallel-stranded guanine tetraplex[40] formed by the hexanucleotide d(TG_4T) and a four-way Holliday junction[41] formed by the DNA sequence d$(CCGGTACCGG)_2$, and found that PROSIT performed reliably and robustly with these structures, too (results not shown).

DNA-protein Complexes. PROSIT is able to identify nucleosides(tides) as individual entities, as members of molecular complexes embedded with other molecules, and as integral subunits of DNA or RNA strands in isolated polymers or polymers bound to proteins. Here we show two examples of DNA-protein complexes.

TATA-box Binding Protein (TBP)/TATA-box Complex. Although archaea have been designated as the third kingdom of life, distinct from eubacteria and eukaryotes,[42] the basal components of transcription in archaea resemble those found in the eukaryotic RNA polymerase II transcriptional system.[43] Archaeal transcription, like the eukaryotic polymerase II transcription, is initiated by an A+T-rich TATA-like segment known as the "boxA" sequence. A 2.1-Å crystal structure of the archaeal transcription complex containing TBP, transcription factor B (TFB) and a DNA fragment has been reported.[44] In Figure 12 we present in graphic form the sugar conformations of the bound DNA calculated by PROSIT for the promoter fragment [d(GC*TTTAAAAA*GTAAGTT)] to illustrate the severe distortion of the eight-base pair box A/TATA element (italics) when bound to TBP in the minor groove.[44] This distortion is clearly apparent by looking at the sugar conformation of this segment as plotted in Figure 12, where the strand beginning with G1 is colored blue and the complementary strand beginning with A18 is colored red. Most of the sixteen sugars in the boxA/TATA element are

```
PDEB08
                    5' 3'
                  1 T
 (192.47, 39.04)    G--C 26  (90.28, 66.39)
 (144.04, 40.80)    A--T    (133.75, 41.90)
 (129.64, 50.36)    T--A    (183.51, 42.77)
 (158.99, 46.54)    A--T    (152.46, 57.24)
 (173.15, 47.34)    G--C    (152.67, 38.60)
  (78.90, 62.93)    C--G    (173.10, 41.37)
 (155.14, 42.88)    G--F 20  (45.41, 41.83)
  (76.48, 57.47)    C--G    (151.08, 46.64)
 (159.28, 38.39) 10 T--A    (150.13, 42.84)
 (142.93, 28.28)    A--T    (113.57, 46.12)
 (145.93, 49.74)    T--A    (151.46, 44.95)
 (154.72, 44.98)    C--G    (184.20, 47.10)
                       T 14  (61.41, 52.43)
                    3' 5'

PD0034
                    5' 3'
                  1 T
  (69.45, 47.49)    C--G 26 (141.98, 55.39)
  (88.69, 53.86)    C--G     (48.88, 57.33)
  (98.94, 52.50)    A--T    (151.88, 44.75)
 (152.05, 44.06)    T--A    (167.36, 44.41)
 (143.18, 58.88)    G--C    (169.67, 45.31)
  (81.12, 73.87)    M--G    (171.10, 44.52)
 (157.01, 49.53)    G--D 20  (52.00, 39.65)
 (119.12, 61.14)    C--G    (185.37, 39.62)
 (147.36, 38.29) 10 T--A    (137.06, 48.81)
 (140.37, 34.03)    G--C    (122.52, 47.42)
 (134.49, 40.23)    A--T    (141.33, 40.33)
 (168.81, 44.73)    C--G      (2.16, 47.23)
                       T 14
                    3' 5'
```

Figure 13. Pseudorotational phase angle *P* and puckering amplitude ν_{max} for the B-DNA helices bound to DNA (cytosine C5)-methyltransferases in the NDB crystal structures PDEB08 and PD0034. F = 5-fluorocytosine, M = 5-methylcytosine, and D = 5,6-dihydro-5-azacytosine.

rather flat, having puckering amplitudes (ν_{max}) in the range of 10°−30°. The small thin solid circle in Figure 12 encloses 13 out of the 16 sugars in the boxA/TATA element, most of which have east conformations with *P* in the range of 70°−120°. Nearly all sugars in the surrounding segments have the usual south conformation of DNA.

DNA (Cytosine C5)-methyltransferases. The predicted mechanism of enzymatic transfer of a methyl group from *S*-adenosyl-L-methionine (AdoMet) to cytosine residues in DNA[45] involves the transient formation of a dihydrocytosine intermediate covalently linked to cysteine at the active site of a DNA (cytosine C5)-methyltransferase (C5-MTase). Such a dihydrocytosine intermediate has been confirmed by crystallographic analysis[46] of M.*HhaI* methyltransferase, AdoMet, and an oligodeoxyribonucleotide (ODN) containing 5-fluorocytosine (F13). Another ODN (DZ13) that contains the transition state analogue 5,6-dihydro-5-azacytosine also showed complete inhibition of methylation by murine DNA C5-Mtase.[47] The X-ray structure of the latter indicates that the binding interaction between the ODN and C5-MTase is similar to that in F13. Both crystal structures show that the target modified cytosine residue is flipped out of the double-stranded helix to become buried in the enzymatically active site.[46,47] Figure 13 shows the pseudorotational parameters for these two DNA helices as calculated by PROSIT. The target 5-fluorocytosine in F13 (PDEB08) has a *P* value of 45.41° and the target 5-azadihydrocytosine in DZ13 (PD0034) has a pseudorotation phase angle of 52.0°, both corresponding to a C4′-exo sugar pucker. The nucleotides surrounding the

target C20 position have the usual south sugar conformation of B-DNA. The nucleotide G8 on the opposing strand is south, while the C or M (5-methylcytosine) residues surrounding G8 have east conformations with *P* in the range of 75°−120°.

SUMMARY

A Pseudo-Rotational Online Service and Interactive Tool (PROSIT) has been developed to perform rapid and automated pseudorotational analysis for nucleosides online. PROSIT is available at http://cactus.nci.nih.gov/prosit/. The online service reads a file from the user's computer and returns the pseudorotational phase angle *P*, puckering amplitude ν_{max} and, if desired, additional torsion angles, the sugar conformation, and 2D or 3D graphic displays of the molecule. The usefulness of the tool has been demonstrated by a survey of the sugar conformations of the nucleosides found in the Cambridge Structural Database. The results confirm that most nucleoside analogues have a preference for either the north or the south conformations in the pseudorotation wheel. The conformations observed for cases falling outside the normal range seem to occur mostly when the sugar moiety is distorted through constraints imposed by additional structural elements, as crystal packing forces alone do not seem to force nucleosides to move into "unusual" regions of the pseudorotation wheel. A similar survey of high-resolution structures of DNA and RNA helices contained the Nucleic Acid Database show that most sugars in A-DNA and RNA adopt a north conformation and cluster within a small area with $0° \leq P \leq 30°$ and $30° \leq \nu_{max} \leq 50°$. On the other hand, sugars in B-DNA have mainly south conformations, albeit *P* values that can be as low as 90°. The results of the analysis of several interesting structures of specific DNA/RNA helices show that in contrast to individual nucleosides, even standard nucleotides can be forced into unusual puckering states when they are part of an oligonucleotide (DNA or RNA). This situation is most frequently encountered when the molecule's conformation is forced to change by environmental factors or protein binding.

ACKNOWLEDGMENT

The authors thank the author of the CACTVS toolkit, Wolf-Dietrich Ihlenfeldt, for assistance during the development of the PROSIT script. For contribution of I.V.F.: "This publication has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract #NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government."

Supporting Information Available: Results of quantum-mechanical calculations in the gas phase for a 2′-*O*, 4′-*C*-methylene bridged nucleoside and list of 72 NDB entries analyzed. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Marquez, V. E.; Ezzitouni, A.; Russ, P.; Siddiqui, M. A.; Ford, H.; Feldman, R. J.; Mitsuya, H.; George, C.; Barchi, J. J. HIV-1 reverse transcriptase can discriminate between two conformationally locked carbocyclic AZT triphosphate analogues. *J. Am. Chem. Soc.* **1998**, *120*, 2780−2789.

(2) Mu, L.; Sarafianos, S. G.; Nicklaus, M. C.; Russ, P.; Siddiqui, M. A.; Ford, H.; Mitsuya, H.; Le, R.; Kodama, E. et al. Interactions of conformationally biased north and south 2′-fluoro-2′,3′-dideoxynucleoside 5′-triphosphates with the active site of HIV-1 reverse transcriptase. *Biochemistry* **2000**, *39*, 11205−11215.

(3) Marquez, V. E.; Russ, P.; Alonso, R.; Siddiqui, M. A.; Hernandez, S.; George, C.; Nicklaus, M. C.; Dai, F.; Ford, H. Synthesis of conformationally restricted carbocyclic nucleosides: The role of the O(4′)-atom in the key hydration step of adenosine deaminase. *Helv. Chim. Acta* **1999**, *82*, 2119−2129.

(4) Marquez, V. E.; Ben-Kasus, T.; Barchi, J. J.; Green, K. M.; Nicklaus, M. C.; Agbaria, R. Experimental and structural evidence that herpes 1 kinase and cellular DNA polymerase(s) discriminate on the basis of sugar pucker. *J. Am. Chem. Soc.* **2004**, *126*, 543−549.

(5) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag New York Inc.: New York, 1984.

(6) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205−8212.

(7) van Wijk, J.; Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Huckriede, B. D.; Westra Hoekzema, A.; Altona, C. *Pseurot 6.3* Ver. 6.3; Leiden Institute of Chemistry: Leiden University, 1999.

(8) Altona, C. Conformational-Analysis of Nucleic-Acids − Determination of Backbone Geometry of Single-Helical Rna and DNA in Aqueous-Solution. *Recl. Trav. Chim. Pays-Bas* **1982**, *101*, 413−433.

(9) Deleeuw, F.; Altona, C. Computer-Assisted Pseudorotation Analysis of five-membered Rings by Means of Proton Spin Spin Coupling-Constants − Program Pseurot. *J. Comput. Chem.* **1983**, *4*, 428−437.

(10) Ihlenfeldt, W. D.; Voigt, J. H.; Bienfait, B.; Oellien, F.; Nicklaus, M. C. Enhanced CACTVS browser of the open NCI database. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 46−57.

(11) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380−388.

(12) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. The Nucleic-Acid Database − a Comprehensive Relational Database of 3-Dimensional Structures of Nucleic-Acids. *Biophys. J.* **1992**, *63*, 751−759.

(13) Ihlenfeldt, W. D.; Takahashi, Y.; Abe, H.; Sasaki, S. Computation and Management of Chemical-Properties in Cactvs − an Extensible Networked Approach toward Modularity and Compatibility. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 109−116.

(14) Weininger, D. Smiles, a Chemical Language and Information-System 0.1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(15) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr., Sect. B* **2002**, *58*, 389−397.

(16) Deleeuw, H. P. M.; Haasnoot, C. A. G.; Altona, C. Nucleic-Acid Constituents 0.13. Empirical Correlations between Conformational Parameters in Beta-D-Furanoside Fragments Derived from a Statistical-Survey of Crystal-Structures of Nucleic-Acid Constituents − Full Description of Nucleoside Molecular Geometries in Terms of 4 Parameters. *Isr. J. Chem.* **1980**, *20*, 108−126.

(17) Latha, Y. S.; Yathindra, N. Stereochemical Studies on Nucleic-Acid Analogues 0.1. Conformations of α-Nucleosides and α-Nucleotides − Interconversion of Sugar Puckers Via O4′-exo. *Biopolymers* **1992**, *32*, 249−269.

(18) Obika, S.; Hari, Y.; Sekiguchi, M.; Imanishi, T. A 2′,4′-bridged nucleic acid containing 2-pyridone as a nucleobase: Efficient recognition of a C·G interruption by triplex formation with a pyrimidine motif. *Angew. Chem., Int. Ed.* **2001**, *40*, 2079−2081.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; et al. *Gaussian 03*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(20) Becke, A. D. Density-Functional Thermochemistry 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(21) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* **1988**, *37*, 785−789.

(22) Vanroey, P.; Taylor, E. W.; Chu, C. K.; Schinazi, R. F. Conformational-Analysis of 2′,3′-Didehydro-2′,3′-Dideoxypyrimidine Nucleosides. *J. Am. Chem. Soc.* **1993**, *115*, 5365−5371.

(23) Balzarini, J.; Herdewijn, P.; Declercq, E. Differential Patterns of Intracellular Metabolism of 2′,3′-Didehydro-2′,3′-Dideoxythymidine and 3′-Azido-2′,3′-Dideoxythymidine, 2 Potent Anti-Human Immunodeficiency Virus Compounds. *J. Biol. Chem.* **1989**, *264*, 6127−6133.

(24) Choi, Y. S.; George, C.; Comin, M. J.; Barchi, J. J.; Kim, H. S.; Jacobson, K. A.; Balzarini, J.; Mitsuya, H.; Boyer, P. L.; et al. A conformationally locked analogue of the anti-HIV agent stavudine. An important correlation between pseudorotation and maximum amplitude. *J. Med. Chem.* **2003**, *46*, 3292−3299.

(25) Exall, A. M.; Jones, M. F.; Mo, C. L.; Myers, P. L.; Paternoster, I. L.; Singh, H.; Storer, R.; Weingarten, G. G.; Williamson, C.; et al. Synthesis from (−)-Aristeromycin and X-ray Structure of (−)-Carbovir. *J. Chem. Soc., Perkin Trans. 1* **1991**, 2467−2477.

(26) Choi, Y. S.; Sun, G. Y.; George, C.; Nicklaus, M. C.; Kelley, J. A.; Marquez, V. E. Synthesis and conformational analysis of a locked analogue of carbovir built on a bicyclo 3.1.0 hex-2-enyl template. *Nucleosides Nucleotides Nucleic Acids* **2003**, *22*, 2077−2091.

(27) Dorr, D. R. Q.; Vince, R. Synthesis and biological evaluation of endocyclic 2′,3′-didehydro-2′,3′-dideoxymethanocarba adenosine. *Nucleosides Nucleotides Nucleic Acids* **2002**, *21*, 665−680.

(28) Yamaguchi, K.; Hayakawa, H.; Ashizawa, H.; Tanaka, H.; Miyasaka, T. The Structure of Cyclonebularine Derivatives. *Acta Crystallogr., Sect. C* **1990**, *46*, 1065−1068.

(29) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA Dodecamer − Conformation and Dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179−2183.

(30) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Crystal-Structure Analysis of a Complete Turn of B-DNA. *Nature* **1980**, *287*, 755−758.

(31) Shui, X. Q.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* **1998**, *37*, 8341−8355.

(32) Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. Crystal structures of B-DNA with incorporated 2′-deoxy-2′-fluoro-arabinofuranosyl thymines: implications of conformational preorganization for duplex stability. *Nucleic Acids Res.* **1998**, *26*, 2473−2480.

(33) Barchi, J. J.; Jeong, L. S.; Siddiqui, M. A.; Marquez, V. E. Conformational analysis of the complete series of 2′ and 3′ monofluorinated dideoxyuridines. *J. Biochem. Biophys. Methods* **1997**, *34*, 11−29.

(34) Sarafianos, S. G.; Das, K.; Tantillo, C.; Clark, A. D.; Ding, J.; Whitcomb, J. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **2001**, *20*, 1449−1461.

(35) Kopka, M. L.; Lavelle, L.; Han, G. W.; Ng, H. L.; Dickerson, R. E. An unusual sugar conformation in the structure of an RNA/DNA decamer of the polypurine tract may affect recognition by RNase H. *J. Mol. Biol.* **2003**, *334*, 653−665.

(36) Wang, A. H. J.; Quigley, G. J.; Kolpak, F. J.; Vandermarel, G.; Vanboom, J. H.; Rich, A. Left-Handed Double Helical DNA − Variations in the Backbone Conformation. *Science* **1981**, *211*, 171−176.

(37) Doi, M.; Inoue, M.; Tomoo, K.; Ishida, T.; Ueda, Y.; Akagi, M.; Urata, H. Structural Characteristics of Enantiomorphic DNA − Crystal Analysis of Racemates of the d(CGCGCG) Duplex. *J. Am. Chem. Soc.* **1993**, *115*, 10432−10433.

(38) Schoning, K. U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. Chemical etiology of nucleic acid structure: The α-threofuranosyl-(3′ → 2′) oligonucleotide system. *Science* **2000**, *290*, 1347−1351.

(39) Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. Crystal structure of a B-form DNA duplex containing (L)-α-threofuranosyl (3′ → 2′) nucleosides: A four-carbon sugar is easily accommodated into the backbone of DNA. *J. Am. Chem. Soc.* **2002**, *124*, 13716−13721.

(40) Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 angstrom resolution. *J. Mol. Biol.* **1997**, *273*, 171−182.

(41) Thorpe, J. H.; Teixeira, S. C. M.; Gale, B. C.; Cardin, C. J. Structural characterization of a new crystal form of the four- way Holliday junction formed by the DNA sequence d(CCGGTACCGG)$_2$: sequence versus lattice? *Acta Crystallogr., Sect. D* **2002**, *58*, 567−569.

(42) Bult, C. J.; White, O.; Olsen, G. J.; Zhou, L. X.; Fleischmann, R. D.; Sutton, G. G.; Blake, J. A.; FitzGerald, L. M.; Clayton, R. A.; et al. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **1996**, *273*, 1058−1073.

(43) Langer, D.; Hain, J.; Thuriaux, P.; Zillig, W. Transcription in Archaea − Similarity to That in Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5768−5772.

(44) Kosa, P. F.; Ghosh, G.; DeDecker, B. S.; Sigler, P. B. The 2.1-angstrom crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6042−6047.

(45) Santi, D. V.; Norment, A.; Garrett, C. E. Covalent Bond Formation between a DNA-Cytosine Methyltransferase and DNA Containing 5-Azacytosine. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6993−6997.

(46) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. D. Hhal Methyltransferase Flips Its Target Base out of the DNA Helix. *Cell* **1994**, *76*, 357−369.

(47) Sheikhnejad, G.; Brank, A.; Christman, J. K.; Goddard, A.; Alvarez, E.; Ford, H.; Marquez, V. E.; Marasco, C. J.; Sufrin, J. R.; et al. Mechanism of inhibition of DNA (cytosine C5)-methyltransferases by oligodeoxyribonucleotides containing 5,6-dihydro-5-azacytosine. *J. Mol. Biol.* **1999**, *285*, 2021−2034.

CI049881+