Tetrahedron Vol. 49, No 5, pp 1043-1052, 1993
Printed in Great Britain

0040-4020/93 $6.00+.00
Pergamon Press Ltd

DETERMINATION AND PREDICTION OF THE ABSOLUTE CONFIGURATION OF DINUCLEOSIDE ALKYLPHOSPHONATES USING CONFORMATIONAL ANALYSIS AND MULTIVARIATE STATISTICS

A.V. Lebedev, A. Frauendorf*, E.V. Vyazovkina and J.W. Engels*

Institute of Bioorganic Chemistry, Siberian Division of the Russian Academy of Sciences,
Novosibirsk-90, 630090, Russia
* Institut für Organische Chemie der J.W.Goethe Universität,
W-6000 Frankfurt am Main 50, Germany

(*Received in Germany* 21 *August* 1992)

Abstract
A new method is presented to predict with high probability the stereochemistry of any synthesized dinucleoside alkylphosphonate on a basis set of at least five dimers (Rp and Sp). This method was developed and tested with all 16 nucleoside methylphosphonate dimers. All methylphosphonate dimers with the naturally occuring bases Ade, Thy, Gua and Cyt were synthesized and analyzed using UV, CD and NMR spectroscopy. From these experimental data parameters could be extracted for use as variables in the subsequent multivariate analysis. Analysis of variance showed which variables could be used to distinguish between the two diastereomers. These selected variables were used as input variables for factor analysis. After varimax rotation two factors were obtained representing the ROE's from the P-Me group to the 5'- and the 3'-nucleoside. These factors were then used as predictor variables in the following discriminant analysis. By that procedure a stereochemical predicition model was developed in which all 32 diastereomers were assigned unambiguously.

INTRODUCTION

Oligonucleoside alkylphosphonates are very attractive nucleotide analogues[1,2] for use as antisense oligonucleotides. In these compounds the negatively charged phosphate oxygen is replaced by an alkyl group (Fig.1), this facilitates their penetration into cells[3]. Their unnatural linkage is also resistant to nucleases[4]. Furthermore, based on electrostatic arguments alkylphosphonates with defined stereochemistry should hybridize stronger with complementary DNA or RNA[5]. On the other hand the occurrence of either Rp or Sp configuration on each chiral phosphorus center is a severe synthetical problem and there is no generally applicable method known to produce oligonucleotide alkylphosphonates with defined stereochemistry with the exception of some nucleoside methylphosphonate dimers[6,7] although there is considerable evidence to support the notion that the stereochemistry about the phosphorus atom affects the ability of the sequence to bind to its target[8].

X-ray analysis so far has been the only reliable and unambiguous method of proving the absolute configuration. But only two dimers could be crystallized and assigned so far[9,10]. Moreover there was some confusion concerning the correlation of the X-ray data and the CD and HPLC data[11].

The stereoregular synthesis of modified oligonucleotides can be achieved using dimer building block synthesis where alkylphosphonate dimers have to be coupled[12]. The stereochemically defined building blocks are synthesized either by stereoselective reactions or are separated by HPLC before use. This approach results in modified oligonucleotides with alternating stereoregular alkylphosphonate and phosphodiester bonds.

At this point the determination of absolute configuration at the chiral phosphorus center of the diastereomeric alkylphosphonate dimer building blocks is a prerequisite. Because there are four bases used to design dimers, synthesis results in 16 dimers times two diastereomers (32 compounds). Methods for a fast

structure determination of new alkylphosphonate derivatives are needed without synthezising all 32 possible combinations of dimers.

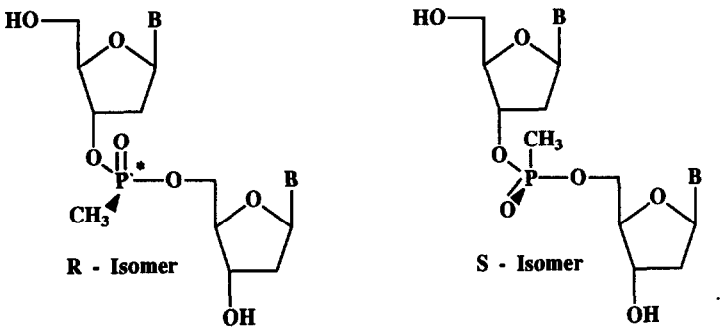

Fig.1 Structure of nucleoside dimers, the * indicates P-chirality

The aim of the present work is to elucidate possible variables and methods that can be used in the prediction of configuration of any synthesized dimer using a small basis set of structurally determined dimers. The validity and reliability of these methods were tested in the case of methylphosphonate dimers. Therefore all 16 deoxyribonucleoside methylphosphonate dimers were synthesized and the diastereomers were separated by silica gel chromatography resulting in 32 compounds. From all dimers UV, CD and ROESY spectra were recorded and analyzed in order to find parameters for further statistical analysis.

EXPERIMENTAL

Synthesis and separation of diastereomers

Deoxynucleoside methylphosphonates were synthesized according to the procedure proposed by Miller[13]. All dimers were purified and separated into individual diastereomers by silica gel chromatography with $CHCl_3$/MeOH or ($CHCl_3$/ethyl acetate)/MeOH gradients. Deprotection of nucleoside base protecting groups and 3′-acetyl group were performed with an anhydrous ethylenediamine/ethanol mixture for 2h at 37 °C . Detritylation was accomplished with 80% acetic acid in the case of 5′-DMT protecting group and 5% CF_3COOH in the case of 5′-MMT protecting group. Final purification of the detritylated dimers was achieved with RP-C18 chromatography using a 0-40% H_2O/CH_3CN solvent gradient[19].

UV, CD and NMR measurements

CD spectra of the fully deprotected dimers were recorded with a JASCO J-600 spectropolarimeter in 1M NaCl, 0.01M NaH_2PO_4, 0.06 mM EDTA (pH 7.0) at 15 °C. UV Hyperchromicity percentage was calculated after complete hydrolysis of the unprotected diastereomer in 1M KOH at 20 °C. ROESY spectra of the fully protected dimers were recorded with Bruker AMX-400, AM 400 and AM 250 spectrometers. Measurements were done at 27 °C after an extensive purge of the sample with N_2. Sam-

ples were prepared as 100 mM solutions in CD_3CN (30 mg of sample in 450 µl). The spectra were calibrated using the internal CH_3CN signal at 1.93 ppm. The details of ROESY experimental parameters were similar to those already described[14]. Normally 512 experiments each of 32 transients were recorded with a relaxation delay of 3 sec. The spectra were presented in phase sensitive mode with baseline correction. The assignment of sugar, base and protecting group signals and their corresponding cross-peaks of the ROESY spectra were made for all dimers. The individual 5′,5′′-protons of both sugars were not assigned.

RESULTS AND DISCUSSION

Conformational analysis

The conformations of all dimers were studied by UV-, CD- and NMR spectroscopy. UV- and CD spectra are very sensitive to base stacking effects. The hyperchromicity measurements suggest that the bases in all dimers are stacked depending on base composition and stereochemistry (Fig.2). So hyperchromicity is increased following the increase of base stacking (purine-purine > purine-pyrimidine > pyrimidine-pyrimidine). This effect is more pronounced for the Rp configuration than for the Sp configuration. Going from the Rp to the Sp series the hyperchromicity effect is drastically reduced in most cases.

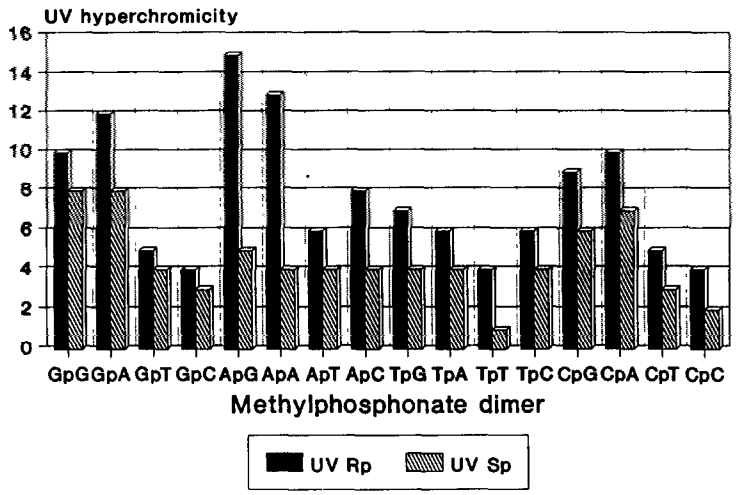

Fig.2 Percent hyperchromicity of all 32 nucleoside methylphosphonate dimers after complete hydrolysis of the unprotected diastereomers in 1M KOH at 20 °C

Compared to UV spectra the magnitude of the molecular ellipticity in dimers is furthermore sensitive to the angle between the transition dipoles of the bases.The value of the molecular ellipticity is greatest for an angle of 45° it diminishes if the angle is 0, 90 or 180°. For that reason CD spectra are not so easily correlated to base composition and stereochemistry of the dimers as UV spectra.

Differences in the extent and mode of base stacking interactions are observed for the individual diastereomers. The profiles of CD spectra for pairs of dimers are qualitatively similar suggesting a B-like conformation. They show a positive band near 280 nm and a negative band near 250 nm. However, the magnitude of molecular ellipticity is higher for the first eluting isomer (Rp) than for the second eluting isomer (Sp) in all cases except Gp(Me)A (Fig.3).

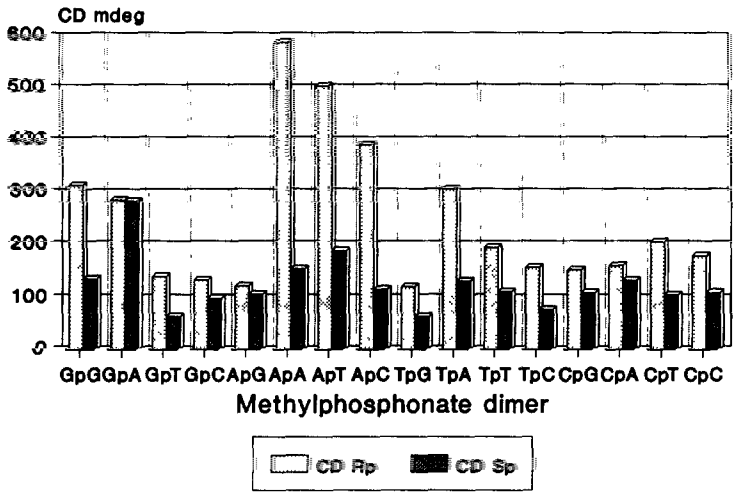

Fig.3 Molecular ellipticity difference of the 280 nm and 250 nm band of the methylphosphanate dimers measured in 1M NaCl, 0.01M NaH_2PO_4, 0.06M EDTA,pH 7.0 at 15 °C

Looking at the NMR shift data, differences were detected for the proton chemical shifts of the P-CH_3 group. In most cases the signal of the first eluting isomer (Rp) was shifted towards higher field compared to the second eluting isomer[20] (data not shown).

For NMR conformational analysis interproton distances were chosen (ROE's). Although analysis can also be done using coupling constants for the determination of torsion angles it should be emphasized that the torsion angle is not the usual Euclidian-type variate because of having both directionality and periodicity. Interatomic distances as molecular descriptors do not bear these complications.

The ROESY data show many interactions[20], the P-CH_3 group interacts with protons of both nucleosides and with the DMT protons (Fig.4).

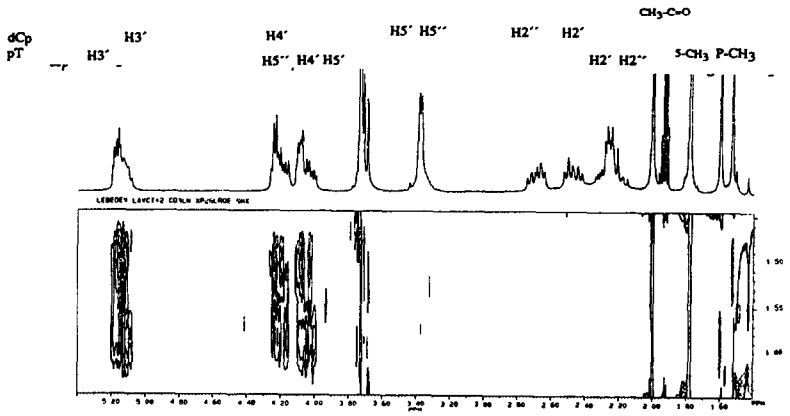

Fig.4 Typical ROESY spectrum of individual diastereomer showing the region of interacions of the P-CH_3 group with deoxyribose protons. Spectrum of Sp diastereomer of (DMT)bzdCp(CH_3)dT(Ac) is presented

It was shown by the work of Bower et al.[15] that in an octamer duplex with one methylphosphonate linkage Rp and Sp configurated methylphosphonates could be distinguished by the NOE of the P-Me group to the 3′-proton of the 5′-nucleoside. There in all cases examined the Sp-duplexes gave an NOE crosspeak between the P-CH_3 and the H3′of the n-1 nucleoside because it is the Sp isomer which has the P-CH_3 bond pointing toward the major groove with the methyl protons in the vicinity of the H3′. This crosspeak was absent in the Rp isomer.

Interestingly Loeschner et al.[14] found that in the case of dimers that crosspeak is present in the Rp as well as in the Sp isomer. Here the crosspeak from the P-CH_3 to the H4′ (n-1) was used as a criterion to distinguish between the diastereomers. It was found that this cross-peak was more intense in the Rp than in the Sp isomer. On the other hand this crosspeak was not detected in the octamer duplex. We found now that even the criterion proposed by Loeschner et al. does not hold in all 32 cases of methylphosphonate dimers.

The reason for this partial failure is that duplexes and dimers are quite different in their conformational behavior. In the case of the duplex the methylphosphonate linkage is forced into a helical conformation with nearly no freedom regardless whether the phosphorus is Rp or Sp configurated.

Methylphosphonate dimers in contrast to oligomer duplexes show a high degree of conformational flexibility. So it is clear that the interatomic distances determined from the ROESY data do not represent distances in one defined conformation but an average distance of many conformational populations in a

rapid equilibrium. This is a difficult situation for the complete analysis of conformational states. Therefore the diastereomers cannot be analyzed due to the existence of different single conformations. Here we favor the concept of two different conformational ensembles, which exist for the two sets of diastereomers. These conformational ensembles cannot be easily predicted and have to be treated statistically. From the above considerations it becomes clear that the high and different conformational flexibility of the dimers explains why no simple criterion could be detected for all possible dimers although fully protected dimers were analyzed by ROE in this study because they were expected to possess a more rigid conformation in organic solvents.

Stereochemical modelling indicated that the interaction of P-CH_3 to H(3´) of the 5´-residue is possible for both diastereomers. Therefore in the case of dimers this interaction cannot serve for the assignment of absolute configuration in any case. In all ROE spectra the interactions of the P-CH_3 to the DMT protons are present indicating relatively short distances. Therefore any changes in the rotational freedom of the DMT group around the C(4´)-C(5´)-bond should influence the rotational freedom around the P-O(3´) bond and furthermore the position of the phosphorus methyl group relative to the 4´and 5´-protons. All data obtained, indicate that the conformation of a diastereomer and consequently of a set of ROE-interactions depends strongly on conformational differences and is dependent on the type of heterocyclic bases in the dimers, be it pyrimidine or purine.

Thus the ROE´s of the P-CH3 groups to the 3´, 4´ and 5´-protons of both nucleosides as well as the CD and UV data were analyzed together in order to find variables for a discrimination of configuration. This was done by Two-Way-Analysis of Variance (ANOVA) using a fixed-effects model.The program SPSS/PC+ was run for all calculations. This method allows the investigation of the effects of the two factors (independent variables) stereochemistry and base composition on a specific ROE, the CD ellipticity or any other dependent variable. In the analysis of variance the observed variability in the sample is subdivided into two components, variability of the observations within a group about the group mean and variability of the group means:

$$Y_{ivz} = m + a_v + b_z + (ab)_{vz} + e_{ivz}$$

where Y is the observed value, m the total mean of the population, a the effect of the first factor (stereochemistry), b the effect of second factor (base composition), (ab) is an interaction term and e the not explained residual. The significance of the results is determined by the F-test. To determine the impor-tance of each factor a Multiple Classification Analysis (MCA) was done. Several measures of association are displayed in Table I. The squared value of h indicates the proportion of variance explained by a given factor. The multiple r value indicated the proportion of variance in the dependent variable accounted for by all factors and interaction terms.

In table I it can be seen that only the UV hyperchromicity shows a significant dependence on both base composition and stereochemistry (ID) where the dependence on base stacking effects ($\eta = 0.65$) is more pronounced than the dependence on stereochemistry ($\eta = 0.52$). The ellipticity of the CD spectra shows a clear dependence on stereochemistry and only a slight not significant dependence on base stacking ef-

fects. That is the case because the CD ellipticity is also dependent on the angle between the transition dipoles of the bases, not covered in the above model.

Tab I. Multiple Classification Analysis (MCA) :

Var	F(ID)	Sign.	F(Base)	Sign.	η(ID)	η(Base)	ρ(mult)
UV	30.7	0.000	24.0	0.000	0.52	0.65	0.84
CD	10.8	0.003	2.16	0.135	0.51	0.32	0.60
HPLC	2.03	0.166	0.75	0.481	n.d.	n.d.	n.d.
δ(P-Me)	1.89	0.180	0.87	0.432	n.d.	n.d.	n.d.
ROE 3′-Np	2.60	0.119	0.00	1.000	n.d.	n.d.	n.d.
ROE 4′-Np	27.1	0.000	0.67	0.519	0.69	0.15	0.71
ROE 5′-Np	13.0	0.001	0.26	0.773	0.56	0.11	0.57
ROE 3′-pN	15.0	0.001	0.69	0.511	0.54	0.16	0.57
ROE 4′-pN	36.1	0.000	2.17	0.135	0.69	0.24	0.73
ROE 5′-pN	0.57	0.903	0.00	1.000	n.d.	n.d.	n.d.

ID= stereochemistry of the dimer, Sign. = significance level reached (All values greater 0 05 are not significant), η = variance of factor ı/total variance
n d. = not determined because values are not significant, F = value due to F-statistical test, Base = base composition

The HPLC retention time is in our understanding not simply related to the stereochemistry or base composition of the dimer. No significant effect is observed although the order of elution in each dimer pair is always the same. So the factors determining HPLC elution time remain obscure in this case.

All ROE′s show no significant effect due to base stacking but the ROE′s of the P-Me groups to H(4′) and H(5′) of the 5′-nucleoside and the ROE′s to the H(3′) and H(4′) of the 3′-nucleoside are significantly dependent on stereochemistry. The ROE between P-Me and H(4′) of the 5′-nucleoside seems to be the most reliable single variable in configurational assignment. This is in good agreement with our earlier work. As a result, four distances were selected and submitted to further analysis.

Factor analysis

To verify whether all these distances are required for a complete description of the dimer system a principle component factor analysis (PCA) was performed to determine the independent variables. Factor analysis is an eigenvalues-eigenvectors method used to find a standardized linear combination of the original variables[16]. This method was already used for the analysis and interpretation of CD spectra of nucleotide dimers[17,18]. The new variables called principal components are orthogonal and oriented, in decreasing order, in the direction of the maximum variability of the data set. The aim of factor analysis is to reduce the number of variables required to describe the covariance relationship of many variables in the system without significant loss of information.

The observable random vector **X** (containing the four selected ROE random variables X_i) has the mean **m** and the covariance matrix **S**. The factor model postulates that **X** is linearily in-dependent upon a few (let´s assume two) unobservable random variables F_1, F_2 called common factors, and four additional sources of variation e called errors. So we get:

$X_1 = m_1 + l_{11}F_1 + l_{12}F_2 + e_1$ (4´-Np)
$X_2 = m_2 + l_{21}F_1 + l_{22}F_2 + e_2$ (5´-Np)
$X_3 = m_3 + l_{31}F_1 + l_{32}F_2 + e_3$ (3´-pN)
$X_4 = m_4 + l_{41}F_1 + l_{42}F_2 + e_4$ (4´-pN) or $\mathbf{X} = \mathbf{m} + \mathbf{L}\mathbf{F} + \mathbf{e}$.

The coefficient l_{ij} is called the loading of the i-th variable on the j-th factor.
Factor analysis usually proceeds in four steps. First the correlation matrix for all variables is computed. In the second step, factor extraction - the number of factors necessary to represent the data and the method of calculating them - must be determined. Here a principal component factor analysis was performed. Due to the Kaiser criterion not to extract factors with eigenvalues less then one, two factors were extracted. These two factors explain 78 % of total variance. Although the factor matrix obtained in the extraction phase indicates the relation-ship between the factors and the individual variables it is usually difficult to identify meaningful factors based on this matrix. So the following rotation phase of the factor analysis attemps to transform the initial matrix into one that is easier to interpret. A variety of algorithms are used for orthogonal rotation to a simple structure. The most commonly used method is the varimax method, which attempts to minimize the number of variables that have high loadings on a factor. The result of a varimax rotation of the two factors is shown in Table II.

Table II: Factor Loadings of the Two Extracted Factors

Variables	Factor 1	Factor 2
Cum. % Variance	52.6	77.5
Var1 ROE (H4´-Np)	0.911	-0.050
Var2 ROE (H5´-Np)	0.830	-0.304
Var3 ROE (H3´-pN)	-0.183	0.851
Var4 ROE (H4´-pN)	-0.120	0.846

As it can be easily seen in Table II , the first factor has high loadings for the ROE´s associated with the P-Me distances to the 5´-nucleoside while the second factor has high loadings for the ROE´s associated with the P-Me distance to the 3´-nucleoside. The two factors represent the ROE´s of the phosphorus methyl group to both nucleosides and were used as new variables in further analysis.

Discriminant analysis

The aim of the work presented here is to test, if the stereochemistry at the phosphorus of a synthesized dimer could be predicted if a small basis set of dimers with determined stereochemistry exists. Therefore in the first phase of analysis all available data, the dimers assigned earlier[14] as well as the two crystal structures, will be used to evaluate a model for confirming the preassignment of configuration.

In discriminant analysis, linear combinations of the two predictor variables (factor1 and factor2) are formed and serve as the basis for classifying dimers into one of the groups (Rp or Sp). In that sense the linear discriminant equation is similar to the multiple linear regression equation[16]:

$$D = B_0 + B_1 F_1 + B_2 F_2 .$$

The coefficients Bi are chosen so that the values of the discriminant function differ as much as possible between the groups.The discriminant scores for each dimer can be seen in Fig.5.

Using the two factor variables and the preassigned dimers $Ap_{Me}T$, $Tp_{Me}T$, $Ap_{Me}A$, $Tp_{Me}A$ and $Cp_{Me}G$ a perfect fit between the preassigned groups and the groups obtained by discriminant analysis was achieved. This proves that the assignment must have been correct. In combination with the known configuration of the second eluting $Ap_{Me}T$ isomer determined by X-ray[9] all configurations can now be determined unambiguously.

At last it was investigated if it is possible to predict the stereochemistry of any given nucleoside methylphosphonate dimer with a small basis set of structurally known compounds used for the estimation of the discriminatory function. For statistical reasons the number of cases used should be at least three to four times greater than the number of variables. As it can be seen from Figure 5 the diastereomers could be clearly separated due to discriminant analysis. All compounds with R configurated phosphorus show positive values of the discriminant function while the Sp configurated compounds show negative values.

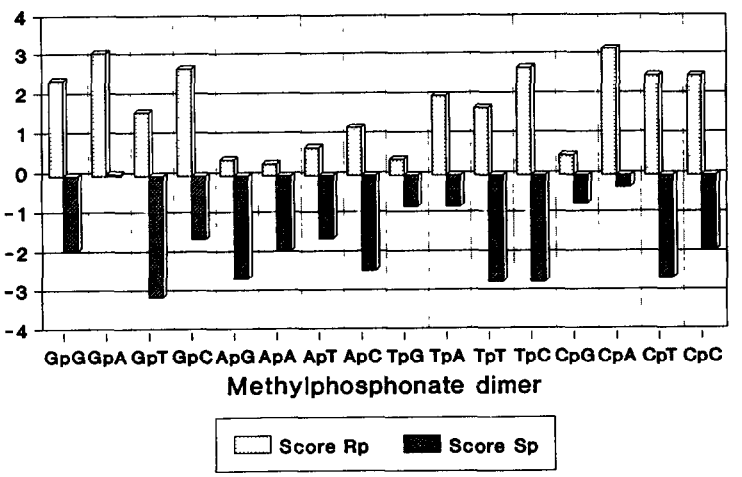

Fig.5 Discriminant scores of all methylphosphonate dimers according to the discriminant function D = -0.397 factor1 + 0.463 factor2

One might wonder why the classification is much worse using CD or UV data although a clear differentiation between the Rp and the Sp diasteromers is always possible (Fig.3). This is due to the fact that the CD and UV data depend much more on other contaminating factors than the ROE data do. In conclusion a direct comparison between two diastereomers of the same base composition is always successful with UV, CD or HPLC methods but an assignment of the stereochemistry of any given dimer is not possible without a comparison with the other stereoisomer.

We have shown here that in the case of highly flexible molecules an unambiguous assignment of configuration on the basis of only a few compounds in combination with simple rules is problematic and will not hold in all cases. So we propose a statistical method for configurational assignment of alkylphosphonate dimers on the basis of ROESY NMR data in combination with multivariate statistical procedures. But there is still some work to do to explain the different ROESY patterns of the Rp and the Sp isomers on a real physical and not only statistical basis. In order to successfully achieve this a full conformational determi-nation of dimers and oligonucleotides has to be derived at.

Acknowledgement - We wish to thank Dr. D. Johnson for his agreement to use the SPSS/PC+ program, Prof. Dr. C. Griesinger for the provision of all NMR pulse programs and Dr. G. Zimmermann for his help in all problems concerning NMR spectra.

REFERENCES

1. Miller, P.S. (1991) Biotechnology, **9**, 358-362.
2. Uhlmann, E., Peyman, A. (1990) Chem. Rev. **90**, 544-584.
3. Agris, C.H., Black, K.R., Miller, P.S., Reddy, M.P., Ts´o, P.O.P., (1986) Biochemistry, **25**, 6268-6275.
4. Smith, C.C., Aurelian, L., Reddy, M.P., Miller, P.S., Ts´o, P.O.P. (1986) Proc. Nat. Acad. Sci. USA, **83**, 2787-2791.
5. Lesnikowsky, Z.J., Jaworska, M., Stec, W.J. (1988) Nucl.Acid Res. **16**, 11675-11689.
6. Engels, J., Jaeger, A. (1982) Angew.Chem.Int.Ed.Engl. **21**, 2010-2015.
7. Cormier, J.F., Pannunzio,T. (1991) Tetrahedron Lett., **32**, 7161-7164.
8. Lesnikowsky, Z.J., Jaworska, M., Stec, W.J. (1990) Nucl.Acid Res., **18**, 2109-2115.
9. Chacko, K.K., Lindner, K., Saenger, W. (1983) Nucl.Acid Res. **11**, 2801 -2814.
10. Han, F., Watt, W.,Callahan, L., Kezdy, F.J., Agarwal, K. (1990) Nucl.Acid Res., **18**, 2759-2767.
11. Callahan, L., Han, F.S., Watt, W., Duchamp, D., Kezdy, F.J., Agarwal, K. (1986) Proc.Nat.Acad.Sci. USA 83, 1617-1621.
12. Herdering, W., Kehne, A., Seela, F. (1985) Helv.Chim.Acta, 68, 2119-2127.
13. Miller, P.S., Reddy, M.P., Murakami, A., Blake, K.R., Lin, S., Agris, C.H. (1986) Biochemistry, **25**, 5092-5097.
14. Löschner, T., Engels, J.W. (1990) Nucl.Acid Res., **18**, 5083-5088.
15. Bower, M., Summers,MF, Regan,,JB, Zon,G, Wilson,WD, (1987) Nucl. Acid Res., **15**, 4915-4930.
16. Johnson, R.A., Wichern, D.W. (1988) Applied Multivariate Statistical Analysis, Prentice Hall Int.Ed. 378.
17. Amaratunga, M., Pancoska, P., Paner, T.M., Benight, A.S. (1990) Nucl.Acid Res., **18**, 577-582.
18. Olsthoorn, C.S., Haasnoot, C.A., Altona, C. (1980) Eur.J.Biochem., **106**, 85-89.
19. Vyazovkina,EV; Lebedev,AV; Komarova,NI (1993) Bioorganicheskaya Khimia, **1**, in press.
20. Vyazovkina,EV, Engels,JW; Lebedev,AV (1993) Bioorganicheskaya Khimia, in press.