Journal of Pharmaceutical & Biomedical Analysis
Vol. 9, Nos 10–12, pp. 1125–1131, 1991
Printed in Great Britain

0731–7085/91 $3.00 + 0.00
Pergamon Press plc

Relationships between nucleotide incorporation rates and molecular parameters obtained by molecular modelling and chromatography*

KLÁRA VALKÓ†‡ and PÉTER SLÉGEL§

‡ *Central Research Institute for Chemistry, Hungarian Academy of Sciences, P.O. Box 17, H-1525, Budapest, Hungary*
§ *Department of Pharmaceutical Chemistry, School of Pharmacy, University of London, London WC1N 1AX, UK*

Abstract: Deoxyuridine derivatives play an important role in pharmaceutical chemistry as they are potential antiviral and antitumour agents. Their pharmacological activity depends on their ability to incorporate into DNA in their triphosphate forms.

High-performance liquid chromatographic (HPLC) retention behaviour of a series of 5-alkyl, alkenyl and alkynyl substituted deoxyuridine derivatives were investigated on reversed-phase stationary phase using various mixtures of methanol and water as mobile phases. The slope and the intercept values of the linear relationships between the logarithmic capacity ratio (log k') values and methanol concentration have been calculated.

Non-polar, non-polar unsaturated and polar surface areas, surface energies, dipole moments, van der Waals radii of the derivatives have been calculated on the bases of molecular mechanics by PC Model approach.

The correlation study of the above-mentioned parameters revealed that hydrophobic and hydrophilic surface areas and the electronic effects of the substituents determine not only the retention behaviour of the derivatives but also their incorporation rate into DNA in their triphosphate forms.

Keywords: *Deoxyuridines; liquid chromatography; molecular modelling; polar and non-polar surface; dipole moment.*

Introduction

A great number of nucleoside derivatives have already been synthesized and tested as potential antiviral and antitumour agents, as reviewed by De Clercq [1]. Their activity can be related to their ability to incorporate into DNA or to influence the biosynthesis of DNA.

A series of 5-substituted 2′-deoxyuridine-5′-triphosphates have been synthesized and their initial incorporation rates into DNA have been measured. A nonlinear relationship between the chain length of the 5-substituent and the incorporation rate was observed [2–4]. The effect of chain branching was more significant when the synthetic poly(dA-dT) template was used in DNA replication reactions catalysed by DNA polymerase enzyme [4].

The chromatographic retention data of the corresponding nucleoside derivatives have already been published not only in reversed-phase systems but also obtained on silica stationary phase by using methanol–ethyl acetate mixtures as mobile phases [5–6]. For the characterization of steric properties of the 5-substituents the inclusion complex stability data of the derivatives with β-cyclodextrins were measured by using a thin-layer chromatographic method [6–7].

The previous studies [5–6] revealed the importance of the electronic, steric and hydrophobic properties measured by reversed-phase chromatography of the compounds or expressed by the octanol–water partition data in the DNA polymerase reactions. The electronic properties of the substituents (their inductive effects) were characterized by the Swain–Lupton [8] parameter (F), while the steric properties were characterized by the molar refractivity (MR) values compiled by Hansch and Leo [8]. The influence of the size and measure of the 5-substituent was more significant when the poly(dA-dT) template was used instead of the random base sequence template (calf thymus DNA).

In this study molecular mechanics [9] calculations have been carried out to determine the three-dimensional structures of the deoxyuridine derivatives and to calculate the non-polar, non-polar unsaturated and polar surface

* Presented at the "Third International Symposium on Pharmaceutical and Biomedical Analysis", April 1991, Boston, MA, USA.

† Author to whom correspondence should be addressed.

areas and energies, the van der Waals radii and dipole moments. The parameters obtained by molecular modelling were correlated to the chromatographic retention data and to the incorporation rate values in order to obtain a better understanding of the behaviour of the compounds in both the chromatographic and the enzyme reaction systems.

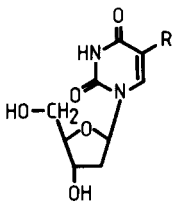

Figure 1
Structure of the deoxyuridine derivatives investigated.

Experimental

The structures of the investigated compounds are shown in Fig. 1. The syntheses of the compounds have been described elsewhere [2–4, 10]. The compounds were chromatographically pure. Conditions for the measurements of the relative initial incorporation rates of the derivatives into calf thymus DNA (RATE%) and poly(dA-dT) (DADT%) have already been published [2–4]. Table 1 shows the structure of the investigated compounds and their measured incorporation rate values.

For the measurements of the reversed-phase retention data an RP-8 (5 μm) column (150 × 4.6 mm i.d.) (Perkin–Elmer, Norwalk, CT, USA) was used with methanol–water mixtures as the mobile phase. The methanol concentration ranged from 40 to 60% (v/v) in 5% steps. A flow-rate of 1.00 ml min^{-1} was applied. Peaks were detected at 260 nm. The dead time was determined by injection of 1% sodium nitrate solution. The logarithmic values of the capacity factors (log k') of the compounds were plotted against the methanol concentration in the mobile phase. For each compound a straight line could be fitted to at least four data points. The correlation coefficients were always higher than 0.97. The slope (RPS) and the intercept (RPI) of the straight line were calculated and listed in Table 1.

The measurements of the inclusion complex stability data with β-cyclodextrin are described elsewhere [6] and the stability constants are listed also in Table 1.

PC Model approach [9] was used to determine the three-dimensional structure of compounds based on energy minimization. After setting up the geometrics of the molecules having the smallest mmx-energy, the non-polar (nopol), non-polar unsaturated (nupol) and polar surface (polsa) areas have been calculated and expressed in $Å^2$. The water solvation shell was considered also in the calculations of

Table 1
Investigated deoxyuridine derivatives, their initial incorporation rate into calf thymus DNA (RATE%) and into poly(dA-dT) (dAdT%) in their triphosphate forms, and their reversed-phase retention data [the slope (RPS) and the intercept (RPI) values of the straight line obtained by plotting the log k' vs methanol concentration, respectively], and their inclusion complex stability data (K) with β-cyclodextrin

No.	R	RATE%	dAdT%	RPI	RPS $\times 10^{-2}$	K
1	H	41	97	0.424	−2.65	0
2	Ethyl	18	59	0.590	−2.39	0
3	Isopropyl	6	1	0.724	−2.39	0
4	Sec-Butyl	8	21	1.299	−2.82	0
5	Tert-Butyl	6	3	1.301	−2.80	1.03 ± 0.31
6	Pentyl	6	19	2.041	−3.53	2.74 ± 0.25
7	Hexyl	4	2	3.001	−4.90	3.12 ± 0.24
8	Vinyl	89	96	0.613	−2.44	1.40 ± 0.27
9	(*E*)-Butenyl	68	80	0.755	−1.56	1.87 ± 0.28
10	(*E*)-Pentenyl	37	50	1.676	−2.87	0.67 ± 0.11
11	(*E*)-Hexenyl	26	43	2.224	−3.30	1.83 ± 0.31
12	(*E*)-Heptenyl	19	2	2.953	−4.09	3.40 ± 0.26
13	(*E*)-Octenyl	12	0.2	3.434	−4.45	4.60 ± 0.24
14	Propynyl	99	69	0.699	−2.69	0.27 ± 0.07
15	Butynyl	74	65	0.704	−2.15	0.66 ± 0.18
16	Hexynyl	26	53	2.287	−4.02	0.70 ± 0.14
17	Heptynyl	21	52	2.297	−4.70	1.60 ± 0.29
18	Octynyl	13	8	3.268	−4.72	2.77 ± 0.19

Table 2
The molecular parameters of the investigated deoxyuridine derivatives calculated by PC model approach. nopol = non-polar surface area (Å^2), nupol = non-polar unsaturated surface area (Å^2), polsa = polar surface area (Å^2), vdw = van der Waals radius, dm = dipole moment

No.	nopol	nupol	polsa	vdw	dm
1	168.0	13.5	194.0	9.17	4.33
2	236.0	2.3	186.4	10.96	4.39
3	272.3	2.1	179.8	11.35	4.47
4	295.7	3.1	184.4	11.99	4.51
5	283.4	4.1	184.6	11.66	4.21
6	324.4	4.1	186.9	12.39	4.30
7	354.6	3.9	185.7	12.77	4.30
8	199.7	32.4	187.7	13.44	4.44
9	285.1	20.3	192.0	14.55	4.48
10	302.1	19.3	191.1	14.98	4.49
11	319.5	18.8	194.1	15.35	4.47
12	339.8	18.9	189.9	15.70	4.48
13	382.1	17.8	192.9	16.07	4.47
14	231.5	28.3	188.9	9.43	4.44
15	278.7	20.1	188.5	10.12	4.37
16	324.0	18.3	188.7	10.96	4.31
17	342.9	22.8	187.1	11.40	4.27
18	351.0	18.4	182.9	11.49	4.30

the accessible polar and non-polar surface areas. The dipole moment (dm) values and the sum of the average van der Waals radii (vdw) of the molecules were also calculated.

The molecular parameters of the deoxyuridine derivatives are listed in Table 2.

The molar refractivity values (MR) and the inductive effects (F) of the substituents were taken from the Hansch–Leo compilation [8] and listed in Table 3. The hydrophobic properties of the compounds in the reversed-phase chromatographic system were characterized by the methanol concentration ($OP\%_0$) which is necessary for achieving log $k' = 0$ retention. The $OP\%_0$ values can be calculated from the slope and intercept (RPS, RPI) values as was described earlier [11]. The $OP\%_0$ values are also listed in Table 3.

Step-wise linear regression analysis has been carried out by the Drugidea^{TM} program system developed for drug design (Chemicro Ltd, Budapest, Hungary). The independent variables were selected according to their significance level considering the increase of the explained variance of the dependent variable. The significance level was always set to 5%. All of the programs were run on an IBM AT compatible personal computer.

Table 3
The molar refractivity (MR) and the inductive effect (F) values of the 5-substituents from the Hansch–Leo compilation [8]. The $OP\%_0$ values were obtained from the reversed-phase retention data by dividing the slope and the intercept values [11]

No.	MR	F	$OP\%_0$
1	1.03	0.00	16.00
2	10.30	0.00	24.68
3	14.96	−0.05	30.27
4	19.61	−0.06	46.06
5	19.63	−0.07	46.46
6	24.26	−0.06*	57.82
7	28.91	−0.06*	61.24
8	10.99	0.07	25.14
9	20.29	0.03	48.41
10	24.94	0.03*	58.40
11	29.59	0.03*	67.40
12	34.24	0.03*	72.19
13	38.89	0.03*	77.17
14	14.20	0.15	25.97
15	18.85	0.15*	32.73
16	28.15	0.15*	56.89
17	32.80	0.15*	48.86
18	37.45	0.15*	69.24

*Estimated data from values of propyl, propenyl, propynyl.

Results and Discussion

The measured initial incorporation rate values, chromatographic retention data, and the β-cyclodextrin inclusion complex stability data are listed in Table 1. Table 2 shows the calculated molecular parameters. The molar refractivity (MR) and the inductive effect (F) values of the 5-substituents obtained from the Hansch–Leo compilation [8] are listed in Table 3. The chromatographic parameters ($OP\%_0$) which are a measure of the hydrophobicity of the compounds [11] are also listed

in Table 3. The $OP\%_0$ values show the volume per cent of the methanol in the hydroorganic mobile phase by which the given compound shows the retention time, which is twice the dead time (i.e. $\log k' = 0$) [11].

The correlation coefficients between the investigated variables are listed in Table 4 in a correlation matrix.

Relationships between chromatographic retention data and molecular parameters

The reversed-phase slope (RPS) and intercept (RPI) values (i.e. the slope and the intercept of the straight line obtained by plotting the logarithmic values of capacity ratio against the methanol concentration of the mobile phase), showed a relatively high correlation coefficient ($r = 0.91$). This result shows that the compounds can be regarded as structurally related according to their behaviour in the chromatographic partition process between the mobile and the stationary phase [12].

The variance of the RPI values could be explained by the non-polar surface areas (nopol) value of 92%, when a parabolic function was calculated. The function is described by equation (1) and the plot is shown in Fig. 2,

$$\text{RPI} = 1.002 \times 10^{-4} \times \text{nopol}^2 - 0.0392\ \text{nopol} + 4.235, \quad (1)$$
$$n = 18;\quad r = 0.961;\quad s = 0.304;\quad F = 89.9,$$

where n is the number of compounds, r is the multiple correlation coefficient, s is the standard error of the estimate, F stands for the Fischer-test value.

Equation (1) shows that the higher the non-polar surface area of the compound the higher the RPI value (i.e. the extrapolated retention of the compound to the 100% water mobile phase concentration). The retention of the polar compounds however did not decrease with the decrease of their non-polar surface, but showed a minimum retention.

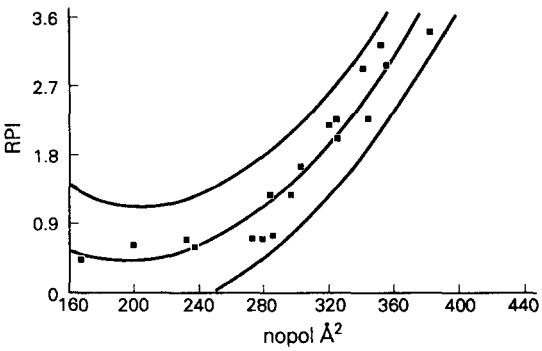

Figure 2
The plot of the non-polar surface area (Å^2) (nopol) against the reversed-phase chromatographic intercept values (RPI). For the parameters see equation (1). The 95% confidence intervals are also indicated.

It was supposed that the inclusion complex formation ability of the compounds with β-cyclodextrin provides information about the size and the polarity of the 5-substituent as the cyclodextrin ring is non-polar inside and it has a certain measure [6]. The step-wise linear regression analysis revealed that the variance of the logarithmic value of the inclusion complex stability constant can be explained by the non-polar unsaturated surface area (nupol), the dipole moment (dm) and the average van der Waals radii (vdw) of the compounds. Equation (2) shows the parameters of the function and the plot of the measured and estimated [from equation (2)] log K values is given in Fig. 3,

$$\log K = 0.021(\pm 0.007)\ \text{nupol} - 4.614(\pm 0.811)\ \text{dm} + 0.258(\pm 0.035)\ \text{vdw} + 16.650, \quad (2)$$
$$n = 18;\quad r = 0.912;\quad s = 0.264;\quad F = 23.2.$$

Table 4
The correlation coefficients of the investigated parameters in a correlation matrix

	RATE%	dAdT%	RPI	K	nupol	nupol	polsa	vdw	dm	$OP\%_0$	MR	F
RATE%	1.00											
dAdT%	0.78	1.00										
RPI	−0.56	−0.66	1.00									
K	−0.29	−0.50	0.82	1.00								
Nopol	−0.58	−0.73	0.89	0.72	1.00							
Nupol	0.76	0.55	0.00	0.13	−0.12	1.00						
Polsa	0.36	0.49	0.02	0.19	−0.11	0.47	1.00					
Vdw	−0.19	−0.31	0.50	0.66	0.52	0.16	0.35	1.00				
Dm	0.28	0.10	−0.16	−0.04	−0.09	0.22	0.27	0.51	1.00			
$OP\%_0$	−0.51	−0.65	0.92	0.79	0.92	0.00	0.12	0.71	0.03	1.00		
MR	−0.44	−0.63	0.93	0.77	0.96	0.12	0.01	0.57	−0.05	0.94	1.00	
F	0.49	0.37	0.12	−0.02	0.09	0.73	0.18	−0.24	−0.10	0.02	0.26	1.00

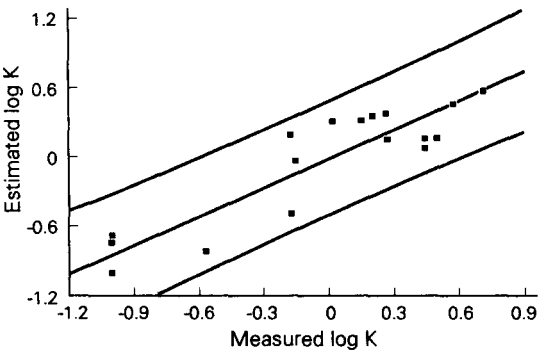

Figure 3
The plot of the measured values of the logarithm of inclusion complex stability constant (log K) against the estimated ones according to equation (2). The 95% confidence intervals are also indicated.

Equation (2) shows that higher van der Waals radii (β weight = 0.941), smaller dipole moment (β weight = 0.731) and higher non-polar unsaturated surface area (β weight = 0.342) are advantageous for the higher inclusion complex stability.

The reversed-phase chromatographic hydrophobicity parameter $OP\%_0$ values showed significant correlations to the non-polar surface area of the nucleosides and to their average van der Waals radii as it is described by equation (3),

$$OP\%_0 = 0.247(\pm 0.029)\ \text{nopol} + 2.734(\pm 0.761)\ \text{vdw} - 58.512, \quad (3)$$
$$n = 18;\ r = 0.957;\ s = 5.73;\ F = 81.5.$$

The plot of the measured and estimated $OP\%_0$ values are shown in Fig. 4. Equation (3) means that a higher methanol concentration is needed in the mobile phase for compounds exerting higher non-polar surface area and van der Waals radii to achieve the same retention (log $k' = 0$).

Relationships between initial incorporation rate values and molecular parameters

When a random sequence template (calf thymus DNA) was used in the polymerase enzyme reactions the initial incorporation rate values of the corresponding nucleotides (RATE%) were dependent on the hydrophobic character of the compounds, while steric properties did not show significant influence on the incorporation rate. The variance of RATE% could be explained by the reversed-phase chromatographic RPI values and by the non-polar unsaturated surface areas (nupol) as described by equation (4),

$$\text{RATE\%} = 2.448(\pm 0.279)\ \text{nupol} - 16.475(\pm 2.553)\ \text{RPI} + 23.037 \quad (4)$$
$$n = 18;\ r = 0.942;\ s = 10.8;\ F = 59.0.$$

The plot of the measured and estimated RATE% values is shown in Fig. 5. The initial incorporation rate of the compounds was facilitated by the higher non-polar unsaturated surface area (β weight of nupol = 0.760). As the RPI values showed a parabolic relation to the non-polar surface areas [see equation (1)], i.e. compounds with higher non-polar surface areas had higher RPI values through a minimum, the negative regression coefficient of RPI in equation (4) means that higher non-polar surface area did not facilitate the DNA incorporation.

In contrast, when a strictly alternating tem-

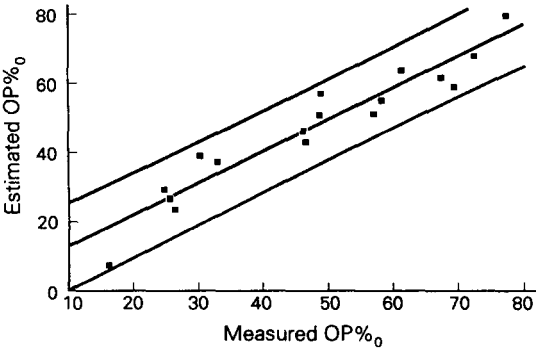

Figure 4
The plot of the measured $OP\%_0$ values (the methanol concentration needed to the retention log $k' = 0$) against the estimated ones according to equation (3). The 95% confidence intervals are also indicated.

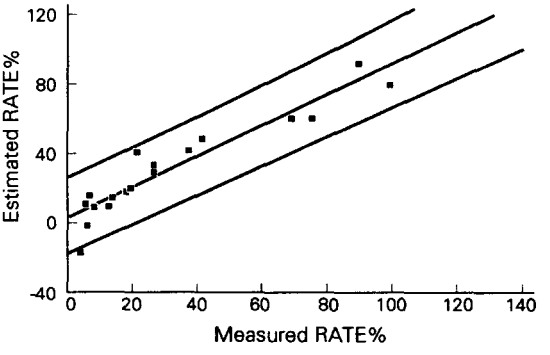

Figure 5
The plot of the measured initial incorporation rate values by using calf thymus DNA as template (RATE%) against the estimated ones according to equation (4). The 95% confidence intervals are also indicated.

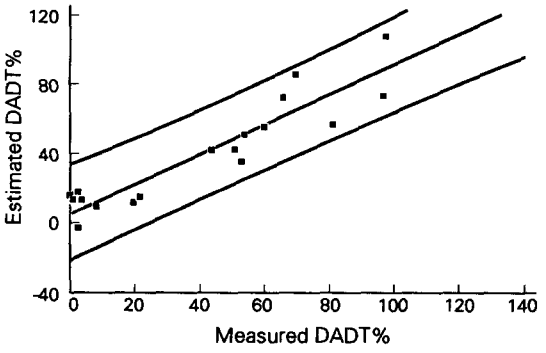

Figure 6
The plot of the measured initial incorporation rate values by using poly(dA-dT) template (DADT%) against the estimated ones according to equation (5). The 95% confidence intervals are also indicated.

plate [poly(dA-dT)] was used in the polymerase reactions the steric properties (MR) and the inductive electronic effects (F) of the 5-substituents were significant in the initial incorporation rate of the nucleotides (DADT%). The polar surface area (polsa) of the compounds played also an important role as described by equation (5),

$$\begin{aligned} \text{DADT\%} &= -2.445(\pm 0.335) \\ &\text{MR} + 208.74(\pm 42.2)F + 3.349(\pm 0.855) \\ &\text{polsa} - 542.3, \qquad (5) \\ n = 18; \quad r = 0.930; \quad s = 13.7; \quad F &= 29.6. \end{aligned}$$

The plot of the measured and estimated [according to equation (5)] DADT% values is shown in Fig. 6. The β weights of the independent variables are as follows: −0.738 for MR, 0.508 for F and 0.393 for polsa. The initial incorporation rate of the derivatives when a strictly alternating template was considered was facilitated by the higher polar surface of the compounds, the stronger inductive electronic effects of the 5-substituents, while the larger size caused steric hindrance of the reaction.

The three-dimensional structures of iso-propyl and sec-butyl deoxyuridine are shown in Figs 7 and 8, respectively. The darker dots show the polar surface of the derivatives. When the strictly alternating template (DADT%) was considered iso-propyl deoxyuridine showed less ability for DNA incorporation than the sec-butyl deoxyuridine. As can be seen on the three-dimensional pictures the non-polar iso-propyl substituent decreases the polar surface region around the oxygens and the nitrogen of the base which are involved in the hydrogen bonds during the base pair formation. The geometry of the sec-butyl substituent allows sterically the base pair formation as it does not decrease the polar surface area involved in the hydrogen bond between base pairs.

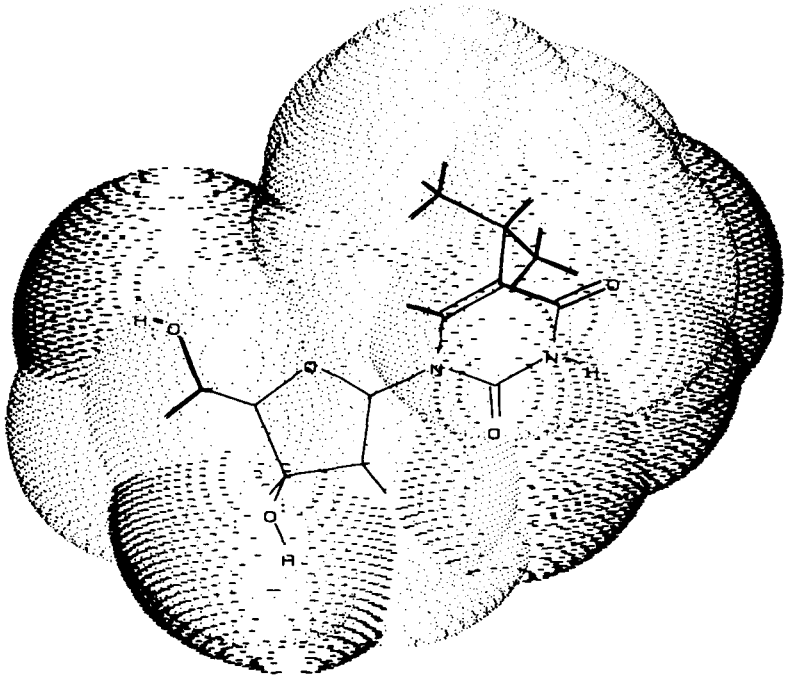

Figure 7
The three-dimensional structure of the *i*-propyl deoxyuridine (compound 3) obtained by molecular modelling. The darker dots indicate the polar surface region of the molecule.

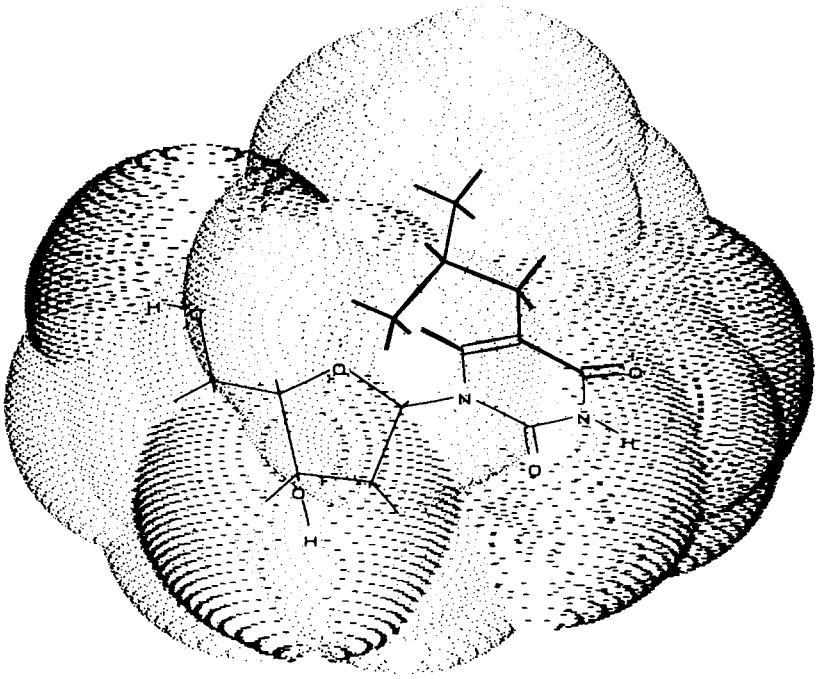

Figure 8
The three-dimensional structure of the sec-butyl deoxyuridine (compound 4) obtained by molecular modelling. The darker dots indicate the polar surface region of the molecule.

Conclusions

The study proved the applicability of the molecular parameters obtained by three-dimensional molecular modelling for explaining the chromatographic behaviour of the nucleoside derivatives and it proved to be a valuable aid in the understanding of the molecular requirements for the DNA incorporation ability of the nucleotides.

Acknowledgement — This research was supported by the Hungarian National Research Foundation (OTKA) Grant No. 2670.

References

[1] E. De Clerq, *Meth. and Find. Exptl. Clin. Pharmacol.* **2**, 253–267 (1980).
[2] J.T. Sági, A. Szabolcs, A. Szemzö and L. Ötvös, *Nucleic Acid Res.* **4**, 2767–2776 (1977).
[3] L. Ötvös, J. Sági, T. Kovács and R.T. Walker, *Nucleic Acid Res.* **15**, 1763–1777 (1987).
[4] L. Ötvös, J. Szécsi, J. Sági and T. Kovács, *Nucleic Acid Res. Symp. Ser.* **18**, 125–129 (1987).
[5] K. Valkó, I. Fellegvári, J. Sági and A. Szemzö, *J. Liq. Chromatogr.* **12**, 2103–2116 (1989).
[6] K. Valkó, T. Cserháti, I. Fellegvári, J. Sági and A. Szemzö, *J. Chromatogr.* **506**, 35–44 (1990).
[7] T. Cserháti, E. Fenyvesi and J. Szejtli, *Acta Biochem. Biophys. Acad. Sci. Hung.* **18**, 60–64 (1983).
[8] C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley, New York (1979).
[9] U. Burkert and N.L. Allinger, *Molecular Mechanics.* American Chemical Society, Washington, D.C. (1982).
[10] A. Szabolcs, J. Sági and L. Ötvös, *J. Carbohydr. Nucleosides Nucleotides* **2**, 197–200 (1975).
[11] K. Valkó, in *Chromatography, the State of the Art* (H. Kalász and L. Ettre, Eds), pp. 739–750. Akadémia, Budapest (1985).
[12] K. Valkó, *J. Liq. Chromatogr.* **10**, 1663–1686 (1987).

[Received for review 30 April 1991; revised manuscript received 3 June 1991]