

Modelling of the binding specificity in the interactions of cationic porphyrins with DNA

Xiaowen Hui, Nohad Gresh and Bernard Pullman*
Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique Associé au CNRS, 13 rue Pierre et Marie Curie, 75005 Paris, France

Received December 21, 1989; Revised and Accepted February 13, 1990

ABSTRACT

A theoretical investigation is performed of the complexes of a tetracationic porphyrin, tetra-(4-N-methylpyridyl)-porphyrin, (T4MPyP), with the hexanucleotides d(CGCGCG)$_2$ and d(TATATA)$_2$, considering the possibility of both the intercalative and the groove binding interactions. These computations demonstrate that T4MPyP manifests a significant preference for intercalation in its complex with d(CGCGCG)$_2$ but for non intercalative binding in the minor groove in its complex with d(TATATA)$_2$. Such a dual binding behaviour of T4MPyP as a function of the sequence to which it is attached is fully consistent with available experimental data. It demonstrates that intercalation and groove binding may be viewed as two potential wells on a continuous energy surface. In agreement with experiment, the computations indicate that in the here considered case the deepest well is associated with intercalation.

INTRODUCTION

Drugs which bind non covalently to DNA are generally considered to do so by two distinct mechanisms : intercalation between base pairs or binding in a groove and are thus classified as intercalators or groove binders (for a recent review see ref. 1). It was occasionally considered that 'outside' binding may represent a transient state for intercalation or a concomitant but much weaker mode of interaction and it was also recognized that bulky substituents may possibly transform an intercalator into a groove binder (see e.g. 2). Essentially, however, the division of 'physically' bound drugs into intercalators or groove binders is an established practice.

Recently two types of antitumor drugs have been shown to exhibit the unaccustomed feature of being able to behave both as intercalators and groove binders in their interaction with synthetic polynucleotides and with DNA. These compounds are a series of large cationic porphyrins (for a recent review see e.g. 3) and DAPI (4',6-diamidino-2-phenylindole) and related systems [4]. The present paper is devoted to the exploration of the factors involved in this particular situation in the former of these compounds.

The fundamental, parent molecule for this series of cationic porphyrins is the meso-tetra-(4-N-methylpyrydil) porphyrin (T4MPyP, fig. 1). Studies with synthetic poly- and oligonucleotides and with DNA have indicated in their great majority [3–9] a sequence selective interaction, intercalative in GC sequences but non intercalative ('outside binding') in AT sequences. The dual interaction is also observed with cationic metalloporphyrins [1, 7–12], the sequence and mode of binding preference depending then strongly, if not decisively, on the nature of the bound metal ion and the porphyrin structure. The present introductory study will be limited to the case of the metal free porphyrin.

Molecular modelling of the interaction of T4MPyP with the dinucleotide monophosphates dCpG and dTpA [13] presented evidence in favour of the feasibility of a full intercalation into the dCpG sequence and the impossibility of such an intercalation into the dTpA sequence, without excluding, however, the possibility of a 'partial' intercalation in the latter case. While obviously significant with respect to the prominent difference of the two sequences towards the acceptance of T4MPyP as an intercalator, this study appears nevertheless as incomplete in the sense that, because of the very short dimensions of the receptor sites, it could not permit a comparative investigation of the competitive, also sequence dependent, groove binding ability of the drug. Such an investigation can only be achieved by considering oligonucleotide receptors of dimensions sufficient to shape the major features of the groove properties. Moreover the use of such longer segments enables also to investigate the influence of flanking base pairs on the reactivity of the intercalation site, a problem which raised recently some questions [14–16]. For these major reasons our own modelling of DNA-T4MPyP interactions uses deoxyhexanucleotides as drug receptors.

PROCEDURE

The computational procedure used in this study is the JUMNA method [17, 18], utilized in a number of previous investigations of this Laboratory on oligonucleotide-ligand complexes (for a review see 1).

* To whom correspondence should be addressed

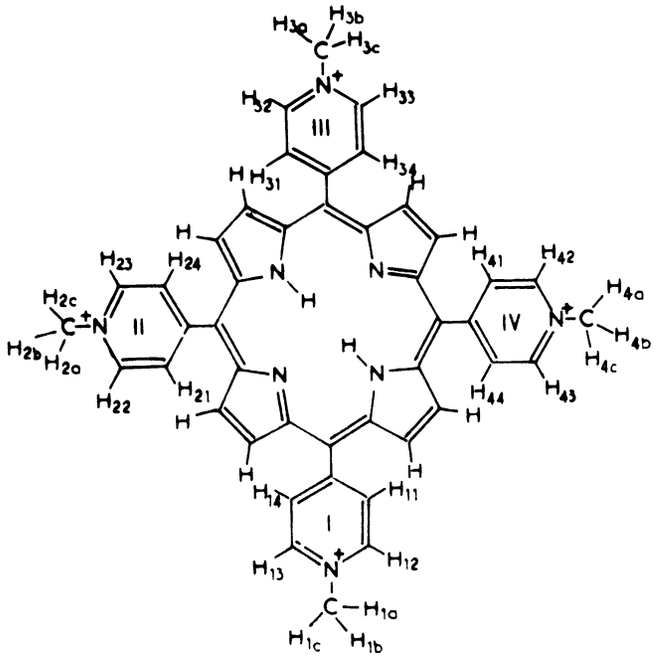

Figure 1. T4MPyP. Structural formula and atom numbering.

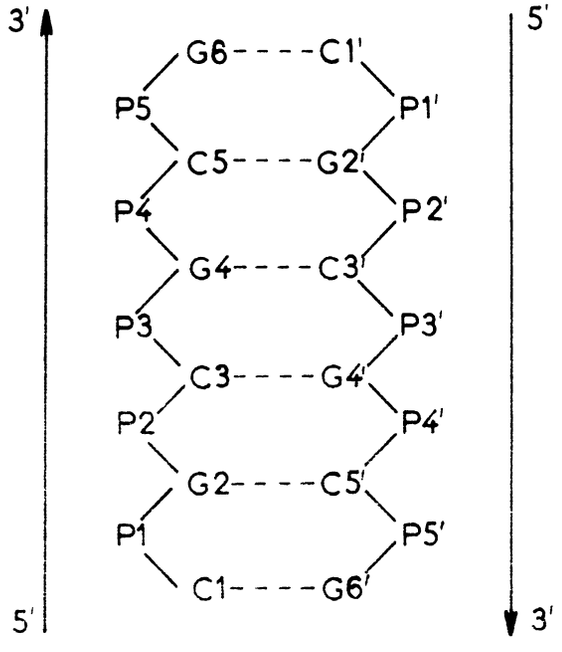

Figure 2. Base numbering of the investigated hexamers, as exemplified on sequence $d(CGCGCG)_2$.

The major part of the computations bears on the interaction of T4MPyP with the alternating sequences $d(CGCGCG)_2$ and $d(TATATA)_2$. The intercalative complexes are considered to occur between the central pyrimidine-purine base pairs, i.e. at step $d(C_3G_4)$ or $d(T_3A_4)$, following the notations of figure 2. They were investigated for both sequences in three distinct orientational possibilities : a symmetrical one with two N-methylpyridinium rings lying in the major groove and two in the minor groove and two asymmetrical ones, one with three N-methylpyridinium rings in the major groove and the fourth in the minor groove and the other with three N-methylpyridinium rings in the minor groove and the fourth in the major groove. These configurations will be denoted as 2M-2m, 3M-1m, and 1M-3m, respectively.

Because of the proposal in [3, 16] that in natural DNAs, as contrasted to poly(dA-dT), intercalation of T4MPyP could take place at a d(TpA) step (whereas in the synthetic polymer only groove-binding occurs), we were led to investigate theoretically the intercalative binding of T4MPyP to the 'mixed' sequence $d(CGTACG)_2$ as well.

The non intercalative complexes of T4MPyP with $d(CGCGCG)_2$ have been investigated both for the major and minor groove binding. Similar binding to $d(TATATA)_2$ was explored solely in the minor groove, on account of the well-known preference of groove binders for this groove in this sequence (for a review see e.g. 1), due principally, as demonstrated abundantly in our Laboratory [1], to the significantly more attractive molecular electrostatic potential in this groove of this sequence [19]. Outside binding of T4MPyP, by which we mean binding esentially to the sugar-phosphate backbone, was investigated only with $d(TATATA)_2$, as no great difference in this type of binding is expected as a function of base sequence.

RESULTS AND DISCUSSION

The results of the computations of the energies of T4MPyP-oligonucleotide interactions are reported in Table I, which lists, for each investigated and energy-minimized complex, the intermolecular oligonucleotide-porphyrin interaction energy E_{inter}, the value of the conformational energy variation of the oligonucleotide, ΔE_{DNA}, with respect to its most stable B-DNA conformation taken as energy zero, the conformational energy variation of the ligand, ΔE_{lig}, the resulting energy balance $\Delta E = E_{inter} + \Delta E_{DNA} + \Delta E_{lig}$, and the difference, δ, of energy balances with respect to the best energy balance taken as energy zero. A common reference of δ is adopted for both intercalating and groove binding interactions.

Tables II and III summarize the most salient interatomic drug-oligonucleotide distances in the best complexes found.

1) Intercalated complexes

The most stable intercalative complex occurs with the $d(CGCGCG)_2$ oligonucleotide demonstrating thus the preference of the drug for CG sites. Moreover the binding of the porphyrin to this oligonucleotide occurs in the following order of relative configurational preferences :

$$2M\text{-}2m > 3M\text{-}1m > 1M\text{-}3m$$

indicating a marked specificity for the 'symmetrical' mode of intercalation. This marked preference for 2M-2m over 3M-1m is dictated by the highly unfavourable ΔE_{DNA} term in the latter configuration. The least favourable energy balance in the 1M-3m configuration, on the other hand, stems from the low intermolecular energy term.

In the most stable 2M-2m complex of T4MPyP with $d(CGCGCG)_2$ (fig. 3), one of the methylpyridinium rings, say ring II, is in the closest proximity to base G4 in the major groove, with stabilizing electrostatic interactions occurring between O_6 and N_7 of this base and atom H_{24} of the porphyrin, on which the cationic charge is partially delocalized. An additional stabilization stems from the vicinity of atom H_{22} of this ring to

TABLE I. Values of the binding energies of T4MPyP to $d(CGCGCG)_2$, $d(TATATA)_2$ and $d(CGTACG)_2$ (see text for definitions). Energies in kcal/mole.

	$d(CGCGCG)_2$			$d(TATATA)_2$			$d(CGTACG)_2$	
Intercalation	2M-2m	3M-1m	1M-3m	2M-2m	3M-1m	1M-3m	2M-2m	3M-1m
E_{inter}	−159.9	−162.1	−129.7	−141.7	−157.3	−109.8	−144.1	−174.0
ΔE_{DNA}	33.5	57.4	33.2	32.2	44.2	32.9	37.2	67.5
ΔE_{lig}	2.3	1.7	0.3	2.4	2.5	1.1	2.0	2.5
ΔE	−124.1	−103.0	−96.2	−107.1	−110.6	−75.8	−104.9	−104.3
δ	0.0	21.1	27.9	17.0	13.5	48.3	19.2	19.8

	$d(CGCGCG)_2$		$d(TATATA)_2$	
Groove Binding	major groove	minor groove	minor groove	sugar phosphate backbone
E_{inter}	−133.1	−116.5	−133.5	−74.9
ΔE_{DNA}	31.1	20.7	15.6	14.4
ΔE_{lig}	2.0	1.9	1.3	1.0
ΔE	−100.0	−93.9	−116.6	−59.5
δ	24.1	30.2	7.5	64.6

TABLE II. T4MPyP-oligonucleotide contacts in interaction complexes (in Å units).

$d(CGCGCG)_2$

Intercalation (2M-2m)

ring	ligand	DNA	Å
	H12	O2(P3)	2.44
I	H1a	O2(P3)	2.61
	H1b	O2(P3)	2.55
	H24	O6(G4)	2.49
II	H24	N7(G4)	2.80
	H22	O1(P2)	2.81
	H31	O6(G4′)	2.34
III	H31	N7(G4′)	2.55
	H34	N3(C3′)	2.95
	H42	O2(P3′)	2.32
IV	H4a	O2(P3′)	2.67
	H4b	O2(P3′)	2.88

Intercalation (3M-1m)

ring	ligand	DNA	Å
	H11	O1(P2)	2.83
	H12	O1(P2)	2.35
I	H12	O1(P3)	2.23
	H1a	O1(P3)	2.97
	H21	O1(P2)	2.47
II	H22	O1(P2)	2.42
	H31	O6(G4′)	2.33
	H31	N7(G4′)	2.31
	H32	O1(P3′)	2.50
III	H32	N7(G4′)	2.61
	H34	N3(C3′)	2.98
	H3a	O1(P3′)	2.70
	H3b	O1(P3′)	2.69
	H41	O2(C3)	2.31
	H41	N3(C3)	2.66
IV	H43	O2(C5)	2.93
	H43	O1′(S5)	2.58
	H44	N3(G4)	2.35
	H4a	O2(P5)	2.52
	H4b	O2(P5)	2.93

$d(TATATA)_2$

Intercalation (3M-1m)

ring	ligand	DNA	Å
	H12	O1(P2)	2.93
	H12	O2(P3)	2.72
I	H14	N7(A4)	2.63
	H1a	O2(P3)	2.95
	H1b	O2(P3)	2.58
	H31	N7(A4)	2.46
	H32	O2(P3′)	2.45
	H33	O1(P2′)	2.33
III	H3a	O2(P3′)	2.61
	H3b	O2(P3′)	2.52
	H3b	O5′(P2′)	2.91
	H41	O2(T3)	2.24
	H41	N3(T3)	2.78
	H41	N1(A4′)	2.78
IV	H42	O2(T3)	2.84
	H42	O2(T5′)	2.52
	H43	O1′(S5)	2.53
	H44	N3(A5)	2.73
	H4a	O1′(S5′)	2.54

Intercalation (2M-2m)

ring	ligand	DNA	Å
	H11	O2(T3)	2.27
	H13	O3′(P4)	2.91
I	H14	N3(A4)	2.79
	H1a	O2(P3)	2.47
	H1b	O2(P3)	2.58
	H24	N6(A4)	2.99
II	H24	N7(A4)	2.87
	H31	N7(A4′)	2.46
III	H34	O4(T3′)	2.34
	H41	O2(T3′)	2.28
	H42	O2(P3′)	2.52
IV	H44	N3(A4′)	2.86
	H4a	O2(P3′)	2.51
	H4b	O2(P3′)	2.50

TABLE III. T4MPyP-oligonucleotide contacts in groove-binding and 'outside' complexes.

	d(CGCGCG)$_2$				d(TATATA)$_2$				d(TATATA)$_2$			
	major groove binding				minor groove binding				binding on the sugar phosphate backbone			
ring	ligand	DNA	Å	ring	ligand	DNA	Å	ring	ligand	DNA	Å
I	H14	N7(G2′)	2.53		H12	O2(T5′)	2.24		H11	O1(P3′)	2.82
					H12	O1′(S6)	2.54		H12	O2(P2′)	2.36
	H22	O1(P2)	2.32		H13	O3′(P3′)	2.45		H12	O3′(P3)	2.98
	H23	N7(G4)	2.65		H14	O5′(P3′)	2.32	I	H1a	O3′(P6)	2.67
II	H24	O6(G4)	2.21	I	H14	O1′(S4′)	2.59		H1a	O3′(P3′)	2.56
	H2a	O1(P2)	2.52		H1a	O2(T5)	2.65		H1b	O3′(P6)	2.88
	H2b	O1(P2)	2.50		H1a	O1′(S6)	2.66				
					H1a	N3(A6)	2.78		H22	O1(P4)	2.19
IV	H44	O1(P2′)	2.18		H1b	O1′(S6)	2.54		H22	02(P4)	2.51
								II	H2a	O2(P4′)	2.81
	H31	N7(G2)	2.86		H42	O1′(S4)	2.46		H2b	O2(P4′)	2.58
	H32	N7(G2)	2.64		H43	O2(T3)	2.45				
	H3b	O6(G2)	2.76		H43	O2(T5′)	2.60				
III	H3b	O6(G6′)	2.90		H44	N3(A4)	2.98				
	H3c	N3(C1)	2.85	IV	H44	O2(T5′)	2.36				
	H3c	O6(G2)	2.65		H4a	O5′(P3)	2.84				
	H3c	O6(G6′)	2.47		H4a	O2(T3)	2.55				
					H4a	O1′(S4)	2.44				
					H4b	O5′(P3)	2.44				
					H4b	O1′(S4)	2.99				
					H4b	O2(T3)	2.65				

O_1 of P2, upstream of the intercalation site. A related proximity in the major groove occurs between ring III stabilized by electrostatic attractions between H_3, and O_6 and N_7. In the minor groove, the anionic oxygen O_2 of P3 of the intercalation site is in simultaneous vicinity to H_{12} and to the N-methylpyridinium hydrogens H_{1a} and H_{1b}. Concomitantly, O_2 of P3′ is in a simultaneous vicinity to H_{42}, H_{4a} and H_{4b} of ring IV. Altogether the four N-methylpyridinium rings are inclined by values in the range 65° − 120°with respect to the porphyrin plane, a range consistent with the average value found in the crystal structure of T4MPyP [13] and with that deduced by us by energy-minimization on T4MPyP. A related range of values was also found in our other energy-minimized complexes of T4MPyP. This situation explains the relatively small values of ΔE_{lig} in the complexes.

A more pronounced tilting of the porphyrin ring with respect to the base pairs, together with its sliding towards the major groove, occurs in the next in stability 3M-1m complex with d(CGCGCG)$_2$. This situation results in the onset of electrostatic interactions in the major groove between atoms H_{21} and H_{22} of ring II and O_1 of P1, one dinucleotide step upstream from the intercalation site. The axis passing through the long axis of rings II and IV (the latter in the minor groove) is approximately parallel to the average dyad axis of the two base pairs of the intercalation site. Additional interactions in the major groove occur between ring I and O_1 of P2 on the one hand, and ring III and O_1 of P3′, O_6 and N_7 of G4′ on the other hand. Such a tilting further enables interactions of ring IV in the minor groove with sites O2 (C5), O_1, (S5) and O_2 (P5), one step downstream of the intercalation site, in addition to interactions occurring between this ring and O_2 and N_3 of C3 as well as N_3 of G4 in the intercalation site. These are facilitated by a bending of base pairs C1′-G6 and G2′-C5 towards the minor groove, in this complex.

Experimental evidence, confirmatory of the preference for the symmetrical selective intercalative binding to CG sequences was provided recently by ^1H and ^{31}P NMR exploration of the T4MPyP-poly(dG-dC)·poly(dG-dC) complex [20].

The intercalation of T4MPyP into the d(TATATA)$_2$ oligonucleotide is at obvious disadvantage with respect to intercalation into the regular CG hexamer. Moreover, the binding configurations to d(TATATA)$_2$ are ranked in the following order of decreasing stabilities :

3M-1m > 2M-2m > 1M-3m

indicating that, if feasible, such an intercalation would be asymmetric. The preference for 3M-1m over 2M-2m is due in this case to a distinctly more favourable E_{inter} term (by 15.6

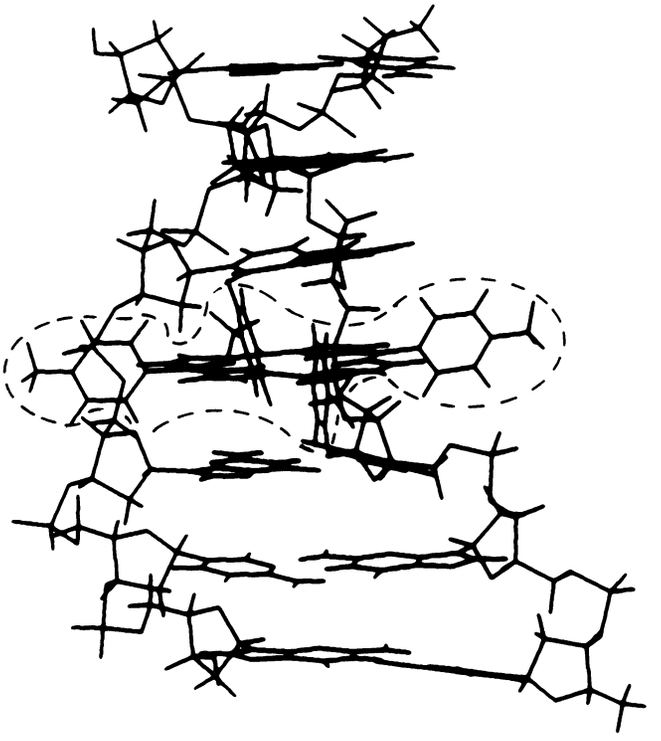

Figure 3. Representation of the intercalation complex 2M-2m of T4MPyP with d(CGCGCG)$_2$.

kcal/mole) overcompensating for the less favourable ΔE_{DNA} term.

It must, however, be noted that a dramatic distortion of the DNA backbone occurs in the 3M-1m complex of T4MPyP with d-$(TATATA)_2$ (fig. 4 and Table II) with a buckle of base pair T3-A4′ of the intercalation site. Ring II in this groove no longer partakes in direct interactions with the oligonucleotide whereas both rings I and III are essentially bound to the backbones of the primed and unprimed strands, respectively. As a consequence of the deformation of the receptor oligomer, a large number of attractive interactions occur in the minor groove between ring IV and sites O_2, N_3, $O_{1'}$, belonging to T3, A4 of the intercalation site, as well as to T5′ downstream of it. At most this mode of interaction of T4MPyP with $d(TATATA)_2$ could be considered as representing a distorted partial intercalation.

The binding energetics of T4MPyP to the mixed sequence $d(CGTACG)_2$, in the two representative arrangements 2M-2m and 3M-1m, do not differ markedly from those in the corresponding complex with the regular sequence $d(TATATA)_2$. These results do not support thus the hypothesis [15] that the inherently unfavourable intercalation of T4MPyP at a d(TpA) step could be facilitated by adjacent GC pairs.

2) Groove binding complexes

In clear distinction to the situation with the intercalative complexes, groove-binding occurs preferentially with the $d(TATATA)_2$ sequence, showing thus a net specificity for AT base pairs. Moreover, as indicated in the Introduction, numerous previous studies (see ref. 1) on groove binders do not leave any doubt that this sequence preference is associated with a parallel preference for its minor groove.

This non intercalative binding of T4MPyP to $d(TATATA)_2$ (fig. 5 and Table III) is stabilized by numerous attractive interactions involving the adjacent rings I and IV. Both are able to interact with sites O_2, $O_{1'}$, N_3 and $O_{3'}$ belonging to both strands, the unprimed strand contributing the most to the binding interactions. Ring II, on the other hand, is far remote from the oligonucleotide. The line connecting the long axis of rings II and IV is nearly colinear with the long axis of base pair T3-A4′, whereas the line connecting rings I and III follows the direction of the helical axis.

A detailed examination of Table I indicates that the most favourable groove binding of T4MPyP to the $d(CGCGCG)_2$ sequence, which is associated with its major groove, is much weaker (by 16.6 kcal/mol) than the groove binding to the $d(TATATA)_2$ sequence. Such a T4MPyP-$d(CGCGCG)_2$ interaction would involve predominantly the two adjacent rings III and IV. Ring II could interact through its trimethylpyridinium with O_1 (P2) end as well as with O_6 and N_7 of G4. Ring III is able to span sites O_6 and N_7 belonging to the two successive base pairs C1-G6′ and G2-C5′. The situation implies, however, a very strong distortion of the receptor DNA.

The binding of T4MPyP to the sole sugar-phosphate backbone of $d(TATATA)_2$ involves a relatively reduced number of stabilizing interactions and concerns solely rings I and II. Hence the particularly small complexation energy of such an arrangement.

3) Energetics of intercalation versus groove-binding interactions

This is the central problem investigated in this exploration. The results of the computations lead to the straightforward indication that *while the intercalative mode of binding is the dominating one for the interaction of T4MPyP with the $d(CGCGCG)_2$ sequence*, in which it is favoured by 24.1 kcal/mole over the most efficient groove binding association, *the reverse is true in the association of this drug with the $d(TATATA)_2$ sequence*, in which the groove binding interaction is by 9.5 kcal/mole more stable than intercalation.

Moreover, the computational results indicate that the *overall most stable interaction is intercalation*, with its associated preference for the CG sequences, a result in agreement also with

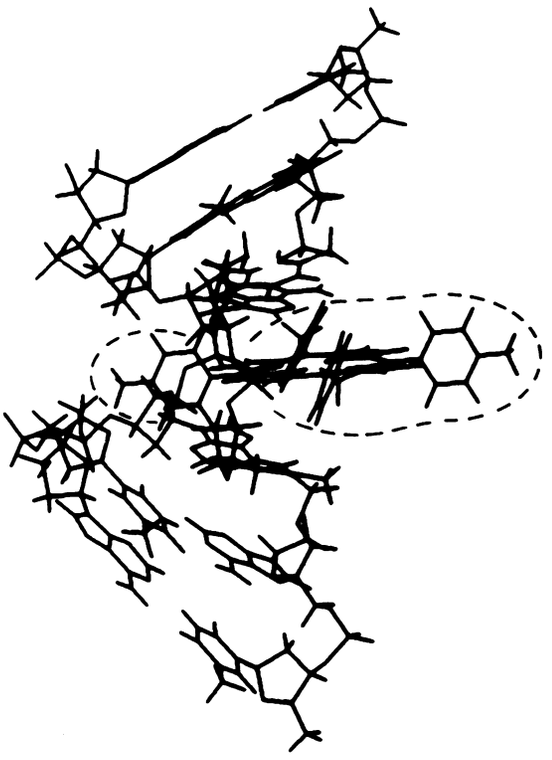

Figure 4. Representation of the intercalation complex 3M-1m of T4MPyP with $d(TATATA)_2$

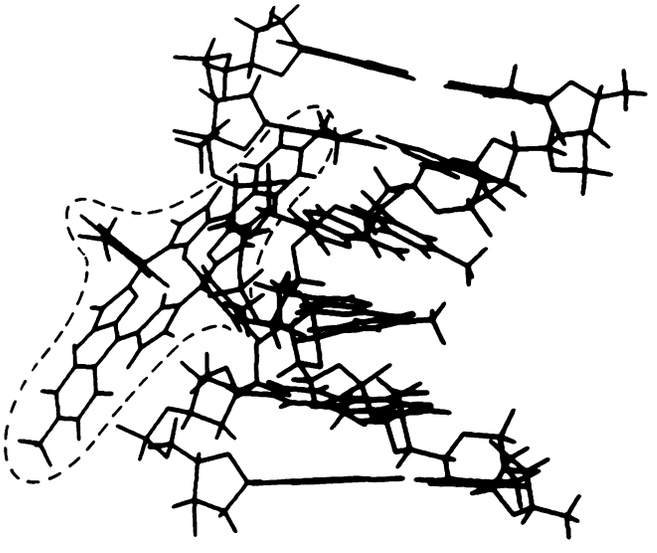

Figure 5 : Representation of the groove-binding complex of T4MPyP with $d(TATATA)_2$.

recent experimental findings of Sari et al. [21], obtained by competition binding experiments using ethidium bromide or one of its dimers.

CONCLUSIONS

To our knowledge, this is the first theoretical demonstration of a differential selection of binding modes, intercalation in $d(CGCGCG)_2$ and minor groove binding to $d(TATATA)_2$, elicited by a DNA-binding ligand and is fully consistent with available experimental data for T4MPyP binding [3–9].

This situation indicates that, as elegantly stated in [22], dealing with the interaction of a series of unfused tricyclic aromatic cations with DNA, 'intercalation and groove binding modes should be viewed as two potential wells on a continuous energy surface'. In the present case the deepest well is associated with intercalation, theory and experiment agreeing. Further explorations will demonstrate whether this is a general situation or whether it may vary as a function of the nature of the associating partners.

ACKNOWLEDGEMENT

This work was supported by the Association for International Cancer Research (Burnell and St. Andrews Universities, United Kingdom) to which the authors wish to express their deep thanks.

REFERENCES

1. Pullman, B. (1989) Advances in Drug Research, **18**, 1–113.
2. Müller, W., Crothers, D.M. and Waring, M.J. (1973) Eur. J. Biochem., **39**, 223–234.
3. Fiel, R.J. (1989) J. Biomol. Struct. Dynam., **6**, 1259–1274.
4. Wilson, W.D., Tanious, F.A., Barton, H.J., Strekowski, L. and Boykin, O.W. (1989) J. Am. Chem. Soc., **111**, 5008–5010.
5. Pasternak, R., Gibbs, E.I. and Villafranca, J.J. (1983) Biochemistry, **22**, 2406–2414.
6. Pasternak, R., Gibbs, E.I. and Villafranca (1983), Biochemistry, **22**, 5409–5417.
7. Kelly, J.M., Murphy, M.J., McConnell, D.J. and Ohlligin, C. (1985) Nucleic Acids Res., **13**, 167–184.
8. Banville, D.L., Marzilli, L.G. and Wilson, W.D. (1983) Biochim. Biophys. Res. Comm. **113**, 148–154.
9. Strickland, J.A., Marzilli, L.G., Gay, K.M., and Wilson, W.D. (1988) Biochemistry, **27**, 8870–8878.
10. Bromley, S.D., Ward, B.W. and Dabrowiak, J.C. (1986) Nucleic Acid Res., **14**, 9133–9148.
11. Ward, B., Skorobogaty, A. and Dabrowiak, J.C. (1986) Biochemistry, **25**, 6875–6883.
12. Ward, B., Skorobogaty, A. and Dabrowiak, J.C. (1986) Biochemistry, **25**, 7827–7833.
13. Ford, K.G., Pearl, L.H. and Neidle, S. (1987) Nucleic Acid Res., 15, 6553–6562.
14. Banville, D.L., Marzilli, G.G., Strickland, J.A. and Wilson, W.D. (1986) Biopolymers, **25**, 1837–1858.
15. Jenkins, B.G., Dissertation, Ph. D., RPMI Division SUNY/Buffalo (1986) quoted in 3.
16. Ford, K., Fox, K.R., Neidle, S. and Waring, M.J. (1987) Nucleic Acid Res., **15**, 2221–2234.
17. Lavery, R. (1988) in 'DNA Bending and Curvature' (Olson, W., Sundaralingam, M., Sarma, M.H. and Sarma, R.H. eds.) Academic Press, New York 191–211.
18. Lavery, R. (1988) in 'Unusual DNA Structures' (Harvey, S. and Wells, R. eds.), Springer Verlag, Pergamon Press, New York) 189–206.
19. Pullman, A. and Pullman, B. (1981) Quarter. Rev. Biophys., **14**, 289–380.
20. Marzilli, G., Banville, D.L., Zon, G. and Wilson, W.D. (1986) J. Am. Chem. Soc., **108**, 4188–4192.
21. Sari, M.A., Battoni, J.P., Mansuy, D. and Le Pecq, J.B., (1986) Biochem. Biophys. Res. Comm., **141**, 643–649.
22. Wilson, W.D., Tanious, F., Watson, R., Barton, M., Strekowska, A., Harden, D. and Strekowski, L. (1989) Biochemistry, **28**, 1984–1992.