NMR-titrations with complexes between ds-DNA and indole derivatives including tryptophane containing peptides

Joachim Sartorius, Hans-Jörg Schneider*

FR Organische Chemie der Universität des Saarlandes, D-66041 Saarbrücken, Germany

Received 31 July 1995; revised version received 25 September 1995

Abstract **It is shown that NMR titrations can be used on a quantitative basis to derive binding constants and binding modes of ds-DNA ligand complexes from several signals. The results are partially at variance with literature conclusions; they can be interpreted by additive and independent contributions of electrostatic interactions between DNA phosphate and ammonium centers in the side chain of the ligands, and of very weak stacking contributions of the indole rings. While gramine and tryptamine are found to intercalate, tryptophane-containing peptides do so only, if the tryptophane is flanked by at least two lysine units.**

Key words: NMR-spectroscopy; ds-DNA; intercalation; Electrostatic interactions; Indole derivatives; Tryptophane; DNA–peptide interactions

1. Introduction

The interaction strength and modes of complexes between nucleic acids and natural as well as synthetic ligands is of obvious significance [1] in biological and medicinal chemistry, and has been analyzed with many physical techniques [2]. The binding modes of several biologically important indole derivatives to DNA have been studied and summarized in particular by Helène et al [3]. Tryptophane (**1d**) or indole (**1a**) alone do not seem to intercalate even with single-stranded DNA, which generally is observed to be a much stronger binder for such small aromatic rings than the double-stranded polymer [3]. However, the naturally occuring amines tryptamine (**1c**) and gramine (**1b**) and the di- and tripeptides of tryptophane with neutral amino acids are believed to intercalate [3e]. Preliminary results in our laboratory concerning the peptides were not in line with this, and led us to apply NMR methods to such complexes. Furthermore, we hoped to obtain by NMR conclusive results also with respect to the binding of tryptophane-containing peptides to ds-DNA, for which Helène et al. [4], mostly on the basis of fluorescence titrations, described 2 step equilibria involving primarily electrostatic groove binding or/and weak intercalation in the second step, characterized by binding constants (for K_2) around or below 1 M^{-1}.

The change of NMR shifts upon complexation in suitable host compounds has been an invaluable tool in the development of supramolecular chemistry [5]. The concentration dependence of the shift changes allows not only to extract binding constants K by appropriate non-linear fitting methods, but also to evaluate the intrinsic complexation induced shift (CIS) for 100% complexation, which can shed light on the geometry of the complexes. The major advantage of NMR as compared to other methods, such as optical spectroscopy, is that several independent signals can be analyzed at the same time, thereby providing more reliable information on both binding affinity and binding modes. In view of these advantages it is surprising that NMR titrations have scarcely been used for the analysis of DNA–ligand complexes. Although NMR has been applied to DNA complexes already in 1972 by Gabbay et al. [6], it became clear that the conclusions from this early work with respect to the claimed intercalation of phenyl derivatives are questionable [7]. Later, Wilson et al. [8] have shown that true intercalators exhibit shielding effects, by the anisotropy of the nucleobases, of up to 1 ppm [9] in some cases, and simultaneous line broadening. In the course of the present study we show for the first time that both NMR shifts and line width can be used in a *quantitative* way for titrations of many ligands even to native calf thymus (CT) DNA.

2. Materials and methods

Calf thymus DNA was purchased from Aldrich and used as sodium salt. It was sonicated according to literature procedures. Solutions were made in D_2O containing 20 mM sodium phosphate, and the pD was adjusted to a value of 7.4. In the cases of L-tryptophane (**1d**) and gramine (**1b**), stock solutions were prepared in DMSO-d_6 because of solubility problems. Titrations were performed in those cases with a constant DMSO content of 10 vol%. This does not influence the stability and conformation of the used duplex in a significant way, as was checked by UV melting studies.

L-Tryptophane and gramine were purchased from Aldrich, tryptamine hydrochloride from Fluka, the peptides Trp-Ala and Trp-Gly-Gly from Sigma, Lys-Trp, Lys-Trp-Lys and Lys-Gly-Trp-Lys as acetates from Bachem; these were used without further purification.

^{1}H-NMR spectra were recorded with a Bruker AM400 (Trp-Ala) or DRX500 spectrometer system (in all other cases). Chemical shifts were measured with respect to an external reference (DSS in D_2O).

Non-linear fitting of the results were executed with the program SIGMAPLOT 5.0 from Jandel Scientific using equations described earlier [10]. The equations for the simulation of the Scatchard plots without and with cooperativity are also described in literature [11].

3. Results and discussion

Fig. 2 demonstrates that the presence of up to 5 mM sonicated CT DNA (base pair concentration) leads with some potential intercalators like tryptophane (**1d**) or the dipeptide Lys-Trp (**2c**) only to shielding effects of up to 0.06 ppm, and to a maximal line broadening of 3–5 Hz. Similar results are obtained with e.g. the peptides **2a** and **2b**, and with many other ligands like phenyl compounds, which are known not to intercalate. In sharp contrast, one observes shielding on *all* indole protons with the indole derivatives tryptamine (**1c**) and gramine (**1b**), but also with the tripeptide Lys-Trp-Lys (**2d**), and the tetrapeptide Lys-Gly-Trp-Lys (**2e**); simultaneously these signals are

*Corresponding author. Fax: (49) (681) 302-4105 or (49) (681) 302-3209.

SSDI 0014-5793(95)01158-7

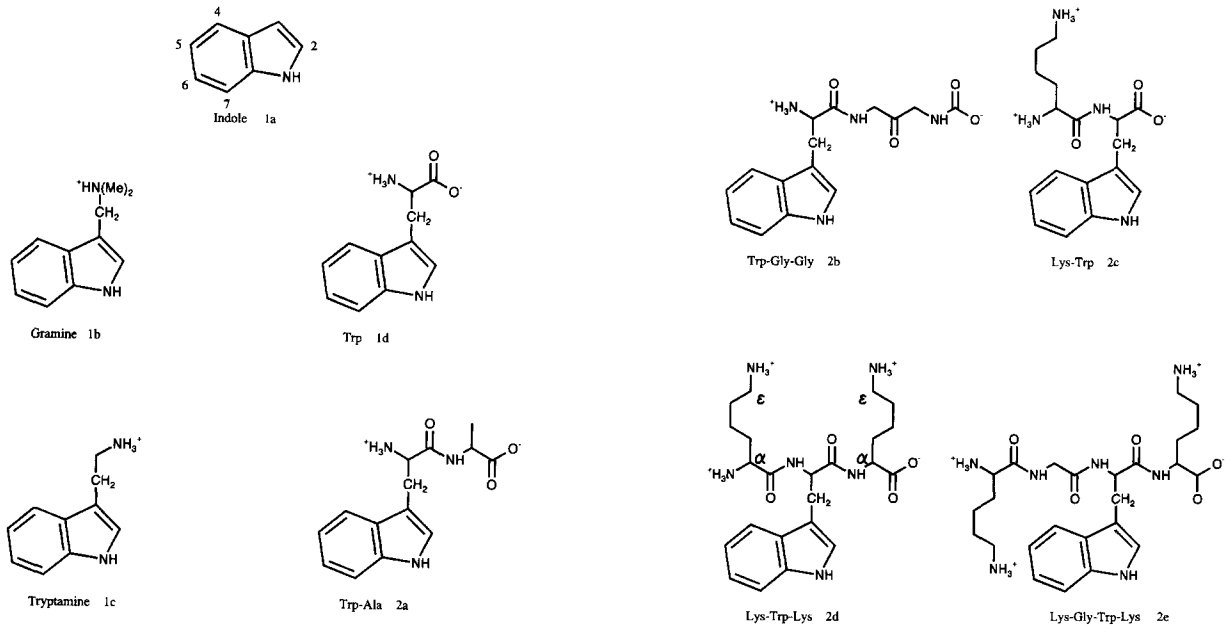

Fig. 1. Structures of ligands.

broadened to up to 25 Hz, or 15 Hz with 5 mM DNA (Fig. 2). Remarkably, the α-CH protons at the N-terminal lysine residues in the tri- and tetrapeptide show **de**shielding by up to 0.2 ppm, without line broadening beyond 5 Hz. This behaviour is atypical for compounds which in this case bind definitely by intercalation; a similar influence of the groove environment on the protonation state of the N-terminal part of DNA bound peptides is also observed for non-intercalating (**2a**) and (**2b**) (0.05 and 0.09 ppm, respectively) with and without a lysine unit (see Fig. 1). A significant lower downfield shift (0.08 ppm, compared to **2d** and **2e**) has upon interaction of Lys-Trp-Lys and **single** stranded poly(A) been reported by Helène [3e], and has been attributed to a pK change for the α-NH_2 of the lysine unit from around 8 up to 8.5.

Evaluation of the NMR shifts according to established procedures [10] by non-linear least square fitting yielded for the first time gratifying consistent values for the equilibrium constants K derived from up to 6 different signals (Fig. 3). The K and CIS values were derived on the basis of a 4:1 stochiometry counting 4 base pairs per ligand at 'saturation', in adherence to the nearest neighbour exclusion principle [12]. The K values decrease by a factor of 2–3 if other models, based on a non-cooperative binding of the ligand to the base pairs or the phosphates were used (stochiometries then bp/ligand = 2/1 or 1/2, respectively; Fig. 3). Noticably, the CIS values change only by less than 10% on the average with the calculational model.

It also was possible for the first time to determine the binding affinity of a groove binder via NMR titration using the downfield shift described above (Fig. 4). Scatchard plots for **2d** instead of non-linear fitting showed, in contradiction to the results by Helène [3e] with related compounds, a strong curvature at very low values for r (ca. 0.1; Fig. 5) which indicates 'neighbour exclusion' and a large negative cooperativity. Because of the small number of experimental points the values for the formal association constant K, the exclusion parameter n and the cooperativity factor w were not fitted. They could, however, at least roughly be estimated by comparison with theoretically derived Scatchard plots [7a,11,12] as being in the

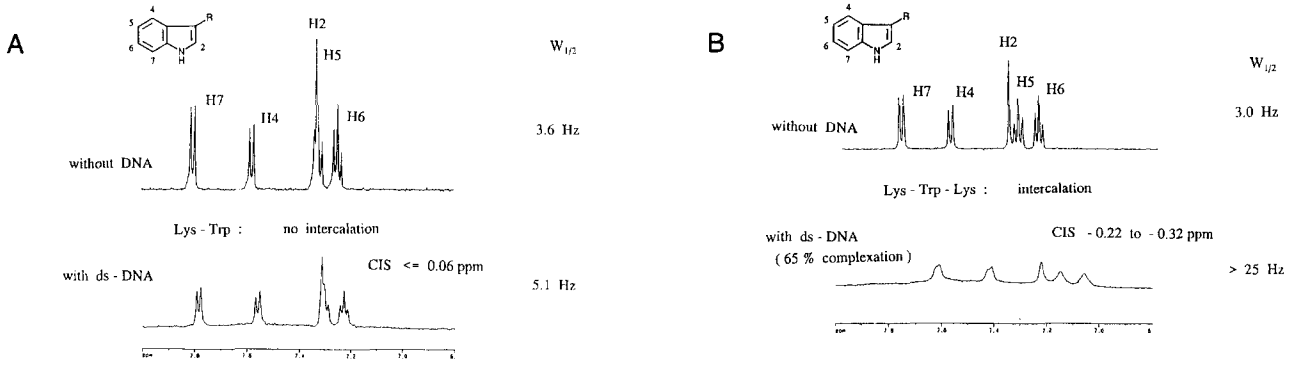

Fig. 2. (a) ^{1}H-NMR spectra of the indole part of ligand **2c** (non-intercalating) (a) without DNA; [**2c**] = 0.42 mM (b) with sonicated CT-DNA; [**2c**] = 0.23 mM/7.5 mM; ($W_{1/2}$ = half-height width of the H2 signal; in [Hz]). (b) ^{1}H-NMR spectra of the indole part of ligands **2d** (intercalating) (a) without DNA; [**2d**] = 0.86 mM (b) with sonicated CT-DNA; [**2d**] = 0.47 mM/[bp] = 4.0 mM; ($W_{1/2}$ = as in Fig. 2a: complexation degree calculated on the basis of next neighbour exclusion principle ([bp]/[ligand] = 4/1)).

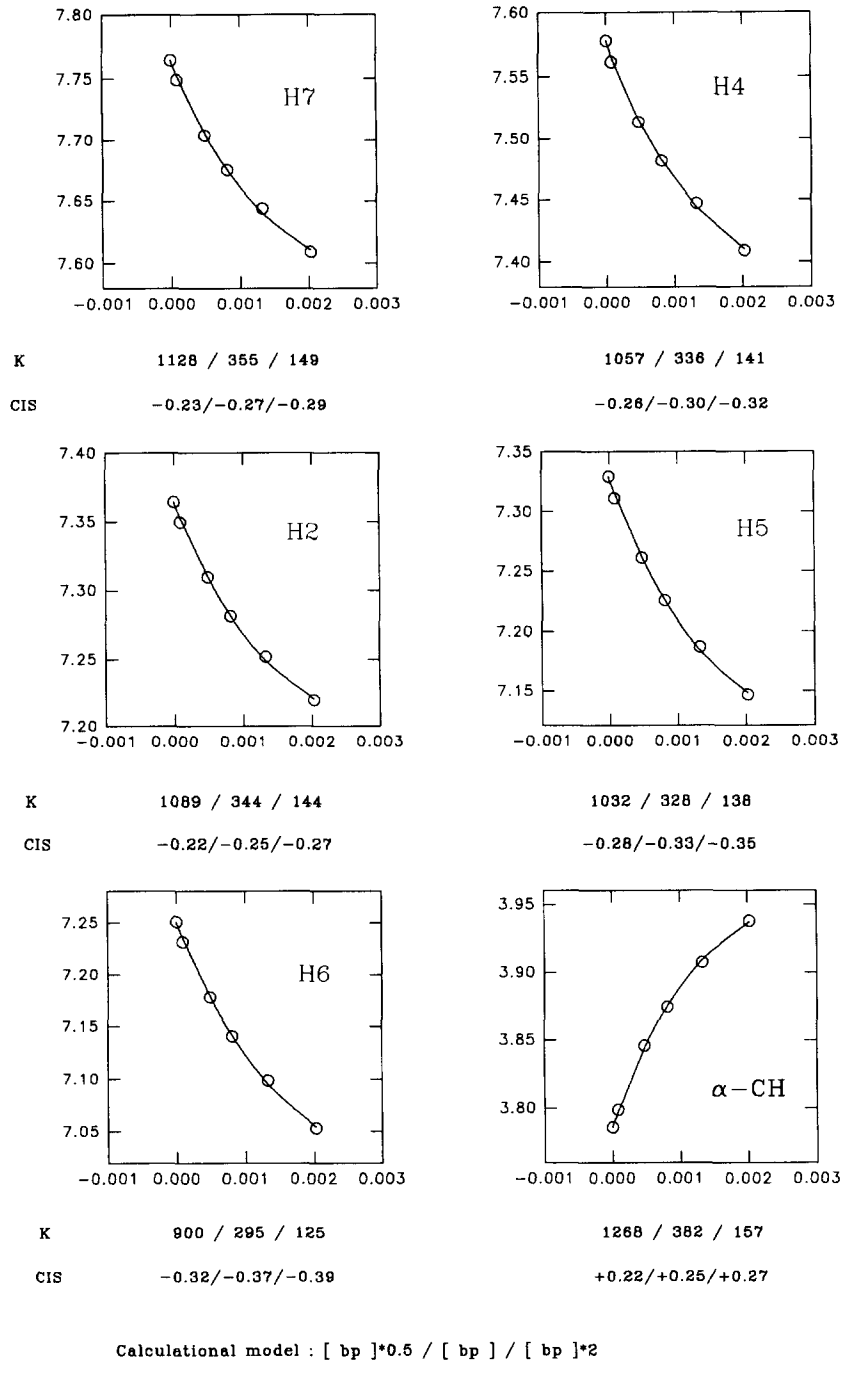

Fig. 3. Plots for determination of the formal association constant K: between ligand **2d** and sonicated CT-DNA for five different protons of the indole part and the α-CH of the N-terminal lysine on the basis of next neighbour exclusion/without neighbour exclusion/with independent behavior of each phosphate. Sequence of protons (from upper left to lower right plot): H7/H4/H2/H5/H6/α-CH (numbering scheme see Fig. 1); measuring conditions see Table 1. CIS = complexation induced shift (for 100% complexation, from line fit), in [ppm].

range of $r = 0.0$–0.3 (Fig. 5). The intercept gives a constant of $K = 293$, comparable to the value obtained for the model without neighbour exclusion (Fig. 3). For the exclusion parameter n it is reasonable to assume a value ≥ 4, which is in the range of 'nearest neighbour exclusion' model or a little above (on the basis of DNA phosphate concentration). According to theory, it is necessary to add a cooperativity term when there is a contribution of electrostatic interactions. The factor w was then estimated to be <0.5 (Fig. 5). The curvature itself indicates as usual the mode of cooperativity.

The equilibrium constants (Table 1), evaluated at minimal ionic strength, roughly agree with literature data [3e] as far as these are available. The K value derived from the N-terminal lysine α-CH proton, which is located closer to the groove than to the base pairs, together with always observed fit to a simple 1:1 equilibrium strongly suggests that these peptides bind to

DNA not in 2 discrete modes, characterized by 2 K values as described by Helène et al. [4] for a related tetrapeptide. The free energies of complexation ΔG also support a complex stabilization by the simultaneous but independent contribution of both electrostatic groove, and intercalative binding. If one attributes the often found [5b,c] value of the 5 kJ/mol and salt bridge to the interaction of one excess positive charge present in the lysine residues (as the terminal $^{+}NH_3$ and COO^{-} charges may cancel), one has to assume a more-or-less negligible stacking contribution of the indole ring. This small value is significantly lower than for related weak intercalators of the naphthalene or quinoline type [13]. The weak stacking interaction of the indole ring is not high enough to destack the nucleobases, if there is no simultaneous help from at least 2 basic amino acid residues. This contrasts to the minimal requirements we found for the above-mentioned naphthalene-type intercalators [13] where *one* charge in the side chain is sufficient. Our findings are at variance with the conclusions of Helène [3] and Gabbay [6b] et al., claiming tryptophane intercalation into ds-DNA of dipeptides and tripeptides possessing no, or only one charge. Noticeably, the constants extracted by Helène et al. [4] for the second (intercalation) step are of the same, almost negligible, size as ours, which would explain why one needed the extremely sensitive fluorescence method to detect them. The observed CIS values are in magnitude and sign similar to other intercalators, and are in line with normal intercalation modes.

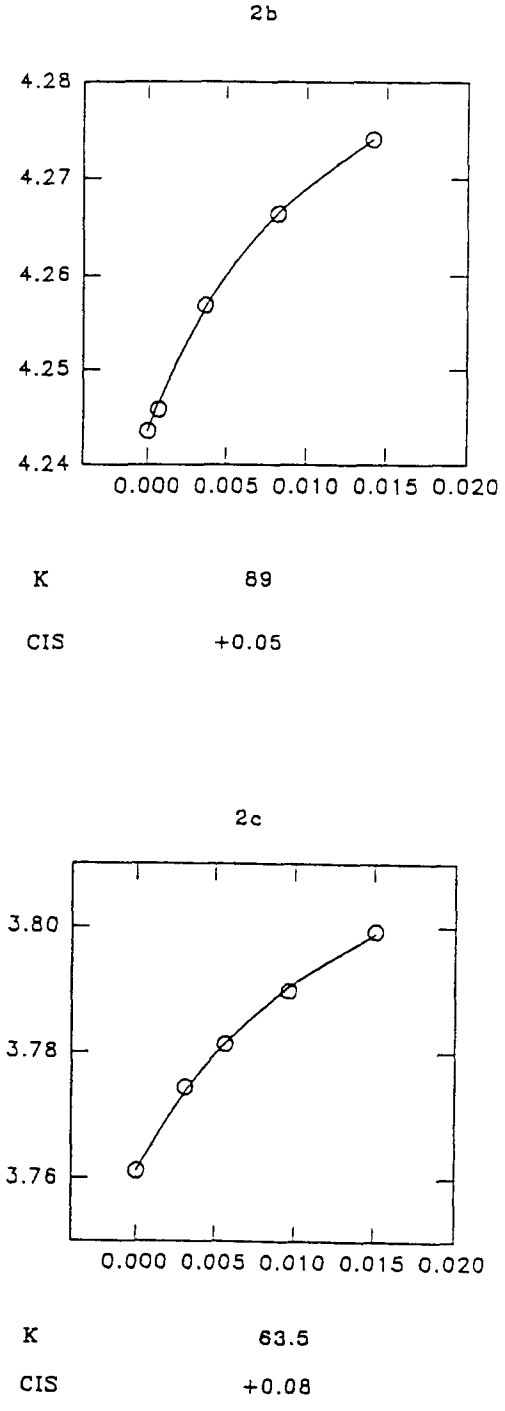

Fig. 4. Plots for determination of the formal association constant K between ligands **2b** and **2c** and sonicated CT-DNA using the α-CH proton of the N-terminal amino acids on the basis of independent binding of each phosphate;for conditions see Table 1; CIS: see Fig. 3.

A possible explanation for the differences in CIS values with **1b** and **1c** in contrast to **2d** and **2e** could be the existence of a second, non-intercalating binding mode being in fast exchange with the intercalative mode. Although this mechanism, suggested by one referee, cannot entirely dismissed, we favour another rationalization based on the substantial difference in the conformational freedom of the indole ring which is much more limited in the peptides by the need of two flanking basic aminoacid residues with a moreover more restricted mobility due to the linking amide bonds. The size and limited flexibility of the peptides may lead here to preferred intercalation from the major groove, leading to less deep intercalation compared to the amines **1b** and **1c**.

The line width $W_{1/2}$ also increases as the shifts do with increasing DNA concentrations (see Fig. 2). The overlapping signals make it possible to reach the non-linear part of the corresponding titration curve only in few cases (Fig. 6). The K value derived from curve fitting of the line width is not far from the one obtained from the shift changes (Fig. 3), in particular if one corrects for the small line broadening which is also produced by pure groove binders. Thus, the change in line width may also be used for DNA titrations, in spite of the difficulties in determining half-widths of broad and overlapping signals. That the origin of the line broadening is the slow tumbling rate of the bound ligand is obvious from the very similar line width for *all* protons at or close to the intercalating aromatic unit. If exchange processes being slow on the NMR time scale would be responsible one would expect a dependence of $W_{1/2}$ on the CIS values, which is not seen. The results demonstrate that $W_{1/2}$ values can be used efficiently to establish intercalation. One can even calculate the 'intrinsic' final line width also in the often occurring situation of heavy overlap by measuring $W_{1/2}$ at lower degrees of complexation (and hence less overlap), by dividing the experimentally observed $W_{1/2}$ by the mole fraction of complexed ligand, as long as one knows the equilibrium constant K from other sources.

4. Conclusions

We have demonstrated that NMR shift titrations can, on the basis of analyzing up to 6 different signals, provide accurate binding constants and intrinsic shift values characterizing the binding geometry, using for the first time a large biopolymer

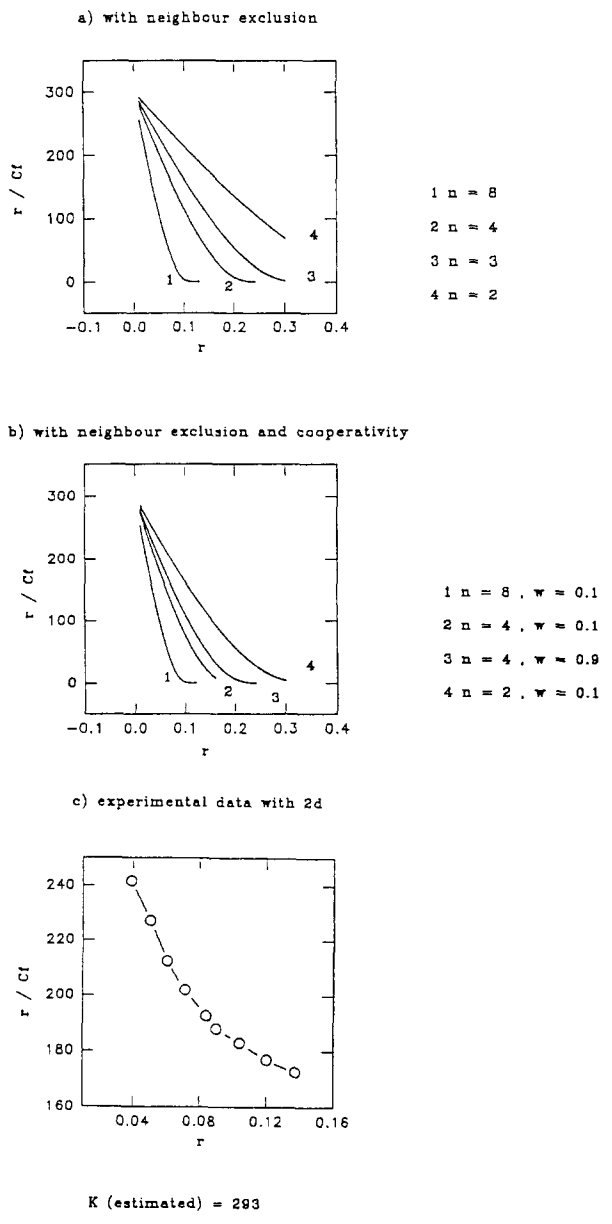

Fig. 5. Scatchard plots for ligand **2d** and sonicated CT-DNA (a) simulated with neighbour exclusion: (1) with $n = 8$; (2) with $n = 4$; (3) with $n = 3$; (4) with $n = 2$; (b) simulated with neighbour exclusion and cooperativity: (1) with $n = 8$ and $w = 0.1$; (2) with $n = 4$ and $w = 0.1$; (3) with $n = 4$ and $w = 0.9$; (4) with $n = 2$ and $w = 0.1$; (c) calculated using experimental data for **2d**; K estimated for the linear part of the plot: $K = 293\ M^{-1}$.

such as calf thymus DNA as host structure. The CIS values as well as the line width are reliable indicators of intercalation. The $W_{1/2}$ values can be measured at low degrees of complexation and then converted to the intrinsic values on the basis of known equilibrium constants. The perfect fit of the titration curves with the tri- and tetrapeptide containing two lysine residues to a model with neighbour exclusion (stochiometry bp/ligand = 4/1), including signals from protons remote or close to the groove, as well as the observed association constants are in line with a simple binding mechanism, in which both electrostatic groove binding with 5 kJ/mol and charge [5b], and stacking with about 1–2 kJ for the indole unit contribute as additive increments at independent binding sites. Tryptamine and gramine, which contain a supporting positive charge at a location suitable for simultaneous groove binding therefore intercalate at low ionic strength, in contrast to the dipeptide Lys-Trp with

Table 1(a)
H-NMR shifts induced by complexation with DNA (CIS) and equilibrium constants K[a]

	Gramine		Tryptamine		Trp	Lys-Trp	
	cis[a,f]	K	cis[a]	K	$\Delta\upsilon$[c,f]	$\delta\upsilon$[c,g]	
Trp2	−0.67	49.5	−0.40	302	−0.01	−0.042	
4	−0.70	50	−0.30	316	−0.013	−0.06	
5	−0.63	56	−0.29	359	−0.012	−0.06	
6	−1.30	29	−0.40	294	−0.013	−0.06	
7	−0.73	46.5	−0.32	330	−0.013	−0.054	
$_\alpha$-CH	–	–	–	–		–[b]	
CH_2	–	–	–	–		−0.013	
1-Lys $_\alpha$-CH	–	–	–	–		+0.08[a]	63[a]
CH_2(1)	−0.08[c,d]	–	−0.07[c,e]	–		–	
CH_2(2)	–	–	−0.012[c,e]	–		–	
CH_3	−0.03[c,d]	–	–	–		–	

[a] CIS in ppm, K in M^{-1} units; in D_2O at pD = 7.4. [Na-phosphate] = 20 mM; 303 ± 0.1 K, with [gramine] = 0.42 mM, [tryptamine] = 0.17 mM, [trp] = 0.4 mM, [Lys-Trp] = 0.23 mM at the end of titration; CT-DNA from [bp] = 0 mM up to [bp] = 9.6 mM resp. 10 mM resp. 7 mM resp. 7.5 mM.
[b] Not determined because of too large overlapping and high multiplicity.
[c] [ppm], shift difference observed at highest DNA concentration used.
[d] [bp] = 9.6 mM.
[e] [bp] = 10 mM.
[f] 10 vol% DMSO-d_6.
[g] [bp] = 7.5 mM.

Table 1(b)
H-NMR shifts induced by complexation with DNA (CIS) and equilibrium constants K[a]

	Lys-Trp-Lys		Lys-Gly-Trp-Lys		Trp-Ala	Trp-Gly-Gly	
	cis[a]	K	cis[a]	K	$\Delta\upsilon$[c,f]	$\delta\upsilon$[c,g]	
Trp2	−0.22	1.09	−0.10	0.68	−0.004	−0.01	
4	−0.26	1.06	−0.12	0.55	−0.01	−0.008	
5	−0.28	1.03	−0.13	0.66	−0.01	−0.004	
6	−0.32	0.90	−0.08	0.68	−0.01	−0.009	
7	−0.23	1.13	−0.10	0.67	−0.01	−0.01	
$_\alpha$-CH	−0.12	–	–			+0.03	
						+0.05[a]	0.089[a]
CH_2	–[b]		–[b]			+0.011	
1-Lys $_\alpha$-CH	+0.22	1.27	+0.20	0.71			
$_\varepsilon$-CH_2	(0.00)[c,d]	–	(0.00)[c,e]				
					Ala:	*Gly:*	
2-Lys $_\alpha$-CH	(0.002)[c,d]	–	(−0.001)[c,e]		CH-0.005	CH_2+0.009/ +0.03	
$_\varepsilon$-CH_2	(0.00)[c,d]	–	(0.00)[c,e]		CH_3 0.00	+0.009/ +0.009	

[a] CIS in ppm, K in $10^{-3}\ M^{-1}$ units; in D_2O at pD = 7.4 [Na-phosphate] = 20 mM; 303 ± 0.1 K, with [Lys-Trp-Lys] = 0.47 mM, [Lys-Gly-Trp-Lys] = 0.21 mM, [Trp-Ala] = 0.6 mM, [Trp-Gly-Gly] = 0.46 mM; CT-DNA from [bp] = 0 mM up to [bp] = 4.0 mM resp. 5.1 mM resp. 3.3 mM resp. 7.1 mM.
[b] Not determined because of too large overlapping and high multiplicity.
[c] [ppm], shift difference observed at highest DNA concentration used.
[d] [bp] = 4.0 mM.
[e] [bp] = 5.1 mM.
[f] [bp] = 3.3 mM.
[g] [bp] = 7.1 mM.

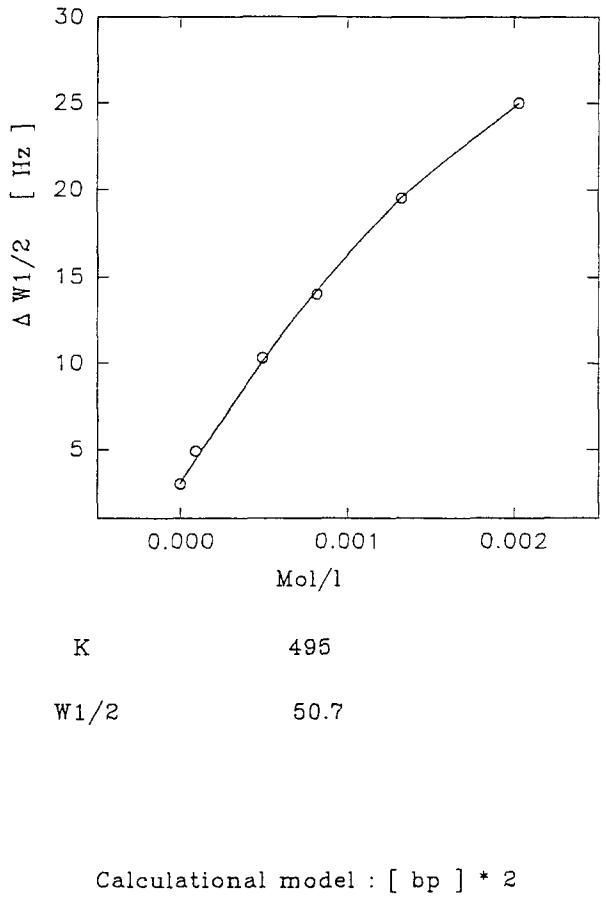

Fig. 6. Plot for the determination of the formal association constant K for ligand **2d** and sonicated CT-DNA using the half-width of the H2 signal of the indole part (see Fig. 1), on the basis of next neighbour exclusion; $W_{1/2}$ = half-width values obtained for 100% complexation.

the additional amide function and the peptides without basic residues. The difference obviously lies in the number and position of the positively charged $^{+}NH_3$-groups which only in the flexible lysine side chain can approach the groove phosphate groups closely enough for the formation of contact ion pairs. Amide functions of neutral aminoacids contribute not sufficiently to binding of the very weak intercalator indole.

Similar limitations of a simultaneous full use of contact at two binding sites, and hence of the full chelate effect have been seen in the binding of aromatic anions in azoniacyclophanes as host compounds. The binding constants in these supramolecular complexes dropped considerably if the spacing between the van der Waals and the ionic contact site was insufficient [14].

Acknowledgements: Our work is supported by the Deutsche Forschungsgemeinschaft, Bonn, and the Fonds der Chemischen Industrie, Frankfurt.

References

[1] Pullman, B. and Jortner, J. (1990) in: Molecular Basis of Specificity in Nucleic Acid–Drug Interactions; The Jerusalem Symposia on Quantum Chemistry and Biochemistry, Vol. 23, Kluwer Academic Publishers, Dordrecht.
[2] Cantor, C.R. and Schimmel, P.R. (1980) in: Biophysical Chemistry of Macromolecules, Part. 3: The Behavior of Biological Macromolecules, W.H. Freeman, San Francisco.
[3] (a) Helène, C. and Maurizot, J.C. (1981) CRC Crit. Rev. Biochem. 10, 213; (b) Helène, C., Montenay-Garestier, T. and Dimicoli, J.-L. (1971) Biochim. Biophys. Acta 254, 349; (c) Brun, F., Toulme, J.-J. and Helène, C. (1975) Biochemistry 14, 558; (d) Helène, C., Dimicoli, J.-L. and Brun, F. (1971) Biochemistry 10, 3802; (e) Dimicoli, J.-L. and Helène, C. (1974) Biochemistry 13, 714.
[4] Rajeswari, M.R., Montenay-Garestier, T. and Helène, C. (1987) Biochemistry 26, 6825; and references cited therein.
[5] (a) Schneider, H.-J., Kramer, R., Simova, S. and Schneider, U. (1988) J. Am. Chem. Soc. 110, 6442; (b) Schneider, H.-J. (1991) Angew. Chem., Int. Ed. Engl. 30, 1417; (c) (1993) Rec. Trav. Chim. Pays-Bas 112, 412 and references cited therein.
[6] (a) Kapicak, L. and Gabbay, E.J. (1975) J. Am. Chem. Soc. 97, 403; (b) Gabbay, E.J., Sanford, K. and Baxter, C.S. (1972) Biochemistry 11, 3429.
[7] (a) Wittingham, M.S. and Jacobsen, A.J. (1982) in: Intercalation Chemistry, Chap. 14, Academic Press, New York; (b) Jacobsen, A.J. and Wang, J.C. (1973) Biochim. Biophys. Acta 335, 49; (c) Quadrifoglio, F., Giancotti, V. and Creszenzi, V. (1976) FEBS Lett. 65, 345.
[8] Chandrasekaran, S., Kusuma, S., Boykin, D.W. and Wilson, W.D. (1986) Magn. Res. Chem. 24, 630.
[9] Strekowski, L., Mokrosz, J.L., Wilson, W.D., Mokrosz, M.J. and Strekowski, A. (1992) Biochemistry 31, 10802.
[10] Wilcox, C.S. (1991) in: Frontiers on Supramolecular Organic Chemistry and Photochemistry; Schneider, H.-J. and Dürr, H. (Eds.), pp. 123–143, VCH Weinheim, similar methods: ref. 5a.
[11] (a) Wilson, W.D. and Lopp, I.G. (1979) Biopolymers 18, 3025; (b) Bloomfield, V.A. and Crothers, D.M. (1974) in: Physical Chemistry of Nucleic Acids, Chap. 7, Harper, New York.
[12] McGhee, J.D. and von Hippel, P.H. (1974) J. Mol. Biol. 86, 469.
[13] Sartorius, J. and Schneider, H.-J. (unpublished results).
[14] (a) Schneider, H.-J. and Theis, I. (1989) Angew. Chem., Int. Ed. Engl. 28, 753; (b) Schneider, H.-J., Blatter, T., Palm, B., Pfingstag, U., Rüdiger, V. and Theis, I. (1992) J. Am. Chem. Soc. 114, 7704.