# The Self-Assembly of a Lipophilic Guanosine Nucleoside into Polymeric Columnar Aggregates: The Nucleoside Structure Contains Sufficient Information To Drive the Process towards a Strikingly Regular Polymer

**Elisabetta Mezzina,[a] Paolo Mariani,[b] Rosangela Itri,[c] Stefano Masiero,[a] Silvia Pieraccini,[a] Gian Piero Spada,[a] Francesco Spinozzi,[b] Jeffery T. Davis,[d] and Giovanni Gottarelli*[a]**

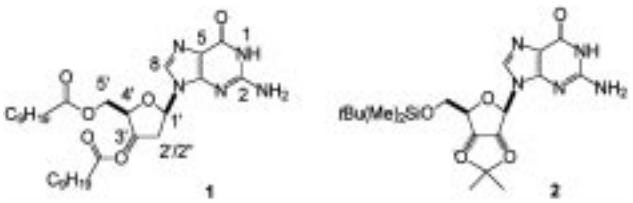
*In memory of André Collet (1945–1999)*

**Abstract:** Lipophilic guanosine derivatives act as self-assembled ionophores. In the presence of alkali metal ions in organic solvents, these G derivatives can form tubular polymeric structures. The molecular aggregates formed by 3′,5′-didecanoyl-2′-deoxyguanosine (**1**) have been characterised by SANS and NMR spectroscopy. The polymer is structured as a pile of stacked G quartets held together by the alkali metal ions that occupy the column's central channel. The deoxyribose moieties, with their alkyl substituents, surround the stacked G quartets, and the nucleoside's long-chain alkyl tails are in intimate contact with the organic solvent. In this polymeric structure, there is an amazing regularity in the rotamers around the glycosidic bond within each G quartet and in the repeat sequence of the G quartets along the columns. In hydrocarbon solvents, these columnar aggregates form lyomesophases of the cholesteric and hexagonal types.

**Keywords:** guanine · nucleobases · liquid crystals · neutron scattering · supramolecular chemistry

## Introduction

Guanylic acids and guanine-rich oligonucleotides form liquid crystalline phases of both the cholesteric and hexagonal types in water.[1] The building blocks for these phases are columnar aggregates based on the guanine quartet (G quartet). The G quartet is formed by the self-assembly of four guanine bases, a process that is templated by alkali metal ions such as $K^+$. We recently reported that lipophilic deoxyguanosine derivatives are able to extract alkali metal ions from water into an organic solvent. This phase transfer of salt occurs via the formation of supramolecular complexes wherein each cation is surrounded by two lipophilic G quartets and coordinated to 8 oxygen atoms (Figure 1). Depending on the experimental conditions, either discrete octamers, $G_8-K^+$, or polymeric aggregates, $(G)_n-(K^+)_m$, can be formed upon coordination of $K^+$ ions by the G quartets.[2–4] The structure of the $K^+$-bound octamer, $\mathbf{1}_8 \cdot$KPic, formed by 3′,5′-didecanoyl-2′-deoxyguanosine **1** and



$K^+$ picrate (KPic) in $CHCl_3$ has been characterised by NMR spectroscopy. In solution, the two G quartets that sandwich the $K^+$ cation adopt a twisted head-to-tail arrangement.[5] Furthermore, the conformations around the glycosidic bonds within the octamer, $\mathbf{1}_8 \cdot$KPic, are quite unique: one G quartet has an unusual all-*syn* conformation, while the glycosidic bonds in the other G quartet are all of the *anti* type.[3]

[a] Prof. G. Gottarelli, Dr. E. Mezzina, Dr. S. Masiero, Dr. S. Pieraccini, Prof. G. P. Spada
Dipartimento di Chimica Organica "A. Mangini"
Università di Bologna
Via S. Donato 15, 40127 Bologna (Italy)
Fax: (+39) 5124-3218
E-mail: gottarel@alma.unibo.it

[b] Prof. P. Mariani, Dr. F. Spinozzi
Istituto di Scienze Fisiche, Università di Ancona
and Istituto Nazionale per la Fisica della Materia
Via Ranieri 65, 60131 Ancona (Italy)

[c] Dr. R. Itri
Instituto de Fisica da Universidade de Sao Paulo
Caixa Postal 66318, Sao Paulo, SP 05315-970 (Brazil)

[d] Prof. J. T. Davis
Department of Chemistry and Biochemistry, University of Maryland
College Park, MD 20742 (USA)



0947-6539/01/0702-0388 $ 17.50+.50/0

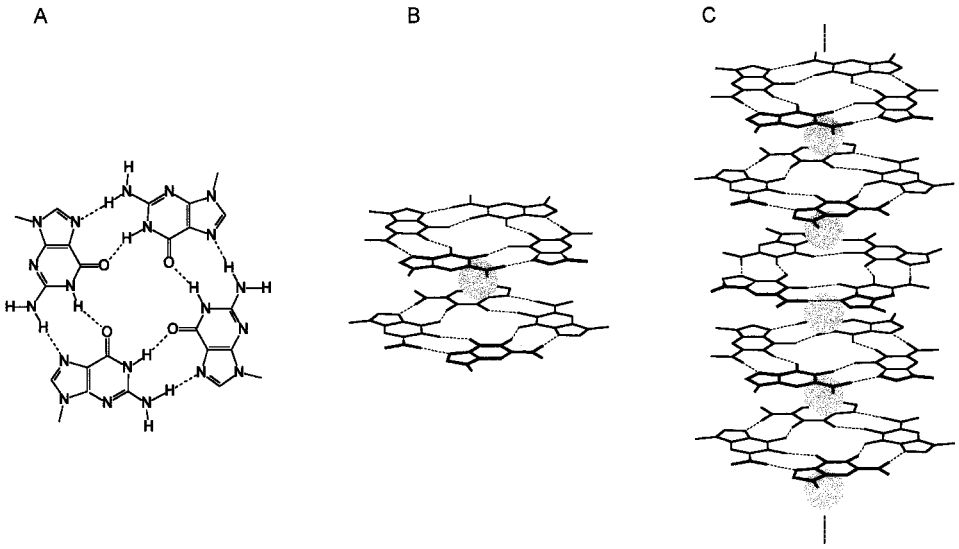



Figure 1. The G quartet (A) and a sketch of the octameric $G_8 \cdot K^+$ (B) and polymeric $(G_4 \cdot K^+)_n$ (C) aggregates.

The crystal structure of the $K^+$ picrate complex of G derivative **2** confirms what was found in solution for the dG derivative **1**. In the solid state, a $K^+$ cation is sandwiched between two G quartets that are arranged in a head-to-tail orientation. Again, the glycosidic torsion angles are homogeneous within each G quartet.[6] Thus, both solution and solid-state studies indicate that discrete $G_8$ octamers can be formed in a diastereospecific manner. As described below, these lipophilic G derivatives can form even higher ordered polymeric structures in organic solvents.

Self-assembly of low molecular weight building blocks into noncovalent, polymeric nanostructures has attracted considerable interest. This attention has been primarily driven by the demonstration that self-assembly is ideally suited for the construction of devices on a molecular scale.[7, 8] As mentioned above, the G-quartet motif is particularly attractive for self-assembly. One of our continuing goals is to utilise the self-assembly of lipophilic G derivatives to build functional nanostructures. Our recent findings have been encouraging. For example, cation-induced aggregation of lipophilic G derivatives gives soluble polymers that display enantioselective discrimination towards binding chiral anions.[4] Also, stacked lipophilic G quartets formed from G **2** in organic solvents constitute a simple and attractive model for a self-assembled ion channel.[6] Finally, we have shown that lipophilic G derivatives can also form unusual chiral liquid crystals, again based on the columnar stacking of individual G quartets.

Because of their potential importance in providing new nanostructures, it is crucial that we develop strategies for characterising the structure of these G-aggregates. In contrast to our detailed NMR and X-ray studies on the smaller $G_8-K^+$ octamers, structural characterisation of the organic-soluble G-quartet polymers has relied solely on the 4:1 G to $K^+$ picrate stoichiometry observed after alkali metal picrate extraction by the lipophilic G derivatives. In this paper, we report the characterisation of G-quartet aggregates using both small-angle neutron scattering (SANS) and NMR spectroscopy. Furthermore, the formation of cholesteric and hexagonal lyomesophases by **1** is observed in hydrocarbon solvents.

**Abstract in Italian:** *Derivati lipofili della guanosina agiscono come ionofori autoassemblanti. In presenza di ioni di metalli alcalini questi derivati possono formare, in solventi organici, strutture polimeriche tubulari. Gli aggregati molecolari formati dalla 3′,5′-didecanoil-2′-desossiguanosina **1** sono stati caratterizzati mediante SANS e spettroscopia NMR. La struttura del polimero è costituita da una pila di G-quartetti sovrapposti tenuta assieme dagli ioni dei metalli alcalini che occupano il canale centrale della colonna. Le unità di desossiribosio, con i loro sostituenti alchilici, circondano i G-quartetti impilati e le code alchiliche a lunga catena del nucleoside entrano in intimo contatto con il solvente organico. In questa struttura polimerica c'è una impressionante regolarità di distribuzione dei rotameri (attorno al legame glicosidico) all'interno di ogni singolo G-quartetto e nella sequenza di ripetizione dei G-quartetti lungo le colonne. In solventi idrocarburici, questi aggregati colonnari formano liomesofasi di tipo colesterico ed esagonale.*

## Results and Discussion

**SANS study**: Two different guanosine samples, obtained upon $CDCl_3$ extraction of potassium picrate from aqueous solutions of different salt concentrations, were studied by SANS (the details of the SANS data treatment are described in the Appendix). For the two samples, the nominal molar ratios of G **1** and $K^+$ picrate in the organic phase, [G]/[K], were about 8:1 and 4:1, respectively. The 8:1 ratio corresponds to 1 $K^+$ ion per 2 G quartets for the first sample, and the 4:1 ratio corresponds to 1 $K^+$ ion per G quartet for the second sample. Under the same extraction conditions, previous $^1$H NMR measurements in chlorinated organic solvents showed either the octamer, $\mathbf{1}_8 \cdot$ KPic, made by two assembled G quartets, or the polymer $(\mathbf{1}_4 \cdot \text{KPic})_n$, made by repeated G-quartet stacking. The relative ratios of octamer $\mathbf{1}_8 \cdot$ KPic and polymer $(\mathbf{1}_4 \cdot \text{KPic})_n$ depended on the amount of $K^+$ that was extracted from the aqueous solution into the organic phase.[2, 3]

Figure 2 shows the SANS profiles obtained from the samples containing the smaller ([G]/[K] = 8, Figure 2a) and the higher ([G]/[K] = 4, Figure 2b) amounts of KPic in solution. Both profiles show an intense signal characterised by a pronounced and well-defined second peak around $Q = 0.24$ Å$^{-1}$ ($Q$ is the scattering vector, see Experimental Section).

The SANS profiles were analysed by means of the global fitting procedure discussed in the Appendix,[9, 10] by considering as particle models cylinders of length $L$ with a central core

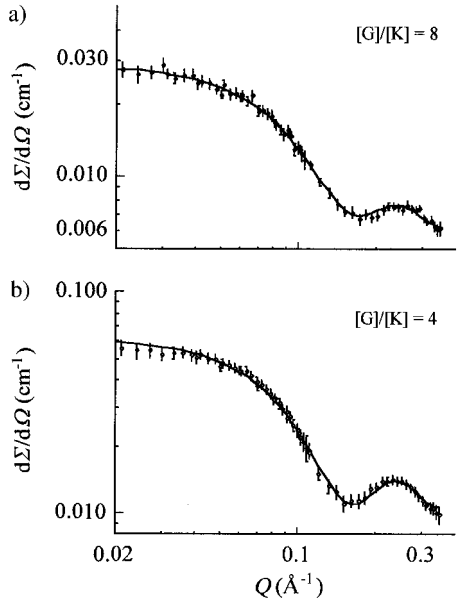



Figure 2. Small angle neutron scattering from the samples containing a) 1 $K^+$ ion per 2 G quartets and b) 1 $K^+$ ion per G quartet. The lines correspond to the best fit to the data (see text for details).

of radius $R_{in}$ and scattering length density $\rho_{in}$, and a concentric shell of thickness $R_{ext} - R_{in}$ of different scattering density $\rho_{ext}$ [see Equation (4)].

In particular, best fits were obtained considering the coexistence of two families of cylindrical particles, with the same two-level scattering densities and radial dimensions, but different lengths $L$: short cylinders (oligomers, with $\nu_o < 20$) and freely rotating in the space, and long cylinders (polymers, with $\nu_p > 20$) preferentially oriented in a plane parallel to the cell walls (see Appendix).

The internal and external radii obtained from the global fit are $R_{in} = 12.6 \pm 0.5$ Å, $R_{ext} = 20.1 \pm 0.5$ Å, in good agreement with the dimensions expected for the G quartet's molecular structure, while the unique concentration results $N_{G\ quartet} = (4.4 \pm 0.2) \times 10^{16}$ $cm^{-3}$.

For the sample that nominally contains 1 $K^+$ ion per 2 G quartets, the best fit parameters show that the fraction of quartets combined as short oligomers is $f = 0.87 \pm 0.06$. This $f$ value means that, under these conditions, approximately 85 % of the G derivative **1** is assembled into small oligomers, while only 15 % of the aggregate is in the polymeric form. The fitted maxima and the standard deviations of the gaussian distributions of lengths are $\nu_o = 2$, $\nu_p = 20$ and $\sigma_o = 3$, $\sigma_p = 8$, for the oligomeric and the polymeric forms, respectively. This result clearly indicates that the condition [G]/[K] = 8 leads to the preferential formation of short aggregates mainly composed of two G quartets, while the long cylinders, which form in small quantity, show a large polydispersity in length.

For the sample prepared with a nominal concentration of 1 $K^+$ ion per G quartet, the quality of the fit is found to be independent on the oligomer fraction in the range $f = 0 - 0.5$, while the fit becomes very bad at higher $f$ values. In practice, the data are very well reproduced considering that almost all the G quartets would be assembled as polymers. In this condition, the parameters of the gaussian length distribution are $\nu_p = 20$ and $\sigma_p = 1$, indicating that the polymeric forms have quite constant length.

The present SANS data then show the facile self-assembly of derivative **1** in $CDCl_3$ in the presence of KPic. The particles show dimensions that fully agree with the association of guanosine derivatives to give G quartets and the further formation of self-assembled species that are composed of at least two G quartets. Moreover, as was the case in our previous NMR studies,[2] a clear dependence of the self-assembly process on the amount of $K^+$ ions present in the organic solvent is evident from this SANS analysis. At the nominal concentration of 1 $K^+$ ion per 2 G quartets, self-assembly of the lipophilic guanosine derivative **1** preferentially leads to short oligomeric aggregates widely distributed as to length ($\nu_o = 2$, $\sigma_o = 3$). Polymers are also present in solution, but their concentration is quite low ($f = 0.87$). In contrast, when 1 $K^+$ ion per G quartet is present in solution, the formation of longer columns is favoured and, from SAXS data, there is no clear evidence that a fraction of G quartets is still assembled in short particles.

**NMR study**: NMR spectroscopy was also used to characterise the polymeric aggregate formed by G **1** and $K^+$ cations in organic solvents. Slow dissolution of potassium picrate into a $CDCl_3$ solution of **1** generates significant changes in the $^1H$ NMR spectrum. The presence of the alkali metal ion induces the splitting of each original signal into two different resonances. As described in our previous paper, the original NMR signals for G **1** gradually disappear and eventually result in the formation of a self-assembled octamer, $\mathbf{1}_8 \cdot$ KPic.[3] The octamer is stable in solution as long as the [G]/[K] ratio of 8:1 is maintained. However, with the decrease of the [G]/[K] ratio from 8:1 upon further dissolution of $K^+$ picrate, additional variations in the $^1H$ NMR spectrum are observed. Thus, upon addition of excess $K^+$ picrate to G **1** the NMR spectrum slowly changes over one week to give new sets of signals, indicating complete formation of a new G aggregate in solution. The stoichiometry of this new self-assembled species was determined as [G]/[K] = 4 by relative integration of the $^1H$ NMR signals for picrate and G **1**. Both the $^1H$ and $^{13}C$ NMR spectra for this new aggregated species showed three sets of signals in a 1:1:1 ratio (see Figure 3). Each NMR peak was assigned and correlated by using a combination of $^1H-^1H$ NOESY, $^1H-^1H$ TOCSY and $^1H-^{13}C$ HMBC data (Tables 1 and 2).

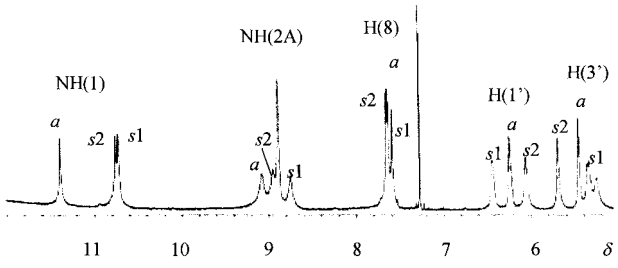



Figure 3. A region of the 400 MHz $^1H$ NMR spectrum at −5 °C of a 0.08 mol $L^{-1}$ solution of **1** in the presence of KPic (0.02 mol $L^{-1}$). Labels *a*, *s*1 and *s*2 refer to signals due to *anti*, *syn*1, and *syn*2 conformers, respectively (see text).



0947-6539/01/0702-0390 $ 17.50+.50/0 *Chem. Eur. J.* **2001**, *7*, No. 2

Table 1. $^1$H NMR chemical shifts for $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ in $CDCl_3$ at room temperature.

| $^1$H | NH(1) | NH(2)A | NH(2)B | H(8) | H(1′) | H(2′) | H(2″) | H(3′) | H(4′) | H(5′)/H(5″)[a] |
|---|---|---|---|---|---|---|---|---|---|---|
| *syn*1 | 10.73 | 8.70 | 4.05 | 7.60–7.62[b] | 6.43 | 2.92 | 2.18 | 5.37 | 4.31 | 4.07; 4.82–4.85[b] |
| *syn*2 | 10.72 | 8.93 | 4.86 | 7.67–7.69[b] | 6.06 | 3.60 | 2.64 | 5.71 | 4.23 | 3.99; 4.47 |
| *anti* | 11.36 | 9.00 | 4.76 | 7.64 | 6.23 | 2.68 | 2.41 | 5.46 | 4.11 | 4.15–4.20[b]; 4.41 |

[a] Not assigned. [b] For spectra measured at different times there is a small chemical shift difference.

Table 2. $^{13}$C NMR chemical shifts for $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ in $CDCl_3$ at room temperature.

| $^{13}$C | C(2)[a] | C(4) | C(5) | C(6)[a] | C(8) | C(1′) | C(2′) | C(3′) | C(4′) | C(5′) |
|---|---|---|---|---|---|---|---|---|---|---|
| *syn*1 | | 151.00 | 116.04 | | 139.50 | 85.61 | 37.39 | 73.18 | 81.60 | 65.88 |
| *syn*2 | | 152.39 | 115.50 | | 141.07 | 86.70 | 37.02 | 76.32 | 81.35 | 63.52 |
| *anti* | | 151.47 | 115.16 | | 134.51 | 82.19 | 37.02 | 75.62 | 82.12 | 62.94 |
| | 152.75, 152.17, 152.03 | | | 159.91, 159.83, 159.27 | | | | | | |

[a] Not assigned.

Table 3. Temperature dependence for the imino NH(1) and amino NH(2) proton chemical shifts.

| $T$ [°C] | NH(1) *anti* | NH(2)A *anti* | NH(2)B *anti* | NH(1) *syn*1 | NH(2)A *syn*1 | NH(2)B *syn*1 | NH(1) *syn*2 | NH(2)A *syn*2 | NH(2)B *syn*2 |
|---|---|---|---|---|---|---|---|---|---|
| 50 | 11.32 | – | – | 10.75 | – | – | 10.66 | – | – |
| 40 | 11.33 | – | – | 10.74 | 8.67 | | 10.68 | – | – |
| 30 | 11.35 | – | 4.78 | 10.74 | 8.69 | 4.06[a] | 10.69 | 8.91 | 4.76 |
| 20 | 11.36 | 9.00 | 4.76 | 10.73 | 8.70 | 4.05[a] | 10.72 | 8.93 | 4.86 |
| 10 | 11.38 | 9.05 | 4.76 | 10.73 | 8.73 | | 10.73 | 8.96 | 5.07 |
| 0 | 11.39 | 9.07 | 4.75 | 10.73 | 8.75 | | 10.75 | 8.96 | 5.17 |
| −5 | 11.40 | 9.09 | 4.75 | 10.73 | 8.76 | 4.04[a] | 10.76 | 8.97 | 5.21 |
| −10 | 11.40 | 9.10 | 4.74 | 10.73 | 8.76 | | 10.76 | 8.97 | 5.25 |
| −20 | 11.42 | 9.14 | 4.72 | 10.73 | 8.79 | | 10.77 | 8.96 | 5.32 |
| −30 | 11.42 | 9.16 | 4.72 | 10.74 | 8.81 | | 10.77 | 8.94 | 5.35 |
| −40 | 11.44 | 9.18 | 4.72 | 10.74 | 8.85 | | 10.74 | 8.96 | 5.40 |
| −45 | 11.44 | 9.20 | 4.70 | 10.75 | 8.86 | | 10.75 | 8.96 | 5.42 |

[a] Obtained from NOESY spectra at 30, 20 and −5 °C.

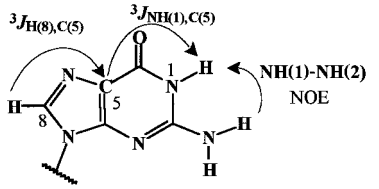
As previously observed for the $\mathbf{1}_8 \cdot \mathrm{KPic}$ octamer, the existence of distinct sets of two amino proton signals separated by more than 4 ppm and the presence of intermolecular H8–NH(2) NOEs indicate that the new polymeric species is assembled from G quartets.[3, 11] The stoichiometry of [G]/[K] = 4 is fully consistent with the columnar polymeric aggregate $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ formed by stacked G quartets that are held together by cation coordination. The structure of $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ in $CDCl_3$ resembles the columnar structures that are proposed to explain the aggregation of G derivatives in water.[1]

The presence of three sets of signals that do not interconvert on the NMR chemical shift time scale and the specific $^1$H–$^1$H NOEs indicate that G **1** residues within $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ adopt three different conformations in solution. The specific conformational isomers are one *anti* rotamer and two different *syn* conformers. These conformers are characterised by their different NOE cross-peak intensities between H(8) and the sugar protons (with 100–300 ms mixing times).[12] The *anti* isomer, whose O(4′)-C(1′)-N(9)-C(4) dihedral angle $\chi$ is typically in the $-180 \pm 90°$ range, is characterised by strong H(8)–H(2′) NOEs and medium intensity NOEs for H(8)–H(1′), H(8)–H(2″), H(8)–H(5′,5″). The *syn* isomers ($\chi = 0 \pm 90°$) are distinguished by their strong H(8)–H(1′) NOEs. A qualitative evaluation of NOE cross-peak intensities suggests that the existence of the two distinct *syn* rotamers is primarily due to two factors: a) different sugar puckers for the syn isomers, apparent from the distinct H(1′)–H(2′,2″) NOESY and TOCSY cross-peaks, and b) a different dihedral angle around the glycosidic C(1′)–N(9) bond angle for the two *syn* isomers. The $\chi$ value is approximately 30° and 50° for the *syn*1 and *syn*2 isomers, respectively, while the *anti* isomer has a $\chi$ value in the range $-120 \pm 10°$.

The $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ aggregate is stable over a temperature range of approximately 100 °C. The $^1$H NMR spectra recorded at different temperatures show small chemical shift changes for the exchangeable protons NH(1) and NH(2), indicating strong hydrogen bonding (Table 3). The NH(1) signals detected at $\delta = 10.72$ and 10.73 at 20° C invert their relative position below room temperature, while the third NH(1) signal ($\delta = 11.3$) moves towards higher frequencies faster than the other two NH(1) signals do at low temperatures. Although they are broad at room temperature, the signals for the three hydrogen-bonded NH(2) protons are sharp and well-separated at −5° C. Finally, one of the non-hydrogen-bonded NH(2) peaks undergoes the largest temperature-dependent change in chemical shift (over 0.5 ppm)

To determine the identity of the different conformational isomers within each G quartet in $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ we followed the procedure previously used to determine the G-quartet

structures in octamer $\mathbf{1}_8 \cdot$ KPic.[3] In that report, the geometry of each G quartet was established on the basis of $^1H-^{13}C$ HMBC and $^1H-^1H$ NOESY experiments. For $\mathbf{1}_8 \cdot$ KPic, one of the G quartets was "all-*anti*" while the other quartet was "all-*syn*". In the present work, we needed to correlate three sets of exchangeable protons NH(1) and NH(2) for $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$. This was achieved by recording a $^1H-^{13}C$ HMBC spectrum at 30 °C to correlate H(8) with its corresponding NH(1) signals by $^3J_{CH}$ coupling to C(5) for the three different rotamers (Figure 4). Although the three H(8) signals ($\delta = 7.61$, 7.63, 7.68) and two of the NH(1) peaks (10.68, 10.72) are very close in chemical shift at 30° C, the assignments were unambiguous because the C(5) signals are so well resolved at that temperature. The C(5) resonances were assigned on the basis of their $^3J_{C(5),H(8)}$ correlations with the *anti*-H(8), *syn*1-H(8) and *syn*2-H(8) protons. In this way the *anti*-NH(1), *syn*1-NH(1) and *syn*2-NH(1) were readily determined by means of $^3J_{C(5),NH(1)}$ coupling constants. Finally, the hydrogen-bonded and non-hydrogen-bonded NH(2) signals were assigned by their strong intramolecular NOEs with NH(1).[3, 11]



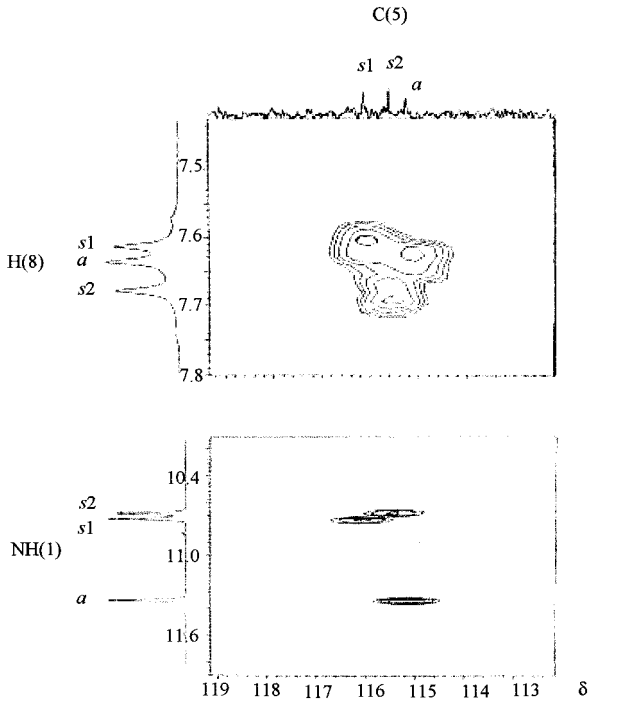
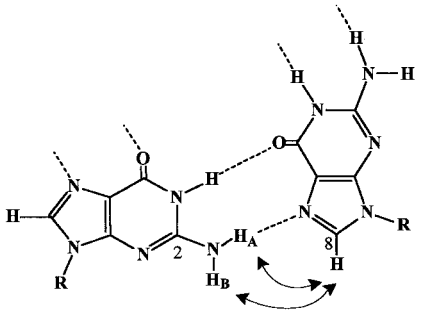




Figure 4. Top: the scheme of correlation between H(8) and NH(1) *by* $^3J$ coupling to C(5). Bottom: a region of the $^{13}C-^1H$ HMBC spectrum recorded at 30 °C of **1** (0.08 mol L$^{-1}$) in the presence of KPic (0.02 mol L$^{-1}$).

Once the various conformers were completely characterised, their arrangement into a G-quartet structure was shown by the NOE cross-peaks between H(8) of one molecule and the NH(2) protons of a neighbouring G **1** base. NOESY spectra at −5° C, a temperature at which the the hydrogen-bonded NH(2) signals for the three rotamers are well resolved, revealed intraquartet H(8)–NH(2) NOEs between conformational isomers of the same type (Figure 5). This key



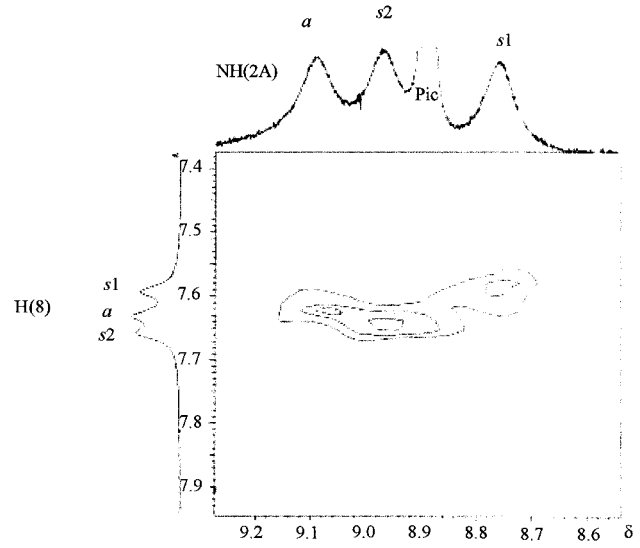



Figure 5. Top: the intraquartet NH(2)–H(8) NOE in a G quartet. Bottom: a region of the 2D NOESY NMR spectrum recorded at −5 °C of **1** (0.08 mol L$^{-1}$) in the presence of KPic (0.02 mol L$^{-1}$).

observation means that each set of $^1H$ NMR signals corresponds to a unique G quartet composed of 4 molecules of the same conformation. Like the $\mathbf{1}_8 \cdot$ KPic octamer, each G quartet within the $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ polymer is homogeneous in terms of its rotamer composition.

The 2D NMR data for $(\mathbf{1}_4 \cdot \mathrm{KPic})_n$ is consistent with a structure composed of three distinct G quartets displaying an all-*anti* (**A**), an all-*syn*1 ($\mathbf{S}_1$) and an all-*syn*2 ($\mathbf{S}_2$) arrangement. The G polymer is then a repetition of a dodecamer building block composed from these three different types of stacked G quartets. The previously described octamer $\mathbf{1}_8 \cdot$ Kpic[3] presumably represents the first step of aggregation. Disregarding the stacking polarity, there are three possible arrangements of three quartets into a dodecamer. These relative arrangements are symbolised as $\mathbf{AS}_1\mathbf{S}_2$, $\mathbf{AS}_2\mathbf{S}_1$ and $\mathbf{S}_1\mathbf{AS}_2$. The analysis of NOESY spectra is most consistent with the $\mathbf{AS}_1\mathbf{S}_2$ arrangement, as interquartet NOEs listed in Table 4 correlate some **A** resonances with $\mathbf{S}_1$ resonances and correlate $\mathbf{S}_1$ signals with



0947-6539/01/0702-0392 $ 17.50+.50/0 *Chem. Eur. J.* **2001**, *7*, No. 2

Table 4. Interquartet NOEs for the two dodecamer models (mixing time 200 ms).

| | NOE | Intensity | **A**(head) – **S**$_1$(tail)/ **S**$_1$(head) – **S**$_2$(tail) | **A**(head) – **S**$_1$(tail)/ **S**$_1$(head) – **S**$_2$(head) |
|---|---|---|---|---|
| 1 | *anti*H(8) – *syn*1H(5′)[a] | strong | + | + |
| 2 | *anti*H(8) – *syn*1H(5″)[a] | medium | + | + |
| 3 | *syn*1H(8) – *anti*H(1′)[a] | medium | + | + |
| 4 | *anti*NH(1) – *syn*1NH(1)[b] | weak[c] | + | + |
| 5 | *syn*1H(8) – *anti*H(1′) | medium | + | + |
| 6 | *anti*H(8) – *syn*1H(3′)[b] | medium | ? | ? |
| 7 | *syn*1H(8) – *syn*2H(3′)[b] | weak | + | – |
| 8 | *syn*1H(8) – *syn*2H(5′) | strong | + | – |
| 9 | *syn*1H(8) – *syn*2H(5″) | medium | + | – |
| 10 | *syn*2H(8) – *syn*1H(1′)[b] | weak | + | + strong |
| 11 | *syn*2H(8) – *syn*1H(2′) | strong | + | not strong |
| 12 | *syn*2H(8) – *syn*1H(2″) | medium | + | + |
| 13 | *syn*2H(8) – *syn*1H(3′) | weak[b] | + | – |

[a] The interquartet NOEs are typical only of an octamer all-*syn* all-*anti* in a head-to-tail arrangement.[3] [b] These interactions are detected at mixing times 400–500 ms. [c] A ROESY experiment (mixing time 100 ms) established that the observed interquartet NOEs were due to through-space magnetisation transfer, and not due to chemical exchange between imino protons.

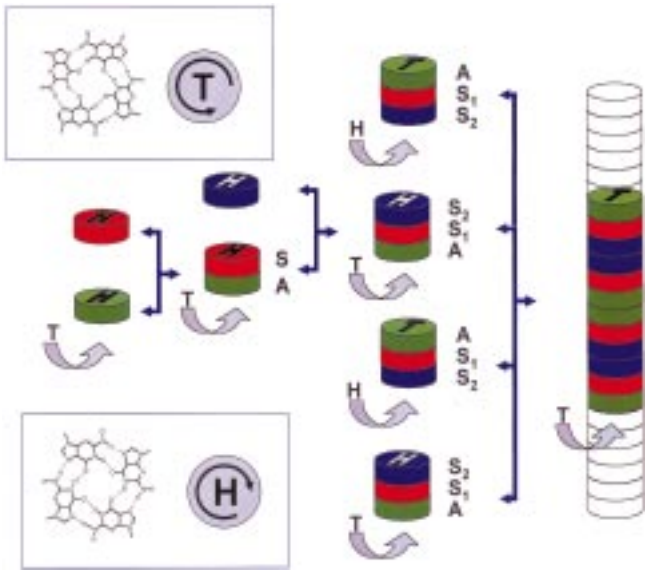



Figure 6. A cartoon of the structure of the columnar aggregate. The three different tetramers A, $S_1$ and $S_2$ are represented in different colours and the two enantiotopic faces of the G quartets are labelled H (head) and T (tail).

those from **S**$_2$. The other two models can be excluded owing to the lack of **A**-**S**$_2$ NOEs.

In addition, interquartet NOEs are important for revealing the relative stereochemical orientation of the three G quartets. Each G quartet has two diastereotopic faces, a so-called head and a tail.[5] In principle, eight ($2^3$) different relative orientations of the G quartets are possible for the **AS**$_1$**S**$_2$ arrangement. Analysis of the NOEs correlating some protons from the **A** quartet with the **S**$_1$ tetramer (entries 1–5 in Table 4) confirms that the **AS**$_1$**S**$_2$ dodecamer is derived from the **AS**$_1$ octamer, whose structure is an all-*anti* G quartet stacked on an all-*syn* G quartet in a head-to-tail arrangement. In particular, the first three entries in Table 4 have already been observed in the NMR study of the relative orientation of the all-*anti* (**A**) and the all-*syn* (**S**) quartets in the **1**$_8$·KPic octamer. In that study, the presence of these interquartet NOEs excluded the possibility of other arrangements[3] such as **A**(head) – **S**(head); **A**(tail) – **S**(head); **A**(tail) – **S**(tail). Therefore, the analysis is restricted to the two models: **A**(head) – **S**$_1$(tail)/**S**$_1$(head) – **S**$_2$(tail) and **A**(head) – **S**$_1$(tail)/**S**$_1$(head) – **S**$_2$(head). All the observed NOE effects are more consistent with the former model.

The interquartet NOEs are also very useful in determining the stacking model for the polymer, that is the self-association pattern of dodecamers. The lack of interactions connecting protons from the **A** and **S**$_2$ G quartets suggests a dodecamer repetition sequence of the type [**AS**$_1$**S**$_2$][**S**$_2$**S**$_1$**A**][**AS**$_1$**S**$_2$]-[**S**$_2$**S**$_1$**A**][**AS**$_1$**S**$_2$], wherein two all-*anti* quartets and two all-*syn*2 quartets are close together, arranged respectively in a tail-to-tail and in a head-to-head orientation (Figure 6).

If chloroform is removed from the solution of the polymeric aggregate and replaced with either hexane or heptane, a viscous liquid is obtained. Starting from a concentration of ca. 6% (*w*/*w*) this liquid becomes birefringent and displays textures typical of lyotropic cholesteric phases. A columnar hexagonal phase is observed at higher concentration (from 16 to ca 40% *w*/*w*). The full characterisation of these liquid crystalline phases will be reported in a separate paper.[13]

## Conclusion

The stereochemical regularity of these columnar polymeric G aggregates is truly amazing. The modified nucleoside **1** must contain all the structural information necessary to drive self-assembly towards the highly ordered polymeric structure schematised in Figure 6 The heterocyclic base contains the cause of the G-quartet formation,[14] namely the two H-bond donors NH(1) and NH(2) and the two H-bond acceptors CO and N(7). The CO groups are also involved in the crucial coordination of the alkali metal ions.[6] The sugar's chirality then dictates the chirality transfer to the pile of G quartets. These stacked G quartets are not in register, but are rotated by ca. 30° to generate a helical structure.[6] The sugar and the C(3′), C(5′) ester groups must also help control the glycosidic bond stereochemistry and the subsequent self-association, probably by steric effects. The fact that the ribo G derivative with the same decanoyl ester groups at C(2′), C(3′) and C(5′) self-assembles to give only an octamer, even with excess KPic present, supports the importance of steric factors in the self-assembly process.[15]

## Appendix

**SANS data analysis**: The macroscopic differential coherent scattering cross-section $d\Sigma/d\Omega(Q)$ (see definitions in the Experimental Part) of a system composed of noninteracting monodisperse randomly oriented particles is given by Equation (1), where $N_p$ is the particle number density and $P(Q)$ is the particle form factor. $P(Q)$ depends on the size, shape and internal features of the scattering particles.

$$\frac{d\sum}{d\Omega}(Q) = N_p P(Q) \quad (1)$$

According to the present aggregation model[2] and the chemical composition, the scattering particles are expected to be cylinders of length $L$ and radius $R_{ext}$, made up of two shells of different scattering length density. The inner cylinder, with radius $R_{in}$ of about 10–12 Å, contains the deoxygua-

nosine core, while the outer shell (of thickness of about 10 Å) contains the lipophilic tails. From the chemical structure and the molecular volumes, scattering length densities of $\rho_{in} = 4.23 \times 10^{10}\ cm^{-2}$ and $\rho_{ext} = -0.423 \times 10^{10}\ cm^{-2}$ have been calculated. The $CDCl_3$ scattering length density is $\rho_0 = 6.33 \times 10^{10}\ cm^{-2}$. On the basis of this model, detailed information on the size and shape of the scattering particles can be obtained by comparing the experimentally observed scattering curve to the form factor calculated for a particular particle structure or for a mixture of known structures. In the present case, for the oligomeric forms (indicated with the subscript $_o$) the $P_o(Q)$ has been modelled as the square of the cylinder axial factor and of the cross-section factor averaged over all possible orientations as Equation (2)[9, 10], where Equation 3 holds. $J_1$ is the first-order Bessel

$$P_o(Q) = \int_0^{\pi/2} [2\pi L \frac{\sin((QL\cos\gamma)/2)}{(QL\cos\gamma} F(Q,\gamma)]^2 \sin\gamma d\gamma \quad (2)$$

$$F(Q,\gamma) = (\rho_{ext} - \rho_0)R_{ext}^2 \frac{J_1(QR_{ext}\sin\gamma)}{QR_{ext}\sin\gamma} + (\rho_{in} - \rho_{ext})R_{in}^2 \frac{J_1(QR_{in}\sin\gamma)}{QR_{in}\sin\gamma} \quad (3)$$

function and $\gamma$ is the angle between the scattering vector **Q** and the longest axis of the cylinder. In these equations, $L$ is related to $\nu$ times the distance between two G quartets (assumed to be 3.4 Å, as observed in lyotropic liquid crystalline phases of guanosine derivatives).[1] In the particular case where the cylinders are quite long and thus can show a preferred orientation around $\gamma = 0$ (i.e. parallel to the cell walls), the axial factor $\sin[(QL\cos\gamma)/2]/[(QL\cos\gamma)/2]$ in Equation (4) goes to zero rapidly, except when $\cos\gamma$ is so small that it compensates for the largeness of $L$.[10] As a consequence, for long, oriented cylinders (polymeric forms indicated with $_p$), Equation (4) holds, where $F(Q,\pi/2)$ is defined in Equation (5).

$$P_p(Q) = 4\pi^2 L^2 F^2\ (Q,\pi/2) \quad (4)$$

$$F(Q,\pi/2) = (\rho_{ext} - \rho_0)R_{ext}^2 \frac{J_1(QR_{ext})}{QR_{ext}} + (\rho_{in} - \rho_{ext})R_{in}^2 \frac{J_1(QR_{in})}{QR_{in}} \quad (5)$$

In our case, the self-assembly process led to oligomeric (short cylinders, $\nu_o < 20$) or to polymeric (long cylinders, $\nu_p > 20$) species according to the amount of alkali metal ions present in the chlorinated organic solvent.[2] The SANS curve analysis was then performed assuming the presence in the organic solution of the two families of particles, disordered oligomers and oriented polymers. Therefore, using $N_{G\ quartets}$ to indicate the number density of tetramers in solution and $f$ to indicate the fraction of G quartets assembled as oligomers, the scattering intensity was modelled according to Equation (6), where the symbol $\langle \ldots \rangle$ indicates an average over a gaussian

$$\left[\frac{d\sum}{d\Omega}\right]_{mod}(Q) = N_{G\ quartets}\left[f\left\langle\frac{P_o(Q)}{\nu_o}\right\rangle + (1-f)\left\langle\frac{P_p(Q)}{\nu_p}\right\rangle\right] + A_0 \quad (6)$$

distribution of the cylinder length $L = 3.4\nu$ Å. In this equation, $A_0$ is a constant background, which mainly takes into account the incoherent scattering.

The two SANS curves were analysed using a global fit procedure[16] based on Equation (6) assuming common values for $R_{ext}$, $R_{in}$ and $N_{G\ quartets}$ but distinct values for $f$ and $A_0$ and for the gaussian maxima and widths ($\nu_o$, $\sigma_o$ and $\nu_p$, $\sigma_p$, for the oligomeric and polymeric forms, respectively).

## Experimental Section

**Synthesis of 2**: Decanoic acid (0.24 g, 1.39 mmol) and 1-adamantanecarbonyl chloride (0.25 g, 1.26 mmol) in 8 mL dry acetonitrile were stirred for 2 h under argon in the presence of freshly distilled triethylamine (0.6 mL, 4.3 mmol). Vacuum-dried 2′-deoxyguanosine (0.15 g, 0.52 mmol) and DMAP (0.015 g 0.12 mmol) were then added and the resulting suspension was stirred overnight at room temperature. TLC on silica gel ($CH_2Cl_2$/MeOH 95:5) showed essentially complete conversion of the starting guanosine. The suspension was then filtered and the precipitate was washed subsequently with acetonitrile, ethanol, Millipore grade water and acetonitrile. The white solid thus obtained was analytically pure **1** (94% c.y.).[2] Before use this material was recrystallised from absolute ethanol.

**SANS experiments**: The small angle neutron scattering experiments were carried out at the Laboratoire Leon Brillouin (Saclay, France) with the small-angle diffractometer PAXE, which is located on a cold neutron guide of the Orphee reactor. The samples were measured at room temperature (ca. 20 °C) in 2 mm thick quartz cells. The neutron wavelength $\lambda$ used was 4.15 Å and the sample to 2D-detector distances were 1.5 m and 5.0 m; this allows us to investigate a scattering $Q$ range between 0.02 and 0.35 $Å^{-1}$, where the modulus of the scattering vector **Q** is defined as $4\pi\sin\theta/\lambda$, with $2\theta$ the scattering angle.

The raw data were corrected for electronic noise and sample-holder signal by measurement of an empty quartz cell. Since the scattering is isotropic, the data were radially averaged to improve statistics. The scattered intensities have been converted into absolute units ($d\Sigma/d\Omega$, in $cm^{-1}$) by a direct determination of the incident flux.[17]

**NMR experiments**: The samples were prepared as follows: solutions were prepared by dissolving 40 mg (0.069 mmol) of **1** and 4.6 mg (0.017 mmol) of KPic in 1 mL of $CDCl_3$ stored over molecular sieves (4 Å). The samples were sonicated for several hours, in order to ensure complete dissolution of potassium picrate. NOESY and HETCOR data were recorded on a Varian Gemini 300 (300 MHz) instrument; HMBC and TOCSY spectra were recorded on a Varian Mercury 400 (400 MHz) spectrometer; $^1H$ and $^{13}C$ spectra were collected at different temperatures on both the instruments; $\delta$(H) values are reported in ppm from the solvent peak in $CDCl_3$ solutions.

The NOESY experiments were recorded at $-5$, 20 and 30 °C with mixing times 50–500 ms in the phase-sensitive mode. The data were collected using a 90° pulse width of 12 μs and a spectral width 4500 Hz. Altogether, 8 repetitions were collected for 256 time increments. The relaxation delay was 2.5 s.

The instrumental settings for the HETCOR experiment were: spectral width 3873 Hz and pulse width 12 μs (90° flip angle) for $^1H$, spectral width 14405 Hz and pulse width 12 μs (90° flip angle) for $^{13}C$. A total of 4096 scans were collected for 48 time increments. The relaxation delay was 2 s.

The HMBC data were collected with the sample nonspinning at 30 °C using a 90° pulse width of 13.5 μs for $^1H$ and 12.5 μs for $^{13}C$. The spectral width was 5394 Hz in the F2 dimension ($^1H$) and 19157 Hz in the F1 dimension ($^{13}C$). A total of 48 repetitions were collected for 1024 time increments. The relaxation delay was 2 s.

The TOCSY experiment was conducted in the phase-sensitive mode using a 90° pulse width of 13.5 μs and a spectral width of 5060 Hz in each dimension. The relaxation delay was 1.5 s and the mixing time was 100 ms. A total of 4 repetitions were collected for 128 time increments.

## Acknowledgement

This work was supported by MURST (Italy, through the National Program—Cofin'99: Structure, Order, Dynamics and Applications of Liquid Crystalline Systems), University of Bologna (funds for selected research topics 1997/99), INFM (Italy), FAPESP (Brazil) and the Department of Energy (US).

[1] G. Gottarelli, G. P. Spada, A. Garbesi in *Comprehensive Supramolecular Chemistry*, *Vol. 9* (Eds.: J.-M. Lehn, J. L. Atwood, D. D. MacNicol, J. A. D. Davies, F. Vögtle, J.-P. Sauvage, M. W. Hosseini), Pergamon, Oxford, **1996**, pp. 483–506.

[2] G. Gottarelli, S. Masiero, G. P. Spada, *J. Chem. Soc. Chem. Commun.* **1995**, 2555–2557.

[3] A. L. Marlow, E. Mezzina, G. P. Spada, S. Masiero, J. T. Davis, G. Gottarelli, *J. Org. Chem.* **1999**, *64*, 5116–5123.

[4] V. Andrisano, G. Gottarelli, S. Masiero, E. H. Heijne, S. Pieraccini, G. P. Spada, *Angew. Chem.* **1999**, *111*, 2543–2544; *Angew. Chem. Int. Ed.* **1999**, *38*, 2386–2388.

[5] The label of "head" and "tail" sides of a G quartet is as defined by Feigon (F. W. Smith, F. W. Lau, J. Feigon, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10546–10550): the head face presents a clockwise rotation of the N−H···O=C hydrogen-bonding pattern (from the donor to the acceptor).

[6] S. L. Forman, J. C. Fettinger, S. Pieraccini, G. Gottarelli, J. T. Davis, *J. Am. Chem. Soc.* **2000**, *122*, 4060–4067.

[7] J. M. Lehn, *Supramolecular Chemistry—Concepts and Perspectives*, VCH, Weinheim, **1995**.
[8] For some recent examples see: C. F. van Nostrum, R. J. M. Nolte, *Chem. Commun.* **1996**, 2385–2392; R. K. Castellano, J. Rebek, Jr., *J. Am. Chem. Soc.* **1998**, *120*, 3657–3663; A. Semenov, J. P. Spatz, M. Moller, J. M. Lehn, B. Sell, D. Schubert, C. H. Weidl, U. Schubert, *Angew. Chem.* **1999**, *111*, 2701–2705; *Angew. Chem. Int. Ed.* **1999**, *38*, 2547–2550; H. A. Klok, K. A. Joliffe, C. L. Shauer, L. J. Prins, J. P. Spatz, M. Moller, P. Timmermann, D. N. Reinhoud, *J. Am. Chem. Soc.* **1999**, *121*, 7154–7155.
[9] O. Glatter, O. Kratky, *Small Angle X-Ray Scattering*, Academic Press, London, **1982**.
[10] F. Carsughi, M. Ceretti, P. Mariani, *Eur. Biophys. J.* **1992**, *21*, 155; F. Carsughi, G. Di Nicola, G. Gottarelli, P. Mariani, E. Mezzina, A. Sabatucci, G. P. Spada, *Helv. Chim. Acta* **1996**, *79*, 220–234.
[11] N. G. Williams, L. D. Williams, B. R. Shaw, *J. Am. Chem. Soc.* **1989**, *111*, 7205–7209; N. G. Williams, L. D. Williams, B. R. Shaw, *J. Am. Chem. Soc.* **1990**, *112*, 829–833; J. Feigon, K. M. Koshlop, F. W. Smith, *Methods in Enzymology*, *Vol. 261* (Ed.: T. L. James), Academic Press, San Diego, **1995**, pp. 225–255.
[12] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley, New York, **1986**, p. 208.
[13] S. Pieraccini, G. Gottarelli, P. Mariani, S. Masiero, L. Saturni, G. P. Spada, *Chirality*, **2000**, in press.
[14] W. Sanger, *Principles of Nucleic Acid Structure*, Springer, New York, **1984**, pp. 315–320.
[15] S. Masiero, J. T. Davis, unpublished results.
[16] J. R. Taylor, *An Introduction To Error Analysis*, University Science, Mill Valley, **1997**.
[17] J. Teixeira in *Structure And Dynamics Of Strongly Interacting Colloids And Supramolecular Aggregates In Solution* (Ed.: S. H. Chen), Kluwer Academic, **1992**, pp. 635–658.

Received: April 7, 2000 [F2407]