Supporting Information

Supramolecular Polymers Based on Quadruplex Formation of Ditopic Guanosine Macromonomers in Non-aqueous Media

Gábor Pintér¹, Gyula Batta¹, Pál Horváth¹, István Löki¹, Tibor Kurtán², Sándor Antus², Sándor Kéki³, Miklós Zsuga^{3*}, Gábor Nagy⁴, János Aradi⁵, Tamás Gunda¹ and Pál Herczegh^{1*}

Department of Pharmaceutical Chemistry, University of Debrecen and Research Group for Chemistry of Antibiotics of the Hungarian Academy of Sciences, Department of Organic Chemistry, University of Debrecen, Department of Applied Chemistry, University of Debrecen, Department of Zoology, University of Debrecen, Department of Biochemistry and Molecular Biology, University Medical School H-4010 Debrecen, Hungary

Synthetic procedures

Synthesis of α,ω -diamino-polyethylene glycol

It was prepared according to the following article:

Mongondry, P.; Bonnans-Plaisance, C.; Jean, M.; Tassin, J. F. *Macromol. Rapid Commun.* **2003**, 24, 1091-1100.

Synthesis of 2',3'-Isopropylidene guanosine (see Scheme 1) and 2',3'-Isopropylidene guanosine hemisuccinate **1**

It was prepared according to the following article:

Defrancq, E.; Leterme, A.; Pelloux, A.; Lhomme, M.F.; Lhomme, J. *Tetrahedron*, **1991**, 5725-5736.

Synthesis of macromonomer **2**

2',3'-Isopropylidene guanosine hemisuccinate (638 mg, 1.5 mmol) was dissolved in dry DMF, Et₃N (312 μ l, 2.25 mmol) and dimethylaminopyridine (55 mg, 0.45 mmol) were added followed by the solution of α,ω -diamino-polyethylene glycol (690 mg, 0.5 mmol in 20 ml dry CH₂Cl₂). The reaction mixture was cooled in an ice-water bath and ethyl-dimethylaminopropyl carbodiimide *HCl (430 mg, 2.25 mmol) was added. It was stirred about 0 °C for 2 hours, than at room temperature overnight. The solvent was evaporated in vacuo and the residue was dissolved in 200 ml of CH₂Cl₂ and was washed with 80 ml of 10 % NaHSO₄ solution, then with saturated NaHCO₃ and the organic phase was dried, and

concentrated in vacuo. Column chromatographic purification resulted **2** as a colorless gel like substance (247 mg, 22.5 %).

Scheme 1. Synthesis of supramolecular macromonomer **2**.

MALDI TOF mass spectrometry

The MALDI MS measurements were performed with a Bruker BIFLEX III mass spectrometer equipped with a TOF analyzer. 19 kV total acceleration voltage was used with pulsed ion extraction (PIE^{TM}). The positive ions were detected both in the reflectron mode. A nitrogen laser (337 nm, 3 ns pulse width, 10⁶-10⁷ W/cm²) operating at 4 Hz was used to produce laser desorption and 100 shots were summed. Samples were prepared with DHB (2,5-Dihydroxybenzoic acid) matrix dissolved in methanol (20 mg/ml). Solutions of α,ω -diamino-polyethylene glycol (**DPEG**) and macromonomer **2** in methanol were made at a concentration of 5 mg/mL. To enhance the cationization, sodium trifluoroacetate (NaTFA) in methanol (2 mg/mL) was added to the corresponding matrix/analyte solutions. The solutions were mixed in a 10:2:1 v/v ratio (matrix:analyte:cationization agent). A volume of 0.5 μ l of these solutions was deposited onto the sample plate (stainless steel), and allowed to air-dry.

As shown in Figure 1. the mass difference between the adjacent peaks is 44 Da, which corresponds to the mass of the repeating unit (ethylene oxide, EO).

Figure 1. MALDI-TOF mass spectra of the starting α,ω -diamino-polyethylene glycol (DPEG) (a) and macromonomer 2 (b). The numbers on the top of peaks show the number of repeat units, i.e., ethylene oxide (EO). The Fig. 1b. inset shows the zoomed MALDI-TOF mass spectrum of macromonomer 2.

The mass series appeared (M) in the MALDI-TOF MS spectra can be expressed by the following equation:

$$M = M_{\text{endgroup}} + nM_{\text{EO}} + M_{\text{cat}} \quad (\text{eq. 1.})$$

where M_{endgroup} , nM_{EO} and M_{cat} are the masses of the endgroup, the repeat unit (i.e. ethylene oxide, 44 Da) and the cation attached to the polymer chains, respectively.

The MALDI-TOF mass spectrum of α,ω -diamino-polyethylene glycol (i.e., the starting polymer in Fig.1.a) shows the exclusive presence of the sodiated, amino-terminated polyethylene glycol oligomers, i.e., $[G1(\text{EO})_n+\text{Na}]^+$ (where n is the number of EO units, and G1 represents the endgroups of the α,ω -diamino-polyethylene glycol) ranging from $n=24$ to $n=41$. The most abundant EO oligomer was found at $n=31$ as shown in Fig 1a.

The mass series calculated by eq. 1. is in good agreement with the observed values. For example, the calculated and the observed mass for α,ω -diamino-polyethylene glycol containing 31 EO units cationized with sodium ion, i.e., $[G1(\text{EO})_{31}+\text{Na}]^+$ is 1447.9 and 1447.8, respectively. (The composition of G1 is $\text{C}_2\text{H}_8\text{N}_2$)

Based on the MALDI-TOF mass spectra the polydispersity (PD), i.e., the value of M_w/M_n (where M_w and M_n are the weight-average and the number-average molecular weight, respectively) can be calculated by eq.2-eq.4.(see ref: Montaudo, G.; Samperi, F. ; Montaudo M. S. *Prog. Polym. Sci.* **2006**, *31*, 277-357):

$$M_n = \sum M_i I_i / \sum I_i \quad (\text{eq.2.})$$

$$M_w = \sum M_i^2 I_i / \sum M_i I_i \quad (\text{eq.3.})$$

$$\text{PD} = M_w / M_n \quad (\text{eq.3.})$$

where I_i is the intensity, M_i is the mass of the i^{th} peaks appeared in the MALDI-TOF mass spectrum.

Using eqs.2.-4. $M_n=1470$ g/mol, $M_w=1480$ g/mol and $\text{PD}=1.007$ were obtained. The value of 1.007 indicates a very narrow molecular weight distribution.

The MALDI-TOF mass spectrum of macromonomer **2** (Figure 1b.) shows two series of peaks. The main series of peaks are due to the presence of the sodiated macromonomer **2** oligomers mass spaced by 44 Da, i.e., $[G2(\text{EO})_n+\text{Na}]^+$ (where G2 denotes the endgroups of macromonomer **2**) and as indicated in Figure 1b. inset. The additional series of peaks with lower intensity originated from the protonated oligomers of macromonomer **2**, i.e., $[G2(\text{EO})_n+\text{H}]^+$ as shown in Figure 1b. inset. Since macromonomer **2** contain basic nitrogen

atoms protonation can occur under MALDI-conditions parallel with the formation of sodiated adduct ions.

As can be seen in Figure 1b., as in the case of the starting α,ω -diamino-polyethylene glycol, oligomers of macromonomer **2** covers a range from $n=24$ to $n=42$ and the most abundant oligomer is also at $n=31$. On the other hand, the mass series calculated by eq. 1. agree well with the measured values, e.g., the calculated and the measured mass for the macromonomer **2** consisting of 31 EO units cationized with sodium ion, i.e., $[G2(EO)_{31}+Na]^+$ is 2258.1 and 2258.2, respectively. (The composition of G2 is $C_{36}H_{46}O_{14}N_{12}$)

The M_n , M_w and PD values of macromonomer **2** were calculated from the the MALDI-TOF mass spectrum presented in Figure 1b. with eqs.2-4. and were found to be 2260 g/mol, 2280 g/mol and 1.009 (very narrow molecular weight distribution), respectively.

Circular dichroism (CD) spectroscopy

General experimental procedures

CD spectra were recorded on a J-810 spectropolarimeter, measured in millidegrees and normalized into $\Delta\epsilon_{\max}$ [$l\ mol^{-1}cm^{-1}$]/ λ [nm] units.

CD data for **2** without addition of K^+ in MeCN [nm ($\Delta\epsilon$): 293 nm (−15.8), 262 nm (12.5), 243 nm (−1.0), 223 nm (2.8), 210 nm (7.2), 191 (−20.0);

$CHCl_3$: 294 nm (−21.7), 264 nm (15.2)

After extraction with potassium picrate in $CHCl_3$: 293 nm (−41.5), 264 nm (30.7),

243 nm (−1.2), 230 (3.8).

NMR analysis

Full $^1H/^{13}C$ assignment of **2** was carried out in $DMSO-d_6$ solution at 300K using HSQC, HMBC (50 ms delay for long-range couplings), TOCSY (15 ms mixing time) and ROESY (100 ms mixing time) experiments. Only "trivial" ROE peaks were detected according to the covalent structure. The NOESY spectrum did not show any appreciable through-space connectivities, probably due to the zero/cross region of this size of a molecule at 500 MHz.

However, new NOESY peaks were observed, e.g. between the CH of the guanine and the OCH₂ of the ribose in the CDCl₃ solution in the presence of picrate, when the aggregates were present.

Assignment	Group	¹³ C (ppm)	¹ H (ppm)
1	NH		10.9 broad
2	C	154.74	
4	C	151.39	
5	C	117.76	
6	C=O	157.76	
8	CH	136.92	7.857
1'	CH	89.13	6.005
2'	CH	84.49	5.241
2-NH2	NH2		6.641
3'	CH	80.02	5.114
4'	CH	85.02	4.252
5'	OCH2	64.88	4.25/4.13
a	C=O	173.48	
b	CH2	29.71	2.49
c	CH2	30.56	2.356
d	C=O	171.65	
e	NH		7.935
f	NCH2	39.46	3.19
g	OCH2	69.92	3.38
h	OCH2 main chain	70.64	3.511
lpr	<u>CH</u> 3	26.14	1.326
lpr	<u>CH</u> 3	27.87	1.522
lpr	<u>C</u> (CH3)2	114.22	

Table 1 For the numbering see the figure 2 (¹H - ¹³C NMR assignment of **2**, referenced to internal TMS = 0 ppm)

Figure 2. The structure of macromonomer **2**

Experimental:

Manufacturer's "ledbpgs2s" stimulated echo DOSY pulse sequence equipped with bipolar and spoil gradient pulses was extended with a 2x2 ms spin echo before detection to suppress baseline distortions. Sine shaped gradient strength was linearly incremented in 32 or 64 steps between 2-95% of the upper limit of 0.5 T/m of the 5 mm inverse broadband probehead of the Bruker DRX 500 spectrometer at 298K. The 90° ^1H pulse was typically 11 μs . Typical DOSY experiments lasted 0.5-1 hrs depending on the number of acquired scans (8-32) and increments. Processed spectra were baseline corrected, intensities of the main OCH_2 peak of the PEG part and the internal reference were processed with an in house written MATLAB code. The triple exponential fitting routine estimates six parameters, and the apparent molecular weights were obtained from the ratio of diffusion coefficients with respect to reference compound. In case of a PEG molecular weight standard of 20 kDa, the uncertainty of the diffusion constants in five independent experiments was ca. 6% of the measured mean. For monoexponential fit of non overlapping PEG signals, the estimated error of the apparent molecular weight was typically smaller than 10%, if we disregard systematic errors caused by the different molecular shape and solvation of the reference compound. However, the absolute molecular masses of the aggregates may have much bigger uncertainties.

In the smaller aggregate ca. 0.5 equivalent, while in the bigger aggregate ca. 1 equivalent picrate was measured from the integration of the signal at 8.9 ppm.

The 1D proton NMR spectrum and DOSY processed diffusion experiments are shown in Figures 3-4.

Figure 3. The proton NMR spectrum of macromonomer **2**

Figure 4. DOSY spectrum of the smaller aggregate. Bound TMS has similar diffusion coefficient as the **2** aggregate. However, no NOE or ROE contacts were observed between TMS and the host molecules.

Figure 5. Fitting the decaying signals of reference phenylthio-2,3,4,6-tetra-*O*-benzyl- α -D-glucopyranoside and multicomponent PEG signals

Dynamic Light Scattering (DLS)

The solution of **2** in CHCl_3 at a concentration of 10 mg/mL was examined by dynamic light scattering (DLS) performed on a Brookhaven Light Scattering device equipped with a BI-9000 digital correlator. The light source was a solid state, vertically polarized laser operating at 533 nm. DLS measurements were performed at 25 °C at scattering angle $\theta = 90^\circ$. The autocorrelation function of the scattered light intensity was acquired in the homodyne mode. For the full particle-size distributions the NNLS (Non-Negative Constraint Least Squares) calculation combined with multiple pass analysis was employed.

Figure 6. Size distribution of aggregates formed from the **2** in CHCl_3 as determined by DLS ($\Theta = 90^\circ$). Intensity and d denote the intensity of scattered light and the diameter of aggregates, respectively.

Quantum mechanical calculations

In order to estimate the geometrical size of the assembly, an ideal molecule consisting of two G₄-quartet moieties and four peg strands were assembled. The PEG strands contained 35 –OCH₂CH₂– subunits, respectively, corresponding to the average molecular weight of the peg used in the synthesis. Clay and Guild (ref.) performed very sound QM investigation of potassium filled stacked G-quartets using ab initio DFT B3LYP QM methods. Obviously this method is not suitable for our large assembly, therefore we had to use combined semiempirical QM and FF methods to obtain a molecule with reasonable geometry. At first single layer and double stack potassium filled G-quartet was build and minimized with the MOPAC QM method (MOPAC package of the Hyperchem 7.1 application using the AM1 Hamiltonian). Energy and geometry minimization gave rise to stable G-quartet structures possessing a very close geometry to those in ref. Clay *et al.* Next the substituted carbohydrate moieties were added followed by new energy minimization. In the next step two potassium filled G-layers were joined with 4 PEG chains consisting of 35 –OCH₂CH₂– subunits each. The whole assembly was minimized again by MM+ force field while the coordinates of the guanosin and carbohydrate moieties were fixed. The resulting ideal barrel with maximal elongation was used for geometry calculations.

The calculated size of the bounding box of molecule X is 159.64 x 37.99 x 37.74 Å (using the VdW radii of the atoms), the length of X calculated from the coordinates is 155.11 Å. If the molecule is not straightened but “squeezed” and the PEG chains are in looped or scrambled conformations, according to basic geometric considerations the width of the molecule can be equal or less than its maximum length.

Ref. : Clay, E.H.; Guild, I.R. *J. Mol. Graph. Model.* **2005**, 24, 138-146.