SEQUENCING OF SHORT RNA OLIGOMERS BY PROTON NUCLEAR MAGNETIC RESONANCE

P. A. HADER, T. NEILSON, D. ALKEMA, E. C. KOFOID* and M. C. GANOZA*

Department of Biochemistry, McMaster University, Hamilton, Ontario L8N 3Z5 and **Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada*

Received 8 October 1981

1. Introduction

Advances have been made in the sequencing of RNA by enzymic [1–3] and chemical [4,5] methods. A limiting feature of these procedures is the quantity of RNA available, unfortunately 2–3 orders of magnitude less than for DNA sequencing [6,7]. These methods can also be unreliable when applied to short RNA fragments, especially sequences containing stretches of guanosine residues where resolution in polyacrylamide gel analysis is ambiguous. DNA sequencing suffers the same drawback [6]. More facile RNA duplex formation [8] creates problems for base-specific endonuclease digestions which are more efficient on single-stranded RNA. Lack of RNA restriction enzymes can be circumvented by recourse to short complementary DNA sequences and the use of ribonuclease H [9] which cleaves within or adjacent to the hybridization site. However, existing RNA secondary structure can inhibit expected RNA:DNA duplex formation and a selection of short DNA sequences must be assayed to achieve a constant fragment pattern.

A direct procedure which complements the shortcomings listed above would be beneficial. Here, we present a non-destructive RNA sequencing technique for short oligoribonucleotides (data for hexaribonucleotides) which is independent methodologically from the existing chemical and enzymatic procedures. This method is fast, 'one step', and can involve little interpretation, since it has the potential to be computerized. A spectroscopic procedure, proton nuclear magnetic resonance (^{1}H NMR) is used which requires a pure sample (500–800 μg) of the RNA fragment to be sequenced. The rapid advancement of instrumentation, in particular, the availability of very high field magnets and more rapid scanning procedures, will dramatically reduce the sample size in the near future. However, the present sample size is convenient for checking oligomers from chemical synthesis. The procedure is based on the fact that the resonances of the non-exchangeable aromatic and anomeric protons of adenosine, guanosine, cytidine and uridine exhibit changes in their particular chemical shifts which depend on the bases immediately surrounding a particular residue in a given sequence. These sequence effects on chemical shifts of protons on a particular residue do not extent significantly beyond the next-nearest neighbour at 70°C [10,11].

A set of ^{1}H NMR chemical shift parameters [12] (derived from the NMR spectra of trinucleoside diphosphates) is available which allows the chemical shifts of protons in longer sequences to be calculated. Consequently, the base sequence in longer oligoribonucleotides can be determined.

2. Materials and methods

The oligoribonucleotides used in this study were synthesized by the phosphotriester method developed in [13]. Complete synthesis details for the preparation of the hexaribonucleotides will appear elsewhere. ^{1}H NMR spectra were obtained in the Fourier transform mode of Bruker WH-90, WM-250 and WH-400 spectrometers equipped with quadrature detection. Probe temperatures were maintained to within ±1°C by a Bruker variable temperature unit and were calibrated by thermocouple measurements. The samples were lyophilized twice from D_2O and dissolved in 100%

Published by Elsevier/North-Holland Biomedical Press
00145793/81/0000–0000/$02.75

D_2O (Aldrich) containing 0.01 M sodium phosphate buffer (pD 7.0) and 1.0 M NaCl. The sample concentrations were 8–14 mM. *t*-Butanol-d was used as an internal reference and the chemical shifts reported in parts per million (ppm) relative to 2,2-dimethyl-2-silapentan-5-sulphonate (DSS). The field frequency lock was provided by the deuterium signal of D_2O.

Sequences of the oligoribonucleotides, CAU$\vec{A}$AUG, AUG$\vec{C}$UA, UCA$\vec{U}$GA and AAG$\vec{C}$UU, were determined up to the 3′-terminal nucleotide as in [4], with the modifications in [14] and the last nucleotide deduced from the known base ratios as confirmed by NMR spectroscopic analysis. Electrophoresis was carried out using 25% polyacrylamide slab gels (20 × 40 cm) containing 7.0 M urea [15]. Analytical gels for sequencing were 0.75 mm thick with 18 wells [6]. Autoradiography used prefogged Kodak XRP-1 film and Cronex Lightening Plus Intensifying screens (Dupont) at −20°C [16].

Polynucleotide kinase (nuclease, free), and ribonuclease *Phy*I were obtained from PL Biochemicals; RNase A and T_1 from Sigma; RNase C and S_1 from Miles and RNase U_2 from Calbiochem.

3. Results and discussion

Tables 1 and 2 compare the observed and calculated ^{1}H NMR chemical shifts of the nonexchangeable aromatic protons and ribose H-1′ protons in oligoribonucleotides, CAU$\vec{A}$AUG, AUG$\vec{C}$UA, UCA$\vec{U}$GA and AAG$\vec{C}$UU, at 70°C. The spectra were recorded at 70°C in dilute solution where all inter- and intrastrand effects are minimized [17,18] and oligomers are assumed to adopt a similar average conformation. Under these conditions the oligoribonucleotides are largely unstacked. The low degree of base stacking [19–21] is indicated by the $J_{1',2'}$ coupling constants which were normally >4.5 Hz. Therefore, at high temperatures and low concentrations, changes in chemical shift of the various protons on a given nucleoside residue are considered to be caused mainly by ring current anisotropy of neighbouring residues [22]. If the chemical shift changes are to be useful in sequencing, the ring current anisotropic effects must be different not only for purines and pyrimidines but also between each base type. This is in fact observed and the addition of a nucleotide pX to the 3′-end of a dinucleoside monophosphate (dimer) produces an X-dependent shielding of the dimer protons, which decreased in the order A > G > C > U [10].

Here, we demonstrate how the chemical shift parameters can be used to calculate the chemical shifts of individual protons, and hence determine the sequence of oligoribonucleotides. The calculated chemical shifts of the oligomers, CAU$\vec{A}$AUG, AUG$\vec{C}$UA, UCA$\vec{U}$GA and AAG$\vec{C}$UU, were determined by solving the appropriate chemical shift equation for a specific proton on a given residue [12]. By adding the computer chemical shift parameters to the observed chemical shifts of the mononucleoside, the sequence specific nearest-neighbour and next-nearest neighbour ring current effects are approximated. The effects for 3′- and 5′-neighbours are accounted for in this manner (see table 1[c]). For example, the sequence-dependent chemical shifts of the protons in the third uridine residue of CAU$\vec{A}$AUG are predicted by adding the correct shift parameters to the chemical shifts observed for uridine. Thus, the 5′- and 3′-ring current effects of the adjacent bases on the third uridine in sequence, CAU$\vec{A}$AUG, are taken into account.

In tables 1 and 2 the calculated and observed chemical shifts for the sequences AUG$\vec{C}$UA, CAU$\vec{A}$AUG, UCA$\vec{U}$GA and AAG$\vec{C}$UU, were compared and found to be in satisfactory agreement, as values different by $\leq \pm 0.040$ ppm are considered acceptable. Sarma and his associates allow the range ±0.10 ppm. Although all 4 sequences contain the same base ratio, 2A:2U:1G:1C, the calculated chemical shifts for each sequence are unique. The oligoribonucleotide base ratio is immediately available from the ^{1}H NMR spectrum, since resonances from protons, specific for a given base, are found consistently in similar regions of the spectrum. Once the base ratio is determined, correct sequence results from fitting the best set of predicted chemical shift values (for a particular sequence) to the observed shifts. This process can be computerized, thereby eliminating the drudgery of finding the correct sequence over a large set of sequences (for example, the set which has the base ratio 2A:2U:1G:1C contains 180 sequences). Thus, the dependence of proton chemical shifts on primary sequence can be exploited to unravel the unknown sequence of an oligoribonucleotide.

This rapid procedure is most convenient for synthetic oligomers since it is non-destructive and confirms the predetermined sequence. Generally, a spectrum of an oligomer can be obtained in 10–20 min. Oligomer length could become a limiting factor

Table 1
Predicted ^{1}H chemical shifts[a] of the protons in CAUAUG and AUGCUA[b]

123456 CAUAUG				123456 AUGCUA			
Proton	Observed[c]	Predicted[d]	Δδ[e]	Proton	Observed[c]	Predicted[d]	Δδ[e]
C(1)H-6	7.643	7.705	+0.062[f]	A(1)H-8	8.276	8.286	+0.010
A(2)H-8	8.326	8.309	-0.017	A(1)H-2	8.214	8.205	-0.009
A(2)H-2	8.154	8.157	+0.003	U(2)H-6	7.757	7.730	-0.027
U(3)H-6	7.685	7.710	+0.025	G(3)H-8	7.961	7.976	+0.015
A(4)H-8	8.310	8.295	-0.015	C(4)H-6	7.757	7.775	+0.018
A(4)H-2	8.131	8.120	-0.011	U(5)H-6	7.757	7.741	-0.016
U(5)H-6	7.685	7.738	+0.053[f]	A(6)H-8	8.397	8.373	-0.024
G(6)H-8	7.945	7.970	+0.025	A(6)H-2	8.236	8.249	+0.013
C(1)H-1'	5.764	5.793	+0.029	A(1)H-1'	6.032	6.042	+0.010
A(2)H-1'	6.002	6.033	+0.031	U(2)H-1'	5.852	5.801	-0.051[f]
U(3)H-1'	5.804	5.767	-0.037	G(3)H-1'	5.858	5.880	-0.022
A(4)H-1'	6.002	6.028	+0.026	C(4)H-1'	5.913	5.875	-0.038
U(5)H-1'	5.822	5.813	-0.009	U(5)H-1'	5.858	5.773	-0.085[f]
G(6)H-1'	5.832	5.874	+0.042[f]	A(6)H-1'	6.084	6.099	+0.015
C(1)H-5	5.894	5.999	+0.105[f]	U(2)H-5	5.817	5.706	-0.111[f]
U(3)H-5	5.736	5.716	-0.020	C(4)H-5	5.896	5.895	-0.001
U(5)H-5	5.751	5.773	+0.022	U(5)H-5	5.817	5.793	-0.024

[a] Chemical shifts are in ppm from DSS in buffered D_2O (pD 7.0) at 70 ± 1°C

[b] Oligoribonucleotides $CA\vec{U}AUG$, $AU\vec{G}CUA$, $UCA\vec{U}GA$ and $AA\vec{G}CUU$, were sequenced by the Donis-Keller method to verify and confirm the sequence

[c] Observed chemical shift assignments were made using the method of incremental analysis; i.e., comparison of the spectra of ApG, ApGpC, ApGpCpU, ApApGpCpU and ApApGpCpUpU. Reference was also made to spectra of related oligomers, for example; comparison of the spectra of ApUpGpC, CpUpA and ApUpGpCpUpA

[d] Values are calculated as follows: for example, U(3)H-6 in CAUAUG

$$\delta^{H\text{-}6}_{U(3)} = \delta^{H\text{-}6}_{U,mon} + \Delta^{3'}_{1,pA} + \Delta^{3'}_{2,pU} + \Delta^{5'}_{1,Ap} + \Delta^{5'}_{2,Cp}$$

where $\delta^{H\text{-}6}_{U,mon}$ is the observed shift in the mononucleoside and Δ terms are parameters reflecting the anisotropic effect and position of the neighbouring base relative to base U [12]

[e] Δδ values are the difference between calculated and observed values

[f] Indicates Δδ values are outside accepted error

Table 2
Predicted ^{1}H chemical shifts[a] of the protons in UCAUGA and AAGCUU[b]

123456 UCAUGA				123456 AAGCUU			
Proton	Observed[c]	Predicted[d]	Δδ[e]	Proton	Observed[c]	Predicted[d]	Δδ[e]
U(1)H-6	7.709	7.745	+0.036	A(1)H-8	8.196	8.188	+0.008
C(2)H-6	7.663	7.744	+0.081[f]	A(1)H-2	8.103	8.083	+0.020
A(3)H-8	8.339	8.344	+0.005	A(2)H-8	8.237	8.231	+0.006
A(3)H-2	8.182	8.191	+0.009	A(2)H-2	8.128	8.103	+0.025
U(4)H-6	7.678	7.665	-0.013	G(3)H-8	7.867	7.902	-0.035
G(5)H-8	7.889	7.875	-0.014	C(4)H-6	7.717	7.781	-0.064[f]
A(6)H-8	8.320	8.330	+0.010	U(5)H-6	7.769	7.851	-0.082[f]
A(6)H-2	8.160	8.199	+0.039	U(6)H-6	7.791	7.836	-0.045[f]
U(1)H-1'	5.911	5.829	-0.082[f]	A(1)H-1'	5.941	5.929	+0.012
C(2)H-1'	5.833	5.827	-0.006	A(2)H-1'	5.941	5.979	-0.038
A(3)H-1'	6.016	6.074	+0.058[f]	G(3)H-1'	5.741	5.827	-0.086[f]
U(4)H-1'	5.813	5.749	-0.064[f]	C(4)H-1'	5.893	5.879	+0.014
G(5)H-1'	5.787	5.761	-0.026	U(5)H-1'	5.880	5.886	-0.006
A(6)H-1'	6.057	6.098	+0.041[f]	U(6)H-1'	5.893	5.921	+0.028
U(1)H-5	5.806	5.810	+0.004	C(4)H-5	5.827	5.860	-0.033
C(2)H-5	5.958	5.985	+0.027	U(5)H-5	5.827	5.832	-0.005
U(4)H-5	5.748	5.687	-0.061[f]	U(6)H-5	5.884	5.892	-0.008

Footnotes as in table 1

because the overlap of resonances makes assignments, and hence sequence determination, difficult. The imminent availability of very high field magnets will also solve this resolution problem, and most resonances of a 24 base DNA duplex have been assigned [23].

References

[1] Simoncsits, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. and Guilley, H. (1977) Nature 269, 833–836.
[2] Silberklang, M., Gillum, A. M. and RajBhandary, U. L. (1977) Nucleic Acids Res. 4, 4091–4108.
[3] Brownlee, G. G. and Cartwright, E. M. (1977) J. Mol. Biol. 114, 93–117.
[4] Donis-Keller, H., Maxam, A. M. and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527–2538.
[5] Peattie, D. A. (1979) Proc. Natl. Acad. Sci. USA 76, 1760–1764.
[6] Sanger, F. and Coulson, A. R. (1975) J. Mol. Biol. 94, 441–448.
[7] Maxam, A. M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.

[8] Alkema, D., Bell, R. A., Hader, P. A. and Neilson, T. (1981) J. Am. Chem. Soc. 103, 2866–2868.
[9] Donis-Keller, H. (1979) Nucleic Acids Res. 7, 179–192.
[10] Everett, J. R., Hughes, D. W., Bell, R. A., Alkema, D., Neilson, T. and Romaniuk, P. J. (1980) Biopolymers 19, 557–573.
[11] Bell, R. A., Everett, J. R., Hughes, D. W., Alkema, D., Hader, P. A., Neilson, T. and Romaniuk, P. J. (1981) Biopolymers 20, 1383–1398.
[12] Hader, P. A., Alkema, D., Bell, R. A. and Neilson, T. (1981) J. Chem. Soc. Chem. Commun. in press.
[13] Werstiuk, E. S. and Neilson, T. (1976) Can. J. Chem. 54, 2689–2696.
[14] Neilson, T., Kofoid, E. C. and Ganoza, M. C. (1980) Nucleic Acids Res. Symp. ser. 7, 313–324.
[15] Neilson, T., Grégoire, R. J., Fraser, A. R., Kofoid, E. C. and Ganoza, M. C. (1979) Eur. J. Biochem. 99, 429–437.
[16] Laskey, R. A. and Mills, A. D. (1977) FEBS Lett. 82, 314–316.
[17] Ts'o, P. O. P., Kondo, H. S., Schweizer, M. P. and Hollis, D. P. (1969) Biochemistry 8, 997–1029.
[18] Martin, F. H., Uhlenbeck, O. C. and Doty, P. (1971) J. Mol. Biol. 57, 201–215.
[19] Lee, C., Ezra, F. S., Kondo, N. S., Sarma, R. H. and Danyluk, S. S. (1976) Biochemistry 15, 3627–3638.
[20] Ezra, F. S., Lee, C., Kondo, N. S., Danyluk, S. S. and Sarma, R. H. (1977) Biochemistry 16, 1977–1987.
[21] Dhingra, M. M. and Sarma, R. H. (1979) in: Stereodynamics of Molecular Systems (Sarma, R. H. ed) p. 15, Pergamon Press, New York.
[22] Giessner-Prettre, C., Pullman, B., Borer, P. N., Kan, L.-S. and Ts'o, P. O. P. (1976) Biopolymers 15, 2277–2286.
[23] Early, T. A. and Kearns, D. R. (1980) Nucleic Acids Res. 8, 5795–5812.