ELSEVIER

Journal of Chromatography A, 864 (1999) 49–57

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

High-performance liquid chromatography separation methods for the analysis of peptide nucleic acids

Yingning Wei[a], Michael Marino[b], Barry Thompson[b], James E. Girard[a,*]

[a]*Department of Chemistry, American University, 4400 Massachusetts Avenue, Washington, DC 20016, USA*
[b]*Center for Medical and Molecular Genetics, Armed Forces Institute of Pathology, Rockville, MD 20850, USA*

Received 15 April 1999; received in revised form 16 September 1999; accepted 17 September 1999

Abstract

An analytical analysis of peptide nucleic acids (PNAs) was carried out by reversed-phase HPLC using a solvent system comprised of aqueous trifluoroacetic acid and acetonitrile. A regression equation was obtained which represents the relationship of the molecular mass, sequence composition and retention time. This equation can be used to estimate the retention time of a known PNA under certain HPLC conditions. In addition to this equation, new HPLC conditions were also optimized which can be used for separation of pure PNAs.

Keywords: Oligonucleotides; Peptide nucleic acids

1. Introduction

Peptide nucleic acids (PNAs) were developed in 1991 by researchers in the Panum Institute and the Riso National Laboratory in Denmark. PNAs are structural mimics of DNA where the regular sugar–phosphate is replaced by repeating *N*-(2-aminoethyl)glycine units (Fig. 1). Thus, a methylene carbonyl linker to this backbone at amino nitrogen connects the standard nucleotide bases in PNAs. This chemistry has three important consequences. First, PNAs are neutral molecules. Second, PNAs are achiral, which avoids the need to develop a stereoselective synthesis. Third, PNA synthesis uses standard benzhydryloxycarbonyl (Bhoc) or 9-fluorenylmethoxycarbonyl (Fmoc) protocols for solid-phase peptide synthesis [2–5].

A standard PNA structure contains all four natural nucleobases. They can hybridize to complementary oligonucleotides obeying the Watson–Crick base-pairing rules. Nucleobase adenine in the PNA sequence can form weak hydrogen bonds with nucleobase thymine in its complementary DNA sequence. The nucleobase cytosine in the PNA sequence can form weak hydrogen bonds with nucleobase guanine with its complementary DNA sequence [6,7]. Because the PNA backbone is uncharged, the PNA–DNA duplex lacks electrostatic repulsion between the two chains. Thus, the PNA–DNA duplex is more stable than the corresponding DNA–DNA duplex and has a higher melting temperature. When a Watson–Crick base-pair mismatch is introduced in the oligonucleotide at any position facing the PNA nucleobases, the melting temperature (T_m) will decrease by 8–20°C. It is also noticed that for virtually all base-pair mismatches, the decrease in thermal stability is greater for the PNA–DNA complex than for the DNA–DNA complex [8–11]. So,

*Corresponding author.

0021-9673/99/$ – see front matter
PII: S0021-9673(99)01013-4

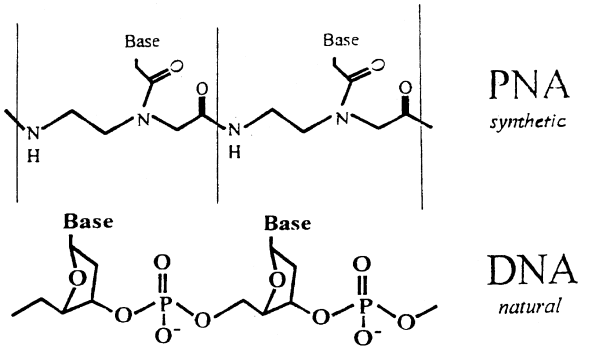

Fig. 1. Comparison of PNA and DNA structure. The number of bonds (in bold) between the nucleotide units in PNA are the same as in the sugar–phosphate unit in DNA (i.e., six), and the number of bonds from the core of the backbone to the base (dashed lines) is also identical to that found in DNA (i.e., three). The design of the PNA originated from a simple computer model of DNA, where the neutral backbone was replaced with an achiral one that had the same inter-base distances as natural DNA, and sufficient flexibility to allow base pairing with RNA and DNA. From Ref. [1].

the sequence discrimination is more efficient for PNA recognizing DNA than DNA recognizing DNA. Because of all these properties of PNA, PNAs provide a great opportunity to develop a DNA affinity capture assay [12–16]. This typing method represents a powerful alternative and improvement to all traditional DNA typing methods.

A DNA affinity capture assay can used for forensic purposes. Recently, a new technique called matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), has been used for HLA-DQα DNA typing [17]. Because of the difficulties of DNA analysis by MALDI-TOF-MS, PNAs were used for DNA typing in this method [18,19]. Although MALDI-TOF-MS is more advanced than the traditional polymerase chain reaction (PCR)-based methods for DNA typing, it is not a perfect technology. The reproducibility of MALDI-TOF-MS is not always adequate. If the laser hits different areas of the same sample, it may give different results. For the same sample, the mass spectra may differ with laser power. Impurities may not appear with a low power laser, but they will appear with a high power laser. Also, MALDI-TOF-MS is not a fundamental instrument that is found in every analytical chemistry laboratory. Because of its high cost (US$200 000) and the shortcomings described, there is a need to develop an alternative method for DNA typing. One method which has shown potential is reversed-phase high-performance liquid chromatography (RP-HPLC) [20,21]. HPLC is a natural analytical technique that offers the potential for cost-effective DNA analysis.

HPLC characterization is also important for PNA analysis. Since there are many failure sequences in synthesized PNAs, all PNA probes need to be HPLC-purified before they can be used for the assay. It is important to know that a certain HPLC peak is the PNA with the desired sequence. Once we knew the relationship between the relative molecular mass of PNA probe and the retention time, the purification can be easily done. After this relationship is established, the HPLC conditions for separation of PNAs can optimized. The performance of the developed RP-HPLC methods were evaluated in terms of resolution, sensitivity and reproducibility. The results obtained by HPLC were confirmed by MALDI-MS.

2. Experimental

2.1. Materials

All PNAs were synthesized on a 2 μmol scale on Expedite PNA columns (PerSeptive Biosystems, Framingham, MA, USA) using an Expedite Nucleic Acids Synthesis System (PerSeptive Biosystems), which had been converted to peptide synthesis conditions. The synthesis was carried out using Merrifield's Nobel Prize winning Fmoc Chemistry peptide synthesis on a solid support and reagents were supplied by PerSeptive Biosystem. A Fmoc group protected each PNA monomer on the amino (N) terminus, which was base-labile. Following deprotection, the PNA strand connected to the solid-support was extended via an amide linkage to the carboxyl (C) terminus of the next monomer. Thus, a PNA molecule was built in the C→N direction. Failure sequences may be generated during each step due to impurities in the monomers or incomplete coupling between the new monomer and the growing PNA strand. Post-synthesis procedures involved cleaving the PNA product from the solid support with a trifluoroacetic acid (TFA)–*m*-cresol (4:1) solution and precipitating the product in ice-cold

diethyl ether. The resulting PNA pellet was dissolved in 0.1% TFA (aqueous).

TFA was peptide synthesis grade and supplied by Applied Biosystems (Foster City, CA, USA). Acetonitrile (ACN) was HPLC grade and was supplied by Sigma (Louis, MO, USA)–Aldrich (Milwaukee, WI, USA). Purified water was obtained with a Milli-Q system from Millipore (Bedford, MA, USA).

2.2. Instruments

A Dionex DX 300 HPLC system (Dionex, Sunnyvale, CA, USA) was used with the variable-wavelength UV detector set at 260 nm. A 150 mm×4.6 mm I.D. reversed-phase Zorbax C_{18} column (DuPont) was used.

2.3. Methods

2.3.1. HPLC purification of crude PNAs

A gradient separation was performed with 0.1% TFA in 100% acetonitrile (solvent A), 0.1% TFA in 1% acetonitrile (solvent B) and 0.08% TFA in 50% acetonitrile (solvent C): 0–5 min 100% B; 5–45 min 91% to 60% B and 9% to 40% C; 45–50 min 100% A. The column temperature was set at 55°C.

2.3.2. HPLC separation of pure PNAs

A step gradient separation was performed: 0–15 min 82% B and 18% C; 15–25 min 82% to 50% B and 18% to 50% C; 25–30 min 100% A; 30–33 min 100% B; 33–43 min 82% B and 18% C. All the separations were carried out at 55°C. The flow-rate was set at 0.6 ml/min.

2.3.3. Stability of PNAs

After the post-synthesis procedure, the resulting PNA was dissolved in 400 μl 0.1% TFA (aqueous). PNAs not used immediately should be stored in the freezer.

PNA oligomers have an affinity for glass surfaces and polystyrene. When working with low concentrations of PNAs, much of the PNA may become bound to the container. Whenever possible, use polypropylene or polyethylene materials for handling and storage of PNAs.

3. Results and discussion

3.1. Analysis of PNAs by RP-HPLC

During the synthesis of PNA probes, some failure sequences are produced. This can be caused by the impurities in the monomer or by failure of the attachment reactions in the PNA synthesis. All these failure sequences will interfere with the PNA applications. So, the PNA probes need to be purified and the failure sequences need to be removed before they can be used in any assay.

To establish the optimum conditions for the analysis of different PNAs by HPLC, an initial trial run was performed starting with 0% C and going to 60% C in 50 min. This experiment allowed an estimate of the percent C that was necessary to elute the very first PNA and also the percent C that eluted the very last PNA. In this experiment, most of the PNAs were eluted in the 20–30 min time range. The relationship between the retention time and the molecular mass of a PNA was examined by the analysis of 20 different PNAs. As shown in Fig. 2, there was a satisfactory linear relation between the retention time and the molecular mass of a PNA. A regression equation was obtained: $y=20.76131-0.00513M_r+1.436339A+2.555066G+1.418646C+2.281181T$, which represents the relationship of the molecular mass, sequence composition and retention time. In this equation, *A*, *G*, *C* and *T* refer to the number of each base in a PNA sequence. Thus, this equation can be used to estimate the retention time of a known PNA under these HPLC conditions.

From this equation, it was found that the retention time is positively correlated to molecular mass. This means that when molecular mass increases, the retention time also increases. The equation also demonstrated the effect of each nucleotide base on the separation. The coefficient for *G* (2.55) and *T* (2.28) are much larger than for *A* (1.43) or *C* (1.41). This means that substitution of *G* or *T* into a PNA structure will lengthen the retention time more than *A* or *C*. So, this equation can be used to estimate the retention time of a particular PNA under specific HPLC conditions. Several other facts were found from this equation. First, when *G* or *T* is absent in a PNA sequence, the actual retention time deviates more from the calculated retention time. This is

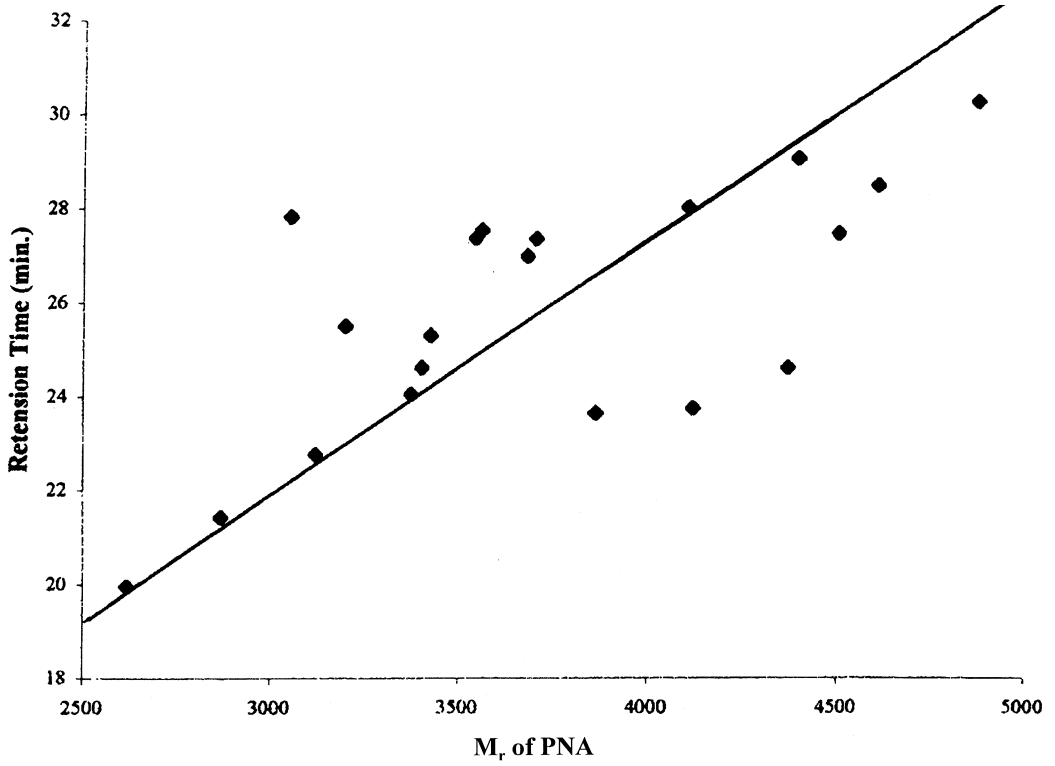

Fig. 2. Relationship between retention time and molecular mass of PNAs. R^2=0.869375.

because that *G* and *T* are more hydrophilic than *A* and *C*. So, when *G* or *T* is absent, the interaction between PNA and column stationary phase will increase, thus, the retention time will increase. Second, from the statistical analysis, all bases have some influence on the retention time. All of them influence the equation. If we know the size of a PNA, then we can estimate its retention time using this equation.

3.2. Optimization on HPLC for the separation of pure PNAs

When solution of the four pure PNAs was analyzed by HPLC using the gradient program, which was described above for purification of the crude PNAs, the four probes had a long elution time. There were no peaks in the first 15 min of the chromatogram. The PNA with 12 nucleobases (*CAG CAA ATT TGG*) was eluted first at the retention time about 16 min. So it was necessary to develop another HPLC method to separate these four PNAs. A new HPLC method for separation of these four PNAs was developed, and one trial run was performed using gradient from 5% to 95% C and 95% to 5% B in 60 min. Column temperature was set to 55°C. The initial and final percentage of solvent C, which contains 0.08% TFA in 50% ACN, was estimated from this trial run using a method described by Kirkland et al. [22].

3.2.1. Flow-rate analysis

From the trial run, it was found that the PNA with 15 nucleobases (*CGC AGA TTT AGA AGA*) and the PNA with 14 nucleobases (*TTT GCC TGT TCT CA*) were the most difficult PNAs to separate. So, these two PNAs were studied first to optimize the separation. For this study, the optimum gradient was found to be: 81% B to 75% B and 19% C to 25% C in 0–10 min, 100% A in 10–15 min. The initial concentration of solvent C and the final concentration of solvent C were decided from the trial run.

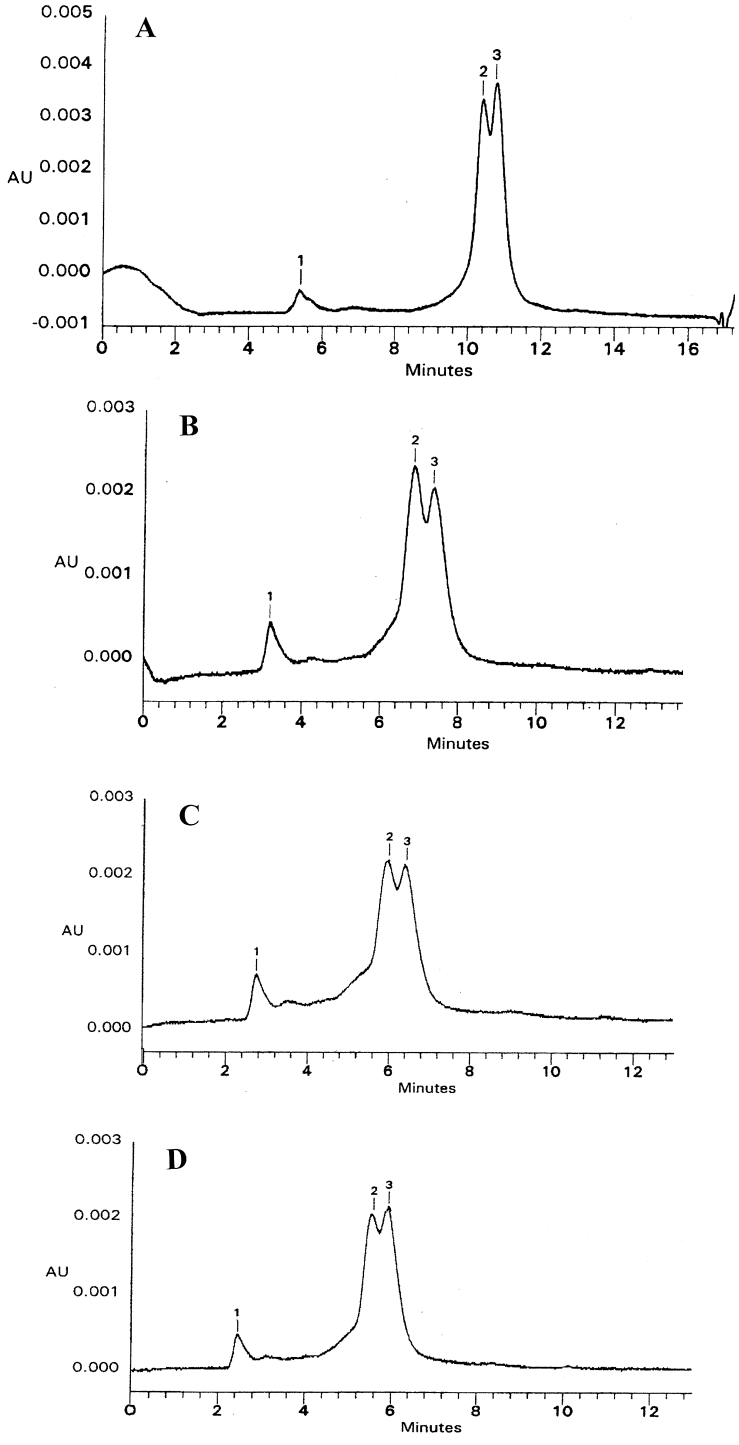

Fig. 3. The effect of flow-rate on the separation of two PNAs. Four different flow-rates were tested and the resolutions were calculated. The best resolution was obtained when the flow-rate was 0.6 ml/min. (A) 0.4 ml/min, R_s=0.32; (B) 0.6 ml/min, R_s=0.38; (C) 0.7 ml/min, R_s=0.30; (D) 0.8 ml/min, R_s=0.26.

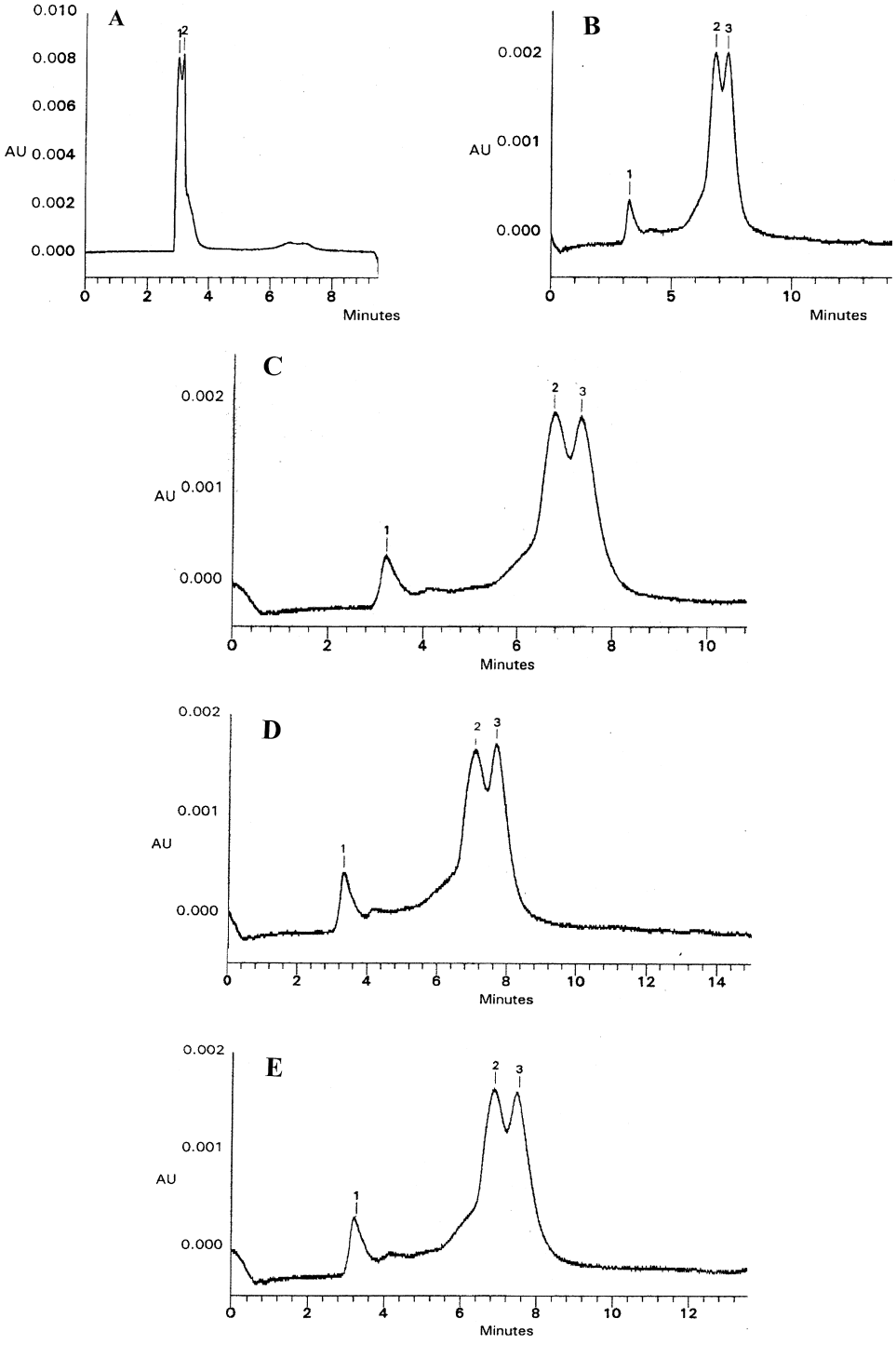

Fig. 4. The effect of gradient steepness on separation. The resolution was initially increased when the gradient steepness of the percentage of solvent C decreased. But when the gradient steepness decreased from 0.4%/min to 0.24%/min, the resolution was barely changed. It meant that the gradient steepness did not have a great effect on improving the separation for these two PNA probes. (A) 1.2%/min, R_s=0.25; (B) 0.6%/min, R_s=0.32; (C) 0.4%/min, R_s=0.36; (D) 0.3%/min, R_s=0.37; (E) 0.24%/min, R_s=0.37.

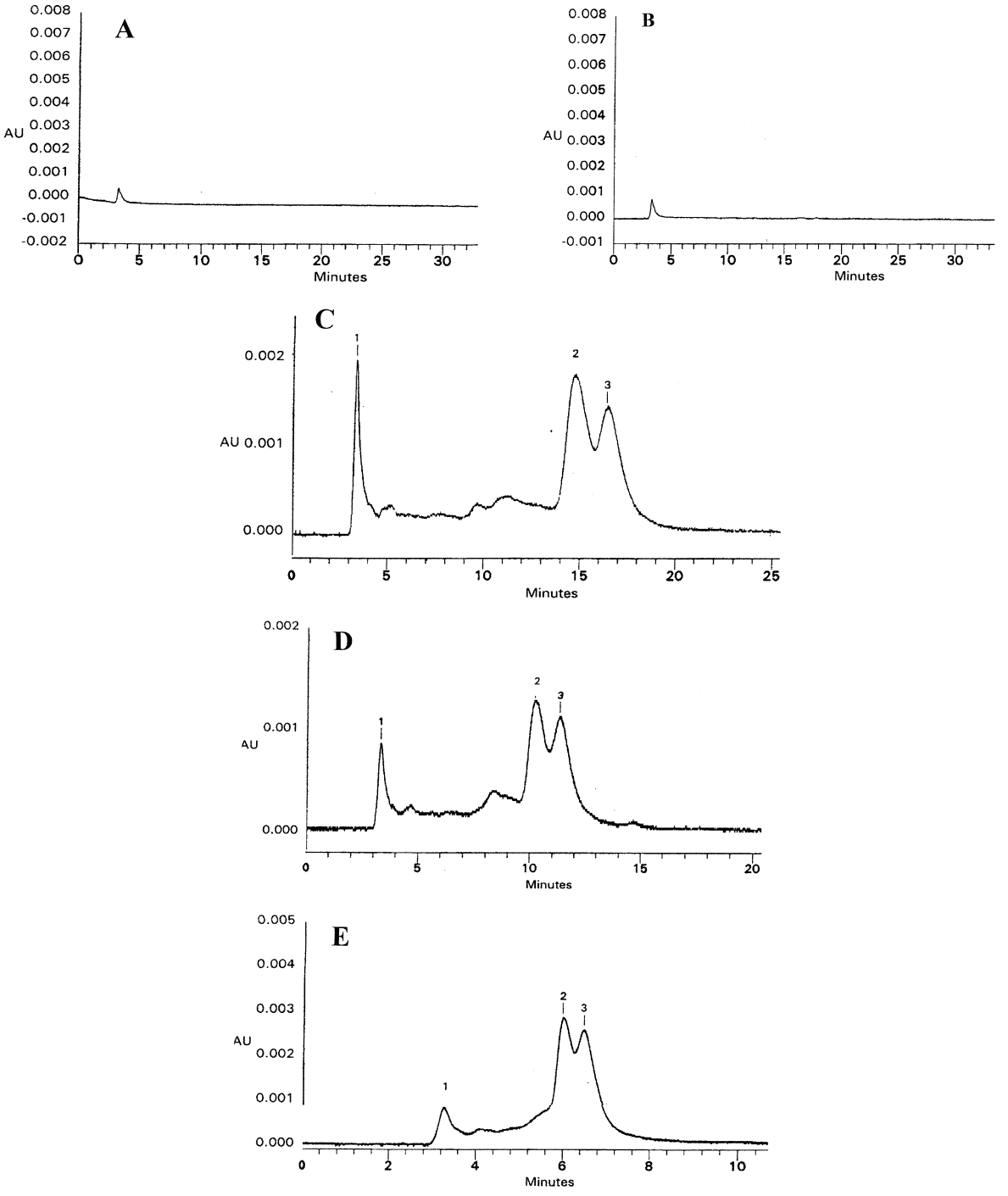

Fig. 5. The isocratic separation of two PNA probes. Solvent C contained 0.08% TFA/50% ACN. Acetonitrile was considered to be the driving force for eluting the PNA probes. When the percentage of solvent C was 10% and 14%, these two PNAs would be eluted after 30 min. The more solvent C that was used, the earlier the probes were eluted. (A) B–C (90:10); (B) B–C (86:14); (C) B–C (83:17), R_s=0.65; (D) B–C (82:18), R_s=0.60; (E) B–C (80:20), R_s=0.41.

Fig. 3 illustrates the effect of flow-rate on the separation. All PNAs were eluted earlier with a faster flow-rate. Resolution, R_s, between the two peaks was calculated based on this equation: $R_s=\Delta t/2(W_1+W_2)$ where Δt is the difference of retention time between the two peaks, W_1 and W_2 are the peak widths for the 15 mer and 14 mer, respectively. As can be seen in Fig. 3, a flow-rate of 0.8 ml/min gave the worst resolution (0.26) and 0.6 ml/min showed the best resolution (0.38).

3.2.2. Effect of gradient

Gradient steepness is usually described in terms of change in % concentration/min. When the gradient steepness is changed, it should change the separation of any two PNAs. The changes to the PNA peaks include the elution time, the shapes of the peaks, and the resolution of the separation. Fig. 4 shows the separation of these two PNAs at different steepness of gradient. The resolution of the separation was calculated. It was found that the resolution increased from 0.25 to 0.36 as gradient steepness decreased. As gradient steepness was reduced from 0.4%/min to 0.24%/min, the resolution R_s only changed from 0.36 to 0.37. This suggested that the steepness of gradient has a small effect on improving the separation for the two PNAs.

3.2.3. Isocratic analysis

Since the gradient steepness had little effect on the separation, isocratic conditions were used. Five different isocratic conditions were studied. Fig. 5 illustrates the effect of changing the percentage of solvent C (0.08% TFA/50% ACN) on the resolution. In the analysis of these peptides, acetonitrile was considered to be the reagent pushing the analytes through the column. The more acetonitrile in the mobile phase, the earlier the PNA analyte would be eluted. When 10% C and 14% C was used for the analysis, the mobile phase was too weak to elute the two PNAs before 30 min. So, no peaks were observed in the chromatogram for 10% C and 14% C. As the percentage of solvent C increased from 17% C to 20% C, the two PNAs were eluted faster. Although the combination of 83% B and 17% C showed the best resolution, the peaks were wide, indicating low efficiency. So, a binary solvent system with 82% B and 18% C was chosen as a compromise of resolution and efficiency as the best conditions for the separation of these two PNAs.

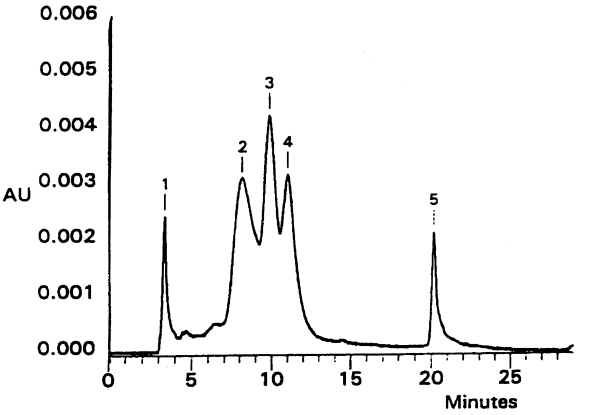

Fig. 6. Fast separation of the four pure PNAs. Peaks: 1=dwell volume of the HPLC system; 2=12mer, *CAG CAA ATT TGG*; 3=15mer, *CGC AGA TTT AGA AGA*; 4=14mer, *TTT GCC TGT TCT CA*; 5=17mer, *TCC ACA GAC TTA GAT TT*.

3.2.4. Fast separation of four PNA probes

After optimal conditions were developed for the separation of the 15-nucleobase and the 14-nucleobase PNA, all four PNAs were analyzed by HPLC using the same isocratic conditions. The four PNAs were well separated by these HPLC conditions. The problem was that the 17-nucleobase PNA (*TCC ACA GAC TTA GAT TT*) eluted very late (24 min) and with a wide peak shape. So, a step gradient was used in order to elute this 17-nucleobase PNA sooner and to improve the peak shape. Fig. 6 shows the separation of these four pure PNAs under the new HPLC conditions developed for the separation of these PNAs.

Acknowledgements

The authors thank Joe Devaney, Ann Davis and Qi Liang for their assistance. Y.W. and J.E.G. thank the staff of the Center for Medical and Molecular Genetics for the use of their facilities and their mass spectrometer. The views stated here are the opinions of the authors and in no way reflect the position of the US Army, US Air Force, or US Department of Defense.

References

[1] Expedite 8900 PNA Chemistry User's Guide, PerSeptive Biosystems, Framingham, MA.
[2] D.R. Corey, TIBTECH 15 (1997) 224.
[3] B. Hyrup, P.E. Nielsen, Bioorg. Med. Chem. 4 (1996) 5.
[4] P.E. Nielsen, M. Egholm, O. Buchardt, Bioconjugate Chem. 5 (1994) 3.
[5] M. Eriksson, P.E. Nielsen, Quart. Rev. Biophys. 29 (1996) 369.
[6] V.V. Demidov, D.I. Cherny, A.V. Kurakin, M.V. Yavnilovich, V.A. Malkov, M.D. Frank-Kamenetskii, S.H. Sonnichsen, P.E. Nielsen, Nucl. Acids Res. 22 (1994) 5218.
[7] S.K. Kim, P.E. Nielsen, M. Egholm, O. Buchardt, R.H. Berg, B. Norden, J. Am. Chem. Soc. 115 (1993) 6477.
[8] T. Bentin, P.E. Nielsen, Biochemistry 35 (1996) 8863.
[9] K.K. Jensen, H. Orum, P.E. Neilsen, B. Norden, Biochemistry 36 (1997) 5072.
[10] K.L. Taneja, BioTechniques 24 (1998) 472.
[11] H.P. Keefe, X. Yao, J.M. Coull, M. Fuchs, M. Egholm, Proc. Natl. Acad. Sci. USA 93 (1996) 14670.
[12] E.A. Lesnik, L.M. Risen, D.A. Driver, M.C. Griffith, K. Sprankle, S.M. Freier, Nucl. Acids Res. 25 (1997) 568.
[13] A.F. Faruqi, M. Egholm, P.M. Glazer, Proc. Natl. Acad. Sci. USA 95 (1998) 1398.
[14] J. Weiler, H. Gausepohl, N. Hauser, O.N. Jensen, J.D. Hoheisel, Nucl. Acids Res. 25 (1997) 2792.
[15] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S.M. Freier, D.A. Driver, R.H. Berg, S.K. Kim, B. Norden, P.E. Nielsen, Nature 365 (1993) 566.
[16] O. Almarsson, T.C. Bruice, Proc. Natl. Acad. Sci. USA 90 (1993) 9542.
[17] P. Jiang-Baucom, J.E. Girard, J. Butler, P. Belgrader, Anal. Chem. 69 (1997) 4894.
[18] P.L. Ross, K. Lee, P. Belgrader, Anal. Chem. 69 (1997) 4197.
[19] T.J. Griffin, W. Tang, L.M. Smith, Nature Biotechnol. 15 (1997) 1368.
[20] P. Jandera, L. Petranek, M. Kucerova, J. Chromatogr. A 791 (1997) 1.
[21] K. Rowland, L. Lennard, J.S. Lilleyman, J. Chromatogr. B 705 (1998) 29.
[22] L.R. Snyder, J.J. Kirkland, J.L. Glajch, Practical HPLC Method Development, 2nd ed., Wiley, New York, 1997.