## Noncovalent Protein–Oligonucleotide Interactions Monitored by Matrix-Assisted Laser Desorption/ Ionization Mass Spectrometry

### Xiaodong Tang,<sup>†</sup> John H. Callahan,<sup>‡</sup> Ping Zhou,<sup>§</sup> and Akos Vertes<sup>\*,†</sup>

Department of Chemistry, The George Washington University, Washington, D.C. 20052, Chemistry Division, Naval Research Laboratory, Washington, D.C. 20375, and Cruachem, Inc., Dulles, Virginia 20166

Positive ion mode matrix-assisted laser desorption/ ionization mass spectrometry (MALDI-MS) was used to explore nonspecific interactions between proteins and oligonucleotides. The formation of noncovalent complexes showed correlation with the type of oligonucleotide bases and with the amino acid composition of the proteins. Among the four DNA homooligomers, abundant protein-nucleic acid complexes were detected for  $pd(T)_n$ . whereas negligible attachment was evident for  $pd(A)_n$ ,  $pd(C)_n$ , and  $pd(G)_n$ . Mixed base sequence nucleic acids (pd(AGCTCAGCTT) and d(TTAGCAGCTT)) also showed affinity to Arg-Lys. The protein affinity of  $pd(T)_n$  turned out to be nonspecific and produced a larger variety of complexes when the number of basic residues in the protein was increased. Complexation of  $pd(T)_n$  with small basic dipeptides (Arg-Lys or His-His) led to significant improvement in the mass resolution for positive ions. For example, the mass resolution of the pd(T)<sub>20</sub>/Arg-Lys complex exhibited about 4 times improvement over pd-(T)<sub>20</sub> alone. The protein-oligonucleotide interactions were also pH and matrix dependent. Lowering the pH from its original value (pH = 1.7) led to diminishing complex related signal, whereas increasing the pH resulted in the appearance of a larger variety of complexes. 2.5-Dihydroxybenzoic acid matrix demonstrated much greater tendency to produce complex ions than did the three other matrix materials we tested. A possible explanation of the observed phenomena was based on pHcontrolled ion pair formation between oligonucleotides and proteins.

New methods for fast and reliable analysis of nucleic acids are being pursued on different avenues by both separation scientists and spectroscopists. The driving force behind this quest is the multitude of potential applications ranging from biomedical research to environmental monitoring. With the rapid development of new soft ionization techniques, such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), mass spectrometry is playing an increasingly important role in analyzing nonvolatile, thermally labile biopolymers.<sup>1-4</sup> A wide

(2) Mann, M.; Meng, C. K.; Fenn, J. B. Anal. Chem. 1989, 61, 1702-1708.
(3) Beavis, R. C.; Chait, B. T. Anal. Chem. 1990, 62, 1836-1840.

4542 Analytical Chemistry, Vol. 67, No. 24, December 15, 1995

variety of biomolecules, including peptides, proteins,<sup>5</sup> oligosaccharides,<sup>6</sup> and oligonucleotides,<sup>7</sup> have been detected at low picomole levels by MALDI-MS. The analysis of oligonucleotide mixtures by this method has also been regarded as a candidate for fast DNA sequencing.<sup>8-11</sup>

In the past 2 years, considerable effort has been devoted to the search for suitable matrices that efficiently ionize oligonucleotides. Becker and co-workers reported that the application of 3-hydroxypicolinic acid matrix substantially enhanced the detection of oligonucleotides.<sup>12</sup> Subsequently, Chen and co-workers at the Oak Ridge National Laboratory discovered that picolinic acid worked well as a matrix for mixed-base oligonucleotides of up to 190 bases.<sup>13</sup> In a related study, the same group successfully detected double-stranded DNA with 246 base pairs by using 3-aminopicolinic acid matrix.<sup>14</sup> Because of the acidic property of the phosphate groups in DNA, negative ions were typically more abundant than positive species. In spite of encouraging progress, the sensitivity and resolution of oligonucleotide analysis were still lower than the corresponding values established for proteins. This was, in part, due to lower ionization efficiencies, the formation of cation adducts, and fragmentation in MALDI.<sup>15-17</sup> The "softness" of MALDI for nucleic acid analysis has been explored in relation to laser wavelength, laser irradiance, the type of mass analyzer, sample preparation, and applied matrices.<sup>18</sup> The nature of the matrix was found to play a key role in determining the

- (5) Hillenkamp, F.; Karas, M.; Chait, B. T.; Beavis, R. C. Anal. Chem. 1991, 63, 1193A-1203A.
- (6) Stahl, B.; Steup, M.; Karas, M.; Hillenkamp, F. Anal. Chem. 1991, 63, 1463– 1466.
- (7) Tang, K.; Allman, S. L.; Chen, C. H.; Chang, L. Y.; Schell, M. Rapid Commun. Mass Spectrom. 1994, 8, 183–186.
- (8) Fitzgerald, M. C.; Zhu, L.; Smith, L. M. Rapid Commun. Mass Spectrom. 1993, 7, 895-897.
- (9) Bai, J.; Liu, Y. H.; Lubman, D. M. Rapid Commun. Mass Spectrom. 1994, 8, 687-691.
- (10) Kenough, T.; Baker, T. R.; Dobson, R. L. M.; Lacey, M. P.; Riley, T. A.; Hasselfield, J. A.; Hesselberth, P. E. *Rapid Commun. Mass Spectrom.* 1993, 7, 195-200.
- (11) Pieles, U.; Zurcher, W.; Schar, M.; Moser, H. E. Nucleic Acids Res. 1993, 21, 3191-3196.
- (12) Wu, K. J.; Steding, A.; Becker, C. H. Rapid Commun. Mass Spectrom. 1993, 7, 142-146.
- (13) Tang, K.; Taranenko, N. I.; Allman, S. L.; Chen, C. H.; Chang, L. Y.; Jacobson, K. B. Rapid Commun. Mass Spectrom. 1994, 8, 673-677.
- (14) Taranenko, N. I.; Tang, K.; Allman, S. L.; Chang, L. Y.; Chen, C. H. Rapid Commun. Mass Spectrom. 1994, 8, 1001-1006.
- (15) Scheider, K; Chait, B. T. Org. Mass Spectrom. 1993, 28, 1353-1361.
- (16) Nordhoff, E.; Cramer, R.; Karas, M.; Hillenkamp, F.; Kirpekar, F.; Kristiansen, K.; Roepstorff, P. Nucleic Acids Research, 1993, 21, 3347-3357.
- (17) Huth-Fehre, T.; Gosine, J. N.; Wu, K. J.; Becker, C. H. Rapid Commun. Mass Spectrom. 1992, 6, 209-213.

0003-2700/95/0367-4542\$9.00/0 © 1995 American Chemical Society

<sup>&</sup>lt;sup>•</sup>Phone: (202) 994-2717. Fax: (202) 994-5873. E-mail: vertes@gwis2.circ. gwu.edu.

<sup>&</sup>lt;sup>†</sup> The George Washington University.

<sup>&</sup>lt;sup>†</sup> Naval Research Laboratory.

<sup>§</sup> Cruachem, Inc.

<sup>(1)</sup> Karas, M.; Hillenkamp, F. Anal. Chem. 1988, 60, 2299-2301.

<sup>(4)</sup> Vertes, A.; Gijbels, R. In Laser Ionization Mass Analysis; Vertes, A., Gijbels, R., Adams, F., Eds.; John Wiley & Sons: New York, 1993; pp 127-175.

softness of volatilization and ionization processes, based on correlations with metastable fragmentation.<sup>18,19</sup>

Interactions between nucleic acids and proteins play an important role in numerous biochemical processes, including DNA replication, recombination, and repair. Despite the new opportunities provided by MALDI and ESI, there have only been sporadic examples of the use of mass spectrometry to study noncovalent binding of biomolecules. Attachment of peptides to other biomolecules, including DNA, was successfully detected by ESI-MS.<sup>20–24</sup> Because of the appearance of multiply charged analyte ions, however, the presence of numerous species in the nucleic acid protein mixtures led to complex mass spectra.

Juhasz and Biemann showed that, due to the lack of multiply charged ions, MALDI was a viable alternative to investigate noncovalent interactions.<sup>25</sup> This study revealed that attachment could be observed between highly acidic DNA and basic peptides. The largest complexes detected in their investigation were those of histone H4 (MW 11 236) and  $d(T)_{10}$  (MW 2980). It was reported that complex formation did not improve either the sensitivity or the resolution in MALDI analysis of oligonucleotides. Recently, Barofsky and co-workers investigated the MALDI-MS of phage T4 gene 32 protein (gp32) attachment to the oligonucleotide pd(T)<sub>20</sub>.<sup>26</sup> This species, however, was formed by UV-induced cross-linking, leading to covalent bonds. The successful detection of protein–oligonucleotide adducts was found to be dependent on the choice of solvents and on the presence of additives mixed with the matrix.

In the light of the importance of protein-oligonucleotide interactions, the objective of this study was to identify the structural features that drive the complex formation. We addressed the question of whether specific amino acid side chains or nucleotide bases were necessary for the observation of complexes. Particularly puzzling was the nature of the interaction that brought and/or kept these relatively large units together throughout the MALDI process. To choose favorable conditions for the observation of complexes, several matrices were investigated, which in turn gave insight into the factors controlling the complexation process. The polyprotic nature of both nucleic acids and proteins dictated that acid-base equilibria governed some of the association processes in the highly acidic matrix solution. Possibilities to utilize protein-oligonucleotide complexes for improving sensitivity and resolution in MALDI analysis of DNA and the nature of protein-oligonucleotide interactions in concentrated matrix solutions will be discussed.

- (18) Nordhoff, E.; Kirpekar, F.; Karas, M.; Cramer, R.; Hahner, S.; Hillenkamp, F.; Kristiansen, K.; Roepstorff, P.; Lezius, A. Nucleic Acids Res. 1994, 22, 2460-2465.
- (19) Preston-Schaffter, L. M.; Kinsel, G. R.; Russell, D. H. J. Am. Soc. Mass Spectrom. 1994, 5, 800–806.
- (20) Camilleri, P.; Haskins, N. J.; New, A. P.; Saunders, M. R. Rapid Commun. Mass Spectrom. 1993, 7, 949–952.
- (21) Haskins, N. J.; Ashcrot, A. E.; Phillips, A.; Harrison, M. Rapid Commun. Mass Spectrom. 1994, 8, 120-125.
- (22) Hsieh, Y. L.; Li, Y. T.; Henion, J. D.; Ganem, B. Biol. Mass Spectrom. 1994, 23, 272–276.
- (23) Cheng, X.; Gao, Q.; Smith, R. D.; Simanek, E. E.; Mammen, M.; Whitesides, G. M. Rapid Commun. Mass Spectrom. 1995, 9, 312–316.
- (24) Smith, R. D.; Light-Wahl, K. J. Biol. Mass Spectrom. 1993, 22, 493-501.
- (25) Juhasz, P.; Biemann, K. Proc. Natl. Acad. Sci. U.S.A. 1994, 91, 4333-4337.
- (26) Jensen, O. N.; Barofsky, D. F.; Young, M. C.; Hippel, P. H.; Swenson, S.; Seifried, S. E. Rapid Commun. Mass Spectrom. 1993, 7, 496-501.
- (27) Olumee, Z.; Sadeghi, M.; Tang, X.; Vertes, A. Rapid Commun. Mass Spectrom. 1995, 9, 744-752.

### **EXPERIMENTAL SECTION**

Instrumentation. The MALDI-MS experiments were carried out on a linear time-of-flight mass spectrometer (TOF 101, Comstock Inc., Oak Ridge, TN), modified to accommodate two laser ports, a viewport, and high accelerating voltage up to 30 kV.27 The high-voltage power supplies (205B-30R, Bertan Associates Inc., Hicksville, NY) provided ripple-free operation (<0.001%), even during the high ion current transients following the laser pulse. The original instrument was fitted with two laser sources: a N2 laser (VSL-337ND, Laser Science Inc., Newton, MA), working at 337 nm (5 ns pulse width), and a dye laser (LPD 3000, Lambda Physik, Goettingen, Germany), pumped at 308 nm by a XeCl-filled excimer laser (Lextra 50, Lambda Physik, Goettingen, Germany). Using Coumarin 540A dye, the tuning range of the dye laser was within 516-608 nm. With the application of a KDP second harmonic generator, the basic harmonic was converted into tunable 12 ns UV pulses in the 275-90 nm range. Laser pulse widths were measured by a 0.5 ns rise time fast photodiode. The pulse energy was regularly checked by a pyroelectric joule meter. The N<sub>2</sub> laser delivered up to 110  $\mu$ J/pulse with <10% intensity fluctuation, whereas the output of the doubled dye laser exceeded  $400 \,\mu$ J/pulse. To obtain good quality spectra, the laser irradiance was controlled by a variable attenuator (935-3-OPT, Newport Corp., Fountain Valley, CA) and had to be kept just above ion generation threshold ( $\sim 10^6$  W/cm<sup>2</sup>).

The pressure was consistently lower than  $10^{-8}$  mbar in the ion source region. A high-capacity diffusion pump (1200 L/s) with a water cooled baffle was situated below the ion source, whereas the analyzer region was pumped with a liquid nitrogen trapprotected diffusion pump. The laser-generated ions were accelerated from the high-voltage probe tip toward a grid kept at ground potential. Following the accelerating region, a large inner diameter (7.13 cm) electrostatic lens maximized the ion transmission. In addition, the interlocking design of lens elements eliminated field penetration and provided excellent field homogeneity. The ions drifted along a 210 cm flight tube and were detected by a dual multichannel plate assembly (Galileo Co., Sturbridge, MA) biased to -1800 V. A 120 MHz preamplifier (9305, EG&G ORTEC, Oak Ridge, TN) was followed by a variable gain amplifier module (6103, Lecroy, Albuquerque, NW) to extend the dynamic range of the detection. A fast transient digitizer (TR8828D, LeCroy, Albuquerque, NM) recorded the ion current with 8 bit resolution. The signal was also monitored by a digitizing oscilloscope for tuning purposes. Data acquisition and analysis were performed on a 486D/33 MHz personal computer running custom made software (TOFWARE, Ilys Software, Pittsburgh, PA). Single-shot mass spectra consisted of 16K data points with 10 ns/ channel resolution. Typically, 30-50 shots were averaged to improve the signal-to-noise ratio.

**Sample Preparation**. Except where noted below, matrices were used without purification. The >98% pure 3-hydroxypicolinic acid (3-HPA), 6-aza-2-thiothymine (ATT), and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid, SA) from Aldrich Chemical Co. (Milwaukee, WI) functioned without interference to the analyte signal. Despite the Sigma Chemical Co. (St. Louis, MO) guaranteed >99% purity for 2,5-dihydroxybenzoic acid (DHB) and for  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA), we had to recrystallize the DHB twice to remove the excessive amount of sodium salt present in the original product. Saturated matrix solutions were prepared fresh every day in 7:3 (v/v) acetonitrile–water

Table 1. Intensity Ratios of Arg-Lys/DNA Complex lons and Protonated DNA lons

|                | intensity ratios |        |        |  |
|----------------|------------------|--------|--------|--|
| DNA sequence   | (1:1)+           | (2:1)+ | (3:1)+ |  |
| $pd(T)_{10}$   | 11.91            | 9.69   | 2.14   |  |
| $pd(T)_{11}$   | 3.50             | 2.11   | 0.09   |  |
| $pd(T)_{20}$   | 1.10             | 0.43   | 0.19   |  |
| $pd(A)_9$      | 0.07             | 0.00   | 0.00   |  |
| $pd(C)_{10}$   | 0.00             | 0.00   | 0.00   |  |
| $pd(G)_{10}$   | 0.00             | 0.00   | 0.00   |  |
| pd(AGCTCAGCTT) | 1.20             | 0.54   | 0.08   |  |

(HPLC grade) solvent. The peptides (Gly-Gly, His-His, Arg-Lys), proteins (bovine insulin, bovine ubiquitin, horse heart cytochrome c, bovine ribonuclease A, chicken egg white lysozyme), and some of the single-stranded 5'-phosphorylated deoxynucleic acids (pd(A)<sub>9</sub>, pd(T)<sub>10</sub>, and pd(C)<sub>10</sub>) were purchased from Sigma Chemical Co. Other oligonucleotides, including pd(T)<sub>20</sub>, pd(T)<sub>40</sub>, pd(T)<sub>60</sub>, pd(G)<sub>10</sub>, and random sequence compounds (pd(AGCTCAGCTT) and d-(TTAGCAGCTT)), were provided by Cruachem, Inc. (Dulles, VA).

The dipeptide and protein stock solutions were prepared in 0.1% trifluoroacetic acid (TFA) with 0.05 and 4  $\times$  10<sup>-4</sup> M concentrations, respectively. Oligonucleotide stock solutions were made in deionized water in  $4 \times 10^{-4}$  M concentration. To vary the molar ratio between dipeptides, insulin, and oligonucleotides, the stock solutions of dipeptides were diluted to 0.005 and 0.0005 M, whereas insulin was diluted to  $3.2 \times 10^{-6}$  M in 0.1% TFA. Once prepared, the protein and oligonucleotide stock solutions were stored in a freezer at -5 °C. No degradation of the analytes was found within 1 month. Generally, 1.5  $\mu$ L protein and 1.5  $\mu$ L oligonucleotide aliquots were mixed in a vial before being combined with 10.0  $\mu$ L of matrix solution on the probe tip. The matrix-to-analyte molar ratio exceeded 1000:1, unless indicated otherwise. Samples were dried in a stream of cold air prior to insertion into the vacuum system. For the experiments involving insulin or ubiquitin, internal mass calibration was applied utilizing the proteins and matrix-related ions. In all other cases, bovine insulin and bovine ubiquitin were used as external standards. Mass accuracy was within  $\pm 0.1\%$  for external and  $\pm 0.05\%$  for internal calibration.

#### **RESULTS AND DISCUSSION**

To explore the nature of protein-oligonucleotide interactions, we investigated the effect of nucleotide base composition, the influence of amino acid makeup of the protein, the matrix dependence, and the effect of pH. The impact of these factors was measured by the number and abundances of different noncovalently bound complexes. We also monitored resolution and sensitivity changes, which had potential implications for the MALDI analysis of nucleic acids by complex formation.

In the discussion of protein-oligonucleotide complexes, the notation of ref 25 was adopted. That is,  $(m:n)^{k+}$  refers to the complex ion  $[mM_p + nM_o + kH]^{k+}$ , where  $M_p$  and  $M_o$  are the molecular masses of the protein and the oligonucleotide, respectively. Variables m and n are their multiplicity, and k refers to the number of charges on the complex ions. For example, (1: 2)<sup>2+</sup> represents the  $[M_p + 2M_o + 2H]^{2+}$  ion.

Effect of Nucleic Acid–Base Composition. Based on the results of Juhasz and Biemann,<sup>25</sup> we expected strong interaction between peptides made of basic residues and certain oligonucleotides. To study the contribution of nucleic acid composition, a



**Figure 1.** MALDI mass spectra of (a) decathymidylic acid,  $pd(T)_{10}$ , and (b) 125:1 mixture of dipeptide Arg-Lys and  $pd(T)_{10}$  using DHB matrix. Peak shapes of (a)  $pd(T)_{10}$  molecular ion and (b) (1:1)<sup>+</sup> complex are shown in the insets. Signal intensity expressed in arbitrary units is denoted by *I*.

dipeptide of strong basic character, Arg-Lys, was selected. Complex formation between Arg-Lys and a series of oligonucleotides was tested using MALDI-MS in DHB matrix. To evaluate the effect of the four different DNA bases, homooligomers were mixed with the dipeptide. As summarized in Table 1, multiple complex formation was observed between  $pd(T)_n$  (n = 10, 11, 20) homologues, whereas little or no complexation occurred with  $pd(A)_9, pd(C)_{10}$ , and  $pd(G)_{10}$ .

In the case of  $pd(T)_n$ , a large variety of complexes were produced containing up to six units of the peptide and four units of the nucleic acid. As a consequence, the MALDI spectra of  $pd(T)_n$  containing only cationized molecules and cluster ions were transformed into spectra showing a series of multiplets. An example is presented in Figure 1, where MALDI-MS of  $pd(T)_{10}$ (Figure 1a) and pd(I)<sub>10</sub>/Arg-Lys (Figure 1b) mixture are compared. The mass spectrum of pd(I)10 exhibits a strong molecular ion peak  $(m/z \ 3061)$  and the corresponding sodiated dimer  $(m/z \ 100 m/z \ 100 m/z)$ 6143), trimer (m/z 9203), and tetramer (m/z 12 265). It is well known that alkali ions attach easily to the phosphate groups in DNA, leading to peak broadening at high masses.<sup>15-17</sup> In the inset of Figure 1a, the  $[M + Na]^+$ ,  $[M + 2Na - H]^+$ , and [M + Na +K - H<sup>+</sup> ions can be distinguished from the molecular ion of  $pd(T)_{10}$ ,  $[M + H]^+$ , accompanied by poor apparent mass resolution  $(\sim 75)$  on all peaks. Throughout this work, the full width at half maximum (50%) was used in determining the mass resolution.

A very different peak shape is observed in the inset of Figure 1b for the  $(1:1)^+$  complex between  $pd(T)_{10}$  and Arg-Lys. The alkalinated peaks are completely absent, resulting in substantial improvement in mass resolution (~240). Also, the signal is now concentrated in a single peak, rather than a multiplet of cationized ions. We speculate that once the dipeptide–oligonucleotide

Table 2. Influence of Basic Residues on Complex Ion Formation between Peptides, Proteins, and  $pd(T)_{10}$  in DHB Matrix

| proteins <sup>a</sup>  | no. of basic residues | complex ion compositions <sup>b</sup>   |
|------------------------|-----------------------|---|
| Gly-Gly                |                       | $(1:1)^+$ (w)   |
| His-His                | 2His                  | $(1:1)^+$ (s), $(1:2)^+$ (m), $(1:3)^+$ (w), $(2:1)^+$ (m), $(2:2)^+$ (m), $(2:3)^+$ (w), $(3:2)^+$ (m), $(3:3)^+$ (w), $(3:4)^+$ (w), $(4:2)^+$ (m), $(4:3)^+$ (w), $(5:2)^+$ (w), $(5:3)^+$ (w) |
| Arg-Lys                | 1Arg, 1Lys            | $(1:1)^+$ (s), $(1:2)^+$ (m), $(1:3)^+$ (w), $(2:1)^+$ (s), $(2:2)^+$ (s),  |
| insulin <sup>c</sup>   | 1Arg, 1Lys, 2His      | $(1:1)^+$ (m)   |
| ubiquitin <sup>d</sup> | 4Arg, 11Lys, 2His     | $(1:1)^+$ (s), $(1:2)^+$ (s), $(1:3)^+$ (m), $(1:4)^+$ (w), $(1:5)^+$ (w)   |
| cvtochrome c           | 2Arg, 18Lys, 3His     | $(1:1)^+$ (m), $(1:2)^+$ (w)  |
| ribonuclease A         | 4Arg, 10Lys, 4His     | $(1:1)^+$ (s), $(1:2)^+$ (m), $(1:3)^+$ (w)   |
| lysozyme               | 11Arg, 6Lys, 1His     | $(1:1)^+$ (s), $(1:2)^+$ (s), $(1:3)^+$ (m), $(1:4)^+$ (w), $(1:5)^+$ (w)   |
|                        |                       |   |

<sup>*a*</sup> The molar ratio of dipeptides (Gly-Gly, His-His, or Arg-Lys) to  $pd(T)_{10}$  was 125:1. <sup>*b*</sup> Abundances are shown after peak assignment: w, weak; m, medium; and s, strong. <sup>c</sup> The molar ratio of insulin to  $pd(T)_{10}$  was 1:125. <sup>*d*</sup> The molar ratio of other proteins (ubiquitin, cytochrome *c*, ribonuclease A, or lysozyme) to  $pd(T)_{10}$  was 1:1.

complexes form, the basic dipeptides occupy the highly acidic sites on the phosphate groups, displacing the alkali ions. Consequently, the ionization of oligonucleotide molecules changes from alkalization to attachment of a molecule which is protonated. Generally, the mass resolution of  $pd(T)_n$  obtained by complex formation in positive ion mode was comparable to the negative ion mode values. This makes viable the collection of good-quality positive ion spectra for DNA analysis, with the inherent benefits of positive ion mass spectrometry.

The addition of peptides to improve DNA MALDI spectra appears to constitute an improvement over existing methods for preparing DNA samples. For example, one of the common methods to reduce cation adduct abundances in MALDI analysis is the use of cation-exchange resins.<sup>28</sup> However, the introduction of resins into the sample disrupts the homogeneity of analytematrix mixture. Thus, "good spots" must be found on the probe in order to obtain good quality spectra. In contrast, by mixing the basic dipeptide, oligonucleotide, and matrix solutions, homogeneity can be easily achieved. Therefore, the complexation method also improves the shot-to-shot reproducibility. One drawback to this method is that peaks of the complex ions become weaker and broader with increasing mass, leading to somewhat lower mass accuracies for clusters of complexes. Due to the notable mass accuracy of MALDI-MS, however, the relative error values were consistently lower than 0.1% for all the investigated ions, even when external calibration was used. Moreover, the peptide:DNA ratio can be adjusted to minimize higher adduct formation.

The MALDI analysis of the Arg-Lys/pd(T)<sub>20</sub> mixture in DHB matrix showed behavior similar to the pd(T)<sub>10</sub> case. Strong (1: 1)<sup>+</sup>, (1:2)<sup>+</sup>, and (1:3)<sup>+</sup> complex ions were observed, with mass resolution of ~275. Again, the resolution was superior to the value for pd(T)<sub>20</sub> peaks measured alone (~75). MALDI mass spectra of larger oligonucleotides, pd(I)<sub>40</sub> and pd(T)<sub>60</sub>, in the presence of Arg-Lys also showed complex formation. The peaks from the Arg-Lys adducts of pd(T)<sub>60</sub> at  $m/z \sim 18000$ , however, were not separated completely due to the mass resolution limitations. Further work is in progress to improve the mass resolution for larger oligonucleotides utilizing the complex formation phenomenon.

Other basic dipeptides (e.g., His-His) gave results comparable to those of Arg-Lys with the different homooligomers. Strong complex formation was observed with  $pd(T)_n$ , whereas little or no complexation occurred with homologues containing the other three bases. The attachment of up to three Arg-Lys groups was observed for the mixed-sequence nucleic acid pd(AGCTCAGCTT). Comparison of the relative complex intensities to the corresponding pd(T)<sub>10</sub> values suggested that complex formation was driven by the thymidylic acid constituents, as will be discussed further below.

Amino Acid Composition Dependence. Noncovalent binding between  $pd(T)_n$  (n = 10, 11, 20, 40, 60) and a selection of peptides (Gly-Gly, His-His, Arg-Lys) and proteins (bovine insulin, bovine ubiquitin, horse heart cytochrome *c*, bovine ribonuclease A, chicken egg white lysozyme) was observed using MALDI-MS. The investigated proteins, their basic residue content, and the detected complexes with  $pd(T)_{10}$  are reported in Table 2.

To uncover the influence of the different amino acid residues, several dipeptides were studied. Strong correlation was found between the presence of basic residues and the tendency to participate in complexes. The mixture of a nonpolar dipeptide, Gly-Gly, and  $pd(T)_{10}$  showed only a very weak response in the mass spectrum corresponding to the  $(1:1)^+$  complex ion. On the other hand, His-His and especially Arg-Lys demonstrated a strong propensity to participate in a large variety of complexes.

The ability of basic side chains to influence complex formation was also observed with larger polypeptides. For example, the noncovalent binding of  $pd(T)_{10}$  to the five medium-size proteins listed in Table 2 revealed large differences in affinity. It was found that the efficacy of complex ion formation depended on, among other parameters, the presence and number of basic amino acid residues.

Three proteins of similar size are listed in Table 2: cytochrome c (MW 12 360.1), lysozyme (MW 14 296.1), and ribonuclease A (MW 13 683). Although cytochrome c had the largest total number of basic amino acid residues (2Arg + 18Lys + 3His =23), we observed only a maximum of two  $pd(T)_{10}$  molecules attached to it in an equimolar mixture. In contrast, up to five pd(T)10 units were added onto a lysozyme molecule under the same conditions, in spite of the lower total number of basic residues (11Arg + 1His + 6Lys = 18). If we assume that complex ions stem from the interactions between basic amino acid residues in the protein and acidic phosphate groups in  $pd(T)_{10}$ , these observations may point to the distinguished role of Arg in the interaction. Indeed, lysozyme has a significantly larger number of Arg residues and more abundant complex peaks than cytochrome c. Of course, other factors not discussed here, such as hydrophobic interactions and protein conformation, may influence the complex formation as well.

<sup>(28)</sup> Nordhoff, E.; Ingendoh, A.; Cramer, R.; Overberg, A.; Stahl, B.; Karas, M.; Hillenkamp, F.; Crain, P. F. Rapid Commun. Mass Spectrom. 1992, 6, 771– 776.



**Figure 2.** Effect of the number of basic residues in the protein on the observation of complexation in MALDI-MS: (a) bovine insulin and  $pd(T)_{10}$  mixture (1:125); (b) bovine ubiquitin and  $pd(T)_{10}$  mixture (1:125); (b) bovine ubiquitin and  $pd(T)_{10}$  mixture (1:125); (c) bovi

In addition to the effect of amino acid side chains, we found that the protein-oligonucleotide molar ratio and the size of the molecules also affected the observation of noncovalent complex formation. The effect of molar ratios on the observation of complex formation was illustrated using insulin and ubiquitin as examples. There are 4 and 17 basic amino acid residues in bovine insulin (1Arg + 1Lys + 2His = 4) and bovine ubiquitin (4Arg + 1Lys + 2His = 4)11Lys + 2His = 17), respectively. Thus, it was expected that the interaction of insulin with the oligonucleotide was weaker. Indeed, no complex ions were observed for an equimolar mixture of insulin and  $pd(T)_{10}$ . To avoid the overwhelming abundance of insulin [MH]<sup>+</sup> ions masking the complex formation, the molar ratio of insulin to pd(T)10 was decreased to 1:125. MALDI-MS response of nucleic acid molecules attached to insulin is shown in Figure 2a. The intensity of the (1:1)+ peak, however, was less than onefourth of the insulin ion abundance. In contrast, an equimolar mixture of ubiquitin and  $pd(T)_{10}$  exhibited extensive complex formation (Figure 2b). Composite ions up to (1:5)<sup>+</sup> appeared, showing intensities comparable to those of the ubiquitin ion.

An additional factor that influenced observation of complexes between proteins and oligonucleotides was the size of the components involved. The larger the size of the nucleotide and/ or protein, the lower the abundance of complex ions. For example, we observed the complexes of Arg-Lys/pd(T)<sub>60</sub> and ubiquitin/pd(T)<sub>20</sub> pairs separately, but no complexes of pd(T)<sub>60</sub> and ubiquitin were detected. The absence of ubiquitin/pd(T)<sub>60</sub> complex peaks from the spectra either was the consequence of the lack of these species or was related to the lower sensitivity of ion detection in the high mass range. More work is needed to address this question.

Role of the Matrix. The detection and the intensity of protein-DNA complex ions were highly dependent on the choice



**Figure 3.** Mass spectra of the equimolar mixture of lysozyme and  $pd(T)_{10}$  using (a) DHB matrix and (b) CHCA matrix.

of matrix. To demonstrate this effect, lysozyme/pd(T)<sub>10</sub> complex detection was compared in four matrices (DHB, CHCA, 3-HPA, and SA). MALDI mass spectra of the equimolar mixture of lysozyme and pd(T)<sub>10</sub> in DHB and CHCA matrices are shown in Figure 3. In DHB matrix (Figure 3a), a broad variety of lysozyme/ pd(T)<sub>10</sub> complex ions (i.e.,  $(1:0)^+$ ,  $(1:1)^+$ ,  $(1:2)^+$ ,  $(1:3)^+$ ,  $(1:4)^+$ , and  $(1:5)^+$ ) were observed, together with the doubly charged species. The  $(1:0)^+$  molecular ion abundance decreased to about one-fifth of its original value with the addition of oligonucleotide, and the  $(1:1)^+$  ion was more abundant than the noncomplexed  $(1:0)^+$  ion. In CHCA matrix, however, the mass spectrum was dominated by strong singly, doubly, and triply charged lysozyme ions, and no complex formation was found (Figure 3b). In 3-HPA matrix, (1: $0)^+$ ,  $(1:1)^+$ ,  $(1:2)^+$  ions were observed, whereas no complexes were found in SA matrix.

Based on these results, protein-oligonucleotide complex formation can be used to differentiate between the tested matrices according to the presence (i.e., DHB and 3-HPA) or absence (i.e., CHCA and SA) of complex ions. This observation resembles the results on the fragmentation patterns of peptides from different matrices,<sup>29,30</sup> indicating that the matrix is a key factor that influences the stability of ions formed in the MALDI process. For example, extensive metastable fragmentation of sperm whale apomyoglobin ions was observed by Karas and co-workers when CHCA was used as a matrix. In contrast, no metastable decomposition was detected using "super" DHB (9:1 mixture of DHB and 2-hydroxy-5-methoxybenzoic acid) or 3-HPA under the same conditions.<sup>29</sup> Chait and co-workers also demonstrated that, in an ion trap mass spectrometer, the ratio of fragment ions to protonated angiotensin III in CHCA or SA was at least 10 times higher than the same ratio in DHB. They attributed this

4546 Analytical Chemistry, Vol. 67, No. 24, December 15, 1995

<sup>(29)</sup> Karas, M.; Bahr, U.; Strupat, K.; Hillenkamp, F.; Tsarbopoulos, A.; Pramanik, B. N. Anal. Chem. 1995, 67, 675-679.

<sup>(30)</sup> Qin, J.; Steenvoorden, R. J. J. M.; Chait, B. T. Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, May 29– June 4, 1994; p 5.

# Table 3. Acid-Base Equilibria and $pK_a$ Values for the Four Nucleotides<sup>31</sup> and Species Ratios<sup>a</sup> in Saturated DHB Solution (pH = 1.7)

|   |                    | $\mathrm{p}K_{\mathrm{a}}$ |                      |   |                 |                     |
|---|--------------------|----------------------------|----------------------|---|-----------------|---------------------|
| compound  | base               | phosphate group            | $[H_2N^+]/[HN]$      | [N <sup>-</sup> ]/[HN]                                      | $[H_2T]/[HT^-]$ | $[T^{2-}]/[HT^{-}]$ |
| adenosine 5'-phosphate<br>guanosine 5'-phosphate<br>cytidine 5'-phosphate | 3.74<br>2.9<br>4.5 | 0.9<br>0.7<br>0.8          | 0.15<br>0.10<br>0.12 | $9.1 	imes 10^{-3} \ 6.5 	imes 10^{-2} \ 1.6 	imes 10^{-3}$ |                 |                     |
| thymidine 5'-phosphate  | 10.0               | 1.6                        |                      |   | 0.83            | $5 \times 10^{-9}$  |

<sup>*a*</sup> HN represents the zwitterionic form of adenosine, guanosine, and cytidine, whereas  $HT^-$  stands for the anionic form of thymidine. Species  $H_2N^+$  and  $H_2T$  are the protonated, whereas species  $N^-$  and  $T^{2-}$  are the deprotonated forms of HN and  $HT^-$ , respectively.

difference to the significantly higher internal peptide ion energies in the case of CHCA and SA matrices.<sup>30</sup>

pH Dependence. Since both the nucleic acid and the protein components of the complex were polyprotic systems, we expected pH to play an important role in the complexation process. The degree of protonation of both the peptides and the oligonucleotides controls their charge states, which in turn affects ionic interactions in the solution phase. Both the solvent used to dissolve the proteins (0.1% TFA, pH = 1.9) and the saturated matrix solutions (DHB, pH = 1.7; 3-HPA, pH = 2.7; CHCA, pH = 2.4; SA, pH = 3.1) were acidic. In addition, because of the relatively high concentration of the matrix, the pH was determined by the matrix solution (DHB saturated in 7:3 (v/v) acetonitrile-water,  $\sim$ 1.8 M). Due to the acidic environment, the amino terminus of the peptide (NH<sub>3</sub><sup>+</sup>,  $pK_a = 8.0$ ) and the basic side chains (ArgH<sup>+</sup>,  $pK_a = 8.0$ ; LysH<sup>+</sup>,  $pK_a = 10.0$ ; HisH<sup>+</sup>,  $pK_a = 6.5$ ) are expected to be protonated. At the same pH, it is expected that the adenosine, guanosine, and cytidine base units are protonated, whereas the thymidine is in the neutral form. The  $pK_a$  values for nucleic acids are summarized in Table 3. According to Table 3, the corresponding phosphate groups are deprotonated for A, G, and C and partially deprotonated for T. Consequently, A. G. and C are effectively neutral at the pH of the DHB solution, whereas T has an overall negative charge.

This hypothesis is supported by studies in which the solution pH is shifted to lower values. To shift the acid-base equilibrium on the phosphate group of 5'-deoxythymidylic acid, we introduced a strong acid into the system. Saturated DHB solution was prepared in 7:3 acetonitrile-water solvent with 1 M HCl in it. In this acidic solution (pH < 1.0), there were essentially no negatively charged phosphate groups in  $pd(T)_n$  (pK<sub>a</sub> = 1.60). To uncover the effect of the negative charge on the phosphate group, complex formation between Arg-Lys and  $pd(T)_{11}$  was observed in the "acidified" and in the regular environment (see Figure 4). Although both the peptide and the nucleic acid components were present in both spectra, the complex ions were abundant only in the absence of the strong acid (Figure 4a). This observation was a definite indication of the role of acid-base processes in the complex formation and pointed to the key role of T in complex formation.

To determine the effect of increasing the pH, the complexation of Arg-Lys with a mixed DNA sequence, d(TTAGCAGCTT), was studied in ATT matrix. In the acidic neat ATT solution (pH = 2.0 for saturated ATT in 3:2 acetonitrile-water solvent), a broad variety of Arg-Lys/DNA complex ions (i.e.,  $(0:1)^+$ ,  $(1:1)^+$ ,  $(2:1)^+$ ,  $(3:1)^+$ ,  $(4:1)^+$ ) were observed (Figure 5a). The most abundant peak corresponded to  $(1:1)^+$ , and the maximum number of Arg-Lys molecules attached to one DNA molecule was equal to the



**Figure 4.** MALDI-MS of the (a) neat, pH = 1.7, and (b) "acidified", pH < 1, mixture of Arg-Lys and  $pd(T)_{11}$  (125:1) in DHB matrix.

number of thymine bases in the sequence. When ATT was prepared in 1:1  $NH_4HCO_3$  buffer-acetonitrile solution (pH = 6.5) and Arg-Lys was dissolved in NH4HCO3 buffer, complex ions up to  $(6:1)^+$  were detected, and the most abundant peak belonged to  $(2:1)^+$  (see Figure 5b). The presence of Arg-Lys in the complex in excess of the number of T bases indicated the possibility of complexation between Arg-Lys and A, G, and/or C at close to neutral pH values. To confirm this hypothesis, experiments with pd(A)<sub>10</sub> in place of d(TTAGCAGCTT) were carried out under identical conditions. In the presence of the NH4HCO3 buffer complex, ion peaks appeared corresponding to  $(1:1)^+$ ,  $(2:1)^+$ , and  $(3:1)^+$  compositions. The intensity of the  $(1:1)^+$  complex ion signal exceeded that of the (0:1)<sup>+</sup> noncomplexed oligonucleotide. Although the MALDI observation of noncovalent complex formation at physiological pH values may have relevance in the exploration of specific protein-oligonucleotide interactions, the significance of the presented data will have to be proved by investigating other systems with well-established solution chemistry.

**Ion Pair Formation.** Based on the observations made here, a simple model of complex formation has been proposed. To unveil the mechanism of complexation, we recall the basic properties displayed by some pyrimidines, such as cytosine, and

Analytical Chemistry, Vol. 67, No. 24, December 15, 1995 4547



**Figure 5.** MALDI-MS of the (a) neat, pH = 2.0, and (b) buffered, pH = 6.5, mixture of Arg-Lys and d(TTAGCAGCTT) (125:1) in ATT matrix.

purines, such as adenine and guanine.<sup>31</sup> It is well known that the protonation sites are N<sup>3</sup> in cytosine, N<sup>1</sup> in adenine, and N<sup>7</sup> in guanine, since positive charge attachment at these nitrogen atoms is facilitated by delocalization. On the other hand, thymine-related structures exist in a tautomeric hydroxy form and lack an amino group. Thus, the protonation of thymine derivatives is hindered because the positive charge cannot be delocalized.

Table 3 lists the  $pK_a$  values for the base and the phosphate groups in adenosine 5'-phosphate, guanosine 5'-phosphate, cytidine 5'-phosphate, and thymidine 5'-phosphate.<sup>31</sup> All four compounds can be viewed as diprotic acids put in an environment where the matrix stabilizes the pH at around 1.7. Based on the data presented in Table 3, one can see that 80–85% of A, G, and C 5'-phosphate are in the zwitterionic form. Of A, G, and C 5'phosphates, only 1, 6, and 0.2% carry a net negative charge, respectively, whereas the remaining part is positive. Thymidine 5'-phosphate, however, shows a very different distribution: 45% of that compound is neutral, 55% has a single negative charge, and the proportion of doubly charged anions is negligible in a pH = 1.7 environment. Thus, it is the presence of the negative charge on T that most likely drives complex formation.

In Chart 1, we depict the fate of the main species for A 5'phosphate and T 5'-phosphate in the presence of protonated peptides. The zwitterionic form, dominant in A, G, and C systems, undergoes a conformational change, leading to an intramolecular ion pair. This species shows only moderate to weak interactions with cationic peptides. The majority (55%) of T 5'-phosphate, however, exhibits a net negative charge, leading to strong ionic interactions with protonated sites on peptide molecules. In spite Chart 1 Intramolecular Ion Pair Formation in the Adenosine 5'-Phosphate Zwitterion (a) and Intermolecular Ion Pair Formation between Arg-Lys and Thymidine 5'-Phosphate (b) in pH = 1.7 Solution<sup>a</sup>



<sup>a</sup> Only the ion pair between the protonated amino group at the N-terminus of Arg-Lys and the negatively charged phosphate group is shown.

of the higher rigidity and spatial hindrance, extended chains of oligonucleotides can undergo similar ion pair formation, even with bulky proteins. The presence of multiple charge centers on peptides may explain how larger complexes (e.g.,  $(1:2)^+$ ,  $(1:3)^+$ ) and doubly, possibly multiply charged ions are formed. As is apparent from Figure 3a, a ubiquitin molecule can combine with up to four pd(T)<sub>10</sub> units in singly and doubly charged forms.

To a smaller degree (~6%), G 5'-phosphate follows this behavior due to its less basic character. On the average, G is about one-tenth as effective in forming an intermolecular ion pair as T. In fact, ubiquitin mixed with  $pd(G)_{10}$  did lead to a very weak complex signal in the MALDI spectrum. Similarly weak complex ion abundances were present, however, in the case of  $pd(A)_9$  and  $pd(C)_{10}$ . The low abundances of these species prevented us from pursuing a quantitative comparison.

#### ACKNOWLEDGMENT

The authors are grateful for the financial support of the National Science Foundation (Grant No. CTS-9212389), used to purchase the excimer laser and part of the dye laser system. In kind contribution from The George Washington University made it possible to acquire the second harmonic generator and laser coupling optics. X.T. was supported in part by the George Washington University Facilitating Fund.

Received for review August 29, 1995. Accepted October 4, 1995.<sup>®</sup>

#### AC9508762

4548 Analytical Chemistry, Vol. 67, No. 24, December 15, 1995

<sup>(31)</sup> Shabarova, Z.; Bogdanov, A. Advanced Organic Chemistry of Nucleic Acids; VCH: Weinheim, Germany, 1994; Chapter 4.

<sup>\*</sup> Abstract published in Advance ACS Abstracts, November 1, 1995.