Nucleobase Amino Acids Incorporated into the HIV-1 Nucleocapsid Protein Increased the Binding Affinity and Specificity for a Hairpin RNA

Tsuyoshi Takahashi,[a] Akihiko Ueno,[a] and Hisakazu Mihara*[a, b]

L-α-amino acids with a nucleobase in the side chain (nucleobase amino acids; NBAs) were used to enhance the function of RNA-binding proteins that recognize structured RNA. These NBAs were utilized in the three-dimensional structure of the protein to enhance RNA binding affinity and specificity as a result of selective recognition of NBAs by RNA bases. NBA units were incorporated at various positions into the HIV-1 nucleocapsid protein NCp7 (residues 1–55), which contains two CCHC-type (Cys-X_2-Cys-X_4-His-X_4-Cys-type; X = an amino acid residue) zinc knuckle domains. The binding ability was evaluated by using the stem-loop (SL)3 region of HIV-1 Ψ-RNA. Visible light absorption measurements revealed that two zinc ions bound strongly and quantitatively to the NBA-NCp7 molecule and to the wild-type NCp7 protein. This result indicates that the incorporation of NBA units composed of L-α-amino acids did not influence the formation of the specific structure of NCp7. Binding analysis with fluorescein-labeled SL3 RNA revealed that incorporation of NBA units into the NCp7 protein at appropriate positions increased its RNA binding affinity and specificity. An NBA-NCp7 protein that possessed cytosine and guanine NBA units at positions 13 and 46, respectively, showed a binding affinity for SL3 RNA ninefold higher than that of wild-type NCp7 as a result of the specific and cooperative interaction of the NBA units with RNA bases. These results clearly demonstrate that inclusion of NBA units in the three-dimensional structure of an RNA-binding protein is a useful strategy for enhancing the function of the protein.

KEYWORDS:

amino acids · chemical ligation · HIV-1 nucleocapsid protein · nucleobases · RNA recognition

Introduction

RNA–protein interaction plays important roles in nature. Many cellular functions, such as transcription, RNA splicing, and translation, depend on the specific interaction of proteins and RNA.[1] RNA-binding proteins form a specific conformation composed of secondary structures such as α helices and β strands and in many cases recognize structured RNA.[1] Single-stranded RNA folds into a compact conformation based on stem and loop structures then forms a specific three-dimensional structure important to its function. Hairpin loops commonly define the binding sites for RNA-binding proteins, which include phage coat proteins, several ribosomal proteins, and transcription terminators/antiterminators.[2–4] The construction of molecules that recognize a specific RNA structure such as a hairpin loop with high affinity and high specificity is useful not only to provide information about the principle of RNA–protein interactions in nature, but also for the design of probes and drugs that target RNA. Therefore, we attempted to enhance the function of an RNA-binding protein by site-specific incorporation of artificial L-α-amino acids with a nucleobase in the side chain (nucleobase amino acids (NBAs); Figure 1a), which can specifically recognize RNA bases. In order to demonstrate the applicability of the NBA units to peptides and proteins, we designed and synthesized short NBA-peptides (17 residues) derived from the HIV-1 Rev protein and evaluated the binding properties of the peptides with HIV-1 RRE IIB RNA.[5, 6] The Rev peptide bound specifically to RRE IIB RNA by forming a simple α-helical structure. Incorporation of an NBA unit into the Rev peptide by replacement of an amino acid with an L-α-amino acid that has a nucleobase in its side chain did not disturb the structure of the α helix. The peptides with an appropriate NBA unit (for example, a cytosine NBA, which is capable of interaction with a guanine base in RRE IIB RNA) bound to the RNA with high affinity and high specificity.[5, 6] Previous findings suggested that the NBA unit in this simple α-helical peptide could be utilized to increase its binding affinity and specificity for the target RNA. To expand the NBA technology to protein engineering, it is first necessary to evaluate the properties of NBAs incorporated into an RNA-binding protein that forms a specific and complex three-dimensional structure.

[a] *Prof. Dr. H. Mihara, Dr. T. Takahashi, Prof. Dr. A. Ueno*
Department of Bioengineering
Graduate School of Bioscience and Biotechnology
Tokyo Institute of Technology
Yokohama 226-8501 (Japan)
Fax: (+81) 45-924-5833
E-mail: hmihara@bio.titech.ac.jp

[b] *Prof. Dr. H. Mihara*
Form and Function, PRESTO
Japan Science and Technology Corporation
Yokohama 226-8501 (Japan)

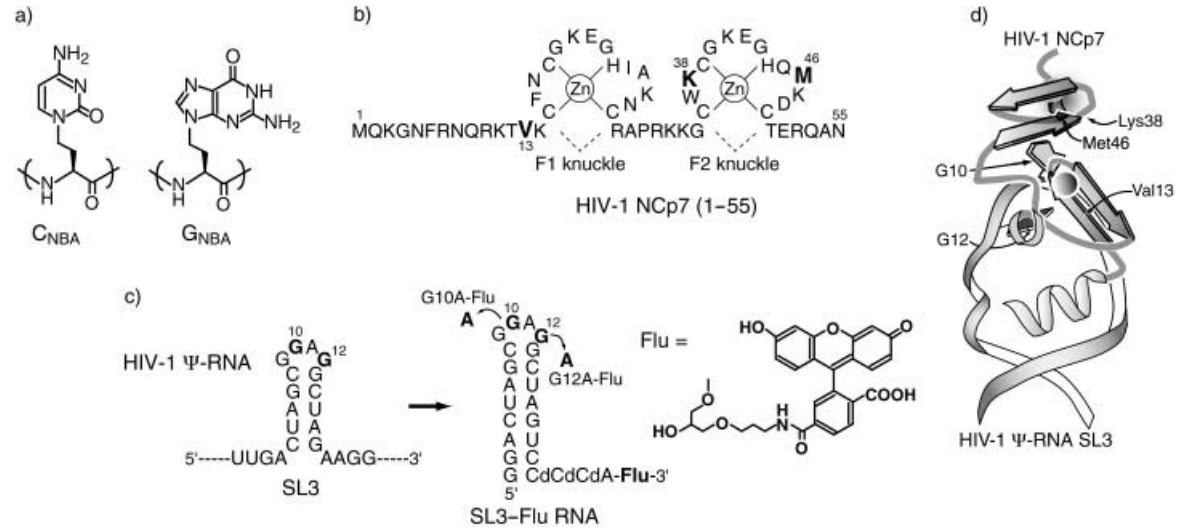


Figure 1. *a) Structure of the C_{NBA} and G_{NBA} units. b) Amino acid sequence of HIV-1 nucleocapsid protein NCp7 (1 – 55). The Val 13, Lys 38, and Met 46 residues of NCp7 were selected as the positions for incorporation of the NBA units. c) Secondary structure of the HIV-1 Ψ-RNA SL3 region and a fluorescein-labeled SL3 (SL3 – Flu) RNA. G = guanine, C = cytosine, U = uridine, A = adenine. d) Schematic representation of the structure of the NCp7 – SL3 complex.*

To achieve this aim, we selected the HIV-1 nucleocapsid 55-residue protein NCp7 (1 – 55), which contains two CCHC-type (Cys-X_2-Cys-X_4-His-X_4-Cys-type; X = an amino acid residue) zinc knuckle domains (Figure 1 b).[7] These domains are critical for viral replication and participate directly in genome recognition and encapsidation.[8] NCp7 recognizes a region of the unspliced viral RNA approximately 120 nucleotides long that is known as the Ψ site.[9–11] The HIV-1 Ψ site contains four stem-loop structures, denoted SL1 to SL4.[12–15] NMR spectroscopy structural analyses have revealed that NCp7 binds to the RNA construct that includes the SL3 region in the Ψ site (SL3 RNA; Figure 1 c, d).[7] Residues Lys 3 to Arg 10 of NCp7 form a 3_{10} helix (a helix with three residues per turn and ten atoms per hydrogen-bonded ring) that binds to the major groove of the RNA stem and also packs against the N-terminal CCHC-type zinc knuckle (F1 knuckle, residues 15 – 28). Binding of NCp7 to SL3 RNA is mediated by specific interactions between the N- and C- (F2 knuckle, residues 36 – 49) terminal CCHC-type zinc knuckles of NCp7 and the G 12 and G 10 bases in the G9-G10-A11-G12 RNA tetra loop, respectively. Based on this three-dimensional information, we attempted to apply the NBA technology to NCp7 by site-specific incorporation of the NBA units into the protein. It was expected that this strategy would lead to the specific interaction of the NBA units in the three-dimensional structure of NCp7 with SL3 RNA and enhance the binding affinity and specificity of interactions between these molecules. A total of 25 protein analogues with NBAs introduced by mutation at various positions were examined for their RNA binding. Here we present several representative examples of NBA-NCp7 proteins and their RNA binding. The findings in this study demonstrate the potential capability of NBAs to act as building blocks for the engineering of RNA-binding peptides and proteins.

Results and Discussion

Design of NBA-NCp7 proteins

A series of NBA-NCp7 proteins was designed based on the structure of the HIV-1 NCp7 protein that binds to the SL3 region in HIV-1 Ψ-RNA.[7] G 12 and G 10 bases in the tetra loop region of the RNA were selected as the target sites for the NBA units. Cytosine- and guanine-containing NBA units C_{NBA} and G_{NBA} (Figure 1 a) were selected for use in the protein because cytosine and guanine moieties have a high potential for interaction with a guanine base. One or two NBA units were incorporated at positions 13, 38, and 46 of NCp7 (Figure 2). In the NMR

Figure 2. *Amino acid sequences of the wild-type NCp7 and NBA-NCp7 proteins with one or two NBA units at positions 13, 38, or 46.*

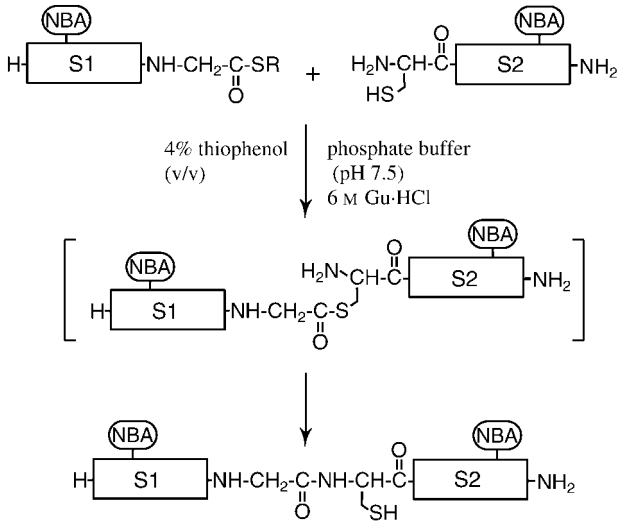
spectroscopy structure of the NCp7 – SL3 complex,[7] the Val 13 residue of NCp7 is close to the G 12 base of the RNA (Figure 1 d). When the cytosine NBA unit is incorporated at residue 13, the cytosine moiety can be positioned at a location in which it is capable of making a Watson – Crick base pair with the G 12 base in the RNA. The Lys 38 and Met 46 residues are located near the G 10 base in SL3 RNA (Figure 1 d). It was expected that a guanine moiety at residue 46 would be superior to a cytosine for specific interaction with the G 10 base since the Met 46 residue in NCp7 is located on the O6- and N7-atom side of the G 10 base in the NMR structure of the NCp7 – SL3 complex. Therefore, if a cytosine base at position 46 could participate in Watson – Crick-type interaction with the G 10 base in SL3 RNA, the structure of the NCp7 – SL3 complex might be disturbed. The Lys 38 residue is positioned such that its side chain is slightly distant from the G 10 base in the RNA compared with the location of the Met 46 side chain. It seems that the G_{NBA} residue, which is composed of a purine ring, interacts with the G 10 base in the RNA more easily than does the C_{NBA} residue, whose composition includes a pyrimidine unit. It was expected that one or two NBA units incorporated into NCp7 at positions 13, 38, and 46 could increase the RNA-binding affinity and specificity of the protein by specific interaction of the nucleobase moieties on the protein with the G 10 and G 12 bases in SL3 RNA.

Protein synthesis

4-(N^4-benzyloxycarbonylcytosin-1-yl)-2*S*-(9-fluorenylmethoxycarbonyl)aminobutyric acid [Fmoc-C_{NBA}(Z)-OH] and 4-(guanin-9-yl)-2*S*-(9-fluorenylmethoxycarbonyl)aminobutyric acid [Fmoc-G_{NBA}-OH] were synthesized[6] for incorporation of the C_{NBA} and G_{NBA} units into NCp7. Wild-type NCp7 and the NBA-NCp7 proteins were synthesized by native chemical ligation[16] with the peptide segment S1, which consists of residues 1 – 35 (used as S1-Gly-COSR; R = CH_2CH_2COOEt), and segment S2, which contained residues 36 – 55 (used as Cys-S2). The segment peptides were synthesized by the solid phase method by using the 9-fluorenylmethoxycarbonyl (Fmoc) strategy.[17] Chemical ligation of S1-Gly-COSR and Cys-S2 (Scheme 1) was carried out in a phosphate buffer (0.1 M; pH 7.5) that contained guanidine hydrochloride (Gu · HCl; 6 M) and thiophenol (4 % v/v).[18] The ligation products were purified by semipreparative reversed-phase (RP) HPLC to give high purity and good yield. Each protein was identified by MALDI-TOFMS and the concentration of each protein was determined by amino acid analysis.

Zinc binding of NCp7 and NBA-NCp7

In order to examine whether the synthetic NCp7 and NBA-NCp7 proteins bind to two zinc ions, the visible absorption spectra were measured in tris(hydroxymethyl)aminomethane (Tris) · HCl buffer (50 mM; pH 7.5). The spectrum of a mixture of NCp7 and $CoCl_2$ in solution was characteristic of an almost symmetrical tetrahedral structure, as shown by the position and the intensity of the bands, which occur at 697, 644, and 616 nm (Figure 3 a).[19] This absorbance by a Co_2 – NCp7 complex disappeared on the addition of 1.0 equiv $ZnCl_2$ for each zinc knuckle domain. These results indicate that the two cobalt ions coordinated to NCp7 in the initial mixture were replaced by two zinc ions and these zinc ions bound to the synthetic NCp7 strongly and quantitatively. In the case of the NBA-NCp7 proteins, the visible absorption spectrum of a mixture of a V13C_{NBA}M46G_{NBA} protein (that is, NCp7 with V 13 replaced by C_{NBA} and M 46 replaced by G_{NBA}) and $CoCl_2$ in solution was observed to be similar to that of NCp7 and $CoCl_2$ (Figure 3 b). In this NBA-NCp7 solution, the absorbance also disappeared upon addition of 1.0 equiv $ZnCl_2$ for each zinc knuckle domain. These results suggest that the zinc ions bound to the V13C_{NBA}M46G_{NBA} protein in a manner similar to that which occurs in the case of NCp7. That is, incorporation of the NBA units into NCp7 could not change the ability of the protein to coordinate with zinc ions and it seems that the NBA-NCp7 protein has an ability equal to that of the wild-type NCp7 to form the specific binding structure.



Scheme 1. *Chemical ligation of S1 – Gly – COSR and Cys – S2. R = CH_2CH_2COOEt.*

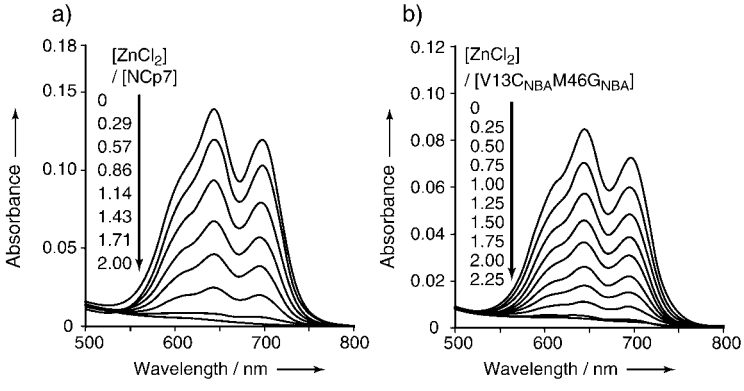


Figure 3. *Visible absorption spectra of (a) NCp7 (140 μM) with $CoCl_2$ (280 μM) and (b) the V13C_{NBA}M46G_{NBA}-NCp7 protein (80 μM) with $CoCl_2$ (160 μM), and the addition of various concentrations of $ZnCl_2$ in 50 mM Tris · HCl buffer (pH 7.5).*

Binding affinity of single-NBA-NCp7 for SL3 – Flu RNA

A fluorescein-labeled RNA (SL3 – Flu; Figure 1c) was prepared by ligation of SL3 RNA and 5′-dCdCdA-fluorescein-3′ DNA (dC = deoxycytosine, dA = deoxyadenine) with T4 RNA ligase[20] for evaluation of the abilities of the proteins to bind with SL3 RNA. The binding properties of the proteins were evaluated by fluorescence anisotropy measurements based on the fluorescein moiety. The fluorescence anisotropy of SL3 – Flu (5.0 nM) was increased by the addition of NCp7 in a buffer solution (pH 7.5; Figure 4). The significant increase of anisotropy is attributed to

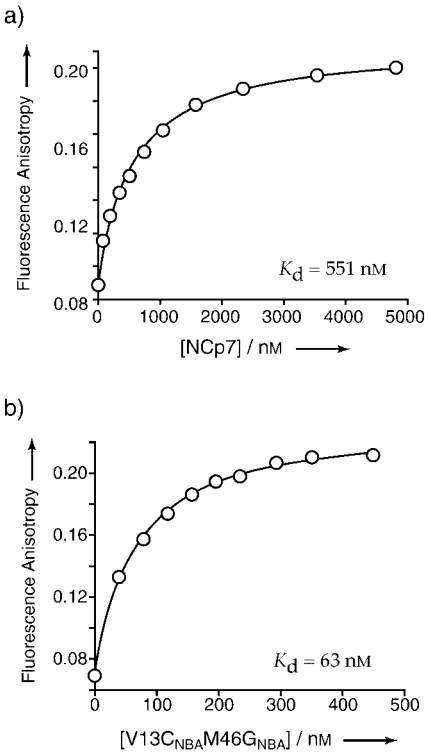


Figure 4. *Fluorescence anisotropy of SL3 – Flu RNA as a function of (a) NCp7 and (b) V13C$_{NBA}$M46G$_{NBA}$ protein concentrations in HEPES buffer (10 mM; pH 7.5) with NaCl (150 mM), MgCl$_2$ (1 mM), and ZnCl$_2$ (50 μM) at 10 °C. [SL3 – Flu] = 5.0 nM.*

the differences in flexibility of the fluorescein moiety and in molecular size between the free SL3 – Flu RNA and the RNA – NCp7 complex. A dissociation constant (K_d) of 551 nM was calculated from the anisotropy increase observed for SL3 – Flu by using a binding equation with 1:1 stoichiometry (for details, see the Experimental Section). The results for the binding of several protein analogues with site-specific mutations with SL3 – Flu RNA are summarized in Table 1.

First, the effect of a single NBA unit in NCp7 was evaluated. The dissociation constant for a V13C$_{NBA}$ protein and RNA was estimated as K_d = 435 nM by fluorescence anisotropy measurements, which is a similar value to that for NCp7 (Figure 5). Further, the V13G$_{NBA}$ protein had a K_d value of 350 nM, which indicates a slightly higher binding affinity for the RNA than that of NCp7. The K38C$_{NBA}$ protein also showed a binding affinity for the RNA (K_d = 441 nM) at a level similar to that of NCp7. These

Table 1. *The dissociation constants of NBA-NCp7 with SL3 – Flu RNA.*

Protein	K_d [nM]	K_d^{NCp7}/K_d^{NBA} [a]
NCp7	551 ± 36	1.00
V13C$_{NBA}$	435 ± 36	1.27
V13G$_{NBA}$	350 ± 32	1.57
K38C$_{NBA}$	441 ± 22	1.25
K38G$_{NBA}$	179 ± 7	3.08
M46C$_{NBA}$	198 ± 25	2.78
M46G$_{NBA}$	167 ± 5	3.30
V13C$_{NBA}$K38G$_{NBA}$	278 ± 12	1.98
V13C$_{NBA}$M46C$_{NBA}$	224 ± 15	2.46
V13C$_{NBA}$M46G$_{NBA}$	63 ± 3	8.75
V13G$_{NBA}$M46G$_{NBA}$	122 ± 9	4.52

[a] K_d^{NCp7}/K_d^{NBA} is the ratio of the dissociation constants of NCp7 and NBA-NCp7 with SL3 – Flu RNA.

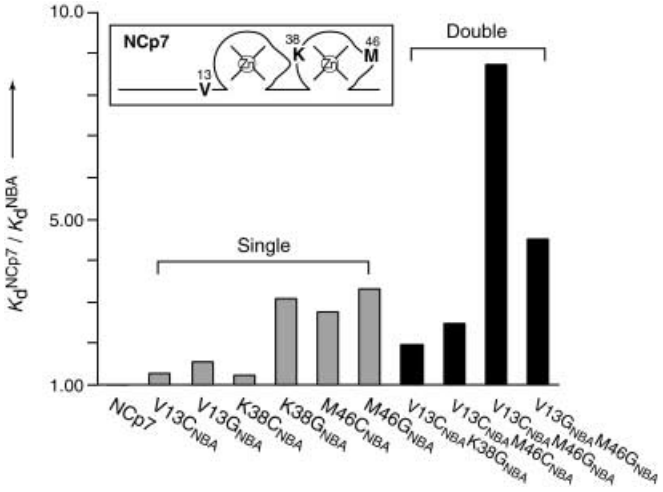


Figure 5. *The relative binding affinities between wild-type NCp7 and NBA-NCp7 proteins with one (gray) and two (black) NBA units and SL3 – Flu RNA. K_d^{NCp7}/K_d^{NBA} is the ratio of the dissociation constants for NCp7 and NBA-NCp7 with SL3 – Flu RNA.*

results indicate that the incorporation of C$_{NBA}$13, G$_{NBA}$13, and C$_{NBA}$38 units into NCp7 is not effective as a way to enhance the RNA-binding ability of the protein. In contrast, the protein with the K38G$_{NBA}$ mutation bound to SL3 – Flu RNA with K_d = 179 nM, a binding affinity 3.1 times stronger than that of NCp7. Further, the dissociation constants of the M46C$_{NBA}$ and M46G$_{NBA}$ proteins with the RNA were 198 and 167 nM, values which indicate affinities 2.8 and 3.3 times higher than that for NCp7, respectively (Figure 5). These results suggest that the G$_{NBA}$38, C$_{NBA}$46, and G$_{NBA}$46 units can work as enhancers for binding to SL3 – Flu RNA, probably as a result of the interaction of the nucleobase moieties with the RNA bases. In the NCp7 – SL3 complex, the Met 46 residue of NCp7 was close to the G 10 base of the RNA.[7] Therefore, guanine and cytosine moieties at position 46 in the protein are functional for interaction with the G 10 base of the RNA. At the Lys 38 site, however, the amino acid side chain is slightly distanced from the G 10 base in the RNA compared with the location of the side chain of the Met 46 residue in the structure of the NCp7 – SL3 complex. It appears that at residue 38 a guanine moiety (the

composition of which includes a purine ring), but not a cytosine group, can interact with the G10 base in the RNA.

Binding affinity of double-NBA-NCp7 for SL3 – Flu RNA

In the case of the single-NBA-NCp7 proteins, an NBA unit at residue 38 or 46 was effective for RNA binding. Although the NBA at position 13 was not effective, incorporation of a second NBA unit at residue 38 or 46 into the single-NBA-NCp7 may improve the ability of the protein to bind to SL3 – Flu RNA. The $V13C_{NBA}M46G_{NBA}$ protein, which possessed C_{NBA} and G_{NBA} units at residues 13 and 46, respectively, bound to SL3 – Flu RNA with $K_d = 63$ nM, an affinity 8.7 times stronger than that of NCp7 (Table 1 and Figure 5). This binding affinity was 6.9 and 2.7 times higher than those of the single mutants $V13C_{NBA}$ and $M46G_{NBA}$, respectively. Although the single-mutation protein $V13C_{NBA}$ bound to the RNA with a strength similar to that of NCp7, the binding affinity of the protein with $V13C_{NBA}M46G_{NBA}$ mutations was significantly increased compared to that of the $M46G_{NBA}$ protein. These results suggest that the cytosine moiety at position 13 in the $V13C_{NBA}M46G_{NBA}$ molecule can make specific contact (such as that in Watson – Crick base pairing) with the base in SL3 – Flu RNA when this contact coincides with the specific interaction between the G_{NBA} 46 moiety of the protein and the G10 base of the RNA. In contrast, the binding affinity of the $V13C_{NBA}M46C_{NBA}$ protein for the RNA was not improved ($K_d =$ 224 nM) compared with that of the protein with an $M46C_{NBA}$ mutation; that is, the C_{NBA} 13 residue cannot in this case contribute to RNA binding. The binding affinity of the $V13C_{NBA}K38G_{NBA}$ analogue for SL3 – Flu RNA was also not improved ($K_d = 278$ nM) compared with the single-mutation protein $K38G_{NBA}$. The $V13C_{NBA}M46G_{NBA}$ protein was an exception for which the cytosine moiety at position 13 worked effectively and cooperatively for RNA binding. In the case of the $V13C_{NBA}K38G_{NBA}$ and $V13C_{NBA}M46C_{NBA}$ proteins, it is possible that interaction of the C_{NBA} 46 or G_{NBA} 38 moiety with the G10 base in the RNA changed the three-dimensional structure of the protein – RNA complex to some extent, such that the C_{NBA} 13 moiety had difficulty interacting with the G12 base in SL3 – Flu RNA. Further, the $V13G_{NBA}M46G_{NBA}$ protein bound to the RNA with $K_d = 122$ nM, an affinity 4.5 and 2.9 times stronger than those of NCp7 and its $V13G_{NBA}$ analogue, respectively. Thus, the G_{NBA} 13 unit also increased the RNA binding affinity in the presence of the specific interaction of G_{NBA} 46 with the RNA base. The binding affinity of the $V13G_{NBA}M46G_{NBA}$ protein for SL3 – Flu RNA, however, was lower than that of the $V13C_{NBA}M46G_{NBA}$ molecule. It was found that the combination of C_{NBA} 13 and G_{NBA} 46 in the protein was most effective for RNA binding by the NBA-NCp7 series.

Binding analysis of NBA-NCp7 with SL3 mutant RNAs

It was revealed that the incorporation into NCp7 of C_{NBA} at position 13 and G_{NBA} at position 46 had significant and cooperative effects on binding to wild-type SL3 – Flu RNA. In order to support the assumption that the specific interaction of the nucleobase moieties in the protein with $V13C_{NBA}M46G_{NBA}$ mutations with the G12 and G10 bases in the tetra-loop region of the wild-type RNA enhanced the RNA binding affinity, we prepared mutant RNAs G12A – Flu and G10A – Flu, in which the G12 and G10 bases of the SL3 – Flu RNA, respectively, were replaced by adenine (Figure 1c).

The binding affinity of the $V13C_{NBA}M46G_{NBA}$ protein for the G12A – Flu mutant RNA ($K_d = 194$ nM) was 3.1 times weaker than that for the wild-type SL3 – Flu RNA (Table 2 and Figure 6). In contrast, NCp7 and the $M46G_{NBA}$ protein showed binding affinities for the mutant RNA ($K_d = 541$ and 171 nM, respectively) similar to those for the wild-type RNA. A comparison of the binding affinities of the proteins with $V13C_{NBA}M46G_{NBA}$ and $M46G_{NBA}$ mutations for the mutant RNA revealed that the effect of the C_{NBA} 13 moiety in the $V13C_{NBA}M46G_{NBA}$ protein was completely removed by the mutation of the G12 base to an A12 one in the RNA. These findings strongly support the suggestion that C_{NBA} 13 can interact specifically with the G12 base in the wild-type RNA, probably in a Watson – Crick base pairing, and that this interaction contributes to an increase in the RNA binding affinity and specificity.

In the case of the G10A – Flu mutant RNA, the $V13C_{NBA}M46G_{NBA}$ protein showed a binding affinity for the mutant RNA with

Table 2. *The dissociation constants of NBA-NCp7 with G12A – Flu and G10A – Flu mutant RNAs.*

Protein	K_d [nM] ($K_d^{mut}/K_d^{SL3[a]}$) G12A – Flu	G10A – Flu
NCp7	541 ± 25 (0.98)	270 ± 10 (0.49)
$M46G_{NBA}$	171 ± 14 (1.02)	121 ± 12 (0.72)
$V13C_{NBA}M46G_{NBA}$	194 ± 17 (3.08)	142 ± 6 (2.25)

[a] K_d^{mut}/K_d^{SL3} is the ratio of the dissociation constants of the proteins with G12A – Flu or G10A – Flu mutant relative to those with SL3 – Flu RNA.

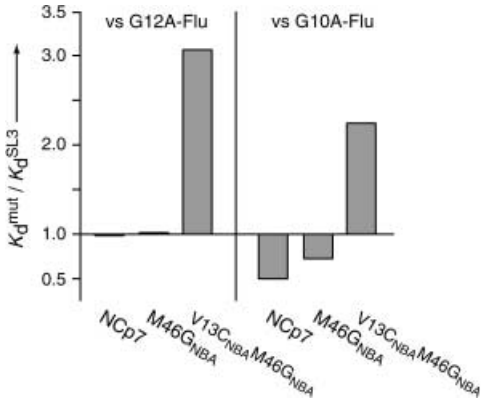


Figure 6. *The relative binding affinities of proteins with wild-type SL3 – Flu RNA and the RNA mutants, G12A – Flu (left) and G10A – Flu (right).* K_d^{mut}/K_d^{SL3} *is the ratio of the dissociation constants for the proteins with the RNA mutants relative to those with SL3 – Flu RNA.*

$K_d = 142$ nM, a value that indicates 2.3 times weaker binding than that for the wild-type RNA (Table 2 and Figure 6). In contrast, the binding affinities of the NCp7 and M46G_{NBA} proteins for the mutant RNA were increased by the G10A mutation ($K_d = 270$ and 121 nM) to values 2.0 and 1.4 times higher than those for the wild-type RNA. The NMR spectroscopy structure of the NCp7–SL3 complex revealed that the G10 base in SL3 RNA interacts with the F2 knuckle protein domain through a hydrophobic cleft formed by the Trp37, Gln45, and Met46 side chains.[7] In the cases of NCp7 and M46G_{NBA}-NCp7, the mutation of the G10 base to an A10 base might increase the effect of the hydrophobic interaction between these amino acid residues and the A10 base of the mutant RNA, which would result in an increase in the binding affinity. In the case of the V13C_{NBA}M46G_{NBA} protein, however, it appears that the mutation of the G10 base to an A10 base weakens the specific and cooperative interaction of the C_{NBA}13 and G_{NBA}46 moieties with the RNA bases. The results revealed that binding of the V13C_{NBA}M46G_{NBA} protein to the wild-type SL3–Flu RNA is more effective and selective than that of NCp7 and the M46G_{NBA} protein.

These mutation analyses with G12A– and G10A–Flu RNA revealed that both the C_{NBA} and G_{NBA} units in the proteins with V13C_{NBA}M46G_{NBA} mutations can contribute significantly to an increase in its binding affinity and specificity for the wild-type SL3–Flu RNA as a result of the specific and cooperative interaction of both the cytosine and guanine moieties in this protein with the G12 and G10 bases in the RNA, respectively.

Conclusion

For the purpose of enhancing the function of HIV-1 NCp7, various NCp7 proteins with NBA units comprised of L-α-amino acids with a nucleobase in the side chain were successfully designed and synthesized by using a native chemical ligation method. The synthetic NCp7 and NBA-NCp7 proteins bound two zinc ions quantitatively. Incorporation of various NBA units into NCp7 successfully enhanced the function of the protein. In particular, the protein with V13C_{NBA}M46G_{NBA} mutations, which possessed C_{NBA} and G_{NBA} units at residue 13 and 46, respectively, showed a rather high binding affinity for SL3–Flu RNA relative to that of NCp7 as a result of the specific and cooperative interaction of both the C_{NBA} and G_{NBA} units with the G12 and G10 bases in the tetra-loop region of SL3 RNA. This study combines a nucleobase function with a protein structure to clearly demonstrate the capability of NBAs as building blocks for the engineering of RNA-binding proteins.

Experimental Section

Materials and methods: All chemicals and solvents were of reagent or HPLC grade. Amino acid derivatives and reagents for peptide synthesis were purchased from Watanabe Chemical (Hiroshima, Japan). MALDI-TOFMS was carried out on a Shimadzu MALDI III mass spectrometer with 3,5-dimethoxy-4-hydroxycinnamic acid as a matrix. HPLC was carried out on a YMC ODS A-302 5C18 column (4.6 × 150 mm; YMC, Tokyo, Japan), or a Cosmosil 5C18 AR-300 column (4.6 × 150 mm or 10 × 250 mm) by employing a Hitachi L-7000 HPLC system. Amino acid analysis was performed by using the phenyl isothiocyanate (PTC) method on a Wakopak WS-PTC column (Wako chemical, Osaka, Japan). Synthetic DNA templates were purchased from Amersham Pharmacia Biotech (Tokyo, Japan). Fluorescein-labeled oligomer DNA (CCA–Flu) was purchased from Espec Oligo Service (Tsukuba, Japan). A RiboMAX large scale RNA production system-T7 was purchased from Promega (Tokyo, Japan). T4 RNA ligase was purchased from Takara Shuzo (Osaka, Japan). Fmoc-C_{NBA}(Z)-OH and Fmoc-G_{NBA}-OH were synthesized according to the procedure previously reported.[6]

Protein synthesis: NCp7 and the NBA-NCp7 proteins were synthesized by the native chemical ligation method reported by Kent et al.[16] by using segment peptides that consisted of residues 1–35 (S1–Gly–COSR; R = CH_2CH_2COOEt) and 36–55 (Cys–S2). S1 fragment peptides were synthesized by the Fmoc solid-phase method[17] with 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and 1-hydroxybenzotriazole hydrate (HOBt·H_2O) as the coupling reagents[17] on 2-chlorotrityl chloride resin[21a, 21b] in an Advanced Chemtech BenchMark model 348 multiple peptide synthesizer. The fully protected S1 fragments were obtained by treatment of the peptide resin with a dichloromethane/2,2,2-trifluoroethanol/acetic acid (1:1:3) solution for 2 h at room temperature.[22] Reaction of each protected peptide with 2-mercaptopropionic acid ethyl ester (10 equiv), 1-ethyl-3-(3′-dimethylaminopropyl)carbodiimide hydrochloride salt (EDC·HCl; 10 equiv), and HOBt·H_2O (10 equiv) gave the fully protected peptide thioester S1–Gly–COSR. Removal of the protecting groups, except for the Z group on the cytosine moiety, was carried out by treatment with trimethylsilyl bromide (TMSBr; 1 M) in trifluoroacetic acid (TFA) solution in the presence of *m*-cresol, ethanedithiol, and thioanisole as scavengers at 0 °C for 1.5 h.[23] The Z group on the cytosine moiety was removed with trimethylsilyl trifluoromethanesulphonate (TMSOTf; 1 M) in TFA in the presence of *m*-cresol, ethanedithiol, and thioanisole at 0 °C for 1.5 h.[24] The product was solidified with diethyl ether in an ice bath. Cys–S2 fragment peptides were synthesized by the Fmoc solid-phase method on Rink amide resin.[25] Removal of the resin and the protecting groups was carried out by the same procedure as described above.

All crude peptides were purified by RP-HPLC on a semipreparative column with a linear gradient of acetonitrile/0.1% TFA. Peptides were identified by their molecular ion peak $(M+H)^+$ by MALDI-TOFMS: *m/z* found (calcd): NCp7(1–35)–COSR, 4153.1 (4153.0); V13C_{NBA}(1–35)–COSR, 4249.2 (4247.1); V13G_{NBA}(1–35)–COSR, 4289.4 (4288.1); NCp7(36–55), 2353.0 (2350.7); K38C_{NBA}(36–55), 2417.8 (2416.7); K38G_{NBA}(36–55), 2458.2 (2457.7); M46C_{NBA}(36–55), 2413.7 (2413.7); M46G_{NBA}(36–55), 2454.2 (2454.7).

NCp7 and the NBA-NCp7 proteins were synthesized by native chemical ligation. S1–Gly–COSR and Cys–S2 were reacted in phosphate buffer (0.1 M; pH 7.5) that contained Gu·HCl (6 M) and thiophenol (4% v/v) at room temperature.[18] This reaction was monitored by analytical RP-HPLC on a Cosmosil 5C18 AR-300 column (2.5 × 150 mm). After 6 h, the reaction was complete and the solution was acidified with 30% aq AcOH and filtered. The product was purified with RP-HPLC on a Cosmosil 5C18 AR-300 column (10 × 250 mm) with a linear gradient of 10–40% acetonitrile/0.1% TFA (30 min). Proteins were identified by their molecular ion peak $(M+H)^+$ by MALDI-TOFMS: *m/z* found (calcd): NCp7, 6370.0 (369.5); V13C_{NBA}, 6463.7 (6463.6); V13G_{NBA}, 6506.9 (6504.6); K38C_{NBA}, 6435.3 (6435.5); K38G_{NBA}, 6475.7 (6475.5); M46C_{NBA}, 6430.6 (6431.5); M46G_{NBA}, 6469.8 (6472.5); V13C_{NBA}K38G_{NBA}, 6570.9 (6570.6); V13C_{NBA}M46C_{NBA}, 6526.8 (6526.6); V13C_{NBA}M46G_{NBA}, 6566.0 (6566.6); V13G_{NBA}M46G_{NBA}, 6608.3 (6607.6).

Preparation of fluorescein-labeled SL3 Ψ RNA and mutants: SL3 RNA was prepared by transcription[26] of a synthetic DNA template (5′-AATTTAATACGACTCACTATAGGACTAGCGGAGGCTAGTCC-3′ and 5′-GGACTAGCCTCCGCTAGTCCTATAGTGAGTCGTATTAAATT-3′) by using the RiboMAX large scale RNA production system-T7 (Promega). The template DNA was degraded by RNase-free DNase I, and then the RNA was purified by Sephadex G-50 gel filtration to give a crude product. The crude product was treated with CCA – Flu by using T4 RNA ligase[20] in Tris · HCl buffer (50 mM; pH 7.5) that contained $MgCl_2$ (10 mM), 1,4-dithiothreitol (DTT; 10 mM), adenosine triphosphate (ATP, 1 mM), bovine serum albumin (BSA; 0.05 %), and dimethyl sulfoxide (DMSO; 10 %). The solution was incubated at 10 °C for 2 h. Ethanol precipitation gave a crude ligation product. The product was purified by PAGE (12 % acrylamide). SL3 – Flu mutants were also prepared by the method described above. Concentrations of RNA were determined by UV spectroscopy.

Visible absorption measurements: Visible absorption spectra were recorded on a Shimadzu BioSpec-1600 spectrophotometer by using a quartz cell with a 1.0 cm pathlength at 25 °C. Proteins were dissolved in Tris · HCl buffer (50 mM; pH 7.5).

Fluorescence anisotropy measurements: Fluorescence anisotropy measurements were performed on a Shimadzu RF-5300PC spectrofluorophotometer by using a quartz cell with a 1.0 cm pathlength at 10 °C in *N*-(2-hydroxylethyl)piperazine-*N*-(2-ethanesulfonic acid) (HEPES) · NaOH buffer (10 mM; pH 7.5) containing NaCl (150 mM), $MgCl_2$ (1 mM), and $ZnCl_2$ (50 μM). Fluorescence anisotropy was calculated from the intensities detected at 520 nm with excitation at 480 nm.

Determination of dissociation constants between proteins and RNA: The binding affinities between the experimental proteins and RNA were determined by fluorescence anisotropy measurements. The protein stock solution (200 – 250 μM in water) contained $ZnCl_2$ (550 μM). A fluorescent RNA (5.0 nM) in HEPES buffer (10 mM; pH 7.5) that contained NaCl (150 mM), $MgCl_2$ (1 mM), and $ZnCl_2$ (50 μM) was titrated with protein solution. After each addition of protein, samples were stirred for 40 s and allowed to equilibrate for 5 min at 10 °C, then fluorescence anisotropy was measured. The dissociation constants of the proteins and RNA were calculated by using Equation (1) and a Kaleida Graph (Synergy Software), in which a 1:1 stoichiometry was assumed.

$$A = (A_b - A_f)([F]_0 + [P]_0 + K_d - (([F]_0 + [P]_0 + K_d)^2 - 4([F]_0[P]_0)^{1/2})/(2[F]_0) \quad (1)$$

$[F]_0$ and $[P]_0$ represent the initial concentrations of fluorescent RNA and protein, respectively, and A, A_b, and A_f are the anisotropy values of each test solution, of the bound fluorescent RNA, and of the free fluorescent RNA, respectively. K_d is the dissociation constant between the protein and RNA.

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Culture, and Sports, Japan. T.T. is a Research Fellow of the Japan Society for the Promotion of Science.

[1] D. E. Draper, *J. Mol. Biol.* **1999**, *293*, 255 – 270.
[2] P. Legault, J. Li, J. Mogridge, L. E. Kay, J. Greenblatt, *Cell* **1998**, *93*, 289 – 299.
[3] J. D. Puglisi, L. Chen, S. Blanchard, A. D. Frankel, *Science* **1995**, *270*, 1200 – 1203.
[4] G. W. Witherell, J. M. Gott, O. C. Uhlenbeck, *Prog. Nucleic Acid Res. Mol. Biol.* **1991**, *40*, 185 – 220.
[5] T. Takahashi, K. Hamasaki, I. Kumagai, A. Ueno, H. Mihara, *Chem. Commun.* **2000**, 349 – 350.
[6] T. Takahashi, K. Hamasaki, A. Ueno, H. Mihara, *Bioorg. Med. Chem.* **2001**, *9*, 991 – 1000.
[7] R. N. De Guzman, Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, M. F. Summers, *Science* **1998**, *279*, 384 – 388.
[8] J. M. Coffin, S. H. Hughes, H. E. Varmus, *Retroviruses*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1997**.
[9] A. Aldovini, R. A. Young, *J. Virol.* **1990**, *64*, 1920 – 1926.
[10] F. Clavel, J. M. Orenstein, *J. Virol.* **1990**, *64*, 5230 – 5234.
[11] M. Poznansky, A. M. L. Lever, L. Bergeron, W. Haseltine, J. Sodroski, *J. Virol.* **1991**, *65*, 532 – 536.
[12] R. D. Berkowitz, S. P. Goff, *Virology* **1994**, *202*, 233 – 246.
[13] F. Baudin, R. Marquet, C. Isel, J.-L. Darlix, B. Ehresmann, C. Ehresmann, *J. Mol. Biol.* **1993**, *229*, 382 – 397.
[14] J. L. Clever, R. A. Taplitz, M. A. Lochrie, B. Polisky, T. G. Parslow, *J. Virol.* **2000**, *74*, 541 – 546.
[15] K. Sakaguchi, N. Zambrano, E. T. Baldwin, B. A. Shapiro, J. W. Erickson, J. G. Omichinski, G. M. Clore, A. M. Gronenborn, E. Appella, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5219 – 5223.
[16] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776 – 779.
[17] W. C. Chan, P. D. White in *Fmoc Solid Phase Peptide Synthesis: A Practical Approach* (Eds.: W. C. Chan, P. D. White), Oxford University Press, New York, **2000**, p. 41 – 76.
[18] P. E. Dawson, M. J. Churchill, M. R. Ghadiri, S. B. H. Kent, *J. Am. Chem. Soc.* **1997**, *119*, 4325 – 4329.
[19] C. Z. Dong, N. Jullian, Y. S. Yang, H. de Rocquigny, M. C. Fournie-Zaluski, B. P. Roques, *J. Am. Chem. Soc.* **1995**, *117*, 2726 – 2731.
[20] J. F. Milligan, O. C. Uhlenbeck, *Methods Enzymol.* **1989**, *180*, 51 – 62.
[21] a) K. Barlos, D. Gatos, J. Kallitsis, G. Papaphotiu, P. Sotiriu, Y. Wenging, W. Schäfer, *Tetrahedron Lett.* **1989**, *30*, 3943 – 3946; b) K. Barlos, D. Gatos, S. Kapolos, G. Papaphotiu, W. Schäfer, Y. Wenqing, *Tetrahedron Lett.* **1989**, *30*, 3947 – 3950.
[22] K. Barlos, O. Chatzi, D. Gatos, G. Stavropoulos, *Int. J. Peptide Protein Res.* **1991**, *37*, 513 – 520.
[23] S. Funakoshi, E. Murayama, L. Guo, N. Fujii, H. Yajima, *J. Chem. Soc. Chem. Commun.* **1988**, 382 – 383.
[24] N. Fujii, A. Otaka, O. Ikemura, K. Akaji, S. Funakoshi, Y. Hayashi, Y. Kuroda, H. Yajima, *J. Chem. Soc. Chem. Commun.* **1987**, 274 – 275.
[25] H. Rink, *Tetrahedron Lett.* **1987**, *28*, 3787 – 3790.
[26] P. J. Romaniuk, O. C. Uhlenbeck, *Methods Enzymol.* **1983**, *100*, 52 – 59.

Received: November 12, 2001 [F318]