Analysis of the Peptides (Prp106—126, MSI-78A, and Oxaldie 1) with the Same Biological Activity by Discrete Fourier Transform: Toward a Selection Rule in Ligand–Receptor Interaction

Naganori NUMAO,*[,a] Hisashi FUJII,[b] Yoshiyuki FUKAZAWA,[b] Ken YOSHIOKA,[b] Mayumi OKADA,[a] and Kazuyoshi TANAKA[c]

[a] *BioFrontier Institute Inc.; 5–4–21 Nishi-Hashimoto, Sagamihara, Kanagawa 229–1131, Japan:* [b] *Kanagawa Industrial Technology Research Institute; 705–1 Shimoimaizumi Ebina, Kanagawa 243–0435, Japan: and* [c] *Department of Molecular Engineering, Kyoto University Graduate School of Engineering; Sakyo-ku, Kyoto 606–8501, Japan.*
Received January 8, 2003; accepted February 13, 2003

Here, we first report a novel method in which the "desired cross-spectrum" of the peptides Prp106—126, MSI-78A, and oxaldie 1 with the same biological activities is obtained by the multiplication of two cross-spectra derived from the RNA sequence and from the cognate amino acid sequence by discrete Fourier transform (DFT), respectively. Based on a well-known method reported previously, we investigated the cross-spectrum by the multiplication of two of three desired cross-spectra. As a result, we found that one prominent peak occurring in the three cross-spectra showed the same frequency when a binary scale was used as a parameter of nucleotide or amino acid in the analysis. Moreover, we examined the relationship between a binary scale and other physicochemical ones. Almost the same results could be reproduced when the absolute elctronegativity scale (or the absolute hardness one) was used, but not in the case of the hydrophobic or electron–ion interacting potential scale reported previously. This indicates that either the absolute electronegativity scale (or the absolute hardness one) or a binary scale, or both is very useful in extracting the information desired for various proteins by the present method from the amino acid and the RNA sequence.

Key words bioinformatic; discrete fourier transform; electronegativity; hardness; binary; prion

We have recently reported[1,2] that three synthetic peptides, Prp106—126 (NH_2-KTNMKHMAGAAAAGAVVGGLG-COOH), MSI-78A (NH_2-GIGKFLKKAKKFAKAFVKILK-K-$CONH_2$), and oxaldie 1 (NH_2-LAKLLKALAKLLKK-$CONH_2$), and their reversed peptides, namely Prp126—106 (NH_2-GLGGVVAGAAAAGAMHKMNTK-COOH), A87-ISM (NH_2-KKLIKVFAKAFKKAKKLFKGIG-$CONH_2$) and 1 eidlaxo ((NH_2-KKLLKALAKLLKAL-$CONH_2$), that are in opposite amino- to carboxy-orientation (*vice versa*) could exert decarboxylase activity for oxaloacetate in the presence of trifluoroethanol (TFE) in aqueous solution (pH 7.0) to give pyruvate and carbon dioxide. The naturally occurring peptides show generally stronger activity than the corresponding reversed peptides. In this, Prp106—126 corresponds to the amino acid number 106—126 of the prion protein sequence (*Homo sapiens*), which may be responsible for conversion of cellular prion protein (PrP^C) to an infective toxic molecule.[3] The secondary structure around the region has been clarified to be the disordered one from the analysis of the three-dimensional structure of recombinant PrP^C by NMR measurement,[4] while the α-helical structure has been postulated by various secondary structure prediction methods.[5] MSI-78A[6] is a synthetic analogue of naturally occurring magainin 2,[7] which is an antibacterial peptide and isolated from the skin of *Xenopus laevis*. Oxaldie 1 has been rationally designed and synthesized to express decarboxylase activity,[8] but the amino acid sequence is homologous to mastparan M (amino acid numbers 1—12: INLKALAA-LAKK) (*Vespa mandarnia*) or partially to mellitin (amino acid numbers 12—23: GLPALISWIKRK) (*Apis mellifera*). The peptides MSI-78A and oxaldie 1 have been known to form amphipathically α-helical polycationic structures as a common factor necessary for expressing activity.

In the field of bioinformatics, the concept of sequence Fourier analysis is not new. Discrete Fourier transform (DFT) analysis has been well known to be able to provide valuable information on the relationship between the characteristic periodicity and bioactivity, higher-ordered dimensional structure, or evolutionary change in the macromolecular sequences of protein, DNA, or RNA. In these studies, the electron–ion interacting potential (EIIP) scale,[9,10] hydrophobic scale,[11] or binary scale[12,13] has usually been used as a parameter.

The informational system method (ISM)[9,10] is a theoretical method using DFT analysis for the informational contents of protein or nucleic acid distribution. It has frequently been emphasized that one characteristic frequency could be extracted from the cross-spectrum of at least two or more protein sequences with the same biological function in this method.[10] The analytical results seemed to hint at novel bioinformatics, excluding the stereochemistry of macromolecules. Furthermore, it has been suggested that the ISM result does not depend on the degree of homology between the amino acid sequences,[10] indicating that the same periodicity is contained in any amino acid sequences with the same biological activity.

In the preceding study,[1] we investigated the ISM in order to elucidate the same biological activity of Prp106—126, MSI-78A, and oxaldie 1 and their reversed peptides (*vide supra*). In the study, it was accepted that the spectra among amino acid sequences of Prp106—126, MSI-78A, and oxaldie 1 showed the same features as those from the reversed peptides Prp126—106, A78-ISM, and 1 eidlaxo, respectively. This indicates that the naturally occurring amino acid sequence and the reversed one have the same periodicity. However, one major characteristic peak that occurred in

* To whom correspondence should be addressed. e-mail: numao-n@sssc.co.jp

the cross-spectrum between Prp106—126 and oxaldie 1 (or MSI-78A and oxaldie 1) was not similar or the same as that between Prp106—126 and MSI-78A. Based on this information, there occurred a question as to the appropriateness of the parameters, namely, the EIIP scale[9,10] assigned to each amino acid which were used in the DFT analysis.

So far, a theoretical manipulation[14] and the experimental results[15—18] to determine a specific interaction between monomeric ribonucleotide and amino acid have been reported, while the preferentially catalytic activities of RNA at the beginning of evolution was suggested,[19] based on the discovery of the enzymic activities in the RNA molecule.[20,21] In the theory,[14] a basic model of the interaction between monomeric amino acid and monomeric nucleic acid has been proposed to determine the definite-sequence polymers, *i.e.*, in terms of polypeptides or polynucleotides. This model is very attractive, although the molecular structure of a given nucleic acid or amino acid is obscure. In experiments,[15] a direct and significant correlation between the amino acid and the anticodonic ribonucleotide monophosphate has been examined. The anticodon theory in the origin of coded protein synthesis has been strongly proposed, based on a relationship between the average hydrophobicity rankings of amino acids and that of their dinucleotide monophosphates, while the final conclusion on the protein synthesis seems to have been determined.[22]

We independently reported that the complementary units (*e.g.*, GA or MXH) of amino acids are evolutionarily conserved as a signal around the active sites of various proteins.[23] Such the complementary units are conserved in Prp106—126 (*vide supra*). The complementary units correspond to something like a restriction site, composed of 6 or 9 bases on DNA. Taking together our previous report with the specific interaction between monomeric amino acid and monomeric nucleic acid described above, it appears worthwhile to investigate the multiplication of the cross-spectrum derived from the mRNA sequence and the cognate amino acid sequence of the peptides (*i.e.*, Prp106—126, MSI-78A, and oxaldie 1) with the same biological activity.

In the present article, the computational results of this concept are reported.

Computational Method

To analyze an RNA (or DNA) sequence and the cognate amino acid sequence by DFT, we employed the program constructed previously,[1] which is almost the same as that of the ISM described above. Our program was built in Microsoft Excel software with Basic language in the Windows PC operating system. The codon usage database on *H. sapiens*, *X. laevis*, *A. mellifera*, and *O. keta* are available in NCBI (http://www3.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=t), but *V. mandarnia* was unavailable.

The DFT procedure is as follows. At first, four kinds of nucleotides and 20 kinds of amino acids are assigned individual parameters. These parameters are on (i) the binary scale, (ii) the average absolute electronegativity scale, (iii) the average absolute hardness scale, *etc*. Details of choice of these parameters are described in the following section. Next, the parameter sequence $\{f_n;\ n=1,2,3,\cdots,2^k\}$ was constructed by conforming to a given nucleotide (and the cognate amino acid) sequence, namely, $a_0a_1a_2a_3\cdots a_N$, where 2^k is the minimum value that the nucleotide number N does not exceed ($2^{k-1}<N\leqq 2^k$). Note that for the values from f_{N+1} to f_{2^k} were set at the mean value of $\{f_n;\ n=1,2,3,\cdots,N\}$ in the nucleotide sequence. On the other hand, for the cognate amino acid sequence, the values from $f_{N/3+1}$ to f_{2^k} were set at the mean value of $\{f_n;\ n=1,2,3,\cdots,N/3\}$ in the amino acid sequence. This prescription originating from the digital signal treatment[24] was employed in the previous study with the EIIP parameterization scheme.[25]

The parameter sequence $\{f_n;\ n=1,2,3,\cdots,2^k\}$ thus obtained is subjected to the DFT procedure based on the following equation.

$$\mathrm{F}(m)=\sum_{n=1}^{2^k} f_n \exp(\mathrm{i}2\pi mn/2^k),\quad m=0,1,2,3,\cdots,2^k/2 \tag{1}$$

Where F(m) is the m-th Fourier coefficients of $\{f_n\}$ and i an imaginary unit. Note that it is sufficient to consider m in the range from 0 to $2^k/2$ ($=2^{k-1}$) due to the usual Fourier transform feature.

Third, the cross-spectra for nucleotides and amino acids, M1n(m) and M2a(m), respectively, from four groups of the parameter sequences {Fi(m); i=1n—4n or 1a—4a} (*vide infra*) are defined as follows,

$$\mathrm{M1n}_{(m)}=\mathrm{F1n}(m)*\mathrm{F2n}(m)*\mathrm{F3n}(m)*\mathrm{F4n}(m) \tag{2}$$

$$\mathrm{M2a}_{(m)}=\mathrm{F1a}(m)*\mathrm{F2a}(m)*\mathrm{F3a}(m)*\mathrm{F4a}(m) \tag{3}$$

Finally, the "desired cross-spectrum" ($\mathrm{DM}_{(m)}$) between the nucleotide sequences and the cognate amino acids is derived by multiplication of each cross-spectrum,

$$\mathrm{DM}_{(m)}=\mathrm{M1n}_{(m)}*\mathrm{M2a}_{(m)} \tag{4}$$

for the assessment of the analogy of peptides with the same biological activities.

Results and Discussion

The aim of this study was to determine the cross-spectrum between the desired spectra of Prp106—126 and MSI-78A, of Prp106—126 and oxaldie 1, and of MSI-78A and oxaldie 1, and to find their resemblance. It is of importance to select appropriate parameterization for each nucleic acid and amino acid prior to proposing a selection rule that may elucidate a specific protein–protein interaction from the sequence. In this section, the analysis results in line with the above-mentioned DFT method are shown itemized by each parameterization scheme.

(i) Binary Scale Parameterization Scheme At first, based on the information previously reported[14,15] in the specific interaction between a nucleotide and an amino acid, the parameters were set for four nucleotides (u, c, a, g) in the naturally occurring mRNA sequence[26] of Prp106—126 in the binary scale (*i.e.*, 0 and 1), that is, one of four nucleic acids takes the value of 1 and other three 0: group 1 (1 for u and 0 for c, a, and g); group 2 (1 for c and 0 for u, a, and g); group 3 (1 for a and 0 for u, c, and g); and group 4 (1 for g and 0 for u, c, and a). {F1n(m)} to {F4n(m)} were constructed by Eq. 1 for any mRNA sequences. The simple classification method was the same as that reported previously by Tiwari *et al.*[12] In this, a sharp peak at frequency f=1/3=0.3333 in the Fourier spectrum has been reported to be a good discriminator of coding potential in a genome.

Under the condition of m=64 (2^6 points in the DFT as described above), we computed {Fi(m); i=1n—4n} for mRNA sequences of Prp106—126, composed of 63 nucleic acids. Then, these were multiplied by Eq. 2 to give a cross-spectrum (Fig. 1a), in which two sharp peaks (f=0.3125 and 0.3438) around at frequency f=0.3333 was observed. Meanwhile, the corresponding virtual antisense mRNA (anti-mRNA) was also computed to give another cross-spectrum (Fig. 1b). Interestingly, both spectra (Fig. 1a and Fig. 1b) were completely the same. This tendency was also observed in the two kinds of RNA sequences of MSI-78A or oxaldie 1 (*vide infra*) as well as other longer nucleotide sequences of various proteins, indicating that the same frequency occurred in both the sense and the antisense strand of RNA (or DNA). In the two RNA sequences of Prp106—126, the same M1n(m) spectra (Fig. 2a and Fig. 2b) were obtained in the in-

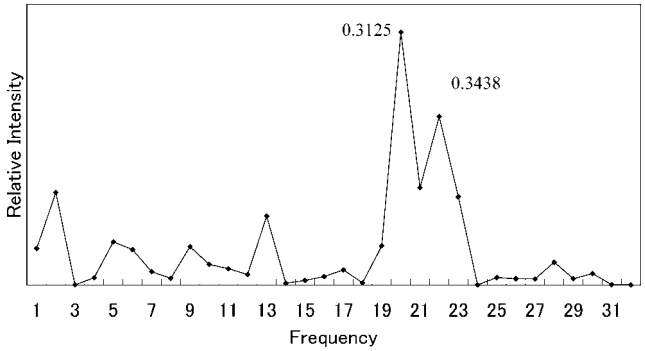

Fig. 1a. The Cross-Spectrum of mRNA of Prp 106—126 Using the Binary-Scale Parameterization of Groups 1—4

The abscissa (m) represents frequencies from 0.0000 to 0.5000 (signifying $2^6/2$) and the ordinate relative intensities (amplitudes) in the spectrum throughout all the figure captions in this article, unless otherwise noted. The number indicated in the figure is the frequency value.

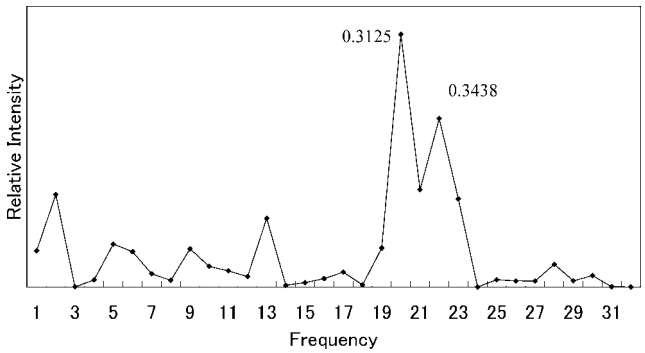

Fig. 1b. The Cross-Spectrum of Anti-mRNA of Prp106—126 Using the Binary-Scale Parameterization of Groups 1—4

See also the caption of Fig. 1a.

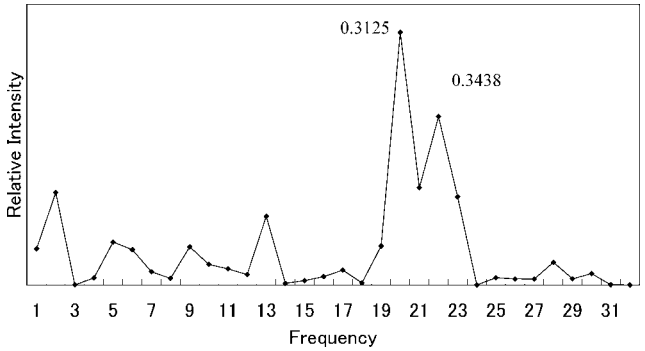

Fig. 2a. The Cross-Spectrum of mRNA of Prp106—126 Using the Binary-Scale Parameterization of Groups 5—8

See also the caption of Fig. 1a.

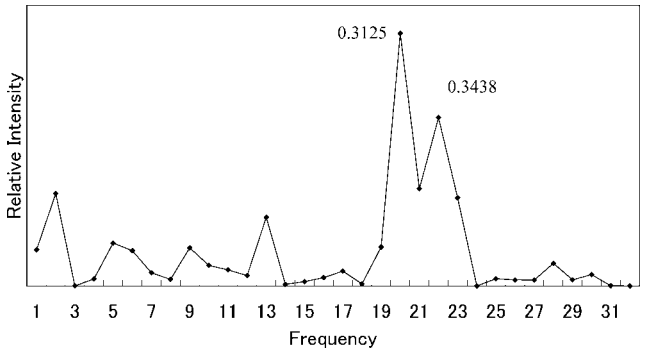

Fig. 2b. The Cross-Spectrum of Anti-mRNA of Prp106—126 Using the Binary-Scale Parameterization of Groups 5—8

See also the caption of Fig. 1a.

terchanged binary scale using group 5 (0 for u and 1 for c, a, and g), group 6 (0 for c and 1 for u, a, and g), group 7 (0 for a and 1 for u, c, and g) and group 8 (0 for g and 1 for u, c, and a). Furthermore, the spectra in Fig. 1 and Fig. 2 are also the same, signifying that the original and its interchanged binary parameters like groups 1 and 5 give the same spectra. Based on the fact that the triplet (*i.e.*, codon) corresponds to a cognate amino acid, the parameter groups 5—8 seem to be more reasonable rather than the groups 1—4.

Next, we determined the parameter group for the amino acids as follows. Based on group 5 (0 for u), we constructed group 9 (1 for N, G, E, P, H, K, A, Q, S, T, R, and D; 0 for other amino acids) by excluding the amino acids having u as the first and the second nucleotide in the codon (or triplet). This concept has come from an indication that the second and third anticodonic bases, namely, the first and second codonic bases, recognize the amino acids.[18] Groups 10—12 were similarly obtained as: group 10 (1 for L, I, N, G, V, E, K, Y, W, M, S, C, F, R, and D; 0 for other amino acids); group 11 (1 for L, G, V, P, A, W, S, C, F, and R; 0 for other amino acids); and group 12 (1 for L, I, N, P, H, K, Y, Q, M, S, T, and F; 0 for other amino acids). Using these parameterization scales, we incrementally calculated $\{Fi(m); i=1a—4a\}$ for 21 amino acids of Prp106—126 and then obtained the cross-spectrum M2a(m) shown in Fig. 3a. Then the desired cross-spectrum $DM_{(m)}$ of Prp106—126 was calculated (see Fig. 4) using this M2a(m) and the above M1n(m) of the corresponding RNA sequence in Fig. 2a. The virtual amino acid sequence translated from the antisense mRNA (5′–3′) sequence was also calculated. The obtained cross-spectrum in Fig. 3b is dissimilar to that (Fig. 3a) of naturally occurring amino acid sequence, unlike the case of RNA (or DNA). Such dissimilarity between naturally occurring and virtual amino acid sequence was observed in various other proteins, including MSI-78A and oxaldie 1.

Finally, the desired cross-spectrum from the mRNA sequence and the naturally occurring amino acid sequence of Prp106—126 was calculated by Eq. 4, which is shown in Fig. 4, in which two characteristic peaks (f=0.0313 and 0.0781) occurred in the range of the lower frequency. They differ from those of either mRNA (f=0.3125 and 0.3438) or amino acid (f=0.0469 and 0.2344), indicating that the desired cross-spectrum is hybridized.

We applied the series of procedures described above to the amino acid sequence of MSI-78A or oxaldie 1, and their corresponding mRNA to synthesize the desired cross-spectrum. For MSI-78A, the triplet nucleotide with the highest occurring frequency in *X. laevis* was used, because MSI-78A is a synthetic analogue of magainin 2. The desired cross-spectrum of MSI-78A is shown in Fig. 7, including a cross-spectrum (Fig. 5) from the mRNA sequence and that (Fig. 6) from the amino acid sequence. In this case, one lysine residue at the C-terminal end of the full length of MSI-78A is removed for convenience of computation, because it is fairly similar to that of MSI-78A itself. In Fig. 7, one characteristic peak occurred in the frequency f=0.0938. On the

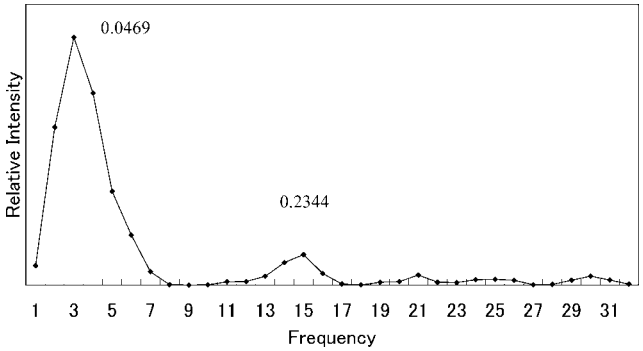

Fig. 3a. The Cross-Spectrum of the Naturally Occurring Amino Acid Sequence of Prp106—126 Using the Binary-Scale Parameterization of Groups 9—12

See also the caption of Fig. 1a.

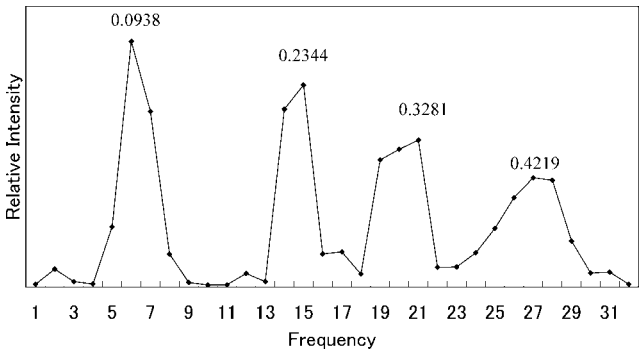

Fig. 3b. The Cross-Spectrum of the Virtual Amino Acid Sequence of Prp106—126 Using the Binary-Scale Parameterization of Groups 9—12

See also the caption of Fig. 1a.

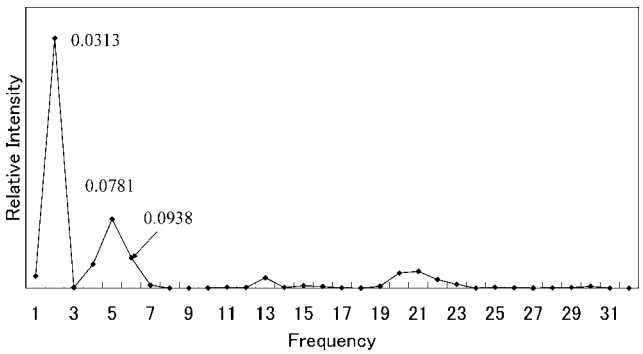

Fig. 4. The Desired Cross-Spectrum of From mRNA and the Cognate Amino Acid (Naturally Occurring) Amino Acid Sequence of Prp106—126, Using the Binary-Scale Parameterization of Groups 5—8 and 9—12, Respectively

See also the caption of Fig. 1a.

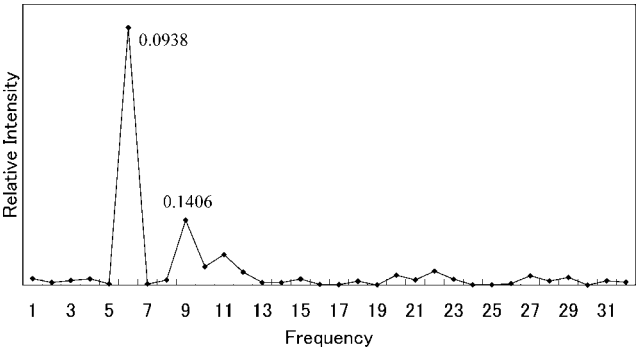

Fig. 5. The Cross-Spectrum of mRNA Represented with All the Highest Occurring Codons of MSI-78A Using the Binary-Scale Parameterization of Groups 5—8

See also the caption of Fig. 1a.

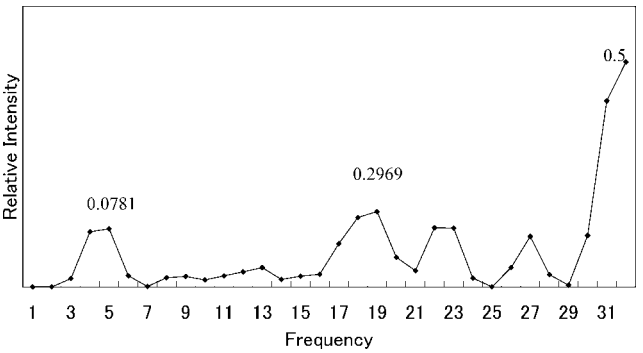

Fig. 6. The Cross-Spectrum of the Amino Acid Sequence of MSI-78A Using the Binary-Scale Parameterization of Groups 9—12

See also the caption of Fig. 1a.

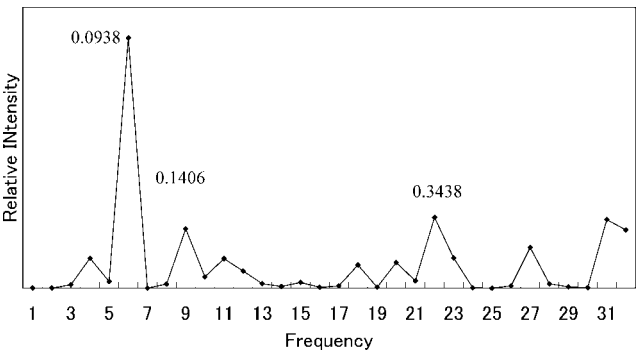

Fig. 7. The Desired Cross-Spectrum of mRNA Represented with All the Highest Occurring Codon and the Cognate Amino Acid Sequence of MSI-78A Using the Binary-Scale Parameterization of Groups 5—8 and 9—12, Respectively

See also the caption of Fig. 1a.

other hand, the peptide oxaldie 1 is completely artificial, and its amino acid sequence is more highly homologous to mastparan M (amino acid number 2—12) (*V. mandarnia*) than mellitin (amino acid number 12—23) (*A. mellifera*). Since, however, the codon usage table of *V. mandarnia* was not available, the codon preference in *A. mellifera* had to be adopted. The desired cross-spectrum of oxaldie 1 is shown in Fig. 10, including a cross-spectrum (Fig. 8) derived from the mRNA and that (Fig. 9) from the amino acid sequence. In Fig. 10, one characteristic peak occurred in the frequency f=0.0938, while it is accompanied by four other major peaks. Furthermore, based on Fig. 5 and Fig. 8, no sharp peaks around frequency f=0.3333 was observed, unlike the case of Prp106—126. This was also indicated in the case of replacing the mRNA sequences of MSI-78A and oxaldie 1 (*vide infra*) with the lowest occurring codon in *X. laevis* and *A. mellifera*, respectively (data not shown). Based on these, a sharp peak at frequency f=0.3333 is not always characteristic in the partial gene.

From the spectra in Fig. 4, Fig. 7, and Fig. 10, we noticed that all the desired cross-spectra are nearly similar in the re-

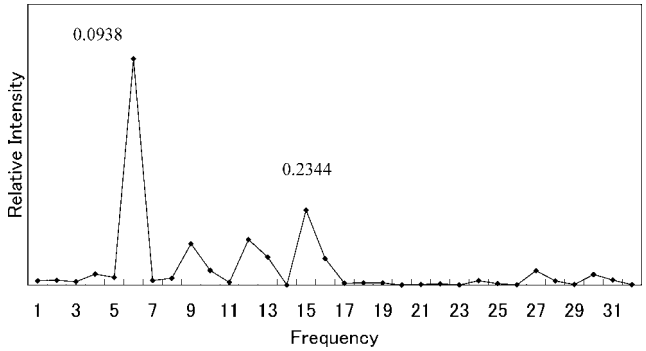

Fig. 8. The Cross-Spectrum of mRNA Represented with All the Highest Occurring Codons of Oxaldie 1 Using the Binary-Scale Parameterization of Groups 5—8

See also the caption of Fig. 1a.

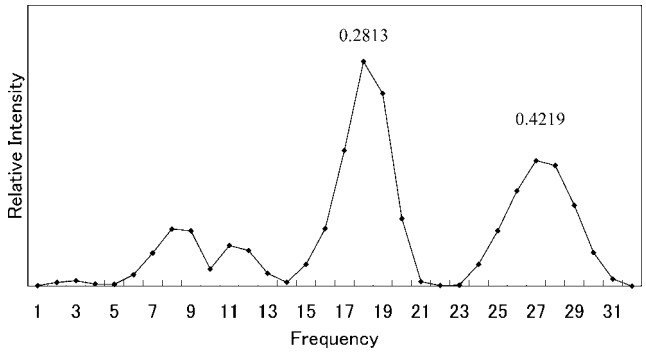

Fig. 9. The Cross-Spectrum of the Amino Acid Sequence of Oxaldie 1 Using the Binary-Scale Parameterization of Groups 9—12

See also the caption of Fig. 1a.

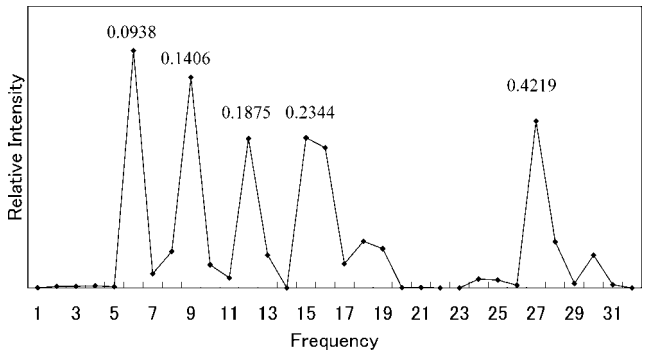

Fig. 10. The Desired Cross-Spectrum of from mRNA Represented with All the Highest Occurring Codons and the Cognate Amino Acid Sequence of Oxaldie 1 Using the Binary-Scale Parameterization of Groups 5—8 and 9—12, Respectively

See also the caption of Fig. 1a.

gion of low frequency. Furthermore, multiplications of these three kinds of desired cross-spectra are shown in Fig. 11 (*i.e.*, Fig. 4*Fig. 7), Fig. 12 (*i.e.*, Fig. 7*Fig. 10), Fig. 13 (*i.e.*, Fig. 10*Fig. 4) and Fig. 14 (*i.e.*, Fig. 4*Fig. 7*Fig. 10). Based on these, the peptides with the same biological activities were clarified to have at least one characteristic peak (f=0.0938) in the cross-spectrum, irrespective of the degree of homology between their amino acid sequences. We also investigated the desired cross-spectra for Prp106—126, MSI-78A, and oxaldie 1 after replacing the corresponding nucleotide se-

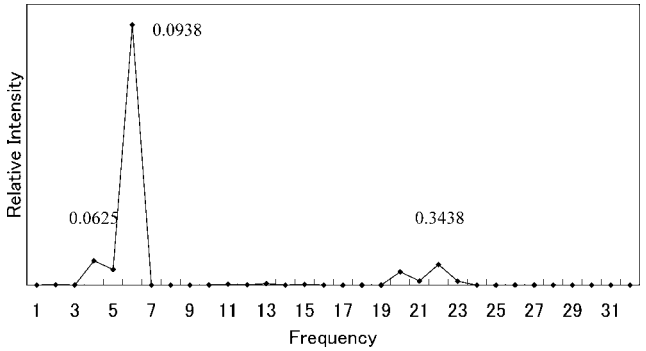

Fig. 11. Multiplication of the Desired Cross-Spectrum of Prp106—126 and that of MSI-78A.

See also the caption of Fig. 1a.

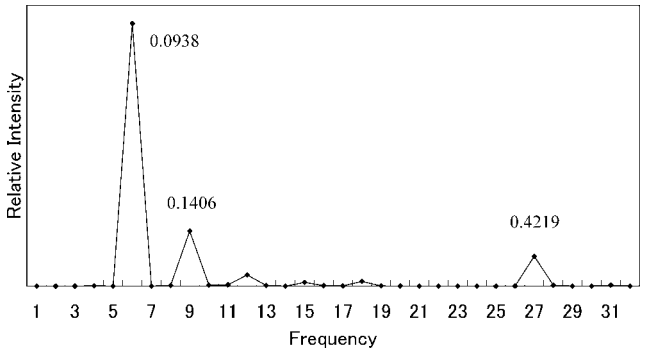

Fig. 12. Multiplication of the Desired Cross-Spectrum of MSI-78A and that of Oxaldie 1

See also the caption of Fig. 1a.

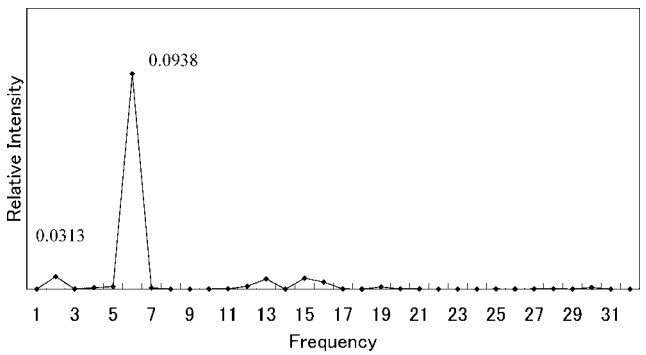

Fig. 13. Multiplication of the Desired Cross-Spectrum of Oxaldie 1 and that of Prp106—126

See also the caption of Fig. 1a.

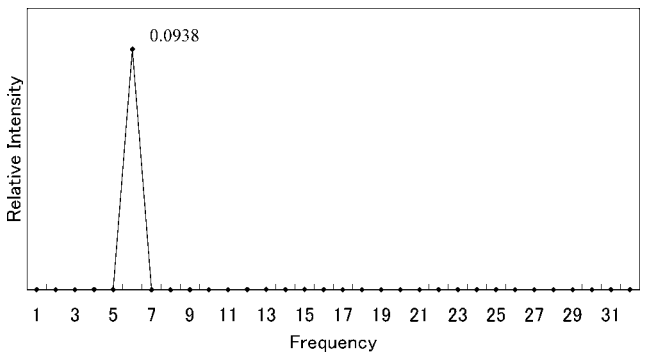

Fig. 14. Multiplication of Three Desired Cross-Spectra (Prp106—126, MSI-78A, and Oxaldie 1)

See also the caption of Fig. 1a.

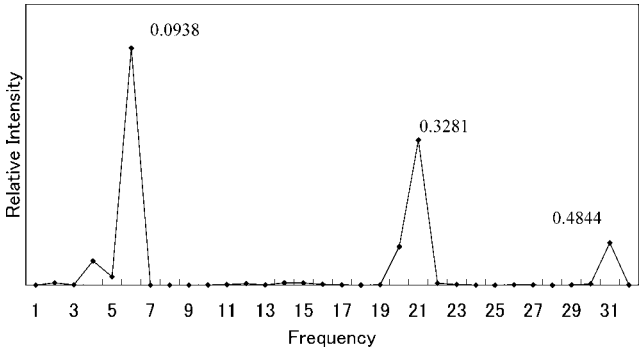

Fig. 15. Multiplication of the Desired Cross-Spectrum of Prp106—126 and for MSI-78 Using the Codons with the Lowest Occurring Frequencies

See also the caption of Fig. 1a.

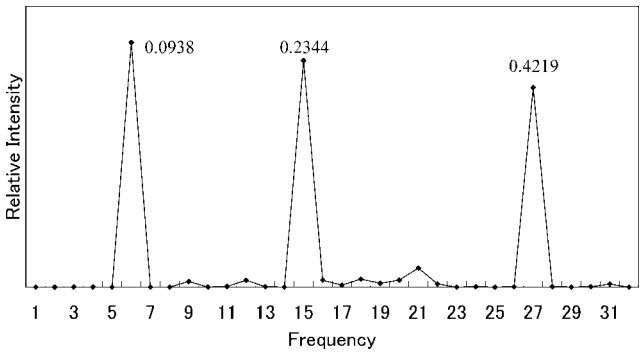

Fig. 16. Multiplication of the Desired Cross-Spectrum of MSI-78A and for Oxaldie 1 Using the Codons with the Lowest Occurring Frequencies

See also the caption of Fig. 1a.

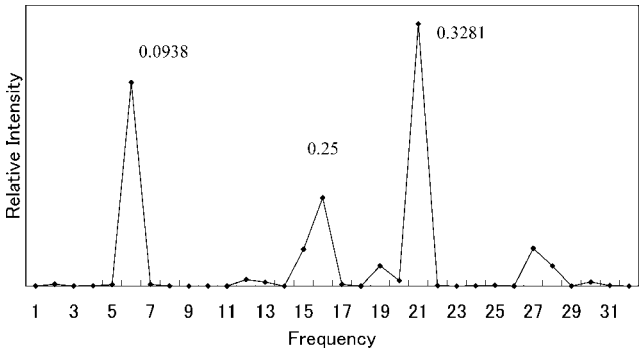

Fig. 17. Multiplication of the Desired Cross-Spectrum of Oxaldie 1 and for Prp106—126 Using the Codons with the Lowest Occurring Frequencies

See also the caption of Fig. 1a.

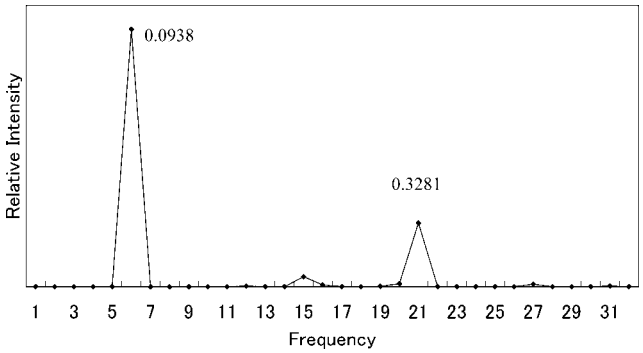

Fig. 18. Multiplication of Three Desired Cross-Spectra (Prp106—126, MSI-78A, and Oxaldie 1), Using the Codons with the Lowest Occurring Frequencies

See also the caption of Fig. 1a.

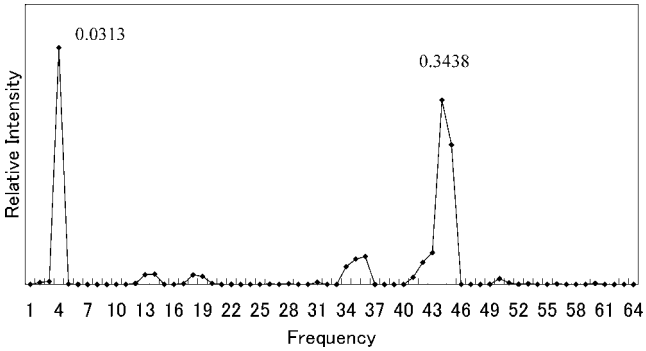

Fig. 19. Multiplication of the Desired Cross-Spectrum of hCT and for sCT Using the Codons with Naturally Occurring Frequencies

The abscissa (m) represents frequencies from 0.0000 to 0.5000 (signifying $2^7/2$) and the ordinate relative intensities (amplitudes) in the spectrum.

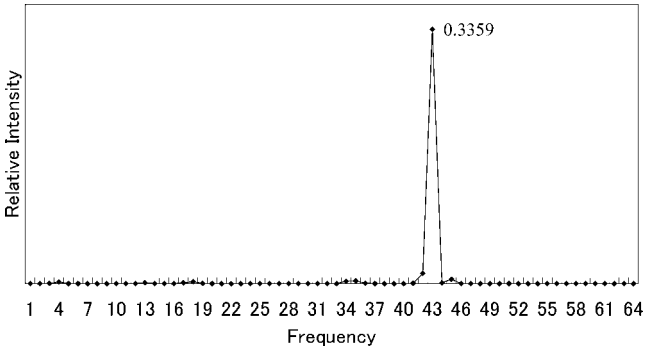

Fig. 20. Multiplication of the Desired Cross-Spectrum of hCT and for sCT Using the Codons with the Highest Occurring Frequencies

See also the caption of Fig. 19.

quence considering the lowest occurring triplet in *H. sapiens*, *X. laevis*, and *A. mellifera*, respectively. The result is shown in Figs. 15—18. This check is of interest since the highest occurring codon is not always used in the naturally occurring mRNA sequence, which encodes polypeptides. As clearly seen in Figs. 15—18, the frequency (f=0.0938) of the prominent peaks are well conserved, though the order of peak height occurring in the multiplied desired cross-spectra is different from those calculated using the highest occurring codon. The comparison of Figs. 11—13 with Figs. 15—17 suggests that the characteristic peak occurring in the present method is fairly susceptible to codon usage.

To verify this, we investigated human calcitonin (NH_2-CGNLSTCMLGTYTQDFNKFHTFPQTAIGVGAP-$CONH_2$) (hCT)[27] and salmon calcitonin (NH_2-CSNL-STCVLGKLSQELHKLQTYPRTNTGSGTP-$CONH_2$) (sCT),[28] in which the region from amino acid number 8 to 22 is the amphipathically α-helical structure,[29] and that from 24 to 32 binds the calcitonin receptor.[30] The cross-spectra of hCT and sCT under three kinds of conditions are shown in Fig. 19, Fig. 20, and Fig. 21, in which the mRNA sequences correspond to their naturally occurring, and the highest and lowest frequency codons in *H. sapiens* and *O. keta*, respectively. Two prominent peaks (f=0.0313 and 0.3438) in Fig. 19 were diminished in both Fig. 20 and Fig. 21, indicating that a significant periodicity of the cognate amino acid sequence depends on that of the naturally occurring mRNA sequence. Moreover, the existence of at least two prominent

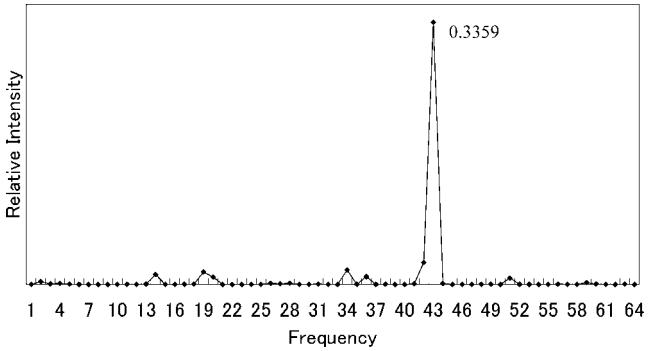

Fig. 21. Multiplication of the Desired Cross-Spectrum of hCT and for sCT Using the Codons with the Lowest Occurring Frequencies

See also the caption of Fig. 19.

peaks was indicated in the cross-spectrum using the naturally occurring mRNA sequence of CTs, though it is presently unknown which the prominent peak is. Several such prominent peaks were observed in the cross-spectrum of tumor necrosis factor-α (TNF-α) and TNF-β, but one peak in epidermal growth factor (EGF) and transforming growth factor α (TGF-α), nerve growth factor (NGF), and neurotrophin 3 (NT3) or interleukin-1α (IL-1α) and IL-1β (data not shown). Therefore more detailed investigation will be required to specify (or extract) the prominent peak.

(ii) Electronegativity and Hardness Scale Parameterization In the course of investigating the relationship between the prominent peaks in DFT analysis and the functional regions on two kinds of CTs using a binary scale, no desired cross-spectrum could be obtained in DFT analysis for the receptor-binding site of sCT, namely, RTNTGSGTP mentioned above, all the amino acids of which belong to group 9. Based on this information, we were interested in replacing a binary scale with any other physicochemical ones such as the EIIP scale, hydrophobic scale, charge scale, or absolute electronegativity and absolute hardness scale as global parameters of nucleotide or amino acid.

The calculation method of the EIIP value[9,10] is pursued to give the mean quasivalence number Z^*[31] for each nucleotide and amino acid side chain.

$$Z^* = \sum_{i=1}^{\text{all the atom}} N_i Z_i \Big/ \sum_{i=1}^{\text{all the atom}} N_i \tag{5}$$

where N_i and Z_i denote the number of the i-th atom and its valence number, respectively. The Z^* scale was first invented to approximate the pseudopotential felt by a conduction electron in bulk metal,[32] while some found that the relation between the Z^* parameterization scale and biological activities was questionable.[33—35] Our separate check also concluded that none of the EIIP, Z^* (parameters not shown), hydrophobic[15] and charge scale (parameters not shown)[36] was appropriate as a parameterization for this purpose in the sense of the similarity of various spectra derived from a binary scale (data not shown).

We considered that the original EIIP scale can be substituted with the absolute electronegativity χ and/or hardness η developed in theoretical chemistry. To do this, the atomic χ values of Pauling,[37] Mulliken,[37] and Pearson[38] and the atomic η values of Pearson[38] were employed. These values were plugged into Eq. 5 as a substitute for Z_i and used as the absolute electronegativity scale or hardness scale. Note that the average operation adopted in Eq. 5 is based on the postulate of geometric mean introduced originally by Sanderson[39—41] and has the same ability of the virtual crystal approximation to represent the average potential in the random-potential medium.[42]

Table 1. Various Scales for Nucleic Acids (na) and Amino Acids (aa)[a]

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
na							
u	0.0561	0.3	2.7055	2.9506	6.9364	5.9918	4.7941
c	0.134	0.35	2.63	2.9487	6.9367	6.1242	4.8176
a	0.126	0.53	2.625	2.929	6.8979	6.205	4.4626
g	0.806	1.1	2.6793	2.9481	6.9407	6.1967	4.6002
t	0.1335	—	2.6221	2.9461	6.9236	5.9836	4.6732
aa							
L	0	3.29	2.3077	2.9396	6.9	5.99	4.9177
I	0	3.64	2.3077	2.9396	6.9	5.99	4.8731
N	0.0036	16.14	2.5475	2.9845	7.0125	6.1288	4.6605
G	0.005	14.79	2.2	3.059	7.18	6.43	4.7394
V	0.0057	7.5	2.305	2.9426	6.907	6.001	4.767
E	0.0058	14.79	2.553	2.9738	6.979	5.931	4.8802
P	0.0198	7.57	2.3167	2.9297	6.8767	5.9533	4.8037
H	0.0242	12.79	2.48	2.9223	6.8709	6.0555	4.4859
K	0.0371	16.21	2.3493	2.9571	6.9453	6.102	4.7587
A	0.0373	12.07	2.2875	2.962	6.9525	6.0725	4.7847
Y	0.0516	4.57	2.446	2.8883	6.7793	5.7393	4.6832
W	0.0548	2.57	2.4217	2.8663	6.7317	5.7594	4.3893
Q	0.0761	14.36	2.4845	2.9695	6.9755	6.0809	4.7937
M	0.0823	6.57	2.33	2.9161	6.8445	5.8318	4.9227
S	0.0829	14.93	2.518	3.0126	7.07	6.074	4.8508
C	0.0829	8.29	2.346	2.8998	6.808	5.686	4.9939
T	0.0941	13.64	2.4425	2.9815	6.9975	6.0288	4.8283
F	0.0946	2.64	2.375	2.865	6.725	5.715	4.9616
R	0.0959	15.93	2.4306	2.9719	6.9871	6.2347	4.5454
D	0.1263	16.29	2.6543	2.9927	7.0229	5.9214	4.9197

a) (a) EIIP scale from ref. 10. (b) Hydrophobic scale from ref. 15. (c) Pauling electronegativity scale from ref. 37. (d) Mulliken electronegativity scale from ref. 37. (e) Pearson electronegativity scale from ref. 38. (f) Pearson hardness scale from ref. 38. (g) The Absolute electronegativity in aqueous conditions.

In addition, we directly obtained molecular χ and η of four bases and 20 amino acids in their neutral forms under the aqueous condition (dielectric constant, 78.5) by the molecular orbital (MO) calculation using the relationships,[43]

$$\chi = -(E_{\text{HOMO}} + E_{\text{LUMO}})/2 \tag{6}$$

$$\eta = (E_{\text{LUMO}} - E_{\text{HOMO}})/2 \tag{7}$$

where E_{HOMO} and E_{LUMO} denote the orbital energies of the highest occupied MO (HOMO) and the lowest unoccupied MO (LUMO), respectively. The semiempirical MO calculation at the PM3 level[44,45] was performed for this purpose using the MOPAC package.[46] All of the parameters obtained in these scales are listed in Table 1.

The desired cross-spectra based on Mulliken's and Pearson's χ were fairly similar, but not the same as, to that from the binary scale. Note that these two kinds of χ values for atoms are essentially the same, because the ratio of Mulliken's χ for atom to Pearson's χ is almost constant[37,38] (or see (d) and (e) in Table 1), while their energy units are different. In addition, both Pearson's χ and η gave the same desired cross-spectra. Although Pauling's χ values for atoms are widely accepted, they showed lesser similarity to the binary scale than Mulliken's χ, based on the relative intensity

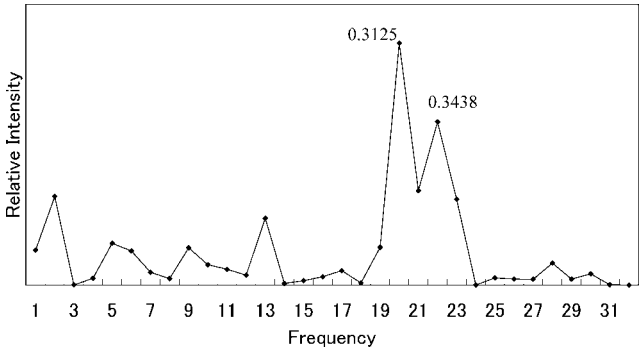

Fig. 22. The Cross-Spectrum of mRNA of Prp106—126 Using Mulliken's Electronegativity

See also the caption of Fig. 1a.

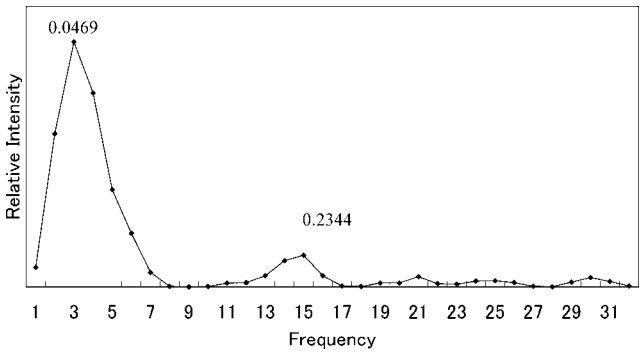

Fig. 23. The Cross-Spectrum of the Amino Acid Sequence of Prp106—126 Using Mulliken's Electronegativity

See also the caption of Fig. 1a.

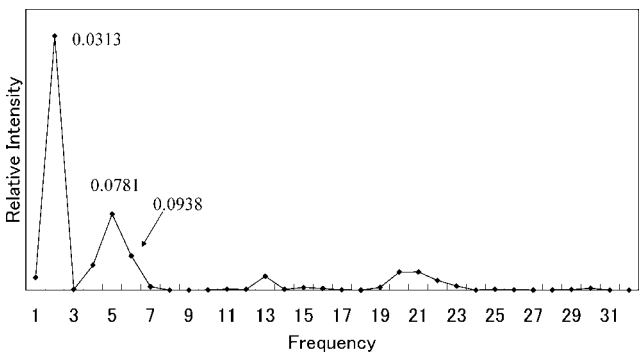

Fig. 24. The Desired Cross-Spectrum from mRNA and the Amino Acid Sequence of Prp106—126 Using Mulliken's Electronegativity, Respectively

See also the caption of Fig. 1a.

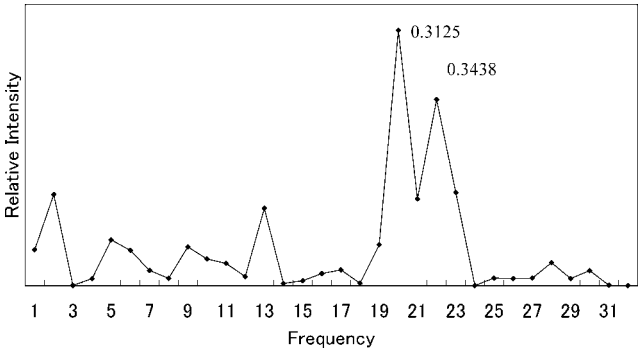

Fig. 25. The Cross-Spectrum of the mRNA Sequence of Prp106—126 Using the Electronegativity Value for a Molecule

See also the caption of Fig. 1a.

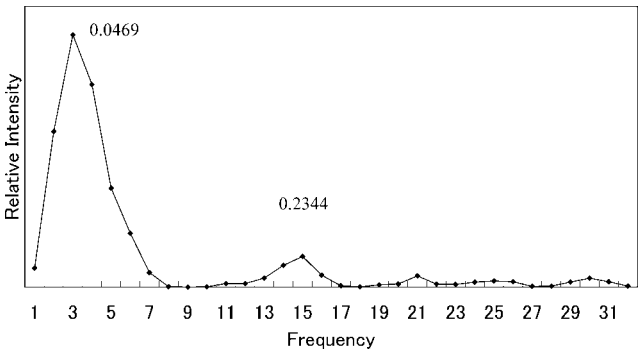

Fig. 26. The Cross-Spectrum of the Amino Acid Sequence of Prp106—126 Using the Electronegativity Value

See also the caption of Fig. 1a.

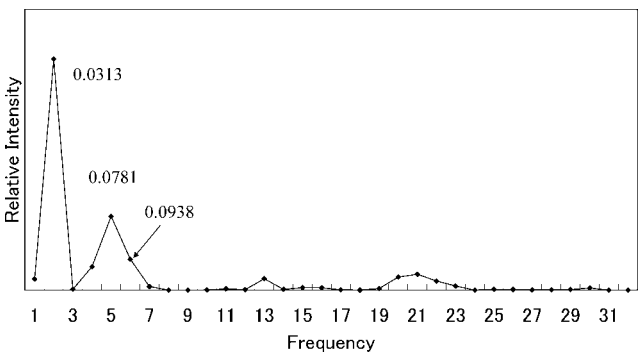

Fig. 27. The Desired Cross-Spectrum from mRNA and the Amino Acid Sequence of Prp106—126 Using the Electronegativity Value for a Molecule, Respectively

See also the caption of Fig. 1a.

of various peaks in the spectrum.

The desired cross-spectrum of Prp106—126 under the condition of Mulliken's χ is shown in Fig. 24. The cross-spectra of the nucleotide sequence and of the amino acid sequence are shown in Fig. 22 and Fig. 23, respectively. All the spectra (Figs. 22—24) were similar to those (Figs. 2, 4, and 7) derived from the binary scale, respectively. In addition, a similar tendency was supported by the semiempirical MO calculation for four bases and 20 amino acids (see Figs. 25—27), while the relative intensity of various peaks was not strictly in accord with the binary, Pauling's, Mulliken's, or Pearson's, parameterization. Based on these, the absolute electronegativity scale is a plausible substitute for the binary scale, although more detailed investigation is required to determine finally which scale is superior for elucidating the same biological activities of proteins.

On the other hand, under the condition of Mulliken's χ the desired cross-spectrum of the receptor-binding region from 24 to 32 (*i.e.* RTNTGSGTP) of sCT could be successfully featured (Fig. 28), unlike the case of the binary scale. The cross-spectrum of the receptor-binding region (*i.e.*, QTAIGVGAP) of hCT and that (*i.e.*, RTNTGSGTP) of sCT is shown in Fig. 29, indicating that one (f=0.3438) of two prominent peaks observed in Fig. 19 is a candidate for the receptor binding region. However, the amino acid sequences

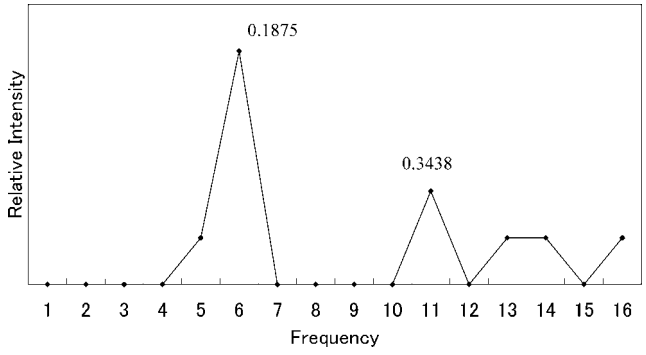

Fig. 28. The Desired Cross-Spectrum of the Amino Acid Sequence (24—32) of sCT Using Mulliken's Electronegativity

The abscissa (m) represents frequencies from 0.0000 to 0.5000 (signifying $2^4/2$) and the ordinate relative intensities (amplitudes) in the spectrum.

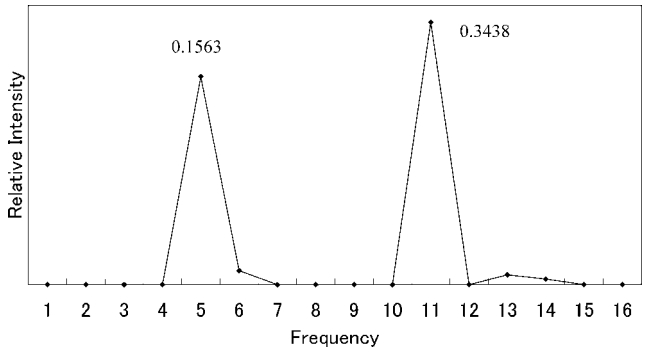

Fig. 29. Multiplication of the Desired Cross-Spectrum of the Amino Acid Sequence (24—32) of hCT and of sCT Using Mulliken's Electronegativity

See also the caption Fig. 28.

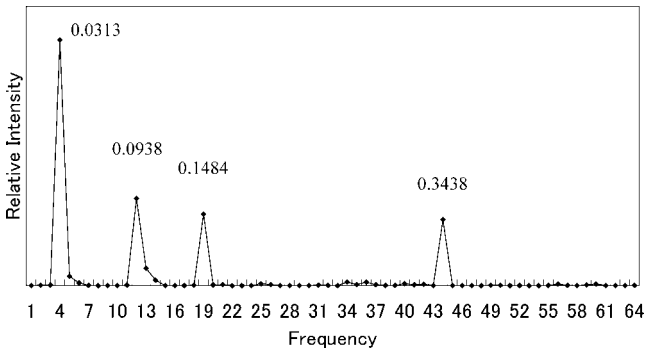

Fig. 30. Multiplication of the Desired Cross-Spectrum of the Amino Acid Sequence (8—32) of hCT and of sCT Using Mulliken's Electronegativity

See also the caption Fig. 19.

from 8 to 32 of the two CTs were involved in another prominent peak (f=0.0313) (Fig. 30). Especially, the C-terminal region of hCT was necessary for its peak (data not shown). In addition, the peak was fairly prominent in hCT (active form) composed of 32aa as well as in procalcitonin (inactive form) composed of 116aa (data not shown). Thus one prominent peak relevant to the receptor binding could not be discriminated.

If a specific interaction between ligand and receptor can be regarded as something like a complex with the same biological activity (or function) in a broad sense, the prominent peak (or peaks) involved in the specific intermolecular interaction may be extracted. Based on this working hypothesis, we are now in the process of constructing a selection rule that can elucidate the ligand–protein interaction, such as prion and 37-kD laminin receptor, hCT and hCT receptor, TNF-α and 55-kD TNF receptor, EGF and EGF receptor, IL-2 and IL-2α receptor, various viruses and their receptors, *etc.* Such a selection rule will be discussed elsewhere, although its theoretical base is as yet unsolved.

Conclusions

We demonstrated DFT analysis to elucidate the same biological activity of Prp106—126, MSI-78A and oxaldie 1 using the mRNA and amino acid sequences. As a result, the prominent peak occurring in the cross-spectrum derived from two of three desired cross-spectra was the same when a binary scale was used as a parameter of nucleotide or amino acid. However, the prominent peak was not always single, like the case of CTs or TNFs. Otherwise, almost the same result could be reproduced when the average absolute elctronegativity scale (or the absolute hardness scale) was used instead of a binary scale. Based on these, various proteins with the same biological activities may be rationally elucidated from both the naturally occurring mRNA and the cognate amino acid sequence, using either a binary scale or the mean absolute electronegativitiy scale (or the mean hardness scale) or both, although more detailed investigation will be required to specify the prominent peak.

Acknowledgments One of the authors (NN) is grateful to Professor T. Fukui at Hoshi University for providing information on the Reference 14. The MO calculation was performed at the Academic Center for Computing and Media Studies, Kyoto University.

References and Notes

1) Numao N., Noguchi N., Eguchi Y., Watanabe S., Fukui T., Kamino T., Shimozono N., Yamazaki A., Kobayashi S., Sasatsu M., *Biol. Pharm. Bull.*, **26**, 229—232 (2003).
2) Numao N., Hirota Y., Iwahori A., Kidokoro S., Sasatsu M., Kondo I., Itoh S., Itoh E., Katoh T., Shimozono N., Yamazaki A., Takao K., Kobayashi S., *Biol. Pharm. Bull.*, **22**, 73—76 (1999).
3) Forioni G., Angeretti N., Chiesa R., Monzani E., Salmona M., Bugiani O., Tagliavini F., *Nature* (London), **362**, 543—546 (1993).
4) Zahn R., Liu A., Luhrs T., Riek R., von Schroetter C., Garcia F. L., Billeter M., Calzolai L., Wider G., Wuthrich K., *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 145—150 (2000).
5) Huang Z., Gabriel J.-M., Baldwin M. A., Fletterick R. J., Prusiner S. B., Cohen F. E., *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 7139—7143 (1994).
6) Iwahori A., Hirota Y., Sampe R., Miyano S., Numao N., *Biol. Pharm. Bull.*, **20**, 267—270 (1997).
7) Zasloff M., *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 5449—5453 (1987).
8) Johnsson K., Allemann R. K., Widmer H., Benner S. A., *Nature* (London), **365**, 530—532 (1993).
9) Veljkovic V., Cosic I., Dimitrijevic B., Lalovic D., *IEEE Trans. Biomed. Eng.*, **BME-32**, 337—341 (1985).
10) Cosic I., *IEEE Trans. Biomed. Eng.*, **41**, 1101—1114 (1993).
11) Eisenberg D., Weiss R. M., Terwilliger T. C., *Faraday Symp. Chem. Ser.*, **17**, 109—120 (1982).
12) Tiwari S., Ramachandran S. Bhattacharya A., Bhattacharya S., Ramaswarmy R., *CABIOS*, **13**, 263—270 (1997).
13) Kutuzova G. I., Frank G. K., Epipova N. G., Makeev V. Yu., Polozov R.V., *Biophysics*, **44**, 218—225 (1999).
14) Fukui K., "Proceedings of the 5th Institute for Fundamental Chemistry Symposium," May 26, 1989 in Japanese.
15) Lacey J. C., Mullins D. W., *Orig. Life*, **13**, 3—42 (1983).
16) Lacey J. C., *Chemtracts-Biochem. Mol. Bio.*, **12**, 398—418 (1999) and references therein.
17) Hobish M. K., Wickramasinghe N. S. M. D., Ponnamperuma C., *Adv. Space Res.*, **15**, 365—382 (1995).

18) Shimizu M., *J. Phys. Soc. Jpn.*, **56**, 43—35 (1987).
19) Gilbert W., *Nature* (London), **319**, 618 (1986) and references therein.
20) Guerrier-Takada C., Gardiner K., Marsh T., Pace N., Altman S., *Cell*, **35**, 849—857 (1983).
21) Kruger K., Grabowski P. J., Zaug A. J., Sands J., Gottschiling D. E., Cech T. R., *Cell*, **31**, 147—157 (1982).
22) Yarus M., *Curr. Opin. Chem. Biol.*, **3**, 260—267 (1999) and references therein.
23) Numao N., Kidokoro S., *Biol. Pharm. Bull.*, **16**, 1160—1163 (1993).
24) Rabiner L., Gold B., "Theory and Application on Digital Signal Processing," Prentice-Hall, Englewood Cliffs, NJ, 1975, pp. 50—59.
25) Cosic I., Nesic D., *Eur. J. Biochem.*, **170**, 247—252 (1987).
26) Kretzschmar H. A., Stowring L. E., Westaway D., Stubblebine W. H., Prusiner S. B., Dearmond S. J., *DNA*, **5**, 315—324 (1986).
27) Craig R. K., Riley J. H., Edbrooke M. R., Broad P. M., Foord S. M., Al-Kazwini S. J., Holman J. J., Marshall I. I., *Biochem. Soc. Symp.*, **52**, 91—105 (1986).
28) Poschl E., Lindley I., Hofer E., Seifert J. M., Brunowsky W., Besemer J., *FEBS Lett.*, **226**, 96—100 (1987).
29) Green F. R., Lynch B., Kaiser E. T., *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 8340—8344 (1987).
30) Inoue A., Shikano M., Komatsu Y., Obata J., Ochiai J., Nishide H., Ito N., Nagao H., Kondo K., Tunemoto D., Hemmi H., Numao N., *Eur. J. Biochem.*, **201**, 607—614 (1991).
31) Veljkovic V., Lalovic D. I., *Experientia*, **33**, 1228—1229 (1977).
32) Veljkovic V., Slavic I., *Phys. Rev. Lett.*, **29**, 105—107 (1972).
33) Herrmann E. Ch., *Experientia*, **35**, 1263—1264 (1979).
34) Rosenblatt D. H., Dacre J. C., *Experientia*, **35**, 567—568(1979).
35) Barness W. S., Levin D. E., *Experientia*, **35**, 565—567 (1979).
36) Fukazawa Y., Fujii H., Yoshioka K., unpublished results.
37) Allen L. C., *J. Am. Chem. Soc.*, **111**, 9003—9014 (1989).
38) Pearson R., *Inorg. Chem.*, **27**, 734—740 (1988).
39) Sanderson R. T., *Science*, **114**, 670—672 (1951).
40) Datta D., *J. Phys. Chem.*, **90**, 4216—4217 (1986).
41) Yang W., Lee C., Ghosh S. K., *J. Phys. Chem.*, **89**, 5412—5414 (1985).
42) Tanaka K., Tsu R., *Phys. Rev.*, **B24**, 2038—2050 (1981).
43) Parr R. G., Pearson R. G., *J. Am. Chem. Soc.*, **105**, 7512—7516 (1983).
44) Stewart J. J. P., *J. Comp. Chem.*, **10**, 209—220 (1989).
45) Stewart J. J. P., *J. Comp. Chem.*, **10**, 221—264 (1989).
46) MOPAC2000 ver.1.3, Fujitsu Ltd., Tokyo, Japan, 2000.