Sequence Fourier Analysis of a Specific Protein–DNA (RNA) Interaction; an Intermolecular Frequency Symmetry

Naganori NUMAO*[,a,1)] and Yoshihito KAMIMOTO[b]

[a] *BioFrontier Institute Inc.; 5–4–21 Nishi-Hashimoto, Sagamihara, Kanagawa 229–1131, Japan: and* [b] *Kanagawa Industrial Technology Center; 705–1 Shimoimaizumi, Ebina, Kanagawa 243–9435, Japan.*

Received May 22, 2009; accepted August 19, 2009; published online August 21, 2009

Here, we first report that a specific protein–DNA (RNA) interaction between p53 protein and divergent I kappa B kinase interacting protein (IKIP)-Apaf1 DNA promoter region, and between poliovirus 3CD protein and the 5′ terminal region of the RNA sequence can be successfully elucidated with a sequence Fourier analysis, alike various specific protein–protein interactions described previously. Based on these, the intermolecular frequency symmetry would be evolutionarily conserved in a specific interaction of the bioactive structure of various long-chain sequences. In addition, the symmetry is fairly sensitive to a substitution (or a point mutation) on the sequence.

Key words symmetry; Fourier analysis; transcription; p53; replication; poliovirus

In the previous study,[2—5)] we could effectively elucidate various specific protein–protein interactions[6)] from both the naturally occurring amino acid (aa) sequence and the corresponding RNA (na) one using the sequence Fourier analysis (SFA), on which one of three kinds scales, that is, the Mulliken's absolute electronegativity (M) scale,[4,5)] the Lacey's relative hydropathy (H) scale[2,3)] and the Garel's relative hydropathy (G) scale[7—9)] (Table 1), was assigned as a parameterization. Both the calculation process[5)] and the criteria[4)] to elucidate the specific interaction had been already reported. In addition, the intermolecular frequency symmetry (IFS; see ref. 9) was indicated to be fairly sensitive to a substitution on the aa sequence.[10—12)]

Meanwhile, as a first trial we investigated a specific interaction between wild type p53 protein, a transcription factor, and the divergent IKIP-Apaf1 promoter DNA sequence.[13—15)] Because the p53 had been already reported to act downstream of the mitogen-activated protein (MAP) cascade,[17,18)] similar to ELK1 reported previously.[2)] To elucidate its specific interaction, three kinds[19)] of frame-shifted virtual aa sequence (170aa in length), as a group based on the genetic code, are separately translated from the double-strand DNA sequence under the condition of a definite window [*i.e.*, 510 base pair (bp) in length].[20)] Of numerously virtual aa sequences, at least two kinds of p53-binding regions (*i.e.*, *I* and *II*) could be finally found on the divergent promoter sequence. One region *I* is a virtual aa sequence (170aa), translated from 91147 (5′ terminus) to 91656 (3′ terminus).[19)] Under the condition of the H scale only (but neither the M nor the G one), one resonant peak (see ref. 9) (f=0.2207) of the desired cross-spectrum[5)] (Fig. 1a) derived from a virtual aa of the region *I* could be overlapped with one resonant peak (f_1=0.2227) of two characteristic ones (see ref. 9) (f_1=0.2227, f_2=0.2856) of the desired cross-spectrum (Fig.

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

na	1)	2)	3)
u	2.9506	0.69	0.3
c	2.9487	0.62	0.35
a	2.929	0.26	1.1
g	2.9481	0.44	0.53
t	2.9461	—	0.94

aa	1)	2)	3)
L	2.9396	3.29	16.71
I	2.9396	3.64	16.36
N	2.9845	16.14	3.86
G	3.059	14.79	5.21
V	2.9426	7.5	12.5
E	2.9738	14.64	5.36
P	2.9297	7.57	12.43
H	2.9223	12.79	7.21
K	2.9571	16.21	3.79
A	2.962	12.07	7.93
Y	2.8883	4.57	15.43
W	2.8663	2.57	17.43
Q	2.9695	14.36	5.64
M	2.9161	6.57	13.43
S	3.0126	14.93	5.07
C	2.8998	8.29	11.71
T	2.9815	13.64	6.36
F	2.865	2.64	17.36
R	2.9719	15.93	4.07
D	2.9927	16.29	3.71

1) M scale; Mulliken's absolute electronegativity scale. 2) H scale; Lacey's relative hydropathy scale. 3) G scale; Garel's relative hdropathy na scale and the Lacey's reversed aa one.

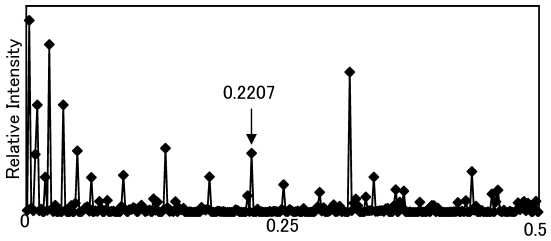

Fig. 1a. The Desired Cross-Spectrum of a Virtual Amino Acid Sequence, Tentatively Translated from the Region *I* (91147 to 91656) on the Same as the Apaf1 Promoter DNA Sequence, under the H Scale

The abscissa represents frequencies from 0.0000 to 0.5000 and the ordinate relative intensities (amplitudes) in the spectrum throughout all the figure captions in this communication. The number indicated in the figure is the resonant frequency value.

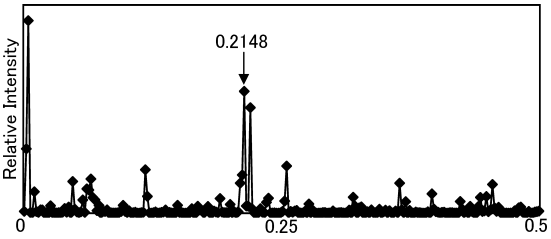

Fig. 1b. The Desired Cross-Spectrum of a Virtual Amino Acid Sequence, Tentatively Translated from the Region *II* (91439 to 90984) on the Opposite Strand to the Apaf1 Promoter DNA One, under the H Scale

See also the caption of Fig. 1a.

* To whom correspondence should be addressed. e-mail: kamimoto@kanagawa-iri.go.jp

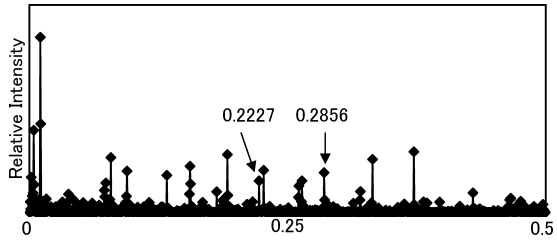

Fig. 2. The Desired Cross-Spectrum of the Wild Type Amino Acid Sequence of p53 Protein under the H Scale

See also the caption of Fig. 1a.

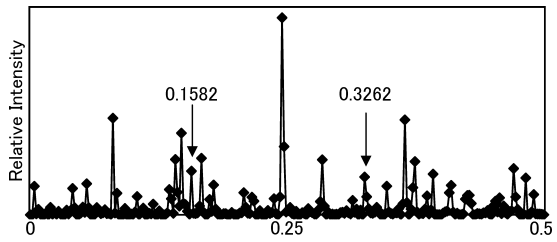

Fig. 3. The Desired Cross-Spectrum of a Virtual Amino Acid Sequence, Tentatively Translated from the Region *III* (53 to 562) on the Poliovirus RNA Sequence, under the M Scale

See also the caption of Fig. 1a.

2) from the wild type p53 (393aa).[9] In this, note that other higher peaks than the resonant (characteristic) peak(s) can be eliminated with the criteria.[4] Such the operation was performed in all other samples shown in this study. The other p53-binding region (*II*) is derived from 91493 (5′ terminus) to 90984 (3′ terminus) (note the numbering direction) on the opposite DNA strand.[19] One resonant peak (f=0.2148) (Fig. 1b) of the region *II* could be overlapped with one resonant peak (f_1=0.2227) (Fig. 2) of two characteristic ones (f_1=0.2227, f_2=0.2856) of the p53 under the H scale only. Interestingly, no specific interaction between the mutant $p53_{V173L}$, as a typical sample,[13] and any virtual aa sequence described above could be observed.

Taking this with two other results,[20] next we had a great interest in investigating a difference in the relationship between of 5′ cloverleaf RNA sequence (virulent type)[21] and of the mutant (a480g) one (attenuated type)[22] of poliovirus 1 genomic RNA (*Picornaviridae*) to the binding poliovirus protein 3CD (644aa; *Picornaviridae*).[23—25] The viral RNA sequence is composed of a single-strand *positive*-polarity and 7441-base in length. The 3CD is directly involved in the poliovirus replication. As a result, it could be elucidated with the M scale only that one resonant peak (f_2=0.3262) of two characteristic ones (f_1=0.1582, f_2=0.3262) (Fig. 3) derived from the 5′ (but not the 3′) virtual amino acid (170aa) sequence, translated from 53 to 562 (region *III*)[19] (NC002058), could be overlapped with one resonant peak (f=0.3291) (Fig. 4) from the desired cross-spectrum derived from the 3CD. Otherwise, no specific interaction between the mutant (a480g) sequence, examined as a typical sample of a single nucleotide polymorphism (SNP) model, and the 3CD could be observed with any scale.

Although test samples exemplified here may not be sufficient to discuss the generality, the IFS is conserved as a necessary condition[26] in a specific protein–DNA (RNA) interaction, alike a specific protein–protein interaction. Thus, it indicates us the existence of something like a Noether's theorem[27] in life science, that is, there is a conservation rule when the symmetry can be found (*vice versa*). In addition, the H scale and the M one are used as a parameter in the initial interaction of transcription and replication process, respectively. A palindrome sequence (or complementary unit)[28] on/near the 5′ (or 3′) termini of the DNA (RNA) sequence (*ex*. region *I*, *II*, *III*, *etc.*) seems to play a role in the process.

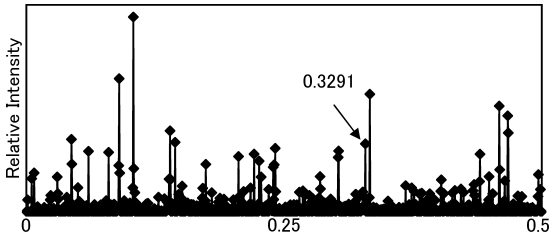

Fig. 4. The Desired Cross-Spectrum of the Amino Acid Sequence of 3CD Protein under the M Scale

See also the caption of Fig. 1a.

Acknowledgements We (N.N.) appreciate very much Y. Asai, T. Nishi, and F. Arai for their support of this research, and Drs. M. Kohara and A. Nomoto for information of poliovirus.

References and Notes

1) Present address: *5–2–17 Hara-machida, Machida, Tokyo 194–0013, Japan*.
2) Numao N., Miyama M., Hanagata N., *Chem. Pharm. Bull.*, **52**, 1026—1028 (2004).
3) Numao N., Fujii H., Fukazawa Y., Hanagata N., Tanaka K., *Chem. Pharm. Bull.*, **52**, 377—379 (2003).
4) Numao N., Fujii H., Fukazawa Y., Tanaka K., *Chem. Pharm. Bull.*, **51**, 1217—1219 (2003).
5) Numao N., Fujii H., Fukazawa Y., Yoshioka K., Okada M., Tanaka K., *Chem. Pharm. Bull.*, **51**, 550—559 (2003) and references therein.
6) The sequences are of *Homo sapiens* used, unless otherwise stated. Various proteins or nucleic acid sequences examined in this study are of the following locus number in NCBI database. [mouse ELK1 (NM007922)/mouse ERK1 (BC029712)], [HIVgp120/CD4], [ELK1 (NM005229)/ERK1 (HSERK1)], [E2F1/pRb], [P53 (HSP53), E2F1/p38α, p38γ (HSU66243), divergent IKIP-Apaf1promoter (AC01328)], [GAL4 (YSCGAL4)/divergent GAL1-10 promoter (YSCGAL)], [T antigen, t antigen/SV40 replication region (NC001669)], [VP1, VP3, 2A, 3C, 3CD/Poliovirus 1 (Mahoney) replication region (NC002058)], [*Eco*RI, DnaA/pBR322], [tobacco ringspot virus], [calcitonin/calcitonin receptor (452aa, 468aa)], *etc.*
7) Garel J. P., Filliol D., Mandel P., *J. Chromatogr.*, **78**, 381—391 (1973).
8) Note that the experimental value (1.10) of adenine (a) only cannot be directly estimated from that (0.26) of the H one, compared with other 3 (u, g, c) nucleotides.
9) In the preceding study,[2] a specific interaction between mouse[6] ELK1 (429aa; f=0.2266) and mouse ERK1 (379aa; f_1=0.2261, f_2=0.2642) could be elucidated with the G scale only (Figs. not shown). In this, the parenthesized numbers signify the total number of the mature amino acids (aa) calculated. The f, f_1 and f_2 signify a characteristic peak (or value) with the SFA, and that the former two (f,f_1) peaks are especially named a resonant peak, respectively. The IFS means that the f, f_1 and f_2 satisfy a definition condition (f=f_1±0.005, 0.5−f=f_2±0.012, f_1+f_2=0.5±0.0176). Further, if the value exists between 0.1660 and 0.1680, between 0.3320 and 0.3340, or between 0.2480 and 0.2520, all such peaks are neglected. A typographical number (f±0.006) in the step 2 in the criteria reported in ref. 4 is corrected with the value (f±0.005).
10) Differing from a wild type ELK1,[2] no specific interaction between a mutant ELK1 (S383A or S389A)[11] and ERK1 could be observed.
11) Cruzalegui F. H., Cano E., Treisman R., *Oncogene*, **18**, 7948—7957

(1999) and references therein.
12) The difference in the interaction between of a wild type protein (*ex.*, p53, prion, APP695) and of the corresponding mutants to the binding partner (*ex.*, p38γ, proteinase K, BACE1B) could be fairly elucidated, respectively.
13) Hofer-Warbinek R., Schmid J. A., Mayer H., Winsauer G., Orel L., Mueller B., Wiesner Ch., Binder B. R., de Martin R., *Cell Death Differ.*, **11**, 1317—1325 (2004) and references therein.
14) Moloni M. C., Hickman E. S., Denchi E. L., Caprara G., Colli E., Cecconi F., Muller H., Helin K., *Nat. Cell Biol.*, **3**, 552—558 (2001).
15) The DNA sequence was boldly regarded as the virtual RNA, because the H scale of t (Table 1) was unknown.[16)]
16) Lacey J. C., Mullins D. W., *Orig. Life*, **13**, 3—42 (1983).
17) Wu G. S., *Cancer Biol. Ther.*, **3**, 156—161 (2004).
18) In addition, a specific interaction between the wild type p53 (393aa; f=0.2856) and p38γ (367aa; f=0.2227, 0.2944) could be elucidated with the H scale only.
19) Three virtual frame-shifted aa sequences are translated from (91146-91655, 91147-91656, 91148-91657) on the same strand of a double-strand Apaf1 promoter DNA region (AC01328) and from (91492-90983, 91493-90984, 91494-90985) on the opposite strand of the promoter (*i.e.*, the same IKIP promoter one) (AC01328), respectively. Similarly, three virtual aa sequences were translated from a single-strand poliovirus genomic RNA sequence (52-561, 53-562, 54-563) (NC002058), respectively. In this, note that the corresponding negative strand was simultaneously translated to verify this analytical method.
20) Using 510 bp only (but neither 255 bp nor 1020 bp), a specific interaction between the wild type GAL4 protein (*S. cerevisiae*) and the divergent GAL1-10 promoter DNA sequence (*S. cerevisiae*) could be rationally elucidated with the H scale only (data not shown). Also, another specific interaction between the T antigen (*polyomaviridae*) and the origin region (*i.e.*, DNA sequence) of simian virus 40 (*polyomaviridae*) could be done with the M scale only (data not shown).
21) Murray K. E., Barton D. J. *J. Virol.*, **77**, 4739—4750 (2003) and references therein.
22) Kawamura N., Kohara M., Abe S., Komatsu T., Tago K., Arita M., Nomoto A., *J. Virol.*, **63**, 1302—1309 (1989).
23) Takahara Y., Ando N., Kohara M., Hagino-Yamasaki K., Nomoto A., Itoh H., Numao N., Kondo K., *Gene*, **79**, 249—258 (1989).
24) No specific interaction between the 3CD and the capsid protein P1 could be observed, differing from the 3C.
25) Brenner S., Based on an abstract presented at the Yakult International Symposium, Yakult Hall, Tokyo, on 13 May 1988, pp. 39—46.
26) 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) and references therein.
27) Kobayashi S., "Symmetry and the Beautiful Universe," ed. by Lederman L. M., Hill C. T., Hakuyosha, Tokyo, 2008, p. 123.
28) Numao N., Kidokoro S., *Biol. Pharm. Bull.*, **16**, 1160—1163 (1993).