

A New Application of Sequence Fourier Analysis to Specific Ligand–Protein Interactions

Naganori NUMAO,^{*,a} Hisashi FUJII,^b
Yoshiyuki FUKAZAWA,^b and Kazuyoshi TANAKA^c

^aBioFrontier Institute Inc.; 5–4–21 Nishi-Hashimoto, Sagami-hara, Kanagawa 229–1131, Japan; ^bKanagawa Industrial Technology Research Institute; 705–1 Shimoimaizumi Ebina, Kanagawa 243–0435, Japan; and ^cDepartment of Molecular Engineering Graduate School of Engineering, Kyoto University; Nishikyo-ku, Kyoto 615–8500, Japan.

Received July 2, 2003; accepted August 4, 2003

A frequency (or sequence)-dependent rule seems to be conserved in a simple and concerted interaction of various human proteins.

Key words bioinformatics; electronegativity; sequence Fourier analysis; calcitonin; poliovirus; tumor necrosis factor- α

We are currently interested in developing a method for effective elucidation of specific ligand-receptor interaction (or signal transduction process) using both the mRNA and the cognate amino acid sequences. To accomplish this, we have investigated various ligand-receptor interactions with the sequence Fourier analysis reported in the preceding papers.^{1,2)}

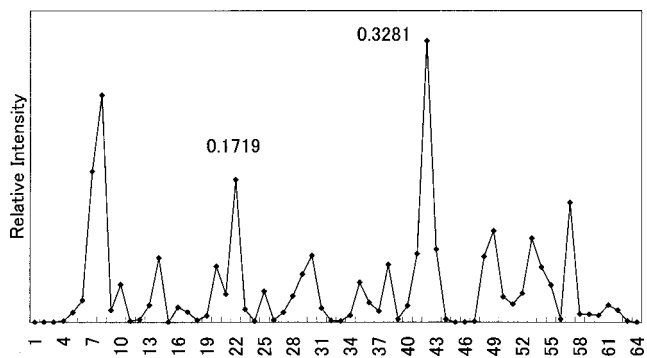


Fig. 1a. The Desired Cross-Spectrum¹⁾ of the Sense Amino Acid Sequence of bACTH

The abscissa represents the 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 frequency value.

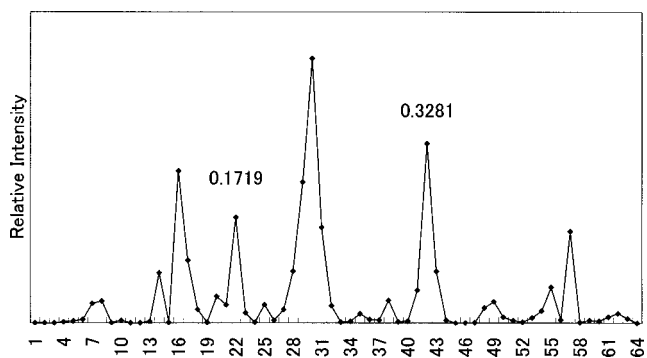


Fig. 1b. The Desired Cross-Spectrum of the Antisense Amino Acid Sequence of bACTH

See also the caption of Fig. 1a.

Of biologically active proteins,³⁾ we first focused on the analytical results (Figs. 1a, b) derived from the naturally occurring (*i.e.*, sense) amino acid sequence of bovine adrenocorticotrophic hormone (ACTH) (bACTH; 24aa),⁴⁾ and from the antisense peptide encoded by a nucleotide sequence complementary to the mRNA for bACTH. Their specific interaction has experimentally been observed by Blalock and his colleagues,⁵⁾ but this work is still controversial. Interestingly, two characteristic peaks ($f=0.1719$ and 0.3281 , where f signifies the frequency value by the sequence Fourier analysis in ref. 1) are observed as major peaks in both the figures.

In the course of investigating various ligand-receptor interactions,⁶⁾ we next happened to notice that two characteristic peaks ($f=0.0313$, 0.4688) in the spectra derived from both the sense- and the antisense-amino acid sequences (Figs. 2a, b, respectively) of human calcitonin (hCT; 32aa) are in close proximity to those ($f=ca. 0.03$, 0.4697) of the receptor (hCTR; 468aa) (Figs. 3a, 3b), of which the existence of four potentially *N*-glycosyl groups on the extracellular region is completely neglected. Two such peaks were also observed in salmon calcitonin I (sCT I; 32aa) (Figs. 4a, b), having fifty times as strong activity as hCT.⁷⁾ Furthermore, only one peak ($f=0.0313$) was observed in hCTR (452aa) deleted 16aa in the putative first intracellular loop of hCTR (468aa). The loop region does not seem to be directly involved in the binding to hCT, however. In addition, the hCTR (468aa) has already been reported to have ten times as strong binding affinity to sCT I as the hCTR (452aa).⁸⁾

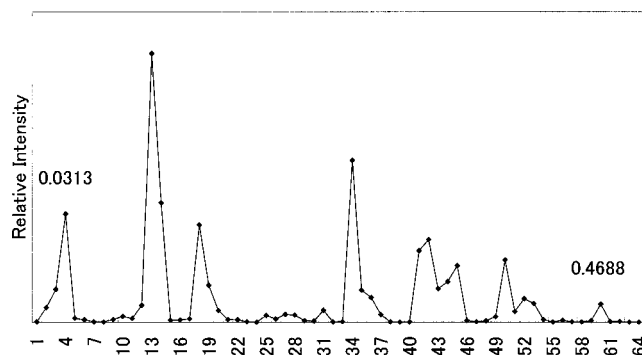


Fig. 2a. The Desired Cross-Spectrum of the Sense Amino Acid Sequence of hCT

See also the caption of Fig. 1a.

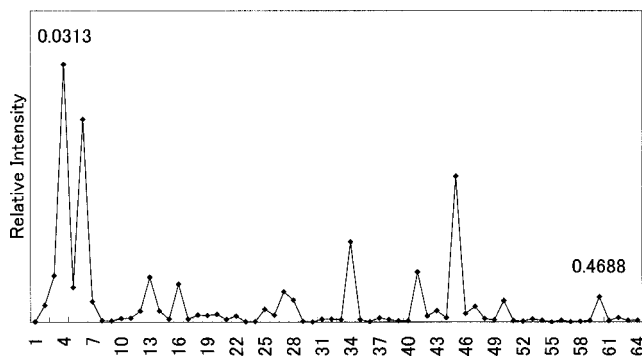


Fig. 2b. The Desired Cross-Spectrum of the Antisense Amino Acid Sequence of hCT

See also the caption of Fig. 1a.

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

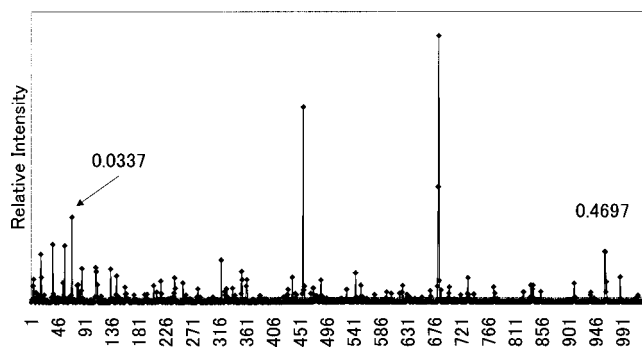


Fig. 3a. The Desired Cross-Spectrum of the Sense Amino Acid Sequence of hCTR

See also the caption of Fig. 1a.

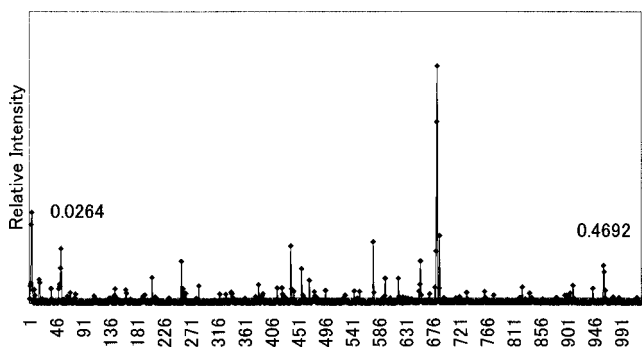


Fig. 3b. The Desired Cross-Spectrum of the Antisense Amino Acid Sequence of hCTR

See also the caption of Fig. 1a.

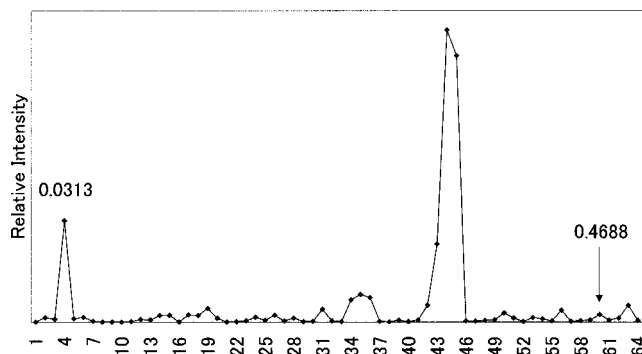


Fig. 4a. The Desired Cross-Spectrum of the sense Amino Acid Sequence of sCT

See also the caption of Fig. 1a.

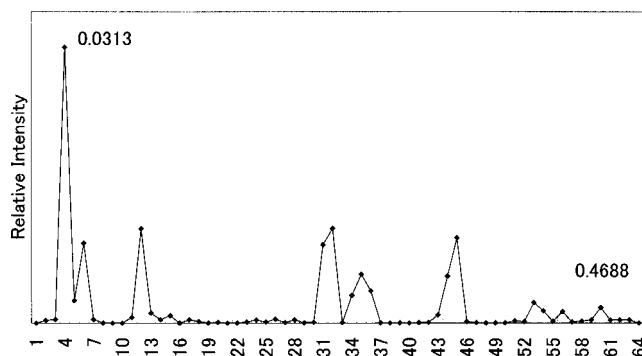


Fig. 4b. The Desired Cross-Spectrum of the Antisense Amino Acid Sequence of sCT

See also the caption of Fig. 1a.

Based on the two examples described above, we tentatively devised a series of the following operations as a hypothetical working procedure, in order to examine the characteristic peak(s) in various specific ligand-receptor interactions^{3,6)} reported previously;

1) The number of peaks selected is tentatively fixed up to the 15 higher ones⁹⁾ derived from the sense- and from the antisense-amino acid of a given protein,¹⁰⁾ respectively.

2) Of the 15 higher peaks from the sense amino acid sequence of a given protein is (are) to be extracted the peak(s) with almost the same f value ($f \pm 0.006$) between the sense- and the antisense-amino acid sequences of a given protein.

3) Of 15 higher peaks from the sense sequence of a given protein is (are) to be extracted the peak(s) almost overlapping ($f \pm 0.012$) with those from the antisense amino acid sequence reflected at the center of the abscissa ($f = 0.2500$).¹¹⁾

4) Continuously, the same procedure as described in 2) or 3) is individually repeated for the concerned characteristic peaks selected from the sense amino acid sequence of a given protein and those from the sense sequence of the specific binding partner, as the application of sense/antisense peptide interactions.¹²⁾

5) Thus, the common peak(s) to specify the interaction between a given protein and the receptor protein is (are) determined from the 15 higher peaks derived from their sense amino acid sequences.

Under the conditions described above, two peaks ($f = ca. 0.03$, $ca. 0.46$) derived from the sense sequence of sCT I or hCTR could be determined from two different directions, that is, from sCT I to hCTR and *vice versa*. In hCT, however, only one direction from hCTR ($f = 0.0337$, 0.4697) to hCT ($f = 0.0313$) is allowed. To assess the potential of this operation, we examined the relationship between CTs (or hCTR) and the various receptors (or ligands).⁶⁾ We found the relationship between CTs and hCTR to be fairly selective, though a similar relationship was observed between one of them and several other proteins including BMPRII or IL2 (data not shown).^{3,6)}

Among various specific ligand-receptor interactions,⁶⁾ we especially note the relationship between VP1 of four coat proteins [VP1 (302aa), VP2 (272aa), VP3 (238aa), VP4 (68aa)]¹³⁾ of poliovirus 1 and the human receptor (CD155; 397aa).¹⁴⁾ The analytical results of VP1 and CD155 are shown in Figs. 5 and 6, respectively. In this case, one ($f = 0.2217$) of the two peaks ($f = 0.2217$, 0.2715) derived from VP1 agrees with that ($f = 0.2305$) from CD155 within the observational error. A similar result was obtained in the VP3/CD155 relationship, but not in VP2 or VP4 (data not shown). Unexpectedly, such a similar result between one of the receptors (*i.e.*, CD155, CD81 or CD120b) and the spike glycoprotein of coronavirus 229E (1158aa) [or SARS (1239aa)] was obtained, while no relation of the latter one with CD13 (967aa) could be successfully elucidated¹⁵⁾ (data not shown).

Furthermore, we describe two exceptional results: i) the relationship between FGF1 (155aa) and the receptor (*i.e.*, FGFR2; 800aa), and ii) that between tumor necrosis factor- α (TNF- α) (157aa) and 75 kd TNFR (CD120b; 439aa). In the former relationship: i) heparan sulfate proteoglycan is directly involved in the ligand-receptor interaction as the third

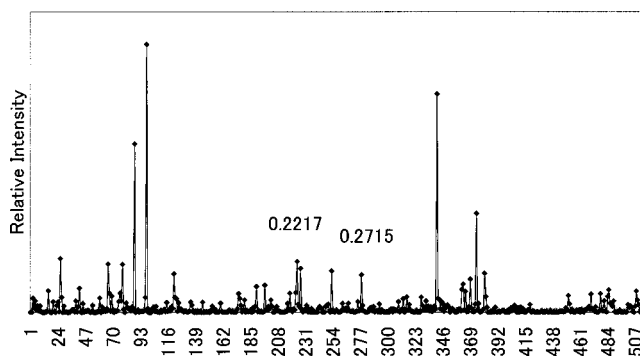


Fig. 5. The Desired Cross-Spectrum of the Sense Amino Acid Sequence of Poliovirus VP1

See also the caption of Fig. 1a.

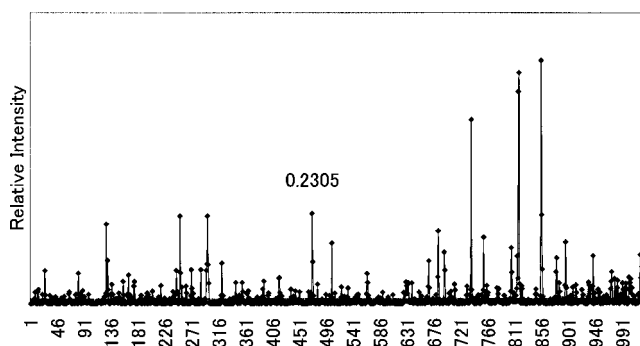


Fig. 6. The Desired Cross-Spectrum of the Sense Amino Acid Sequence of CD155

See also the caption of Fig. 1a.

compound.¹⁶⁾ Based on the analytical results of FGF1 and FGFR2 (data not shown), no characteristic peaks could be found in their 15 higher peaks, as in the relationship between FGF1 and FGFR1 (799aa), FGF2 (156aa) and FGFR1, or FGF2 and FGFR2 (data not shown). The present method thus would not be able to explain multi to multi reaction. A similar example is the relationship between human prion and laminin receptor 1.¹⁷⁾ Otherwise, in the case of ii), one direction from CD120b ($f=0.0845$ and 0.4214) to TNF- α ($f=0.4277$) is allowed (figures not shown), but one (*i.e.*, the above condition 2) of the two-symmetry operations (*i.e.*, the above conditions 2 and 3) for CD120b itself is unsatisfied. This result would have something to do with the ligand passing¹⁸⁾ indicated as a biological role of CD120b. In addition, no relationship between 55 kd TNFR (*i.e.*, CD120a) and TNF- α was directly observed.¹⁹⁾

The number of test samples investigated here may still not be sufficient to discuss the generality of a specific ligand-protein interaction. In addition, the elucidation accuracy by this method depends on the length of the amino acid sequence used. Nevertheless, based on the analytical results a frequency (or sequence)-dependent rule seems to be conserved in a simple and concerted interaction of other human pro-

teins, except those used in this study.

Acknowledgements One of the authors (NN) is grateful to Mr. Y. Asai at BFI for his support of this research.

References and Notes

- 1) Numao N., Fujii H., Fukazawa Y., Yoshioka K., Okada M., Tanaka K., *Chem. Pharm. Bull.*, **51**, 550—559 (2003).
- 2) All the data in the present sequence Fourier analysis were obtained using Mulliken's electronegativity scale as the parameters for amino acids as well as nucleotides (see ref. 1).
- 3) Numao N., Kidokoro S., *Biol. Pharm. Bull.*, **16**, 1160—1163 (1993).
- 4) The parenthesized numbers signify the total number of the amino acids calculated.
- 5) Bost K. L., Smith E. M., Blalock J. E., *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 1372—1375 (1984).
- 6) The content of each parenthesis means (ligand/receptor), and their sequences are of homo sapiens used, unless otherwise stated; (HIVgp120/CD4, CD195), [poliovirus capsid proteins (VP1, VP2, VP3, VP4)/CD155], (IL-2/CD25, CD122, CD132), (TNF- α /CD120a, CD120b), (FasL/CD95), (INS/CD220), (IFN γ /CDw119), (APP/BACE1), [CD230(prion)/laminin receptor 1], (glucagon/GCGR), (EGF, TGF- α /EGFR), (NPPA/NPR1), (angiotensin II/AGTR1), (somatostatin/SSTR1, SSTR2), (hCT, sCT/hCTR), (ACTH/MC2R), (GH/GHR), (NGF/p75, TRKa, TRKb, TRKc), (IL-6/CD126, CD130), [glycoprotein (hepatitis C virus, H77 strain)/CD81], (IL-1 β /CD121a, CDw121b), (FGF1, FGF2/FGFR1, FGFR2, FGFR3), (BMP2, BMP4, BMP7/BMPRIA, BMPRIb, BMPRII), (GCSF/GCSFR), (EPO/EPOR), (SCF/c-kit), (TGF- β /TGF- β RI, TGF- β RII), (HGF/HGFR), (UK/CD87), [spike glycoprotein (coronavirus 229E or SARS)/CD13], *etc.*
- 7) Kapurniotu A., Taylor J. W., *J. Med. Chem.*, **38**, 836—847 (1995).
- 8) Gorn A. H., Rudolph S. M., Flannery M. R., Morton C. C., Weremowicz S., Wang T. Z., Krane S. M., Goldring S. R., *J. Clin. Invest.*, **95**, 2681—2691 (1995).
- 9) The selection number is tentatively resolved by the different preference order of the peak ($f=0.4687$) in the spectrum derived from hCT or sCT, because of their different activities.
- 10) Employing its active form in the absence of *N*- or *O*-glycosyl group performed the calculation for a given protein.
- 11) This operation is equal to change of "the original f value" to "0.5—the original f value." Although the rationalization of this operation is currently unclear, this seems to be substantial to the sense-antisense interaction analysis.
- 12) Heal J. R., Roberts G. W., Raynes J. G., Bhakoo A., Miller A. D., *ChemBioChem*, **3**, 136—151 (2002), and references therein.
- 13) Takahara Y., Ando N., Kohara M., Hagino-Yamasaki K., Nomoto A., Itoh H., Numao N., Kondo K., *Gene*, **79**, 249—258 (1989).
- 14) Rieder E., Gorbalenya A. E., Xiao C., He Y., Baker T. S., Kuhn R. J., Rossman M. G., Wimmer E., *Dev. Biol. (Basel)*, **105**, 111—122 (2001), and references therein.
- 15) Breslin J. J., Mork I., Smith M. K., Vogel L. K., Hemmila E. M., Bonavia A., Talbot P. J., Sjostrom H., Noren O., Holmes K. V., *J. Virol.*, **77**, 4435—4438 (2003).
- 16) Ornitz D. M., Xu J., Colvin J. S., McEwen D. G., MacArthur C. A., Coulter F., Gao G., Goldfarb M., *J. Biol. Chem.*, **271**, 15292—15297 (1996).
- 17) Hundt C., Peyrin J.-M., Haik S., Gauczynski S., Leucht C., Rieger R., Riley M. L., Desly J.-P., Dormont D., Lasmezas C. I., Weiss S., *EMBO J.*, **20**, 5876—5886 (2001).
- 18) Tartaglia L. A., Pennica D., Goeddel D. V., *J. Biol. Chem.*, **268**, 18542—18548 (1993).
- 19) Lie B.-L., Tunemoto D., Hemmi H., Mizukami Y., Fukuda H., Kikuchi H., Kato S., Numao N., *Biochem. Biophys. Res. Commun.*, **88**, 503—509 (1992).