Application of the Jellyfish Model

I. Virulency of Some Mammalian Viruses

By Yukio KIHU,^{*)}^{†)} Ryuichi YAMADA,^{*)} Yasunori OKABE,^{**)} Christopher PALMER,^{*)}
and Yoshio OKADA, M.J.A.

(Communicated Oct. 13, 1997)

Abstract : In contrast to natural sequence of trypsin, functional sites of polar sequence of trypsin can be predicted only in the presence of cleavage site (CL) of the substrate, surface protein-F of viruses having strong or intermediate virulencies (Jellyfish model). In the case of less-virulent viruses, however, things are not so simple as virulent ones and sometimes DEV model fails to predict the functional sites which are already assigned empirically by virologists as the cleavage site just upstream to the long span of hydrophobic amino acids in the natural sequence. Actually, there are lots of weak virulent viruses in nature. Simple mathematical calculation supports this possibility which suggests a limitation in the application of our models. Mutation and selection are considered to control these situations (Appendix).

Key words : Virus virulency; DEV model; Jellyfish model; cleavage site; hydrophobic amino acid span.

Introduction. In 1988, Kiho presented the DEV model¹⁾ for the analysis of the primary amino acid sequence of trypsin. Its functional system is assigned, through which experimentally known active sites of trypsin, 40-H, 84-D, 177-S, are predicted by the DEV model.^{2),3)} Since then, our analyses have been extended to investigate the interaction with the substrate⁴⁾ including proteins other than trypsin.⁵⁾ A simple hypothesis^{1),6)} is the starting point of our model, in which two parameters, *DEV* and *DD*, are conceived. *DEV* is a measure of any possible functional activity in the protein and is defined by its primary amino acid sequence. *DD* is a measure of the spatial distance between two sites (site B and site E in our usage) in the higher dimensional (HD) structure of the protein. The functional site in the HD-structure is then defined by *DEV/DD*. Fortunately, the active sites are located near some of the functional sites derived from the DEV model. Thus the functional system of the protein can be defined.

Superficially, the story of the DEV model is interesting, reflecting the complex nature of the biological process. However, it is not a straightforward matter to get a reasonable conclusion which does not contradict the experimental results. For example, various strategies, all of which are derived from the DEV model, are often

required for the prediction of the active sites of protein. One difficulty is in the situation where some functional activity is low and more or less hidden. Then its expression by the DEV model needs the additional use of a variety of procedures. Assignment of the functional site around 40-H in trypsin is a typical example. In this case, a couple of strategies including “connector” and “interactor”^{2),4)} are needed to reach the goal. Furthermore, in addition to the natural sequence of trypsin, a smaller number of alphabet system including polar/non-polar sequences and/or sequences written by nucleic acid base (A, T, G, C) are used.^{3),6)-10)} Exact but less specific locations for the functional sites are often found. Another example of a difficulty is shown in this report, which is an evaluation of some viruses of weak virulency.

It is known that some mammalian viruses (paramyxo-, orthomyxo- and retroviruses) have one specific site in the surface protein-F which is more or less sensitive to trypsin, and this proteolysis is essential for subsequent viral infection¹¹⁾ and cell fusion.¹²⁾ Although there are many arginine residues, R (or lysine, K) in the F-protein, the proteolysis is restricted to only one cleavage site R (K). This is considered to be due to a long span (10–20 amino acids) of hydrophobic amino acids which are located just downstream to the above cleavage site.¹³⁾ The sensitivity to trypsin seems to correlate with their virulency.¹⁴⁾ These experimental results are investigated by the DEV model^{4),10)} and by rough energy calculations based on

^{*)} Ishihara Sangyo Kaisha, Ltd., 1–3–15 Edobori, Nishiku, Osaka 550, Japan.

^{**)} University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan.

^{†)} Correspondence to: Y. Kiho.

Table I. DEV analysis of trypsin with or without viral portion

Virus	No.	CL	Phob	Site B	Site E	DEV	DD	DEV/DD	Rank	Active sites predicted
no	1	—	—	B(T)	<i>e</i>	2.0	66	30	1	84-D
SV-5	2	+	—	B(T)	<i>e</i>	2.0	66	30	2	84-D
	3	+	+	B(V)	E	5.9	60	99	1	CL, 40-H, 177-S
				B(T)	E*	2.1	66	32	2	84-D, 40-H, 177-S
				B(V)	E	6.0	60	100	1	CL, 40-H, 177-S
NDV-v	4	+	—	B(T)	<i>e</i>	2.0	66	30	2	84-D
	5	+	+	B(V)	E	4.3	63	68	1	CL, 40-H, 177-S
				B(T)	E*	2.1	66	32	2	84-D, 40-H, 177-S
				B(V)	E	4.0	63	70	1	CL, 40-H, 177-S
NDV-av	6	+	—	B(T)	<i>e</i>	2.0	66	31	1	84-D
	7	+	+	B(V)	(201—)	1.3	62	20	—	CL
				B(T)	E*	2.1	66	32	—	84-D, 40-H, 177-S
				B(V)	(201—)	1.3	62	21	—	CL

CL: Cleavage site, five-amino-acid sequence, v1–v5, including a cleaved amino acid R (or K) at the end of v5. Phob: Downstream to CL, rich in hydrophobic amino acids, usually v6–v20. B(T): Site B having maximum DEV/DD in trypsin, exclusively at 79-NTGNN-83. The active site 84-D is the nearest neighbor. B(V): Site B at v1–v5. The peptide bond just after v5 is cleaved by trypsin. Site E: Several types are observed as follows; E=(26–30), (35–39), (64–68), (177–181), (v9–v20). Part of the last one appears when Phob is added for the analysis. The active sites (40-H and 177-S) related regions are underlined. *e*=(65–72), (90–94), (132–136). Other site Es are described by the sequence observed.

Table II. DEV analysis on viruses of intermediate virulency

Virus	CL	Phob	Site B	Site E	DEV	DD	DEV/DD	Rank	Active sites predicted
ReSV	+	—	B(T)	<i>e</i>	2.0	66	31	2	84-D
			B(V)	E	3.5	68	51	1	CL, 40-H, 177-S
	+	+	B(T)	<i>e</i>	2.0	66	30	2	84-D
			B(V)	E	3.4	68	50	1	CL, 40-H, 177-S
FPV	+	—	B(T)	<i>e</i>	2.0	67	30	2	84-D
			B(V)	E	2.6	73	36	1	CL, 40-H, 177-S
	+	+	B(T)	<i>e</i>	2.1	66	31	2	84-D
			B(V)	E	2.7	74	37	1	CL, 40-H, 177-S
VISV	+	—	B(T)	<i>e</i>	2.0	66	30	2	84-D
			B(V)	E	2.6	74	35	1	CL, 40-H, 177-S
	+	+	B(T)	<i>e</i>	2.1	66	31	2	84-D
			B(V)	E	2.7	74	36	1	CL, 40-H, 177-S
HIV	+	—	B(T)	<i>e</i>	2.0	66	30	1–1	84-D
			B(V)	E	2.3	75	30	1–2	CL, 40-H, 177-S
	+	+	B(T)	<i>e</i>	2.1	66	32	1–1	84-D
			B(V)	E	2.4	75	32	1–2	CL, 40-H, 177-S
A/Aichi	+	—	B(T)	<i>e</i>	2.0	66	30	1	84-D
			B(V)	E	1.8	81	23	—	CL, 40-H, 177-S
	+	+	B(T)	<i>e</i>	2.1	66	31	1	84-D
			B(V)	E	1.9	81	24	—	CL, 40-H, 177-S

S from B(V), in the presence of CL. Addition of Phob is not necessary in these cases. Information on site E and the predictable active sites are common among these viruses tested. Also DD values in B(V) are generally larger than those in B(T), which means that the spatial distance between CL and the two active sites is short. This situation may facilitate the first step of trypsin catalysis, *i.e.*, electron transfer from 177-S to viral CL. Most

DEV/DD values are larger in B(V) than those in B(T), except in the case of A/Aichi. However, the final conclusion is not affected. Table III shows the results on each virus whose CL has one or two hydrophobic amino acids. Therefore, DEV analysis is performed after these amino acids have been replaced by G. In contrast to Table II, correct assignment can be accomplished only in the presence of CL and Phob, except for HVJ and A/PR (data

Table III. DEV analysis on each virus whose CL has a hydrophobic amino acid

Virus	No.	CL	Phob	Site B	Site E	DEV	DD	DEV/DD	Rank	Active sites predicted
A/Japan	1	+	—	B(T)	<i>e</i>	2.0	66	31	1	84-D
				B(V)	(201—)	1.2	65	19	—	CL
	2	+	+	B(T)	E*	2.0	66	32	1	84-D, 40-H, 177-S
				B(V)	(201—)	1.3	64	20	—	CL
MMTV	3	+	—	B(T)	<i>e</i>	2.0	66	31	1	84-D
				B(V)	(158—)	1.0	55	—	—	CL
	4	+	+	B(T)	E	2.1	67	32	1	84-D, 40-H, 177-S
				B(V)	(27—), (74—), (93—), (117—)	1.0	57	18	—	CL
HVJ	5	+	—	B(T)	<i>e</i>	2.0	66	31	1	84-D
				B(V)	(201—)	0.4	42	9	—	CL
	6	+	+	B(T)	<i>e</i>	2.0	66	31	1	84-D
				B(V)	(201—)	0.4	49	8	—	CL
A/PR	7	+	—	B(T)	<i>e</i>	2.0	66	31	1	84-D
				B(V)	(201—)	1.0	78	13	—	CL
	8	+	+	B(T)	<i>e</i>	2.1	66	32	1	84-D
				B(V)	(201—)	1.1	76	14	—	CL
A/PR SIQSR	9	+	—	B(T)	<i>e</i>	2.0	66	31	1	84-D
				B(V)	(85—), (138—)	1.9	81	23	—	CL
	10	+	+	B(T)	E*	2.1	68	32	1	84-D, 40-H, 177-S
				B(V)	(178—), (229—)	2.0	81	25	—	CL, (177-S)

Table IV. DEV analysis of HTLV-1

No.	CL	Phob	Site B	Site E	DEV	DD	DEV/DD	Rank	Active sites predicted
1	+	—	B(T)	<i>e</i>	2.0	66	31	2	84-D
			B(V)	(85—), (138—)	3.1	70	44	1	CL, 84-D
2	+	+	B(T)	<i>e</i>	2.1	66	31	2	84-D
		(G6WG2)	B(V)	(85—), (138—)	3.1	70	45	1	CL, 84-D
3	+	+	B(T)	<i>e</i>	2.1	66	31	2	84-D
		(G6)	B(V)	(85—), (138—)	3.1	70	45	1	CL, 84-D
4	+	+	B(T)	<i>e</i>	2.1	66	31	2	84-D
		(G7 or G8)	B(V)	(85—), (138—)	3.1	70	45	1	CL, 84-D
5	+	+	B(T)	E*	2.1	66	32	2	84-D, 40-H, 177-S
		(G9)	B(V)	(178—), (229—)	3.2	70	45	1	CL, 84-D
6	+	+	B(T)	E*	2.1	66	32	2	84-D, 40-H, 177-S
		(G10)	B(V)	(178—), (229—)	3.2	70	45	1	CL, 84-D
7	+	—	B(T)	E1(182—)	1.9	60	32	2	84-D, 177-S
				E2(44—)	1.9	58	—	—	40-H
			B(V)	E1(201—)	2.2	53	43	1	CL
				E2(187—)	2.2	52	—	—	
8	+	***	B(T)	E1(182—)	2.0	59	33	2	84-D, 177-S
				E2(44—)	2.0	58	—	—	40-H
			B(V)	E1(201—), (8—12)	2.2	53	41	1	CL
				E2(187—)	2.2	52	—	—	

Details are noted in Table I. For Phob, several sequences are analyzed. Natural sequences of trypsin and viral portion (nos. 7 and 8) are also analyzed.

** Natural sequence of HTLV-1-Phob is AVPVAVWVLS.

Table V. Site Bs having correct E in ReSV with CL

Site B	Site E	<i>DEV</i>	<i>DD</i>	<i>DEV/DD</i>
40-44	E 40-H	1.55	81	19
170-174	E 171-D, near 177-S	1.52	83	18
197-201	E	1.70	86	20
198-202	E	2.06	81	26

entries nos. 5, 6, 7 and 8). The same conclusion is obtained without this replacement (data not shown), except A/PR. In the case of A/PR in which CL is natural SIQSR (data entries nos. 9 and 10), prediction of the active sites can be done in the same way as for other viruses in Table III. We do not know the underlying principle which can explain these situations, except that the results are affected mathematically by the formula of *DEV* mentioned above in Materials and methods. Characteristics of the viruses cited in Table III are similar to those of NDV-av in Table I. Viewing from the *DEV* model, these viruses are considered to have weak virulency.

Table IV is the result for HTLV-1 which is found to be the one most difficult to predict by *DEV* analysis. The data entries nos. 1 and 2 indicate that our *DEV* analysis failed to predict the active sites, 40-H and 177-S. We considered that the problem was in W(v12). Thus we tried v6-GGGGGG-v11 as Phob (W is not included) (data entry no. 3). However the prediction of the active site failed again. W was then interpreted as a hydrophobic amino acid (this is only partly true) and analyzed. The data entry no. 4 indicates that addition of Phob, G7 or G8, still fails to predict well. However, addition of Phob, G9 or G10, successfully predicts the active sites (data entries nos. 5 and 6). These facts suggest that (i) W should be treated as a hydrophobic amino acid and (ii) the size of Phob should be larger than nine amino acids. These findings in themselves indicate the complex and delicate nature of the biological process and suggest that very small differences can give rise to different issues, sometimes dramatically. However, these results are obtained with the polar sequence of trypsin and the viral portion in the above condition. The data entries nos. 7 and 8 are the results on the natural sequence. As reported previously,^{2),3)} assignment of functional sites through which the definition of the functional system and the prediction of the active sites of trypsin can be done with its natural sequence. The same results are obtained when CL with or without Phob are added, although with the natural sequence it is not so simple and not so exact as with the polar sequence.

The three active sites of trypsin can be identified from its natural sequence. From the polar sequence, however, it was not possible to predict two of them, 40-H and 177-S. This difficulty is overcome by the addition of the substrate portion to the polar trypsin. This is the main issue of the Jellyfish model. The effect of CL-Phob on various viruses are described above. In the case of the virulent viruses, the situation is fairly simple and the Jellyfish model explains well the results of the calculation. In the case of the weak virulent viruses, especially when hydrophobic amino acids are in CL, complex situations are observed. The essential point of the model is whether the site E is a correct E or not, as only from E the active sites, 40-H and 177-S, can be predicted. What is then necessary for site E to be the correct E? In order to solve this problem a similar calculation as above is performed with the polar trypsin after some of its portion is removed. The correct E are (26-30), (35-39), (64-68), (177-181) and Phob. All of these regions have the same amino acid composition, four Gs and one S, but (64-68) is remote from the active site, and so it is removed from the polar trypsin and then analyzed with various viruses. Change from E to *e* is observed which is shown as E* in Tables I, III and IV. Interestingly, they are all E for B(T) and not for B(V). One exception is observed with MMTV (data entry no. 4 in Table III). The reason for these changes is in the formula of *DEV* but we have no biochemical explanation for this yet. For the other kinds of site E, for example, *e* does not change when (64-68) is removed.

In contrast to the polar sequence, removal of (64-68) from the natural sequence of trypsin shows a big effect on site E (data not shown). Again, we have no biochemical explanation, but it suggests that (64-68) is a hidden function and somehow important for trypsin function.

Appendix 1. *DEV* value depend upon $n(z)$ (see the formula in Materials and methods). When two or more identical amino acids (z) are in the X region, we get a high *DEV* value for this region, to which high functional activity is assigned. *DD*, another parameter, is a measure of the spatial distance in the HD-structure: high *DD* values suggest that two sites (B and E) are located close to each other. Thus when both *DEV* and *DD* values are high, functional activity assigned to the X region can be expressed in a biochemical experiment. In the case of virulent viruses, SV5 and NDV-v, the number of R in CL are 5 and 4, respectively. Therefore, their high *DEV* values make their analysis simpler to perform. For example, identification of the functional sites and/or application of the Jellyfish model are easier in comparison to those of weak virulent viruses.

Another problem is that the number of these difficult cases is greater than that of the easier ones. Calculation may clarify the situation. Suppose that, out of 20 kinds of amino acids, we want to choose 5 amino acids. If repeating the amino acid is allowed, the number of possible combinations is huge, ${}_{20}H_5=42,504$. If repetition is not allowed, the number is still large, ${}_{20}C_5=15,504$. Now, if the number of repeated amino acid in X is 5, then the number of cases should be only 20. Only SV5 belongs to this category and its *DEV* value is high (Table I). Repetition of 4 amino acids is the second category which has $20 \times 20 = 400$ cases and there is only NDV-v with high *DEV* value. In the third category, 3 amino acids are repeated, the number of cases is 20 times ${}_{19}H_2=3,800$ and two viruses, ReSV and HTLV-1 are found whose *DEV* values are over 3 (Tables II and IV). Most of other viruses tested are lower categories with low *DEV* values. Fortunately, the virulence roughly parallels the *DEV* value, although the number of such viruses and difficulties in *DEV* analysis increase. In terms of phenomenology, both mutation and selection are two fundamental concepts for control of the situation. However, we do not know the underlying mechanism which is still controversial and challenging. Generally speaking, mutation is the essential force which gives rise to a diversity of biological organisms. Occasionally, a strong virulent mutation appears suddenly but after for a while, it may disappear and become hidden. Regulation by an attenuated virus¹⁷⁻²⁰ to control a strong virulent one is an additional mechanism to suppress the strong growth power of the latter and the dynamic society results in diversity. A mathematical approach²¹ to describe this situation, tells us "chaos".

Appendix 2. In the case of some intermediate or weak virulent viruses, even in the presence of CL-Phob, the prediction of 40-H and 177-S fails. For example, B(T) in Table II or B(V) in Table III. In general, these site Bs have no correct E. However, there are other site Bs than B(T) that give a correct prediction. As an example, the results on ReSV with CL is shown in Table V. Here, the first site B (40-44) contains 40-H and the second one (170-174) is close to 177-S. Although the *DEV* values are low, it suggests that the relationship between 40-H and 177-S is retained even when the B(T) can not predict 40-H and 177-S. Furthermore, similar data can be obtained in all the cases mentioned above, where B(T) or B(V) fails to predict 40-H and 177-S. Thus in reality, with the help of other strategies, all the active sites of the polar trypsin can be predicted in the presence of any CL of weak virulent

viruses. This leads to the extension of the Jellyfish model, which will be described in the next report.

Acknowledgments. We would like to say many thanks to Mr. Ohya, N., Ms. Noda, T. (Ishihara Sangyo Kaisha, Ltd.) and Ms. Kichinaka, I. (Japan Information Processing Service Co., Ltd.) for their valuable help in computer technology including manuscript preparation.

References

- 1) Kiho, Y. (1988) Cell Str. Func. **13**, 387-405.
- 2) Kiho, Y. (1995) Proc. Japan Acad. **71B**, 45-50.
- 3) Kiho, Y., Bhandari, G., and Okada, Y. (1996) Proc. Japan Acad. **72B**, 11-15.
- 4) Kiho, Y. (1995) Proc. Japan Acad. **71B**, 51-56.
- 5) Kiho, Y., Miyata, K., Ubasawa, A., Bhandari, G., Mizuno, Y., Iwai, T., and Okada, Y. (1995) Proc. Japan Acad. **71B**, 244-249.
- 6) Kiho, Y., Iwai, T., and Hida, T. (1996) Proc. Japan Acad. **72B**, 85-90.
- 7) Kiho, Y., Haga, T., and Oshima, K. (1996) Proc. Japan Acad. **72B**, 91-94.
- 8) Kiho, Y., Ubasawa, A., Okabe, Y., and Iwai, T. (1996) Proc. Japan Acad. **72B**, 185-190.
- 9) Kiho, Y., Katori, K., Ohshima, K., and Okada, Y. (1996) Proc. Japan Acad. **72B**, 191-196.
- 10) Kiho, Y., Bhandari, G., Ubasawa, A., Palmer, C., and Ikeda, S. (1997) Proc. Japan Acad. **73B**, 74-78.
- 11) Homma, H. (1971) J. Virol. **8**, 619-629; **9**, 829-835; **12**, 1457-1465.
- 12) Okada, Y. (1988) Current Topics in Membrane and Transport **32**, 297-336.
- 13) Goething, M. J., White, J. M., and Waterfield, M. D. (1978) Proc. Natl. Acad. Sci., U.S.A. **75**, 2737-2740.
- 14) Nagai, Y., Klenk, H. D., and Rott, R. (1976) Virology **72**, 492-508.
- 15) Okada, Y., Shima, Y., Shimamoto, T., Kusaka, N., and Kiho, Y. (1989) Cell Str. Func. **14**, 707-719.
- 16) Kiho, Y., Miyata, K., and Okada, Y. (1989) Cell Str. Func. **14**, 721-730.
- 17) Oshima, N., Goto, T., and Sato, R. (1964) Hokkaido Nat. Agric. Exp. Stn. Bull. **83**, 87-99.
- 18) Kiho, Y., and Nishiguchi, M. (1984) Microbiol. Immunol. **28**, 589-599.
- 19) Sabin, A. B. (1956) J. Am. Med. Assoc. **162**, 1589-1596; **164**, 1216-1223; (1973) J. Biol. Stand **1**, 115-118.
- 20) Ueda, S., Takahashi, M., Minekawa, Y., Ogino, T., Suzuki, N., Yamanishi, K., Baba, K., Okuno, Y., Konobe, T., Sasada, T., and Takaku, K. (1970) Biken J. **13**, 111-116; **14**, 253-258.
- 21) May, R. M. (1974) Science **186**, 645-647.