Basic construction of the functional system in a protein

III. Information transfer in a closed circuit

By Yukio KIH^{*,†)} Christopher PALMER,^{*} Hiroyuki MORI,^{*} Yoshio OKADA, M.J.A.,
and Tsuruji IWA^{**)}

(Contributed by Yoshio OKADA, M. J. A., June 15, 1999)

Abstract: From sites B and E in the DEV model, a closed circuit can be made in which the active sites of trypsin are found. In the presence of a substrate, the hydrophobic amino acid sequence near to the cleavage site of the substrate enters into the trypsin closed circuit. This closed circuit is considered to be a channel for information transfer. Information transfer can start from anywhere in the molecule and enters into the closed circuit. The mode of transfer is unique to biological catalysis, in contrast to chemical catalysis which operates by a one way transfer of information.

Key words: DEV model; information transfer; biological-catalyst; chemical catalyst.

Introduction. In the previous two papers, the proteolytic functional system of trypsin and chymotrypsin were investigated.^{1),2)} Originally, the functional system of trypsin was readily assigned by our DEV model, strategy A.^{1),3)} However, application of strategy A to proteins other than trypsin in order to discover their functional system was difficult and some simplification of the original sequence of the protein was needed for the assignment of the functional system. In the case of chymotrypsin,²⁾ the functional system was assigned with a model chymotrypsin, made by removal of the portion which interacts primarily with the substrate, surface protein of HVJ-TR2. Also in the case of trypsin, by making a couple of simplifications of the original sequence to construct a nonpolar sequence,⁴⁾ that with less alphabet,⁵⁾ and modified trypsin (replacing of human essential amino acids by G),¹⁾ assignment of the functional system was performed as well as in the original natural sequence. In these investigations, we designated the proteolytic functional system as the basic system of trypsin or chymotrypsin, in the general sense that an enzymatic functional system is the basic function of an enzyme protein.

An interesting outcome from these investigations is that the component of the basic system appears not as

an independent entity in a protein, but rather it appears to relate with other portions of a protein functionally and/or structurally.²⁾ Interactor sites (responsible for interaction of trypsin and its substrate), and connector sites (connecting functional sites in trypsin), reported in the previous paper,⁶⁾ are also considered to be in this category. These situations suggest that the functional system makes a kind of network in a protein relating with other portion of the molecule. In this third paper of the series to investigate "what is the basic construction of a functional system in a protein?", we attempt to interpret the basic constitution of a protein by the above mentioned functional network. For this purpose, a network was made from the site B and E. It was found that a closed circuit was made on which components of the basic functional system and active sites are located. In short, the basic constitution of protein, proteolytic functional system of trypsin, makes a closed circuit. In our previous report,⁷⁾ information transfer in a protein was discussed in comparison with that in an electric circuit. The above mentioned closed circuit is one of the issues concerning this problem.

Materials and methods. Details of DEV analysis can be referred to in the previous reports.³⁾ A network is constructed from site B and E derived from DEV analysis on a protein. As an example, a trypsin network is made as follows. Table I shows the result of the DEV analysis on trypsin (DD > 60). DEV, a regional variance of amino acids deviated from pseudo-random distribution that is derived from the amino acid composition in a pro-

^{*}) Ishihara Sangyo Kaisha, Ltd., 1-3-15 Edobori, Nishi-ku, Osaka 550-0002, Japan.

^{**)} Kansai Medical University, Hirakata, Osaka 573-1136, Japan.

^{†)} Correspondence to: Y. Kiho.

Table I. Intramolecular interaction of trypsin analysed by DEV model (DD>60)

Site B	Site E	DEV	DD	P
1-5	94-98	0.98	64.4	63
1-5	123-127	0.98	60.8	60
2-6	28-32	1.02	61.7	63
2-6	94-98	1.02	61.7	63
3-8	16-20	1.08	61.9	67
3-8	68-72	1.08	62.6	68
5-9	16-20	0.78	60.5	47
6-11	132-136	0.69	62.8	43
9-13	43-47	0.78	63.7	50
9-13	116-120	0.78	65.9	51
9-13	183-187	0.78	67.1	52
12-16	27-31	1.23	61.5	76
12-16	117-131	1.23	61.5	76
16-20	196-200	0.73	67.5	49
28-32	196-200	0.76	64.4	49
28-32	202-206	0.76	63.3	48
28-32	204-208	0.76	63.3	48
39-43	27-31	0.92	60.7	56
39-43	117-121	0.92	60.7	56
46-50	122-126	0.86	60.9	52
50-54	68-72	0.79	60.1	47
53-57	183-187	1.14	63.1	72
53-57	184-188	1.14	60.8	69
54-58	41-45	1.42	60.1	85
54-58	124-129	1.42	60.8	86
54-58	183-187	1.42	64.4	91
54-58	195-199	1.42	60.7	86
66-70	187-191	1.18	62.3	74
77-82	44-48	1.18	60.8	72
77-82	182-186	1.18	63.8	75
79-83	182-186	1.91	60.3	115
80-84	182-186	1.22	62.0	76
81-85	182-186	1.18	64.2	76
83-87	182-186	0.90	61.9	56
91-95	187-191	1.18	62.3	74
92-96	201-205	1.19	64.8	77
94-98	202-208	0.76	63.3	48
100-104	154-158	1.12	61.0	68
100-104	186-190	1.12	63.3	71
102-106	6-11	0.80	63.8	51
102-106	154-158	0.80	60.6	48
102-106	196-200	0.80	61.7	49
104-108	8-12	0.93	60.6	56
107-111	201-205	0.91	63.3	58
108-112	187-191	1.21	61.0	74
108-112	201-205	1.21	63.9	77
120-124	68-72	1.34	64.9	87
120-124	101-105	1.34	62.4	84
120-124	133-137	1.34	63.6	85
121-125	101-105	1.03	62.0	64
123-127	187-191	0.80	67.1	54
126-130	197-201	1.36	60.2	82
126-130	208-214	1.36	61.5	84
136-140	44-48	1.15	62.6	72
136-140	188-192	1.15	62.6	72
143-147	8-12	1.49	64.4	96
143-147	201-205	1.49	65.0	97
144-148	8-12	1.51	63.7	96
144-148	187-191	1.51	61.9	93
144-148	201-205	1.51	64.2	97
146-151	8-12	1.42	67.5	96
146-151	154-158	1.42	62.0	88
146-151	187-191	1.42	65.7	93
154-158	101-105	0.71	61.9	44
154-158	133-137	0.71	64.2	46
164-169	68-72	1.12	60.2	67
167-171	16-20	1.11	60.0	67
178-182	28-32	1.32	62.6	83
178-182	94-98	1.32	62.6	83
178-182	219-223	1.32	61.1	81
185-189	97-101	0.99	60.1	59
185-189	219-223	0.99	61.9	61
187-191	123-127	0.66	60.2	40
192-196	207-211	1.09	60.0	65
196-200	16-20	0.70	66.4	46
196-200	28-32	0.70	61.5	43
197-201	189-194	0.78	60.3	47
198-202	116-120	1.15	61.9	71
200-204	101-105	1.05	61.0	64
202-208	28-32	0.71	60.7	43
202-208	94-98	0.71	60.7	43
205-209	17-22	1.18	62.2	73
205-209	27-31	1.18	64.2	76
205-209	117-181	1.18	64.2	76
209-213	43-47	1.13	61.3	69
209-213	184-188	1.13	61.3	69
211-215	116-120	0.84	61.2	51
213-217	17-22	1.19	61.2	73

P = product of DEV and DD.

tein, is a type of information about site B. DD is an inverse distance between two sites (site B and E), in a three dimensional structure of a protein.³⁾ If information transfer occurs between these two sites B and E,⁷⁾ (it is highly possible that this assumption is valid), easier transfer can be attained when both DEV and DD are larger values. In the final column of the Table I, a product of DEV and DD is shown ($P = DEV \times DD$), from which information transfer between site B and site E is evaluated. The range of P values are between 40 and 115. Fig. 1(a) is a graphical representation of the network connecting sites B and E having large P values (>80). Fig. 1(b) is the simplified representations of the possible networks derived from the network with P values more than 90. The active sites, 40-H, 84-D and 177-S are related. Further analyses are described later (Results and discussion).

In this paper, several proteins are analyzed and their networks are presented. These are chymotrypsin, elastase, RNase, lysozyme, with their substrates in addition to trypsin.

Results and discussion. Fig. 2(a) shows a network in which P values are between 70 and 80. A couple of new interactions are added to the above network (Fig. 1(b)). The resulting new network is shown in Fig. 2(b). Two characteristics are noted: (i) the site (41-48), active site 40-H region, and the site (77-85), active site 84-D region, are directly related. (iii) sites (41-48), (77-85), (182-192) and (136-140) make a closed circuit in which the first three sites are near three active sites. This is shown by the bold line in Fig. 2(b). The concept that an active site region in a protein can make a closed circuit in which transfer of biological information may occur is an interesting and radical idea. This suggests that in a protein some information transfer is restricted in a type of closed circuit like an attractor in a complex system. Then our problem is, from where does an information start and to where does it give an output. At present, we have no biochemical information on this issue.

For the first problem, i.e. the start of information transfer, the following procedure was taken to obtain a network and its relation to the closed circuit was examined. Table II (a) shows some results of the DEV analysis on trypsin, site B (1-5). In the DEV model, only positive DD values are concerned. Furthermore, the site E which is close to the site B and has large P value (local maximum in P) was chosen. The resulting site E is thus the site E (9-13). Next, this site E is assumed to be the next site B and again its nearby site E having locally maximum P value is obtained. It is the site (17-21) ($P = 33$).

Fig. 1. Trypsin network I. (a) Network connecting sites B and E by large P (product of DEV and DD). $P \geq 90$ (solid line), $80 \leq P < 90$ (dotted line). (b) Network ($P \geq 90$).

Fig. 2. Trypsin network II. (a) Network connecting sites B and E ($70 \leq P < 80$). (b) Network ($P \geq 70$).

The same procedure as above is repeated and the third site (39-43) obtained which partly overlaps with the region (41-48) in the closed circuit. Thus the network starting from the site (1-5) enters into the closed circuit at (41-48) region eventually, which is shown in Table II (b)-(1). Similar networks are obtained and several examples are shown in Table II (b). For example, case (2) indicates the starting site B (2-6). Again it enters into the closed circuit at the same site (39-43). Surprisingly, all starting site Bs from (1-5) to (36-40) were found to enter into the closed circuit similarly as above. The case (3), starting site B (52-56), enters into the closed circuit at (77-81). Alternatively, in the reverse direction, it also enters into the closed circuit at (44-48). As shown in the case (4), information transfer

in the reverse direction is essential, although P values of some transfer steps are fairly low. The same procedure as above was applied to the entire region of the primary sequence of trypsin and it was found that all site Bs entered into the closed circuit and there is no site B which is independent of the closed circuit. These results indicate that information transfer may start anywhere in the primary sequence of trypsin and can enter into the closed circuit. Obviously, it depends on the magnitude of the P value, and it is generally the low values for a network outside of the closed circuit which may reflect the unique nature of a biological situation (i.e. a biological catalyst in contrast to a chemical catalyst), predicted by the DEV model.⁹⁾ It suggests that a protein in vivo is usually in a structural fluctuation, taking a loose and flexible structure in which mainly mechanical energy is operating. By chance encounters with substrate, it forms a somewhat more compact structure in which an ordinary physico-chemical interaction may occur, and its function is then expressed. This idea is presented as the Jellyfish model in our previous report.^{4),8)} In a chemical reaction, the necessary starting materials and energy are afforded for the reaction to initiate and proceed in an appropriate environment. As a whole, this one way transfer of information (not a closed circuit) is sufficient for a chemical reaction to occur. On the contrary, in biological situation, the above notion is somehow related to the basic construction principle of a protein.

Table III shows the results of DEV analysis on the interaction of trypsin with its substrate, NDV-F surface protein.⁶⁾ As a source of site B and site E, regions around sites B shown in Fig. 2(b) and entire regions of the substrate protein, respectively are used. Interestingly, sites E from NDV-F are also restricted to two regions which are (97-102) and (360-367). The former is the hydrophobic amino acid region downstream from the cleavage site, 91-R of NDV-F protein and is considered to be one of the most important sites interacting with trypsin.⁶⁾ Adding this new information to the above network, we can derive trypsin-substrate network shown in Fig. 3. It indicates that two NDV-F portions including the hydrophobic amino acid sequence near the cleavage site can interact with the closed circuit of trypsin. Thus the latter circuit is enlarged by a portion of the substrate to ensure the interaction between trypsin and its substrate, without violating the original form of the closed circuit. The cyclic nature of trypsin function has been presented by a different approach in the previous report.⁴⁾

The same procedures as above are applied to pro-

Table II. Intramolecular Interaction of Trypsin analysed by DEV model

(a) Site Es for the site B (1-5) are searched around the region (1-21).

Site B	Site E	DEV	DD	P
1-5	1-5	0.98	-100.3	
⋈	2-6	⋈	-84.2	
⋈	3-7	⋈	-69.0	
⋈	4-8	⋈	⋈	
⋈	5-9	⋈	-17.1	
⋈	6-10	⋈	11.7	11
⋈	7-11	⋈	11.7	11
⋈	8-12	⋈	-3.5	
⋈	9-13	⋈	23.3	23
⋈	10-14	⋈	-2.8	
⋈	11-15	⋈	-10.0	
⋈	12-16	⋈	-43.1	
⋈	13-17	⋈	12.6	12
⋈	14-18	⋈	17.1	17

(b) Sites B-E network (numbers under arrows are P values. Chymotrypsin sequence is underlined>.

(1) Starting from (1-5) → (9-13) → (17-21) → (39-43)
23 33 29
(2) Starting from (2-6) → (16-20) → (23-27) → (28-32) → (39-43)
46 24 55 29
(3) Starting from (52-56) → (68-72) → (77-81)
↓ 52 48 41
(44-48)
(4) Starting from (219-223) → (202-206) → (195-199) → (156-160) → (162-166)
30 26 11 28
→ (170-174) → (177-181) → (183-187)
10 21 11
(5) model protein, starting from (1-5) → (9-13) → (17-21) → (29-33) → (38-42)
25 32 14 10

teins other than trypsin. Fig. 4 shows the chymotrypsin network. As in the case of trypsin, a closed network was found. Three active sites of chymotrypsin, 42-H, 87-D and 180-S, are in or nearby the component sites in the network. HVJ-TR2 surface protein²⁾ was used as a substrate for chymotrypsin. Hydrophobic amino acid sequence (120-133), downstream to the cleavage site 116-I, relates the chymotrypsin network at (48-54) and (80-85), as shown in Fig. 4. Fig. 5 shows the elastase network. Elastase is a similar serine protease as chymotrypsin, hydrolyzing the peptide bond just after hydrophobic amino acid residue of small size. (chymotrypsin has an affinity for hydrophobic amino acid residue of large size.) Elastase is composed of 240 amino acids. Its active sites are 45-H, 93-D and 188-S. Unfortunately in this case, the specific substrate such as in the case of trypsin or chymotrypsin is not known. (A specific substrate is one in which only one specific site is attacked by the enzyme, for example, only 91-R in NDV-F protein is cleaved by trypsin, although it has

Table III. Interaction of Trypsin with NDV-F (site B = trypsin portion shown in Fig. 2 (b), site E = NDV-F). (DEV>1, P>70)

Site B (trypsin)	Site E (NDV-F)	DEV	DD	P
52-56	98-103	1.27	60.5	77
54-58	360-364	1.42	62.6	89
79-83	97-101	1.91	60.8	116
79-83	362-366	1.91	61.2	117
135-139	97-101	1.55	62.2	96
135-139	265-269	1.55	61.1	95
135-139	363-367	1.55	62.2	96
199-203	97-103	1.19	64.0	76
199-203	363-367	1.19	64.0	76

Fig. 3. Trypsin:NDV-F network.

many other R scattered in the NDV-F protein.) Accordingly, an artificial substrate, NGQLSV, was used in this case. Its derivation is as follows. To the elastase sites shown in Fig. 5, artificial peptides (usually of 5 amino acids) made out of 20 different kinds of amino acids obtained by random selection were added to calculate their DD values. From the peptides obtained, we have chosen NGQLSV which relates with a couple of sites in the elastase network at large DD value. Thus a similar network as in trypsin and/or chymotrypsin is obtained as shown in Fig. 5.

From Fig. 6 to Fig. 8, networks of several proteins are shown. RNase T1 (Fig. 6) and lysozyme (Fig. 7) both have similar networks as above. Probably due to insufficient data, we could not obtain a closed circuit for several proteins shown in Fig. 8. However, with an artificial substrate, a closed circuit could be presented as shown in Fig. 8(a) for subtilisin, (b) pepsin, (c) phospholipase A2 and (d) lactate dehydrogenase. Summing these results, we can say that a protein has a closed circuit in which functional information can be transferred. Thus this circuit is considered to be the basic constitution of a protein.

Appendix-I. The relation between the specificity

Fig. 4. Chymotrypsin network. 230 aa (42H*, 87D**, 180S***).

Fig. 5. Elastase network. 240 aa (45H*, 93D**, 188S***).

Fig. 6. RNase T1 network. 104 aa (40H*, 58E**, 77R***).

Fig. 7. Lysozyme (egg) network. 129 aa (35E*, 52D**, 62W***).

Fig. 8. Networks of some protein. (a) Subtilisin network. 275 aa (32D*, 64H**, 221S***). (b) Pepsin network. 327 aa (32D*, 72S**, 215D***). (c) Phospholipase A₂ network. 124 aa (28Y*, 48H**, 49D). (d) Lactate dehydrogenase network. 329 aa (149C*, 176H**).

and the basic constitution of trypsin. As described above, trypsin cleaves the peptide bond at R and chymotrypsin has an affinity for large hydrophobic amino acid residues. Hedstrom *et al.*¹¹⁾ have pointed out that the S1 binding pocket of trypsin (171-DSCQGDS, 192-SWGS GC and 203-PGV), determines the specificity of ester hydrolysis, whereas specific amide hydrolysis requires both the S1 binding site and two loops, L1 (165-YLEGG) and L2 (198-AQKNK). Proteolytic reaction (amide hydrolysis), is more specific than ester hydrolysis. Actually, by site-directed mutagenesis, these authors have succeeded in converting trypsin to a chymotrypsin-like protease by modification of trypsin-S1,

-L1 and -L2 to chymotrypsin-S1 (174-SSCMGDS, 199-SWGSSTC, 210-PGV), -L1 (170-ASGV) and -L2 (206-STST). (In contrast to amide hydrolysis, ester hydrolysis is of low specificity and replacement of only S1 site is enough to convert their ester hydrolytic activity from trypsin to chymotrypsin or vice versa.)

What kind of specificity does the closed circuit of trypsin have? From Table I, interactions involving S1, L1 and L2 are chosen and are shown in Fig. 9(a). Relation between the specificity region (S1, L1 and L2) and the closed circuit is shown in Fig. 9(b). Adding the information from Table III, these results indicate that the specificity region can relate with the closed circuit of trypsin

Fig. 9. Relation between biochemical specificity and closed circuit of trypsin. S1 = (171-177), (192-197), (203-205); L1 = (165-169); L2 = (198-202). (a) Interactions involving S1, L1 and L2. (b) Trypsin closed circuit relating with S1, L1 and L2.

which would stimulate the highly specific enzymatic reaction. (In contrast, the functional system of trypsin derived from the DEV model, strategy A or simple closed circuit lack such specificity as above.)

Appendix-II. In order to know whether the above notion, “information transfer may start anywhere in the primary sequence of trypsin and can enter into the closed circuit”, is applicable to any peptide sequence or not, a network of model protein, trypsin portion (1-13) plus chymotrypsin (14-39) plus trypsin (40-223) was examined and the result was shown in Table III (b)-(5). Starting from trypsin (1-5), through chymotrypsin sequence (dotted underlined), information transfer enters into (38-42), trypsin closed circuit. Thus, the transfer may be possible but P values are lower in the chymotrypsin sequence (dotted underline), compared with the pure trypsin sequence (see case (1) in Table III (b)). This suggests that the transfer is not so easy. Thus the P value is an important parameter when we evaluate an information transfer.

Acknowledgments. We owe many help to Drs. A. Ubasawa and K. Miyata for their critical reading of the manuscript and profound discussion.

References

- 1) Kiho, Y., Palmer, C., Ito, M., Iwai, T., and Okada, Y. (1999) Proc. Japan Acad. **75B**, 144-149.
- 2) Kiho, Y., Okada, Y., and Iwai, T. (1999) Proc. Japan Acad. **75B**, 150-155.
- 3) Kiho, Y. (1995) Proc. Japan Acad. **71B**, 45-50.
- 4) Kiho, Y., Palmer, C., Katori, K., and Iwai, T. (1997) Proc. Japan Acad. **73B**, 176-181.
- 5) Kiho, Y., Ubasawa, A., Okabe, Y., and Iwai, T. (1996) Proc. Japan Acad. **72B**, 185-190.
- 6) Kiho, Y. (1995) Proc. Japan Acad. **71B**, 51-56.
- 7) Kiho, Y., Katori, K., Oshima, K., Iwai, T., and Okada, Y. (1996) Proc. Japan Acad. **72B**, 191-196.
- 8) Kiho, Y., Bhandari, G., Ubasawa, A., Palmer, C., and Ikeda, S. (1997) Proc. Japan Acad. **73B**, 74-78.
- 9) Kiho, Y., Iwai, T., Migita, T., Okada, Y., Okabe, Y., Ochi, M., and Katori, K. (1996) Proc. Japan Acad. **72B**, 95-100.
- 10) Kiho, Y., Yamada, R., Okabe, Y., Palmer, C., and Okada, Y. (1997) Proc. Japan Acad. **73B**, 175-176.
- 11) Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) Science **255**, 1249-1253.