Functional Word in Protein

VI. Recognition

By Yukio KIHU,^{*)}^{†)} Guru BHANDARI,^{**)} Aiko UBASAWA,^{***)} Christopher PALMER,^{**)}
and Shigero IKEDA^{****)}

(Communicated by Yoshio OKADA, M.J.A., May 12, 1997)

Abstract: Interaction between trypsin and its substrate, NDV-F, is investigated by two approaches, in which the natural sequence and the polar/non-polar sequence are used respectively. In the first approach that we call “stepwise recognition”, addition of the connector-Es of trypsin resulted in stronger interaction between trypsin and NDV-F than that reported previously.²⁾ The polar sequence of trypsin could predict only a functional site near 84-D. In the second approach, however, when the cleavage site of NDV-F is added to the polar sequence of trypsin, the other active sites of trypsin, 40-H and 177-S, can be predicted. We call it “sudden recognition by chance encounters”. The results suggest that the function of trypsin is recognized only when the enzyme encounters its substrate by chance.

Key words: Trypsin; NDV-F; stepwise recognition; connector-E; sudden recognition by chance encounters; Jellyfish model.

Introduction. From the DEV model,¹⁾⁻⁴⁾ we have suggested that the higher dimensional (HD) structure of trypsin is rather flexible in its natural state.⁵⁾ It must be distinct from the structure obtained by X-ray crystallography, which is derived from the crystallized state of trypsin. The flexible HD-structure is derived from a structural model (made of wire and rod) in which a parameter *DD*, a measure of the distance between two sites (site-B and site-E in our usage) in the protein, is a decisive factor for making the model.⁶⁾ Analysis by the DEV model on trypsin catalysis^{2),6)} or consideration of the information transfer in it⁵⁾ support the above suggestion. Additionally, some fluctuations are added to these processes until the final step is reached, i.e., a substrate (mainly we used NDV-F) site is cleaved by trypsin. In energetics, mechanical energy is suggested to be involved in such a flexible structure.⁵⁾ Deducing from *DD* values, ordinary physicochemical type energies, those of ionic, van der Waals or hydrogen bondings, seem to be unlikely factors in these situations

except in the final step. From our calculation, *DD* values should be larger than 65 for physicochemical type energies to be a factor.

The DEV model shows us the complex nature of the biological system. We need various strategies to identify the functional system of a protein^{1),3)} or to analyze its interaction with a substrate.^{2),6)} These ambiguous and complex situations can be improved by taking more simplified approaches: (i) analysis on a small protein model made of several portions of parent protein^{2),4),6)} and/or (ii) using a smaller alphabet system instead of twenty alphabets in natural protein.^{4),5),7)-9)} The results obtained from these artificial conditions are clear. However, we still do not know about the situation just prior to the final step of site cleavage.

In this report, we focus our investigation on the problem, “How is the final step attained?” and present two approaches to elucidate this problem: one is *stepwise recognition* and the other is *sudden recognition by chance encounters*. The former is based on our previous reports^{2),4)-6),8)} and the latter is a new idea which we call “the jellyfish model”.

Materials and methods. Details on the DEV model and its analytical methods can be referred to from the previous reports.^{1),7)} Amino acid sequences of trypsin and its substrate, NDV-F (F-protein of Newcastle disease virus) is reported previously.²⁾ Details on the smaller

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

^{**)} Ricerca, Inc., Painesville, Ohio, U.S.A.

^{***)} Woodward Laboratory, Naracho, Aoba-ku, Yokohama 227, Japan.

^{****)} Emeritus Professor at Osaka University, Kasugaoka, Ibaraki 568, Japan.

^{†)} Correspondence to: Y. Kiho.

Table I. Intramolecular interaction ($DD \geq 60$) of the model proteins shown in Fig. 1

	No.	Site-B	Site-E	DEV	DD	Comments
(a)	1	6-VVSAA	25-NDIMG	1.35	60.5	near 40-H and 84-D
	2	10-AHCYK	1-GSLIN	0.94	62.0	40-H and HS-connector
	3	20-PNTLN	15-SGIQV	0.95	64.9	near 84-D and HD-connector near 40-H
	4	22-TLNND	47-QGIVS	0.95	64.9	84-D and HD-connector near 177-S
	5	23-LNNDI	42-VCSGK	0.82	66.1	84-D and HD-connector near 177-S
(b)	6	3-LINVV	34-QGDSG	0.98	63.7	near 40-H and 177-S
	7	3-LINVV	44-SGKQG	0.98	63.7	near 40-H and near 177-S
	8	5-NVVSAA	45-GKQGI	0.96	62.7	near 40-H and HD-connector near 177-S
	9	20-PNTLN	15-SGIQV	1.31	64.6	near 84-D and HS-connector
	10	20-PNTLN	47-QGIVS	1.31	64.6	near 84-D and HD-connector near 177-S
	11	22-TLNND	47-QGIVS	1.28	66.6	84-D and HD-connector near 177-S
	12	35-GDSGG	2-SLINV	1.33	64.4	177-S and HS-connector
	13	35-GDSGG	17-IQVPN	1.33	63.5	177-S and HD-connector near 84-D
	14	6-VVSAA	v20-LITGN	1.38	60.6	near 40-H and NDV-F
	15	22-TLNND	v12-IGSVA	1.28	67.6	84-D and NDV-F (hydrophobic sequence)
	16	38-GGPVV	v25-ISNAL	1.18	69.8	near 177-S and NDV-F
	17	v4-QRRFI	v13-GSVAL	1.28	60.2	intra-NDV-F
	18	v21-ITGNI	v13-GSVAL	0.86	63.8	intra-NDV-F

(a) Trypsin portion (1–51).

(b) NDV-F portion (v1–v29) is added to (a).

The active sites are underlined.

alphabet system are previously described for polar and non-polar sequences^{4),7),8)} and for three alphabet (A, C, S).^{5),9)} Construction of the wire and rod model is described in the previous reports.^{3),4),6)}

Results and discussion. In the assignment of functional sites and the prediction of the active sites of trypsin by the DEV model, two steps are considered to reach the goal. The first step is a simple outcome of the mathematical calculation by the DEV model in which any information on the physicochemical nature of amino acids or protein is not considered necessary to assign the functional sites and then to define the functional system of trypsin. In the second step, active sites of trypsin, 40-H, 84-D and 177-S, which have been determined already by biochemical experiments, are found by the DEV model near or inside the functional sites mentioned above.¹⁾ This idea is realized more clearly if we classify the twenty kinds of amino acids into two categories: polar and non-polar amino acids from which polar and non-polar amino acid sequences of trypsin are made. Usually, in a polar amino acid sequence, all hydrophobic amino acids are replaced by glycine (G) to negate the physicochemical nature of these amino acids and accentuate the polar amino acids. For a non-polar sequence, all polar amino acids are replaced by glutamic acid (E) to accentuate the non-polar amino acids. (Choice of amino acid for replacement is arbitrary.)

It is found that functional sites can be assigned in the

non-polar sequence of trypsin (similarly as in the natural sequence) but not in the polar one.⁴⁾ By combining these three functional sites to make a small protein model, the active sites in the original trypsin can be predicted. The non-polar sequence of the small protein model does not work here. This procedure works only for a small polar protein. (Later, this point is modified in the interaction of polar trypsin with the substrate.) These results suggest that the first step, which the hydrophobic amino acids are responsible for, is followed by the second step in which the polar amino acids are involved. This difference between the first step and the second step can also be observed in their *DD* values. The former has smaller *DD* values than those of the latter. Accordingly, the HD-structure of trypsin is flexible and that of the small polar protein model mentioned above is compact by which the three active sites come closely together each other and their prediction can be achieved.⁴⁾

To investigate the interaction between trypsin and its substrate, NDV-F, we used two approaches, (A) and (B), in which the natural sequence and the polar/non-polar sequence, respectively, are used.

(A) *Stepwise recognition.* In the previous report,²⁾ the two step nature of the interaction between trypsin and NDV-F is presented, i.e. a bimolecular interaction followed by monomolecular interaction. Then the resulting possible interacting complex made of functional sites of both

Fig. 1. Amino acid sequence of the model proteins. The connector-Es are underlined. The other portions are derived from the functional sites of trypsin and NDV-F. The trypsin portions, (1–5), (6–19), (20–28) and (29–51), correspond to the original regions of trypsin, (27–31), (35–48), (78–84) and (169–191), respectively. The NDV-F portions, (v1–v17), (v18–v24) and (v25–v29), correspond to the original regions, v(86–102), v(241–247) and v(449–453), respectively. The border between trypsin and NDV-F is indicated by a vertical dotted line. The active sites of trypsin and the cleavage site of NDV-F are indicated under the sequence by the number of the respective original molecule.

Fig. 2. Polar sequence of trypsin (a) and a portion of NDV-F including the cleavage site and its downstream, hydrophobic amino acid sequence (b). The active sites and the cleavage site are double-underlined.

Fig. 3. The structure of the model protein made of site-Bs and site-Es, shown in Table III. The active sites of trypsin and the cleavage site of NDV-F are double-underlined. The border between the portions of trypsin and NDV-F is marked by a vertical dotted line.

molecules is considered. In terms of *DD* values, however, we still do not know sufficiently the details of the interaction because of the flexible HD-structural nature of the complex; except the interaction of 84-D containing region with hydrophobic region of NDV-F (Fig. 4(c) in the previous report²⁾). The difficulty can be overcome by introduction of connector-Es,¹⁾ as shown in Fig. 1 and Table I. The amino acid sequence of the small protein

Table II. Assignment of functional sites in the polar sequence of trypsin (a) and the model protein (b)

(a) Polar sequence of trypsin

		DEV	DEV/DD	DD
Site-B	79-NTGNN-83	2.02	30	
Site-E1	(65–72), (90–94), (132–136)			66.4
Site-B	144-SDSSC-148	2.41	30	
Site-E1	(85–90), (138–143)			81.3

(b) Polar sequence of the model protein (trypsin + NDV-F portion)

		DEV	DEV/DD	DD
Site-B1	v2-RRQRR	4.40	70	
Site-E1	(26–30), (35–39), (140–144), (177–181), (v10–v17)			63.3
Site-B2	79-NTGNN-83	2.08	32	
Site-E1	(26–30), (35–39), (64–68), (140–144), (177–181), (v10–v17)			65.7

model derived from various portions of trypsin and NDV-F is shown in Fig. 1. Note that three connector-Es of trypsin are added here: (1-GSLIN-5), (15-SGIQV-19) and (43-CSGKQGIIVS-51) which relates 40-H to 177-S, 40-H to 84-D and also 40-H to 84-D, respectively in the original molecule.¹⁾ By addition of these connectors, improved information can be obtained. Intra-molecular interactions of the model protein are shown in Table I(b). As a control, calculation is made with only the trypsin portion (1–51) of the model protein and its result is shown in Table I(a). Data Nos. 1–5 and data Nos. 6–13 are intra-trypsin interactions. Comparing these data, it is found that most of the data in Table I(a) effectively disappear (*DD* is decreased) in Table I(b). Only data Nos. 3 and 4 are conserved which correspond to No. 9 and 11, respectively in Table I(b). The data Nos. 6–8, 10, and 12–13 are newly appeared (*DD* is increased) in Table I(b). This indicates that intra-trypsin interactions are changed to make a well defined system by the interaction with NDV-F. Data Nos. 14–16 represent new interactions between trypsin and NDV-F. Three active site regions of trypsin appear as site-B. Especially, 84-D and 177-S regions (but not 40-H region), interact fairly strongly with NDV-F functional sites with large *DD* values (data Nos. 15–16). Data Nos. 17–18 are intra-NDV-F interactions. These data are distinct from those reported previously²⁾ and this is due to the addition of connector-Es of trypsin to the model protein (Fig. 1).

(B) Sudden recognition by chance encounters (jellyfish

Table III. Intramolecular interaction ($DD \geq 60$) of the model protein, made of Figs. 2(a) and 2(b).
The active sites and the cleavage site are double-underlined

No.	Site-B	Site-E	<i>DEV</i>	<i>DD</i>	Comments
1	40-HCYKS	84-DGGGG	1.63	78.4	40-H and 84-D
2	40- <u>H</u> CYKS	177- <u>S</u> GGGG	1.63	80.8	40-H and 177-S
3	40- <u>H</u> CYKS	v6- <u>R</u> GGGG	1.63	78.4	40-H and NDV-F (cleavage site)
4	80-TGNN <u>D</u>	35-GGSGG	1.32	68.4	84-D and near 40-H
5	80-TGNN <u>D</u>	177-SGGGG	1.32	68.4	84-D and 177-S
6	80-TGNN <u>D</u>	v6- <u>R</u> GGGG	1.32	68.4	84-D and NDV-F (cleavage site)
7	80-TGNN <u>D</u>	v12- <u>G</u> SGGG	1.32	68.4	84-D and NDV-F (hydrophobic sequence)
8	170-KDSC <u>Q</u>	35-GGSGG	1.60	82.4	near 177-S and near 40-H
9	170-KDSCQ	84-DGGGG	1.60	67.5	near 177-S and 84-D
10	170-KDSCQ	177- <u>S</u> GGGG	1.60	82.4	near 177-S and 177-S
11	170-KDSCQ	v6- <u>R</u> GGGG	1.60	80.4	near 177-S and NDV-F (cleavage site)
12	170-KDSCQ	v13- <u>G</u> SGGG	1.60	82.4	near 177-S and NDV-F (hydrophobic sequence)

Table IV. Intramolecular interaction ($DD > 65$) of the model protein, shown in Fig. 3. Cases with large *DD* values are shown. The active sites and cleavage site are double-underlined

Site-B	Site-E	<i>DEV</i>	<i>DD</i>	250/ <i>DD</i>
5-GHCYK	1-GGSGG	1.21	72.1	3.5
5-G <u>H</u> CYK	27-SGGGG	1.21	72.1	3.5
5-G <u>H</u> CYK	v10-v17 (all are SG ₄)	1.21	72.1	3.5
6-HCYKS	v6-RGGGG	2.05	83.8	3.0
11-TGNN <u>D</u>	1- <u>G</u> SGGG	1.53	69.8	3.6
11-TGNN <u>D</u>	27-SGGGG	1.53	69.8	3.6
11-TGNN <u>D</u>	v10-v17	1.53	69.8	3.6
20-KDSC <u>Q</u>	v6-RGGGG	1.92	88.7	2.8
v2-RRQRR	1- <u>G</u> SGGG	4.16	70.4	3.6
v2-RRQRR	27-SGGGG	4.16	70.4	3.6
v2-RRQRR	v10-v17	4.16	70.4	3.6
v7-v13	20-KDSCQ	1.30	82.6	3.0

model). The jellyfish, a weak swimmer, is usually bobbing on or near the water's surface. Swimming is the result of the regular contraction of subumbrellar muscles which are controlled by the sensory organs at the margins. After contraction, the muscles relax due to the elasticity of the jelly. The jellyfish possesses extensive marginal tentacles which if touched can cause painful stings to man. Both jellyfish and man are independent entities until, for example, the man's hand meets the jellyfish. A type of interaction suddenly takes place between them (the jellyfish and the man's hand). During the investigations on trypsin catalysis by the DEV model, we perceived an analogy between trypsin and jellyfish if we assume that jellyfish stinging cells correspond to the active sites of trypsin and the sting mechanism corresponds to the electron movement from 177-S of trypsin to the cleavage

site of NDV-F (which corresponds to the man's hand portion). In detail, the formation of the transition state from the original form in trypsin catalysis which is mainly operated by mechanical energy⁶⁾ can be perceived to be like the regular contraction of jellyfish muscles. Visualized information transfer shown in Fig. 3(a) in the previous report,⁵⁾ i.e., via a flexible structure in which dynamic water movement facilitates information transfer can be considered analogous to a jellyfish swimming in the sea. Actually, in terms of *DD*, the HD-structure of trypsin is a rather flexible one which stimulates the idea of this analogy.

Why is it that the functional system in the polar sequence of trypsin could not be assigned? Fig. 2(a) shows the polar sequence of trypsin in which the most probable functional site-B, having a maximum *DEV/DD* value, and its site-Es, giving two largest *DD* values to the site-B, are obtained. The result, shown in Table II(a), indicates that there are two site-Bs having the maximum *DEV/DD* (=30) in this molecule: 79-NTGNN-83 and 144-SDSSC. The former is a close neighbor of the active site, 84-D (as in the natural sequence¹⁾). However, no site-Es can be found near the active site, 40-H. Thus the functional system cannot be assigned. Table II (b) shows that when the region containing the cleavage site of NDV-F (Fig. 2(b)) is added to the tail of the polar sequence of trypsin (Fig. 2(a)), the site B has been changed to the cleavage site region (v2-RRQRR, in which the last R correspond to the cleavage site, 91-R in the original NDV-F). Site-Es for this new site-B contains (35-39) and (177-181) (underlined in the table) around which there are 40-H and 177-S, respectively. Furthermore, there is the site-B2, (79-83), which appeared in (a), and is one of the two site-Bs having the maximum *DEV/DD* (=30). In this case, the site-E1 for

Fig. 4. Two possible structures of the model protein (Fig. 3).

this site-B2 contains (35–39) and (177–181). (The site-B, (144–148), is disappeared because of the low *DEV/DD* in this molecule (b).) The same result is obtained as above when only v-RRQRR is added to the tail of the polar trypsin. Now, the assignment of the functional system of trypsin can be accomplished when a target site in the substrate contacts with trypsin in the polar sequence. It suggests that the function of trypsin is recognized in the presence of its substrate. This idea may offer another possible approach for the assignment of functional site of the protein. Further information is obtained from HD-structure. Intramolecular interaction of the above model protein is shown in Table III in which only site-Bs around active sites of trypsin are described. Site-Bs and their site-Es are combined to make a model protein (Fig. 3). Intramolecular interaction of this model protein is searched (Table IV) from which wire and rod model is made (1 cm/amino acid for wire and 250/DD cm for rod). Two possible models are shown in Fig. 4. Compact HD-structure can be made and the involvement of physico-chemical energy is now possible for consideration.

To summarize the results so far, prediction of the active sites of trypsin can be achieved as follows. (i) By natural (or non-polar) sequence of trypsin to get its functional system, through which, after rewriting in the polar sequence, active sites can be predicted. (ii) By polar sequence of trypsin to obtain one of the active sites (84-D). Then, after adding the cleavage site of the substrate, another two active sites (40-H, 177-S) can be predicted. This result suggests that our understanding of trypsin function can be accomplished with its substrate in which flexible structures of trypsin and its substrate encounter by chance in a very sophisticated way to make the reaction complex.

This leads us to two somewhat complex problems: (i) “What is the chance of random encounters?” This is one of the biggest mysteries in science. From this study, we feel that now we are one step closer to an answer. (ii) “What is the artificial sequence of trypsin?” Although it is a kind of abstract strategy that scientists are familiar with, we would like to ask “whether these polar or non-polar sequences really exist or not.” Anyhow, we need a kind of hypothetical regulation mechanism to further advance our understanding of this complex problem.

Acknowledgments. We would like to say many thanks to Dr. Y. Okada, M. J. A., for the helpful discussion and invaluable assistance that he has always given.

References

- 1) Kiho, Y. (1995) Proc. Japan Acad. **71B**, 45–50.
- 2) Kiho, Y. (1995) Proc. Japan Acad. **71B**, 51–56.
- 3) Kiho, Y., Miyata, K., Ubasawa, A., Bhandari, G., Mizuno, Y., Iwai, T., and Okada, Y. (1995) Proc. Japan Acad. **71B**, 244–249.
- 4) Kiho, Y., Bhandari, G., and Okada, Y. (1996) Proc. Japan Acad. **72B**, 11–15.
- 5) Kiho, Y., Katori, K., Ohshima, K., and Okada, Y. (1996) Proc. Japan Acad. **72B**, 191–196.
- 6) Kiho, Y., Iwai, T., Migita, T., Okada, Y., Okabe, Y., Ochi, M., and Katori, K. (1996) Proc. Japan Acad. **72B**, 95–100.
- 7) Kiho, Y., Iwai, T., and Hida, T. (1996) Proc. Japan Acad. **72B**, 85–90.
- 8) Kiho, Y., Haga, T., and Oshima, K. (1996) Proc. Japan Acad. **72B**, 91–94.
- 9) Kiho, Y., Ubasawa, A., Okada, Y., and Iwai, T. (1996) Proc. Japan Acad. **72B**, 185–190.