Application of the Jellyfish Model

II. Higher Dimensional Structure of Trypsin

By Yukio KIHU,^{*,†)} Christopher PALMER,^{*} Kaoru KATORI,^{**)} and Tsuruji IWAI^{***)}

(Communicated by Yoshio OKADA, M.J.A., Oct. 13, 1997)

Abstract: How information transfer occurs in trypsin in a water environment, is analyzed by the DEV model. Three possible trypsin structures, each being flexible, are involved, from which a dynamic cycle under biological regulation involving a trypsin substrate and surrounding water is proposed.

Key words: Trypsin; polar/non-polar sequence; protein folding; information transfer.

Introduction. In our previous report,¹⁾ information transfer in a protein was compared with that in an electric circuit. Both systems have a similar basic structure, that is a linear sequence of amino acids in a protein and a linear line of metal conductor. However, each system is quite distinct in their usage of an insulator to achieve their purpose. In the electric circuit, water must be excluded, so the conductor line is enclosed by an insulator. In the case of protein, there are two kinds of elementary unit (amino acid): one in a polar amino acid and the other is a non-polar amino acid. The former acts like a conductor and the latter corresponds to an insulator and they are arranged hand in hand as the primary amino acid sequence. Investigation on how the amino acids are arranged in the primary sequence led us to make the DEV model from which assignment of functional sites and of the functional system of trypsin and other proteins,^{2)–4)} and finally the active sites of trypsin which are experimentally known, are predicted by the DEV model.⁵⁾

As schematically shown in Fig. 3 of the previous report,¹⁾ because of the intervention by hydrophobic amino acids in the linear arrangement of amino acids including hydrophilic ones, nonlinear information transfer may be considered. In reality, by making the higher dimensional (HD) structure which is a kind of protein folding, two sites (site B and site E in our usage) can approach each other closely even though they are far apart in the primary amino acid sequence. Fortunately, the DEV model could assign functional sites from these sites B and E, in which *DD* was

used as an inverse measure of the spatial distance between them.²⁾ Then *DD* was used to construct the HD model of protein.⁴⁾ The resulting structure of trypsin was found to be loose (small *DD* values) and is distinct from that obtained by X-ray crystallography of crystallized protein. Furthermore, under normal conditions, trypsin of a loose structure is considered to be surrounded by lots of water molecules. These situations are represented in the Fig. 3 in the previous report.¹⁾ In this report, we discuss “how the information transfer can be achieved in trypsin” under these conditions.

Materials and methods. The basic idea and the method of construction of the HD-structure of the protein are described in our previous reports.^{4),6)} In addition to the natural sequence of trypsin, the polar and non-polar sequences of trypsin were studied from which their HD-structural models were made. Then natural and the polar sequences are shown in previous reports,^{3),9)} the non-polar sequence is shown in Fig. 1 in this report. Although the polar and non-polar sequences are model ones, they often serve a useful purpose in DEV analysis.^{1),5),7)–10)} This is due to the fact that the *DD* values obtained from these sequences are generally higher than those from the natural sequence and then compact HD-structures are obtained from the polar and/or non-polar sequences. In general, these characteristics are common to smaller numbered alphabetical systems including polar and non-polar sequences and the assignment of functional sites can be done more exactly but less specific.⁸⁾ The polar and non-polar sequences, however, are essentially model ones. “How can these sequences be realized in the information transfer in a protein?” This is the theme of discussion related to one mentioned above. Details of the DEV model and the analytical methods can be referred to

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

^{***)} Hirosaki University, Faculty of Humanity, Hirosaki 036, Japan.

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

^{†)} Correspondence to: Y. Kiho.

```

      10      20      30
IVGEEEEGAE EVPEEVELEE GEEFEGGELI
      40      50      60
EEEEVVAAAE EEEEGIEVEL GEEEIEVVEG
      70      80      90
EEEFIEAAES IVEPEFEPEE LEEEIMLIEL
     100     110     120
EEAAELEEEV ESVELPEEEA EAGEMELIEG
     130     140     150
EGEEEEEGEE EPEVLEELEA PILEEEEEEE
     160     170     180
EEPGEIEEEM FEAGFLEGGE EEEEGEEGGP
     190     200     210
VVEEGELEGI VEEGEGEAAE EEPGVVEEVE
     220
EEVEEIEEEI AEE

```

Fig. 1. Primary amino acid sequence of non-polar trypsin.

Polar amino acids in the primary amino acid sequence are replaced by E.

Table I. The *DEV* and *DD* values of site B with site E, obtained from the natural sequence of trypsin

Site B	Site E	<i>DEV</i>	<i>DD</i>	500/ <i>DD</i> (cm)
25	70	1.15	61	8.2
66	189	1.19	64	7.8
93	189	1.19	64	7.8
120	70	1.35	65	7.7
138	46	1.16	63	7.9
138	190	1.16	63	7.9
149	10	1.43	66	7.6
166	70	1.11	62	8.1
180	30	1.33	62	8.1
207	29	1.18	64	7.8
207	119	1.18	64	7.8
CL	29	3.44	51	9.8
CL	119	3.44	51	9.8
CL	184	3.44	51	9.8
Phob	18	1.30	64	7.8
Phob	135	1.30	64	7.8
79	Phob	1.21	65	
218	Phob	1.20	64	

(*DEV*>1, *DD*>60) Location of sites B and E is shown as the number of the central amino acids in each site. The last column indicates the length of the rod which is used for making the model.

the previous report.^{2),7)}

Results and discussion. The *DEV* analysis was performed on three sequences of trypsin. The results are shown in Table I (natural sequence, *DEV*>1 and *DD*>60), Table II (polar sequence, *DEV*>1 and *DD*>80) and Table III (non-polar sequence, *DEV*>1 and *DD*>80). From these data, wire-and-rod models for their HD-structures were made which are shown in Fig. 2 (natural), Fig. 3 (polar) and Fig. 4 (non-polar). Without restriction on the *DEV* and/or *DD* values described above, the number of data is so large that the HD-structure is difficult to construct. Generally, *DD* values are higher in the polar and non-polar sequences than in the natural sequence of the

Table II. The *DEV* and *DD* values of site B with site E, obtained from the polar sequence of trypsin

Site B	Site E	<i>DEV</i>	<i>DD</i>	500/ <i>DD</i> (cm)
40	E	1.57	80	6.3
125	168	1.63	90	5.6
125	181	1.63	90	5.6
146	187	2.28	82	6.1
150	189	1.70	85	5.9
172	E	1.52	83	6.0
197	E	1.70	86	5.8
CL	E	4.25	63	7.9

(*DEV*>1, *DD*>80) E=28, 37, 66, 142, 179. The other details are the same as Table I.

Table III. The *DEV* and *DD* values of the site B with the site E, obtained from the non-polar sequence of trypsin

Site B	Site E	<i>DEV</i>	<i>DD</i>	500/ <i>DD</i> (cm)
3	E1	1.22	85	5.8
28	E2	1.16	86	5.8
42	189	1.29	87	5.7
129	189	1.29	87	5.7
140	E3	1.03	86	5.8
146-150	189	1.29	87	5.7
180	E4	2.52	84	6.0
CL	189	1.29	87	5.7
Phob	42	1.89	90	5.6
Phob	125	1.89	90	5.6

(*DEV*>1, *DD*>80) E1=41, 109, 200. E2=208-217. E3=21, 43, 61, 123, 171-175. E4=42, 125, 148, 172 and CL. The other details are the same as Table I.

protein. Accordingly, a natural HD-structure is looser than the polar and non-polar HD-structures. By making restrictions on polarity, the resulting HD-structure is rough, but still the essential functional aspects of the protein are obtained. (The HD-structure of papain⁴⁾ is a typical example.) This is due to the fact that we have chosen the site Bs having larger *DEV* values (*DEV*>1) and the corresponding site E with the largest *DD* value in the *DEV* analysis.

Looking at the results for Figs. 2, 3 and 4 (These are made with *Lw* (length of the wire) equal to 0.5 cm/amino acid and *Lr* (length of the rod) equal to 500/*DD* cm, although the rod is not shown in the figure for simplicity.), roughly speaking, the HD-structures shown in Fig. 2 (natural) and Fig. 4 (non-polar) have a round shape like a ball. (The 3D model shows this more clearly than the figures.) The radius is larger in Fig. 2 than that in Fig. 4. In contrast, the HD-structure shown in Fig. 3 (polar) is somewhat like a crushed ball and a portion of its interior is

Fig. 2. Natural HD-structural model. The black circles indicate the location of the active sites, 40-H, 84-D, or 177-S. N and C mean the N-terminus and the C-terminus, respectively. The single and double lines indicate for CL and Phob, respectively. A $\sim 90^\circ$ clockwise rotation around the z-axis followed by a $\sim 90^\circ$ clockwise rotation about the x-axis results in the lower figure.

squeezed out. Small black circles in the figures indicate the three active sites of trypsin: 40-H, 84-D, and 177-S. Unfortunately, the *DD* values shown in the tables are not conclusive for making the models. Although some restriction has been made, several possible structures can be obtained. In terms of the direction of the wire in the model, both clockwise and counterclockwise foldings are possible and inevitable for the construction of the HD-structure. For example, three possible structures for the non-polar sequence are shown in Fig. 4. Practically, however, starting from the N-terminus of the molecule, the number of possible structures increases at first but later gradually decreases by subsequent interactions and their *DD* values. Furthermore, the area of triangle HDS is another useful criterion from the viewpoint of information transfer. Then the area of triangle HDS made from the three active sites, can be measured and the result is shown in Table IV. Reflecting the above mentioned image, the area of triangle HDS made from the natural sequence is larger than those obtained from the polar and non-polar sequences and the triangle HDS of minimum area is from the polar sequence of trypsin. From a physico-chemical standpoint, the minimum area of the triangle HDS made from the polar amino acid sequence should facilitate the easy movement of electrons which favor enzymatic

Fig. 3. Polar HD-structural model of trypsin. The black circles indicate the location of the active sites, 40-H, 84-D, or 177-S. N and C mean the N-terminus and the C-terminus, respectively. The single and double lines indicate for CL and Phob, respectively. A $\sim 90^\circ$ clockwise rotation around the z-axis followed by a $\sim 90^\circ$ clockwise rotation about the X-axis results in the lower figure.

Fig. 4. Non-polar HD-structural models of trypsin.

Table IV. Areas of the triangle HDS composing of the three active sites in Figs. 2, 3 and 4

Sequence	Source	Area (cm ²)
Natural	Fig. 2	26.2
Polar	Fig. 3	4.1
Non-polar	Fig. 4	16.2(A), 18.2(B), 22.0(C)

reaction of trypsin to occur.

The key issue is: "how can the information transfer be attained" and "how can the polar and non-polar sequences be used to realize this process". From the above discussion, we assume that both the polar and non-polar sequences are real and we start from the HD-structure derived from the natural sequence (natural HD-structure, Fig. 2) and our goal is the polar HD-structure (Fig. 3). Due to low *DD* values, the natural HD-structure is so loose that ordinary physico-chemical bondings are difficult to operate. Especially in the aqueous environment, for example, the ionic interaction between two polar amino acids is weakened, due to the high dielectric constant of water. These considerations do not support a possibility of direct transfer from the natural HD-structure to the polar one. However, if we introduce water into our consideration, the above problem can be treated by thermodynamics. Energy transformations in a biological system is governed by the following equation:

$$\Delta G = \Delta H - T \Delta S$$

in which ΔG is the change of free energy in a system and its environment, ΔH is the change in its enthalpy and ΔS is the change in its entropy. Most biological processes occur at a constant pressure and a constant temperature T . In the case of a protein, the folding favours a decrease of free energy of the molecule and its surroundings. Thus the interaction will decrease the enthalpy and/or increase the entropy of the system. Practically, positive ΔS is mainly explained by the structural change due to water in the surroundings. As a result, a water-driven-hydrophobic association will occur and we can visualize the transfer from natural HD-structure to the non-polar HD-structure. Here, hydrophobic interaction among the non-polar amino acids is stimulated and the surrounding water is excluded. Figs. 2 and 4 reflect the situation in their *DD* values: the former is a loose while the latter is a more compact structure and its area of triangle HDS is smaller than that of the natural one. This raises the possibility that an ordinary physico-chemical interaction can operate here in non-polar HD-structures. Thus the change in the HD-

structure from Fig. 4 to Fig. 3 becomes worthy of consideration.

As the non-polar HD-structure is more compact than that of the natural one, it is possible that ordinary physico-chemical interactions among nearby amino acids would occur in the non-polar HD-structure. Then it is possible that the polar HD-structure is derived from the non-polar one. This notion is understandable when we compare Fig. 4 with Fig. 3. The area of triangle HDS in the polar HD-structure is smaller than that in the non-polar one, so the biochemical reaction can take place easily. However, the total volume of these structures is not so different from each other, suggesting a gradual change from Fig. 4 to Fig. 3, except in the change in the area of triangle HDS. These data suggest that the transformation from the non-polar HD-structure to the polar one is not so drastic as, for example, an overall change from "hydrophobic outside and hydrophilic inside" to "hydrophobic inside and hydrophilic outside". Actually from Fig. 3, the associated hydrophobic portion seems to be conserved but somewhat squeezed out from the interior of the non-polar HD-structure. Hydration of the hydrophobic portion may operate here like in poly (VPGVG).¹¹⁾ Thus the change in volume from non-polar to polar HD-structure is a rather modest one, that is, smaller than that from natural to non-polar one.

From the above discussion, we theorize that the change from the start (natural HD-structure) to the goal (polar HD-structure) would occur through the non-polar HD-structure. (Such a two-step process has been frequently suggested in our previous reports.^{2),7)} Furthermore, two model sequences, the polar and non-polar sequence can be realized by an introduction of water into the system under consideration. In a water-rich medium, we must think about two phases: natural and polar HD-structures. The former is loose and the latter is compact. Difficulty in conversion between these two structures is overcome by introducing the non-polar HD-structure, and any change may occur under mild conditions. Thus we think that each of these three phases is not the individual entity which is in a stable state, but each one is in flexible state and mutually convertible along the pathway discussed above, depending on water content in the environment. Needless to say, various kinds of fluctuation may be included.

An important comment should be added here, that is—all figures (Figs. 2, 3 and 4) are drawn from the *DD* values obtained in the presence of CL-Phob of NDV-F, cited from the previous report.⁹⁾ Without CL-Phob, the assignment of functional sites of the polar trypsin (40-H, 84-D and 177-S) is difficult. Addition of CL-Phob,

however, made assignment possible and so we postulated the Jellyfish model. The sites interacted with CL-Phob are indicated in the figures. (The small vertical single and double line indicate for CL and Phob, respectively). The *DD* values for these interactions are shown in the tables. Even there, the relationship between the 40-H and 177-S regions is somewhat conserved in the above transformation. (See *Appendix* in the previous report¹⁰⁾). Gradual and mild transformation among three HD-structures described in this report is considered as another example of the Jellyfish model. We would like to extend the model to depict the metastable system in which natural, non-polar and/or polar sequences are involved and these components are dynamically transformable with each other.

Appendix. What is next? After NDV-F is cleaved by trypsin, both components separate from each other and proceed along their own pathway, i.e. trypsin is returned to its original form and NDV goes on to the infection process. When the new substrate molecule is added to trypsin, the above-mentioned trypsin-CL-Phob complex is again formed in which the above mentioned processes are repeated. Here, we would like to appreciate this multi-component and flexible cycle as one of the plausible biological systems *in vivo*. The natural HD-structure in itself is not functional even in the presence of its substrate. This is because the reacting components are separated too far from each other especially in the water environment. To be functional, trypsin must take a more compact form by the removal of water in which hydrophobic amino acids are responsible to make it (non-polar HD-structure). Water-deficiency is also a suitable condition for proceeding to the next reaction and then the polar HD-structure is formed. After the reaction, trypsin must return to its original state for which water should be present, otherwise the trypsin-substrate complex cycle cannot continue. Thus this cycle includes water as an important driving force by which the cyclization is an important strategy for maintaining the biological process. Although the natural HD-structure is functionally inactive, it is an important component as the starting point of this cycle. Inclusion of an inactive but important component in the system is not unusual in biological processes.

In protein synthesis, mRNA is destroyed when its role is complete and further new protein synthesis takes place by newly formed mRNA. Here, the resulting inactive component is removed from the reaction system. However, at least two points are distinct from the ordinary chemical reaction (successive reaction). First, the information on mRNA is stored in DNA. If the latter is included in the reaction system, a successive protein synthesis is

considered to be a biological cycle. Second, the amount of the product (protein) is regulated in a sophisticated way in which environmental conditions including any perturbation added is an important factor in this reaction cycle. Probably, the principle operating here is that the energy required to make the necessary proteins is minimized. Various species-dependent half lives of mRNA, for example the long life in silkworm or the short life in bacteria, reflects these situations. In contrast, a chemist can design the successive reaction simply based on from the expected amount of the product he required.

Although complex, biological regulation has to be sophisticated and mild to economize energy and materials. This is born from the beginning of life itself. Fortunately, due to biological catalysis, vigorous conditions are not necessary in the biological system. Unfortunately, however, it easily suffers from external perturbation. For the two examples above, trypsin is inactivated by a trypsin inhibitor and the half life of mRNA increases by the addition of chloramphenicol (protein synthesis inhibitor).¹²⁾ In both cases, the biological cycle mentioned above would be destroyed, even though the external perturbation may attack somewhat indirectly to the system, sometimes beyond the scientist's expectation. In the first example, the inhibitor is used to destroy trypsin when it is unacceptable. The second work was done almost forty years ago. It suggested the intervention of RNA (later called mRNA) between DNA and protein in the central dogma proposed by Watson and Crick.¹³⁾

Acknowledgments. We would like to express many thanks to Dr. Y. Okada, M.J.A., and Dr. S. Ikeda, Emeritus Professor at Osaka University for the profound discussions from philosophy to science, also to Ms. I. Kichinaka, Japan Information Processing Service Co., Ltd. for the manuscript preparation.

References

- 1) Kiho, Y., Katori, K., Ohshima, K., and Okada, Y. (1996) *Proc. Japan Acad.* **72B**, 191–196.
- 2) Kiho, Y. (1995) *Proc. Japan Acad.* **71B**, 45–50.
- 3) Kiho, Y. (1995) *Proc. Japan Acad.* **71B**, 51–56.
- 4) Kiho, Y., Miyata, K., Ubasawa, A., Bhandari, G., Mizuno, Y., Iwai, T., and Okada, Y. (1995) *Proc. Japan Acad.* **71B**, 244–249.
- 5) Kiho, Y., Bhandari, G., and Okada, Y. (1996) *Proc. Japan Acad.* **72B**, 11–15.
- 6) Kiho, Y., Iwai, T., Migita, T., Okada, 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., Ubasawa, A., Okabe, Y., and Iwai, T. (1996) *Proc.*

- Japan Acad. **72B**, 185–190.
- 9) Kiho, Y., Bhandari, G., Ubasawa, A., Palmer, C., and Ikeda, S.
(1996) Proc. Japan Acad. **73B**, 74–78.
- 10) Kiho, Y., Yamada, R., Okabe, Y., Palmer, C., and Okada, Y.
(1997) Proc. Japan Acad. **73B**, 170–175.
- 11) Urrey, D. W. (1989) In Protein Folding (eds. Gierasch, L. M.,
and King, J.). American Association for the Advancement of
Science, pp. 63–71.
- 12) Kiho, Y. (1960) VIRUS **10**, 93–98 (in Japanese with Summary
written in English).
- 13) Watson, J., and Crick, F. (1953) Nature **171**, 737–738.