

## PREDICTION OF SOME SPATIAL STRUCTURES OF PROINSULIN BY HYDROPATHIC MASS

Rong-Qiao HE\*<sup>1</sup>, Yang Liu<sup>1</sup>, Ying Liu<sup>1,2</sup>

Lab Visual Information Processing, Institute of Biophysics, The Chinese Academy of Sciences, 15 Da Tun Rd, Chaoyang District, Beijing 100101, China<sup>1</sup>. Dipartimento di Fisiologia e Biochimica, Sezione di Biologia Cellulare e Dello Sviluppo Universita' Degli Studi di Pisa<sup>2</sup>.

**Abstract** Hydropathic mass (HM) has been used to display the characterization of hydrophilicity and hydrophobicity of proinsulin. The  $\alpha$ -helix of the B chain (residues 9-19), which is located at the inner part of the crystalline insulin, possesses a strong positive HM. The proteolytic sites of proinsulin for insulin maturation are located in the most negative HM regions. This suggests that they are exposed to the exterior of the molecule, which contributes to the digestion by proteases during insulin maturing.

**Key Words:** proinsulin, insulin, hydropathic mass, conformation, crystal structure

### Introduction

It is well known that insulin is used in the treatment of human diabetes and has become one of the most important therapeutic agents to medicine; it has prolonged countless lives. About 30 years ago, Adams et al. demonstrated the structure of rhombohedral 2 zinc insulin crystals [1], followed by several further reports on the structure with an electron density map which was calculated at a resolution of 2.5 and 1.8 Å in the 1970s [2-3]. After that, the structure of 2 zinc pig insulin crystals at 1.5 Å resolution was described by Baker *et al.* [4]. In 1995, Kaarsholm & Ludvigsen [5] reported a solution structure of the insulin monomer at high resolution determined by NMR. So far, research on the crystalline structure of insulin has been continually carried out for different applications in medicine or pharmaceuticals [6-7]. Although the spatial structure of insulin has been widely studied, little is known about the crystalline or solution structure of proinsulin [8-9].

---

Corresponding Author. Tel. 086 10 64889876; Fax. 086 10 64877837. Email. herq@sun5.ibp.ac.cn  
Abbreviations: hydropathic mass (HM)

Kyte and Doolittle (1982) demonstrated the hydrophobic index of each amino acid and evaluated the hydrophilicity and hydrophobicity of a protein along its amino acid sequence [10]. Although the method is not unique and embodies principles that have long been appreciated, its simplicity and graphic nature make it a very useful tool for evaluation of protein structure. It appears that the hydrophobic mass should be used as a criterion to evaluate the hydrophobic characterization of peptides and it should provide some important information about locations of residues. Proteins from homeobox genes are important to embryo development [11-15], we used hydrophobic mass to analyze divergences among the homeodomains of Hox clusters [16]. The results suggested that Hox genes are correlated to Van Baer's law. This paper is concerned with the prediction of some spatial conformations of proinsulin with hydrophobic mass.

### Materials and Methods

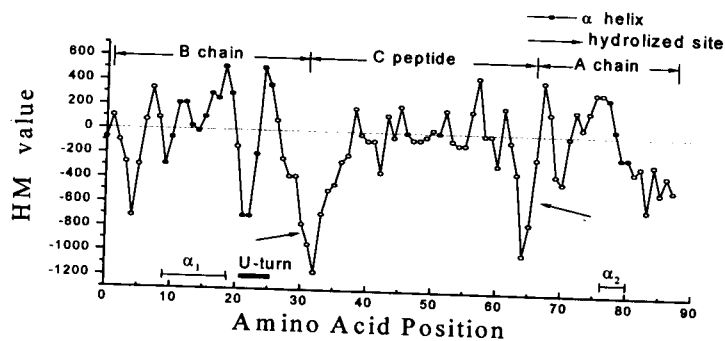
The sequence of the human proinsulin used in this paper was referred to the Protein Data Bank from Brookhaven National Laboratory, USA [17]. The hydrophobic mass of the peptide, which was described before [16], was calculated as follows:

$$\eta = (h_n m_n + h_{n-1} m_{n-1} + h_{n-2} m_{n-2}) \quad (n \geq 1) \quad (1)$$

where  $h$ ,  $m$ ,  $\eta$  and  $n$  represent hydrophobic index, formula mass of side chain, hydrophobic mass in the peptide and position of amino acid residue from N- to C-terminus, respectively. The data were plotted as described previously [16]. As we know, each amino acid residue has chances to interact with other residues of a peptide. However, chances for a residue to interact with the two neighboring residues both up- and down-stream are much more than those with the remoter ones. Therefore, we used the sum of hydrophobic masses of the two vicinal residues to follow such an interaction.

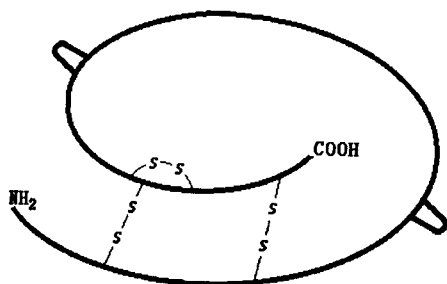
### Results and Discussion

According to Kyte and Doolittle [10], the hydrophobic mass should be such a criterion that can be used for estimation of the location of an amino acid residue in a protein. In general, the larger the criterion of an amino acid residue is, the more chances it approaches to the interior of the molecule, and *vice versa*. Human proinsulin contains 86 residues and its HM is shown in Figure 1. The N-terminal region (residues 1-5) possesses a strong negative HM, and this region should be exposed to the exterior. The  $\alpha$ -helix of B-chain (residues 9-19) is located in a positive HM region, and it should be buried in the molecule. These are coincident with the results of the crystalline X-ray [2-3,18]. Furthermore, the U-turn (Gln-21, Arg-22 and Gly-23) with a strong negative HM is exposed to the exterior. In contrast, the amino acid residues either up- or down-stream with great positive HM may approach to the inner part of the molecule. This difference between hydrophobicity and hydrophilicity



might be one reason for formation of the U-turn. In addition, the  $\alpha$ -Helix of A chain (from A-12 to A-15) may also approach to the inner part of insulin because of its positive HM.

**Fig. 1.** Map of hydrophobic mass of proinsulin. The proinsulin sequence is referred to the Protein Data Bank from Brookhaven National Laboratory, USA (Brousseau et al. 1982). Arrow represents the proteolytic digestion site, when proinsulin is converted into mature insulin.



**Fig. 2.** Diagram of proinsulin structure. The arrow points out the proteolytic site between B chain and C peptide as well as C peptide and A chain. The two bulge regions should be accessible to degradation by proteases *in vivo*.

Mature insulin is formed from its larger precursor proinsulin by proteolytic processing. Removal of a 23-residue signal fragment at the N-terminus of preproinsulin and formation of three disulfide bonds produces proinsulin. Further proteolytic digestion removes the C peptide, *in vivo*, releasing mature insulin containing A and B chains. Now we can see, both digestion sites for A and B chains are in situ at the regions enriched with Arg residues. The C-terminus of the B chain (Thr-30) and the N-terminus of the A chain (Gly-66) are just in situ at the regions where hydrophobic masses are the most negative. We suppose that Thr-30 and Arg-31 as well as Arg-65 and Gly-66 should be exposed at the molecule surface, even though crystal structure of proinsulin has not yet been reported. It is such an exposure that contributes to recognition and digestion of the peptidyl

bonds by the specific proteases in the pancreatic  $\beta$ -cells during insulin maturing (Fig. 2). Digestion of peptidyl bonds by proteases requires the exposure of the specific residues [19]. Exposure of the proteolytic sites of the molecule should be a reason why proinsulin is susceptible to proteases.

#### Acknowledgements

This project is supported in part by the 95-Key Foundation of the Chinese Academy of Sciences (CAS), Foundation of Director of CAS and The National Nature Foundation of China (No. 39970236).

## References

- [1] Adams, M.J., Baker, E.N., Blyndell, T.L., Harding, M.M., Dodson, E.J., Hodgkin, D.C., Dodson, G.G., Rimmer, b., Vuayan M. and Sheat, S. (1969) *Nature*, **224**, 491-495.
- [2] Group of insulin crystal structure. (1972) *Science China*, **1**, 3-20.
- [3] Group of insulin crystal structure. (1974) *Science China*, **6**, 519-611.
- [4] Baker, E.N., Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Hodgkin, D.M., Hubbard, R.E., Isaacs, N.W., Reynolds, C.D. (1988) *Philos. Trans R. Soc. Lond. B Biol. Sci.*, **319**, 369-456.
- [5] Kaarsholm, N.C., Ludvigsen, S. (1995) *Receptor*, **5**, 1-8.
- [6] Jacoby, E., Hua, Q.X., Stern, A.S., Frank, B.H., Weiss, M.A. (1996) *J. Mol. Biol.*, **258**, 136-57.
- [7] Corper, A.L., Stratmann, T., Apostolopoulos, V., Scott, C.A., Garcia, K.C., Kang, A.S., Wilson, I.A. and Teyton, L. (2000) *Science*, **288**, 505-511.
- [8] Kaarsholm, N.C., Ko, H.C. and Dunn, M.F. (1989) *Biochemistry*, **28**, 4427-35.
- [9] Hua, Q.X., Hu, S.Q., Jia, W., Chu, Y.C., Burke, G.T., Wang, S.H., Wang, R.Y., Katsoyannis, P.G. and Weiss, M.A. (1998) *J. Mol. Biol.*, **277**, 103-18.
- [10] Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105-132.
- [11] McGinnis W and Kuziora M. (1994) *Scientific American* **270**, 58-70.
- [12] McGinnis W., Levine M., Hafen E., Kuroiwa A and Gehring W.J. (1984a) *Nature* **308**, 428-433.
- [13] Gehring W.J. (1994) A history of the homeobox, in *Guidebook to the Homeobox Genes* 1st Ed by Duboule D. Oxford Univ. Press, Oxford, pp3-10.
- [14] Hua, Q. (1999) *Progress in Biochem. and Biophys.*, **27**, 199-201.
- [15] Liu, Y. and He, R.Q. (2000) *Progress in Biochem. and Biophys.*, **28**, In press.
- [16] He, R.Q. and Yao, J.L. (2000) *Protein and Peptide Letters*, In press.
- [17] Brousseau, R., Scarpulla, R., Sung, W., Hsiang, H.M., Narang, S.A. and Wu, R. (1982) *GENE* **17**, 279-289.
- [18] Liang, D. C. (1981) *Proteins. Biol. Med.* (U.S.-China Conf.), 25-42.
- [19] Botfield, M. C., Jancso, A., and Weiss, M. A. (1992) *Biochemistry*, **31**, 5841-5848.

Received on May 5, 2000, accepted on May 24, 2000