Volume 231, number 1, 47–50 FEB 05731 April 1988

Localization of the pro-sequence within the total deduced primary structure of human β-hexosaminidase B

John Stirling, Amy Leung[†], Roy A. Gravel[†x] and Don Mahuran[†§]

MRC Human Genetics Disease Research Group, Department of Biochemistry, Kings College London, Kensington Campus, London, England, [†]Research Institute, Hospital for Sick Children, 555 University Ave, Toronto, [§]Department of Clinical Biochemistry and *[x]Department of Medical Genetics, University of Toronto, Ontario, Canada*

Received 25 January 1988

The β subunit of β-hexosaminidase (β-*N*-acetylhexosaminidase, EC 3.2.1.52) is synthesized in the rough endoplasmic reticulum as a prepropolypeptide. After the loss of the signal peptide and formation of an enzymatically active dimer, the pro-enzyme is either secreted from the cell or transported into the lysosome for processing to its mature form. In order to characterize the early posttranslational events we have purified nearly 1 mg of pro-hexosaminidase B from the NH_4Cl containing medium of fibroblasts derived from a patient with the infantile form of Tay-Sachs disease. The partial N-terminal sequence was mapped to a position 42 residues C-terminal to the first in-frame ATG (Met residue) and 79 residues N-terminal to the known mature N-terminus. This position corresponds to that predicted for the cleavage of a 17 amino acid signal peptide generated through the use of the third rather than the first in-frame ATG as the initiation site for protein synthesis.

Lysosome; Structure; Biosynthesis; Posttranslational processing; Signal peptide

1. INTRODUCTION

There are two major isozymes of human lysosomal β-hexosaminidase (Hex), Hex A, $\alpha(\beta_a\beta_b)$, and Hex B, $2(\beta_a\beta_b)$. While the mature α subunit contains a single polypeptide chain, the mature β subunit is composed of two nonidentical chains (β_a and β_b) derived from a common pro-β precursor [1,2]. As is typical of glycoproteins, the components of Hex are synthesized in the rough endoplasmic reticulum (RER) as prepropolypeptides. The pre-sequence, the signal peptide, is cleaved shortly after entry into the RER. The mature subunits are produced through the proteolytic processing of the resulting propolypeptides, already formed into enzymatically active dimers, once they enter the lysosome (review [3]).

We have previously determined the primary structure of the prepro-β polypeptide deduced from the nucleotide sequence of a cDNA clone and established the structures of the β_a and β_b components of the mature β subunit by N-terminal sequencing of human placental Hex B [4]. However, the sequence of the signal peptide (the pre sequence) and its cleavage site (the N-terminus of the pro-β chain) were not determined experimentally. They were predicted from von Heijne's 'rules' established from a compilation of published signal sequences and cleavage sites [5,6]. An underlying assumption in these rules is that the first in-frame ATG is the site of initiation of protein synthesis. There are, however, precedents for the utilization of downstream ATGs [7]. In the case of the β subunit there are three candidate ATG sequences. These code for three M (Met) residues N-terminal to the experimentally determined N-terminus of the mature β_b chain (localized to the N-terminal half of the pro-β chain) [4]. In order to extend our experimental data to more N-terminal sequences, we have examined the N-terminus of the pro-β

Correspondence address: D. Mahuran, Research Institute, Hospital for Sick Children, 555 University Ave, Toronto, Ontario M5G 1X8, Canada

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

polypeptides of pro-Hex B secreted from cultured fibroblasts. This sequence corresponds to the site of cleavage of the signal peptide and provides a more focussed prediction of the initiator M residue and sequence of the signal peptide.

2. MATERIALS AND METHODS

2.1. *Isolation of pro-Hex B*

The method of Creek et al. [8] was used for the isolation of near mg amounts of purified pro-Hex B. Briefly, fibroblasts (11490 cm^2 roller bottles) from a patient with the classical, infantile form of Tay-Sachs disease were grown to confluency in supplemented minimal essential medium, α-MEM, containing 10% fetal calf serum (like other such patients, cells from this patient do not contain any detectable α chain mRNA [9,10]). The medium was replaced with supplemented [8] serum-free α-MEM containing 250 μg/ml human serum albumin (Hex free) and NH_4Cl (10 mM), the latter to induce the secretion of Hex. This medium (45 ml/flask) was collected and replaced daily for three weeks. The pro-Hex B in the pooled medium was purified by a single step procedure utilizing the specific Hex-affinity column (80 ml) previously described for the isolation of Hex A and B from human placenta [1,11].

2.2. *Characterization of the pro-β polypeptide chain*

The fractions of the pH 8.5 eluant from the affinity column were assayed for Hex activity [1]. Active fractions were pooled and concentrated to 5 ml using a 50 ml Amicon pressure cell fitted with a YM-10 membrane. The concentrated protein was dialyzed against 0.1 M NH_4Cl and freeze-dried. The freeze-dried powder was dissolved in 200 μl of H_2O and separated into two samples of 10 and 190 μl. Each sample was redried. The smaller sample was dissolved in sample buffer [12] containing 3% SDS and 25 mM dithiothreitol. It was heated to 60°C for 20 min. The M_r of the reduced polypeptide chain was determined using the standard Laemmli SDS-PAGE system [12]. The partial N-terminal amino acid sequence of the larger sample (15 residues) was determined in a Beckman 890C sequenator. Phenylthiohydantoin derivatives of the resulting amino acids were identified by HPLC [13].

The probability score 'S' predicting the presence of a signal peptide and the position of its cleavage site was calculated using a weighted matrix method of von Heijne [6]. A computer program based on this method for use with the Apple Macintosh was kindly provided to us by Michael Richards.

3. RESULTS AND DISCUSSION

3.1. *Purification of pro-Hex B*

The amount of pro-Hex B necessary to obtain a reliable partial N-terminal sequence was obtained using the method of Creek et al. [8]. Based on activity measurements and the specific activity of the purified placental enzyme [1], approx. 1.2 mg of pro-Hex B was contained in the pooled medium from the Tay-Sachs fibroblasts. About 50% of this was recovered in a pure form from the Hex affinity column.

3.2. M_r *of pro-Hex B*

An analytical sample ($\cong$25 μg) of the pooled protein eluted from the Hex-affinity column was analyzed by SDS-PAGE. The reduced sample produced a single major band corresponding to an M_r of 65000 (fig.1). This is similar in size to the $\cong$60 kDa species previously reported for the secreted, pro-form of the β polypeptide chain from

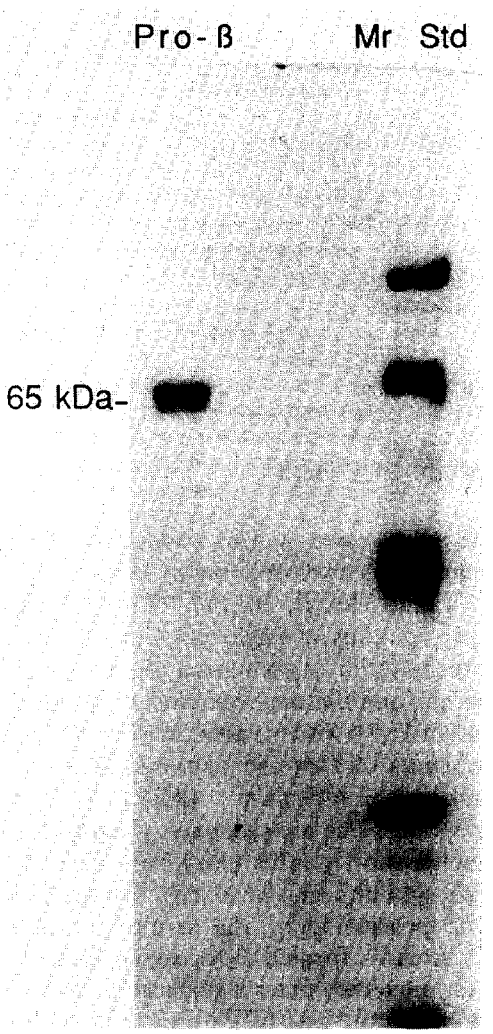

Fig.1. SDS-PAGE of reduced samples of affinity purified pro-Hex B (pro-β) and the proteins (94, 67, 30, and 20.1 kDa) from the Pharmacia 'Low molecular mass standard' kit (kDa standard). The resulting 10% gel was stained with Coomassie blue.

human fibroblasts isolated by immunoprecipitation [14,15], demonstrating that other, more processed forms of the enzyme were not secreted in appreciable amounts during the course of the experiment. Further, the use of mutant fibroblasts (Tay-Sachs) that do not synthesize the α subunit also assured that only the pro-β polypeptide dimer, pro-Hex B, would be secreted.

3.3. *The partial N-terminal sequence of the pro-β polypeptide chain*

The first 15 turns from the sequenator were analyzed and a single major sequence found, ARAPSVSAKPGPALW. This sequence corresponded exactly to residues 43–57 (using the first in-frame M residue as number 1 [4]) in the total deduced amino acid sequence of the prepro-β chain (fig.2). Thus, the secreted form of the enzyme is defined by a single polypeptide species with the N-terminus A (Ala) 43 residues from the first in-frame M residue of the deduced amino acid sequence of the prepro-β chain. We suggest that this site corresponds to the site of cleavage of the signal peptide. The removal of an additional 79 amino acids (residues 43–121) is required to generate the N-terminus of the β_b chain of the mature lysosomal β subunit [4].

It was unexpected to find the N-terminus of the pro-β chain so far carboxyl to the first in-frame M residue. Either the signal peptide corresponds to an exceptionally long 42 amino acids or one of two ATGs more 3′ is the initiation codon. Since most of the signal peptides compiled by von Heijne consist of 16–26 residues and the longest contained 30 [5], the latter explanation seems the more likely. Examination of the other two 3′ ATG codons shows that if the second ATG (M, residue 13, fig.2) were used it would result in a 30 residue signal peptide. In the case of the third ATG (M, residue 26, fig.2) a 17 amino acid signal peptide would be generated. Thus, initiation at the third ATG would best fit the normal length expectation.

Further data supporting the above hypothesis come from the observation that the signal peptides found by von Heijne to have abnormal lengths were devoid of permissible signal sequences within the first 16–26 residues [5]. A previously reported examination of residues 1 to 35 identified several permissible signal sequences that could be associated with the initiation at the first [4] or second ATG codons (unpublished). Thus, it would be unprecedented to utilize either of these ATGs and have cleavage at the observed site.

The method used for the above analysis was von Heijne's weighted matrix calculation [6]. This utilizes a moving window that calculates the probability, S, that a given string of 15 amino acids could serve as a signal peptide. If the observed N-terminus of the pro-β chain resulted from the action of the signal peptidase on the pre-β sequence initiated from the third ATG codon, then this series of calculations should produce a peak value

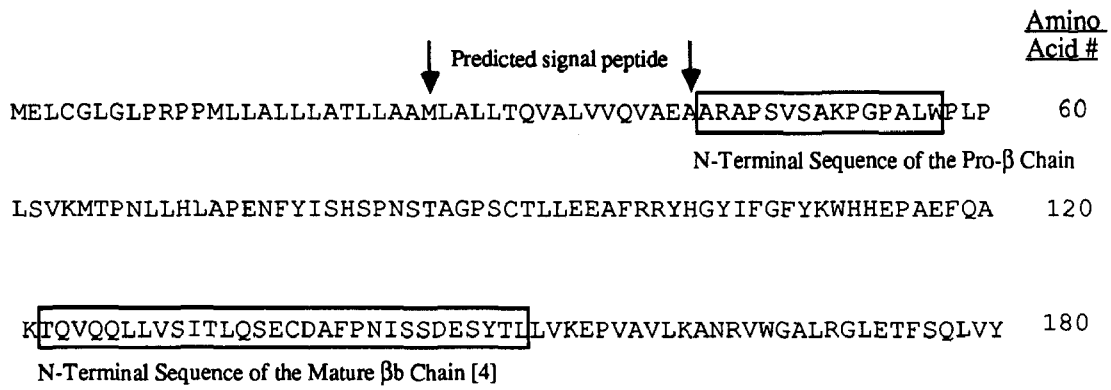

Fig.2. Localization of the pro-β (A, residue 43) and mature-β N-termini (T, residue 122) within the partial deduced amino acid sequence from the 5′ end of a cDNA clone coding for the prepro-β polypeptide chain of Hex [4]. The deduced sequences corresponding to the experimentally determined partial N-terminal sequence of each form of the β chain are indicated by boxes. The area between the putative initiation codon (M, residue 26) and the unique cleavage site (A, residue 42) for a signal peptide predicted using von Heijne's rules [6] is indicated by arrows.

for S (which must be >3.4 to be significant) when the moving window places residue 42, A, and residue 43, A, at the cleavage site of the putative signal sequence. We therefore examined the residues from the third M, residue 26, to the known N-terminus of the mature protein, residue 122 (T, fig.2). An S value of 6.4 was generated when the above A-A sequence was aligned in the hydrolysis site. This was the only position of the window at which a significant S value was obtained. Therefore, the observed pro-β-N terminus is coincident with the unique cleavage site predicted in a signal peptide obtained using the third ATG codon as the initiation site for protein synthesis.

Another criterion for the prediction of the initiation site is based on a consensus sequence surrounding the initiation codon [7]. The sequence is CC{A or G}CCATG. While more than 50% of 211 sequences analyzed match three or four of the nucleotides 5′ to the ATG, only 6 failed to have a purine at the −3 position [7]. The nucleotides 5′ to the first three in-frame ATGs are, respectively: (1) CGGCC*ATG*; (2) CGCCC*ATG*, and (3) CGGCG*ATG*, where matches to the consensus sequence are underlined. Thus, only the first and third sites fulfill the '−3′ rule'.

In conclusion, this report has clearly identified the N-terminus of the secreted pro-form of the β polypeptide chain and localized it within the total deduced primary structure of the prepro-polypeptide, completing the definition of the primary structure of the pro-β chain. It has also revealed an interesting problem related to the pre sequence. The data, based on von Heijne's [6] and Kozak's [7] rules, suggest that the third rather than the first ATG codon is the initiation site for protein synthesis. Further work will be necessary to fully substantiate this conclusion.

Acknowledgements: We wish to thank William Sly for his suggestion of the procedure we used to isolate the large amount of pro-Hex B needed for sequence analysis. We would also like to thank C. Hew, M. Whiteside, and S. Joshi for determining the amino acid sequence of the pro-enzyme. This work was supported by Medical Research Council of Canada Grant PG-4 and a travel grant to J.S. from Wellcome Trust, UK.

REFERENCES

[1] Mahuran, D.J. and Lowden, J.A. (1980) Can. J. Biochem. 58, 287–294.
[2] Mahuran, D.J., Tsui, F., Gravel, R.A. and Lowden, J.A. (1982) Proc. Natl. Acad. Sci. USA 79, 1602–1605.
[3] Von Figura, K. and Hasilik, A. (1986) Annu. Rev. Biochem. 55, 167–193.
[4] Mahuran, D.J., Neote, K., Klavins, M.H., Leung, A. and Gravel, R.A. (1988) J. Biol. Chem., in press.
[5] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–27.
[6] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[7] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[8] Creek, K.E., Fischer, D. and Sly, W. (1983) Methods Enzymol. 98, 290–300.
[9] Korneluk, R.G., Mahuran, D.J., Neote, K., Klavins, M.H., O'Dowd, B.F., Tropak, M., Willard, H.F., Anderson, M.-J., Lowden, J.A. and Gravel, R.A. (1986) J. Biol. Chem. 261, 8417.
[10] Myerowitz, R., Piekarz, R., Neufeld, E.F., Shows, T.B. and Suzuki, K. (1985) Proc. Natl. Acad. Sci. USA 82, 7830–7834.
[11] Geiger, B., Ben-Yoseph, Y. and Arnon, R. (1974) FEBS Lett. 45, 276–281.
[12] Laemmli, U.K. (1970) Nature 227, 680.
[13] Peckett, M., Scott, G., Davies, P., Wang, N., Jacki, S. and Hew, C. (1984) Eur. J. Biochem. 143, 55–58.
[14] Hasilik, A. and Neufeld, E.F. (1980) J. Biol. Chem. 255, 4937–4945.
[15] Tsui, F., Mahuran, D.J., Lowden, J.A., Mosmann, T. and Gravel, R.A. (1983) J. Clin. Invest. 71, 965–973.