

Establishment of a heterologous system for the expression of *Canavalia brasiliensis* lectin: a model for the study of protein splicing

Waldery Melgaço Bezerra¹, Cristina Paiva da Silveira Carvalho², Renato de Azevedo Moreira³ and Thalles Barbosa Grangeiro¹

¹Laboratório de Citogenética e Genética Molecular, Departamento de Biologia, Bloco 906, Centro de Ciências, Universidade Federal do Ceará, Av. Humberto Monte, s/n, 60455-970 Fortaleza, CE, Brasil

²Centro de Ciências da Saúde, Universidade Estadual do Ceará, Av. Paranjana, 1700, 60740-000 Fortaleza, CE, Brasil

³Departamento de Bioquímica e Biologia Molecular, Bloco 907, Universidade Federal do Ceará, Fortaleza, CE, Brasil

Corresponding author: T.B. Grangeiro

E-mail: thalles@ufc.br

Genet. Mol. Res. 5 (1): 216-223 (2006)

Received January 10, 2006

Accepted February 17, 2006

Published March 31, 2006

ABSTRACT. During its biosynthesis in developing *Canavalia brasiliensis* seeds, the lectin ConBr undergoes a form of protein splicing in which the order of the N- and C-domains of the protein is reversed. To investigate whether these events can occur in other eukaryotic organisms, an expression system based on *Pichia pastoris* cells was established. A DNA fragment encoding prepro-ConBr was cloned into the vector pPICZB, and the recombinant plasmid was transformed in *P. pastoris* strain GS115. Ten clones were screened for effective recombinant protein production. Based on Western blot analysis of the two clones with the highest level of protein expression: 1) diffuse high-molecular mass immunoreactive bands were produced as early as 24 h after induction; 2) a single-, high-molecular mass protein was secreted into the medium, and 3) a significant fraction of the recombinant poly-

peptides that cross-reacted with anti-ConBr antibodies comprised a band of approximately 34.5 kDa. Diffuse protein bands with high molecular masses are attributed to hyperglycosylation at the single potential N-glycosylation site located in the linker peptide of prepro-ConBr. In contrast, native ConBr is made up of three polypeptides, the intact α chain (aa 1-237) and the fragments β (aa 1-118) and γ (aa 119-237), which have apparent molecular masses of 30, 16 and 12 kDa, respectively. Apparently, the yeast *P. pastoris* is not able to carry out all the complex post-translational proteolytic processing necessary for the biosynthesis of ConBr.

Key words: *Canavalia brasiliensis* lectin, Protein splicing, Heterologous, Recombinant, *Pichia pastoris*

INTRODUCTION

Lectins are a structurally heterogeneous group of carbohydrate-binding proteins comprising distinct families of evolutionarily related proteins (Peumans et al., 2001). The most thoroughly investigated lectins are those from plants, in particular the Leguminosae. Legume lectins are a large group of structurally similar proteins with distinct carbohydrate specificities. Concanavalin A (ConA), found in the seeds of *Canavalia ensiformis*, was the first lectin to be sequenced (Cunningham et al., 1975; Wang et al., 1975) and to have its three-dimensional structure determined by X-ray crystallography (Becker et al., 1975; Hardman and Ainsworth, 1976). The many biochemical, biophysical, and structural studies carried out in ConA make this protein the best-characterized plant lectin. Indeed, ConA is the prototype of legume lectins.

Protein splicing is a post-translational event, in which the translated protein undergoes cleavage, resulting in the loss of an internal spacer sequence, while the remaining flanking domains are rejoined via a peptide bond (Cooper and Stevens, 1993; Wallace, 1993). *C. ensiformis* lectin is synthesized by a unique mechanism in maturing seeds. Based on information from pulse-chase analysis, cDNA cloning, and amino acid sequencing of the processing intermediates, it is known that ConA undergoes a form of protein splicing in which the order of the N- and C-domains of the protein is reversed (Carrington et al., 1985; Herman et al., 1985; Bowles et al., 1986; Chrispeels et al., 1986; Faye and Chrispeels, 1987).

ConBr, the glucose/mannose-binding lectin found in *C. brasiliensis* seeds (Moreira and Cavada, 1984), has 99% amino acid sequence identity with ConA (Grangeiro et al., 1997). These lectins display the same primary carbohydrate binding specificity for residues of D-glucose, D-mannose and derivatives, and both proteins possess enhanced affinities for the branched chain trimannoside, 3,6-di-O-(α -D-mannopyranosyl)-D-mannose (Dam et al., 1998), which is found in the core region of all asparagine-linked (N-linked) carbohydrates. Furthermore, ConBr and ConA essentially recognize the same set of binding epitopes on the structure of the trimannoside (Dam et al., 2000). The amino acids at positions 58 (glycine in ConBr and aspartic acid in ConA) and 70 (glycine in ConBr and alanine in ConA) are the only residues that differ between ConA and ConBr (Grangeiro et al., 1997; Sanz-Aparicio et al., 1997). Based on a comparison of

the amino acid sequence of the precursor of ConBr (prepro-ConBr), deduced from the cDNA, with the established amino acid sequence of the lectin mature subunit (Grangeiro et al., 1997), the same mechanism of protein splicing observed during ConA biosynthesis is also involved in the production of mature ConBr (Figure 1).

Heterologous expression of prepro-ConBr in the yeast *Pichia pastoris* would help to determine whether the lectin maturation machinery is specific to *C. brasiliensis* or is conserved in other eukaryotic cells. We report here an initial characterization of *P. pastoris* clones transformed with the coding sequence for prepro-ConBr.

MATERIAL AND METHODS

Cell strains, plasmids and cell cultures

P. pastoris strain GS115 cells and pPICZB expression plasmid were obtained from Invitrogen. Medium compounds were purchased from Amersham Biosciences and from Oxoid. All medium formulations and abbreviations are described in the EasySelection *Pichia* Expression Kit manual, version F (www.invitrogen.com). All the cells were stored in 10% (v/v) glycerol at -20°C.

Expression vector construction

A DNA fragment encoding prepro-ConBr was amplified by PCR using *C. brasiliensis* genomic DNA as a template. The oligonucleotide primers were 5' (forward): 5'-GAATTCAAAAATGTCTGCCATCTCAAAGAAATC-3' and 3' (reverse): 5'-GCTCTAGAACACGGTAGCAAT-3'. *EcoRI* (forward) and *XbaI* (reverse) sites were included at the 5' end of each primer. Primer sequences were designed in such a way to place the open-reading frame for prepro-ConBr in frame with the C-terminal tag. In addition, the design of the forward primer placed the start codon ATG as part of a yeast consensus sequence (Romanos et al., 1992). Primer sequences were based on the published sequence of ConBr gene (Grangeiro et al., 1997), which is deposited in GenBank (accession number Y13904). The amplification was carried out in a 25- μ L reaction volume containing 50 ng template DNA, 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 1.5 mM MgCl₂, 200 mM of each dATP, dCTP, dGTP, and dTTP (Amersham Biosciences), 0.5 μ M of each primer and 2.5 units of Taq DNA polymerase (Amersham Biosciences). PCR reactions were carried out in an MJ Research (Watertown, MD) PTC-200 thermocycler programmed for an initial denaturation (5 min at 95°C) followed by 30 cycles of 1 min at 94°C, 1 min at 45°C and 2 min at 72°C. PCR products were purified from 1% agarose gel slices and cloned into the pPICZB plasmid vector (Invitrogen), using the restriction sites incorporated into the primer sequences.

Pichia pastoris transformation and induction of expression of the recombinant lectin

The recombinant plasmid was linearized by restriction with *SacI* (*SstI*) and transformed into *P. pastoris* strain GS115 by electroporation. Transformants were selected on YPD-sorbitol plates containing 200 μ g/mL zeocin (Invitrogen) and incubated at 30°C. The integration of prepro-ConBr coding sequence in the genome of *P. pastoris* was confirmed by PCR, using genomic

DNA samples from 20 individual zeocin-resistant clones as templates and primers specific for the coding sequence, as described in the previous section. Ten clones were then screened for the expression of recombinant lectin on a small scale using flask cultures. A single colony from each clone was grown overnight in 5 mL MGYH medium containing 200 µg/mL zeocin, spun down, resuspended in sterile distilled water and inoculated in 25 mL MMH medium. The cultures were incubated at 30°C with continuous shaking (250 rpm) and fresh 100% methanol was added twice a day to a final concentration of 0.5% (v/v). Culture samples (1 mL) were taken at various stages (0, 24, 48, 72, 96, and 120 h) during cell growth in the presence of methanol, and immediately centrifuged (12,000 g for 10 min). The supernatants and cell pellets were frozen in liquid nitrogen and stored at -20°C, for later analysis by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot.

SDS-PAGE and Western blot analysis

SDS-PAGE was done as described by Laemmli (1970), using 15% slab gels. Western blot analysis (Towbin et al., 1979) was carried out in nitrocellulose membranes (Hybond™ C Extra, Amersham Biosciences, USA) using rabbit anti-ConBr antibodies and goat anti-rabbit IgG conjugated to horseradish peroxidase (Calbiochem, USA). Rabbit antibodies directed against ConBr were prepared by immunization with native ConBr purified by affinity chromatography on Sephadex G-50, as previously described (Moreira and Cavada, 1984). Antibodies specific to ConBr were purified from the serum of immunized animals using a ConBr-Sepharose column. Western blotting procedures followed standard methods (Sambrook et al., 1989).

RESULTS AND DISCUSSION

A PCR product encoding the prepro-ConBr was amplified from *C. brasiliensis* genomic DNA and cloned into pPICZB, a plasmid vector designed for intracellular expression in *P. pastoris*. As shown in Figure 1, the amplified DNA fragment encodes a polypeptide with 290 amino acids (aa), comprising a 29-aa signal peptide, the C-terminal domain (aa 119-237) found in the mature lectin, a 15-aa linker peptide, the N-terminal domain (aa 1-118) of the native protein, and a 9-aa C-terminal extension (Grangeiro et al., 1997). Integration of prepro-ConBr coding sequence in the yeast genome was confirmed by PCR in 20 clones of *P. pastoris* selected in the presence of zeocin. This yielded a single 870-bp PCR product, which is the expected size of the prepro-ConBr coding sequence (data not shown).

Ten transformed colonies of *P. pastoris* were randomly chosen and checked for effective recombinant protein production. Both cells and supernatants from methanol-induced cultures were analyzed by Western blot for each clone. In this screening, clonal variation in the expression of the recombinant lectin was observed, with two of 10 clones showing no detectable expression of rConBr (data not shown). Clonal variation in heterologous expression of *Phaseolus vulgaris* lectin E in *P. pastoris* has also been observed (Baumgartner et al., 2002). Western blot analysis was made of the two clones with the highest ConBr expression level. It was evident that upon induction a protein with an apparent molecular mass of about 34.5-kDa, which reacted with anti-ConBr antibodies, was produced (Figure 2). This immunoreactive band was not detected in the cell culture just before induction. Production of this 34.5 kDa band was detected in the cells 48-72 h after induction with methanol, and it remained until 120 h (Figure 2).

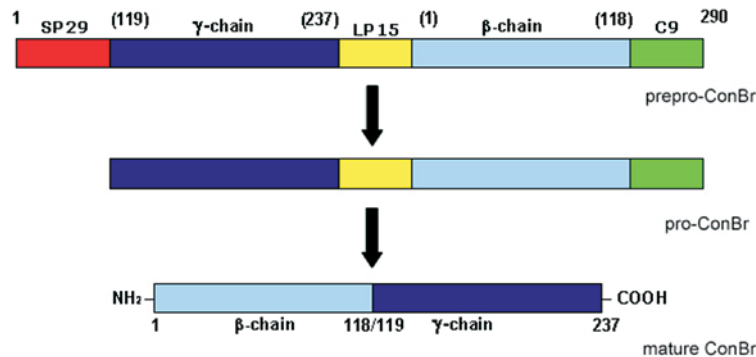


Figure 1. Primary structural organization of prepro-ConBr, pro-ConBr and the mature protein, ConBr, a glucose/mannose-binding lectin from the seeds of *Canavalia brasiliensis*. The prepro-sequence (290 amino acids) is comprised of a 29-aa signal peptide (SP29), the C-terminal domain (aa 119-237) found in the mature lectin (γ -chain), a 15-aa linker peptide (LP15), the N-terminal domain (aa 1-118) of the native protein (β -chain), and a 9-aa C-terminal extension (C9). The precursor of ConBr (pro-ConBr) is cleaved into two distinct chains that are religated by transpeptidation to form the mature lectin (30-kDa α -chain).

In addition, as early as 24 h after induction, multiple immunoreactive bands with higher molecular masses were also detected in the cells of both clones, and they lasted until 120 h (Figure 2). These diffuse bands or smears at higher molecular masses are diagnostic of protein glycosylation (Raemaekers et al., 1999). Indeed, prepro-ConBr contains a single-potential N-glycosylation site located in the 15-aa linker peptide (Figure 1). However, native ConBr purified from *C. brasiliensis* seeds is not glycosylated. In the case of the homologous lectin ConA, it has been shown that the oligosaccharide side chain is removed during the transport of the glycosylated precursor via the Golgi into the protein storage vesicles of cotyledonary cells (Herman et al., 1985; Faye and Chrispeels, 1987). N-linked oligosaccharide side chains found on glycoproteins expressed in *Pichia* are mainly of the high-mannose type, and most of them have 8-14 mannose residues; thus they would be expected to increase the molecular mass of a glycoprotein by approximately 2 kDa (Cregg et al., 1993). Consequently, the high-molecular mass bands that reacted with anti-ConBr antibodies appear to be hyperglycosylated unprocessed ConBr.

A single-, high-molecular mass band that cross-reacted with anti-ConBr antibodies was also detected in the supernatants of the induced cultures (Figure 2). This suggests that ConBr's native signal peptide directed the secretion of a soluble form of the lectin.

In the pPICZB vector, the recombinant lectin is produced as a fusion protein containing C-terminal tags, the *myc* epitope and a polyhistidine (6xHis) sequence, which allow detection of the expressed protein and its purification by immobilized metal ion affinity chromatography, respectively. The *myc* epitope and the polyhistidine tag contribute 2.5 kDa to the size of the expressed protein (Invitrogen). In addition, most of the high-mannose oligosaccharide side chains present in *Pichia* glycoproteins contribute 2 kDa to the mass of the expressed polypeptide (Cregg et al., 1993). Taking these findings into account, the 34.5-kDa polypeptide that is recognized by anti-ConBr antibodies can be explained by the removal of both the native signal peptide as well as the single-N-linked oligosaccharide chain, when prepro-ConBr is expressed in *P. pastoris*. Therefore, the 34.5-kDa band would correspond to pro-ConBr fused with the C-terminal tags. On the other hand, SDS-PAGE of purified native ConBr has shown that the

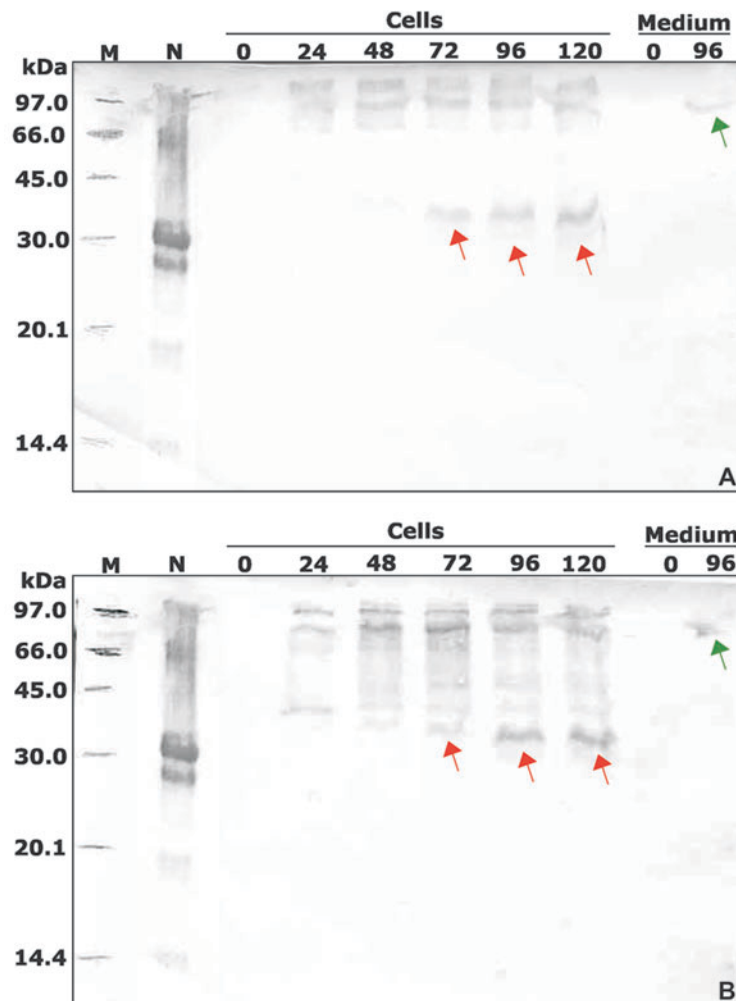


Figure 2. Immunodetection of recombinant ConBr expressed in *Pichia pastoris* GS115. The red arrows indicate a protein with an apparent molecular mass of about 34.5 kDa, which reacted with anti-ConBr antibodies. The cell cultures from two transformed *P. pastoris* clones (A and B) were induced with 0.5% (v/v) methanol and the production of the recombinant protein was analyzed by Western blot. Culture samples were taken at various stages (0, 24, 48, 72, 96, and 120 h) during induction, spun down, and the cell pellets were submitted to SDS-PAGE and Western blot. The culture medium at 0 and 96 h after induction was also sampled and analyzed. Native ConBr (lane N) was included for comparison. The apparent molecular masses of lectin subunits were estimated using standard proteins (lane M): phosphorylase b, 97.0 kDa; bovine serum albumin, 66.0 kDa; ovalbumin, 45.0 kDa; carbonic anhydrase, 30.0 kDa; trypsin inhibitor, 20.1 kDa, and α -lactalbumin, 14.4 kDa (Amersham Biosciences). A single-, high-molecular mass band that cross-reacted with anti-ConBr antibodies was also detected in the supernatants of the induced cultures (green arrows).

mature lectin consists of a major 30-kDa band and small proportions of 16- and 12-kDa fragments (Moreira and Cavada, 1984). N-terminal amino acid sequencing of electroblotted samples, as well as comparisons of the amino acid sequence of the precursor of ConBr (prepro-ConBr) deduced from the cDNA with the known amino acid sequence of the lectin mature subunit (Grangeiro et al., 1997), indicated that ConBr precursor is post-translationally processed by a protein-splicing mechanism. In this mechanism, the precursor of ConBr is cleaved into two

distinct chains that are religated by transpeptidation to form the mature lectin (30 kDa α -chain). Since this process is not quantitative, the 16- and 12-kDa fragments (β - and γ -chains, respectively) are most probably unligated peptides rather than proteolytic products of the 30-kDa polypeptide. Therefore, the difference in the electrophoretic profiles between native ConBr and the recombinant 34.5-kDa polypeptide produced in *P. pastoris*, as revealed by Western blotting, is attributed to the absence of further proteolytic processing of pro-ConBr.

In conclusion, when expressed in *P. pastoris*, prepro-ConBr undergoes a series of transformations, including hyperglycosylation and secretion into the medium. Processing of a fraction of these polypeptides produces a molecular species that has a molecular mass similar to pro-ConBr. No further processing of this precursor was detected. Apparently, the yeast *P. pastoris* is not able to carry out all the complex post-translational proteolytic processing that occurs during the biosynthesis of ConBr.

ACKNOWLEDGMENTS

Research supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP), and Banco do Nordeste do Brasil (BNB).

REFERENCES

- Baumgartner P, Raemaekers RJM, Durieux A, Gatehouse A, et al. (2002). Large-scale production, purification, and characterization of recombinant *Phaseolus vulgaris* phytohemagglutinin E-form expressed in the methylotrophic yeast *Pichia pastoris*. *Protein Expr. Purif.* 26: 394-405.
- Becker JW, Reeke Jr GN, Wang JL, Cunningham BA, et al. (1975). The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. *J. Biol. Chem.* 250: 1513-1524.
- Bowles DJ, Marcus SE, Pappin DJ, Findlay JB, et al. (1986). Posttranslational processing of concanavalin A precursors in jack bean cotyledons. *J. Cell Biol.* 102: 1284-1297.
- Carrington DM, Auffret A and Hanke DE (1985). Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* 313: 64-67.
- Chrispeels MJ, Hartl PM, Sturm A and Faye L (1986). Characterization of the endoplasmic reticulum-associated precursor of concanavalin A. Partial amino acid sequence and lectin activity. *J. Biol. Chem.* 261: 10021-10024.
- Cooper AA and Stevens TH (1993). Protein splicing: excision of intervening sequences at the protein level. *Bioessays* 15: 667-674.
- Cregg JM, Vedvick TS and Raschke WC (1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology* 11: 905-910.
- Cunningham BA, Wang JL, Waxdal MJ and Edelman GM (1975). The covalent and three-dimensional structure of concanavalin A. II. Amino acid sequence of cyanogen bromide fragment F3. *J. Biol. Chem.* 250: 1503-1512.
- Dam TK, Cavada BS, Grangeiro TB, Santos CF, et al. (1998). Diocleinae lectins are a group of proteins with conserved binding sites for the core trimannoside of asparagine-linked oligosaccharides and differential specificities for complex carbohydrates. *J. Biol. Chem.* 273: 12082-12088.
- Dam TK, Cavada BS, Grangeiro TB, Santos CF, et al. (2000). Thermodynamic binding studies of lectins from the diocleinae subtribe to deoxy analogs of the core trimannoside of asparagine-linked oligosaccharides. *J. Biol. Chem.* 275: 16119-16126.
- Faye L and Chrispeels MJ (1987). Transport and processing of the glycosylated precursor of concanavalin A in jack bean. *Planta* 170: 217-224.
- Grangeiro TB, Schriefer A, Calvete JJ, Raida M, et al. (1997). Molecular cloning and characterization of ConBr, the lectin of *Canavalia brasiliensis* seeds. *Eur. J. Biochem.* 248: 43-48.
- Hardman KD and Ainsworth CF (1976). Structure of the concanavalin A-methyl alpha-D-mannopyranoside

- complex at 6-A resolution. *Biochemistry* 15: 1120-1128.
- Herman EM, Shannon LM and Chrispeels MJ (1985). Concanavalin A is synthesized as a glycoprotein precursor. *Planta* 165: 23-29.
- Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.
- Moreira RA and Cavada BS (1984). Lectin from *Canavalia brasiliensis* (Mart.): isolation, characterization and behavior during germination. *Biol. Plant.* 26: 113-120.
- Peumans WJ, Van Damme EJ, Barre A and Rouge P (2001). Classification of plant lectins in families of structurally and evolutionary related proteins. *Adv. Exp. Med. Biol.* 491: 27-54.
- Raemaekers RJM, Muro L, Gatehouse JA and Fordham-Skelton AP (1999). Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris* correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. *Eur. J. Biochem.* 265: 394-403.
- Romanos MA, Scorer CA and Clare JJ (1992). Foreign gene expression in yeast: a review. *Yeast* 8: 432-488.
- Sambrook J, Fritsch EF and Maniatis T (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY, USA.
- Sanz-Aparicio J, Hermoso J, Grangeiro TB, Calvete JJ, et al. (1997). The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological properties from concanavalin A. *FEBS Lett.* 405: 114-118.
- Towbin H, Staehelin T and Gordon J (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* 76: 4350-4354.
- Wallace CJ (1993). The curious case of protein splicing: mechanistic insights suggested by protein semisynthesis. *Protein Sci.* 2: 697-705.
- Wang JL, Cunningham BA, Waxdal MJ and Edelman GM (1975). The covalent and three-dimensional structural of concanavalin A. I. Amino acid sequence of cyanogen bromide fragments F1 and F2. *J. Biol. Chem.* 250: 1490-1502.