DNA MOLECULES OPTIMIZED SEQUENCES THAT ENCODE THE IA-2IC ANTIGEN, RNA MOLECULES, EXPRESSION VECTORS, TRANSFORMED CELLS, METHOD OF PREPARING THE IA-2IC ANTIGEN, POLYPEPTIDE OF THE HUMAN IA-2IC ANTIGEN, AND IN VITRO PROCEDURE AND KIT FOR THE DIAGNOSIS OF AUTOIMMUNE DIABETES

Publication Classification

Int. Cl. .......................... C12Q 1/68; C07H 21/04
U.S. Cl. .................................. 435/6; 536/23.2

ABSTRACT

Optimized DNA molecules that encode human IA-2ic, wherein at least one of the codons starting at positions 2747 and 2816 of the complete IA-2 gene is CGT. Said molecules are useful for detecting insulin-dependent (type 1) diabetes mellitus, and may be fused to auxiliary proteins or peptides (for example, His-tag). RNA molecules encoded by the optimized DNA molecules, expression vectors comprising the optimized DNA molecules and cells transformed with said DNA molecules (particularly Escherichia coli cells). Also a method of producing human IA-2ic antigen that comprises growing cells transformed with the said optimized nucleic acid molecules under conditions suitable for producing said protein, the polypeptide obtained through that procedure, and an in vitro diagnostic method and a diagnostic kit using it.
FIGURE 1

<table>
<thead>
<tr>
<th>NS</th>
<th>Extracellular domain</th>
<th>TM</th>
<th>Intracellular domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td></td>
<td>579</td>
<td>601</td>
</tr>
</tbody>
</table>
Figure 5A

The graph shows the CD spectra of two compounds, IA-2icH₆ and IA-2ic(AD7DG)H₆.

- **IA-2icH₆**
- **IA-2ic(AD7DG)H₆**

The spectra are plotted against wavelength (nm) on the x-axis and [θ] × 10⁻³ (deg cm² dmol⁻¹) on the y-axis.
FIGURE 5 B

\[ [\theta] \text{ (deg cm}^2 \text{ dmol}^{-1}) \]

- IA-2icH\textsubscript{6}
- IA-2ic(AD/DG)H\textsubscript{6}

Wavelength (nm)
DNA MOLECULES OPTIMIZED SEQUENCES 
THAT ENCODE THE IA-2C ANTIGEN, RNA 
MOLECULES, EXPRESSION VECTORS, 
TRANSFORIZED CELLS, METHOD OF 
PREPARING THE IA-2C ANTIGEN, 
POLYPEPTIDE OF THE HUMAN IA-2C ANTIGEN, 
AND IN VITRO PROCEDURE AND KIT FOR THE 
DIAGNOSIS OF AUTOIMMUNE DIABETES

[0001] The present invention relates to the production of antibodies useful for the immunodiagnosis of diabetes mellitus. More particularly, the invention relates to optimized DNA sequences that encode the IA-2c antigen.

[0002] Insulin-dependent (type 1) diabetes mellitus is characterized by the autoimmune destruction of insulin-producing P cells. During the prodrome of type 1 diabetes, a variety of autoantibodies, called markers, is generated, some of which persist long after the complete development of the disease, and which are a valuable aid for its early detection [1-4]. The best characterized markers are those directed against glutamic acid decarboxylase (GAD) [2, 4-6], tyrosine phosphatase IA-2 (also known as ICA512) [7, 11], and insulin [12]. Reliable and highly reproducible radioligand binding assays (RBA) have been developed for detecting these markers.

[0003] About 70% of type 1 diabetic patients have GAD markers (GADA) [6], whereas the percentage of positivity for the anti-IA-2 marker (IA-2A) is somewhat less [20]. The percentage of patients having at least one of these two markers is about 90% [15], making these tests very valuable tools for both basic research and practical medical intervention [14].

[0004] IA-2, a 106-kDa protein similar to members of the tyrosine-phosphatase family (PTPases), is a 979 amino-acids membrane protein with intracellular (C-terminal), extracellular (N-terminal), and transmembrane (residues 579-603) domains, and a 34-residue N-terminal signal sequence [21] (see FIG. 1). IA-2 is expressed in cells of the neuroendocrinial lineage as pituitary neurons and β cells of the pancreatic islet [22, 23]. After various post-translational processing, including extensive proteolytic cleavage and glycosylation, IA-2 is transformed into several products with apparent molecular weights ranging 40 to 130 kDa [21]. In addition, the messenger RNA of IA-2 is spliced, generating a form that lacks the exon encoding for the transmembrane domain [24].

[0005] IA-2 lacks tyrosine-phosphatase activity [19]. However, it can be artificially rendered active by replacing residues Ala 877 and Asp 911 by Asp and Ala, respectively [25]. The lack of enzymatic activity poses the question as to what is the biological function of the protein. To this respect, it was recently proven that IA-2-gene deficient mice have depressed insulin release in glucose tolerance tests [26].

[0006] Most IA-2 epitopes are located in the intracytoplasmic domain (residues 604 to 979) [27-29], which comprises the juxtamembrane area (residues 604-670) and the PTPase-like domain (residues 670-979). The most widely used antigen constructions for the detection of IA-2A are ICA512bdc (which includes residues 236-556 and 630-979, and which lacks amino-acids spanning 557 to 629 [30]), and IA-2c (residues 604-979), which comprises the entire cytoplasmic portion [8]. Sensitivity of IA-2c for IA-2A detection is slightly higher than that of ICA512bdc, due to the absence in the latter of epitopes comprised in the juxtamembrane portion [24].

[0007] Because most IA-2A are directed against discontinuous epitopes [31], it is essential that the antigen used in the diagnostic test maintains its native conformation. Native IA-2c has been produced in prokaryotic systems, as a fusion protein with, on the one hand, a N-terminal peptide that carries a biotinylation site [32, 33] or, on the other, with glutation S-transferase (GST) [25, 31]. The yield of properly folded IA-2c using these constructions was very low; furthermore, antibody accessibility to the N-terminus of IA-2A might be compromised by the fusion with these rather large polypeptides.

[0008] There is therefore a need of practical methods for the large-scale production of autoantigens that would serve in the diagnosis of type 1 diabetes. The present invention fulfills this need by providing an inexpensive method of producing IA-2c as a recombinant protein in Escherichia coli with high yield.

DESCRIPTION OF THE INVENTION

[0009] The present invention comprises novel nucleic acid molecules designed to improve the expression of IA-2c, a fundamental antigen in autoimmune diabetes, necessary for IA-2 autoantibodies detection in prokaryotic systems.

[0010] In one aspect, the present invention refers to DNA molecules comprising optimized nucleotide sequences. The most important optimized sequences are SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and parts of them.

[0011] As another object, the present invention provides a chimeric DNA molecule comprising any of the sequences SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or parts of them, fused with sequences encoding an auxiliary peptide.

[0012] Also constitutes an object of the invention DNA molecules encoded by the optimized DNA molecules, either alone or fused to one or more auxiliary peptides.

[0013] In another object of the invention, expression vectors are provided that comprise the DNA molecules of the present invention. In them, the DNA molecules of the invention are within the context of one or more regulatory sequences that direct the expression of the IA-2c antigen in bacterial cell. Preferably, the expression vectors provided are derived from the pET plasmid (Novagen).

[0014] A host cell transformed with DNA molecules like those of the present invention constitutes another object of the present invention. The transformed cell is preferably a prokaryotic cell and, particularly Escherichia coli.

[0015] Also constitutes another aspect of the present invention a method of producing IA-2c comprising growing of cells transformed either by a DNA molecule like those of the invention, or by an expression vector like those of the invention, under suitable conditions for the production of said polypeptide. In a preferred embodiment, the transformed cells are of Escherichia coli.

[0016] The products encoded by the DNA molecules of the present invention and produced using the methods of the present invention are also part of the invention, as well as a procedure for in vitro diagnosis of autoimmune diabetes that
comprises contacting a sample of the patient’s body fluids with the product encoded by the DNA molecules of the invention. Finally, another object of the present invention is a diagnostic test for detecting autoimmune diabetes that comprises the product encoded by the DNA molecules of the invention.

DESCRIPTION OF THE FIGURES

[0017] FIG. 1 schematically depicts the complete sequence of IA-2. The N-terminal sequence signal (NS) and transmembrane domain (TM) are highlighted.

[0018] FIG. 2 is a schematic representation of IA-2c encoding sequences and their predicted protein products. The names refer to the corresponding plasmid and encoded peptide, and the numbers refer to the base pair position in the complete IA-2 gene or to the amino-acid residue in the complete sequence of the IA-2 peptide.

[0019] FIG. 3(A) shows a SDS-PAGE revealing the expression level of IA-2c encoded by pITACA (whole cell lysates before and after induction, lanes 1 and 2, respectively); IA-2c encoded by pITACAe (whole cell lysates before and after induction, lanes 3 and 4, respectively); IA-2cH5 encoded by pITCAH5 (whole lysate of induced bacteria and purified product, lanes 5 and 6, respectively); and IA-2c(AD/DH)H5 encoded by pITCAH5 (AD/DH)H5 (whole lysate of induced bacteria and purified product, lanes 7 and 8, respectively). FIG. 3(B) shows a Western blot of whole lysate of bacteria expressing IA-2c revealed with a pool of sera from IA-2A-positive patients. And FIG. 3(C) shows a Western blot of purified 1A-2cH5 and IA-2c(AD/DH)H5 (lanes 2 and 3, respectively). In all panels, arrows indicate the electrophoretic mobility of ovalbumin (43 kDa).

[0020] FIG. 4 shows the results of a radioimmunooassay (RIA) using different concentrations of IA-2c variants against a radioactive tracer [35S]IA-2c. The tracer was incubated with an IA-2c specific rabbit polyclonal serum and different concentrations of each of the IA-2c variants. The ratio of bound-to-free (b/f) [35S]IA-2c as a function of the logarithm of the molar concentration of free antigen (F) was plotted in a coordinates system. The solid square represents b/f in the absence of unlabeled antigen. IA-2cH5 and IA-2c(AD/DH)H5 concentrations in solution were calculated based on the absorbance at 280 nm. The concentration of IA-2c in crude extract was estimated by SDS-PAGE densitometry, using a calibration curve obtained with known concentrations of IA-2cH5. Kassp for each variant are summarized in TABLE 2.

[0021] FIG. 5 shows some representative spectra of circular dichroism in the far-UV (A) and near-UV (B) spectra of IA-2cH5 and IA-2c(AD/DH)H5. Protein concentration was 0.7 μM (far UV spectrum) or 10 μM (near-UV spectrum). In both cases, the proteins were dissolved in 50 mM sodium phosphate, 200 mM NaCl, pH 6.5.

[0022] FIG. 6 shows a scheme of a size-exclusion chromatography of IA-2cH5 (solid line) and reference proteins (dotted line). Peaks 1-5 correspond to thyroglobulin (670 kDa), bovine IgG (158 kDa), ovalbumin (44 kDa), and myoglobin (17 kDa), respectively.

[0023] FIG. 7 shows the correlation between bcELISA enzimoiimmunoassay and the radioligand binding assay (RBA). Signals are expressed as standard deviation units (SD units). Statistical parameters were r=0.53 and P<0.001. The linear regression line is shown. Control patients’ data were not included in the graph, and all points correspond to patients’ sera.

[0024] FIG. 8 is a listing of primers used during the cloning and mutagenesis processes carried out in the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0025] In the description of the present invention a number of terms current in the field of genetic engineering are used. To facilitate understanding the scope of the present invention, some of these terms are defined below.

[0026] Within the context of the present invention, “optimized nucleotides sequence” refers to a nucleotides sequence from a gene that has been modified with the purpose of increasing the expression levels in a specific biological system.

[0027] The term “expression” refers to the production of a protein as a result of transcription of the gene encoding it in a messenger RNA molecule (mRNA), and its subsequent translation into a polypeptide chain. As known by experts in the art, the expression levels of genes are affected by regulatory elements (promoters, polyadenylation sites, etc.) present in the gene, and factors related to the aforementioned transcription and translation.

[0028] “tARNAGG/AGG” refers to arginine transfer RNA that possesses a specific anticodon for the AGG codon.

[0029] “Regulatory sequences” refers to the sequences of nucleotides located before, after, or within the gene, which control the transcription and/or translation of the gene. Examples of regulatory sequences are promoters, transcription termination sequences, and polyadenylation sites.

[0030] The term “promoter” refers to the sequence of nucleotide (normally located before the S' end of the gene reading frame) that regulates the gene transcription. These sequences are necessary to direct expression. As known by experts in the art, the position of a promoter with respect to the translation initiation site depends on the type of promoter, both in the case of natural ones as well as in those artificially added, and its efficacy to promote transcription depends on the presence of the molecules with which they interact in a biological system.

[0031] “Chimeric DNA molecule” is a portion of DNA formed by the fusion of at least two DNA fragments, in an array previously nonexistent in nature. Whenever chimeric DNA molecules form a reading frame, proteins encoded by this type of molecule are called “fusion proteins” or “chimeric proteins”.

[0032] The expression “operatively bound” refers to nucleotide sequences situated on a same nucleic acid molecule, which are associated in a way that the function of one is affected by the other. For example, a promoter is operatively bound to a gene when it is capable of affecting the expression of that gene, and means that this gene is under the transcriptional control of the promoter.

[0033] “Transfection” refers to the stable introduction of a DNA fragment carrying a functional gene, within an organism that previously lacked that gene.
“PCR” refers to the polymerase chain reaction.

“APP” refers to a peptide or protein auxiliary for expression and folding; these APP are used to express and/or purify proteins as fusion products.

“Primers” refers to polynucleotides necessary to initiate PCR.

Compared to previous attempts, the present invention comprises DNA molecules that allow an important increase in the production yield of IA-2eC. Since IA-2eC is a major antigen in autoimmune diabetes and is necessary for detection of autoantibodies against IA-2 (IA-2A), this invention also constitutes a contribution to the development of IA-2A detection techniques. Further, this invention provides DNA molecules with optimized nucleotide sequences encoding IA-2eC.

During the development of the present invention, when cells Escherichia coli were transformed with pTICA—derivative of pET containing the natural encoding sequence for IA-2c inserted in sites Nde I and BamHI, as shown in FIG. 2—the induced cultures showed extremely low IA-2c expression (FIG. 3 A, lanes 1 and 2). These results confirmed prior reports on the expression of IA-2c [33]. When analyzing the nucleotide sequence of the IA-2c gene, two AGG codons were found starting, respectively, at positions 2747 and 2816, according to the numbering of the base pairs in the complete IA-2 gene reported in [9] (SEQ ID NO-4), which have low frequency of use in Escherichia coli, as defined in [34, 35]. Thus, translation of the natural IA-2c gene may have been hindered by insufficient quantities of tARNAG/AuG.

Degeneracy of the genetic code enables a single amino-acid to be encoded by more than one codon. For example, six different codons encode serine, and four different codons do so for alanine. However, the frequency of use of the codons for each amino-acid varies according to the group of organisms involved. When comparing the usage frequency distribution of codons characterizing different organisms, it is observed that as the organisms are less phylogenetically related, they will differ more in the usage frequency of codons for the different amino-acids.

Apparently, these differences in the usage frequency of codons affect the expression levels of genes by regulating the elongation rate of peptides. This argument is supported by two lines of experimental evidence. First, the rate of polypeptide synthesis depends on the nature of the codons being translated, as well as the initial kinetics for the formation of the ternary complex of transfer RNA (tRNA). Secondly, the frequency distribution of codons in the cell tRNA tends to follow the frequency distribution of codons in mRNA. In this way, when a heterologous gene from a very phylogenetically distant species is inserted into a cell, the differences in the usage frequency of alternative codons may lead to a significant waste of the cell’s expression capacity. It follows then that using optimized nucleotide sequences— in which the frequencies of codons reflect those frequencies found naturally in the host organism—is useful and advantageous for high-level expression of recombinant proteins.

Nevertheless, the genetic expression does not depend solely on the coding nucleotide sequences, but also on various types of interactions between defined portions of DNA and proteins (in turn encoded by other genes), and on various feedback mechanisms within and between cells. A typical eukaryotic gene consists of different, operationally distinguishable portions (exons and introns, promoters, enhancers, and other control elements), so constituting a complex system of hardly predictable behavior when attempting expression outside its natural context.

Degeneracy of the genetic code, as already mentioned, is a phenomenon whereby various codons (commonly known as alternative codons) encode to a same amino-acid. This allows introducing variations in the nucleotide sequence of a gene without modifying the amino-acid sequence of the expressed polypeptide. Said variations (called silent mutations) may affect the stability of the mRNA the tendency of the mRNA to form secondary structures in segments important for translation initiation, finally affecting the expression level of the polypeptide encoded by the gene. On the other hand, despite that alternative codons specify the same amino-acid, these are not equally used. As a result of this difference in use of the alternative codons, translation of the mRNA is slower in those mRNA zones where less-frequently-used codons are more abundant, whereas where those frequently-used are more abundant, translation is faster. This differential use varies for each species, with the level of use of each of the alternative codons being known for many species.

Despite knowing the level of use of alternative codons in many biological systems, replacing less-frequently-used ones with those frequently-used does not assure an increased expression of the encoded polypeptide. This is due in part to the contribution of other factors like the aforementioned mechanisms of interaction. It is also necessary to take into account that pause in the translation produced by several less-frequently-used codons is a synergistic phenomenon, which may mean that replacing just one of these codons may have no appreciable effect on the expression rate of the polypeptide. In fact, in the case of the present invention, replaced codons were not the only ones less-frequently-used, although they were the ones of lowest usage among those in this category. Moreover, it was observed that the presence of less-frequently-used codons in the mRNA sequence fulfills a physiologic function, facilitating folding of the growing polypeptide; so that the replacement of a less-frequently-used codon by a frequently-used may disturb the folding of the protein and, consequently, its expression in the soluble form.

On the other hand, it is important to mention that, in the case of the present invention, the encoded polypeptide is a fragment of a protein (spanning amino-acids 604 to 979). Because of the lack, to date, of structural studies on this polypeptide, the possibility that the selected portion of protein were unable to fold autonomously could not be discarded in advance. If this occurred, the appearance of the correctly-folded polypeptide in solution could be an energetically unfavorable event, which could be the cause of the low yield of previous attempts [33]. The true magnitude of this problem can be appreciated considering the numerous cases in which the absence or addition of a few amino-acids in the N— O C-terminal ends of a polypeptide are crucial for the overall folding of the protein and its solubility.

Finally, the expression level of a protein may be affected by the conditions under which the biological system is developed as, for example, those of the culture medium.
(temperature, induction time, presence of certain nutrients), that have little or nothing to do with the nucleotides sequence of the gene.

Therefore, despite that inspection of the sequence of nucleotides that encode IA-2ic revealed the presence of less-frequently-used codons, it was not possible to predict, without further research, that their replacement by frequently-used codons would improve expression.

This invention provides DNA molecules with nucleotide sequences that encode IA-2ic wherein the natural, less-frequently-used in bacteria AGG codons, were replaced by CGT codons which are frequently-used in Escherichia coli (Table 1). As a result, the present invention includes DNA molecules that comprise the following optimized sequences: SEQ ID NO:1, a nucleotide sequence that encodes IA-2ic, wherein both natural codons have been replaced; SEQ ID NO:2, which encodes IA-2ic and wherein the only one replaced is codon AGG starting at position 2747 (the numbers correspond to the base pair position in the complete IA-2 gene—SEQ ID NO:4—reported in [9]); and SEQ ID NO:3 in which only codon AGG starting at position 2816 was replaced; and nucleotide sequences with at least 50% of identity with those, in which at least one of the codons initiating at positions 2747 and 2816 is CGT.

<table>
<thead>
<tr>
<th>Triplet</th>
<th>CTT</th>
<th>CGC</th>
<th>OGA</th>
<th>CGG</th>
<th>AGA</th>
<th>AGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>20.7</td>
<td>21.1</td>
<td>3.7</td>
<td>5.7</td>
<td>2.7</td>
<td>1.6</td>
</tr>
</tbody>
</table>

*In number of occurrences per thousand.

These replacements introduce no changes in the amino-acid sequence of IA-2ic, but, as has been demonstrated, they lead to an important increment in the protein expression levels when those sequences are used to produce IA-2ic, for example, in Escherichia coli. For instance, experiments yielded 80 mg of protein per liter of culture, whereas with the native gene, the amount of protein was not quantifiable by SDS-PAGE; at the same time, prior attempts published by other research groups [36] reported yields of 1 mg of protein per liter of culture using the native gene sequence. In addition, in Western blot assays, RIA and ELISA, the antigen thus obtained was able to bind to anti-IA-2 specific antibodies, either from immunized animals or from patients sera samples, proving to be fully reactive from an immunological point of view.

To facilitate the purification procedures of IA-2ic peptides encoded by the optimized molecules of the present invention, chimeric DNA molecules were developed, which are also part of the present invention. In the latter, the 3'-end of the optimized DNA molecules encoding IA-2ic is linked with sequences encoding an APP. The addition of the APP does not affect the high expression yield achieved by optimizing the IA-2ic gene sequence. A preferred APP for use in these chimeric DNA molecules is a His-tag (frequently referred to as H). Once expressed in an adequate biological system, the chimeric genes of this invention allow obtaining fusion proteins formed by a IA-2ic antigen bound to an APP (for example, His-tag), which may be easily purified through chromatographic methods. The fusion proteins obtained in this way keep their immunological properties intact.

Sequences encoding APP may be linked to the 3'-end of optimized nucleotide sequences encoding to IA-2ic previously cloned in a plasmid. Therefore, another object of the present invention is an expression vector containing an optimized nucleotide sequence that encodes IA-2ic antigen. Said expression vectors may further comprise one or more operatively-bound regulatory sequences directing the expression of the encoded product in a bacterial cell. In addition, said vectors may contain a sequence that encodes an APP bound to one of the ends (in this case, 3') of the nucleotides sequence that encodes IA-2ic. These vectors are evidently useful for the transfection of cells, which, once modified, will produce, under appropriate suitable culture conditions, the IA-2ic peptides, alone or bound to APP. In one of its preferred embodiments, the vector is a plasmid derived from the pET plasmid, which is commercially available and possesses the promoter of bacteriophage T7 RNA polymerase. These plasmids are particularly suitable for the transformation of bacterial cells. They are particularly useful for transfecing Escherichia coli cells.

All the genetic constructions of the present invention may be prepared by techniques and procedures well known to those expert in the art. For example, both DNA and useful for the production of proteins, may be synthesized by methods involving biological systems. For DNA, cloning the gene in a plasmid capable of autonomous replication inside bacteria, is an example of a method for obtaining a good quantity of specific DNA. To obtain specific RNA in vitro, a DNA molecule—which must contain the sequence that it is desired to transcribe along with the necessary regulatory sequences—is incubated in a medium containing the biosynthetic apparatus necessary to synthesize RNA. This biosynthetic apparatus is generally found in commercial systems like rabbit reticulocyte lysates [53], or wheat germ extracts [54]. Subsequent to obtaining RNA and DNA molecules of native sequences, the optimized sequence may be obtained, for example, through directed mutagenesis. The DNA and RNA constructions may be optionally obtained by chemical synthesis, through in vitro assembly of the constituent nucleotides. This procedure may be carried out manually using well-established techniques [55], or through automated chemical synthesis using one of the several commercially available instruments.

A further aspect of this invention is a transformed host cell containing a DNA molecule that comprises an optimized nucleotide sequence encoding IA-2ic antigen. Since the DNA molecules of the present invention were optimized for expressing IA-2ic in prokaryotic systems, the host cells are preferably bacterial cells. Among the different bacterial cells, Escherichia coli cells are preferred.

The transformed cells of the present invention may be obtained by methods well-known to the expert in the art. For example, DNA sequences of the present invention may be introduced into host cells using vectors. Although the vectors are preferably plasmids, any other type of vector can be used as, for example, bacteriophages, cosmids, phagemids, YACs, etc. The use of vectors may be avoided resorting to transformation of cells by direct insertion of the IA-2ic gene of the present invention into the cell genome.
preferably a bacterial cell. Another transformation procedure that allows omitting the use of vectors consists of obtaining the mRNA (by isolation or chemical synthesis), after incubating it with an intracellular bacterial extract, thus expressing the protein in vitro. Therefore, mRNA molecules derived from transcription of the optimized DNA sequences is another aspect of the invention, as well as chimeric DNA molecules provided through this invention.

Another aspect of the present invention is a method of preparing IA-2c by culturing under appropriate conditions, cells transformed with the DNA molecules of the present invention. Once the peptides have been expressed, they can be purified (for example, by chelated-metal affinity chromatography [37]) for their subsequent use in the diagnosis of autoimmune diabetes. Consequently, both the polypeptides produced by preparing IA-2c of the present invention, and the in vitro diagnostic procedures employing them (in which a sample of the patient’s serum, plasma or blood is contacted with the polypeptide) are part of the present invention. Finally, another aspect of the present invention is a diagnostic kit for detecting autoimmune diabetes that comprises the polypeptide of the invention.

The following examples are presented to clarify the nature of the present invention. Their purpose is to illustrate the development and application of the invention, and should not be considered as limiting its scope.

EXAMPLE 1

Obtaining DNA Constructions

Total RNA was obtained by extraction with phenol-chloroform [38] from human pancreas samples obtained for diagnostic or surgical purposes. Single-strand DNA was prepared by reverse transcription using primer icaR (FIG. 8) and reverse transcriptase of the Moloney leukemia virus (Promega, Madison, USA). The previous reaction product was subjected to PCR [39] with primers icaR and icaF using Taq polymerase (Promega). As a result, a DNA molecule was obtained spanning position 1462 to position 3055 of the complete IA-2 gene [40]. This molecule was cloned in a PGEM-T Easy plasmid (Promega) generating pG-ICA (FIG. 2).

Using pG-ICA, a DNA fragment of 1627 base pairs (bp) was amplified by PCR with primers icNdC and 3095BamH, using Pfu DNA polymerase (Promega) in the PCR. The product was cloned in pGEM3zf (Promega) previously digested with Sma I. A fragment of 1215 bp was excised from the resulting plasmid with Nde I and BamHI I, which was cloned in a pET9b plasmid (Novagen), generating plasmid pTICA.

Subsequent derivatives of pTICA, described in FIG. 2, were generated by directed mutagenesis. To generate pTICAα (FIG. 2), a DNA fragment was amplified by PCR using pTICA and primers 2797c and 3095BamH. The product was digested with Pst I and Nco I, and cloned at the same sites as pTICA. To generate pTICAβ, two PCR were carried out using pTICAα as a template with pairs of primers 2741cF-3095BamH and 2741cR-icNdC. The products underwent another PCR with primers icNdC and 3095BamH, generating a fragment, which, once digested with Pst I and Nco I, was cloned at the same sites as pTICAα. To add the sequence encoding the His-tag, a PCR was carried out with primers icNdC and 3′ThsBam using pTICAα as the template. The product was digested with Pst I and BamHI I, and cloned at the same sites as pTICAα and pTICAβ, thus resulting in plasmids pTICAαH3 and pTICAβH3, respectively (for simplicity, FIG. 2 shows only pTICAαH3 and its derivative). To generate the amino-acid mutations, PCR was carried out on pTICAαH3 using primers 2797cDG and 3′ThsBam. The product was digested with Pst I and BamHI I, and cloned at the same sites as pTICAϕ, thus resulting in plasmid pTICAϕ(DG)H3 (not shown in FIG. 2). Using this plasmid as a template, a new PCR was carried out with primers A877D and 3′ThsBam, and the product was digested with Pvu II and BamHI I and cloned at the same sites as pTICAϕ, thus resulting in plasmid pTICAϕ(AD/DG)H3.

All the DNA sequences encoding IA-2 or its variants, were confirmed by sequencing (DNA Sequencing Facility of the Cancer Research Center, University of Chicago, USA).

EXAMPLE 2

Synthesis and Purification of IA-2c and its Variants

BL21 (DE3)pLysS Esherichia coli bacteria transformed with plasmids pTICA, pTICAα, pTICAβ, or pTICAϕ(DG)H3, were grown at 37° C. in Luria-Bertani (LB) nutrient medium (100-200 mL), supplemented with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol to A600 nm=1. Protein expression was induced for about 16 h at 20° C. with 10 mM IPTG (Fluka). After induction, the bacteria were collected by centrifugation and resuspended in 10 mL of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, pH 6.5), supplemented with 10 mM PMSF (Sigma, St Louis, Mo.), and lysed by compression-decompression with a french press (Spectronics Instruments, Inc., NY). The soluble fraction of the lysate was isolated by centrifugation (12000 g, 15 min, 4° C.). If dealing with a variant of IA-2c containing His-tag, the soluble fraction was supplemented with 10 mM imidazole, pH 6.5, and was mixed with 2 mL of agarose Ni-NTA (QIAGEN) preequilibrated with the same solution. After incubating for 1 h at 4° C., the suspension was poured into a 1.5 cm×5.0 cm column and washed five times with a lysis buffer containing 40 mM imidazole, pH 6.5. Then, matrix-bound protein was eluted with lysis buffer containing 200 mM imidazole, pH 6.5. Finally, the solution with which the protein was eluted was exchanged to 50 mM sodium phosphate, pH 6.5, or 50 mM MES, pH 6.5 by dialysis for 20 h at 4° C. When a greater purity was required, proteins were purified by ion-exchange chromatography in a FPLC system (Pharmacia-LKB Biotechnology). In this case the solution used was 30 mM HEPES, pH 6.5 and the samples were loaded into a MonoQ HR 5/5 column (Pharmacia-LKB Biotechnology), from which the protein was eluted with an increasing ionic-strength gradient to 30 mM HEPES, pH 6.5, 500 mM NaCl.

EXAMPLE 3

Assessment of the Expression Level by SDS-PAGE

Protein samples were first denatured at 100° C. for 5 min in a buffer solution containing 1% sodium dodecyl sulfate, 4% β-mercaptoethanol, 0.25% bromophenol blue,
0.3 M Tris Cl, pH 6.8. Then 10-20 μL were placed in each of the lanes located at the upper edge of a discontinuous gel of 10% polyacrylamide as described in [41]. These were then subjected to a 30 mA current for a time whereby the bromophenol reached the bottom edge of the gel (2-3 h). Finally, the proteins present in the gel were dyed with 0.5% Coomassie brilliant blue R250 in an aqueous solution of 50% methanol, 10% acetic acid. The results are shown in FIG. 3A.

**EXAMPLE 4**

Immunoreactivity Analysis by Western Blot

[0062] SDS-PAGE gels obtained as in the prior example were contacted with nitrocellulose membranes. The proteins contained in the gel were transferred to the membrane by a 100 mA cross current. Then, the membranes were separated and incubated with a blocking solution (2% deissicated low-fat milk in 50 mM sodium phosphate, 125 mM NaCl, pH 6.5 for 2 h). This membrane was then incubated for 2 h with a specific anti-IA-2c rabbit antibody in incubation solution (blocking solution, 0.5% Tween 20). The membrane was then incubated with rabbit immunoglobulins specific mouse antibodies conjugated to horseradish peroxidase, diluted 1/1000 in incubation solution. This was then incubated with a solution of 0.3% α-chloronaphthol in 10% methanol and in the presence of H₂O₂ until visualization of the bands. Between incubations, five washes with 50 mM sodium phosphate, 125 mM NaCl, 0.5% Tween 20, pH 6.5, were always carried out. The results are shown in FIG. 3B-C, and indicate the immunoreactivity of variants IA-2cH₆ and IA-2c(AD/DG)H₆.

**EXAMPLE 5**

Immunoreactivity Analysis by RIA and ELISA

[0063] Radioimmunoassays (RIA) were carried out in 60 μL of solution containing about 10000 dpm of [³⁵S]IA-2c and different concentrations (from 50 to 3000 nM) of IA-2c (in a soluble fraction of bacterial lysate), pure IA-2cH₆ or pure IA-2c(AD/DG)H₆, and aliquotes of an anti-IA-2c specific polyclonal rabbit serum in buffer solution for RIA (50 mM sodium phosphate, 150 mM NaCl, pH 6.5, 0.05% Tween 20, 1 mg/mL bovine serum albumin). Following one-week incubation at 4°C, 50 μL of a 50% suspension of protein A-Sepharose 4B FF (Sigma, St. Louis, Mo.) were added to a buffer solution for RIA, and this was incubated for 2 h at 4°C in an end-over-end shaker. After sedimentation, protein A-Sepharose was washed twice with buffer solution for RIA containing 0.35 M NaCl and resuspended in 100 μL of 1% sodium dodecyl sulfate. The supernatants of each reaction were transferred to appropriate vials to quantify the radioactivity in a liquid scintillation counter. The results were adjusted by least squares method to the following general equation:

\[ y = \frac{a}{1 + e^{-x}} \]

[0064] where b and f refer to the radioactive antigen (tracer bound and free, respectively). F refers to the total "free" antigen, and a and c are auxiliary parameters. From the adjusted curve an average affinity constant Kₐₙ₉ was derived, which may be calculated as 1/(concentration of non-radioactive antigen) at the value of b=f=0.5 [42]. The results, shown in TABLE 2 and in FIG. 4, demonstrate that the variants of IA-2c of the present invention bind to specific antibodies.

**TABLE 2**

<table>
<thead>
<tr>
<th>Variant</th>
<th>Kₐₙ₉ (M⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA-2c</td>
<td>2.2 (1.3-3.6)</td>
</tr>
<tr>
<td>IA-2cH₆</td>
<td>4.2 (1.9-8.2)</td>
</tr>
<tr>
<td>IA-2c(AD/DG)H₆</td>
<td>4.2 (1.7-10.0)</td>
</tr>
</tbody>
</table>

*Displacement data are shown in FIG. 4. Kₐₙ₉ values were calculated as described in [42] and 95% confidence intervals are shown in parenthesis.

[0065] In addition, by carrying out an ELISA, the recombinant antigens obtained through the present invention proved useful as diagnostic aid in autoimmune diabetes. Since the preliminary results with conventional single-well tests suffered from an unacceptably high unspecific signal, an assay format was developed, which we call bcELISA, where the nonspecific signal is subtracted from the final color development. The percentages of positivity for the bcELISA and the radioligand binding assay (RBA) were 54 and 62% respectively, while none of the techniques gave positive values among the control group (see TABLE 3).

[0066] It is worth mentioning that the results of the bcELISA were similar to those of the best ELISA published previously [13, 52]. Sera from 4 of 32 patients were positive by RBA and negative by bcELISA. Inversely, the bcELISA detected IA-2A in one negative RBA patient. The correlation between bcELISA and RBA was significant (r=0.53, P<0.001; TABLE 3 and FIG. 7). However, the slope of the regression line differed from the unit, reflecting differences between the principles of both techniques [42, 43]. Further, the conformation of the immobilized and chemically-modified antigen in ELISA may differ from the soluble radioactive antigen in RBA. Therefore, ELISA and RBA methods do not usually show comparable values for each patient, and the results may only be judged by the overall agreement in total positivity.

**TABLE 3**

<table>
<thead>
<tr>
<th>Analysis of IA-2A of sera from diabetic patients*</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBA</td>
</tr>
<tr>
<td>n</td>
</tr>
<tr>
<td>Patients</td>
</tr>
<tr>
<td>Controls</td>
</tr>
</tbody>
</table>

*The reference RBA was carried out as described earlier [44, 45], using [³⁵S]IA-2c synthesized in vitro.  The bcELISA was carried with biotinylated as-2cH₆. Percentages are in parentheses.  The two methods showed no significant differences in detecting positive samples (Z = 0.706, P = 0.24).

[0067] ELISA formats are relatively inexpensive and may be executed by technicians in simply equipped hospital laboratories. ELISA require no radioactive material. On the
contrary, RBA is the kind of assay that is carried out in well-equipped research or diagnostic centers, is expensive, and requires highly skilled personnel and radioactive materials disposal service. With these considerations in mind, and even if in a larger study it were slightly less sensitive than RBA, the performance of bcELISA in detecting IA-2A can be judged as satisfactory, thus the new method holds great promise for large scale screening of samples in preventive medicine.

**EXAMPLE 6**

Analysis of the Convormation by Circular Dichroism Spectroscopy

- The appropriate conformation of the His-tag containing variants of IA-2ic of this invention was confirmed by circular dichroism spectroscopy. The Circular Dichroism Spectroscopy measurements were carried out at 20°C on a Jasco 810 spectropolarimeter (Jasco Corporation, Japan) equipped with a pelletier-effect device for temperature control. The instrument was calibrated with (+) 10-camphorsulfonic acid following the manufacturer’s instructions. Scan speed was set to 20 and 50 nm/min (for near UV and far UV, respectively) with a 1-s response time and 1-nm bandwidth. Measurements in the near-UV were carried out in 1-cm cells containing protein in a concentration of 10-12 μM in 50 mM sodium phosphate, pH 6.5. On their part, 1-cm cells were used for the far-UV, containing protein concentration of 0.6-0.8 mM in 2.6 mM sodium phosphate, 200 mM NaF, pH 6.5. Six spectra were consecutively recorded, being their averages subjected to a smoothing process by the Savitzky-Golay method [46] with a 10-point-moving-window and then graphed. The figures show the trace of at least two measurements from different protein preparations. The far-UV results (shown in FIG. 5A) indicated that both variants of IA-2ic are similar in secondary structure, with a predomi-

**EXAMPLE 7**

Size-Exclusion Chromatography

- Hydrodynamic dimensions and protein aggregation state were determined by size-exclusion chromatography in an FPLC system equipped with a Superose 12 column (Pharmacia-LKB Biotechnology), using 100 mM sodium phosphate, pH 6.5, as elution solution. The hydrodynamic volume was calculated by comparison with proteins of a known size as indicated in FIG. 6. The results were consistent with an adequately folded, monomeric state.

**EXAMPLE 8**

Controlled Digestion with Trypsin

- IA-2icH$_{2}$ in a 30 mM concentration was incubated with trypsin in a 100:1 weight of protein-to-protease ratio in 50 mM sodium phosphate, pH 6.5 at 37°C: or 4°C. Samples were withdrawn at different times and analyzed by SDS-PAGE. At 4°C, the digestion produced two principal fragments of about 40 and 38 kDa. Treatment at 37°C resulted in complete digestion. There are reports of hot proteolytic points in this protein. IA-2 forms in vivo a product of 40 kDa by specific proteolysis at the N-terminus [24, 50] in cell-free extract, the principal proteolytic

**SEQUENCE LISTING**

<160> NUMBER OF SEQ ID NOS: 15

<210> SEQ ID NO: 1
<211> LENGTH: 1746
<220> TYPE: DNA
<230> ORGANISM: Artificial Sequence
<220> Feature:
<223> OTHER INFORMATION: Nucleotide sequence that encodes the intracellular domain IA-2ic. Base position 1=1868; Base position 2=1869; etc.

<400> SEQUENCE: 1

catattgagg aagattgtgg gccgcaagac aagggagccc tggcagctct gggtcctgag  60

gggccctag tgtcactaca ctttgaattg caggacgtgt gccgacgaga ctaggcccaq  120

eaagcttcttg tcaaccgggt caggggttca cggagcccct caggggttgg caggggttcc  180
tccagattca ggaagccgag ccaagcccag ccagcgcctc acagacagac ccagtctcgg  240
tgtgaggac cgcggcagcc acaattgacc atcctccacc gggccccagt tctgctattc  300
atgaggattc aagttgaggaa cggagcggcc tggcagcaag agtggaaggg cctctgctgc  360
taaacagcgcag cagaaacacg ctagcgccag gcgcagggag gggccacat cagaaagac  420
-continued

cggcatcttg acttcttgcc ctatgaccat ggcgcctas aaactgaagt ggcagcagc 490
cctttcggg aagggagcat gcaagcagct caaacattgc cagatcaagc agtaatgca 540
gcctaaat gcaagcagct caaacattgc cagatcaagc agtaatgca 560
tgggagacct ggcctcgagt cattcagctt cctacaagcct cgggtgaga ctatgcaagc 660
cgatgtgacc gctatgccc gagttgaggt gctttctctt accagctata tgggtgcaac 720
cctgtgttg aagccatact ggtgacagct tttctggtgc ggaagttctt cttatcagcc 780
gtcacagcct cgagcagcct cagacactct ggtgacagct ggcagcagc 840
gcgacagcct gtcacagcct cagacactct ggtgacagct ggcagcagc 900
cgggcctgct cttcagcatct cattcggtct ctggtgaga ctatgcaagc 960
ttcacatcgtc tcgacacgcct ccagacactct ggtgacagct ggcagcagc 1020
gtcgacacc ccagtctgtg gccaagcagcct gtcacagcct cagacactct 1080
ttcttggatt gctttcagct cttcggtgac ggtgacagct ggcagcagc 1140
cgtgacacc cggtggtcccg ctgggagctc ggtgacagct ggcagcagc 1200
agatctgtg gtcacagcct cagacactct ggtgacagct ggcagcagc 1260
cctttctggtg cagatcagcct cgggagcctc ggggagccac gggcctggtg 1320
cgggcctgct cttcagcatct cattcggtct ctggtgaga ctatgcaagc 1380
cgggacagcct cggcagcagcct cagacactct ggtgacagct ggcagcagc 1440
gggtttgttg cacatcggctc tgggtgaga ctatgcaagc 1500
ttcacatcgtc tcgacacgcct ccagacactct ggtgacagct ggcagcagc 1560
cgggcctgct cttcagcatct cattcggtct ctggtgaga ctatgcaagc 1620
ggtgtttgttg cacatcggctc tgggtgaga ctatgcaagc 1680
aaccattttaa aaaaaaactc attttggtg gttggtggttg 1740
cgcttg 1746

<210> SEQ ID NO 2
<211> LENGTH: 1746
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nucleotide sequence that encodes for intracellular domain of IA-2lc. Base position 1-1868; 2-1869; etc.

<400> SEQUENCE: 2

cctatgccc ggcagcagct cgggagcctc tgggtgaga ctatgcaagc 60
ggggagccagt gtcacacgcct cgggtgaga ctatgcaagc 120
agatcagcct cgggagcctc tgggtgaga ctatgcaagc 180
ttcacatcgtc tcgacacgcct cgggagcctc tgggtgaga ctatgcaagc 240
ttcacatcgtc tcgacacgcct cgggagcctc tgggtgaga ctatgcaagc 300
agatcagcct cgggagcctc tgggtgaga ctatgcaagc 360
ttcacatcgtc tcgacacgcct cgggagcctc tgggtgaga ctatgcaagc 420
agatcagcct cgggagcctc tgggtgaga ctatgcaagc 480
cctttcggg aagggagcat gcaagcagct caaacattgc cagatcaagc agtaatgca 540
-continued

gctactag ccaagcaggg ccagcgtgcc cataacctctg cagacttacct gcagatggtg 600
tggagacgc gtgcacacc ctaggtcatg ctggcaccgc tggcggagga tgtgtgcag 660
cagtgccac gctacaggg gcagtgaggt gcctgctctg accactgtaa tggagttgac 720
tctgctgtgc aagccatcgt cctgggagac acctgctgta gaggctctta cctgaagcac 780
gtgcagacac aggagacgag cagctcaccg catctgccct toctagctcg gcggagagag 840
ggcacaccc cctccacagcg gcgctgtgctg gatccocgcc ctaaggtgaa cagctgcctc 900
cgagagcgcct ccctgccccct cagctgctcg tgcagttggt gtcgagggag gcagggccac 960
tcctctcctc tagcattggt ccctgccccccc atggcgaaag gatgtgagga gatggagacc 1020
gtgcagacac cccgagcagc ctggcagctg cgggtgctgct gttctcctct tgggacgagc 1080
tttggcagttt ctggcagcgc cggcgagggag gatgtgagga gatgtgagga ccctcctcct 1140
cgagacccct ctgggagccc tggcgaggca gcgcagacgc tgccttctct tgcctgtgtg 1200
ggcctcacgcc ctcctgctcc cccagccgc cctcttgccgc atgctcagcgc 1260
cctctagaga gcgcgagagc ggagcagacc gcaggggagc gcggcagcgc cgctgagctg 1320
cgagacgcc ggcgtcagcc aggccagggc acgccagcgc gggtctgtcg gcggagagag 1380
agggagaagc atgggtgaga attccaccc cagcggctct ctcggcctct toctagctcg 1440
ggctgtccct tgcctagctg ggggtccttg gcgtctccag accatctcg ccggtgtctcg 1500
tgtcatcag ctagggaga tgggtggtgag ggtgctccga cccagtcccc cctgctcctc 1560
agccggccgc ctsgcctgta ctcgctagct ggcgcttgtc ctcctctccgc ctcctgggcc 1620
gcgtctgc accacccct gcacagcgcg tgtcaagctt ctogctctct gcggacagac 1680
acatctgta cacaaaacct cttttgctac aagttgagaa aagttgtggt gttgtctgtg 1740
cgctg 1746

<210> SEQ ID NO 3
<211> LENGTH: 1746
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nucleotide sequence that encodes for the intracellular domain of IA-2ic. Base position 1=1968; base position 2=1969; etc.

<400> SEQUENCE: 3

ctatgcccgg aagatgcccg gcacagacac aaggggcccc ctaggcgctc cgggggctgg 60

gggcagcct gtagcacttc tcttgaactc cggccgcggc cggcggagc cagctgctgc 120

cagtgccac gctacaggg gcagtgaggt gcctgctctg accactgtaa tggagttgac 180
tctgctgtgc aagccatcgt cctgggagac acctgctgta gaggctctta cctgaagcac 240
gtgcagacac aggagacgag cagctcaccg catctgccct toctagctcg gcggagagag 300
ggcacaccc cctccacagcg gcgctgtgctg gatccocgcc ctaaggtgaa cagctgcctc 360
cgagagcgcct ccctgccccct cagctgctcg tgcagttggt gtcgagggag gcagggccac 420
tcctctcctc tagcattggt ccctgccccccc atggcgaaag gatgtgagga gatggagacc 480
gtgcagacac cccgagcagc ctggcagctg cgggtgctgct gttctcctct tgggacgagc 540
ggcctcacgcc ctcctgctcc cccagccgc cctcttgccgc atgctcagcgc 600
cctctagaga gcgcgagagc ggagcagacc gcaggggagc gcggcagcgc cgctgagctg 660

tggagacgc gtgcacacc ctaggtcatg ctggcaccgc tggcggagga tgtgtgcag 720
cagcggccct ccgatctgcc acactgctcc tgcgtgcctc acctagttta tgcggctgac 720
catcggcgc aggactgcgc gacactgcag gcacgctctg tcccctgcgc gacgagcag 780
gtcggcggcc cccccggtc cgtcgccgct ctctcgctct ggcggcagc 840
gcgcggcgcc ctctcgccgc gcagccggct ggtggcctct gcgtggccg 900
cgacggcttg tccggcccaa ccctgtcctc tccggtgctc tccggtgctc 960
cgacggcttg tccggcccaa ccctgtcctc tccggtgctc tccggtgctc 1020
gtcggcggcc cccccggtc cgtcgccgct ctctcgctct ggcggcagc 1080
tttggttgg cccggccgac agctggctgc gcggcgtgac 1140
cagcggccct ccgatctgcc acactgctcc tgcgtgcctc acctagttta tgcggctgac 1200
agcactgcag tcctcgccgc gcagccggct ggtggcctct gcgtggccg 1260
cgtcgccgct ctctcgctct ggcggcagc 1320
cgccggcttg gcgccggcgc gcagccggct ggtggcctct gcgtggccg 1380
tttggtttgc gcgacggtgc gcgacggtgc gcgacggtgc gcgacggtgc 1440
ggcgcgctg tggcgcgctg tggcgcgctg tggcgcgctg tggcgcgctg 1500
agcactgcag tcctcgccgc gcagccggct ggtggcctct gcgtggccg 1560
gtcggcggcc cccccggtc cgtcgccgct ctctcgctct ggcggcagc 1620
agcactgcag tcctcgccgc gcagccggct ggtggcctct gcgtggccg 1680
agcactgcag tcctcgccgc gcagccggct ggtggcctct gcgtggccg 1740
agcactgcag tcctcgccgc gcagccggct ggtggcctct gcgtggccg 1746

<210> SEQ ID NO: 4
<211> LENGTH: 3613
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nucleotide sequence of the complete IA-2 gene.
<400> SEQUENCE: 4
    acggccccct ccggtggtgc gcggcgctgc ctctcgccct cccccggtcc 60
ggacggcag aagcactgcag gcacgctcag gcgtggcctc tggcgcgctg 120
    ccgacgccgc ccctctgctgc ccctctgctgc ccctctgctgc ccctctgctgc 180
tgcggcggcc cccccggtc cgtcgccgct ctctcgctct ggcggcagc 240
tttggttgg cccggccgac agctggctgc gcggcgtgac 300
    gcgcggcttg gcgccggcgc gcagccggct ggtggcctct gcgtggccg 360
tgcggccgct ctctcgccct tccggtgctc tccggtgctc tccggtgctc 420
tggcgcgctcg gcgaccgcgc gcagccggct ggtggcctct gcgtggccg 480
tgcggccgct ctctcgccct cccccggtc cgtcgccgct ctctcgccct 540
ggcgcgctcg gcgcggcgcc cccccggtc cgtcgccgct ctctcgccct 600
cgccggcgcc cccccggtc cgtcgccgct ctctcgccct cccccggtc 660
cgcgcgctcg gcgcggcgcc cccccggtc cgtcgccgct ctctcgccct 720
    gcgcggcgcc cccccggtc cgtcgccgct ctctcgccct cccccggtc 780
cgcgcgctcg gcgcggcgcc cccccggtc cgtcgccgct ctctcgccct 840
ccccctgcaca toctacgggg acotcacaag gctctcaacct gcccagcttt ttcagactc 900
tgggtcatct tacctgcacc aaggaggcc gcaacacacag gcggcaggg ggggggctg 960
gcagacgca gagccgcag gcggggcagaga cctcccagcag gggccccag ggggaggg 1020
aactcgagac agggaacagt ccgaagcgc tggcagcttt gcggcgagag ccggctgctgca 1080
gagggggcgc ctgtggtgctgg cggggctcagag tggcagcttg ccgtcaggtgca cccctcagca 1140
gctctccccat cctctgcaccctt ctgtcagctct actgccccag ggtgcaagag gaactcaggg 1200
agggggtttag aatgtgtgag ctgctatctg aaaaacatgt gcggggcagcc cggggcag 1260
agacacgca cagctcaggcg cccgcacatgc ccccgtctgt gcagccgacc ctggcagccc 1320
tacctccagt gagcagcag cagttggggga cctgctctcc cctgagcttc ccaagctgc 1380
cagcagctcc cggcagactct tctcttctgc gcagagaggc ccccgggacc agagccaggg 1440
caggggctga ccagcagcctcc cagggggcag agcagcagag gatattgctct acatctgcag 1500
ttagcagcag ccccagtgcc cgggctccag ggctgcacgcttc tggcagctgc tggcagctgc 1560
tggtcgcagct ctgcagggcca gttcatctcc acctcagttgg ccggccagag cggggcag 1620
cccgctggc gccacttctccgg aacccctctgc tttgggtctg cgtgaccacaag aacccctcag 1680
ggggaggttt gcagactcaggg gcagcagggc gctctataag ttcggcagag cggggcag 1740
agggggggag gcaagctgg cccctcctccc aacctggac gcaacactcc ccccggggctc 1800
aggtgttgctt ctttgaggttc cgggaggggg ggcggcgagg tgggctccct ctgggctgctg 1860
ggcgctgctgctt gcagccgagg gcggcaggcc cggcaggcag cccgcttgctc ccgggctgctg 1920
gccgggaggg gggcctgttat acaactatctg gcggtatctag gcggccagag gcgcgcagcc 1980
gggcagcagc cttcggcagcag cgggggcaag ggggccagcc cggggcagag cggggcag 2040
tgtgcctgg gcagcagcagcag cggcagccag ctggggccag ggggccctct gcggccagcc 2100
ggtctgggctt gggaggggct cccctcctccat g CCCCTCACAC GTTGCAGCT 2160
ncgataggg gcgacatccc tgggacgtcc ggcgggcttg gcggcaggtct ggagggctgc 2220
ctttcctgg ccagacgag ccgggcctgc tcggcggcg ccgggctgctc ccgggctgctc 2280
acccgccttg gggcctgtgt cccctccttc ctggcagctgc ctgggctgctg ctgggctgctg 2340
gggggctct gggggcccgtatttgccttt gcggggctttt gcggggctttt gcggggctttt 2400
gttggggtttt ggggggggcctt ggttcagctc cgggctgctc cgggctgctc cgggctgctc 2460
tgtggcgttt tgggctgctc gcggccagcc ctggggctgc cgggggctttt gcggggctttt 2520
tgcctgctgc cttgggggctt gcgctgctc cgggctgctc cgggctgctc cgggctgctc 2580
gggggggtttt gggggggggg gcggggctgctc ggcgggctgc cgggctgctc cgggctgctc 2640
ccgggggcgg gggcagccag gcggggctgc cgggctgctc cgggctgctc cgggctgctc 2700
ccggggtttt gggggggggg gcggggctgctc ggcgggctgc cgggctgctc cgggctgctc 2760
tgggctgctgc cttgggggctt gcgctgctc cgggctgctc cgggctgctc cgggctgctc 2820
cctgctgctgc cttgggggctt gcgctgctc cgggctgctc cgggctgctc cgggctgctc 2880
tgctgctgc cttgggggctt gcgctgctc cgggctgctc cgggctgctc cgggctgctc 2940
ggggggggg gcggggctgctc ggcgggctgc cgggctgctc cgggctgctc cgggctgctc 3000
cccctgctgc cttgggggctt gcgctgctc cgggctgctc cgggctgctc cgggctgctc 3060
ctgggggttt tgggctgctc gcggggctgctc ggcgggctgc cgggctgctc cgggctgctc 3120
ctcagccccct ccctgaagag tcaggaagg ggagcggaga gggccgagc tgcctcgccct 3180
cgcagcggag atctccgagc ccagggcact cctagggggt gctgcaagca 3240
ggagagacgg asagacatt gcgtacaggt ccctccgctg cctaccattc 3300
cgtgcgtggt ctcctcagct ctcctcgggg ggtcgaggg ttcgcctgac acctgtgtgt 3360
gttcccactg gctgcaaggt ggagagaggt ggtgaggggt cgcagcagc ccggtccctt 3420
gtttcctcg ccggccgctg cccctgagtc acctctgggc ggtgtccgctt cctgggtcct 3480
agcggcggcc tgtgcggccc acctctcccc cagctgcgtgc ctaacctcctc tctttctgcc 3540
gcagagagc atctggcaga aaacactact ttgtgccagt ggtggtgtgt ttcgtgtgtt 3600
gtctgctgctg ttcgtgtgtt 3613

<210> SEQ ID NO 5
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (2797c) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 5
ctgcagtgat ggtgctgggg gcgtacc 26

<210> SEQ ID NO 6
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (2797cDG) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 6
cctcagtggt ggtgctgggg gcgtacc 26

<210> SEQ ID NO 7
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (27410F) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 7
ttcggccgta agtgagacaa gttgtcag 27

<210> SEQ ID NO 8
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (27410R) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 8
tggtcaccctc aggcgggaag tcagcagcgtg 28
-continued

<223> OTHER INFORMATION: Initiator polynucleotide (A877D) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 9
cctagcagct tacacttcct cagctgccg gatgaggta caacg 45

<210> SEQ ID NO 10
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (IoNde) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 10
catatcggc acacgacaa ggaagc 26

<210> SEQ ID NO 11
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (IoC) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 11
cagtgagag taggtacaca gagatg 26

<210> SEQ ID NO 12
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (IoF) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 12
gcgcgacaa gacacgagc g 21

<210> SEQ ID NO 13
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (3'ThisBam) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 13
ggatcttcag tgtgaggat ggtgagggct gcgcgcgcc acacgcctgg gaggcccctt 60
gag 63

<210> SEQ ID NO 14
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Initiator polynucleotide (3995BamHI) used during the cloning and mutagenesis processes.

<400> SEQUENCE: 14
cagtgagag taggtccca gagatg 26

<210> SEQ ID NO 15
1. A DNA molecule comprising an optimized nucleotide sequence that encodes IA-2ic antigen, wherein said sequence is SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, or nucleotide sequences with at least a 50% identity with SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, wherein at least one of the codons starting at bases 2747 and 2816 is CGT.

2. The DNA molecule as in claim 1, wherein said sequence comprises SEQ ID NO:1.

3. The DNA molecule as in claim 1, wherein said sequence comprises SEQ ID NO:2.

4. The DNA molecule as in claim 1, wherein said sequence comprises SEQ ID NO:3.

5. A chimeric DNA molecule comprising a DNA molecule of claim 1, fused with a sequence that encodes a peptide or protein which is auxiliary of the expression and/or folding.

6. The chimeric DNA molecule as in claim 5, wherein the auxiliary peptide is His-tag.

7. An RNA molecule, encoded by a DNA molecule of claim 1.

8. An expression vector, comprising the DNA molecule of claim 1.

9. The expression vector of claim 8, wherein the DNA molecule, is operatively bound to one or more regulatory sequences that direct expression of the IA-2ic antigen in a bacterial cell.

10. The expression vector of claim 9, wherein said expression vector is a plasmid derived from the plasmid pET.

11. A transformed host cell, comprising a DNA molecule of claim 1.

12. The cell like that of claim 11, wherein said cell is a prokaryotic cell.

13. The cell like that of claim 12, wherein said cell is an Escherichia coli cell.

14. A method of producing a polypeptide of human IA-2ic antigen, which comprises growing cells transformed with a nucleic acid molecule of claim 1, under conditions suitable for the production of said polypeptide.

15. The method of claim 14, wherein the transformed cells are Escherichia coli cells.

16. A method of producing a polypeptide of human IA-2ic antigen, which comprises the stages of:

   a) Transforming the Escherichia coli cells with a plasmid of claim 10;

   b) Growing the transformed cells at 37° C.;

   c) Inducing protein expression with IPTG;

   d) Collecting the cells;

   e) Lysing the collected cells; and

   f) Isolating the polypeptide.

17. A polypeptide of human IA-2ic antigen, wherein said polypeptide is produced using the method of claim 14.

18. An in vitro diagnostic procedure for autoimmune diabetes which comprises contacting a sample of a patient's body fluids with the polypeptide of claim 17.

19. A diagnostic kit for detecting autoimmune diabetes comprising the polypeptide of claim 17.

20. The diagnostic procedure of claim 18, further comprising an ELISA assay wherein a final correct signal is determined by subtracting nonspecific signal from a final measured color.

* * * * *