Autoantibodies and anti-microbial antibodies: Homology of the protein sequences of human autoantigens and the microbes with implication of microbial etiology in autoimmune diseases

Peilin Zhang, MD., Ph.D.
PZM Diagnostics, LLC
Charleston, WV 25301

Correspondence:

Peilin Zhang, MD., Ph.D.
PZM Diagnostics, LLC.
500 Donnally St., Suite 303
Charleston, WV 25301
Email: pzmllc@gmail.com
Tel: 304 444 7505

Abstract

Autoimmune disease is a group of diverse clinical syndromes with defining autoantibodies within the circulation. The pathogenesis of autoantibodies in autoimmune disease is poorly understood. In this study, human autoantigens in all known autoimmune diseases were examined for the amino acid sequences in comparison to the microbial proteins including bacterial and fungal proteins by searching Genbank protein databases. Homologies between the human autoantigens and the microbial proteins were ranked high, medium, and low based on the default search parameters at the NCBI protein databases. Totally 64 human protein autoantigens important for a variety of autoimmune diseases were examined, and 26 autoantigens were ranked high, 19 ranked medium to bacterial proteins (69%) and 27 ranked high and 16 ranked medium to fungal proteins (66%) in their respective amino acid sequence homologies. There are specific autoantigens highly homologous to specific bacterial or fungal proteins, implying potential pathogenic roles of these microbes in specific autoimmune diseases. The computational examination of the primary amino acid sequences of human autoantigens in comparison to the microbial proteins suggests that the environmental exposure to the commensal or pathogenic microbes is potentially important in pathogenesis of a majority of autoimmune diseases, providing a new direction for further experimental investigation in searching for new diagnostic and therapeutic targets for autoimmune diseases.

Introduction

Autoimmune disease is characterized by the presence of circulating autoantibodies. Some autoantibodies, for example, anti-smith antibodies, are more specific and pathognomonic to a specific autoimmune disease such as systemic lupus erythematosus (SLE)(1, 2). Others such as anti-nuclear antibody, are non-specific and present in many clinical diseases or even normal healthy individuals (1, 2). The questions of how and why these autoantibodies are generated in patients remain largely unanswered. Microbiome study demonstrated the presence of trillions of various microbes within the body, and there is an intimate symbiotic relationship between these microbes and the human host in various aspects of human tissue and organ functions (3-5). In addition, there are numerous species of microbial DNA within the blood circulation without the culturable microbes (6). The presence of the microbial DNA in blood without culturable microbes raises two possibilities: the blood culture methods are insensitive to detect the microbes but the microbes are present within the circulation (7), or alternatively the intact microbes are destroyed by the host immune system but the microbial DNA and proteins are present in the circulation, eliciting the human immune response to these microbial DNA/proteins. The human immune responses to the microbial DNA and/or proteins may manifest as anti-microbial antibodies with potential cross-reactivity to human tissues through molecular mimicry (8, 9). In the process of identifying the infectious agents in Crohn's disease and Sjogren's syndrome, we have identified a group of microbial proteins from the commensal or pathogenic microbes reacting to patients' plasma, and we showed that there were elevated antibodies within the circulation of patients reacting to the microbial proteins (8). Furthermore, we showed that there is a cross reactivity of anti-microbial antibodies to the human tissues. It is reasonable to assume that the anti-microbial antibodies in response to the microbes, either pathogenic or commensal, can be detected as autoantibodies in autoimmune diseases, given the presence of cross-reactivity of anti-microbial antibodies against human tissues (8).

In current study, I took one step further to examine the primary amino acid sequences of all known human autoantigens in autoimmune diseases and compared these amino acid sequences to the microbial proteins including the bacterial and fungal proteins,

using information from the Genbank and BLAST search tools publically available at National Center for Biotechnology Information (NCBI) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Our results are surprising and more than two third of the human autoantigens important for a variety of autoimmune diseases are homologous to the microbial proteins including bacterial and fungal proteins. These microbial proteins may elicit the human antibody responses with potential cross-reactivity to the human tissues, leading to specific human tissue damage and autoimmune diseases. The presence of anti-microbial antibodies in circulation in autoimmune diseases suggests a potential new mechanism of autoantibody production and autoimmune diseases. The computational analysis of the primary protein sequences for homology represents the initial step toward understanding the production of autoantibodies in autoimmune diseases.

Methods:

All known human autoantigens associated with human autoimmune diseases are listed in Table 1. The protein (amino acid) sequences of these human autoantigens were searched from the Genbank (https://www.ncbi.nlm.nih.gov/protein/?term=), and the specific amino acid sequences and/or specific Genbank accession numbers were used for BLASTP search against the bacterial database (Bacteria (taxid:2)) and fungal database (Fungi (taxid:4751)) using the previously defined search parameters (default parameters). The homology was ranked as low, medium and high using previously defined (default) parameters, and denoted in color from NCBI (blue — low, medium— pink, red — high) (Figure 1). Screenshot photos were taken to demonstrate the homology between the human proteins with the various microbial proteins with colored graphical illustrations and the specific amino acid alignments between the human autoantigens and the perspective microbial proteins (bacterial and/or fungal). The specific homology scores from the default NCBI search algorism including compositional matrix adjustment, positive identity and gaps were not used because of the lack of expertise from the author's perspective. The low, medium and high homology definitions based on the colored coded graphics were used to estimate the similarity of

two protein sequences from two separate species. Unique specific microbes were also denoted in Table 1 for specific homologous human autoantigens. Each and all human autoantigen with their respective search data are illustrated in a book to be published later (in preparation). The current writing is the summary of all the search data with a goal to point to an entirely different direction for further experimental investigation for the mechanism of autoantibodies and autoimmune diseases.

Results:

1. **Genbank search, definition of low, medium and high homology between the human autoantigens and the microbial proteins**

There are three examples of BLASTP searches for definitions of low, medium and high homology between the amino acid sequences of human autoantigens and the microbial proteins (Figure 1). Using the Genbank accession number for each individual human autoantigen (SSA52/Ro52, Genbank accession number AAA36581, and Histone H2A, Genbank accession number P20671, anti-Jo1 autoantigen, Genbank accession number P07814) and BLASTP search, three matched results were shown in Figure 1. The low, medium, and high homologies between the SSA52/R052, Histone H2A and Jo1 and the microbial proteins were illustrated by blue, pink and red colored graphics followed by the amino acid sequence alignments with the respective microbial proteins from specific bacterium or fungi. It is worth noting that antibody production requires a small stretch of amino acid (epitope) within an appropriate antigenic structure. The examination of the amino acid sequences can only show the homology and alignment between the human proteins and the microbial proteins without predicting the three-dimensional structure. As a general rule, the experimental significance of this homology search is limited to the knowledge of computational prediction, and the predictive information requires vigorous experimental validation to be clinically relevant. The importance and clinical significance of these human autoantigens in the specific autoimmune diseases is beyond the scope of this study.

2. Microbial proteins reactive to human plasmas in Crohn's disease and Sjogren's syndrome are highly homologous to human protein homologues:

We have previously identified a panel of bacterial proteins reactive to human plasmas from patients with Crohn's disease and Sjogren's syndrome, and we demonstrated that there were elevated anti-microbial antibodies within the circulation of these patients (8). We also demonstrated that the specific antibodies against the microbial proteins produced in vitro can cross react to human tissues. Currently, the amino acid sequences of these microbial proteins were used to search Genbank human protein database to compare the microbial proteins with human proteins (Figure 2). This is a reverse exercise of the methods in Figure 1. The query sequences from the bacterial proteins were used to search the human protein database (Homo sapiens (taxid:9606)). The homologies between the microbial proteins EF-G, ATP5a from *Staphylococcus aureus /pseudintermedius (10)*, Hsp65 from *mycobacterium avium subspecies hominissuis/Mycobacterium tuberculosis*, and EF-Tu from *Escherichia coli* to the human proteins were significantly high, and the antibody against the microbes can cross react to the human tissues as shown previously (8). The clinical significance of these microbial proteins and their respective antibodies within the circulation in Crohn's disease and Sjogren's syndrome have been previously discussed (8, 9).

3. Homology of all human autoantigens to microbial proteins

Using the same principle and search methods with identical search parameters, the known human autoantigens important for human autoimmune diseases were examined, and the results were listed in Table 1. Totally 64 protein autoantigens in a variety of autoimmune diseases were examined against the microbial protein databases including bacterial database (Bacteria (taxid:2)) and fungal database (Fungi (taxid:4751)). There were 25 autoantigens highly homologous to the bacterial proteins, 29 highly homologous to the fungal proteins. 19 human autoantigens showed medium homology to the bacterial proteins and 16 to fungal proteins. Two examples of high protein sequence homology between the human autoantigens and the microbial proteins

(bacterial and fungal) are shown in Figure 3. Combining the high and medium homology groups, there are 68.8% of human autoantigens showing medium to high homology to the bacterial proteins and 69% to the fungal proteins. Phospholipids can also be antigenic in autoimmune diseases such as Guillain-Barre syndrome or anti-phospholipid antibody syndromes, but the phospholipids serve as haptens in antibody response, and the haptens usually need to combine with carrier proteins to be antigenic (11). The phospholipids can derive from plasma membranes of eukaryotic or prokaryotic cells, and it is difficult to determine the sources of these phospholipids in these patient populations. It is also unclear if the patients with anti-phospholipid antibodies carry other circulating anti-microbial antibodies against other microbial proteins.

There is other important information that is medically relevant to clinical management of a variety of patients with autoimmune diseases, and this information requires experimental validation. Human autoantigen U5 ribonuclear protein in SLE is highly homologous to that of the *Candida albicans*. Human proinsulin sequence is homologous to a specific hypothetical protein of *Enterococcus faecium (12)*. Histidine tRNA ligase of anti-Jo1 antibody important for inflammatory myopathy is highly specific to that of *Clostridium,* and the antigens from anti-smooth muscle antibodies in autoimmune hepatitis were found to be specific to those of *Escherichia coli*. Specific clinical management plan can be devised if these relationships between the specific autoantigens and specific microbial proteins are experimentally validated.

Among the autoantigens highly homologous to the microbial proteins (red colored), most commonly seen were enzymes with catalytic functions in both human host cells and the microbial cells. The structural proteins in cytoskeleton are also common, and these structural proteins are well conserved across the microbial or all species with important functions in cell division and cell mobility (Table 2). Others proteins with various functions such as nucleoproteins, regulatory proteins, immune related proteins and ion channels are also identified. It is conceivable that the human immune responses to these microbial proteins lead to antibody production and these anti-

microbial antibodies will cross-react to the human tissue through molecular mimicry, leading to collateral human tissue damage.

Discussion:

It is well known that the B-cells and plasma cells are critical for autoimmune diseases (13). Through examination of all human autoantigens against the bacterial and fungal protein databases, approximately two third of the human autoantigens are significantly homologous to the microbial proteins including bacterial and fungal proteins in their respective primary amino acid sequences. The highly homologous protein sequences between the human host and the microbes suggest a reasonable probability that the autoantibodies in autoimmune diseases are derived from the host immunity against the microbes present in the human body, commensal or pathogenic, bacterial or fungal in origins. It is now known that human host is colonized by trillions of commensal microbes and these commensal microbes are intimate components of human development after birth (3-5). Exposure to environmental microbes including bacteria and fungi help develop normal immunity to prevent pathogenic infections, regulate human metabolic activity and various tissue functions (3-5).

It is noteworthy that no viral proteins are examined for autoantigens as the viral immunity is vastly different from the bacterial or fungal immunity. Anti-viral antibodies are protective against the subsequent viral infection, forming the basis of modern medical vaccination. However, exceptions are well-known that the viruses can evade the human immune system by hiding in the intracellular compartments and reactivated in response to the body stress conditions or immune compromised conditions. The examples of reactivation of Epstein-Bar virus (EBV) and herpes simplex virus in immune compromised patients are well-documented, although the mechanism of reactivation is yet to be proven. Specific anti-viral antibodies with cross reactivity to human tissues remain to be a possible mechanism for autoimmune diseases.

Production of antibody in vivo in response to a specific antigen is well studied, and a large scale production of antibody for pharmaceutical industry is performed routinely. However, the mechanism of the production of specific autoantibody in vivo, such as anti-CCP antibody (anti-CCP, cyclic citrullinated peptide, ACPA) in rheumatoid arthritis (RA) is still intriguing. Citrullination or deimination of protein is to convert the aiming acid arginine to amino acids citrulline. Citrulline is an unusual amino acid and it is not one of the standard 20 amino acids encoded by DNA in the genetic code. Citrullination is the result of post-translational modification by a group of enzymes arginine deiminases or peptidylarginine deiminase (PADs). Deimination of proteins will change the hydrophobicity since arginine is positively charged and citrulline is not, and the protein Citrullination leads to abnormal folding of proteins and its functions. Citrullination of proteins will also induce abnormal immune response by generating anti-citrullinated protein antibodies, leading to autoimmune disease such as RA and multiple sclerosis (MS). There are multiple PADs in human with enzymatic deiminase activities. Human PADs distribute in a variety of tissues in a tissue specific fashion, and these enzymes likely play important roles in cellular functions and signal transduction. There are also many bacterial deiminases that share significant homologies with human proteins. Bacterial deiminases, from either commensal or pathogenic, can potentially citrullinate human proteins, leading to abnormal citrullinated proteins to induce human immune response and autoimmune diseases. Alternatively, bacterial deiminases can directly induce human immune response upon entering the human body, and anti-deiminase antibody generated against the bacterial deiminases can cross-react to the human enzymes (deiminases), leading to autoantibodies and autoimmune diseases, although neither the reports of anti-deiminase autoantibodies in human diseases are documented nor any effort is made to discover anti-deiminase antibody in autoimmune diseases. However, in mouse model, a recent study demonstrated that the immunized mice with the murine or the human PAD2 or PAD4 enzyme can generate anti-citrullinated peptide anybody, and these antibodies against citrullinated peptides only occur with the enzyme (PAD) bound to the targets (14). It appears that the immunogenic epitope is the peptide-enzyme complex. Furthermore, the bacterial PAD from porphyromonase gingivalis (periodontitis) is important in periodontal disease and it

is also important for RA (15). At this point, the data to support the presence of anti-bacterial PAD antibodies or anti- PAD autoantibodies in autoimmune diseases are lacking, although it is plausible with a new mechanism of disease process.

The present study of the autoantigens and the microbial proteins is a computational prediction and requires vigorous experimental validation. Specific epitopes eliciting antibody productions in vivo are usually small peptides, and BLASTP analysis could miss some small areas of amino acids that are antigenic. Antigenicity of any given protein is not random. Epitope mapping of human autoantigens may help answer these questions. Combination of direct comparison of the human autoantigens and the microbial proteins with specific antigenicity mapping can help anticipate the epitopes of any given proteins and facilitate the specific epitope mapping. These computational tools are widely available in public and make it possible to study the specific autoimmune diseases for better diagnostics and therapeutics. Many microbial proteins identified by the BLASTP with homology to the human autoantigens are hypothetical proteins from the known or unknown microbes, and significant effort is required to characterize these unknown microbes (bacteria or fungi) and the functions of their hypothetical proteins so that clinical significance of these microbial proteins can be determined. Some hypothetical proteins are derived from the well-known common microorganisms such as *E. coli*, *Enterobacters* or *Candida albicans*, and the functions of these microbes within the body and disease state is yet to be completely understood.

Conclusion:

More than two third of the autoantigens important for human autoimmune diseases are homologous to the microbial proteins, suggesting a majority of the autoantibody production are against the microbial protein origin. Further experimental validation is required for better understanding of the autoantibodies and the autoimmune diseases.

Figure legends:

Figure 1. Definition of protein sequence homology with colored graphic illustration and the sequence alignment between the human autoantigens and the microbial proteins. The BLASTP search parameters were of the default setting, and the colored graphical illustrations were from the NCBI.

Figure 2. Highly homologous bacterial proteins identified in Crohn's and Sjogren's patients to human proteins by the default BLASTP search with graphic illustrations and protein sequence alignments. Individual microbial proteins and their Genbank accession numbers were listed on top of the graphics.

Figure 3. Representative high homology alignments between the human autoantigens and the microbial proteins (SSA60 and mitochondrial protein M2 complex E2 component with their Genbank accession numbers).

Table 1. Human autoantigens in autoimmune diseases and homology to the microbial proteins by BLASTP analysis. All common human autoantigens were listed with their respective search data against the bacterial and fungal protein database. The unique microbes (bacterial or fungi) were also denoted.

Table 2. High homology between the human autoantigens and the microbial proteins in categories based on the protein structures and functions.

Conflict of interest

PZM Diagnostics, LLC is a private clinical laboratory registered in the State of West Virginia, USA. The author is the co-founder and the stake owner of the laboratory.

References

1. Kumar V, Abbas, AK., Aster, JC. Robbins and Cotran Pathologic Basis of Disease - 9th edition. 9th Ed. ed.: Elsevier; 2015.
2. Kasper D, Fauci, AS., Hauser, SL., Longo, DL., Jameson, JL., Loscalzo, J. Harrison's Principles of Internal Medicine. 19th Ed. ed.: McGraw Hill Education; 2015.
3. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nat Med. 2018 Apr 10;24(4):392-400.
4. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, et al. Microbiome-wide association studies link dynamic microbial consortia to disease. Nature. 2016 Jul 7;535(7610):94-103.
5. Young VB. The role of the microbiome in human health and disease: an introduction for clinicians. Bmj. 2017 Mar 15;356:j831.
6. Kowarsky M, Camunas-Soler J, Kertesz M, De Vlaminck I, Koh W, Pan W, et al. Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. Proc Natl Acad Sci U S A. 2017 Sep 05;114(36):9623-8.
7. Zhang P, Minardi LM, Kuenstner JT, Zekan S, Kruzelock R. Extracellular Components in Culture Media of Mycobacterium Avium Subspecies and Staphylococci with Implications for Clinical Microbiology and Blood Culture. American Journal of Infectious Diseases and Microbiology. 2016;4(6):112-7.
8. Zhang P, Minardi, LM., Kuenstner, JT., Zekan, SM., Zhu, F., Hu, YL., Kruzelock, R. Cross–reactivity of antibodies against microbial proteins to human tissues as basis of Crohn's disease and Sjogren's syndrome. http://biorxivorg/content/early/2017/03/13/116574. 2017.
9. Zhang P, Minardi LM, Kuenstner JT, Zekan SM, Kruzelock R. Anti-microbial Antibodies, Host Immunity, and Autoimmune Disease. Front Med (Lausanne). 2018;5:153.
10. Somayaji R, Priyantha MA, Rubin JE, Church D. Human infections due to Staphylococcus pseudintermedius, an emerging zoonosis of canine origin: report of 24 cases. Diagn Microbiol Infect Dis. 2016 Aug;85(4):471-6.
11. Kochibe N, Nicolotti RA, Davie JM, Kinsky SC. Stimulation and inhibition of anti-hapten responses in guinea pigs immunized with hybrid liposomes. Proc Natl Acad Sci U S A. 1975 Nov;72(11):4582-6.
12. Niegowska M, Paccagnini D, Burrai C, Palermo M, Sechi LA. Antibodies against Proinsulin and Homologous MAP Epitopes Are Detectable in Hashimoto's Thyroiditis Sardinian Patients, an Additional Link of Association. PLoS One. 2015;10(7):e0133497.
13. Aschermann S, Lehmann CH, Mihai S, Schett G, Dudziak D, Nimmerjahn F. B cells are critical for autoimmune pathology in Scurfy mice. Proc Natl Acad Sci U S A. 2013 Nov 19;110(47):19042-7.
14. Arnoux F, Mariot C, Peen E, Lambert NC, Balandraud N, Roudier J, et al. Peptidyl arginine deiminase immunization induces anticitrullinated protein antibodies in mice with particular MHC types. Proc Natl Acad Sci U S A. 2017 Nov 21;114(47):E10169-E77.
15. Gabarrini G, de Smit M, Westra J, Brouwer E, Vissink A, Zhou K, et al. The peptidylarginine deiminase gene is a conserved feature of Porphyromonas gingivalis. Sci Rep. 2015 Sep 25;5:13936.

Table 1: Human autoantigens and homologies to microbial proteins in autoimmune diseases

Autoimmune disease	Autoantibody	Autoantigen	Bacterial homology (high, medium, low)	Fungal homology (high, medium, low)	Unique to bacteria	Unique to fungi	Unique microbes
Systemic lupus erythematosus	Anti-SSA/Ro	Ribonucleoproteins	SSA60 High SSA52 Low	SSA60 High SSA52 Low			
	Anti-ds-DNA	Double-stranded DNA					
	Anti-Smith	snRNP core proteins	U5 snRNP 100K High	U4/U6-U5 complex High		U5	Candida
	Anti-histone	Histone	Medium	Medium			
	Anti-thrombin/lupus anti-coagulant	Thrombin	Medium	Medium			
	Anti-ubiquitin	Ubiquitin	High	High			Acinetorbacter, streptomyces
Sjogren's syndrome	Anti-SSA/Ro	Ribonucleoproteins					
	Anti-SSB/La		Medium	Medium			
CREST syndrome	Anti-Centromere	Centromere	CENP-A, B, C Medium	Medium			
Inflammatory myopathy	Anti-Jo1	Histidine-tRNA ligase	High	High			Clostridium
Primary biliary cirrhosis	Anti-p62	Nucleoporin 62	No bacterial homology	Medium		p62	
	Anti-Sp100	Sp100 nuclear antigen	No bacterial homology	Low			
	Anti-gp210	Nucleoporin 210	No bacterial homology	No fungal homology			
	Anti-mitochondria	Mitochondria	M2, High, p52 High	High, High			
	ATP-dependent DNA helicase 2 subunit 2 (Lupus Ku autoantigen p86, p70)		No bacterial homology	High		Ku	
Systemic sclerosis	Anti-topoisomerase	Type I topoisomerase (Scl-70)	High	High			
Celiac disease	Anti-tTG	Transglutaminase	Medium	Low	tTG		
	Anti-actin	Actin	High	High			
Dermatitis herpetiformis	Anti-eTG	Transglutaminase	Medium	Low	eTG		
Rheumatoid arthritis	RF	IgG	High	High			
	Anti-CCP	Cyclic citrulinated peptide	PAD1, High	Medium	CCP		
Autoimmune hepatitis	Liver kidney microsomal type 1	P450 2D6	High	Medium	Liver kidney microsome 1 (P450 2D6)		
	Ant-soluble liver/liver-pancreas antibody	UGA suppressor tRNA protein	High	High			Chlamydia
	Anti-smooth muscle	Smooth muscle	SMA, myosin, tropomyosin, F-actin High	High			Escherichia coli
Granulomatosis with polyangiitis	C-ANCA	Leukocyte Protease 3	Medium	High			
		Autoantigen 56kD (CAP-50, P50559)	High	High			

Table 1: Human autoantigens and homologies to microbial proteins in autoimmune diseases

Microscopic polyangiitis, eosinophilic granulomatosis, systemic vasculitides	p-ANCA	Myeloperoxidase	High	High			
Hashimoto's thyroiditis	Anti-TPO	Thyroid peroxidase (microsomal)	High	High			
	Anti-TG	Thyroglobulin	Medium	Medium			
		Graves's soluble carrier protein, P16260	Medium	High		Graves' P16260	
Graves' disease	Anti-TSH	TSH receptor	Low	Low			
Myasthenia gravis	Anti-AchR	Nicotine acetylcholine receptor	Low	No fungal homology			
	Anti-MUSK	Muscle specific kinase	Medium	Medium			Salmonella
Lambert-Eaton myasthenic syndrome	Anti-VGCC	Voltage-gated calcium channel (P/Q type)	Medium	High		Calcium channel	
Limbic encephalitis, Isaac's syndrome (autoimmune neuromyotonia)	Anti-VGKC	Voltage-gated potassium channel	Medium	Medium			
Polymyositis	Anti-SRP	Signal recognition particle	High	High			
Scleromyositis		Exosome complex					
Limbic encephalitis, encephalomyelitis, subacute sensory neuronopathy	Anti-Hu (ANNA-1)	Neuronal nuclear proteins	Hu-D, Medium	Medium			
	Anti-Ma		No bacterial homology	No fungal homology			
Paraneoplastic cerebellar degeneration	Anti-Yo	Cerebellar Purkinje cells	No bacterial homology	Low			
	Anti-Tr	Glutamate receptor	No bacterial homology	No fungal homology			
Opsocionus myoclonus syndrome	Anti-Ri (ANNA-2)	Neuronal nuclear proteins					
Stiff person syndrome, Diabetes type I	Anti-GAD	Glutamate decarboxylase	High	High			
	Anti-amphiphysin	amphiphysin	No bacterial homology	Medium		amphiphysin	
Optic neuropathy, chorea	Anti-CRMP-5	Collapsin response mediator protein 5	High	High			
Sydenham's chorea, PANDAS		Basal ganglia neuron	Dopamine D2 receptor, Medium	Low			
Anti-NMDA receptor encephalitis	Anti-NMDA receptor	N-methyl-D-aspartate receptor (NMDA)	Medium	No fungal homology			
Neuromyelitis optica (Devic's syndrome)	Anti-NMO	Aquaporin-4	Medium	Medium			
		Human interferon gamma	No bacterial homology	Medium		IFN-G	
		BP230 (BPAG1)	No bacterial homology	High (first 400 aa)		BP230	
		BP180 (BPAG2)	Low (60 to 138 aa identical to E. coli	Low	BP180		E. coli

Table 1: Human autoantigens and homologies to microbial proteins in autoimmune diseases

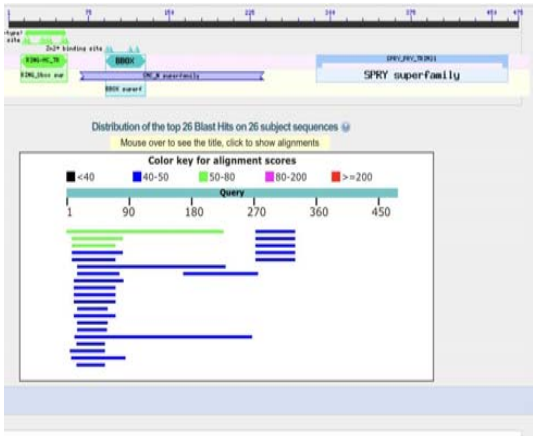
			hypothetical protein)				
Nephrotic and nephritis		Factor H	Low (40-50 aa identical to E. coli hypothelical protein	No fungal homology			E. coli
		Complement C3	High	No fungal homology	C3		Klebsiella
		C1q receptor/Calreticulin	High	High			Kangiella spongicola
		ACE (angiotensin converting enzyme)	High	High			
Neurodegenerative diseases		Presenillin 1	Low (50 aa identical to E. coli hypothetical protein	High		Presenilin	
		Nacastrin	Low	Medium			
		MMP1	Low	Medium			
Type 1 diabetes		Insulin	Medium (44 aa identical to Enterobacter faecium)	Low	Insulin		Enterococcus faecium
		Prohormone convertase	Low	High		Prohornmone C	
Mixed connective tissue	Anti-RNP	Ribonucleoprotein					
Antiphospholipid syndrome	Anti-phsopholipid	Phospholipid					
Miller-Fisher syndrome	Anti-ganglioside	Gangliosides GQ1B					
Acute motor axonal neuropathy		Ganglioside GD3					
Multifocal motor neuropathy with conduction block (MMN)		Ganglioside GM1					
			Totally 64 protein antigens				
			25 high	29 high			
			19 medium	16 medium			
			68.8%	69.20%			

Table 2: Protein classification of the autoantigens and microbial proteins

	Autoantibody	Autoantigen (Genbank accession #)	Bacterial homology (high)	Fungal homology (high)
Enzymes	Anti-Jo1	Histidine-tRNA ligase (P07814)	High	High
	Anti-mitochondria	Mitochondria (M2 and p52) (P10515, P11182)	High	High
	Anti-topoisomerase	Type I topoisomerase (Scl-70) (AAL05624)	High	High
	Liver kidney microsomal type 1	Liver kidney microsome 1 (P450 2D6) (AIA09571)	High	Medium
	C-ANCA	Leukocyte Protease 3 (P24158)	Medium	High
		Autoantigen 56kD (CAP-50, P50559)	High	High
	p-ANCA	Myeloperoxidase (AAA59863)	High	High
	Anti-TPO	Thyroid peroxidase (microsomal) (AAA61217)	High	High
		ACE (angiotensin converting enzyme) (AAR03504)	High	High
		Presenillin 1 (AAB46371)	Low (50 aa identical to E. coli hypothetical protein	High
		Prohormone convertase (insulin) (AAQ89322)	Low	High
	Anti-GAD	Glutamate decarboxylase (CAA49554)	High	High
Structural proteins	Anti-smooth muscle	Smooth muscle (SMA, myosin, tropomyosin, F-actin)(P62736, EAW66154, AAB59509, 5TBY-F)	High	High
	Anti-ubiquitin	Ubiquitin (CAA28495)	High	High
	Anti-CRMP-5	Collapsing response mediator protein 5 (Q9BPU6)	High	High
		BP230 (BPAG1) (Q03001)	No bacterial homology	High (first 400 aa)
Immune protein		Complement C3b (NP_000055)	High	No fungal homology
		C1q receptor/Calreticulin (P19474)	High	High
Nucleoproteins	Anti-SSA/Ro (SSA60)	Ribonucleoproteins (NP_001035828)	High	High
	Anti-Smith (U5 snRNP 100K, U4/U6-U5 complex)	snRNP core proteins (AAB87902, SC5314)	High	High
		Graves's soluble carrier protein, P16260	Medium	High
Ion channel	Anti-VGCC	Voltage-gated calcium channel (P/Q type) (O00555)	Medium	High
	Anti-SRP	Signal recognition particle (NP_003127)	High	High

Figure 1

SSA52 --- Low



hypothetical protein A6F71_09265 [Cycloclasticus sp. symbiont of Poecilosclerida sp. M]

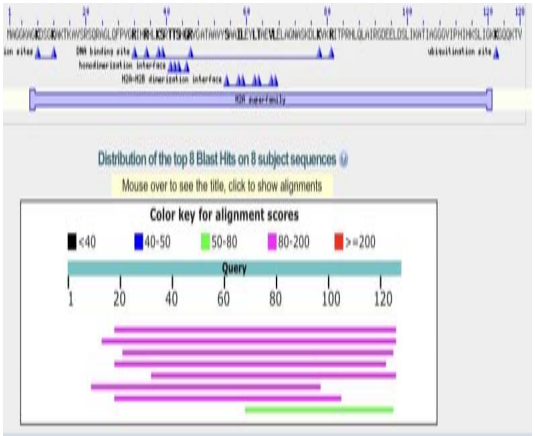
Sequence ID: ORU94389.1 Length: 689 Number of Matches: 1

Range 1: 1 to 243 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
72.4 bits(176)	6e-10	Compositional matrix adjust.	63/247(26%)	114/247(46%)	26/247(10%)

```
Query  1    MASAARLTMMWEEVTCPICLDPFVEPVSIECGHSFCQECISQVG----KGGGSV-CPVCR  55
            MA  A L ++ E++ C +CLD + +P  ++C H FC++C+  +G    +G  S+ CP CR
Sbjct  1    MAEEA-LKIIEEQLKCSVCLDIYTDPKLLQCFHVFCRKCLVPLGVRDQQGQLSLTCPTCR  59

Query  56   Q----------RFLLKN-LRPNRQLANMVNNLKEISQEAREGTQGERCAVH-GERLHLF  102
            +            FL+ N L  ++L N +          +    Q  RC+ H GE L+ +
Sbjct  60   KVTPRTVAGLQSAFLINNLLEAHKKLLNPLGKGNTTPPTSASAVQ--RCSEHAGEDLNFY  117

Query  103  CEKDGKALCWVCAQSRKHRDHAMVPLEEAAQEYQEKLQVALGELRRKQELAEKLEVEIAI  162
            C+   K +C  CA S    +H     ++A ++Y E++ ++L  +  ++++  K   E+
Sbjct  118  CQTCQKLICMECAHSMCG-NHEFGFFQDAFEKYTEEMML SLTPMENQEDVVTKALRELKT  176

Query  163  KRADWKKTVETQKSRIHAEFVQQKNFLVEEEQRQLQELEKDEREQLRILGEK----EAKL  218
               +     ET +  IH  F   +  L   E   + +L     + +L+ L  +    E  L
Sbjct  177  HHEEISDQRETIEKNIHIAFRSVREALTVRETELIGQLNHMTQIKLKGLAGQIDRIETTL  236

Query  219  AQQSQAL  225
             AQ   L
Sbjct  237  AQLKSCL  243
```

Histone --- Medium



Download GenPept Graphics

hypothetical protein A7M48_18550 [Acinetobacter baumannii]

Sequence ID: OIC85867.1 Length: 124 Number of Matches: 1

Range 1: 16 to 123 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
130 bits(328)	7e-37	Compositional matrix adjust.	69/109(63%)	82/109(75%)	2/109(1%)

```
Query  19   SRSQRAGLQFPVGRIHRHLKSRTTSHGRVGATAAVYSAAILEYLTAEVLELAGNASKDLK  78
            SRS RAGLQFPVGRIHR L+     +  RVGA A VY AA++EYL AEVLELAGNA++D K
Sbjct  16   SRSNRAGLQFPVGRIHRLLRKGNYAE-RVGAGAPVYLAAVMEYLAAEVLELAGNAARDNK  74

Query  79   VKRITPRHLQLAIRGDEELDSLIK-ATIAGGGVIPHIHKSLIGKKGQQK  126
              RI PRHLQLAIR DEEL+ L+   TIA GGV+P+I   L+ KK ++K
Sbjct  75   KTRIIPRHLQLAIRNDEELNKLLSGVTIAQGGVLPNIQAVLLPKKTEKK  123
```

Download GenPept Graphics

hypothetical protein [Pseudooceanicola sp. 157]

Sequence ID: WP_100165103.1 Length: 159 Number of Matches: 1

See 1 more title(s)

Range 1: 25 to 137 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
120 bits(301)	2e-32	Compositional matrix adjust.	67/114(59%)	83/114(72%)	2/114(1%)

Jo1 (aminoacyl-tRNA synthetase) -- High

proline--tRNA ligase [Clostridium pasteurianum]

Sequence ID: WP_003447938.1 Length: 480 Number of Matches: 1

See 7 more title(s)

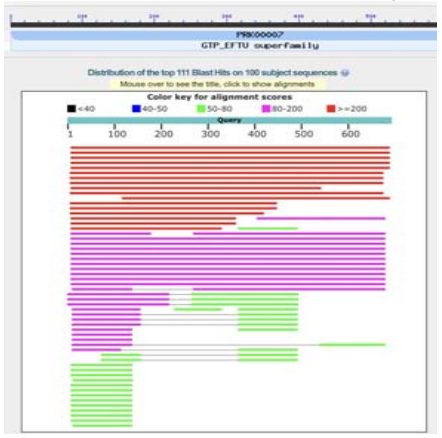
Range 1: 17 to 480 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
461 bits(1185)	9e-145	Compositional matrix adjust.	238/494(48%)	324/494(65%)	32/494(6%)

```
Query  1021  EENLADWYSQVITKSEMIEYHDISGCYILRPWAYAIWEAIKDFFQAEIKKLGVENCYFPM  1080
             +E+ A WY+ ++ K+E+++Y  + GC ILRP+AYAIWE I+ + D + K+ G EN Y P+
Sbjct  17    DEDFAQWYTDIVKKAELVDYSSVKGCVILRPYAYAIWENIQKYLDGKFKEKGHENVYMPL  76

Query  1081  FVSQSALEKEKTHVADFAPEVAWVTRSGKTELAEPIAIRPTSETVMYPAYAKWVQSHRDL  1140
             + +S L+KEK HV  FAPEVAWVT+ G  EL E + +RPTSET+      YAK VQS++DL
Sbjct  77    LIPESLLQKEKDHVEGFAPEVAWVTRGGNDELTEKLCVRPTSETLFCEHYAKIVQSYKDL  136

Query  1141  PIKLSQWCNVVRWEFKHPQPFLRTREFLWQEGHSAFATMEEAAEEVLQILDLYAQVYEEL  1200
             P  SQWC+VVRWE K  +PFLRT EFLWQEGH+  AT EEA  E +Q+L++YA  E +
Sbjct  137   PKLYSQWCSVVRWE-KTTRPFLRTTEFLWQEGHTIHATKEEAAEATIQMLNIYADFCENM  195

Query  1201  LAIPVVKGRKTEKEKFAGGDYTTTIEAFISASGRAIQGGTSHHLGQNFSKMFEIVFEDPK  1260
             LAIPVVKGRKTEKEKFAG + T TIE+ +   G+A+Q GTSH  GQNF+K F I F D
Sbjct  196   LAIPVVKGRKTEKEKFAGAEATTTIESLMH-DGKALQSGTSHFFGQNFAKAFGIQFAD--  252

Query  1261  IPGEK-QFAYQNSWGLTTRTIGVMTMVHGDNMGLVLPPRVACVQVVIIPCGITNALSEED  1319
               EK ++  Q SWG+TTR IG + MVHGD+ GL +PPR+A  QVVI+P          +
Sbjct  253   -KNEKLEYVNQTSWGMTTRIIGAIIMVHGDDEGLKMPPRIAPTQVVIVPIA--------QH  304

Query  1320  KEALIAKCNDYRRRLLSVNIRVRADLRDNYSPGWKFNHWELKGVPIRLEVGPRDMKSCQF  1379
             KE +I K  + + R+  V  RV+ D+ D  +PGWKFP +E+KG+P+RLEVGP+DM+  Q
Sbjct  305   KEGVIEKATELKDRIAKV-ARVEWDISDK-TPGWKFSEYEMKGIPVRLEVGPKDMEKNQV  362

Query  1380  VAVRRDTGEKLTVAENEAETKLQAILEDIQVTLFTRASEDLKTHHVVAHTMEDFQKILDS  1439
             V VRRDT EK+ V  +E ETK+  +LEDI  ++   A++  +A TME+F  I+++
Sbjct  363   VLVRRDTREKIFVPMDELETKIPELLEDIHKSMLEAATKAREEKTNIATTMEEFNDIVEN  422

Query  1440  GK-IVQIPFCGEIDCEDWIKKTTARDQDLEPGAPSMGAKSLCIPFKPLCELQPGAKCVCG  1498
              V+  +CG+ DCED IK+ T              GA S CIPF+   + +    CVC
Sbjct  423   KTGFVKAMWCGDKDCEDKIKEDT-------------GATSRCIPFE---QEELSDTCVCC  466

Query  1499  ENPAKYYTLFGRSY  1512
              AE+   +GR+Y
Sbjct  467   GKKAKHNVYWGRAY  480
```

Figure 2

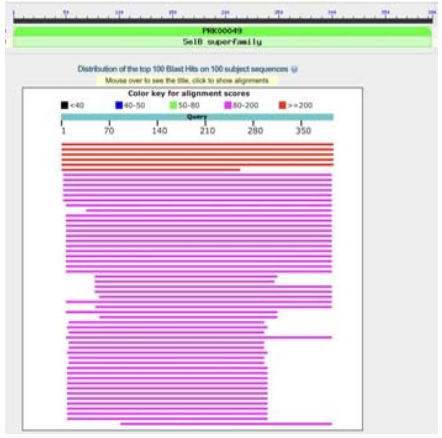
EF-G (WP_014614745)



Hsp65 (P9WPE6)

ATP5a (KZK18041)

EF-Tu (AAA50993)



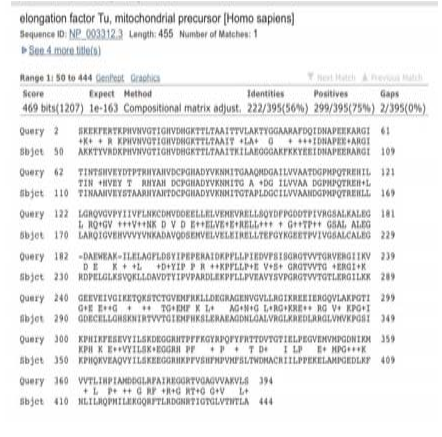
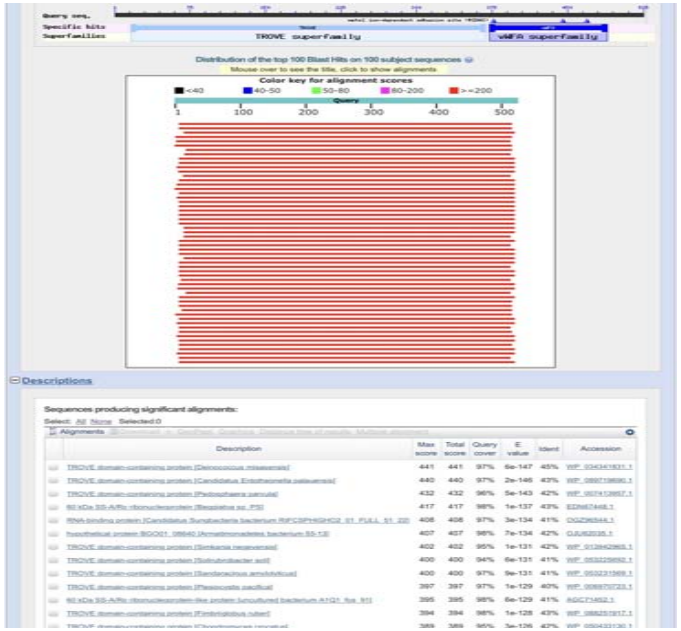


Figure 3

SSA60/Ro60 (NP_001035828)



TROVE domain-containing protein [Deinococcus misasensis]

Sequence ID: WP_034341831.1 Length: 525 Number of Matches: 1

Range 1: 14 to 505 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
441 bits(1133)	6e-147	Compositional matrix adjust.	231/513(45%)	311/513(60%)	24/513(4%)

```
Query  7    QMQPL-NEKQIANSQDGYVWQVTDMNRLHRFLCFGSEGGTYYIKEQKLGLENAEALIRLI  65
            Q QP+    Q AN   GY W+V+D  +L RFL  GSEGGT+Y+ EQKL ++NA+A+ R I
Sbjct  14   QNQPIPGSNQKANHAGGYAWEVSDWTQLDRFLILGSEGGTFYVGEQKLTVQNAQAVQRCI  73

Query  66   EDGRGCEVIQEIKSFSQEGRTTKQEPMLFALAICSQCSDISTKQAAFKAVSEVCRIPTHL  125
            E   G   ++ I   S+ GR  K +P LFALA+C+    D  T++AA +A+  V RI THL
Sbjct  74   EQD-GIRTVERIAEISESGRAPKNDPALFALAMCASFGDEKTRKAALEALPRVARIGTHL  132

Query  126  FTFIQFKKDLKESMKCGMWGRALRKAIADWYNEKGGMALALAVTKYKQRNGWSHKDLLRL  185
            F F++F +  +        WGR LR+AIADWYN+K   ALAL   KY+QR+GW+H+D LRL
Sbjct  133  FHFMEFIQQFRG------WGRGLRRAIADWYNQKDLKALALQAVKYRQRDGWTHRDALRL  186

Query  186  SHLKPSSEGLAIVTKYITKGWKEVHELYKEKALSVETEKLLKYLEAVEKVKRTRDELEVI  245
            +H   SEG   + ++IT+          E     L+ +EA ++V++      E +
Sbjct  187  AHPVAPSEGHQKLYRWITRD-------------QFEPATGLELIEAFKQVQQASVP-EAV  232

Query  246  HLIEEHRLVREHLLTNHLKSKEVWKALLQEMPLTALLRNLGKMTANSVLEPGNSEVSLVC  305
             LI+ HRL RE L T  L   EVW+ALL  MPL A++RNL  MT   +L P +    LV
Sbjct  233  KLIKRHRLPREALPTELLVRPEVWEALLPHMPLEAMVRNLANMTRVGLLTPMSDASRLVQ  292

Query  306  EKLCNEKLLKKARIHPFHILIALETYKTGHGLRGKLKWRPDEEILKALDAAFYKTFKTVE  365
             KL N + + KARIHP  IL AL TY  G G+RG  +W P ++++ ALD AFY  F  V
Sbjct  293  RKLGNSEAILKARIHPIKILAALRTYAGGQGMRGHGQWTPVQQVVDALDGAFYSAFGNVT  352

Query  366  PTGKRFLLAVDVSASMNQRVLGSI--LNASTVAAAMCMVVTRTEKDSYVVAFSDEMVPCP  423
            PTGKR +LA+DVS SM    +  I  L     +AA+ +V   TE    V+ FS ++VP
Sbjct  353  PTGKRIMLALDVSGSMEWGHIAGIPGLTPRVASAALALVTASTEAQHMVMGFSHQLVPIG  412

Query  424  VTTDMTLQQVLMAMSQIPAGGTDCSLPMIWAQKTNTPADVFIVFTDNETFAGGVHPAIAL  483
            ++    L  VL   + ++P GGTDC+LPM+ A K N P D FIV+TD+ET+ G VHPA AL
Sbjct  413  ISPRQRLDDVLNTVGRVPMGGTDCALPMLHALKENIPVDAFIVYTDSETWYGNVHPAQAL  472

Query  484  REYRKKMDIPAKLIVCGMTSNGFTIADPDDRAL  516
            + YR++M I AKLIV GM SN F+IADP D  +
Sbjct  473  QRYRREMGIAAKLIVVGMQSNRFSIADPRDGGM  505
```

E2 component p70 (P10515) (M2)

pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase [Alphaproteobacteria ba

Sequence ID: OFW69273.1 Length: 407 Number of Matches: 2

See 2 more title(s)

Range 1: 3 to 404 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
399 bits(1024)	2e-130	Compositional matrix adjust.	209/430(49%)	284/430(66%)	29/430(6%)

```
Query  219  MQVLLPALSPTMTMGTVQRWEKKVGEKLSEGDLLAEIETDKATIGFEVQEEGYLAKILVP  278
            +++L+PALSPTMT G + +W KK G+K+S GD++AEIETDKAT+  E  +EG L KIL+P
Sbjct  3    IEILMPALSPTMTEGNLVKWHKKEGDKVSAGDVVAEIETDKATMEVEAVDEGTLGKILIP  62

Query  279  EGTRDVPLGTPLCIIVEKEADISAFADYRPTEVTDLKPQVPPPTPPPVAAVPPTPQPLAP  338
            EGT +V + +P+ +I+EK  D     +Y+              PTP       P   P+ P
Sbjct  63   EGTENVKVNSPIALILEKGEDKKVLENYKA------------PTPAAAQKEEPVSAPV-P  109

Query  339  TPSAPCPATPAGPKGRVFVSPLAKKLAVEKGIDLTQVKGTGPDGRITKKDIDSFVPSKVA  398
            T ++P P T     GR+  SPLAK+LA EK IDL QV+G+GP GRI K+DID+FVP   A
Sbjct  110  TMASPAP-TIVLSTGRIVASPLAKRLATEKNIDLRQVQGSGPHGRIVKQDIDTFVPGSAA  168

Query  399  PAPAAVVPPTGPGMAPVPTG-VPTDIPISNIRRVIAQRLMQSKQTIPHYYLSIDVNMGEV  457
              A          P  TG ++ D P+SN+RRVIAQRL +SKQT+PH+YL++D  +  +
Sbjct  169  RGHAL----------PTHTGPLYQDKPVSNMRRVIAQRLTESKQTVPHFYLTLDCEIDAL  218

Query  458  LLVRKELNKILEGRSKISVNDFIIKASALACLKVPEANSSWMDTVIRQNHVVDVSVAVST  517
            L  R+ +N       K++VNDF++KA ALA   VP+AN+SW    IR      D+SVAV+
Sbjct  219  LAARQSINSHF--NVKVTVNDFVLKAVALALQDVPDANASWRGETIRYYTTSDISVAVAI  276

Query  518  PAGLITPIVFNAHIKGVETIANDVVSLATKAREGKLQPHEFQGGTFTISNLGMFGIKNFS  577
              GLITPIV  A+ K + TI+ +V +L  KA+EG+L+P EFQGG+F++SNLGMFGIK F
Sbjct  277  EDGLITPIVKMANFKSLLTISEEVKTLVQKAKEGRLKPEEFQGGSFSVSNLGMFGIKQFE  336

Query  578  AIINPPQACILAIGASEDKLVPADNEKGFDVASMMSVTLSCDHRVVDGAVGAQWLAEFRK  637
            AIINPPQ CILA+GA E +  P   E G  VA++M+ +LS DHRVVDG+VG++ L  F+K
Sbjct  337  AIINPPQGCILAVGAGEKR--PVVKEGGLAVATVMTCSLSVDHRVVDGSVGSKLLQAFKK  394

Query  638  YLEKPITMLL  647
            Y+E P  +L+
Sbjct  395  YIENPALLLV  404
```