UNITED STATES PATENT AND TRADEMARK OFFICE



Biological Sequence Searching

Christopher Babic Supervisory Patent Examiner, Technology Center 1600 USPTO March 2, 2022

UNITED STATES PATENT AND TRADEMARK OFFICE

Today's topics

- Pre-examination
- USPTO search tools & resources
- Claim language & search requests
- USPTO contacts



Pre-examination

Sequence compliance

- Patent applications that contain disclosures of unbranched:
 - Nucleotide sequences with <u>10 or more specifically defined and</u> <u>enumerated</u> residues;
 - Amino acid sequences with <u>4 or more specifically defined and</u> <u>enumerated</u> residues;

<u>Must conform to CFR §§ 1.821 -1.825 (see MPEP § 2422)</u>

• Pre-examination staff review newly submitted applications for compliance with sequence listing requirements prior to search and examination by a patent examiner

Sequence listings

SEQUENCE LISTING

<110>	Universitaetsklinikum Regensburg
<120>	NOVEL IL-3 ANTIBODIES AND THEIR USE IN DIAGNOSIS AND TREATMENT OF DISEASES OR MALFUNCTIONS ASSOCIATED WITH ELEVATED LEVELS OF IL-3
<130>	095229-0919889
<140>	US 14/400,954
<141>	2014-11-13
<160>	13
<170>	Patentin version 3.5
<210>	1
<211>	133
<212>	PRT
<213>	Homo sapiens
(220)	
(220)	Human TL 2 (without signal nontide)
<222>	(1)(133)
<400>	1
Ala Pr 1	o Met Thr Gln Thr Thr Pro Leu Lys Thr Ser Trp Val Asn Cys 5 10 15
Ser As	n Met Ile Asp Glu Ile Ile Thr His Leu Lys Gln Pro Pro Leu 20 25 30
Pro Le	u Leu Asp Phe Asn Asn Leu Asn Gly Glu Asp Gln Asp Ile Leu 35 40 45

• 37 CFR 1.821 requires that all applications containing covered nucleotide and/or amino acid sequences have a "Sequence Listing" that separately lists all the sequences.



Example

USPTO search tools & resources

Automated Biotechnology Sequence Search System (ABSS)

- In-house USPTO system designed to search electronic sequence listing data submitted by applicants
- Supports searching of molecular sequences (i.e., nucleic acids and amino acids) using data stored from both applicant submissions and public/commercial databases of published sequence information



ABSS database search

- Examiners and contractors conduct sequence homology searches, i.e., most similar to one or more sequences of interest in a patent application sequence listing
- ABSS finds the highest scoring local alignment between the query sequence and each sequence in the data set
- Reports database sequences that produced the highest scoring local alignments with the query sequence (separate homology search reports for each query sequence and data set)



ABSS interface

	ABSS database search	NCBI Blast search	Seq-to-seq alignment	User manual
	Self-Se	rvice ABSS Dat	abase Search	
Searches entered usin	g this interface are submitted <u>di</u> STIC-EIC1600 does <u>not</u> monitor, To re	rectly to ABSS. Examiners are res , track, record, or document searc quest an ABSS search from STIC-	ponsible for monitoring any sear ches submitted using this interfac EIC1600, <u>click here</u>	ches submitted using this interface. te in any way.
		Step 1: Get Sequen	ces	
	Examiner ID: Application Number		Get Sequences	
		Step 2: Select Seque	nces	
	Sequences:			
			A	
	Limit of 10 sequence	ces per application and no sequence ca	n be larger than 10,000 base pairs.	
		Step 3: Select Data	base	
		Prior Art	~	
	Standard All Protein	Search - Nucleic acids and prote Search - Translate any nucleic acid	ins to same type databases sequences to protein	
	 All Nucleic 	Acid Search - Back-translate any pr	otein sequences to nucleic acid	
		Submit Search		
	STIC-EIC1600 does not	maintain this interface. Report an	y problems with this interface to	PASM

ABSS data sets

Data set	Туре	Description
GenEmbl	na	NCBI GenBank database: all sections except "pat", "est" and "gss"
EST	na	NCBI GenBank database: "est" and "gss" sections
Geneseq_NA		Sequences from <u>Geneseq</u> , a proprietary database of biosequences covered in patents from multiple issuing authorities
Geneseq_AA	аа	
UniProt	аа	The <u>UniProt</u> knowledge base of amino acid sequences (both Swiss-Prot and TrEMBL sections)
PIR	аа	The last available version (2004) of the PIR-PSD data set
Pending_Patents_NA_Main Pending_Patents_NA_New	na	Sequences from certified sequence listings from pending patent applications available in ABSS.
Pending_Patents_AA_Main Pending_Patents_AA_New	аа	Note: the _Main data set is reconstituted twice a year; sequences from listings certified between reconstitutions are added to the _New set.
Published_Applications_NA_Main Published_Applications_NA_New	na	Sequences from published sequence listings.
Published_Applications_AA_Main Published_Applications_AA_New	аа	Note: the _Main data set is reconstituted twice a year; sequences from listings published between reconstitutions are added to the _New set.
Issued_Patents_NA	na	Sequences from sequence listings of issued applications.
Issued_Patents_AA	аа	Note: these set are reconstituted twice a year; sequences from applications issued between reconstitutions do not appear in these data sets.

<u>Search</u>

- Prior art
- Double patenting

These databases are produced by and accessible to only USPTO employees. Information in these databases is confidential. They contain sequences from all pending applications that have valid CRFs.



ABSS search capabilities

- Protein query vs. protein databases
- Nucleic acid query vs. nucleic acid databases
- Protein query vs. nucleic acid databases (back-translation)
- Nucleic acid query vs. protein databases
- Fragment of query sequence
- Score-over-length
- Oligomer search
- Length-limited
- Limited by % query match
- Sequences containing alternate or excluded residues

Advanced search techniques



NCBI Blast

ABSS database search	NCBI Blast search	Seq-to-seq alignment	User manual				
NCBI	Blast Search Su	Ibmission					
Caution: this site incorporates links to an outs	ide public search domain and should	anly be used for sequences from p	ublished applications				
Fetch a sequence (of a pending application with a PGPub):							
Application #: Fetched sequence will appear here.							
From:							
To: Clear Fetch							
	Submit NCBI Blast req	uest:					
NIH) U.S. National Library of Medicine NCBI	National Center for Biotechnology Informatio	n.:	Sign in to NCBI				
BLAST [®]		Home Recent Results Save	ed Strategies Help				
Get the latest research information l Find NCBI SARS-CoV-2 literature, see	from NIH: https://www.nih.gov/co quence, and clinical content: http:	ronavirus ://www.ncbi.nlm.nih.gov/sars-	cov-2/				
Basic Local Alignment Searc	h Tool	AST+ in Docker and on the cloud: Web	inar on December 9,				
bLAST i mos regions or similarity devicent biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. W BLAST - in Docker that is ready to use locally and in the cloud.							
	Wed, 02	Dec 2020 12:00:00 EST	More BLAST news				
Web BLAST							
77.2005	blastx translated nucleotide ► protein						
Nucleotide BLAST	tblastn protein > translated nucleotide	Protein BL/	AST				

1:1 Sequence Alignment

ABSS database search		NCBI Blast search		Seq-to-seq alignment	User manual
ABSS seq	uen	ce-to-sequ	en	ce alignment	
Sequence 1:		O Amino acid O Nucleic	acid	í.	
Application #: SEQ ID NO: From: To:	OR				
Clear Sequence 1					
Sequence 2: Application #: SEQ ID NO: From: To:	OR	• Amino acid • Nucleic	acid		
Clear Sequence 2		Clear all Submi			
Alignment result will appear here	when read	ty.			

• Multiple sequence alignments also available



Other search resources

- Multiple sequence alignments: EMBOSS/Showalign, CLUSTALW, GCGPileUp
- Variable residues: JEMBOSS/Fuzzpro
- STN CAS Registry, USGENE, PATGENE

-Uncommon amino acids -Blocking groups or end groups -Cyclic molecules -"D" amino acids -Variable residues -Structure searches



Claim language & sequence search requests

Standard search

Claim: A nucleic acid comprising the nucleotide sequence chosen from SEQ ID NO:1-9.

Request: SEQ ID NO: 1-9 vs. nucleic acid ABSS data sets.



Fragment search

Claim: A protein comprising the amino acid sequence selected from SEQ ID NO: 11-19 or residues 6-29 of SEQ ID NO:12.

Request: SEQ ID NO: 11-19 and residues 6-29 vs. amino acid ABSS data sets.



Length-limited search

Claim: An antisense oligonucleotide consisting of less than 50 nucleotides of SEQ ID NO:1.

Request: SEQ ID NO: 1, limiting the length of matching sequences to 49 nts or less.



Back-translated search

Claim: The nucleotide sequence encoding the protein having the amino acid sequence SEQ ID NO:11 or residues 8-232 of SEQ ID NO:11.

Request: SEQ ID NO: 11 (AA) and residues 8-232 of Seq 11 (AA) vs. nucleic acid ABSS data sets.



Percent identity search

Claim: A nucleic acid comprising a nucleotide sequence which has greater than 90% identity, similarity, or homology with one of the nucleotide sequences chosen from SEQ ID NO:1-9.

Option 1: (few hits): SEQ ID NO: 1-9 (standard).

Option 2: (many hits): SEQ ID NO: 1-9 w/ >89% query match or 100 scores and alignments, whichever is fewer.

Option 3: Score-over-length (SOL) search (depending on length of query sequences).

Score-over-length search

Claim: A probe or primer, 15-50 nucleotides long, which hybridizes at least 80% with the nucleic acid sequence of SEQ ID NO:1.

Request: SOL of SEQ ID NO: 1

minimum length = 15
maximum length = 50
SOL cutoff = 80%

• Parameters may be adjusted to accommodate claim breadth, etc.



Oligomer search

Claim: A peptide sequence comprising at least 10 consecutive amino acids of SEQ ID NO:13.

Request: oligomer of SEQ ID NO: 13. \blacktriangleright word size = 10 (all hits > = 10 aa)

Retrieves sequences that contain the longest contiguous series of 100% matches.



Sequence fusions

Claim: A protein sequence comprising SEQ ID NO: 2 and SEQ ID NO: 4

Request: fusion sequence comprising SEQ ID NO: 2 and SEQ ID NO: 4.

Any orderGap penalty set to minimum



Variable residue searches

Claim: A peptide comprising SEQ ID NO:22, where residues 2-4 may be isoleucine, valine, or leucine, residue 6 may be tryptophan or phenylalanine, residue 9 may be glycine or alanine, and residues 12-14 may be any amino acid.

Request: SEQ ID NO: 22 w/ alternate residues as claimed.

• Undefined residues and point mutations (e.g. SNPs) are searched similarly.



USPTO contacts

Sequence Help Desk

Assists with questions regarding compliance with 37 CFR 1.821-1.825 (sequence rules) or use of PatentIn or Checker software programs. Local: 571-272-2510
SequenceHelpDesk@uspto.gov

Christopher Babic: <u>christopher.babic@uspto.gov</u>

Phone: 571-272-8507



