

Motif sequence searching

This card summarizes STN motif symbols and characters as well as types of motif searches on STN.

STN motif symbols and characters

Use this symbol:	When you want to:	Example ⁽¹⁾	Possible Answers
^	Search at the beginning or the end of a sequence	=> S ^MCGIL/SQSP => S VCDS^/SQSP	"MCGIL....." ".....VCDS"
[]	Specify alternate residues	=> S LGP[VL]/SQSP	LGPV LGPL
[-] or [~]	Exclude one or more residues	=> S PTGK[-H]/SQSP => S PTGK[~H]/SQSP	PTGKACCD
{#,#} {# - #} {#}	Repeat preceding residue(s)	=> S GG(FL){1,3}/SQSP => S GG(FL){1-3}/SQSP => S GG(FL){3}/SQSP	GGFL GGFLFL GGFLFLFL
.	Specify gap(s) in the sequence	=> S SY.RPG/SQSP => S SY...RPG/SQSP	SYARPG SYAAARPG
	Specify alternate residues	=> S ACD KLM/SQSP	ACD KLM
		=> S A(CD KL)M/SQSP	ACDM AKLM
?	Repeat residue(s) zero or one time	=> S FLRR(RP)?K/SQSP	FLRRK FLRRRPK
*	Repeat residue(s) zero or more times	=> S KLK(WD)*N/SQSP	KLKN KLKWDN KLKWDWDN
+	Repeat residue(s) one or more times	=> S AQP+/SQSP	AQPP AQPPP AQPPPP
		=> S (AQP)+/SQSP	AQPAQP AQPAQPAQP AQPAQPAQPAQP
&	Join multiple sequence fragments together as one	=> S ACDKLM&KLKWDN/SQSP	ACDKLMKLKWDN

(1) To initiate sequence code match in CAS REGISTRYSM, use SEARCH or S, i.e., S DSDGP/SQEP.
To initiate sequence code match in USGENE, DGENE, or PCTGEN, use RUN GETSEQ, i.e., RUN GETSEQ DSDGP/SQEP.

STN sequence code match types

Search Type	Amino Acids	Nucleic Acids	Example
EXACT	/SQEP	/SQEN	=> S DSDGP/SQEP => S GGAATT/SQEN
EXACT FAMILY	/SQEFP	----	=> S DSDGP/SQEFP
SUBSEQUENCE	/SQSP	/SQSN	=> S DSDGP/SQSP => S GGAATT/SQSN
SUBSEQUENCE FAMILY	/SQSFP	----	=> S DSDGP/SQSFP

STN amino acid family substitution definitions

Groups	Amino Acids
Neutral Weak Hydrophobic	Alanine (Ala, A) Glycine (Gly, G) Proline (Pro, P) Serine (Ser, S) Threonine (Thr, T)
Acid-Amines Hydrophilic	Aspartic Acid (Asp, D) Asparagine (Asn, N) Glutamic Acid (Glu, E) Glutamine (Gln, Q)
Basic Hydrophobic	Arginine (Arg, R) Histidine (His, H) Lysine (Lys, K);
Hydrophobic	Isoleucine (Ile, I) Leucine (Leu, L) Methionine (Met, M) Valine (Val, V)
Aromatics	Phenylalanine (Phe, F) Tryptophan (Trp, W) Tyrosine (Tyr, Y)
Cross Linking	Cysteine (Cys, C)



A division of the
American Chemical Society

March 2009
CAS2567-0309

CAS Customer Center
Phone: 800-753-4227 (North America)
614-447-3700 (worldwide)
Fax: 614-447-3751
E-mail: help@cas.org
Internet: www.cas.org