

Sequence Symbols

GCG uses the letter codes for amino acid codes and nucleotide ambiguity proposed by IUB (Nomenclature Committee, 1985, Eur. J. Biochem. 150; 1-5). These codes are compatible with the codes used by the EMBL, GenBank, and PIR data libraries.

Nucleotides

The meaning of each symbol, its complement, and the Cambridge (Staden) equivalents are shown below.

<u>IUB/GCG</u>	<u>Meaning</u>	<u>Complement</u>	<u>Staden/Sanger</u>
A	A	T	A
C	C	G	C
G	G	C	G
T/U	T	A	T
M	A or C	K	5
R	A or G	Y	R
W	A or T	W	7
S	C or G	S	8
Y	C or T	R	Y
K	G or T	M	6
V	A or C or G	B	not supported
H	A or C or T	D	not supported
D	A or G or T	H	not supported
B	C or G or T	V	not supported
X/N	G or A or T or C	X	-/X
(Gap Character) .	not G or A or T or C	.	not supported

Amino Acids

Here is a list of the standard one-letter amino acid codes and their three-letter equivalents. The synonymous codons and their depiction in the IUB codes are shown. You should recognize that the codons following semicolons (;) are not sufficiently specific to define a single amino acid even though they represent the best possible backtranslation into the IUB codes!

<u>IUB Symbol</u>	<u>3-letter</u>	<u>Meaning</u>	<u>Codons</u>	<u>Depiction</u>
A	Ala	Alanine	GCT, GCC, GCA, GCG	!GCX
B	Asp, Asn	Aspartic, Asparagine	GAT, GAC, AAT, AAC	!RAY
C	Cys	Cysteine	TGT, TGC	!TGY
D	Asp	Aspartic	GAT, GAC	!GAY
E	Glu	Glutamic	GAA, GAG	!GAR
F	Phe	Phenylalanine	TTT, TTC	!TTY
G	Gly	Glycine	GGT, GGC, GGA, GGG	!GGX
H	His	Histidine	CAT, CAC	!CAY
I	Ile	Isoleucine	ATT, ATC, ATA	!ATH
K	Lys	Lysine	AAA, AAG	!AAR
L	Leu	Leucine	TTG, TTA, CTT, CTC, CTA, CTG	
!TTR, CTX, YTR; YTX				
M	Met	Methionine	ATG	!ATG
N	Asn	Asparagine	AAT, AAC	!AAY
P	Pro	Proline	CCT, CCC, CCA, CCG	!CCX
Q	Gln	Glutamine	CAA, CAG	!CAR
R	Arg	Arginine	CGT, CGC, CGA, CGG, AGA, AGG	
!CGX, AGR, MGR; MGX				
S	Ser	Serine	TCT, TCC, TCA, TCG, AGT, AGC	!TCX, AGY; WSX
T	Thr	Threonine	ACT, ACC, ACA, ACG	!ACX
V	Val	Valine	GTT, GTC, GTA, GTG	!GTX
W	Trp	Tryptophan	TGG	!TGG
X	Xxx	Unknown		!XXX
Y	Tyr	Tyrosine	TAT, TAC	!TAY
Z	Glu, Gln	Glutamic, Glutamine	GAA, GAG, CAA, CAG	!SAR
*	End	Terminator	TAA, TAG, TGA	!TAR, TRA; TRR