

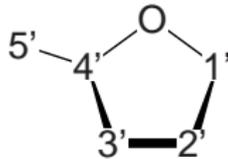
# *Introduction to Computational Biology; An Evolutionary Approach: Appendices*

Bernhard Haubold & Thomas Wiehe

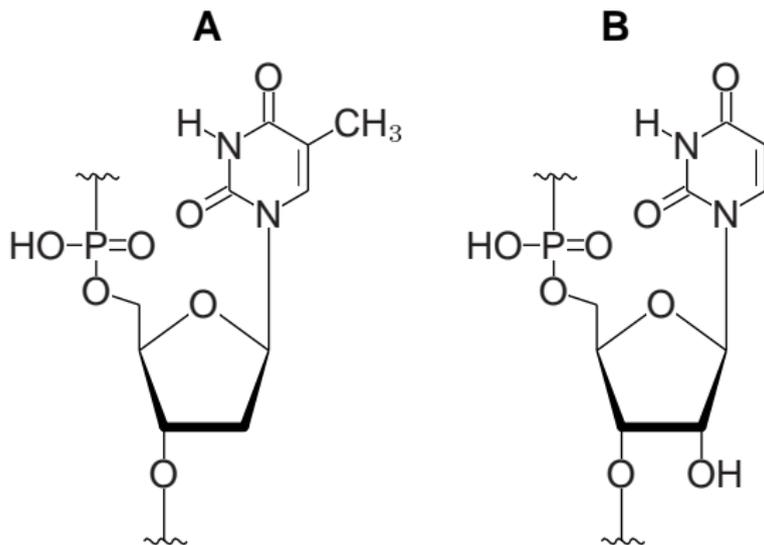
# Nomenclature of Nucleic Acid Constituents

Base	Structure	Chemical Class	Nucleoside	Symbol
adenine		purine	adenosine	A
guanine		purine	guanosine	G
cytosine		pyrimidine	cytidine	C
thymine		pyrimidine	thymidine	T
uracil		pyrimidine	uridine	U

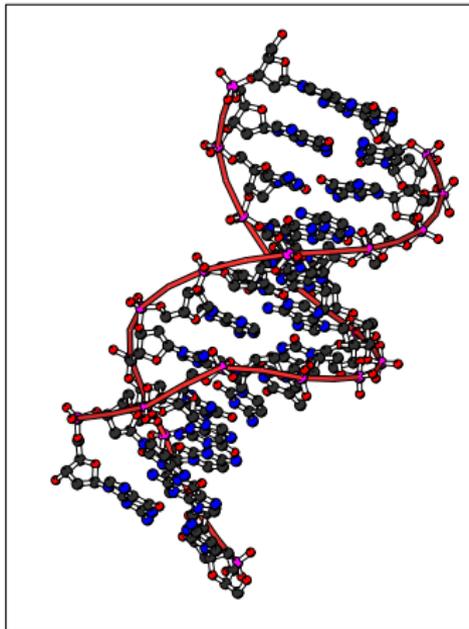
# Carbon Atoms in Ribose



# DNA & RNA



# Structure of DNA



# Ambiguity Codes for DNA

Symbol	Meaning
A	A
B	TGC
C	C
D	TGA
G	G
H	TCA
K	TG
M	CA
N	TGCA

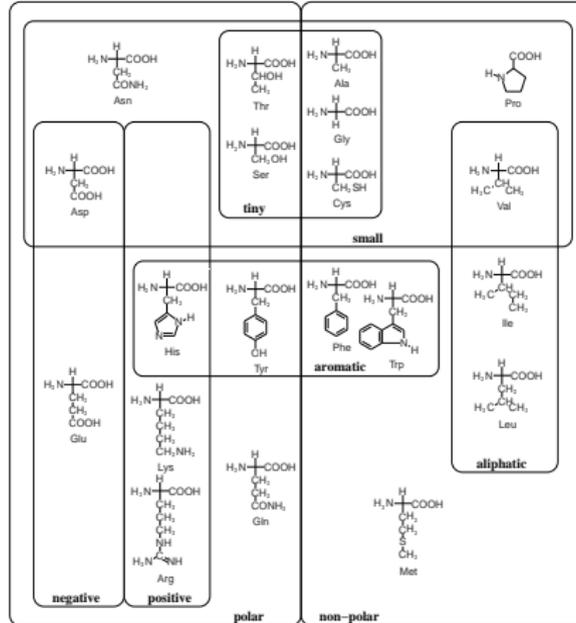
Symbol	Meaning
R	GA
S	GC
T	T
U	T
V	GCA
W	TA
X	TGCA
Y	TC

# Amino Acid Nomenclature

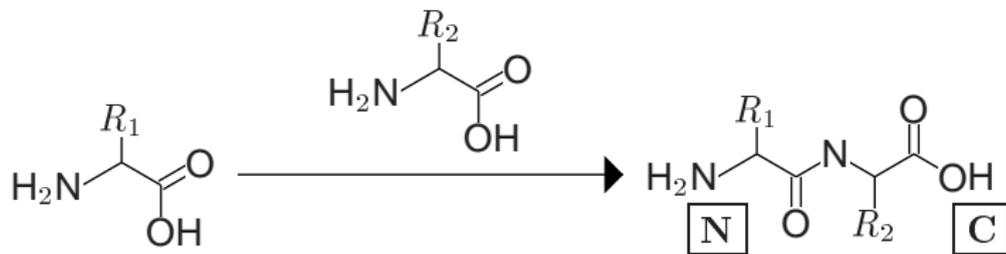
One	Three	Name
A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine

One	Three	Name
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

# Physico-Chemical Properties of Amino Acids



# Synthesis of Peptide Bond



# Standard Genetic Code

5' end	second position				3' end
	T	C	A	G	
<b>T</b>	Phe/F	Ser/S	Tyr/Y	Cys/C	T
	Leu/L		Ter/*	Ter/*	A
		Pro/P	His/H	Arg/R	G
					Gln/Q
Ile/I	Thr/T	Asn/N	Ser/S	C	
Met/M		Lys/K	Arg/R	A	
<b>C</b>	Val/V	Ala/A	Asp/D	Gly/G	T
			Glu/E		C
	<b>A</b>	Gly/G	Gly/G		A
G					
<b>G</b>	Val/V	Ala/A	Asp/D	Gly/G	T
			Glu/E		C
<b>A</b>	Val/V	Ala/A	Asp/D	Gly/G	A
			Glu/E		G
<b>G</b>	Val/V	Ala/A	Asp/D	Gly/G	T
			Glu/E		C

# Bioinformatics Applications

Program	Purpose	Reference	Website
BLAST	Database search	[2]	<a href="http://www.ncbi.nlm.nih.gov/BLAST">www.ncbi.nlm.nih.gov/BLAST</a>
clustalw	Multiple sequence alignment	[12]	<a href="http://ftp-igbmc.u-strasbg.fr/pub">ftp-igbmc.u-strasbg.fr/pub</a>
EMBOSS	Various bioinformatics tasks	[10]	<a href="http://emboss.sourceforge.net">emboss.sourceforge.net</a>
DnaSP	Analyze DNA sequence polymorphisms	[11]	<a href="http://www.ub.es/dnasp">www.ub.es/dnasp</a>
FASTA	Database search	[9]	<a href="http://fasta.bioch.virginia.edu">fasta.bioch.virginia.edu</a>
GenScan	Gene prediction	[3]	<a href="http://genes.mit.edu/GENSCAN.html">genes.mit.edu/GENSCAN.html</a>
gff2aplot	Plotting sequence comparisons	[1]	<a href="http://genome.imim.es/software">genome.imim.es/software</a>
ms	Coalescent simulations	[6]	<a href="http://home.uchicago.edu/rhudson1">home.uchicago.edu/rhudson1</a>
MUMmer	Genome alignment	[4]	<a href="http://www.tigr.org/software">www.tigr.org/software</a>
PHYMLIP	Phylogeny reconstruction	[5]	<a href="http://evolution.genetics.washington.edu">evolution.genetics.washington.edu</a>
SGP-2	Gene prediction	[8]	<a href="http://www1.imim.es/software/sgp2">www1.imim.es/software/sgp2</a>
Slam	Gene prediction	[7]	<a href="http://bio.math.berkeley.edu/slam">bio.math.berkeley.edu/slam</a>

# Bioinformatics Databases

Name	Purpose	Website
GenBank	primary sequence repository	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>
EMBL	primary sequence repository	<a href="http://www.ebi.ac.uk/embl">www.ebi.ac.uk/embl</a>
DDBJ	primary sequence repository	<a href="http://www.ddbj.nig.ac.jp">www.ddbj.nig.ac.jp</a>
SwissProt	curated protein sequences	<a href="http://www.expasy.org/sprot">www.expasy.org/sprot</a>
PDB	macromolecular structures	<a href="http://www.rcsb.org/pdb">www.rcsb.org/pdb</a>
INTERPRO	protein motifs	<a href="http://www.ebi.ac.uk/interpro">www.ebi.ac.uk/interpro</a>

# References

- 1 J. F. Abril, R. Guigó, and T. Wiehe.  
git2gap1.0rc: plotting sequence comparisons.  
*Bioinformatics*, 19:2477–2479, 2004.
- 2 S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.  
Basic local alignment search tool.  
*Journal of Molecular Biology*, 215:403–410, 1990.
- 3 C. B. Burge and S. Karlin.  
Prediction of complete gene structures in human genomic DNA.  
*Journal of Molecular Biology*, 288:78–94, 1997.
- 4 A. L. Delcher, S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg.  
Alignment of whole genomes.  
*Nucleic Acids Research*, 27:2369–2376, 1999.
- 5 J. Felsenstein.  
PHYLIP (phylogeny inference package), 1993.
- 6 R. R. Hudson.  
Generating samples under a Wright-Fisher neutral model of genetic variation.  
*Bioinformatics*, 18:337–338, 2002.
- 7 L. Pachter, M. Alexandersson, and S. Cawley.  
Applications of generalized pair Hidden Markov Models to alignment and gene finding problems.  
In *Proceedings of the Fifth Annual Conference on Computational Molecular Biology (RECOMB 2001)*, pages 241–248. ACM Press, New York, 2001.
- 8 G. Parra, P. Agarwal, J. F. Abril, T. Wiehe, J. Fickett, and R. Guigó.  
Comparative gene prediction in human and mouse.  
*Genome Research*, 13:108–117, 2003.
- 9 W. R. Pearson and D. J. Lipman.  
Improved tools for biological sequence comparison.  
*Proceedings of the National Academy of Sciences, USA*, 85:2444–2448, 1988.
- 10 P. Rice and A. Bleasby.  
EMBOSS: The European Molecular Biology Open Software Suite.  
*Trends in Genetics*, 16:276–277, 2000.
- 11 J. Rozas, J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas.  
DnaSP, DNA polymorphism analyses by the coalescent and other methods.  
*Bioinformatics*, 19:2496–2497, 2003.
- 12 J. D. Thompson, D. G. Higgins, and T. J. Gibson.  
CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice.  
*Nucleic Acids Research*, 22:4673–4680, 1994.