BIOBITS NUCLEIC ACID (BN) KITS

GENERAL NOTES TO ACCOMPANY ALL KITS

(1) Codes

A one-letter code is in current use for the five major bases found in DNA and messenger (m-) RNA. The transfer (t-) RNA's contain other purine and pyrimidine bases and so many substituted bases that it is not possible to name all with single letters. Recommended international coding for these and other bases has been published as these notes were being written (J. Mol. Biol., 1971, 55, 299; Arch. Biochem. Biophys., 1971, 142, 1). The coding used herein is that shown in 'Atlas of Protein Sequence and Structure' 1969, edited by Margaret O. Dayoff (National Biomedical Research Foundation, 11200 Lockwood Drive, Silver Spring, Maryland 20901, U.S.A.) and discussed on pp. 89-101 and D299/230. The Biobits colour coding is described with the one and three-letter codes in the 'Biobits and Atomunits' manual which accompanies every kit; see pp. 17 and 10.

(2) Sequences

No complete DNA or m-RNA has been sequenced. Many t-RNA's have been sequenced completely. They all contain some 70-90 bases which can be hydrogen-bonded into the accepted clover leaf secondary structure as shown in the accompanying figure; all are numbered from the 5' phosphate end through to the 3' free CH end which carries the aminoacid being transferred. Various 5S RNA's derived from ribosomes but of unknown function have been sequenced. These all seem to be 120 residues long, to contain only the four usual RNA bases and are numbered from the 5' phosphate.

(3) Co-ordinates

No 3D structures have yet been deduced by crystallographic methods and so no co-ordinates are known. As these eventually become available they will be supplied with a frame for 3D construction of models.

(4) Models available and model building

a. DNA As no DNA has been sequenced, models can only be built of a generalised DNA. Such models will show the standard base pairing, double helix, etc., as described in the manual p.39.

b. DNA-RNA Transcription is readily observed from this model.

c. t-RNA's Sets of five baker's yeast or E. coli t-RNA's allow the construction and comparison of these molecules. Obviously all such molecules must be fairly similar in 3D as they must fit the ribosome during protein synthesis. They must differ adequately in order neither to pick up and transfer a wrong aminoacid nor to fit in any but the correct position on the m-RNA. A study of such a set may be valuable in hypothesising on the correct 3D nature of the 2D clover leaf structure. In the manual on p. 10, xanthosine is shown as 'X' but this base has not yet been found in any RNA. In the table enclosed with these notes, the letter 'X' is used alone to indicate either a base or unknown structure or, as in the case of 'XC', a known base with an unknown substituent. Each t-RNA kit is supplied with the correct aminoacid monomer set and with lengths of messenger RNA backbone for hydrogen-bonding onto the triplet anticodons.

d. 5S RNA's A set of two such RNA's has been made available to aid preliminary building studies of secondary structures.

d. A protein synthesis model will shortly be available.

(5) The model designers and manufacturers would be interested to hear of any comments, suggestions for further kits, reactions of staff and students, as well as any information you think might be worth passing on to other users of these kits.

(6) Much of the data reproduced here has been obtained from the 'Protein Atlas' (see 1 above). We are pleased to be able to record our debt to those concerned for permission to use their data. Models for any molecular structure described in the Atlas or elsewhere can be obtained on request provided that a full reference is given.

Kit BN 1. - This is a general introductory kit to illustrate the general nucleic acids and their major properties. It permits simultaneous building of a 25 pair DNA double helix, a five-segment DNA-RNA transcription unit at one end of the DNA helix and a complete Ala t-RNA. Also included are various monomer nucleosides to illustrate synthesis and sequencing techniques.

Kit BN 2 - Five complete t-RNA's derived from Bakers Yeast (Ala, Ser 2, Phe, Tyr, Val.)

Kit BN 3 - Five complete t-RNA's from E. coli. (N-Formyl-Met, Met, Phe, Tyr, Val.)

Both t-RNA kits allow the construction of five different molecules in the clover-leaf, hydrogen-bonded form. For each RNA there is included its specific aminoacid which can be attached to the free 3' end, and a piece of m-RNA backbone which can be correctly coded for the aminoacid and hydrogen-bonded to the anticodon (nodoc) triplet. Each backbone is supplied as a continuous section -- if the section is too long, the extra length can be cut off after ensuring that the structure is correct. t-RNA sequences are shown in the enclosed table. To simplify the construction of the correct primary sequences, each horizontal sequence in the table should be lightly coloured with different colours. The table of abbreviations is included for use with the sequences and diagrams of clover leafs. Bases underlined in the t-RNA table are those present in the hydrogen-bonded arms; the numbers below the underlined sections indicate which lengths bond to which. To clarify this, compare the underlined lengths with the clover leaf diagrams for the bakers yeast t-RNA's. Unknown substituents labelled 'X' in the table are all indicated by a white, small sphere bit.

Kit BN 4 Two complete 5S RNA's - E. coli and human KB. Each backbone is 120 residues long and built on one wire to ensure easy manipulations during building of hypothetical secondary structures.

Additional Bits - not shown on pp 10-11 of manual

As more and more tRNA's are sequenced it is being found that they contain minor or unusual purine or pyrimidine bases. Furthermore the major five bases have various other substituents not previously listed. Occasionally the structure has not yet been determined and this is shown by an x or X (see abbreviations). The following updates the available Bits.

X = unknown base = large flat-facetted green, pale mottled
S = thiouridine = small flat-facetted green, pale mottled

small spheres		multifacetted thick	
7 Methyl	brown	NP	light blue
3 Methyl	pearl white	SP	dark blue
	lilac	TCP	pink
		unknown substituent on any base	white
			orange
			mauve
			ivory
Formyl-(Met)	white tube		(beige)

GENETIC CODE: Codon → anticodon → aminoacid.

Codons 5'→3'	Anticodon 3'→5' yeasts	Anticodon 3'→5' E. coli	Aminoacid	Codons 5'→3'	Anticodon 3'→5' yeasts	Anticodon 3'→5' E. coli	Aminoacid
GG ucag			Gly	UG uc..			Cys
GC ucag	C G I		Ala	AU ...g		U A xC	Met
GU ucag	C A I	C A X	Val	AU ...g		U A G	f-Met
CU ucag		G A C	Leu	UC ucag	A G I		Ser
UU ..ag			Leu	AG uc..			Ser
AU uca.	U A I		Ile	AC ucag			Thr
GA uc..	C U G		Asp	CC ucag			Pro
GA ..ag			Glu	UU uc..	A A omG	A A G	Phe
AA uc..			Asn	UA uc..	A PSU G	A U xG	Tyr
CA ..ag			Gln	UG ...g		A C C	Trp
CA uc..			His	UA ..ag			End
AA ..ag			Lys	UG ..ag			End
AG ..ag			Arg				
CG ucag			Arg				

The table shows every codon for each aminoacid. Each anticodon probably reacts to only one of the possible codons. The anticodon is written in the 3'→5' direction to highlight the anti-parallel base pairing.

Sequences of 5S RNA

5S RNA - ESCHERICHIA COLI (2 STRAINS)

BASE COMPOSITION

41	GUANINE	G	36	CYTOSINE	C
23	ADENINE	A	20	URACIL	U

TOTAL NO. OF BASES = 120

```
                5         10        15        20        25        30
  1 U G C C U G G C G G C G U A G C G C C G U G G U C C C A C
 31 C U G A C C C C A U G C C G A A C U C A G A A G U G A A A C
 61 G C C G U A G C G C C C G A U G G U A G U G U G G G G U C U C
 91 C C C A U G C G A G A G U A G G G A A C U G C C A G G C A U
```

BROWNLEE,G.G., SANGER,F. AND BARRELL,B.G., VOL.34, PP.379-412, 1968 J. MOL. BIOL.,

5S RNA - HUMAN KB CARCINOMA CELLS

BASE COMPOSITION

39	GUANINE	G	33	CYTOSINE	C
22	ADENINE	A	26	URACIL	U

TOTAL NO. OF BASES = 120

```
                5         10        15        20        25        30
  1 G U C U A C G G C C A U A C C A C C C U G A A C G C G C C C
 31 G A U C U C G U C U G A U C U C G G A A G C U A A G C A G G
 61 G U C G G G C C U G G U U A G U A C U U G G A U G G G A G A
 91 C C G C C U G G G A A U A C C G G G U G C U G U A G G C U U
```

NUCLEOTIDE SEQUENCE OF KB CELL 5S RNA
FORGET,B.G. AND WEISSMAN,S.M., SCIENCE, VOL. 158, 1967 PP.1695-1699,

Issue 3809 29 Dec

Phenylalanine 1 – Baker's Yeast	G	C	G	G	A	U	U	U	A	NMG	C	U	C	-	-	A	G	DHU	-	DHU
Phenylalanine 1 – Wheat	G	C	G	G	G	G	A	U	A	NMG	C	U	C	-	-	A	G	DHU	-	DHU
Tyrosine 1 – Baker's Yeast	C	U	C	U	C	G	G	U	A	NMG	C	C	A	-	-	A	G	DHU	-	DHU
Tyrosine 1 – Torula Yeast	C	U	C	U	C	G	G	U	1MG	NMG	C	C	A	-	-	A	G	DHU	-	DHU
Tyrosine 1 – *Escherichia coli*	G	G	U	G	G	G	G	XU	U	C	C	C	G	-	-	A	G	-	-	C
Tyrosine 2 – *Escherichia coli*																				
Serine 1 – Baker's Yeast	G	G	C	A	A	C	U	U	G	G	C	NAC	G	-	-	A	G	DHU	-	-
Serine 2 – Baker's Yeast																				
Serine 1 – Rat	G	U	A	G	U	C	G	U	G	G	C	NAC	G	-	-	A	G	DHU	-	-
Alanine 1 – Baker's Yeast	G	G	G	C	G	U	G	U	1MG	G	C	G	C	G	U	A	G	DHU	-	C
Alanine 2 – Baker's Yeast																				
Valine 1 – Baker's Yeast	G	G	U	U	U	C	G	U	1MG	G	U	C	PSU	-	-	A	G	DHU	-	C
Valine 1 – Torula Yeast	G	G	U	U	C	C	G	U	1MG	G	U	C	PSU	-	-	A	G	DHU	-	DHU
Isoleucine 1 – Torula Yeast	G	G	U	C	C	C	U	U	G	G	C	C	C	-	-	A	G	DHU	-	DHU
Methionine 1 – *Escherichia coli*	G	G	C	U	A	C	G	S	A	G	C	U	C	-	-	A	G	DHU	-	DHU
N-Formyl-Methionine 1 – *Escherichia coli*	C	G	C	G	G	G	G	S	G	G	A	G	C	-	-	A	G	C	C	U
N-Formyl-Methionine 2 – *Escherichia coli*																				
						1						2								
Phenylalanine 1 – *Escherichia coli*	G	C	C	C	G	C	U	C	A	G	C	G	G	DHU	C	G	A	S	C	C
Valine – *E coli*	G	G	G	U	G	A	U	S	A	G	C	U	C	-	-	A	G	-	C	DHU

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
Phe 1 – B.Y.	G	G	G	-	-	A	G	A	G	C	DMG	C	C	A	G	A	OMC	U	DMG	A	A	X	A
Phe 1 – Wh.	G	G	G	-	-	A	G	A	G	C	DMG	PSU	C	A	G	A	OMC	U	DMG	A	A	X	A
Tyr 1 – B.Y.	DMG	G	DHU	DHU	DHU	A	A	G	G	C	DMG	C	A	A	G	A	C	U	G	PSU	A	NPA	A
Tyr 1 – T.Y.	DMG	G	DHU	DHU	DHU	A	A	G	G	C	DMG	PSU	C	A	G	A	C	U	G	PSU	A	NPA	A
Tyr 1 – *E.c.*	DMG	G	C	C	A	A	A	G	G	G	A	G	C	A	G	A	C	U	XG	U	A	XA	A
Tyr 2 – *E.c.*																							
Ser 1 – B.Y.	DMG	G	DHU	DHU	-	-	A	G	G	C	DMG	A	A	A	G	A	PSU	U	I	G	A	NPA	A
Ser 2 – B.Y.																							
Ser 1 – Rt.	DMG	G	DHU	DHU	-	-	A	G	G	C	DMG	A	PSU	G	G	A	3MC	U	I	G	A	NPA	A
Ala 1 – B.Y.	G	G	DHU	-	-	A	G	C	G	C	DMG	C	U	C	C	C	U	U	I	G	C	MI	PSU
Ala 2 – B.Y.																							
Val 1 – B.Y.	G	G	DHU	DHU	-	A	U	G	G	C	A	PSU	C	U	G	C	PSU	U	I	A	C	A	C
Val 1 – T.Y.	G	G	DHU	C	-	A	U	G	G	C	A	PSU	C	U	G	C	PSU	U	I	A	C	A	C
Ile 1 – T.Y.	G	G	DHU	DHU	-	A	A	G	G	C	DMG	PSU	G	G	U	G	C	U	I	A	U	TCP	A
Met 1 – *E.c.*	DMG	G	DHU	DHU	-	A	G	A	G	C	A	C	A	U	C	A	C	U	XC	A	U	XA	A
F-Met 1 – *E.c.*	G	G	DHU	-	-	A	G	C	U	C	G	U	C	G	G	G	OMC	U	C	A	U	A	A
F-Met 2 – *E.c.*																							
								2					3						ANTICODON				
Phe – *E.c.*	A	G	DHU	A	G	G	U	A	G	A	G	C	A	G	G	G	G	G	A	PSU	U	LG	A
Val – *E.c*	G	G	-	-	G	A	G	A	G	C	A	C	C	U	C	C	C	U	X	A	C	MA	A

	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	
	PSU	5MC	U	G	G	A	G	7MG	-	-	-	-	-	-	-	-	-	U	C	5MC	U	G	U	G	Phe 1 – B.Y.
	PSU	C	U	G	A	A	G	7MG	-	-	-	-	-	-	-	-	DHU	C	G	C	G	G	G	G	Phe 1 – Wh.
	PSU	C	U	U	G	A	G	A	-	-	-	-	-	-	-	-	DHU	5MC	G	G	G	C	G	G	Tyr 1 – B.Y.
	PSU	C	U	G	A	A	C	A	-	-	-	-	-	-	-	-	DHU	5MC	G	G	G	C	G	G	Tyr 1 – T.Y.
	PSU	C	U	G	C	C	G	U	-	C	A	U	C	G	A	C	U	U	C	G	A	A	G	G	Tyr 1 – *E.c.*
												···C	A···												Tyr 2 – *E.c.*
	PSU	C	U	U	U	OMU	G	G	G	C	U	C	U	G	C	C	C	G	5MC	G	C	A	G	G	Ser 1 – B.Y.
												···U···													Ser 2 – B.Y.
	OMP	C	C	A	U	OMU	G	G	G	G	U	3MC	U	C	C	C	C	G	5MC	G	C	A	G	G	Ser 1 – Rt.
	G	G	G	A	G	A	C	-	-	-	-	-	-	-	-	-	-	U	C	U	C	C	G	G	Ala 1 – B.Y.
																		···DHU···							Ala 2 – B.Y.
	G	C	A	G	A	A	C	-	-	-	-	-	-	-	-	-	DHU	5MC	C	C	C	A	G	G	Val 1 – B.Y.
	G	C	A	G	A	A	C	-	-	-	-	-	-	-	-	-	-	5MC	C	C	C	A	G	G	Val 1 – T.Y.
	C	G	C	C	A	A	G	A	-	-	-	-	-	-	-	-	DHU	5MC	A	G	C	A	G	G	Ile 1 – T.Y.
	PSU	G	A	U	G	G	G	7MG	-	-	-	-	-	-	-	-	X	C	A	C	A	G	G	G	Met 1 – *E.c.*
	C	C	C	G	A	A	G	A	-	-	-	-	-	-	-	-	U	C	G	U	C	G	G	G	F-Met 1 – *E.c.*
								···7MG···																	F-Met 2 – *E.c.*
		3							5						5						4				
	A	SPA	A	PSU	C	C	C	C	G	U	7MG	-	-	-	-	-	-	7MG	C	C	U	U	G	G	Phe – *E.c.*
	G	G	A	G	G	G	G	7MG	-	-	-	-	-	-	-	-	-	U	C	G	C	C	G	G	Val – *E.c*

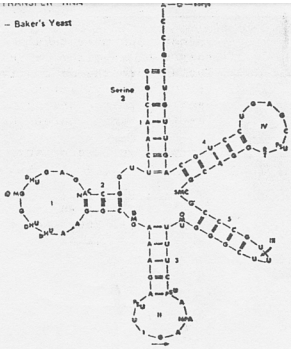
	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	
	RT	PSU	C	G	MA	U	C	C	A	C	A	G	-	A	A	U	U	C	G	C	A	C	C	A	Phe 1 – B.Y.
	RT	PSU	C	G	MA	U	C	C	A	C	G	C	-	U	C	A	C	C	G	C	A	C	C	A	Phe 1 – Wh.
	RT	PSU	C	G	MA	C	U	C	G	C	C	C	-	C	C	G	G	G	G	A	A	C	C	A	Tyr 1 – B.Y.
	RT	PSU	C	G	MA	A	U	C	G	C	C	C	-	C	C	G	A	G	G	A	A	C	C	A	Tyr 1 – T.Y.
	RT	PSU	C	G	A	A	U	C	C	U	U	C	-	C	C	C	C	A	C	C	A	C	C	A	Tyr 1 – *E.c.*
																									Tyr 2 – *E.c.*
	RT	PSU	C	A	A	A	U	C	C	U	G	C	-	A	G	U	U	G	U	C	G	C	C	A	Ser 1 – B.Y.
				···G···		···G···																			Ser 2 – B.Y.
	RT	PSU	C	G	MA	A	U	C	C	U	G	C	-	C	G	A	C	U	A	C	G	C	C	A	Ser 1 – Rt.
	RT	PSU	C	G	A	U	U	C	C	G	G	A	-	C	U	C	G	U	C	C	A	C	C	A	Ala 1 – B.Y.
																									Ala 2 – B.Y.
	RT	PSU	C	G	MA	U	C	C	U	G	G	G	G	C	G	A	A	A	U	C	A	C	C	A	Val 1 – B.Y.
	RT	PSU	C	G	MA	U	C	C	U	G	G	G	-	C	G	A	A	A	U	C	A	C	C	A	Val 1 – T.Y.
	RT	PSU	C	G	MA	U	C	C	U	G	C	U	-	A	G	G	G	A	A	C	A	C	C	A	Ile 1 – T.Y.
	RT	PSU	C	G	A	A	U	C	C	G	G	U	-	C	G	U	A	G	C	C	A	C	C	A	Met 1 – *E.c.*
	RT	PSU	C	A	A	A	U	C	C	G	G	C	-	C	C	C	C	G	C	A	A	C	C	A	F-Met 1 – *E.c.*
																									F-Met 2 – *E.c.*
								4									1								
	RT	PSU	C	G	A	U	U	C	C	G	A	G	-	U	C	C	G	G	G	C	A	C	C	A	Phe – *E.c*

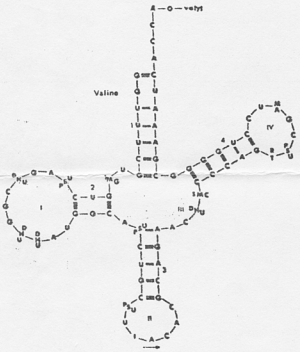
Abbreviations

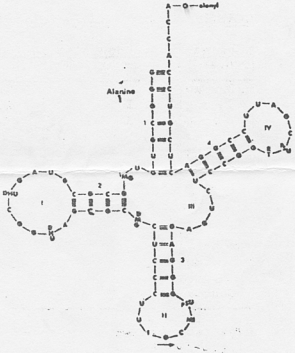
The unmodified and modified nucleosides in the aligned sequences are designated by abbreviations using 1 to 3 characters for each nucleoside. The standardized rules on nucleoside abbreviations could not be followed completely because some are incompatible with computer type. As is conventional, the left end of the sequence is that bearing a 5'-phosphate and the right end is that bearing a 3'-hydroxyl group.

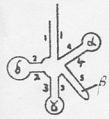
A	–	ADENOSINE	G	–	GUANOSINE
DMA	–	N(6)-DIMETHYLADENOSINE	DMG	–	N(2)-DIMETHYLGUANOSINE
MA	–	1-METHYLADENOSINE	1MG	–	1-METHYLGUANOSINE
NPA	–	N(6)-ISOPENTENYLADENOSINE	NMG	–	N(2)-METHYLGUANOSINE
SPA	–	N(6)-ISOPENTENYL-2-METHYLTHIO-ADENOSINE	OMG	–	2'-O-METHYLGUANOSINE
			7MG	–	N(7)-METHYLGUANOSINE
TCP	–	N-(PURIN-6-YLCARBAMOYL)-THREONINE RIBOSIDE	XG	–	GUANOSINE DERIVATIVE
XA	–	ADENOSINE DERIVATIVE	I	–	INOSINE
			MI	–	1-METHYLINOSINE
C	–	CYTIDINE			
3MC	–	3-METHYLCYTIDINE	RT	–	RIBOTHYMIDINE
5MC	–	5-METHYLCYTIDINE			
NAC	–	N(4)-ACETYLCYTIDINE	U	–	URIDINE
OMC	–	2'-O-METHYLCYTIDINE	S	–	THIOURIDINE
XC	–	CYTIDINE DERIVATIVE	DHU	–	5,6-DIHYDROURIDINE
			OMP	–	2'-O-METHYLPSEUDOURIDINE
			OMU	–	2'-O-METHYLURIDINE
			PSU	–	PSEUDOURIDINE
			XU	–	URIDINE DERIVATIVE
			X	–	UNKNOWN

– Baker's Yeast









α = TψC arm
β = variable arm
γ = Anticodon
δ = DHU arm