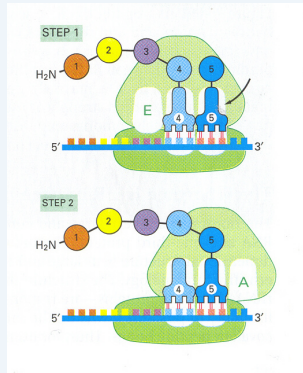
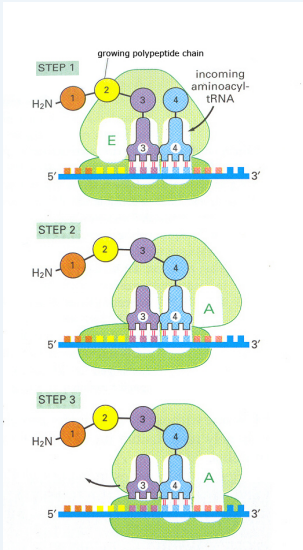


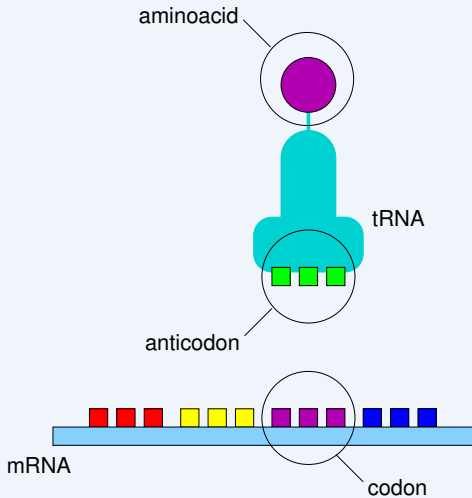
Codon Bias with PRISM

2IM24/25, Fall 2007

from RNA to protein

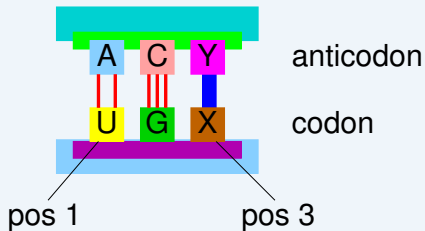


mRNA vs. tRNA



codon-anticodon matching

- ▶ Watson-Crick base pairing A–U and C–G
- ▶ binding first two nucleotide pairs strict
third pair may be non-strict



the genetic code

Ala	GCU, GCC, GCA, GCG	Leu	UUA, UUG, CUU, CUC, CUA, CUG
Arg	CGU, CGC, CGA, CGG, AGA, AGG	Lys	AAA, AAG
Asn	AAU, AAC	Met	AUG
Asp	GAU, GAC	Phe	UUU, UUC
Cys	UGU, UGC	Pro	CCU, CCC, CCA, CCG
Gln	CAA, CAG	Ser	UCU, UCC, UCA, UCG, AGU, AGC
Glu	GAA, GAG	Thr	ACU, ACC, ACA, ACG
Gly	GGU, GGC, GGA, GGG	Trp	UGG
His	CAU, CAC	Tyr	UAU, UAC
Ile	AUU, AUC, AUA	Val	GUU, GUC, GUA, GUG

codon–amino acid correspondence

PRISM: a probabilistic modelchecker

- ▶ discrete-time and continuous-time Markov chains
- ▶ reactive module-type of modelling language
`[label] guard -> rate : command`
- ▶ specifications in PCTL and CSL
- ▶ **see** <http://www.cs.bham.ac.uk/~dxp/prism/>

CSL properties

- ▶ *The program terminates successfully with probability 1.*

$$\mathcal{P}_{\geq 1}[\text{true } \mathcal{U} \text{ terminated}]$$

- ▶ *In the long-run, chances for frost are less than 30%.*

$$\mathcal{S}_{<0.3}[\text{true } \mathcal{U} \text{ temperature} = 0]$$

- ▶ *Once molecule A exceeds 10000,
will molecule B exceed 3000 within 23 to 25 hours?*

$$\mathcal{P}_{=?}[(A > 10000) \Rightarrow (\text{true } \mathcal{U}_{[23,25]} (B > 3000))]$$

modelling mRNA

```
module mRNA
```

```
s : [0..6] init 1;
```

```
cnt : [0..N] init 1;
```

```
ready : bool init false;
```

```
// Nx 3 codon mRNA CGA-GGG-AAG for protein Arg-Gly-Lys
```

```
[cga] s=1 -> ONE : s'=2; // CGA codon for Arg
```

```
[arg] s=2 -> ONE : s'=3; // Arg added to AA-chain
```

```
[ggg] s=3 -> ONE : s'=4; // GGG codon for Gly
```

```
[gly] s=4 -> ONE : s'=5; // Gly added to AA-chain
```

```
[aag] s=5 -> ONE : s'=6; // AAG codon for Lys
```

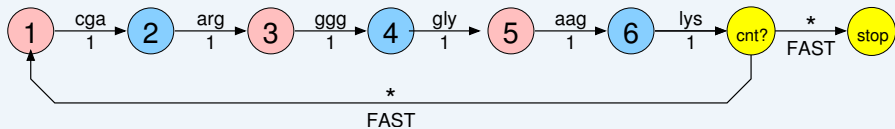
```
[lys] s=6 -> ONE : s'=0; // Lys added to AA-chain
```

```
[ ] s=0 & cnt<N -> FAST : s'=1 & cnt'=cnt+1;
```

```
[ ] s=0 & cnt=N -> FAST : ready'=true;
```

```
endmodule
```


modelling mRNA



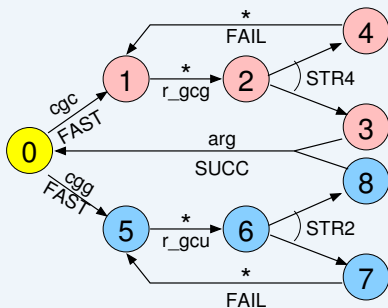
modelling iso-acceptance

```
// Arginine iso-acceptance: CGC-GCG, CGG-GCU
```

```
[cgc]  s_arg=0  -> FAST   : s_arg'=1;
[ ]    s_arg=1  -> r_gcg  : s_arg'=2;
[ ]    s_arg=2  -> STR4*FAST : (s_arg'=3) +
                (1-STR4)*FAST : (s_arg'=4);
[arg]  s_arg=3  -> SUCC   : s_arg'=0;
[ ]    s_arg=4  -> FAIL   : s_arg'=1;

[cgg]  s_arg=0  -> FAST   : s_arg'=5;
[ ]    s_arg=5  -> r_gcu  : s_arg'=6;
[ ]    s_arg=6  -> STR2*FAST : (s_arg'=7) +
                (1-STR2)*FAST : (s_arg'=8);
[arg]  s_arg=7  -> SUCC   : s_arg'=0;
[ ]    s_arg=8  -> FAIL   : s_arg'=5;
```

modelling iso-acceptance



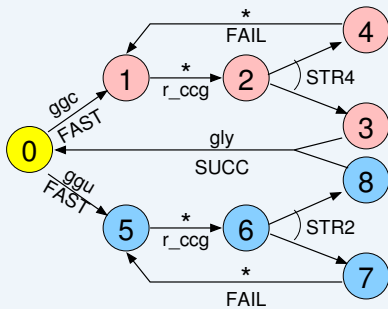
modelling wobble-acceptance

```
// Glycine wobble-acceptance: GGC-CCG and GGU-CCG
```

```
[ggc]  s_gly=0  -> FAST   : s_gly'=1;
[ ]    s_gly=1  -> r_ccg  : s_gly'=2;
[ ]    s_gly=2  -> STR4*FAST : (s_gly'=3) +
                (1-STR4)*FAST : (s_gly'=4);
[gly]  s_gly=3  -> SUCC   : s_gly'=0;
[ ]    s_gly=4  -> FAIL   : s_gly'=1;

[ggu]  s_gly=0  -> FAST   : s_gly'=5;
[ ]    s_gly=5  -> r_ccg  : s_gly'=6;
[ ]    s_gly=6  -> STR2*FAST : (s_gly'=7) +
                (1-STR2)*FAST : (s_gly'=8);
[gly]  s_gly=7  -> SUCC   : s_gly'=0;
[ ]    s_gly=8  -> FAIL   : s_gly'=5;
```

modelling wobble-acceptance



modelling mixed acceptance

```
// Glycine mixed acceptance: GGG-CCC and GGG-CCU
```

```
[ggg] s_gly=0 -> FAST : s_gly'=1;
```

```
[ ] s_gly=1 -> r_ccc : s_gly'=2;
```

```
[ ] s_gly=2 -> STR4*FAST : (s_gly'=3) +  
                (1-STR4)*FAST : (s_gly'=4);
```

```
[gly] s_gly=3 -> SUCC : s_gly'=0;
```

```
[ ] s_gly=4 -> FAIL : s_gly'=1;
```

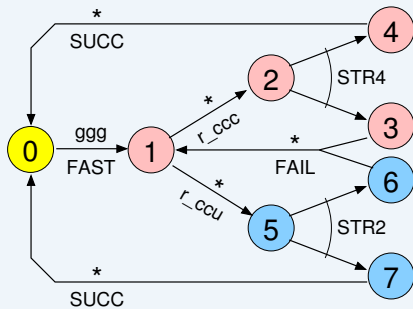
```
[ ] s_gly=1 -> r_ccu : s_gly'=5;
```

```
[ ] s_gly=5 -> STR2*FAST : (s_gly'=6) +  
                (1-STR2)*FAST : (s_gly'=7);
```

```
[gly] s_gly=6 -> SUCC : s_gly'=0;
```

```
[ ] s_gly=7 -> FAIL : s_gly'=1;
```

modelling mixed acceptance



a toy experiment

	codon	anticodon	binding
Arginine	CGC	GCG	4
	CGU	GCG	2
	CGG	GCC	4
	CGG	GCU	2
Glycine	GGC	CCG	4
	GGU	CCG	2
	GGG	CCC	4
	GGG	CCU	2

translation times for iso, wobble and mixed acceptance

iso-acceptance

	codon	anticodon	binding
Arginine	CGC	GCG	4
	CGU	GCG	2
	CGG	GCC	4
	CGG	GCU	2
Glycine	GGC	CCG	4
	GGU	CCG	2
	GGG	CCC	4
	GGG	CCU	2

mRNA: 10x CGC–GGC–CGG–GGG

wobble acceptance

	codon	anticodon	binding
Arginine	CGC	GCG	4
	CGU	GCG	2
	CGG	GCC	4
	CGG	GCU	2
Glycine	GGC	CCG	4
	GGU	CCG	2
	GGG	CCC	4
	GGG	CCU	2

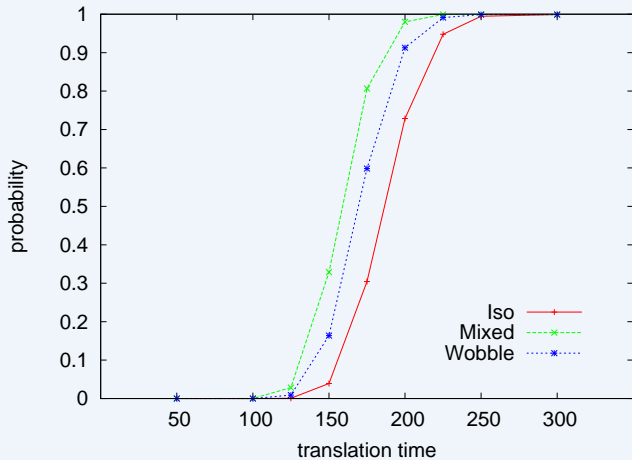
mRNA: 10x CGC–GGC–CGU–GGU

mixed-acceptance

	codon	anticodon	binding
Arginine	CGC	GCG	4
	CGU	GCG	2
	CGG	GCC	4
	CGG	GCU	2
Glycine	GGC	CCG	4
	GGU	CCG	2
	GGG	CCC	4
	GGG	CCU	2

mRNA: 20x CGG–GGG

results



$\mathcal{P}_{=?} [\text{true } U_{\leq n} \text{ ready}]$

real-life experiments

KEGG database for *Saccharomyces cerevisiae*

```
atggcgtcagtaacagaacaattcaacgatattattagcttatactcaacaaaattggaa  
cacacatctttgaggcaagattcaccagagtaccagggattattactttccacgatcaag  
aaattattaacttaaaaacagcaatttttgacaggttggcattgttcagtaactaatgag  
accattgatgatgtgtctactgcttccatcaaatttctagcagttgattactatttagga  
ttattgatatcaagacgacagtcgaatgattcggatgttgctcaaaggcagtcctatgaaa  
ttgatttacctgaaaaaaagcgttgaatctttcattaatttctctgacactattgcaggat  
tataagcttctagatcctttggttggtgaaaaactaggtaacttcaaggatcgttataac  
cctcagcttagcgaattgtacgcgcaacccaaaaataacaaaagatttatctggagcacag  
ttgaagagaaaagaaaagattgagctattccagcgcataaagaaattagcacaaaactg  
cactgcttgagttggaattaaaaaacaacgacgaggaccacgaccatgatgaattacta  
agagaactatatttgatgaggttacatcactttagtcttgatacgatataacaacattgaa  
cagaatattttgaatgtgaaatgctctctaatttctcaaaaattccgtacatgaagtc  
aatcatcaggtactcagatacgaagaagaatcgaatgatgatgatccactggttttacc  
gataaattagagaatataaataagccattgatagacaaaaaaggccaagtcttgaggaa  
ttcacgcttgcgacaaaaggcaacaactgcaacaaaaagtcgaggatgatggccaat  
ggaccaacaatgtcgggtggaggaaatttttagataaaagagtttgaagaaggtcgcgctctt  
caaggtggcgaagaaccagagcaagcaccagatgaagaaaacatggactggcaagataga  
gaaacctataaagctcgtgagtgaggacgagttcaaggaaagtcagctaaaggaagcgg  
aataccatgaatagaggatag
```

TAP42-gene: essential protein in TOR signaling pathway

tRNA profile

Ala	AGC 11	GGC	CGC	TGC 5		
Gly	ACC	GCC 16	CCC 2	TCC 3		
Pro	AGG 2	GGG	CGG	TGG 10		
Thr	AGT 11	GGT	CGT 1	TGT 4		
Val	AAC 14	GAC	CAC 2	TAC 2		
Ser	AGA 11	GGA	CGA 1	TGA 3	ACT	GCT 2
Arg	ACG 6	GCG	CCG 1	TCG	CCT 1	TCT 11
Leu	AAG	GAG 1	CAG	TAG 3	CAA 10	TAA 7
Phe	AAA	GAA 10				
Asn	ATT	GTT 10				
Lys	CTT 14	TTT 7				
Asp	ATC	GTC 16				
Glu	CTC 2	TTC 14				
His	ATG	GTG 7				
Gln	CTG 1	TTG 9				
Ile	AAT 13	GAT	TAT 2			
Met	CAT 10					
Tyr	ATA	GTA 8				
Cys	ACA	GCA 4				
Trp	CCA 6					

anticodon availability for *S. cerevisiae*

challenge

a program for codon bias experiments with Prism

- flexible, maintable & user-friendly
- available on Linux machines `poema` and `olifant`
- command-line and interactive generation
- cockpit, batch facilities, graphical output
- documentation, user manual