Fungal Genetics Reports

Volume 46 Article 5

Codon bias in the ß-lactam procucer Acremonium chrysogenum

K Jekosch
Lehrstuhl für Allgemeine und Molekulare Botanik

U Kück
Lehrstuhl für Allgemeine und Molekulare Botanik

Recommended Citation
Jekosch, K., and U. Kück (1999) "Codon bias in the ß-lactam procucer Acremonium chrysogenum," *Fungal Genetics Reports*: Vol. 46, Article 5. https://doi.org/10.4148/1941-4765.1233

Codon bias in the ß-lactam procucer Acremonium chrysogenum

Abstract

In this paper we compile the G/C-content, the codon bias, and consensus sequences for the translation initiation and the intron splicing sites from 19 nuclear genes of the major ß-lactam antibiotic producer *Acremonium chrysogenum*. Our data are compared with those from other filamentous fungi, such as *Aspergillus nidulan, Neurospora crassa,* and *Sordaria macrospora*.

Number 46, 1999

Codon bias in the ß-lactam producer *Acremonium chrysogenum*

Kerstin Jekosch and Ulrich Kück, Lehrstuhl für Allgemeine und Molekulare Botanik,
D-44780 Bochum, Germany

In this paper we compile the G/C-content, the codon bias, and consensus sequences for the translation initiation and the intron splicing sites from 19 nuclear genes of the major ß-lactam antibiotic producer *Acremonium chrysogenum*. Our data are compared with those from other filamentous fungi, such as *Aspergillus nidulans*, *Neurospora crassa*, and *Sordaria macrospora*.

The filamentous fungus *Acremonium chrysogenum* is the industrially most important producer of the ß-lactam antibiotic cephalosporin C (reviewed by Brakhage 1998, Microbiol. Mol. Biol. Rev. 62:547-585). Sequence characteristics of this imperfect fungus are of great interest as prerequisite for the improvement of antibiotic biosynthesis. We present a comprehensive sequence analysis of all nuclear genes, published so far. We provide a compilation of codon usage, consensus sequences for translation initiation sites, and regulatory sequences relevant for intron splicing.

Genes used for the present analysis are 1) cystathionine beta-synthase (Acc # E08842), 2) homoserine 0-acetyltransferase (Acc # E08840), 3) cystathionine gamma-lyase (Acc # E08276), 4) CPC biosynthesis related gamma gene (Acc # E06692), 5) CPC biosynthesis related alpha gene (Acc # E06691), 6) alkaline protease, *alp* (Acc # D00923), 7) glyceraldehyde-3-phosphate dehydrogenase (Acc # E03375), 8) actin (Acc # E03374), 9) phosphoglycerate kinase E (Acc # 03373), 10) beta isopropylmaleate dehydrogenase, *leu2* (Acc # E01906), 11) orotidine 5´-phosphate decarboxylase, *pyr4* (Acc # X15937), 12) δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase (Acc # M33522), 13) isopenicillin N synthase (Acc # S39881), 14) desacetoxycephalosporin C synthetase/hydroxylase, *pcbAB* (Gutiérrez et al. 1991, J. Bacteriol. 173:2354-2365), 15) desacetylcephalosporin C acetyltransferase, *cefG* (Acc # M91649), 16) esterase C, *estC* (Japan patent: Matsuda et al. 1992, Hei 4-144688), 17) beta tubuline, *βtub* (Acc # X72789), 18) transcriptional repressor CREA, *creA* (unpublished), 19) transcription factor CPCR1, *cpcR1* (Acc # AJ132014). No. 12 to 16 and 18 to 19 are involved in ß-lactam biosynthesis. The modes of action of genes 4 and 5 are not yet known.

The average G/C content is 61.5 ± 2.3%, which is higher than in *A. nidulans* (about 50 %, Lloyd and Sharp 1991, Mol. Gen. Genet. 230:288-294), *N. crassa* (54.1%, Edelmann and Staben 1994, Exp. Mycol. 18:70-81), and *S. macrospora* (56.7 %, Pöggeler 1997, Fungal Genet. Newsl. 44:41-44).

Similar to other fungi we found a bias for codons with a C in the third position (Table 1). While the termination codon TAA is preferred in *N. crassa* and *S. macrospora*, there is no such preference in *A. nidulans* and only little bias towards this codon in *A. chrysogenum*. The six least used codons in *A. chrysogenum* are TTA (Leu), TGT (Cys), AGA (Arg), ATA (Ile), GTA (Val) and AAA (Lys). Compared with the other three filamentous fungi, the same bias for rarely used and preferred codons is found in *A. chrysogenum*. An exception is found when the codons for serine are considered. Obviously, there is no significant preference for any one of the six possible tripletts.

The cephalosporin C-biosynthetic genes show different preferences for individual codons which correlates with their rate of expression as concluded from our own data (unpublished). The highly expressed *pcbC* and *cefEF* genes show only 15 and 17% low usage codons, while genes with a low transcriptional level, such as *pcbAB*, *cefG* and *estC*, contain 27 to 53% low usage codons. This correlation of the occurrence of rare codons and expression has already been reported for other organisms, e.g. *Saccharomyces cerevisiae*, *Escherichia coli*, *Drosophila melanogaster* (Zhang et al. 1991, Gene 105:61-67), *S. macrospora* (Pöggeler 1997) and *A. nidulans* (Lloyd and Sharp 1991, Mol. Gen. Genet. 230:288-294) and was supposed to affect translation rates.

The consensus sequence of the translation initiation context in *A. chrysogenum* does not significantly differ from *S. macrospora* and *N. crassa*. It should be noted that the open reading frame of the *pcbAB* gene starts with an GTG. The translation startpoint of the *estC* gene has not been identified yet. Following the ATG start codon, the three fungi show a preference for GCN, coding for alanine.

The average intron length in *A. chrysogenum* is 95 nt and this agrees well with the average intron length in *S. macrospora* (88 nt) but contrasts with that in *N. crassa* (63 nt). The average distance between the branch site and the 3´-splice site is 14 nt, and it is between 8 and 30 nt in length. In *S. macrospora* and *N. crassa* this distance varies from 12 to 22 nt and 14 to 30 nt, respectively. Similar the intron donor and acceptor consensus sequences show no significant deviation from those of the other fungi (Table 2). One exception is the *cefG* gene, the only ß-lactam biosynthesis gene containing introns with striking differences in the intron regulatory sequences (5´ intron donor G^TAGGTA and C^GGTGAG, 3´ intron acceptor AAAGT^, TGCTA^).

Acknowledgments: This work was supported by a grant of Hoechst Marion Roussel (Frankfurt a.M., Germany).

Table 1: Analysis of codon usage on the basis of 19 nuclear gene sequences. The codon usage values are given as whole numbers (N), as percent usage determined for each amino acid (%), and as relative synonymous codon usage (RSCU).

Codon	AS	N	%	RSCU
GCT	Ala	154	16	1.48
GCC		466	52	4.47
GCA		101	11	0.97
GCG		178	20	1.71
CGT	Arg	71	12	0.68
CGC		248	41	2.38
CGA		58	10	0.56
CGG		122	20	1.17
AGA		41	7	0.39
AGG		109	18	1.05
AAT	Asn	84	22	0.81
AAC		301	78	2.89
GAT	Asp	163	28	1.56
GAC		428	72	4.11
TGT	Cys	21	18	0.20
TGC		99	82	0.95
CAA	Gln	69	18	0.66
CAG		321	82	3.08
GAA	Glu	99	17	0.95
GAG		493	83	4.73
GGT	Gly	163	21	1.56
GGC		414	53	3.97
GGA		94	12	0.90
GGG		113	14	1.08
CAT	His	89	28	0.85
CAC		234	72	2.25
ATT	Ile	95	19	0.91
ATC		363	72	3.48
ATA		46	9	0.44
TTA	Leu	9	1	0.09
TTG		82	8	0.79
CTT		114	12	1.09
CTC	Leu	394	41	3.78
CTA		58	6	0.56
CTG		313	32	3.00
AAA	Lys	48	12	0.46
AAG		362	88	3.47
ATG	Met	187	100	1.79
TTT	Phe	81	22	0.78
TTC		280	78	2.69
CCT	Pro	103	17	0.99
CCC		252	41	2.42
CCA		74	12	0.71
CCG		186	30	1.78
TCT	Ser	82	10	0.79
TCC		209	25	2.01
TCA		86	10	0.83
TCG		189	22	1.81
AGT		75	9	0.72
AGC		210	25	2.02
ACT	Thr	80	12	0.77
ACC		283	44	2.72
ACA		94	15	0.90
ACG		190	29	1.82
TGG	Trp	124	100	1.19
TAT	Tyr	64	19	0.61
TAC		259	81	2.49
GTT	Val	101	14	0.97
GTC		380	52	3.65
GTA		48	7	0.46
GTG		202	28	1.94
TAA	Term.	8	42	2.54
TAG		5	26	1.57
TGA		6	32	1.94

Table 2: *A. chrysogenum* (*Ac*) consensus sequences compared with those of *Sordaria macrospora* (*Sm*) and *Neurospora crassa* (*Nc*). Abbreviations: org. = organism, subscribed numbers indicate the percentage of occurence.

org.	translation initiation context											
Ac		C_{44}					A_{39}	A_{39}				
	C_{67}		N	N	C_{61}	A_{61}			A_{94}	T_{100}	G_{95}	G_{53}
		C_{53}										
		A_{33}					C_{39}	C_{39}				
Sm			A_{38}				C_{50}					G_{50}
	C_{75}	C_{50}		N	C_{63}	A_{88}		A_{63}	A_{100}	T_{100}	G_{100}	
		C_{50}										
			G_{38}				A_{38}					T_{35}
Nc							A					
	C	N	N	N	C	A		A	A	T	G	G
		C										
							C					
	consensus 5´intron donor											
Ac	G_{41}	^	G_{95}	T_{91}	A_{64}	A_{68}	G_{91}	T_{64}				
Sm	G_{67}	^	G_{100}	T_{100}	A_{72}	A_{61}	G_{83}	T_{72}				
Nc	G	^	G	T	A	A	G	T				
	consensus branch site											
Ac				A_{50}								
	G_{64}	C_{100}	T_{100}		A_{95}	C_{64}	C_{41}					
					G_{45}							
Sm	A_{56}			A_{56}								
		C_{100}	T_{100}		A_{94}	C_{78}	N					
	G_{44}			G_{33}								
Nc	A			A		C						
		C	T		A		C					
	G			G		A						
	consensus 3´intron acceptor											
Ac	A_{32}	G_{36}	C_{55}									
				A_{91}	G_{91}							
	T_{32}	A_{41}	T_{27}									
Sm	G_{33}	A_{39}	C_{56}									
				A_{100}	G_{100}							
	A_{27}	T_{39}	T_{44}									
Nc	A	A	T									
				A	G							
	T	T	C									