FEBS Letters 399 (1996) 78–82 FEBS 17893

Non-random usage of 'degenerate' codons is related to protein three-dimensional structure

Alexei A. Adzhubei[1,a], Ivan A. Adzhubei[b], Igor A. Krasheninnikov[b], Stephen Neidle[a,*]

[a]*CRC Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK*
[b]*Department of Molecular Biology, Faculty of Biology, Lomonosov Moscow State University, 119899 Moscow, Russia*

Received 15 October 1996

Abstract **We report an analysis of a novel sequence-structure database of mammalian proteins incorporating nucleotide sequences of the exon regions of their genes together with protein sequence and structural information. We find that synonymous codon families (i.e. coding the same residue) have non-random codon distribution frequencies between protein secondary structure types. Their structural preferences are related to the third, 'silent' nucleotide position in a codon. We also find that some synonymous codons show very different or even opposite structural preferences at the N- or C-termini of structure fragments, relative to those observed for their amino acid residues.**

Key words: Synonymous codon usage; mRNA sequence; Protein secondary structure; Sequence-structure relationship; Cotranslational folding

1. Introduction

It has been possible for some while to produce a protein in large quantities by introducing its gene, incorporated in a proper genetic construct, into a micro-organism and then inducing its expression. However, there are instances where expression of eukaryotic proteins in bacteria produces proteins that are biologically inactive. Frequently, they form insoluble aggregates, which have to be renatured artificially in order to regain similarity in structure and biological activity with native analogues [1]. It has been shown that highly expressed genes utilize a distinct, species-specific subset of synonymous codons, which is tailored for effective mRNA translation [2,3]. This leads to species-specific codon bias which may interfere in the expression of foreign genes [4]. Consequently, the best expression is achieved when the constructed gene is optimized for a target organism codon usage. Since it has been suggested that protein folding can proceed cotranslationally [5,6], the optimization may also include the tuning of codon pattern along mRNA to a particular translation kinetics necessary to ensure the proper folding of a nascent polypeptide. In particular, codon context variations along mRNA affect translation speed and uniformity, either due to disparity in the rates of translation of different codons [7–9], or by introducing 'slow' regions of mRNA in which a ribosome has to move over mRNA local secondary structure elements [10]. There are indications that synonymous codon usage might be biased towards rare codons in segments connecting domains and regular secondary structure blocks [9,11,12], accompanied by lower rates of translation of inter-domain regions [11,13]. An opposite view is presented in [14], the authors find no correspondence between the distribution of rare codons and positions of structural blocks in proteins. The analysis performed in this work examines possible correlations between synonymous codon usage in amino acid residues, and protein secondary structure.

2. Materials and methods

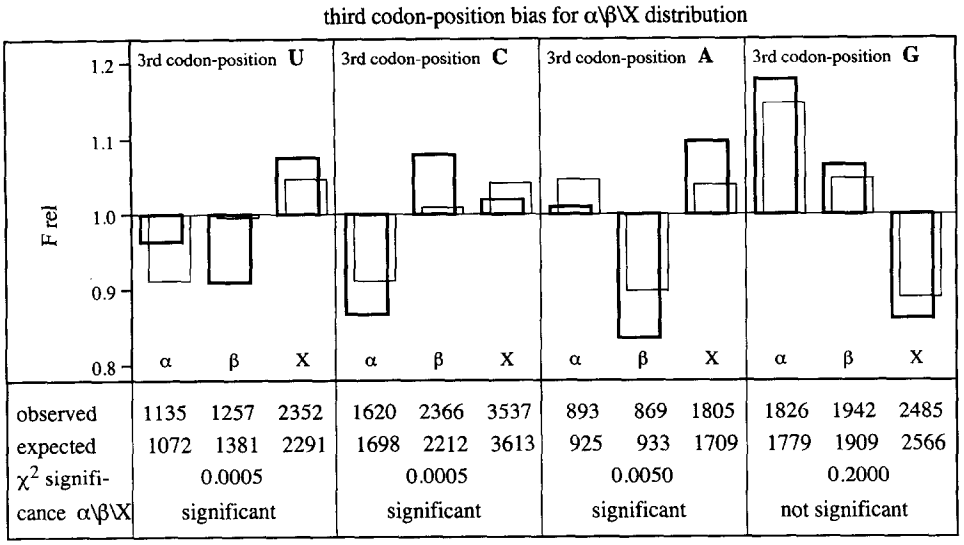
The sequence-structure database used in this analysis incorporates exactly matching codon sequences – amino acid sequences – secondary structure – internal geometric parameters of peptide groups (ϕ,ψ angles) – PDB coordinates for 109 mammalian proteins. The non-redundant dataset of the three-dimensional structures of mammalian proteins was compiled using the PDB Browser facility and a locally mirrored Brookhaven PDB database (Table 1). Amino acid sequences were extracted from PDB records and checked for homology against the rest of the database by the program FASTA. Proteins that had sequence identity $> 50\%$ in any of the pairwise alignments with other sequences in the database were removed. Protein sequences from the PDB files were scanned for homology against the NCBI 'nr' (formerly GenPept, the translated version of GenBank) database using BLAST at NCBI via e-mail server, yielding amino acid sequences with the highest similarity scores (95–100%) in the 'nr' database. The corresponding nucleotide sequences were then extracted from GenBank and the best matching sequence identified according to locally performed TFASTA alignments and additional selection criteria. Each protein sequence in the sequence-structure database was aligned with the corresponding coding nucleotide sequence and substituted with the sequence of codons. Cluster analysis of the relative synonymous codon usage distances between sequences [15] was used to monitor the uniformity of the database in terms of synonymous codon usage. The final database displays a statistically significant GC bias which corresponds to the data of the Codon Usage Database [16] (Table 2). Protein secondary structure classes in the three-state model were assigned by the program DSSP [17]. The class α (α-helix) corresponds to the DSSP structure type H, class β (β-structure) includes types E and B, class X incorporates all other DSSP structure types including unstructured residues. N- and C-termini are the additional structure classes enclosing amino acid residues in potentially important positions, respectively, in the 'start' and 'end' regions of regular secondary structure blocks. They were set to include six consecutive residues for both 'start' and 'end' of each α-helix (four inside+two outside a structure block) and five consecutive residues for each β-strand (three inside+two outside a structure block). Protein structures included in the database have resolution better or equal to 2.5 Å (Table 1), with only three exceptions: amino acid sequence identity is below 30% for 99.5% of the total non-redundant set of pairwise alignments. The final codon sequence dataset has lower level of sequence similarity, with no pairwise alignments above 30% sequence identity. Codon occurrence statistics for secondary structures were subsequently calculated according to Adzhubei and Sternberg [18].

Observed frequencies $f_{obs}(cdn,ssi)$ represent the number of occurrences of a codon *cdn* in a synonymous codon family, for a secondary structure classification *ssi*. For a three-state model used in this analysis $ssi = \alpha, \beta, X$. Expected frequencies $f_{exp}(cdn,ssi)$ were calculated for

*Corresponding author. Fax: (44) (181) 643-1675.
E-mail steve@iris5.icr.ac.uk

[1] *A.A.A. and I.A.A. are joint first authors.*

PII S0014-5793(96)01287-2



	3rd codon-position U			3rd codon-position C			3rd codon-position A			3rd codon-position G		
observed	1135	1257	2352	1620	2366	3537	893	869	1805	1826	1942	2485
expected	1072	1381	2291	1698	2212	3613	925	933	1709	1779	1909	2566
χ^2 significance α\β\X	0.0005 significant			0.0005 significant			0.0050 significant			0.2000 not significant		

Fig. 1. Observed and expected distributions of nucleotides in the third (silent) codon position. Observed frequencies (shown in bold lines) show the combined contribution of amino acid usage and synonymous codon bias in different secondary structure types. Expected frequencies (shown in hairlines) give the hypothetical pattern based on a random occurrence of synonymous codons in secondary structures, but reflect amino acid preferences in secondary structures. F_{rel} is the relative occurrence frequency of codons with a given nucleotide in the third position in a particular structure type. $F_{rel} = 1.0$ indicates no preference, with values above 1.0 corresponding to overrepresented and below 1.0 to underrepresented bases. For all practical purposes F_{rel} can be considered identical to the Chou-Fasman [34] secondary structure propensities. Structure types in a three-state model are denoted as α, α-helix (25%); β, β-structure (29%); X, all other residues (46%). χ^2 significance levels for the α\β\X distribution were calculated from the given observed and expected frequencies.

$h \times k$ tables where h is the number of synonymous codons in a family and k represents the number of structural classifications. $f_{exp}(cdn,ssi) = f(fam,ssi) \times P(fam,cdn)$, where $f(fam,ssi)$ is the occurrence frequency of a secondary structure type *ssi* in a given amino acid codon family *fam* and $P(fam,cdn)$ is the probability of a codon *cdn* determined by the existing synonymous codon bias in this family. Expected frequencies thus correspond to the null hypothesis of a random synonymous codon usage in structure classifications for the encoded amino acid, corrected for different occurrence frequencies of a given amino acid in α\β\X and for synonymous codon bias in the amino acid family. Synonymous codon observed and expected frequencies for some amino acids are given in Fig. 2. χ^2 were subsequently computed according to the formula $\chi^2 = \Sigma(f_{obs}-f_{exp})^2/f_{exp}$ [19], where the sum is taken over all cells in a $h \times k$ table. χ^2 and significance levels for the rejection of the null hypothesis were calculated for all codon families that met the criterion of $f_{exp}(cdn,ssi) \geq 10.0$ for all $f_{exp}(cdn,ssi)$. Frequencies for U, C, A and G in the third codon-position represent sums over all codons with the identical third base. $F_{rel} = f(obs, exp; ssi)/f(exp; no\ ssi\ bias)$ where $f(obs, exp; ssi)$ is an observed or expected frequency as defined above and $f(exp; no\ ssi\ bias)$ is the expected frequency of a codon or a third base calculated as shown above but assuming the random occurrence (no preference) of amino acids in α\β\X structure classes. For the latter case $f(fam,ssi)$ is calculated as $f(fam) \times P(ssi)$, where $f(fam)$ is the occurrence frequency of residues in the amino acid family *fam* and $P(ssi)$ is the probability of a secondary structure classification *ssi* in the database.

3. Results and discussion

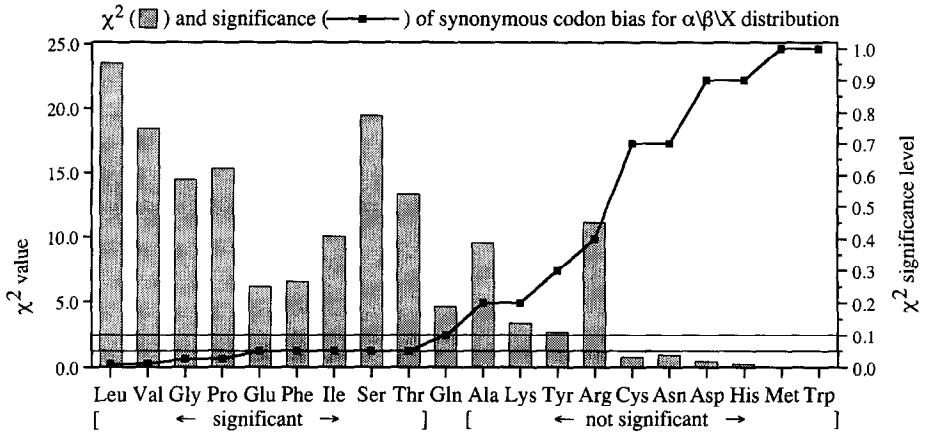
The first phase of the project involved construction of the integrated sequence-structure database. The database contained secondary structure assignments for the experimentally determined three-dimensional structures of 109 non-homologous mammalian proteins, that have been aligned with the full nucleotide sequences of the coding regions of their genes (or mRNAs). This database was used to calculate the actual codon occurrence frequencies in the secondary structure of each protein, in accordance with a three-state model incorporating α-helices (α), β-sheets (β) and other (X) structure classes. A χ^2

Table 1
The non-redundant set of protein structures used in the compilation of the integrated sequence-structure database

PDB entries (chain identifiers are given in parentheses)				Resolution if below 2.5 Å
1aap(A)	1abm(A)	1ads	1ald	
1ang	1ant(I)	1bet	1bpb	
1cbs	1cd8	1cka(A)	1cks(B)	
1cll	1crb	1csk(A)	1dfn(A)	
1dlh(AB)	1drf	1dyn(A)	1esl	
1fna	1frp(A)	1fru(A)	1ggt(A)	
1glq(A)	1gmf(A)	1guh(A)	1hcg	
1hcn(AB)*	1hcq(A)	1hdr	1hdx(A)	*2.6 Å
1hfc	1hlc(A)	1hle	1hml	
1hmp(A)	1hnf	1hsa(A)	1hsb(B)	
1hti(A)	1htr(BP)	1hul(A)	1hup	
1ice(AB)	1ifc	1ilk	1ilr(1)	
1lcf	1lki	1lpb(B)	1lpe	
1lya(AB)	1lz1	1mld(A)	1mup	
1nsk(R)	1pk4	1pod	1ppb(HL)	
1ppf(E)	1psn	1rbp	1rcb	
1rfb(A)*	1rhp(A)	1rto(A)	1rtp(1)	*3.0 Å
1sac(A)	1shf(A)	1spd(A)	1ten	
1tld	1tnf(A)	1tnr(AR)	1ton	
1tpk(A)	1tta(A)	1ubq	1ula	
1vca(A)	1zaa(C)	2ach(AB)	2ada	
2bb2	2cba	2cpl	2ctb	
2fke	2gst(A)	2hhb(AB)	2hhm(A)	
2hmb	2hnp	2ldx	2pfl	
2ran	2tgi	3cd4	3est	
3grs	3il8	3mdd(A)	4fgf	
4gcr	4ilb	5pti	6rlx(AB)	
7api(AB)*				*3.0 Å

Source:	Total:
human (*Homo sapiens*); bovine (*Bos taurus*); horse (*Equus caballus*); mouse (*Mus musculus*); porcine (*Sus scrofa*); rat (*Rattus rattus* and *Rattus norvegicus*)	proteins 109 residues 22157 codons[a] 22087

[a]No codon assignments were made for 70 individual sequence positions where amino acid or nucleotide sequence information was incomplete.



synonymous codon families			structure classification					
			α		β		X	
amino acid	codons	f obs	f obs	f exp	f obs	f exp	f obs	f exp
Leu	CUU	204.0	71.0	67.1	56.0	71.4	77.0	65.5
	CUC	400.0	112.0	131.5	164.0	140.1	124.0	128.4
	CUA	115.0	43.0	37.8	48.0	40.3	24.0	36.9
	CUG	863.0	298.0	283.8	295.0	302.2	270.0	277.1
	UUA	98.0	34.0	32.2	31.0	34.3	33.0	31.5
	UUG	245.0	75.0	80.6	80.0	85.8	90.0	78.7
family		1925.0						
Val	GUU	264.0	75.0	62.1	105.0	127.4	84.0	74.5
	GUC	400.0	83.0	94.0	204.0	193.0	113.0	113.0
	GUA	129.0	21.0	30.0	61.0	62.2	47.0	36.4
	GUG	751.0	184.0	176.6	375.0	362.4	192.0	212.1
family		1544.0						
Gly	GGU	265.0	42.0	31.0	37.0	50.2	186.0	183.8
	GGC	621.0	75.0	72.7	133.0	117.6	413.0	430.7
	GGA	391.0	38.0	45.8	66.0	74.0	287.0	271.2
	GGG	371.0	38.0	43.4	76.0	70.2	257.0	257.3
family		1648.0						
Pro	CCU	280.0	35.0	29.0	41.0	39.0	204.0	212.0
	CCC	370.0	28.0	38.3	66.0	51.6	276.0	280.1
	CCA	253.0	30.0	26.2	22.0	35.3	201.0	191.5
	CCG	101.0	11.0	10.5	11.0	14.1	79.0	76.5
family		1004.0						
Glu	GAA	596.0	196.0	198.4	126.0	143.3	274.0	254.3
	GAG	864.0	290.0	287.6	225.0	207.7	349.0	368.7
family		1460.0						
Phe	UUU	395.0	100.0	97.5	152.0	169.9	143.0	127.6
	UUC	565.0	137.0	139.5	261.0	243.1	167.0	182.4
family		960.0						
Ile	AUU	378.0	103.0	104.8	170.0	174.0	105.0	99.2
	AUC	621.0	168.0	172.2	305.0	285.9	148.0	162.9
	AUA	137.0	44.0	38.0	48.0	63.1	45.0	35.9
family		1136.0						
Ser	UCU	253.0	51.0	46.2	52.0	61.6	150.0	145.2
	UCC	374.0	72.0	68.3	110.0	91.1	192.0	214.6
	UCA	156.0	21.0	28.5	39.0	38.0	96.0	89.5
	UCG	72.0	12.0	13.1	25.0	17.5	35.0	41.3
	AGU	213.0	42.0	38.9	41.0	51.9	130.0	122.2
	AGC	394.0	69.0	72.0	89.0	95.9	236.0	226.1
family		1462.0						
Thr	ACU	301.0	72.0	60.2	80.0	103.1	149.0	137.8
	ACC	586.0	106.0	117.1	223.0	200.6	257.0	268.3
	ACA	320.0	64.0	64.0	106.0	109.6	150.0	146.5
	ACG	119.0	23.0	23.8	45.0	40.7	51.0	54.5
family		1326.0						

Fig. 2. Significance of deviation from random usage in secondary structures for synonymous codons in amino acid families, and their observed and expected frequencies. χ^2 values and significance levels were calculated separately for each family. For nine amino acids the null hypothesis of the random synonymous codon distribution between α\β\X classifications was rejected at the significance level ≤0.050. Observed (f_{obs}) and expected (f_{exp}) frequencies of synonymous codons are listed for these families.

approximation [19] was used to evaluate the statistical significance of synonymous codon bias in the database, for the distribution between the three structure classes.

Synonymous codons in an amino acid family differ in the third nucleotide position, which is termed 'silent' since nucleotide substitutions in it do not affect the type of amino acid encoded. Expected frequencies were calculated taking into account the known synonymous codon bias in families and different amino acid frequencies in the secondary structure classes, but assuming the random synonymous codon usage in structure types for each family. Fig. 1 shows the distribution of codons relative to the four possible bases in the third

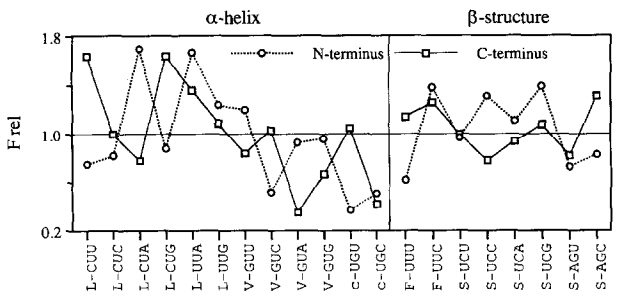


Fig. 3. Relative observed frequencies (F_{rel}) for synonymous codons in families with significant deviations from random usage for the distribution between N-termini, core and C-termini of the regular secondary structures. χ^2 significance levels are 0.010 for Leu (L), 0.025 for Val (V), 0.050 for Css (c, cystine residues, S-S bonds formed), 0.001 for Phe (F), 0.025 for Ser (S).

codon position. Expected frequencies reflect the pattern which would have arisen if, assuming random synonymous codon usage, the contribution of existing amino acid structural preferences was solely responsible for the third-base bias in the secondary structures. Deviations from random occurrence in the secondary structure classes, for synonymous codons in the amino acid families are shown in Fig. 2. This figure also lists observed and expected frequencies for some codon families, as used in the χ^2 calculations.

All three secondary structure classes show distinct patterns of the third base usage. Notably, the origin of the bias displayed by G differs from that observed for the other bases in the third codon position. Since amino acids residues are coded by sets of codons that can differ in the number of synonyms each ending with a specific base (examples are given in Fig. 2), the non-equal occurrence of residues in secondary structures gives rise to the expected third base bias which corresponds to random synonymous codon usage (Fig. 1). The overall distributions of codons that have U, C and A in their third position are non-random between the α, β and X structure classes. They significantly alter the expected bias in secondary structures for these bases, resulting in the observed third base bias. In contrast with this, the bias for G in the third codon position is solely based on the difference in amino acid frequencies between secondary structures, since the additional synonymous codon bias is not significant. In α-helices the synonymous codon contribution to U and A third position bias pushes the observed frequencies closer to F_{rel}=1.0 (no bias), while it enhances the high negative C bias. A contrasting pattern is observed for the β-structure where both negative third position U and positive C bias result solely from the synonymous codon contribution, which also markedly increases the already high expected negative bias for A in the third codon position.

The overall distribution of the full set of 61 sense codons in α\β\X classes deviates highly significantly from the expected random occurrence for amino acid codon families at the 0.0005 significance level (estimated from the total χ^2 value). At the level of individual codon families (Fig. 2), 9 families (comprising 35 codons) show significant non-random occurrences of synonymous codons in the α\β\X distribution. Thus Fig. 2 shows the existence of statistically significant secondary structure preferences for synonymous codons in individual codon families, in addition to the third codon position bias related to secondary structure types (Fig. 1). For most of the codons the deviations from random usage are in opposite directions in α-helices and β-sheets, i.e. codons overrepresented in α-helices are underrepresented in β-sheets and vice versa.

The frequency of occurrence of some synonymous codons depends on their position relative to secondary structure fragment boundaries. Although our database is not large enough for a full statistical analysis of these frequency deviations, the amino acid families which display the most pronounced statistically significant synonymous codon bias in terms of the distribution between N-terminal, C-terminal and core regions of a particular secondary structure (X structure class is also included) can be identified (Fig. 3). It is known that a number of residues occur more frequently at secondary structure termini [20–22]. We now see that there are actually quite distinct relative frequencies of codons within particular families. Some are position-sensitive as for example Leu and Phe, whereas other families do not show specific preferences for secondary structure termini.

The significant preference for different synonymous codons in secondary structure classes suggests the existence of a link between nucleic acid coding sequence and protein three-dimensional structure. It is unlikely to result from evolution pressure operating solely on the sequence level since the bias in secondary structures is observed in addition to any existing synonymous codon usage bias in families. Our results are in agreement with the increasingly supported view of protein folding as a process occurring cotranslationally [5,6,9,23–27]. According to this hypothesis a newly synthesized polypeptide chain starts to fold into the elements of secondary and higher level structure before the ribosome finishes translation of its mRNA. The protein molecule thus can be folded in its final native conformation (and display activity) before or immediately after the release of the nascent polypeptide chain from the ribosome [25,27]. If protein folding proceeds cotranslationally, the codon context may affect protein structure. The hypothesis is difficult to prove experimentally and only a few recent studies address this problem [28]. However, there is indirect experimental evidence supporting this hypothesis. The majority of genes expressed in the most popular *E. coli* or *S. cerevisiae* expression systems have originated from higher eukaryotes, where genes are characterized by a distinct pattern of codon usage; their pattern is quite different from both *E. coli* and *S. cerevisiae* codon usage [15,16,29,30]. One possible explanation of the appearance of inactive proteins after expression in these systems is that the process of protein overexpression itself causes such a severe alteration of the micro-organism biosynthetic machinery that it leads to inacti-

Table 2
Nucleotide frequencies in the third codon position

Third nucleotide	f_{obs}	f_{exp}	% obs	% exp
U	4744.0	5595.6	21.5	25.3
C	7523.0	5595.6	34.1	25.3
A	3567.0	5251.7	16.1	23.8
G	6253.0	5644.1	28.3	25.6
G+C	13776.0	11239.7	62.4	50.9
A+U	8311.0	10847.3	37.6	49.1

f_{obs}, observed frequencies in the database; f_{exp}, expected frequencies calculated on the basis of random synonymous codon usage in amino acid families (no synonymous codon bias). The χ^2 calculations show that the bias is statistically significant at the 0.0005 significance level.

vation and aggregation of the produced protein. However, it has been shown for several proteins that when their genes were synthesized de novo or mutated in such a way that their codons were replaced by synonyms preferably used in the target micro-organism (retaining exactly the same sequence of amino acids), the yield of the active product was substantially increased [31,32], up to 100-fold and more [33]. Our results suggest that in order to achieve a high level of expression of active protein, codon secondary structure preferences must be taken into account in addition to the species-specific synonymous codon usage optimization.

Acknowledgements: We thank Dr. Ross D. King, Dr. Anton A. Komar and Dr. Evgeniy N. Kuznetsov for valuable discussions, and Dr. Natalya Chernyaeva for indispensable help. This work was supported by the Cancer Research Campaign and in part by the Russian Fund for Basic Research. I.A.A. received a fellowship from The Royal Society under the 'Exchanges with the FSU' scheme. The full data and results of statistical analysis will be available via the CRC Biomolecular Structure Unit WWW Home Page on http://www.bms-unit.icr.ac.uk/.

References

[1] Nilsson, B. and Anderson, S. (1991) Annu. Rev. Microbiol. 45, 607–635.
[2] Ikemura, T. (1985) Mol. Biol. Evol. 2, 13–34.
[3] Andersson, S.G.E. and Kurland, C.G. (1990) Microbiol. Rev. 54, 198–210.
[4] Kane, J.F. (1995) Curr. Opin. Biotechnol. 6, 494–500.
[5] Hardesty, B., Kudlicki, W., Odom, O.W., Zhang, T., McCarthy, D. and Kramer, G. (1995) Biochem. Cell Biol. 73, 1199–1207.
[6] Kolb, V.A., Makeyev, E.V., Kommer, A. and Spirin, A.S. (1995) Biochem. Cell Biol. 73, 1217–1220.
[7] Bonekamp, F., Andersen, H.D., Christensen, T. and Jensen, K.F. (1985) Nucleic Acids Res. 13, 4113–4123.
[8] Wolin, S.L. and Walter, P. (1988) EMBO J. 7, 3559–3569.
[9] Krasheninnikov, I.A., Komar, A.A. and Adzhubei, I.A. (1991) J. Protein Chem. 10, 445–454.
[10] Chaney, W.G. and Morris, A.J. (1979) Arch. Biochem. Biophys. 194, 283–291.
[11] Krasheninnikov, I.A., Komar, A.A. and Adzhubei, I.A. (1989) Biokhimia (Moscow) 54, 187–200.
[12] Krasheninnikov, I.A., Komar, A.A. and Adzhubei, I.A. (1989) Dokl. Akad. Nauk. SSSR 305, 1006–1012.
[13] Thanaraj, T.A. and Argos, P. (1996) Prot. Sci. 5, 1594–1612.
[14] Brunak, S. and Engelbrecht, J. (1996) Proteins 25, 237–252.
[15] Sharp, P.M., Tuohy, T.M. and Mosurski, K.R. (1986) Nucleic Acids Res. 14, 5125–5143.
[16] Nakamura, Y., Wada, K., Wada, Y., Doi, H., Kanaya, S., Gojobori, T. and Ikemura, T. (1996) Nucleic Acids Res. 24, 214–215.
[17] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[18] Adzhubei, A.A. and Sternberg, M.J.E. (1993) J. Mol. Biol. 229, 472–493.
[19] Dowdy, S. and Wearden, S. (1991) Statistics for Research, John Wiley and Sons, New York.
[20] Argos, P. and Palau, A. (1982) Int. J. Peptide Protein Res. 19, 380–393.
[21] Richardson, J.S. and Richardson, D.C. (1988) Science 240, 1648–1652.
[22] MacArthur, M.W. and Thornton, J.M. (1991) J. Mol. Biol. 218, 397–412.
[23] Crombie, T., Swaffield, J.C. and Brown, A.J.P. (1992) J. Mol. Biol. 228, 7–12.
[24] Komar, A.A., Kommer, A., Krasheninnikov, I.A. and Spirin, A.S. (1993) FEBS Lett. 326, 261–263.
[25] Makeyev, E.V., Kolb, V.A. and Spirin, A.S. (1996) FEBS Lett. 378, 166–170.
[26] Chen, W., Helenius, J., Braakman, I. and Helenius, A. (1995) Proc. Natl. Acad. Sci. USA 92, 6229–6233.
[27] Kudlicki, W., Kitaoka, Y., Odom, O.W., Kramer, G. and Hardesty, B. (1995) J. Mol. Biol. 252, 203–212.
[28] Komar, A.A. and Jaenicke, R. (1995) FEBS Lett. 376, 195–198.
[29] Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori, T. and Ikemura, T. (1990) Nucleic Acids Res. 18 (Suppl.), 2367–2411.
[30] Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, r43–r74.
[31] Williams, D.P., Regier, D., Akiyoshi, D., Genbauffe, F. and Murphy, J.R. (1988) Nucleic Acids Res. 16, 10453–10467.
[32] Martin, S.L., Vrhovski, B. and Weiss, A.S. (1995) Gene 154, 159–166.
[33] Mohsen, A.-W.A. and Vockley, J. (1995) Gene 160, 263–267.
[34] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.