Appendix A

Data related to Experiments

A.1 Codon Usage Bias in bacterial genomes Supplementary Data

A.1.1 Importance of the codons estimated in randomized gene sets of *E. coli* genome

Figure A-1 presents the importance of the codons estimated in randomized gene sets of *E. coli* genome in terms of Z-scores using the Boruta algorithm [101]. The algorithm was run 1000 times, and the distribution of Z-scores for the codons was presented in the form of box plots. Codons contributing towards classifying HeG and LeG are shown with gray boxes, codons having negligible contribution are shown with white boxes. The distributions of Z-score for shadow boxes are shown in black.

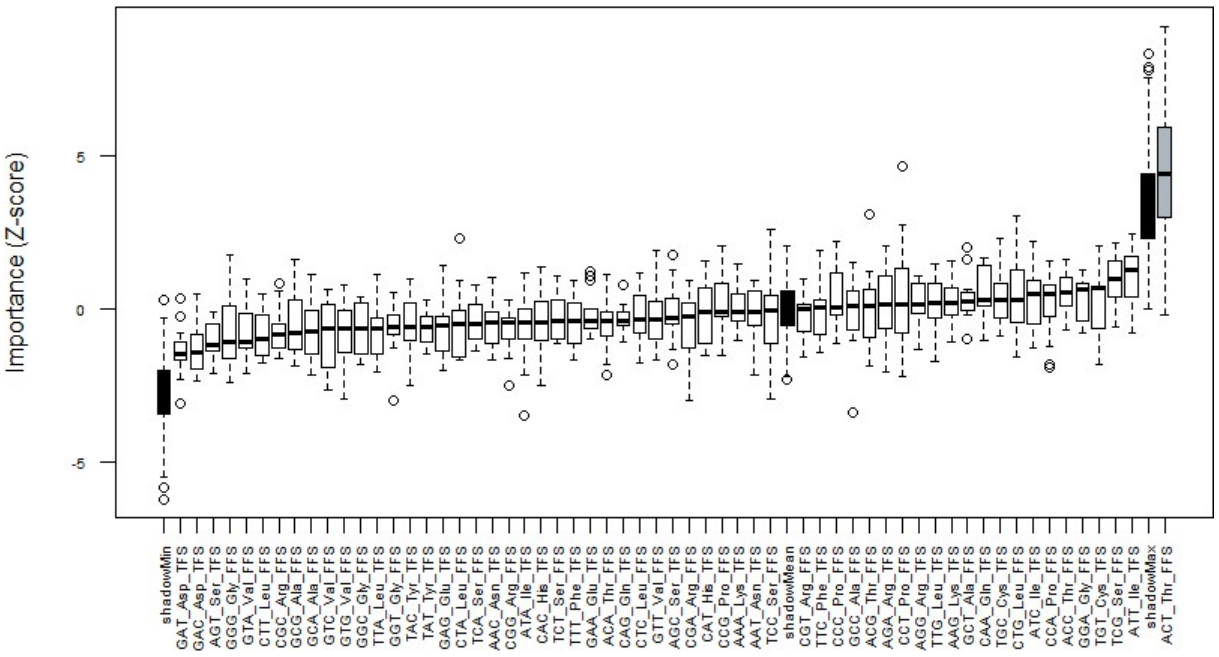

Figure A-1: Importance of the codons estimated in randomized gene sets of *E. coli* genome

A.1.2 List of high and low expression genes considered in this study

Github link for A.1.2: `https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_2/Chapter%202_Appendix%20A.1.2.pdf`

A.1.3 Details of 683 bacteria species

Github link for A.1.3: `https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_2/Chapter%202_Appendix%20A.1.3.pdf`

A.1.4 Important codons (marked as 'C') of 683 bacteria species

Github link for A.1.4: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_2/Chapter%202_Appendix%20A.1.4.xlsx

A.2 Single nucleotide polymorphism in bacterial genomes Supplementary Data

A.2.1 Finding single nucleotide polymorphism from the sequence alignments

The detailed procedure we followed can be explained with an example as shown below in Figure A-2. Let us consider the alignment of 10 hypothetical sequences containing 21 nucleotides each. Before finding mutations, we generated a reference sequence. At each position in the sequence alignment, occurrences of the four nucleotides are counted, and the nucleotide with the maximum occurrence is considered to be the nucleotide in that position in the reference sequence. Now, if there is (are) any other nucleotide (s) at a particular position with a non-zero count, then we considered that a mutation. For example, at position 1, there are two mutations G $\rightarrow$ C and G $\rightarrow$ A. We wrote a script in Python to find out these mutations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
>SA1	G	A	T	C	C	T	G	T	A	T	G	C	C	A	C	T	C	G	A	**T**	**C**
>SA2	**A**	A	T	C	C	T	G	T	A	T	G	C	C	A	C	T	C	G	A	**T**	**A**
>SA3	G	A	T	C	C	**C**	G	T	A	T	G	C	C	A	C	T	C	G	A	T	T
>SA4	**A**	A	T	**T**	C	T	G	T	A	T	G	C	C	A	C	T	C	G	A	T	T
>SA5	G	A	T	C	C	T	**A**	T	A	T	G	C	C	A	C	T	C	**A**	A	T	T
>SA6	**C**	A	T	C	C	T	G	T	A	T	G	C	C	A	**T**	T	C	G	A	T	T
>SA7	G	A	T	C	C	T	G	T	A	T	**A**	C	C	A	C	T	C	G	A	**T**	**A**
>SA8	G	A	T	C	C	T	G	T	A	T	G	C	C	A	C	T	C	G	A	T	T
>SA9	G	A	T	C	C	T	G	T	A	T	G	C	C	**G**	C	T	C	G	A	T	T
>SA10	G	A	T	C	C	**C**	G	T	A	T	G	C	C	A	C	T	C	G	A	T	T
>Ref	**G**	**A**	**T**	**C**	**C**	**T**	**G**	**T**	**A**	**T**	**G**	**C**	**C**	**A**	**C**	**T**	**C**	**G**	**A**	**T**	**T**
$Count_A$	**2**	**10**	**0**	**0**	**0**	**0**	**1**	**0**	**10**	**0**	**1**	**0**	**0**	**9**	**0**	**0**	**0**	**1**	**10**	**0**	**2**
$Count_T$	**0**	**0**	**10**	**1**	**0**	**8**	**0**	**10**	**0**	**10**	**0**	**0**	**0**	**0**	**1**	**10**	**0**	**0**	**0**	**10**	**7**
$Count_G$	**7**	**0**	**0**	**0**	**0**	**0**	**9**	**0**	**0**	**0**	**9**	**0**	**0**	**1**	**0**	**0**	**0**	**9**	**0**	**0**	**0**
$Count_C$	**1**	**0**	**0**	**9**	**10**	**2**	**0**	**0**	**0**	**0**	**0**	**10**	**10**	**0**	**9**	**0**	**10**	**0**	**0**	**0**	**1**
Mutation (s)	**G→ C, G→ A**			**C→ T**		**T→ C**	**G→ A**				**G→ A**		**C→ T**	**A→ G**	**C→ T**			**G→ A**			**T→ A, T→ C**

Figure A-2: Determining single nucleotide polymorphism from the sequence alignments

A.2.2 GC and AT skews in chromosomes

Github link for A.2.2: `https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.2.docx`

A.2.3 Phylogeny of the five bacteria *Ec*, *Kp*, *Se*, *Sa* and *Sp* constructed using *rpoB* and *rpoC* gene sequence

Github link for A.2.3: `https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.3.docx`

A.2.4 Polymorphism frequency distribution among strains of the bacteria using rpoB and rpoC gene sequences

Github link for A.2.4: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.4.docx

A.2.5 (G+C)% in whole genome sequence (WGS), at IRs and FFS of the five amino acids in genome, as well as in the high expression genes (HEGs) of LeS and LaS

Github link for A.2.5: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.5.docx

A.2.6 Polymorphism spectra at FFS of five amino acids in the leading and the lagging strands of five bacteria

Github link for A.2.6: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.6.docx

A.2.7 Comparison of polymorphism frequency at WGS, HEGs and IRs

Github link for A.2.7: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.7.docx

A.2.8 Correlation of genome size with genome G+C% in bacterial groups

Github link for A.2.8: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.8.docx

A.2.9 Intra-species polymorphism in *rpoB* and *rpoC* genes and inter-species substitution frequency in *rpoB* and *rpoC* comparison between *Ec*, *Kp*, and *Se* as well as between *Sa* and *Sp*

Github link for A.2.9: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.9.docx

A.2.10 Nucleotide composition at FFS of different amino acids in the high expression genes (HEGs) of five bacteria

Github link for A.2.10: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_3/Chapter%203_Appendix%20A.2.10.docx

A.3 RNA Secondary Structure Estimation Supplementary Data

A.3.1 Details of 25 Pseudoknotted Structure

Github link for A.3.1: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_4/Chapter%204_Appendix%20A.3.1.xlsx

A.3.2 Gene wise performance metrices

Github link for A.3.2: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_4/Chapter%204_Appendix%20A.3.2.docx

A.3.3 Count of RNAs, mean, standard deviation, minimum value, 25% or Q1, 50% or Q2, 75% or Q3, maximum value, Interquartile range (IQR) of Accuracy, Precision, Recall, Specificity and F1 Score of the 25 pseudoknotted structures

Github link for A.3.3: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_4/Chapter%204_Appendix%20A.3.3.docx

A.4 Base substitutions in tRNA genes Supplementary Data

A.4.1 Substitutions in predicted tRNA Secondary structure

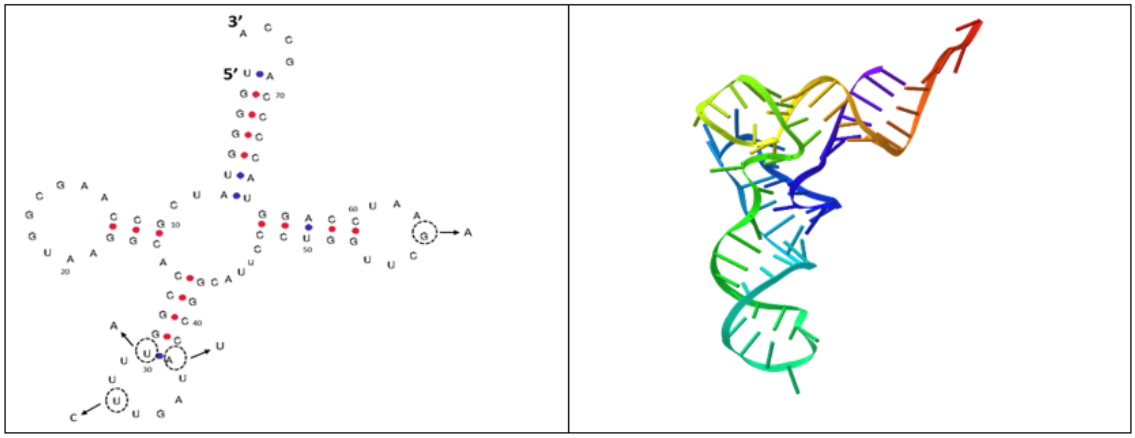

Figure A-3: Substitutions in predicted tRNA Secondary structure of *Ec* Gln tRNA with UUG anti-codon

Figure in the left panel represents secondary structure (2D) (clover leaf model) of *Ec* Gln tRNA with UUG anti-codon and Figure in the right panel represents the three-dimensional (3D) (L-shaped) view of the same tRNA. The 2D structure is obtained using tRNAscan-SE On-line software (Lowe and Chan 2016). The substitutions U→A, U→C, A→U and G→A at 30th, 33rd, 38th and 56th positions, respectively are also depicted in the figure. The 3D view of the tRNA is obtained with the help of two web servers Vfold3D[242][251] and iCn3D[230] as described in Materials and Methods. We have used V-R spectrum for the tRNA colouring where the violet and red are the 5'and 3' end respectively. The blue and orange portion are the D-loop and T-loop respectively. The green portion is the anticodon loop.

A.4.2 Base substitution frequency in non-compensatory and compensatory stem

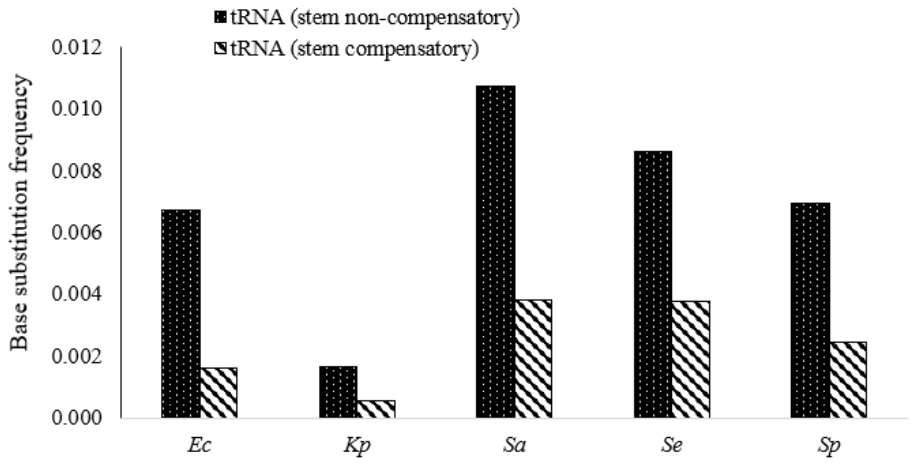

Figure A-4: Base substitution frequency in non-compensatory and compensatory stem

Histogram presenting the compensatory and non-compensatory base substitution frequency values in stem regions in tRNA genes. The values are significantly different (p-value $< .05$). The x-axis presents the five bacteria *Escherichia coli* (*Ec*), *Klebsiella pneumoniae* (*Kp*), *Salmonella enterica* (*Se*), *Staphylococcus aureus* (*Sa*) and *Streptococcus pneumoniae* (*Sp*).

A.4.3 G:T mispairing and A:C mispairing energy calculation study

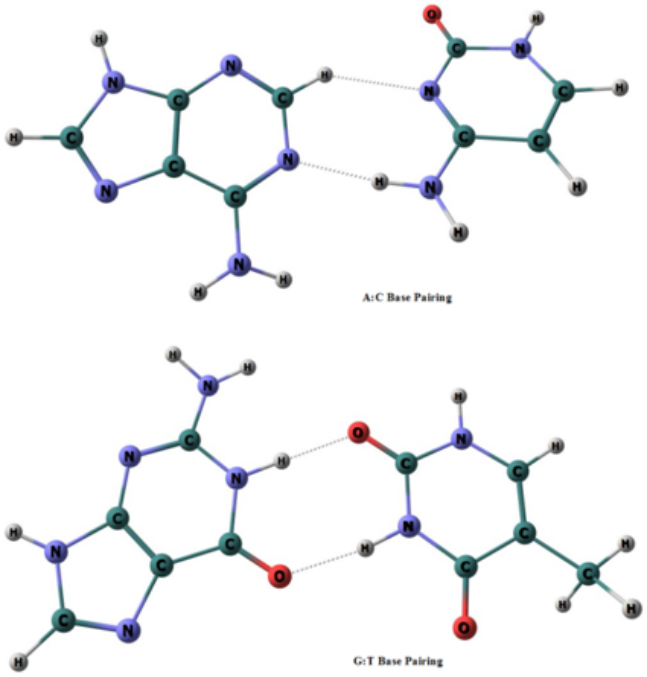

Figure A-5: G:T mispairing and A:C mispairing energy calculation study G:T mispair and A:C mispair, study was done by DFT calculations using M06-2X/6-311++G(d,p) level of theory as implemented in the software GAUSSIAN 09. The calculated interaction energy for G:T pair is -24.51 kJ/mol whereas that for A:C pair is 2.37 kJ/mol. The positive interaction energy shows unfavorable interaction between A and C while favorable for the interaction between G and T.

A.4.4 Amino(A/C) → Keto (G/T), Keto (G/T) → Amino (A/C) polymorphism frequencies IRs of five bacteria

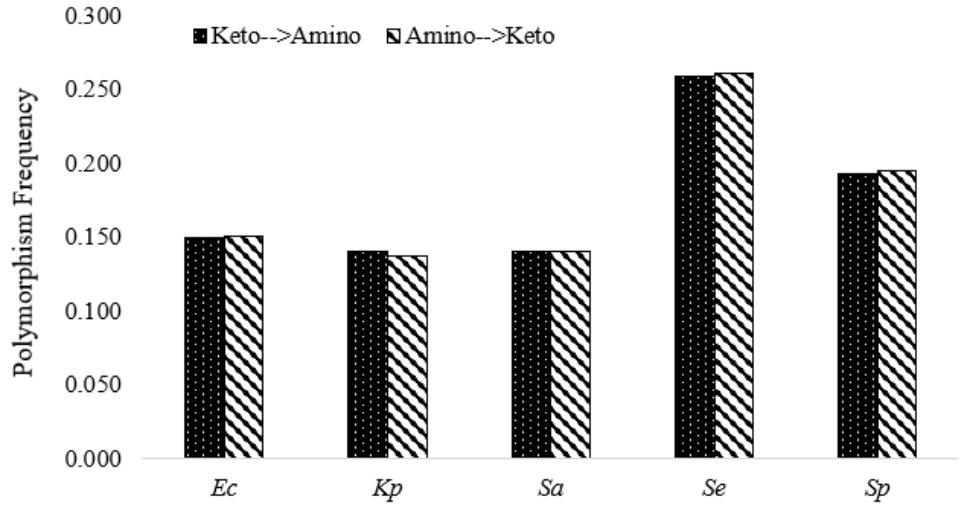

Figure A-6: Amino(A/C) → Keto (G/T), Keto (G/T) → Amino (A/C) polymorphism frequencies IRs of five bacteria

Histogram presents normalized values of Keto → Amino and Amino (A/C) → Keto (G/T), polymorphism frequencies in IRs. The x-axis presents the five bacteria *Escherichia coli* (*Ec*), *Klebsiella pneumoniae* (*Kp*), *Salmonella enterica* (*Se*), *Staphylococcus aureus* (*Sa*) and *Streptococcus pneumoniae* (*Sp*). From the polymorphism pattern it can be observed that, there is no significant difference between Keto → Amino and Amino (A/C) → Keto (G/T) transitions in the IRs (p-value > 0.01).

A.4.5 Details of tRNA studied in dot bracket notation

Github link for A.4.5: `https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_5/Chapter%205_Appendix%20A.4.5.xlsx`

A.4.6 List of tRNA genes of five bacteria

Github link for A.4.6: `https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_5/Chapter%205_Appendix%20A.4.6.docx`

A.4.7 Proportionate fold increase of ti as well as tv in the IRs than tRNA

Github link for A.4.7: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_5/Chapter%205_Appendix%20A.4.7.docx

A.4.8 Energy calculation of different base pairing

Github link for A.4.8: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_5/Chapter%205_Appendix%20A.4.8.docx

A.4.9 Isoacceptor Details

Github link for A.4.9: https://github.com/CBBILAB/Appendices_Thesis_Piyali_Sen/blob/main/Chapter_5/Chapter%205_Appendix%20A.4.9.xlsx

Appendix B

Publication list

The thesis contributions led to the following publications.

Journals

1. Sen, P., Aziz, R., Deka, R.C., Feil, E.J., Ray, S.K., and Satapathy, S.S. Stem region of tRNA genes favors transition substitution towards keto bases in bacteria. *Journal of Molecular Evolution*, 90(1):114-123, 2022. (Journal url:https://www.springer.com/journal/239)

2. Sen, P., Kurmi, A., Ray, S.K., and Satapathy, S.S. Machine learning approach identifies prominent codons from different degenerate groups influencing gene expression in bacteria. *Genes to Cells*, 27:591-601, 2022. (Journal url: https://onlinelibrary.wiley.com/journal/13652443)

3. Sen, P., Aziz, R., Das, S., Namsa, N.D., Deka, R.C., Feil, E.J., Ray, S.K., and Satapathy, S.S. Single nucleotide polymorphism at the four-fold degenerate sites, but not at the intergenic regions, explains the difference among bacterial genomes regarding nucleotide compositional strand asymmetry. (*Under Review*).

4. Sen, P., and Satapathy, S.S. A Review on Recent Advancements in RNA

Secondary Structure Prediction Algorithms Using Deep Learning Methods. (*Under Review*).

Conference

1. Sen, P., Waris, A., Ray, S.K., and Satapathy, S.S. A web portal to calculate Codon Adaptation Index (CAI) with organism specific reference set of high expression genes for diverse bacteria species, In *International conference on Intelligent Computing and Smart Communication (ICSC'19)*, Springer, 2019.

2. Sen, P., Tula, D., Ray, SK., and Satapathy, SS. Bhardwaj, K., Das, A., and Patra, S. SatapathyEstimating RNA Secondary structure by maximizing stacking regions, In *International Conference on Computer Communication and Internet Of Things (ICCCIOT'20)*, Springer, 2020.

Bibliography

[1] Abdullah, T., Faiza, M., Pant, P., Akhtar, M. R., and Pant, P. An analysis of single nucleotide substitution in genetic codons-probabilities and outcomes. *Bioinformation*, 12(3):98, 2016.

[2] Abe, H. and Aiba, H. Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabilization. *Biochimie*, 78(11-12):1035–1042, 1996.

[3] Agris, P. F. Bringing order to translation: the contributions of transfer RNA anticodondomain modifications. *EMBO Rep.*, 9(7):629–635, 2008.

[4] Agris, P. F., Narendran, A., Sarachan, K., Väre, V. Y. P., and Eruysal, E. The importance of being modified: the role of RNA modifications in translational fidelity. *Enzym.*, 41:1–50, 2017.

[5] Akashi, H. Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics*, 136(3):927–935, 1994.

[6] Aziz, R., Sen, P., Beura, P. K., Das, S., Tula, D., Dash, M., Namsa, N. D., Deka, R. C., Feil, E. J., Satapathy, S. S., et al. Incorporation of transition to transversion ratio and nonsense mutations, improves the estimation of the number of synonymous and non-synonymous sites in codons. *DNA Research*, 29(4):dsac023, 2022.

[7] Balbi, K. J., Rocha, E. P., and Feil, E. J. The temporal dynamics of slightly

deleterious mutations in Escherichia coli and Shigella spp. *Mol. Biol. Evol.*, 26(2):345–355, 2009.

[8] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., and Nielsen, H. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–424, 2000.

[9] Bao, C., Loerch, S., Ling, C., Korostelev, A. A., Grigorieff, N., and Ermolenko, D. N. mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding. *Elife*, 9:e55799, 2020.

[10] Baruah, V. J., Satapathy, S. S., Powdel, B. R., Konwarh, R., Buragohain, A. K., and Ray, S. K. Comparative analysis of codon usage bias in crenarchaea and euryarchaea genome reveals differential preference of synonymous codons to encode highly expressed ribosomal and rna polymerase proteins. *Journal of genetics*, 95:537–549, 2016.

[11] Baulin, E., Yacovlev, V., Khachko, D., Spirin, S., and Roytberg, M. URS DataBase: universe of RNA structures and their motifs. *Database*, 2016, 2016.

[12] Bellaousov, S., Reuter, J. S., Seetin, M. G., and Mathews, D. H. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*, 41(Web Server issue), 2013.

[13] Bennetzen, J. L. and Hall, B. D. Codon selection in yeast. *Journal of Biological Chemistry*, 257(6):3026–3031, 1982.

[14] Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R., and Stadler, P. F. Rnaalifold: improved consensus structure prediction for rna alignments. *BMC bioinformatics*, 9(1):474, 2008.

[15] Beura, P. K., Sen, P., Aziz, R., Satapathy, S. S., and Ray, S. K. Transcribed intergenic regions exhibit a lower frequency of nucleotide polymorphism than

the untranscribed intergenic regions in the genomes of escherichia coli and salmonella enterica. *Journal of Genetics*, 102:22, 2023.

[16] Bobay, L.-M. and Ochman, H. Impact of recombination on the base composition of bacteria and archaea. *Mol. Biol. Evol.*, 34(10):2627–2636, 2017.

[17] Bonnet, E., Rzewski, P., and Sikora, F. Designing rna secondary structures is hard. *Journal of Computational Biology*, 2020.

[18] Botzman, M. and Margalit, H. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome biology*, 12 (10):1–11, 2011.

[19] Breiman, L. Random forests. *Mach. Learn.*, 45(1):5–32, 2001.

[20] Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics*, 129(3):897–907, 1991.

[21] Byun, Y. and Han, K. Pseudoviewer3: generating planar drawings of large-scale rna structures with pseudoknots. *Bioinformatics*, 25(11):1435–1437, 2009.

[22] Calonaci, N., Jones, A., Cuturello, F., Sattler, M., and Bussi, G. Machine learning a model for rna structure prediction. *NAR genomics and bioinformatics*, 2(4):lqaa090, 2020.

[23] Cao, Y., Fu, L., Wu, J., Nie, Q., and Xie, X. UFold: Fast and Accurate RNA Secondary Structure Prediction with Deep Learning. *bioRxiv*, 2020.

[24] Castro, E., Benz, A., Tong, A., Wolf, G., and Krishnaswamy, S. Uncovering the Folding Landscape of RNA Secondary Structure with Deep Graph Embeddings. *arXiv Prepr. arXiv2006.06885*, 2020.

[25] Charif, D., Thioulouse, J., Lobry, J., and Perrière, G. Online synonymous codon usage analyses with the ade4 and seqinr packages. *Bioinformatics*, 21 (4):545–547, 2005.

[26] Charneski, C. A., Honti, F., Bryant, J. M., Hurst, L. D., and Feil, E. J. Atypical AT skew in Firmicute genomes results from selection and not from mutation. *PLoS Genet*, 7(9):e1002283, 2011.

[27] Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L., and McAdams, H. H. Codon usage between genomes is constrained by genome-wide mutational processes. *Proc. Natl. Acad. Sci.*, 101(10):3480–3485, 2004.

[28] Chen, T. and Guestrin, C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 13-17-Augu:785–794, 2016.

[29] Chen, X., Li, Y., Umarov, R., Gao, X., and Song, L. Rna secondary structure prediction by learning unrolled algorithms. *arXiv Prepr. arXiv2002.05810*, 2020.

[30] Chen, X., Li, Y., Umarov, R., Gao, X., Song, L., Tech, G., and Financial, A. Rna S Econdary S Tructure P Rediction. pages 1–19, 2020.

[31] Chewapreecha, C., Marttinen, P., Croucher, N. J., Salter, S. J., Harris, S. R., Mather, A. E., Hanage, W. P., Goldblatt, D., Nosten, F. H., and Turner, C. Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLoS Genet*, 10(8):e1004547, 2014.

[32] Clarke, B. Darwinian evolution of proteins. *Science*, 168(3934):1009–1011, 1970.

[33] Danaee, P., Rouches, M., Wiley, M., Deng, D., Huang, L., and Hendrix, D. bpRNA: large-scale automated annotation and analysis of RNA secondary structure. *Nucleic Acids Res.*, 46(11):5381–5394, 2018.

[34] Darty, K., Denise, A., and Ponty, Y. Varna: Interactive drawing and editing of the rna secondary structure. *Bioinformatics*, 25(15):1974, 2009.

[35] Davis, B. D. Transcriptional bias: a non-Lamarckian mechanism for substrate-induced mutations. *Proc. Natl. Acad. Sci.*, 86(13):5005–5009, 1989.

[36] Dietel, A.-K., Merker, H., Kaltenpoth, M., and Kost, C. Selective advantages favour high genomic AT-contents in intracellular elements. *PLoS Genet.*, 15 (4):e1007778, 2019.

[37] Do, C. B., Woods, D. A., and Batzoglou, S. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, 22(14): 90–98, 2006.

[38] Du, M.-Z., Zhang, C., Wang, H., Liu, S., Wei, W., and Guo, F.-B. The gc content as a main factor shaping the amino acid usage during bacterial evolution process. *Frontiers in microbiology*, 9:2948, 2018.

[39] Duchêne, S., Ho, S. Y. W., and Holmes, E. C. Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models. *BMC Evol. Biol.*, 15(1):1–10, 2015.

[40] Elias, R. and Hoksza, D. Traveler: a tool for template-based rna secondary structure visualization. *BMC bioinformatics*, 18(1):1–10, 2017.

[41] Eun, H.-M. Enzymes and Nucleic Acids. In Eun, H.-M. B. T. E. P. f. R. D. N. A. T., editor, *Enzymol. Prim. Recomb. DNA Technol.*, pages 1–108. Academic Press, San Diego, 1996.

[42] Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.*, 27(8): 861–874, 2006.

[43] Fechter, P., Rudinger-Thirion, J., Florentz, C., and Giege, R. Novel features in the tRNA-like world of plant viral RNAs. *Cell. Mol. Life Sci. C.*, 58(11): 1547–1561, 2001.

[44] Fersht, A. R. and Knill-Jones, J. W. DNA polymerase accuracy and spontaneous mutation rates: frequencies of purine. purine, purine. pyrimidine,

and pyrimidine. pyrimidine mismatches during DNA replication. *Proc. Natl. Acad. Sci.*, 78(7):4251–4255, 1981.

[45] Forsdyke, D. R. and Mortimer, J. R. Chargaff's legacy. *Gene*, 261(1):127–137, 2000.

[46] Francino, M. P. and Ochman, H. Strand asymmetries in DNA evolution. *Trends Genet.*, 13(6):240–245, 1997.

[47] Frank, A. C. and Lobry, J. R. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene*, 238(1):65–77, 1999.

[48] Frisch, M., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., and Petersson, G. gaussian 09, Revision d. 01, Gaussian. *Inc., Wallingford CT*, 201, 2009.

[49] Fuglsang, A. The effective number of codons revisited. *Biochemical and biophysical research communications*, 317(3):957–964, 2004.

[50] Gaillard, H., Herrera-Moyano, E., and Aguilera, A. Transcription-associated genome instability. *Chem. Rev.*, 113(11):8638–8661, 2013.

[51] Galtier, N. and Lobry, J. R. Relationships between genomic G+ C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.*, 44(6):632–636, 1997.

[52] Garey, M. R. and Johnson, D. S. *Computers and intractability*, volume 29. wh freeman New York, 2002.

[53] Gavril, F. Algorithms on circular-arc graphs. *Networks*, 4(4):357–369, 1974.

[54] Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, 2003.

[55] Gholami, H., Mohammadifar, A., Golzari, S., Kaskaoutis, D. G., and Collins, A. L. Using the boruta algorithm and deep learning models for mapping land susceptibility to atmospheric dust emissions in iran. *Aeolian Research*, 50:100682, 2021.

[56] Gojobori, T., Li, W.-H., and Graur, D. Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.*, 18(5):360–369, 1982.

[57] Gouy, M. and Gautier, C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.*, 10(22):7055–7074, 1982.

[58] Grantham, R., Gautier, C., and Gouy, M. Codon frequencies in 119 individual genes confirm corsistent choices of degenerate bases according to genome type. *Nucleic Acids Res.*, 8(9):1893–1912, 1980.

[59] Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pave, A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.*, 8(1):197–197, 1980.

[60] Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.*, 9(1):213–213, 1981.

[61] Greenacre, M. J. Theory and applications of correspondence analysis. 1984.

[62] Grigoriev, A. Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res.*, 26(10):2286–2290, 1998.

[63] Grosjean, H. and Fiers, W. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene*, 18(3):199–209, 1982.

[64] Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.*, 36(Web Server issue):70–74, 2008.

[65] Grundy, F. J. and Henkin, T. M. From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit. Rev. Biochem. Mol. Biol.*, 41(6):329–338, 2006.

[66] Hall, M. N., Gabay, J., Débarbouillé, M., and Schwartz, M. A role for mRNA secondary structure in the control of translation initiation. *Nature*, 295(5850): 616–618, 1982.

[67] Han, K. and Byun, Y. PseudoViewer2: visualization of RNA pseudoknots of any type. *Nucleic Acids Res.*, 31(13):3432–3440, 2003.

[68] Hershberg, R. and Petrov, D. A. General rules for optimal codon choice. *PLoS Genet*, 5(7):e1000556, 2009.

[69] Hershberg, R. and Petrov, D. A. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet*, 6(9):e1001115, 2010.

[70] Hickey, D. A. and Singer, G. A. C. Genomic and proteomic adaptations to growth at high temperature. *Genome Biol.*, 5(10):1–7, 2004.

[71] Higgs, P. G. RNA secondary structure: physical and computational aspects. *Q. Rev. Biophys.*, 33(3):199–253, 2000.

[72] Higgs, P. G. A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. *Biology direct*, 4(1):1–29, 2009.

[73] Hildebrand, F., Meyer, A., and Eyre-Walker, A. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet*, 6(9):e1001107, 2010.

[74] Hofacker, I. L. and Stadler, P. F. Automatic detection of conserved base pairing patterns in rna virus genomes. *Computers & chemistry*, 23(3-4): 401–414, 1999.

[75] Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A. *Science*, 147(3664):1462–1465.

[76] Holt, K. E., Wertheim, H., Zadoks, R. N., Baker, S., Whitehouse, C. A., Dance, D., Jenney, A., Connor, T. R., Hsu, L. Y., and Severin, J. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. *Proc. Natl. Acad. Sci.*, 112(27):E3574–E3581, 2015.

[77] Hor, C.-Y., Yang, C.-B., Chang, C.-H., Tseng, C.-T., and Chen, H.-H. A Tool preference choice Method for RnA secondary structure prediction by sVM with statistical Tests. *Evol. Bioinforma.*, 9:EBO–S10580, 2013.

[78] Hou, Y.-M. and Schimmel, P. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature*, 333(6169):140–145, 1988.

[79] Hsu, W.-L. The coloring and maximum independent set problems on planar perfect graphs. *Journal of the ACM (JACM)*, 35(3):535–563, 1988.

[80] Hudelot, C., Gowri-Shankar, V., Jow, H., Rattray, M., and Higgs, P. G. RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. *Mol. Phylogenet. Evol.*, 28(2):241–252, 2003.

[81] Huffman, J. Brief Papers. *Brain Cogn.*, 32(2):273–344, 1996.

[82] Ikemura, T. Correlation between the abundance of escherichia coli transfer rnas and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the e. coli translational system. *J. Mol. Evol.*, 151(3):389–409, 1981.

[83] Ikemura, T. Correlation between the abundance of yeast transfer rnas and the occurrence of the respective codons in protein genes: differences in synonymous codon choice patterns of yeast and escherichia coli with reference

to the abundance of isoaccepting transfer rnas. *J. Mol. Evol.*, 158(4):573–597, 1982.

[84] Ikemura, T. Codon usage and trna content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, 2(1):13–34, 1985.

[85] Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J., and Frishman, D. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics*, 9(1):1–17, 2008.

[86] Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.-L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J. H., and Ackerman, S. L. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science*, 345(6195):455–459, 2014.

[87] Ito, K. Strolling Toward New Concepts. *Annu. Rev. Microbiol.*, 70:1–23, 2016.

[88] Jia, L. and Sun, T. RNA secondary structure prediction based on long-range interaction and Support Vector Machine. *Acad. J. Comput. Inf. Sci.*, 3(3), 2020.

[89] Jinks-Robertson, S. and Bhagwat, A. S. Transcription-associated mutagenesis. *Annu. Rev. Genet.*, 48:341–359, 2014.

[90] Jow, H., Hudelot, C., Rattray, M., and Higgs, P. G. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. *Mol. Biol. Evol.*, 19(9):1591–1601, 2002.

[91] Karlin, S., Campbell, A. M., and Mrazek, J. Comparative DNA analysis across diverse genomes. *Annu. Rev. Genet.*, 32(1):185–225, 1998.

[92] Keller, I., Bensasson, D., and Nichols, R. A. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genet.*, 3(2):e22, 2007.

[93] Kerpedjiev, P., Hammer, S., and Hofacker, I. L. Forna (force-directed rna): simple and effective online rna secondary structure diagrams. *Bioinformatics*, 31(20):3377–3379, 2015.

[94] Kim, N. and Jinks-Robertson, S. Transcription as a source of genome instability. *Nat. Rev. Genet.*, 13(3):204–214, 2012.

[95] Knies, J. L., Dang, K. K., Vision, T. J., Hoffman, N. G., Swanstrom, R., and Burch, C. L. Compensatory evolution in RNA secondary structures increases substitution rate variation among sites. *Mol. Biol. Evol.*, 25(8):1778–1787, 2008.

[96] Knudsen, B. and Hein, J. Pfold: Rna secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.*, 31(13):3423–3428, 2003.

[97] Koonin, E. V. and Novozhilov, A. S. Origin and evolution of the genetic code: the universal enigma. *IUBMB life*, 61(2):99–111, 2009.

[98] Kriner, M. A. and Groisman, E. A. RNA secondary structures regulate three steps of Rho-dependent transcription termination within a bacterial mRNA leader. *Nucleic Acids Res.*, 45(2):631–642, 2017.

[99] Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, 35(6):1547–1549, 2018.

[100] Künstner, A., Nabholz, B., and Ellegren, H. Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. *Genome Biol. Evol.*, 3:1381–1389, 2011.

[101] Kursa, M. B. and Rudnicki, W. R. Feature selection with the Boruta package. *J Stat Softw*, 36(11):1–13, 2010.

[102] Kursa, M. B., Jankowski, A., and Rudnicki, W. R. Borutaa system for feature selection. *Fundam. Informaticae*, 101(4):271–285, 2010.

[103] Lari, K. and Young, S. J. Applications of stochastic context-free grammars using the inside-outside algorithm. *Comput. Speech Lang.*, 5(3):237–257, 1991.

[104] Lassalle, F., Périan, S., Bataillon, T., Nesme, X., Duret, L., and Daubin, V. GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genet*, 11(2):e1004941, 2015.

[105] Lewis, C. A., Crayle, J., Zhou, S., Swanstrom, R., and Wolfenden, R. Cytosine deamination and the precipitous decline of spontaneous mutation during Earth's history. *Proc. Natl. Acad. Sci.*, 113(29):8194–8199, 2016.

[106] Li, J., Xu, C., Wang, L., Liang, H., Feng, W., Cai, Z., Wang, Y., Cong, W., and Liu, Y. Psrna: Prediction of small rna secondary structures based on reverse complementary folding method. *Journal of bioinformatics and computational biology*, 14(04):1643001, 2016.

[107] Li, J., Xu, C., Liang, H., Cong, W., Wang, Y., Luan, K., and Liu, Y. RGRNA: prediction of RNA secondary structure based on replacement and growth of stems. *Comput. Methods Biomech. Biomed. Engin.*, 20(12):1261–1272, 2017.

[108] Liaw, A. and Wiener, M. Classification and regression by randomForest. *R news*, 2(3):18–22, 2002.

[109] Libri, D., Stutz, F., McCarthy, T., and Rosbash, M. RNA structural patterns and splicing: molecular basis for an RNA-based enhancer. *Rna*, 1(4):425–436, 1995.

[110] Lind, P. A. and Andersson, D. I. Whole-genome mutational biases in bacteria. *Proc. Natl. Acad. Sci.*, 105(46):17878–17883, 2008.

[111] Liu, Q., Ye, X., and Zhang, Y. A hopfield neural network based algorithm for RNA secondary structure prediction. *First Int. Multi- Symp. Comput. Comput. Sci. IMSCCS'06*, 1:10–16, 2006.

[112] Liu, Q., Ye, X., and Zhang, Y. A hopfield neural network based algorithm for rna secondary structure prediction. In *First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06)*, volume 1, pages 10–16. IEEE, 2006.

[113] Lobachev, K. S., Shor, B. M., Tran, H. T., Taylor, W., Keen, J. D., Resnick, M. A., and Gordenin, D. A. Factors affecting inverted repeat stimulation of recombination and deletion in saccharomyces cerevisiae. *Genetics*, 148(4): 1507–1524, 1998.

[114] Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.*, 13(5):660–665, 1996.

[115] Lobry, J. R. and Sueoka, N. Asymmetric directional mutation pressures in bacteria. *Genome Biol.*, 3(10):1–14, 2002.

[116] Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, 6(1):1–14, 2011.

[117] Lowe, T. M. and Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.*, 44(W1): W54–W57, 2016.

[118] Lu, W., Cao, Y., Wu, H., Huang, H., and Ding, Y. Research on RNA secondary structure prediction based on decision tree. In *Int. Conf. Intell. Comput.*, pages 430–439. Springer, 2019.

[119] Lu, W., Tang, Y., Wu, H., Huang, H., Fu, Q., Qiu, J., and Li, H. Predicting

RNA secondary structure via adaptive deep recurrent neural networks with energy-based filter. *BMC Bioinformatics*, 20(Suppl 25):1–10, 2019.

[120] Lu, W., Cao, Y., Wu, H., Ding, Y., Song, Z., Zhang, Y., Fu, Q., and Li, H. Research on RNA secondary structure predicting via bidirectional recurrent neural network. *BMC Bioinformatics*, 22:1–19, 2021.

[121] Lu, W., Chen, X., Zhang, Y., Wu, H., Shen, J., Zhou, N., and Ding, Y. Research on RNA Secondary Structure Prediction Based on MLP. In *Int. Conf. Intell. Comput.*, pages 336–344. Springer, 2021.

[122] Lu, Z. J., Gloor, J. W., and Mathews, D. H. Improved RNA secondary structure prediction by maximizing expected pair accuracy. *Rna*, 15(10): 1805–1813, 2009.

[123] Lyons, D. M. and Lauring, A. S. Evidence for the selective basis of transition-to-transversion substitution bias in two RNA viruses. *Mol. Biol. Evol.*, 34 (12):3205–3215, 2017.

[124] Maiti, M., Nauwelaerts, K., Lescrinier, E., Schuit, F. C., and Herdewijn, P. Self-complementary sequence context in mature miRNAs. *Biochem. Biophys. Res. Commun.*, 392(4):572–576, 2010.

[125] Mann, H. B. and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, pages 50–60, 1947.

[126] Mao, K., Wang, J., and Xiao, Y. Prediction of RNA secondary structure with pseudoknots using coupled deep neural networks. *Biophys. Reports*, 6 (4):146–154, 2020.

[127] Mao, K., Wang, J., and Xiao, Y. Length-dependent deep learning model for rna secondary structure prediction. *Molecules*, 27(3):1030, 2022.

[128] Mathews, D. H. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *Rna*, 10(8):1178–1190, 2004.

[129] Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.*, 101(19):7287–7292, 2004.

[130] McClain, W. H. and Foss, K. Changing the identity of a tRNA by introducing a GU wobble pair near the 3'acceptor end. *Science*, 240(4853):793–796, 1988.

[131] McInerney, J. O. Replicational and transcriptional selection on codon usage in borrelia burgdorferi. *Proc. Natl. Acad. Sci.*, 95(18):10698–10703, 1998.

[132] McLean, M. J., Wolfe, K. H., and Devine, K. M. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J. Mol. Evol.*, 47(6):691–696, 1998.

[133] Monschke, K. A. and Arslan, A. N. Rna secondary structure graphical rendering library. In *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–6. IEEE, 2018.

[134] Montange, R. K. and Batey, R. T. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.*, 37:117–133, 2008.

[135] Mugal, C. F., Grünberg, H.-H. von, and Peifer, M. Transcription-induced mutational strand bias and its effect on substitution rates in human genes. *Mol. Biol. Evol.*, 26(1):131–142, 2009.

[136] Muto, A. and Osawa, S. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci.*, 84(1):166–169, 1987.

[137] Novembre, J. A. Accounting for background nucleotide composition when measuring codon usage bias. *Mol. Biol. Evol.*, 19(8):1390–1394, 2002.

[138] Novoa, E. M., Pavon-Eternod, M., Pan, T., and Pouplana, L. R. de. A role for trna modifications in genome structure and codon usage. *Cell*, 149(1): 202–213, 2012.

[139] Nussinov, R. and Jacobson, A. B. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci.*, 77(11):6309–6313, 1980.

[140] O'Connor, M. tRNA imbalance promotes 1 frameshifting via near-cognate decoding. *J. Mol. Biol.*, 279(4):727–736, 1998.

[141] O'Connor, M. Imbalance of tRNAPro isoacceptors induces+ 1 frameshifting at near-cognate codons. *Nucleic Acids Res.*, 30(3):759–765, 2002.

[142] Osawa, S., Jukes, T. H., Watanabe, K., and Muto, A. Recent evidence for evolution of the genetic code. *Microbiological reviews*, 56(1):229–264, 1992.

[143] Palidwor, G. A., Perkins, T. J., and Xia, X. A general model of codon bias due to gc mutational bias. *PLoS One*, 5(10):e13431, 2010.

[144] Park, C., Chen, X., Yang, J.-R., and Zhang, J. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci.*, 110(8):E678–E686, 2013.

[145] Peden, J. and Sharp, P. Codonw: correspondence analysis of codon usage, 2005.

[146] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011.

[147] Pereira, M., Francisco, S., Varanda, A. S., Santos, M., Santos, M. A. S., and Soares, A. R. Impact of tRNA modifications and tRNA-modifying enzymes on proteostasis and human disease. *Int. J. Mol. Sci.*, 19(12):3738, 2018.

[148] Peselis, A. and Serganov, A. Structure and function of pseudoknots involved in gene expression control. *Wiley Interdiscip. Rev. RNA*, 5(6):803–822, 2014.

[149] Plotkin, J. B., Robins, H., and Levine, A. J. Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci.*, 101(34):12588–12591, 2004.

[150] Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, 20(1):110–121, 2010.

[151] Powdel, B. R., Satapathy, S. S., Kumar, A., Jha, P. K., Buragohain, A. K., Borah, M., and Ray, S. K. A study in entire chromosomes of violations of the intra-strand parity of complementary nucleotides (Chargaff's second parity rule). *DNA Res.*, 16(6):325–343, 2009.

[152] Powdel, B. R., Borah, M., and Ray, S. K. Strand-specific mutational bias influences codon usage of weakly expressed genes in escherichia coli. *Genes to Cells*, 15(7):773–782, 2010.

[153] Proctor, J. R. and Meyer, I. M. Cofold: thermodynamic rna structure prediction with a kinetic twist. *arXiv preprint arXiv:1207.6013*, 2012.

[154] Pruitt, K. D., Tatusova, T., and Maglott, D. R. Ncbi reference sequence (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, 33(suppl_1):D501–D504, 2005.

[155] Puigbò, P., Bravo, I. G., and Garcia-Vallve, S. Caical: a combined set of tools to assess codon usage adaptation. *Biology direct*, 3(1):1–8, 2008.

[156] Qasim, R., Kauser, N., and Jilani, T. Secondary structure prediction of rna using machine learning method. *International Journal of Computer Applications*, 975:8887, 2010.

[157] Quan, L., Cai, L., Chen, Y., Mei, J., Sun, X., and Lyu, Q. Developing parallel ant colonies filtered by deep learned constrains for predicting RNA secondary structure with pseudo-knots. *Neurocomputing*, 384:104–114, 2020.

[158] Rafi, R. H., Morshed, I., Shishir, M. A. R., and Mahdi, T. OpenCL: Finding Minimal Free Energy of RNA Secondary Structure Based on Multiple GPUs. *Int. J. Comput. Sci. Inf. Secur.*, 18(3), 2020.

[159] Raghavan, R., Kelkar, Y. D., and Ochman, H. A selective force favoring increased G+ C content in bacterial genes. *Proc. Natl. Acad. Sci.*, 109(36): 14504–14507, 2012.

[160] Ran, W. and Higgs, P. G. The influence of anticodoncodon interactions and modified bases on codon usage bias in bacteria. *Mol. Biol. Evol.*, 27(9): 2129–2140, 2010.

[161] Ray, S. K. and Goswami, I. Synonymous codons are not the same with respect to the speed of translation elongation. *Curr. Sci.*, 110(9):1612–1614, 2016.

[162] Ray, S. K. and Goswami, I. Synonymous codons are not the same with respect to the speed of translation elongation. *Current Science*, 110(9):1612–1614, 2016.

[163] Ray, S. K., Baruah, V. J., Satapathy, S. S., and Banerjee, R. Cotranslational protein folding reveals the selective use of synonymous codons along the coding sequence of a low expression gene. *Journal of genetics*, 93:613–617, 2014.

[164] Reuter, J. S. and Mathews, D. H. Rnastructure: software for rna secondary structure prediction and analysis. *BMC bioinformatics*, 11(1):129, 2010.

[165] Reuter, S., Török, M. E., Holden, M. T. G., Reynolds, R., Raven, K. E., Blane, B., Donker, T., Bentley, S. D., Aanensen, D. M., and Grundmann, H. Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. *Genome Res.*, 26(2):263–270, 2016.

[166] Reyes, A., Gissi, C., Pesole, G., and Saccone, C. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.*, 15(8):957–966, 1998.

[167] Rice, P., Longden, I., and Bleasby, A. Emboss: the european molecular biology open software suite. *Trends in genetics*, 16(6):276–277, 2000.

[168] Rocha, E. P. C. The replication-related organization of bacterial genomes. *Microbiology*, 150(6):1609–1627, 2004.

[169] Rocha, E. P. C. and Danchin, A. Ongoing evolution of strand composition in bacterial genomes. *Mol. Biol. Evol.*, 18(9):1789–1799, 2001.

[170] Rocha, E. P. C. and Feil, E. J. Mutational patterns cannot explain genome composition: are there any neutral sites in the genomes of bacteria? *PLoS Genet*, 6(9):e1001104, 2010.

[171] Rocha, E. P. C., Danchin, A., and Viari, A. Universal replication biases in bacteria. *Mol. Microbiol.*, 32(1):11–16, 1999.

[172] Rocha, E. P. C., Touchon, M., and Feil, E. J. Similar compositional biases are caused by very different mutational effects. *Genome Res.*, 16(12):1537–1547, 2006.

[173] Rossetti, G., Dans, P. D., Gomez-Pinto, I., Ivani, I., Gonzalez, C., and Orozco, M. The structural impact of DNA mismatches. *Nucleic Acids Res.*, 43(8):4309–4321, 2015.

[174] Roth, A., Anisimova, M., and Cannarozzi, G. M. Measuring codon usage bias. *Codon evolution: mechanisms and models*, pages 189–217, 2012.

[175] Saha, S. K., Goswami, A., and Dutta, C. Association of purine asymmetry, strand-biased gene distribution and PolC within Firmicutes and beyond: a new appraisal. *BMC Genomics*, 15(1):1–26, 2014.

[176] Satapathy, S., Ray, S., Sahoo, A., Begum, T., and Ghosh, T. Codon usage bias is not significantly different between the high and the low expression genes in human. *Int. J. Mol. Genet. Gene Ther*, 1(1), 2015.

[177] Satapathy, S. S., Dutta, M., Buragohain, A. K., and Ray, S. K. Transfer RNA gene numbers may not be completely responsible for the codon usage bias in asparagine, isoleucine, phenylalanine, and tyrosine in the high expression genes in bacteria. *J. Mol. Evol.*, 75(1-2):34–42, 2012.

[178] Satapathy, S. S., Powdel, B. R., Dutta, M., Buragohain, A. K., and Ray, S. K. Selection on GGU and CGU codons in the high expression genes in bacteria. *J. Mol. Evol.*, 78(1):13–23, 2014.

[179] Satapathy, S. S., Powdel, B. R., Dutta, M., Buragohain, A. K., and Ray, S. K. Selection on ggu and cgu codons in the high expression genes in bacteria. *Journal of molecular evolution*, 78(1):13–23, 2014.

[180] Satapathy, S. S., Powdel, B. R., Buragohain, A. K., and Ray, S. K. Discrepancy among the synonymous codons with respect to their selection as optimal codon in bacteria. *DNA Res.*, 23(5):441–449, 2016.

[181] Satapathy, S. S., Sahoo, A. K., Ray, S. K., and Ghosh, T. C. Codon degeneracy and amino acid abundance influence the measures of codon usage bias: improved Nc (Nc) and ENCprime (N c) measures. *Genes to Cells*, 22 (3):277–283, 2017.

[182] Sato, K., Akiyama, M., and Sakakibara, Y. RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.*, 12(1):1–9, 2021.

[183] Savill, N. J., Hoyle, D. C., and Higgs, P. G. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics*, 157(1):399–411, 2001.

[184] Schroeder, J. W., Randall, J. R., Hirst, W. G., O'Donnell, M. E., and Simmons, L. A. Mutagenic cost of ribonucleotides in bacterial DNA. *Proc. Natl. Acad. Sci.*, 114(44):11733–11738, 2017.

[185] Sen, P., Waris, A., Ray, S. K., and Satapathy, S. S. A Web Portal to Calculate Codon Adaptation Index (CAI) with Organism Specific Reference Set of High Expression Genes for Diverse Bacteria Species. In *Int. Conf. Intell. Comput. Smart Commun. 2019*, pages 319–325. Springer, 2020.

[186] Sen, P., Aziz, R., Deka, R. C., Feil, E. J., Ray, S. K., and Satapathy, S. S. Stem Region of tRNA Genes Favors Transition Substitution Towards Keto Bases in Bacteria. *J. Mol. Evol.*, pages 1–10, 2022.

[187] Seplyarskiy, V. B., Kharchenko, P., Kondrashov, A. S., and Bazykin, G. A. Heterogeneity of the transition/transversion ratio in Drosophila and Hominidae genomes. *Mol. Biol. Evol.*, 29(8):1943–1955, 2012.

[188] Shah, P. and Gilchrist, M. A. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci.*, 108(25):10231–10236, 2011.

[189] Shaheen, A. and Iqbal, J. Spatial distribution and mobility assessment of carcinogenic heavy metals in soil profiles using geostatistics and random forest, boruta algorithm. *Sustainability*, 10(3):799, 2018.

[190] Sharp, P. M. and Li, W.-H. Codon usage in regulatory genes in escherichia coli does not reflect selection for rarecodons. *Nucleic Acids Res.*, 14(19): 7737–7749, 1986.

[191] Sharp, P. M. and Li, W.-H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of molecular evolution*, 24: 28–38, 1986.

[192] Sharp, P. M. and Li, W.-H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, 15(3):1281–1295, 1987.

[193] Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F., and Sockett, R. E. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.*, 33(4):1141–1153, 2005.

[194] Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F., and Sockett, R. E. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.*, 33(4):1141–1153, 2005.

[195] Sharp, P. M., Emery, L. R., and Zeng, K. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. B Biol. Sci.*, 365(1544):1203–1212, 2010.

[196] Shen, B., Zhang, H., Li, C., Zhao, T., and Liu, Y. Deep Learning Method for RNA Secondary Structure Prediction with Pseudoknots Based on Large-Scale Data. *J. Healthc. Eng.*, 2021, 2021.

[197] Singh, J., Hanson, J., Paliwal, K., and Zhou, Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat. Commun.*, 10(1), 2019.

[198] Singh, J., Paliwal, K., Zhang, T., Singh, J., Litfin, T., and Zhou, Y. Improved RNA secondary structure and tertiary base-pairing prediction using evolu-

tionary profile, mutational coupling and two-dimensional transfer learning. *Bioinformatics*, 2021.

[199] Sinha, N. K. and Haimes, M. D. Molecular mechanisms of substitution mutagenesis. An experimental test of the Watson-Crick and topal-fresco models of base mispairings. *J. Biol. Chem.*, 256(20):10671–10683, 1981.

[200] Sohmen, D., Chiba, S., Shimokawa-Chiba, N., Innis, C. A., Berninghausen, O., Beckmann, R., Ito, K., and Wilson, D. N. Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling. *Nat. Commun.*, 6(1):1–10, 2015.

[201] Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H., et al. The bioperl toolkit: Perl modules for the life sciences. *Genome research*, 12(10):1611–1618, 2002.

[202] Stern, S., Weiser, B., and Noller, H. F. Model for the three-dimensional folding of 16 s ribosomal rna. *J. Mol. Evol.*, 204(2):447–481, 1988.

[203] Stoletzki, N. and Eyre-Walker, A. Synonymous codon usage in Escherichia coli: selection for translational accuracy. *Mol. Biol. Evol.*, 24(2):374–381, 2007.

[204] Stoltzfus, A. and Norris, R. W. On the causes of evolutionary transition: transversion bias. *Mol. Biol. Evol.*, 33(3):595–602, 2016.

[205] Sueoka, N. Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. *Proc. Natl. Acad. Sci.*, 47(8):1141–1149, 1961.

[206] Sueoka, N. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci.*, 85(8):2653–2657, 1988.

[207] Sueoka, N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol.*, 40(3):318–325, 1995.

[208] Sun, X., Yang, Q., and Xia, X. An improved implementation of effective number of codons (n c). *Mol. Biol. Evol.*, 30(1):191–196, 2013.

[209] Supek, F. and Vlahoviček, K. Inca: synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics*, 20(14):2329–2330, 2004.

[210] Supek, F., Škunca, N., Repar, J., Vlahoviček, K., and Šmuc, T. Translational selection is ubiquitous in prokaryotes. *PLoS genetics*, 6(6):e1001004, 2010.

[211] Suzuki, T. and Kamiya, H. Mutations induced by 8-hydroxyguanine (8-oxo-7, 8-dihydroguanine), a representative oxidized base, in mammalian cells. *Genes Environ.*, 39(1):1–6, 2017.

[212] Takefuji, Y. and CHEN, L. Parallel algorithms for finding a near-maximum independent set of. *IEEE Transactions on Neural Networks*, 1(3):263, 1990.

[213] Taufer, M., Licon, A., Araiza, R., Mireles, D., Van Batenburg, F. H. D., Gultyaev, A. P., and Leung, M.-Y. PseudoBase++: an extension of PseudoBase for easy searching, formatting and visualization of pseudoknots. *Nucleic Acids Res.*, 37(suppl_1):D127–D135, 2009.

[214] Tharwat, A. Classification assessment methods. *Appl. Comput. Informatics*, 2020.

[215] Thorpe, H. A., Bayliss, S. C., Hurst, L. D., and Feil, E. J. Comparative analyses of selection operating on nontranslated intergenic regions of diverse bacterial species. *Genetics*, 206(1):363–376, 2017.

[216] Tinoco, I., Uhlenbeck, O. C., and Levine, M. D. Estimation of secondary structure in ribonucleic acids. *Nature*, 230(5293):362–367, 1971.

[217] Topal, M. D. and Fresco, J. R. Complementary base pairing and the origin of substitution mutations. *Nature*, 263(5575):285–289, 1976.

[218] Tsukiyama, S., Ide, M., Ariyoshi, H., and Shirakawa, I. A new algorithm for generating all the maximal independent sets. *SIAM Journal on Computing*, 6(3):505–517, 1977.

[219] Tubbs, A. and Nussenzweig, A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell*, 168(4):644–656, 2017.

[220] Turner, D. H. and Mathews, D. H. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.*, 38(suppl_1):D280–D282, 2010.

[221] Turner, D. H., Sugimoto, N., and Freier, S. M. Rna structure prediction. *Annual review of biophysics and biophysical chemistry*, 17(1):167–192, 1988.

[222] Van Batenburg, F., Gultyaev, A. P., and Pleij, C. W. An apl-programmed genetic algorithm for the prediction of rna secondary structure. *Journal of theoretical Biology*, 174(3):269–280, 1995.

[223] Van Leuven, J. T. and McCutcheon, J. P. An AT mutational bias in the tiny GC-rich endosymbiont genome of Hodgkinia. *Genome Biol. Evol.*, 4(1): 24–27, 2012.

[224] Loon, B. van, Markkanen, E., and Hübscher, U. Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA Repair (Amst).*, 9(6):604–616, 2010.

[225] Van Neste, C., Umarov, R., Li, Y., Salhi, A., Kuwahara, H., and Gao, X. NNfold: RNA Secondary Structure Prediction by Deep Learning with an Architecture Imposing Contextual Constraining. *Available SSRN 3813288*, 2021.

[226] Vogel, F. and Kopun, M. Higher frequencies of transitions among point mutations. *J. Mol. Evol.*, 9(2):159–180, 1977.

[227] Wald, N., Alroy, M., Botzman, M., and Margalit, H. Codon usage bias in prokaryotic pyrimidine-ending codons is associated with the degeneracy of the encoded amino acids. *Nucleic Acids Res.*, 40(15):7074–7083, 2012.

[228] Wang, H.-c. and Hickey, D. A. Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic Acids Res.*, 30(11):2501–2507, 2002.

[229] Wang, H.-C., Xia, X., and Hickey, D. Thermal adaptation of the small subunit ribosomal RNA gene: a comparative study. *J. Mol. Evol.*, 63(1): 120–126, 2006.

[230] Wang, J., Youkharibache, P., Zhang, D., Lanczycki, C. J., Geer, R. C., Madej, T., Phan, L., Ward, M., Lu, S., and Marchler, G. H. iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics*, 36(1):131–135, 2020.

[231] Wang, L., Liu, Y., Zhong, X., Liu, H., Lu, C., Li, C., and Zhang, H. DMFold: A novel method to predict RNA secondary structure with pseudoknots based on deep learning and improved base pair maximization principle. *Front. Genet.*, 10(MAR):1–12, 2019.

[232] Wang, L., Liu, Y., Zhong, X., Liu, H., Lu, C., Li, C., and Zhang, H. Dmfold: A novel method to predict rna secondary structure with pseudoknots based on deep learning and improved base pair maximization principle. *Frontiers in genetics*, 10:143, 2019.

[233] Wang, L., Zhong, X., Wang, S., Zhang, H., and Liu, Y. A novel end-to-end method to predict RNA secondary structure profile based on bidirectional LSTM and residual neural network. *BMC Bioinformatics*, 22(1):1–15, 2021.

[234] Wang, Y., Liu, Y., Wang, S., Liu, Z., Gao, Y., Zhang, H., and Dong, L. Attfold: Rna secondary structure prediction with pseudoknots based on attention mechanism. *Frontiers in Genetics*, 11:612086, 2020.

[235] Willmott, D. *Recurrent Neural Networks and Their Applications to RNA Secondary Structure Inference*. 2018.

[236] Willmott, D., Murrugarra, D., and Ye, Q. Improving RNA secondary structure prediction via state inference with deep recurrent neural networks. *Comput. Math. Biophys.*, 8(1):36–50, 2020.

[237] Wright, F. The effective number of codons used in a gene. *Gene*, 87(1):23–29, 1990.

[238] Wu, C.-I. and Maeda, N. Inequality in mutation rates of the two strands of DNA. *Nature*, 327(6118):169–170, 1987.

[239] Wu, G., Culley, D. E., and Zhang, W. Predicted highly expressed genes in the genomes of streptomyces coelicolor and streptomyces avermitilis and the implications for their metabolism. *Microbiology*, 151(7):2175–2187, 2005.

[240] Xia, X. An improved implementation of codon adaptation index. *Evolutionary Bioinformatics*, 3:117693430700300028, 2007.

[241] Xia, X. Dambe7: New and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.*, 35(6):1550–1552, 2018.

[242] Xu, X., Zhao, P., and Chen, S.-J. Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One*, 9(9):e107504, 2014.

[243] Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y., and Johnson, C. H. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature*, 495(7439):116–120, 2013.

[244] Yan, Z., Hamilton, W. L., and Blanchette, M. Neural representation and generation for RNA secondary structures. *arXiv Prepr. arXiv2102.00925*, 2021.

[245] Yan, Z., Hamilton, W. L., and Blanchette, M. D. Integrated pretraining with evolutionary information to improve rna secondary structure prediction. *bioRxiv*, 2022.

[246] Yu, C.-H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., and Liu, Y. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Molecular cell*, 59(5):744–754, 2015.

[247] Zhang, H., Zhang, C., Li, Z., Li, C., Wei, X., Zhang, B., and Liu, Y. A new method of RNA secondary structure prediction based on convolutional neural network and dynamic programming. *Front. Genet.*, 10(MAY):1–12, 2019.

[248] Zhang, K. and Lv, Y. A Multiobjective RNA Secondary Structure Prediction Algorithm Based on NSGAII. In *2018 IEEE SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov.*, pages 1450–1454, 2018.

[249] Zhang, L., Zhang, H., Mathews, D. H., and Huang, L. Threshknot: Thresholded probknot for improved RNA secondary structure prediction. *arXiv Prepr. arXiv1912.12796*, 2019.

[250] Zhang, Z., Wang, J., Wang, J., Wang, J., and Li, Y. Estimate of the sequenced proportion of the global prokaryotic genome. *Microbiome*, 8(1): 1–9, 2020.

[251] Zhao, C., Xu, X., and Chen, S.-J. Predicting RNA structure with Vfold. In *Funct. Genomics*, pages 3–15. Springer, 2017.

[252] Zhao, F., Yu, C.-h., and Liu, Y. Codon usage regulates protein structure

and function by affecting translation elongation speed in drosophila cells. *Nucleic Acids Res.*, 45(14):8484–8492, 2017.

[253] Zhao, Y., Wang, J., Zeng, C., and Xiao, Y. Evaluation of RNA secondary structure prediction for both base-pairing and topology. *Biophys. reports*, 4 (3):123–132, 2018.

[254] Zok, T., Antczak, M., Zurkowski, M., Popenda, M., Blazewicz, J., Adamiak, R. W., and Szachniuk, M. Rnapdbee 2.0: multifunctional tool for rna structure annotation. *Nucleic Acids Res.*, 46(W1):W30–W35, 2018.

[255] Zuckerkandl, E. and Pauling, L. Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins*, pages 97–166. Elsevier, 1965.

[256] Zuker, M. Prediction of rna secondary structure by energy minimization. In *Computer analysis of sequence data*, pages 267–294. Springer, 1994.

[257] Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13):3406–3415, 2003.

[258] Zuker, M. and Sankoff, D. RNA secondary structures and their prediction. *Bull. Math. Biol.*, 46(4):591–621, 1984.

[259] Zuker, M. and Stiegler, P. Optimal computer folding of large rna sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, 9(1): 133–148, 1981.