

DNA Codon Recognition by a Cubane Wire: In Silico Approach

Mahmoud Mirzaei^{a,b,1}, Nasser Hadipour^c, Oğuz Gülseren^d

^aBioinformatics Research Center, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran

^bBiosensor Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

^cDepartment of Chemistry, Tarbiat Modares University, Tehran, Iran

^dDepartment of Physics, Bilkent University, Ankara, Turkey

Abstract: DNA codons, consisting of triplet nucleotides (NTs), could play important roles for RNA transcription and protein translation in living systems. Therefore, their recognition could be seen important for diagnosis and therapy purposes. Based on triplet sequence formations of Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) NTs, 64 codons were investigated in this work regarding their complexation with a molecular cubane (CUB) wire. To achieve this aim, each of singular 64 codons and CUB were optimized to be prepared for docking processes of complex formations. Hence, 64 complexes of codon-CUB were docked to see the recognition potency of CUB wire versus each of DNA codons. Interestingly, the obtained docking scores indicated that the CUB could work specifically versus the DNA codons, in which G-rich and A-rich triples were seen to be more favorable for complexation with CUB in comparison with other C-rich and T-rich triplet codons. Moreover, the results indicated that not pure G triplet but GAG codon was the most favorable one to be recognized by the CUB wire. However, pure T triplet was the worst one for such complex formations. The results of this work remarkably indicated that the CUB wire could work for recognition process of DNA codons from each other and such recognition could be very much specified for each of G-rich and A-rich codons, in which GAG codon was the best one among all the 64 investigated codons.

Keywords: Codon; DNA; Triplet; Cubane; In silico.

1. Introduction

DNA discovery by the pioneering work of Watson and Crick in 1953 introduced new ways of investigating mysteries inside the biological systems opening molecular biology indeed [1]. Further works indicated that the cells of all living systems could share almost a common process to synthesize proteins based on encoded genes information [2]. In this process, genetic informative DNA is transcribed into RNA to be subsequently translated to proteins. DNA and RNA both include Adenine (A), Guanine (G) and Cytosine (C) nucleobases, in which Thymine (T) is included only in DNA and Uracil (U) is included only in RNA [3]. Macromolecular proteins are composed of long chains of amino acids (AAs), which are directly

built up by the transcribed RNA codons from the initial genetic DNA codons [3-7]. Each codon contains three nucleotides (NTs) in a joint triplet sequence, which corresponds to a specific AA [8]. Table 1 represents 64 DNA codons responsible for biosynthesizing proteins based on specific AAs translations. There are 20 AAs in living systems, in which each of them could be translated by one or up to six codons. The problems of processes of formations, transcriptions and translations are still unsolved for codons and the topic is an important subject of genetic research areas [9, 10]. In addition to positive benefits of such codons for building up the living systems, viruses formations and functions are negative points. Both of DNA and RNA viruses could synthesize proteins for destructive functions

¹ Corresponding Authors

e-mail: mdmirzaei@pharm.mui.ac.ir

in living systems, in which recognition or inhibition of their corresponding codons could be an important task of anti-virus developments [11]. New corona virus is an example of DNA virus with huge negative impact on human life worldwide [12]. Recognition of DNA codons sequences of such viral infection could work dual function of diagnosis and therapy in early stages [13]. This hypothesis is very much important at the molecular scale diagnosis and therapy, in which it will go further in future upon achievements of such research investigations.

Carbon atom has had always important roles in all field of science and technology, in which introduction of nanocarbon materials opened many more important features for this distinguished atom [14]. Various spherical shapes and electronic properties of nanocarbon materials encouraged researchers to investigate their potency for novel applications in different aspects, in which

investigating for the life systems has been almost the most important goal [15-19]. Cubane (CUB) is a cubic structure of carbon atom, in which it could work either as the initiator of other materials or as the single standing chemical structure [20]. Regarding the achievements of recent research works, formations of CUB related structures have been seen as important topics of further investigations [21, 22]. Moreover, its linear molecular scale formation has been also seen possible for working in single-standing structure [23]. Considering characteristic features, it could be expected that CUB could work to recognize other substances e.g., DNA codons. Based on such expectation, interactions of CUB with all 64 DNA codons were investigated in this work employing the in silico approach. The major goal of this work was to show the ability of CUB for DNA codons recognition at the molecular scale.

Table 1: List of DNA codons and their specified amino acids (AAs).*

1 st NT	2 nd NT								3 rd NT
	A		G		C		T		
	Codon	AA	Codon	AA	Codon	AA	Codon	AA	
A	AAA	Lys	AGA	Arg	ACA	Thr	ATA	Ile	A
	AAG		AGG		ACG		ATG	Met	G
	AAC	Asn	AGC	Ser	ACC		ATC	Ile	C
	AAT		AGT		ACT		ATT		T
G	GAA	Glu	GGA	Gly	GCA	Ala	GTA	Val	A
	GAG		GGG		GCG		GTG		G
	GAC	Asp	GGC		GCC		GTC		C
	GAT		GGT		GCT		GTT		T
C	CAA	Gln	CGA	Arg	CCA	Pro	CTA	Leu	A
	CAG		CGG		CCG		CTG		G
	CAC	His	CGC		CCC		CTC		C
	CAT		CGT		CCT		CTT		T
T	TAA	Stop	TGA	Stop	TCA	Ser	TTA	Leu	A
	TAG		TGG	Trp	TCG		TTG		G
	TAC	Tyr	TGC	Cys	TCC		TTC	Phe	C
	TAT		TGT		TCT		TTT		T

*Abbreviations of AAs: Alanine (Ala), Arginine (Arg), Asparagine (Asn), Aspartic acid (Asp), Cysteine (Cys), Glutamic acid (Glu), Glutamine (Gln), Glycine (Gly), Histidine (His), Isoleucine (Ile), Leucine (Leu), Lysine (Lys), Methionine (Met), Phenylalanine (Phe), Proline (Pro), Serine (Ser), Threonine (Thr), Tryptophan (Trp), Tyrosine (Tyr), Valine (Val).

2. In Silico Details

This work was done by employing in silico approach, which is known as a useful technique for molecular scale investigations [24]. All molecular models of DNA codons were drawn and optimized by the HyperChem program employing Amber molecular mechanics method [25]. In this step, optimized 3D models of DNA codons were prepared for further investigations (Table 1). Next,

the small molecular wire model of CUB (Fig. 1) was drawn by the GaussView program and it was optimized at the B3LYP/3-21G* level of density functional theory (DFT) as implemented in the Gaussian program [26]. Doing these steps prepared each of the codons to be the target structure and the CUB to be the ligand structure to be involved in the molecular docking processes. To perform docking processes, all individual targets were examined

regarding the potency of CUB ligand to make codon-CUB complexes. All the docking processes were performed at the HDock webservice [27]. The results of docking scores were summarized in Table 2 and the first rank codon-CUB complex was shown in Fig. 2. The graphical presentations of all other complexes were summarized in Fig. 4s of supplementary file. It is indeed an advantage of computer-based *in silico* works to investigate the structural systems at the lowest molecular scales to generate insightful information for their characteristic features [28-31].

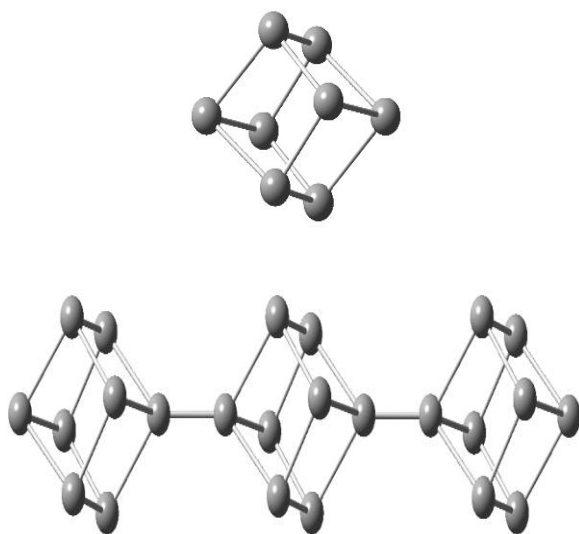


Fig. 1: Singular cubane (CUB) and wire.

3. Results and discussion

We investigated the potency of molecular CUB wire to make possible recognition of DNA codons from each other. First, we prepared the optimized structures of both of CUB and codons to be included in the next docking processes. The CUB wire was built up as covalently attached triplet as shown in Fig. 1. Indeed, it was expected to have a small ligand molecule for participating in interactions with the target molecule, DNA codons. The optimization process yielded the stabilized structure of CUB wire to be considered for the ligand role of docking processes. In the next step, the molecular models of DNA codons were drawn and optimized to yield the stabilized structures to be considered for the target role docking processes. All possibilities of 5'→3' formations of triplet NTs to make DNA codons were considered to have total

number of 64 codon models (Table 1). By preparing each of optimized singular ligand (CUB) and target (codon) counterparts, the docking processes were performed to examine codon-CUB complex formations. The obtained results of docking score (S) and complex rank (R) were summarized in Table 2. By values of S, the ranking of complexes were done to have R for comparing the favorability of codon-CUB complex formations. A quick look at the results of Table 2 could show that the complexes detected different interaction environments for the complex formations regarding the obtained different S values.

Moreover, it could be found that the NT building block of codon was very much important to obtain good value of S for complex formation. Indeed, it could be mentioned that the value of S could show somehow the strength of interaction, in which more negative value implied for higher strength of interactions. As could be seen by the representative GAG-CUB complex in Fig. 2 and other codon-CUB complexes in Fig. 4s of supplementary file, the interactions were all in non-covalent states but strong enough for complex formations. Further analyses of the results could indicate that A and G contained codons worked better than pure C and T ones, in which the values of S were better for the complexes of former ones than latter ones. However, pure A and G codons did not work very well in comparison with GAG, as the first rank codon-CUB complex formation (Fig. 2).

Interestingly, those other combinations of A and G were seen to be very much better than other complexes. For much more detailed examination of results, it could be found that the G contained codons in combination with each of A, C and T could work even better than all other combinations of NTs. The weakest DNA codon complexation with CUB was found for TTT with the lowest value of S and placing in the 64th R of codons. As could be concluded here, the G contained DNA codons were very much suitable to be recognized by the CUB wire with better S values in comparison with other types of codons for complexation with the CUB. The major problem of this work was to recognize the DNA codons by mean of the CUB wire, in which the illustrated results of Fig. 3 could approve the achievement of this work regarding the solution of mentioned problem. Indeed, it is very much important to recognize different codon

triplets based on their role for RNA transcription and protein translation.

Here with the results, it could be seen that the CUB could make problem mainly for G and A containing codons than other C and T codons. The trend could mean that, the recognition process worked for diagnosis first and therapy next. G-rich and A-rich sequences might be separated from those of other types by such recognizing processes, in which the -rich implies for formation of a codon, all or most of all, from unique type of NT. Moreover, advantages of recognition were seen for G-rich codons than A-rich codons. The codon itself is almost a semi-small molecule, in which it requires another small molecule for recognition process. The investigated molecular CUB wire was seen useful for such recognition process at the molecular scale.

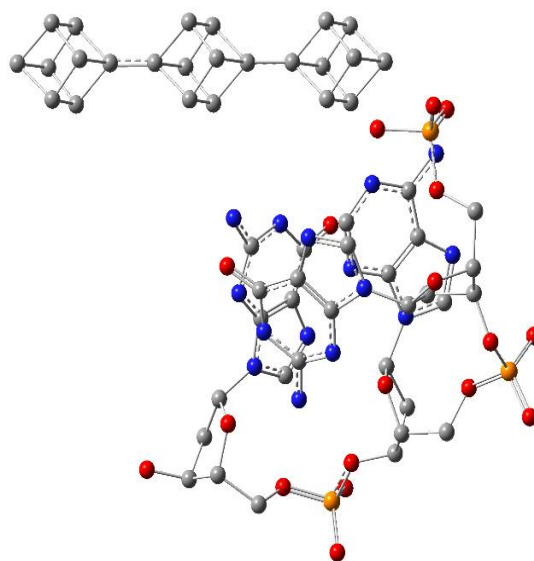


Figure 2. GAG-CUB complex, first rank.

Table 2: Codon-CUB complexes scores (S) and ranks (R).

			CUB								
Codon	S	R	Codon	S	R	Codon	S	R	Codon	S	R
AAA	-35.04	5	AGA	-34.77	8	ACA	-32.37	28	ATA	-32.28	30
AAG	-35.16	4	AGG	-34.85	7	ACG	-35.69	2	ATG	-32.68	22
AAC	-33.23	15	AGC	-33.09	17	ACC	-30.66	43	ATC	-30.27	49
AAT	-33.03	19	AGT	-32.62	23	ACT	-31.04	39	ATT	-30.90	40
GAA	-32.61	25	GGA	-32.80	20	GCA	-31.36	35	GTA	-30.42	47
GAG	-36.70	1*	GGG	-35.50	3	GCG	-32.70	21	GTG	-33.04	18
GAC	-31.91	31	GGC	-33.14	16	GCC	-30.85	41	GTC	-29.95	50
GAT	-31.38	34	GGT	-30.63	44	GCT	-29.17	54	GTT	-32.62	24
CAA	-34.58	9	CGA	-33.59	14	CCA	-30.39	48	CTA	-29.88	52
CAG	-34.90	6	CGG	-33.79	13	CCG	-32.58	27	CTG	-27.48	58
CAC	-31.17	36	CGC	-31.14	37	CCC	-30.48	46	CTC	-26.75	60
CAT	-31.51	32	CGT	-31.40	33	CCT	-27.76	56	CTT	-27.41	59
TAA	-33.90	12	TGA	-32.61	26	TCA	-30.61	45	TTA	-26.71	61
TAG	-34.42	10	TGG	-32.32	29	TCG	-33.99	11	TTG	-28.30	55
TAC	-31.09	38	TGC	-29.92	51	TCC	-27.63	57	TTC	-24.98	63
TAT	-30.77	42	TGT	-29.40	53	TCT	-25.88	62	TTT	-24.69	64

*See Fig. 2 for graphical representation of this complex. For other complexes, see Fig. 4s of supplementary file.

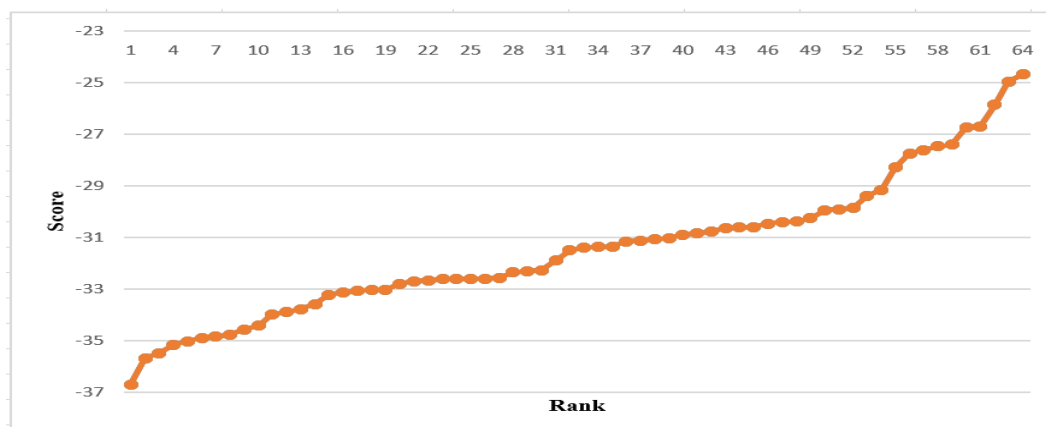


Figure 3. Score vs. Rank for the codon-CUB complexes.

4. Conclusion

This work was done by employing in silico approach to investigate recognition process of DNA codons by a molecular CUB wire. The singular molecules of CUB and each of 64 DNA codons were optimized first and their codon-CUB complexation through docking processes were investigated next. The remarkable achievements based on the obtained results could be summarized in the following notes. First, the possibility of formation of singular CUB and codon were achieved. Second, different modes of codon-CUB complexes were found based on their docking scores. Third, both of G-rich and A-rich triplet codons worked better than C-rich and T-rich ones in codon-CUB complex formations. Fourth, the G-rich codon, not pure G triplet but GAG, was seen to be the best codon for complexation with the CUB in comparison with other ones. Fifth, pure T triplet was the worst codon for complexation with the CUB in comparison with other ones. And finally, the investigated CUB wire could be proposed for recognition of DNA codons based on the obtained different values of docking scores.

Acknowledgements

The support of this work by the research council of Isfahan University of Medical Sciences under grant number 294218 is acknowledged.

Conflict of Interests

There is no conflict of interests for the authors.

References

- [1] Watson JD, Crick FH. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*. 1953;171:737-8.
- [2] Shu JJ. A new integrated symmetrical table for genetic codes. *Biosystems*. 2017;151:21-6.
- [3] Tan SC, Yiap BC. DNA, RNA, and protein extraction: the past and the present. *Journal of Biomedicine and Biotechnology*. 2009;2009:574398.
- [4] Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C. A new and updated resource for codon usage tables. *BMC Bioinformatics*. 2017;18:1-0.
- [5] Simmons MP. Relative benefits of amino-acid, codon, degeneracy, DNA, and purine-pyrimidine character coding for phylogenetic analyses of exons. *Journal of Systematics and Evolution*. 2017;55:85-109.
- [6] Stein CK, Pawlyn C, Chavan S, Rasche L, Weinhold N, Corken A, Buros A, Sonneveld P, Jackson GH, Landgren O, Mughal T. The varied distribution and impact of RAS codon and other key DNA alterations across the translocation cyclin D subgroups in multiple myeloma. *Oncotarget*. 2017;8:27854.

- [7] Rapino F, Delaunay S, Rambow F, Zhou Z, Tharun L, De Tullio P, Sin O, Shostak K, Schmitz S, Piepers J, Ghesquière B. Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature*. 2018;558:605-9.
- [8] Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. *Nature*. 1961;192:1227-32.
- [9] Karlin S, Mrázek J. What drives codon choices in human genes?. *Journal of Molecular Biology*. 1996;262:459-72.
- [10] Alexaki A, Kames J, Holcomb DD, Athey J, Santana-Quintero LV, Lam PV, Hamasaki-Katagiri N, Osipova E, Simonyan V, Bar H, Komar AA. Codon and codon-pair usage tables (CoCoPUTs): facilitating genetic variation analyses and recombinant gene design. *Journal of Molecular Biology*. 2019;431:2434-41.
- [11] Groenke N, Trimpert J, Merz S, Conradie AM, Wyler E, Zhang H, Hazapis OG, Rausch S, Landthaler M, Osterrieder N, Kunec D. Mechanism of virus attenuation by codon pair deoptimization. *Cell Reports*. 2020;31:107586.
- [12] Dilucca M, Forcelloni S, Georgakilas AG, Giansanti A, Pavlopoulou A. Codon usage and phenotypic divergences of SARS-CoV-2 genes. *Viruses*. 2020;12:498.
- [13] Lopes A, Vanvarenberg K, Préat V, Vandermeulen G. Codon-optimized P1A-encoding DNA vaccine: toward a therapeutic vaccination against P815 mastocytoma. *Molecular Therapy-Nucleic Acids*. 2017;8:404-15.
- [14] Iijima S. Carbon nanotubes: past, present, and future. *Physica B*. 2002;323:1-5.
- [15] Mirzaei M, Kalhor HR, Hadipour NL. Covalent hybridization of CNT by thymine and uracil: A computational study. *Journal of Molecular Modeling*. 2011;17:695-9.
- [16] Mirzaei M, Yousefi M. Computational studies of the purine-functionalized graphene sheets. *Superlattices and Microstructures*. 2012;52:612-7.
- [17] Pagar T, Ghotekar S, Pansambal S, Pagar K, Oza R. Biomimetic synthesis of CuO nanoparticle using *Capparis decidua* and their antibacterial activity. *Advanced Journal of Science and Engineering*. 2020;1:133-7.
- [18] Faramarzi R, Falahati M, Mirzaei M. Interactions of fluorouracil by CNT and BNNT: DFT analyses. *Advanced Journal of Science and Engineering*. 2020;1:62-6.
- [19] Mirzaei M. Effects of carbon nanotubes on properties of the fluorouracil anticancer drug: DFT studies of a CNT-fluorouracil compound. *International Journal of Nano Dimension*. 2013;3:175-9.
- [20] Eaton PE, Cole TW. Cubane. *Journal of the American Chemical Society*. 1964;86:3157-8.
- [21] Biegasiewicz KF, Griffiths JR, Savage GP, Tsanaktsidis J, Priefer R. Cubane: 50 years later. *Chemical Reviews*. 2015;115:6719-45.
- [22] Bernhard SS, Locke GM, Plunkett S, Meindl A, Flanagan KJ, Senge MO. Cubane cross-coupling and cubane-porphyrin arrays. *Chemistry - A European Journal*. 2018;24:1026-30.
- [23] Deville C, Folkjær M, Reinholdt P, Hvid MS, Lamagni P, Borup K, Sun Z, Lauritsen JV, McKee V, Jensen KM, Lock N. Cubes on a string: A series of linear coordination polymers with cubane-like nodes and dicarboxylate linkers. *Nanoscale*. 2020;12:11601-11
- [24] Mirzaei M. Science and engineering in silico. *Advanced Journal of Science and Engineering*. 2020;1:1-2.

- [25] Froimowitz M. HyperChem: a software package for computational chemistry and molecular modeling. *Biotechniques*. 1993;14:1010-3.
- [26] Frisch MJ et al., Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT, 2013.
- [27] Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*. 2017;45:365-73.
- [28] Behzadi H, Hadipour NL, Mirzaei M. A density functional study of ^{17}O , ^{14}N and ^2H electric field gradient tensors in the real crystalline structure of α -glycine. *Biophysical Chemistry*. 2007;125:179-83.
- [29] Mirzaei M, Hadipour NL, Abolhassani MR. Influence of C-doping on the B-11 and N-14 quadrupole coupling constants in boron-nitride nanotubes: A DFT study. *Zeitschrift für Naturforschung A*. 2007;62:56-60.
- [30] Samadi Z, Mirzaei M, Hadipour NL, Khorami SA. Density functional calculations of oxygen, nitrogen and hydrogen electric field gradient and chemical shielding tensors to study hydrogen bonding properties of peptide group (OC-NH) in crystalline acetamide. *Journal of Molecular Graphics and Modelling*. 2008;26:977-81.
- [31] Farahbakhsh Z, Zamani MR, Rafienia M, Gulseren O, Mirzaei M. In silico activity of AS1411 aptamer against nucleolin of cancer cells. *Iranian Journal of Blood and Cancer*. 2020;12:95-10