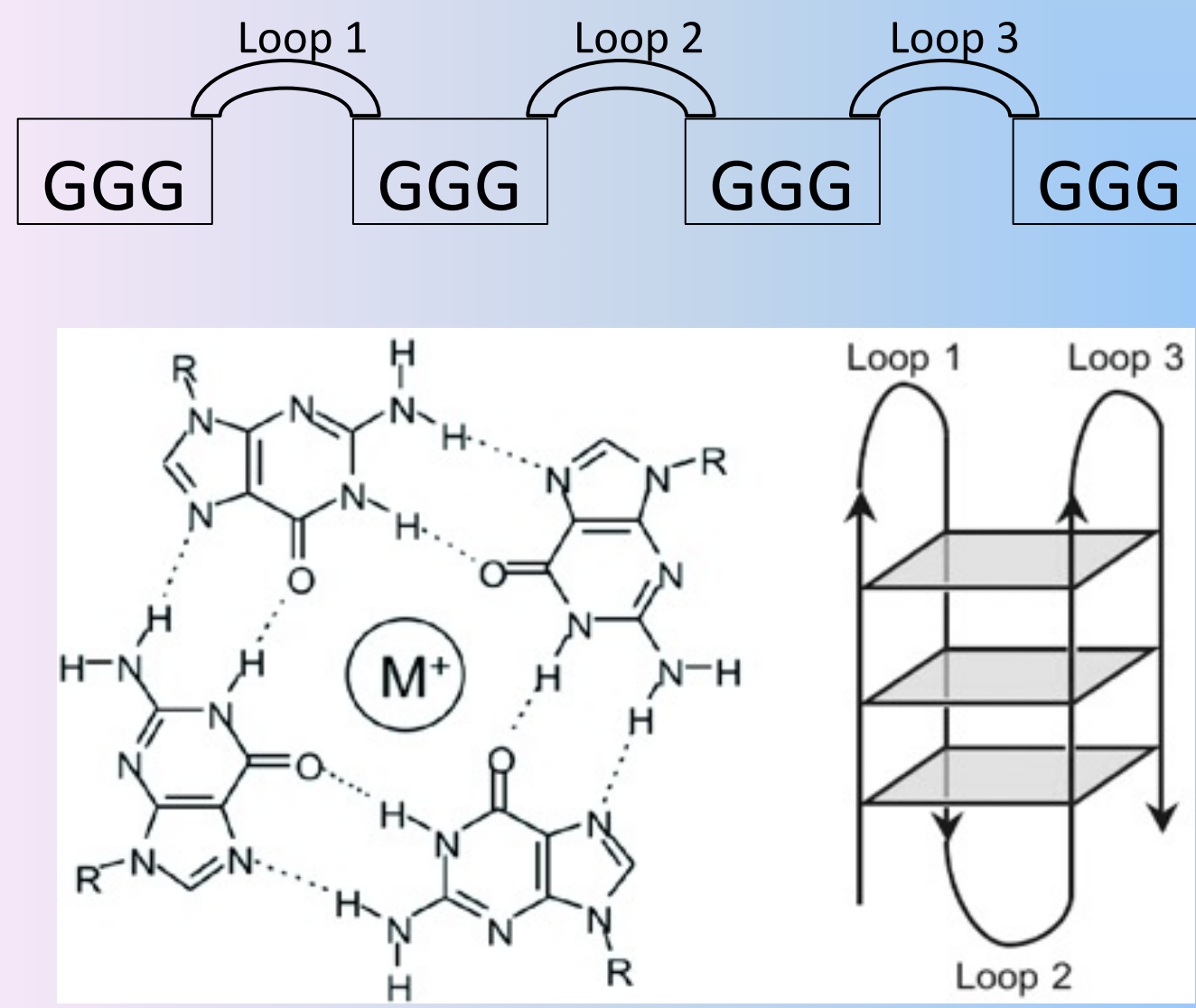


ABSTRACT

Non-canonical forms of DNA like guanine quadruplex (G4) elements play important roles in regulating transcription and translation through interactions with their protein partners. Nucleoside Diphosphate Kinases (NDPKs) are multifunctional enzymes, some of them have been identified as G4-binding proteins, such as human NM23-H2 and maize ZmNDPK1, settling down an evolutionary conservation of this NDPKs feature. TcNDPK1 is the *Trypanosoma cruzi* homolog of NM23-H2 and ZmNDPK1 and different studies from our group associate TcNDPK1 with nuclear functions, for instance, *in vitro* nuclease and DNA-binding properties and, more recently, participation in DNA damage response machinery probably through the regulation of DNA repair enzymes expression. Very little is known about G4 elements in trypanosomes. In this work, considering that they might function as regulatory elements or as binding sites for NDPKs enzymes, an *in silico* search to identify putative G4 elements in the genome of *T. cruzi* was initiated. Homologous chromosomes Dm28c_6, Chr30P and TCC_10 from DM28c, CL Brener and TCC strains respectively, were analysed using the online available algorithms G4Hunter and QGRS Mapper. 49, 41 and 47 canonical G4 elements were identified by both programmes for each chromosome respectively, which were distributed heterogeneously all along the chromosomes. Interestingly, more than 90 % of the G4 elements localize in the coding strand, about 65% are intergenic and 48% are in UTR regions. In addition, 80% are conserved among the three chromosomes. These results predict that G4 elements are present in the genome of *T. cruzi* and suggest that they may act as regulatory elements for gene expression where different proteins, e.g. TcNDPK1, could bind and exert its functions. Further investigations are needed to confirm these predictions.

INTRODUCTION

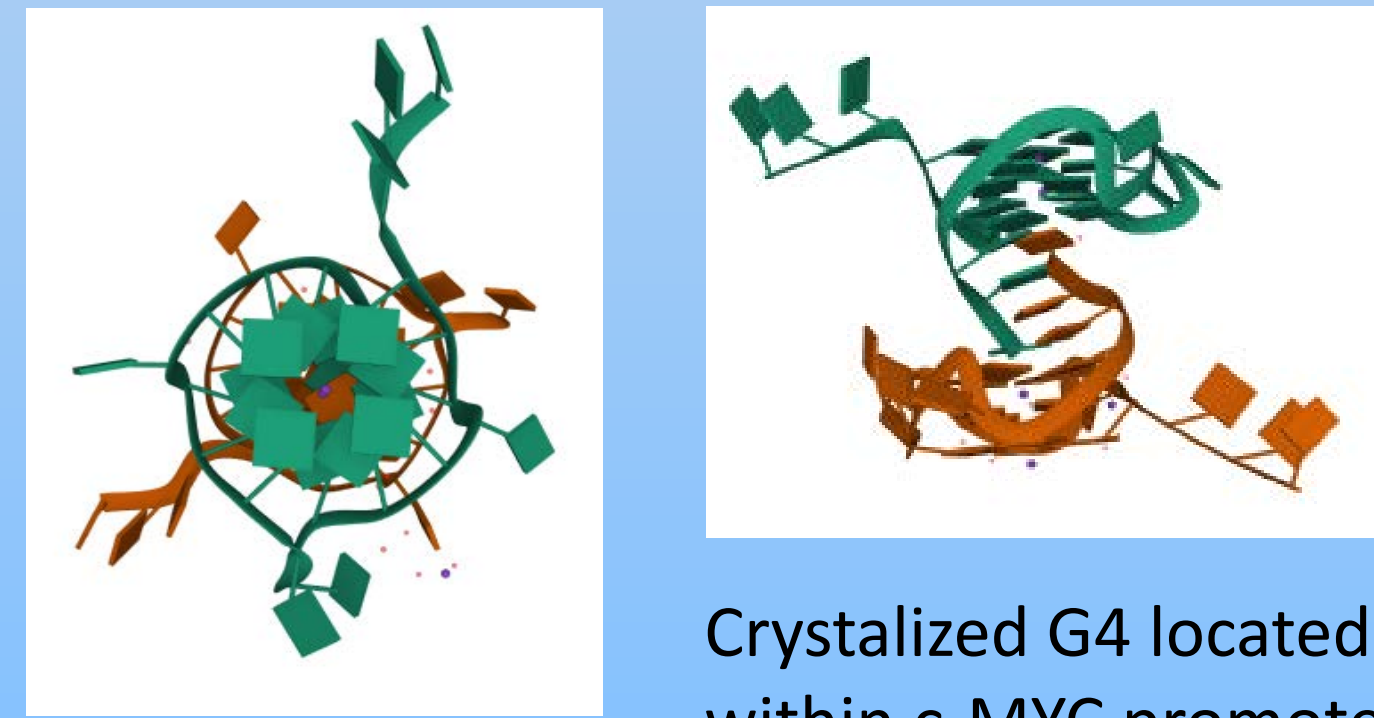
Picture of a representative canonical G4



Source: Wikipedia, Huppert 2005

G-quadruplex (G4) elements are secondary nucleic acid structures formed by stacked guanine tetrads joined through Hoogsteen type base pairing.

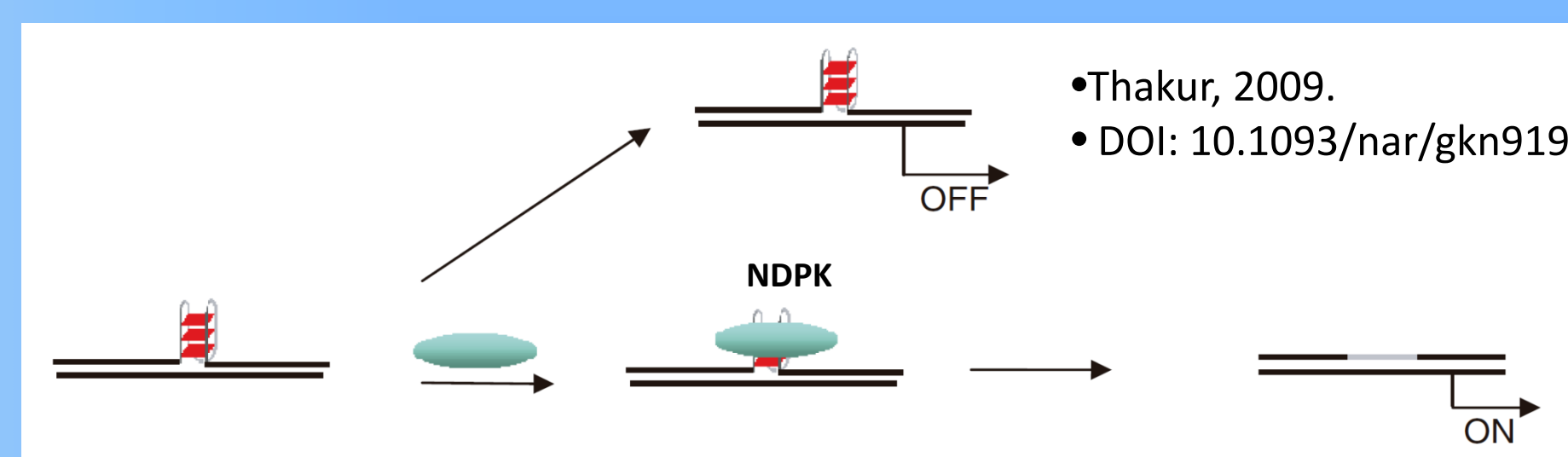
G4 of c-MYC promoter



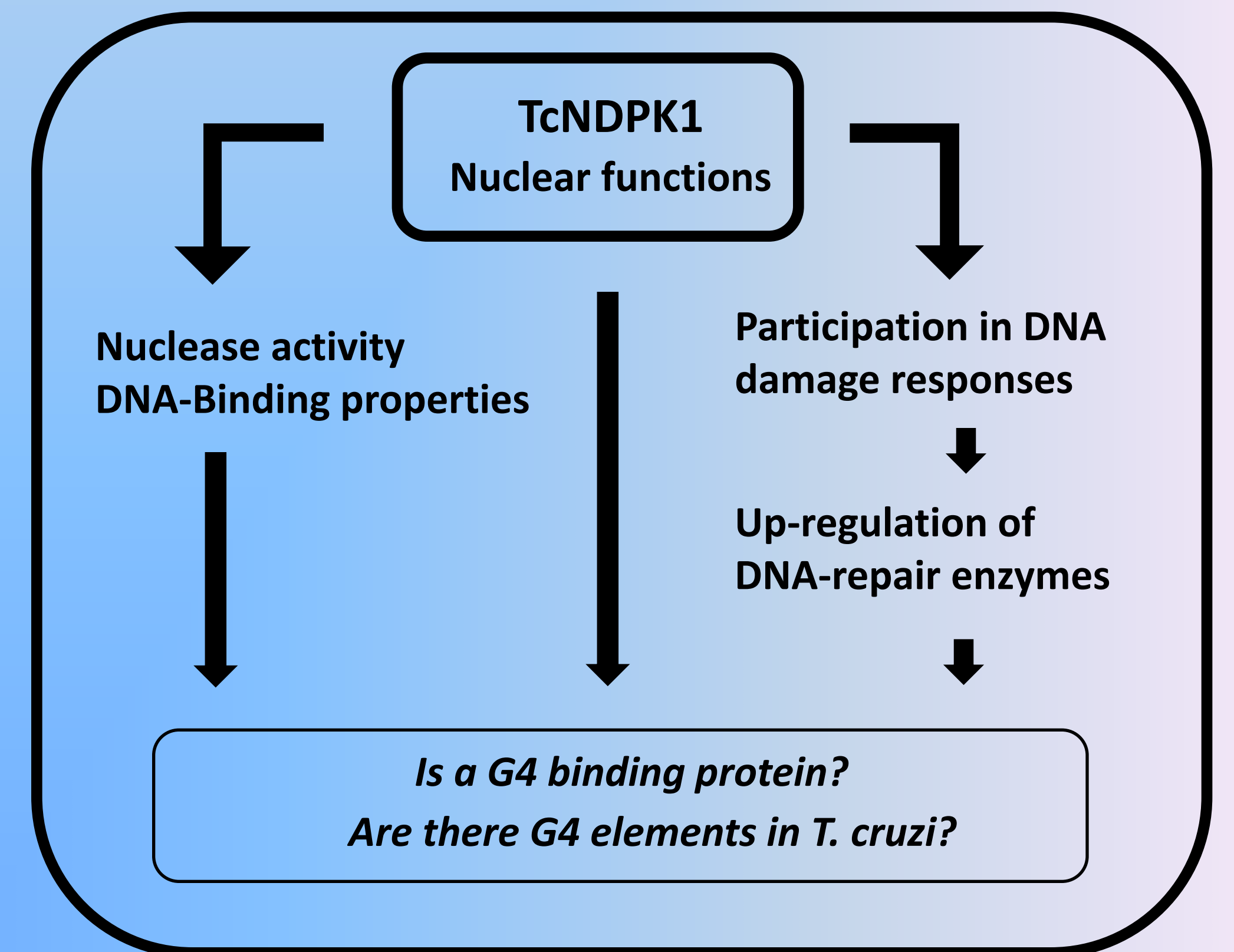
Crystallized G4 located within c-MYC promoter

5' TGAGGGTGGGTAGGGTGGGTAA 3' PDB: 6AU4

Proposed mechanism for NDPK mediated gene regulation



Trypanosoma cruzi



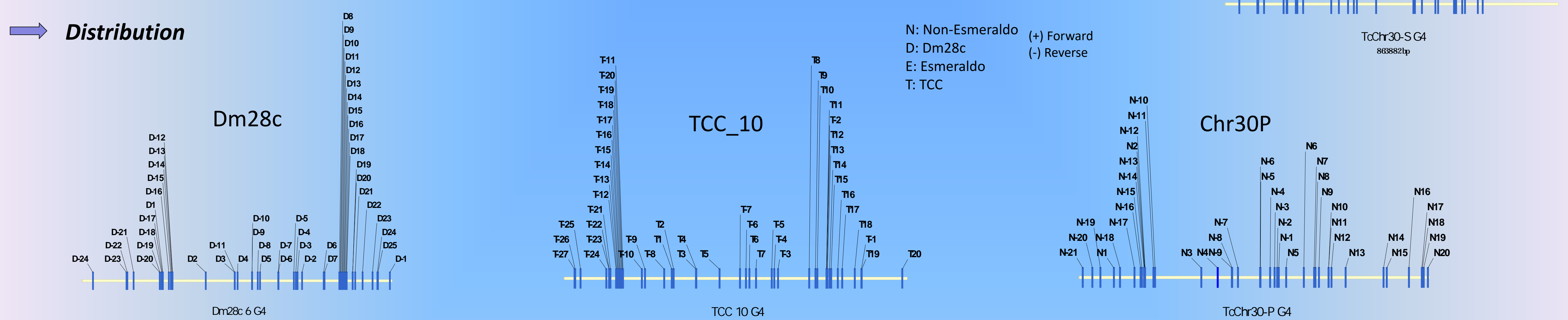
RESULTS

Genome/Chromosome	pb	%GC	QGRS Mapper		G4Hunter							
			Lengh/minG/loop	#G4	Window/Threshold							
					30/3/3-10	25/1.4	30/1.4	25/1.5	25/1.75			
#G4	#overlapped	#G4	Frequency	#G4	Frequency	#G4	Frequency	#G4	Frequency			
Dm28c	53273765	51,6	—	—	43522	0,8 / 1000 bp	24905	0,5 / 1000 bp	34021	0,6 / 1000 bp	12756	0,2 / 1000 bp
Dm28c_6	814295	51,3	49	578	926	1,1 / 1000 bp	4810	0,6 / 1000 bp	611	0,8 / 1000 bp	245	0,3 / 1000 bp
TCC	87058484	51,7	—	—	67028	0,8 / 1000 bp	38280	0,4 / 1000 bp	49700	0,6 / 1000 bp	19339	0,2 / 1000 bp
TCC_10	809670	51,4	47	644	702	0,9 / 1000 bp	4210	0,5 / 1000 bp	532	0,7 / 1000 bp	206	0,3 / 1000 bp
CL Brener Non-Esmeraldo	32529072	50,7	—	—	23109	0,7 / 1000 bp	13353	0,4 / 1000 bp	17973	0,6 / 1000 bp	7090	0,2 / 1000 bp
Chr30P	863882	50,9	41	157	668	0,8 / 1000 bp	3880	0,4 / 1000 bp	498	0,6 / 1000 bp	219	0,3 / 1000 bp
CL Brener Esmeraldo	32529070	50,4	—	—	21441	0,7 / 1000 bp	12057	0,4 / 1000 bp	16972	0,5 / 1000 bp	6560	0,2 / 1000 bp
Chr30S	863882	—	31	369	—	—	—	—	—	—	—	—

Summary of putative G4 elements obtained with QGRS Mapper and G4Hunter web interface softwares applied to Dm28c, TCC and CL Brener genomes and the homologues chromosomes Dm28c_6, TCC_10 and Chr30P/S. Due to limits on the length of the sequences to be analysed, QGRS Mapper was restricted only to the homologous chromosomes.

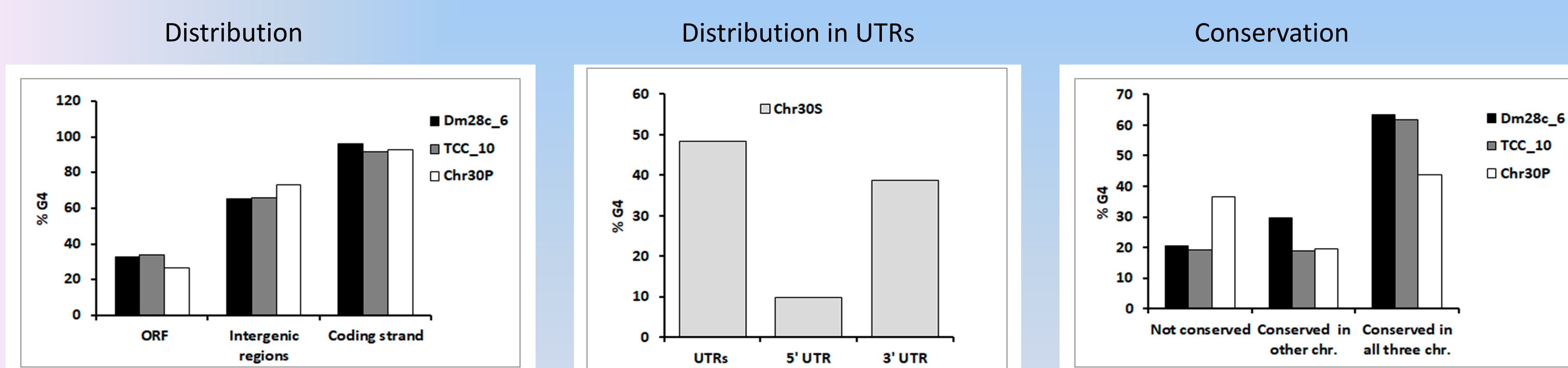
ANALYSIS OF CANONICAL NON-OVERLAPPED G4 ELEMENTS

Distribution



Canonical not overlapped G4 elements retrieved by QGRS Mapper were identified and located in each chromosome, showing that G4 are heterogeneously distributed with higher frequencies near the ends.

Features



Canonical not overlapped G4 elements are mostly located in the coding strands (about 90%) and in intergenic regions (about 65%). Additionally, an analysis made on CL Brener Esmeraldo Chr30S UTRs (information available only for this strain and haplotype), revealed that 48% of the G4 are located in the UTRs but mostly in the 3'UTRs. Finally, the majority of the canonical G4 are conserved in the three chromosomes, about 20-35% of the G4 are restricted only to each chromosome.

DISCUSSION

These results predict that G4 elements are present in the genome of *T. cruzi* and conserved among different *T. cruzi* strains belonging to different DTUs. This suggests that G4 elements may act as regulatory elements for gene expression where different proteins, e.g. TcNDPK1, could bind and exert its functions. The present study constitutes an initial survey of putative G4 elements in *T. cruzi*, an interesting research field since trypanosomatids possess singular gene expression/regulation machinery, different from other eukaryotes. Further and deeper investigations are needed to confirm our predictions.

FUNDING ANPCyT: PICT 2012-0559, PICT 2018-01871, PICT 2018-000585, PICT 2019-02622, PICT 2018-1801
Global Challenges Research Fund (GCRF), Grant: MR/P027989/1.