CONFIRMATION OF PERFECT G-QUADRUPLEXES IN MYCOBACTERIA TUBERCULOSIS BY ALPHAFOLD 3 ARTIFICIAL INTELLIGENCE

Limanskaya O.Yu.[1], Balak O.K.[2], Limanskii A.P.[3]

[1]NSC "Institute of Experimental and Clinical Veterinary Medicine" of NAAS of Ukraine, Kharkiv, Ukraine; [2]Kharkiv National Medical University, Kharkiv, Ukraine; [3]Institute of Physiological Active Compounds, Kharkiv, Ukraine.

The problem

G-quadruplexes (G4s) are non-canonical structures formed in guanidine-rich sequences of nucleic acids, stabilized by stacking interactions of planar G-tetrads and monovalent metal ions.

Until now, only experimental methods have been used to confirm structure of such alternative elements.

Introduction

The goal of the present study was the determination those G4s from a number of putative perfect G-quadruplexes formed by four G-quartets in the sense strand of *Mycobacterium tuberculosis* (MTB) H37Rv DNA, the 3D models of which were confirmed by the AlphaFold 3 artificial intelligence.

Methods

QGRS Mapper software was used for searching motifs of G4s. 3D models of G4s were built by the AlphaFold 3 (J. Abramson et al, 2024) for 10 K+ ions. For a reliability of the 3D model prediction in AlphaFold 3 the pIDDT parameter was used. A pIDDT value for a high level of model confidence should be over 90 on a scale from 0 to 100.

Results

11 conserved putative G-quadruplexes of four G-tetrads were found in the DNA sense strand of MTB H37Rv isolate with complete genome (Table 1).

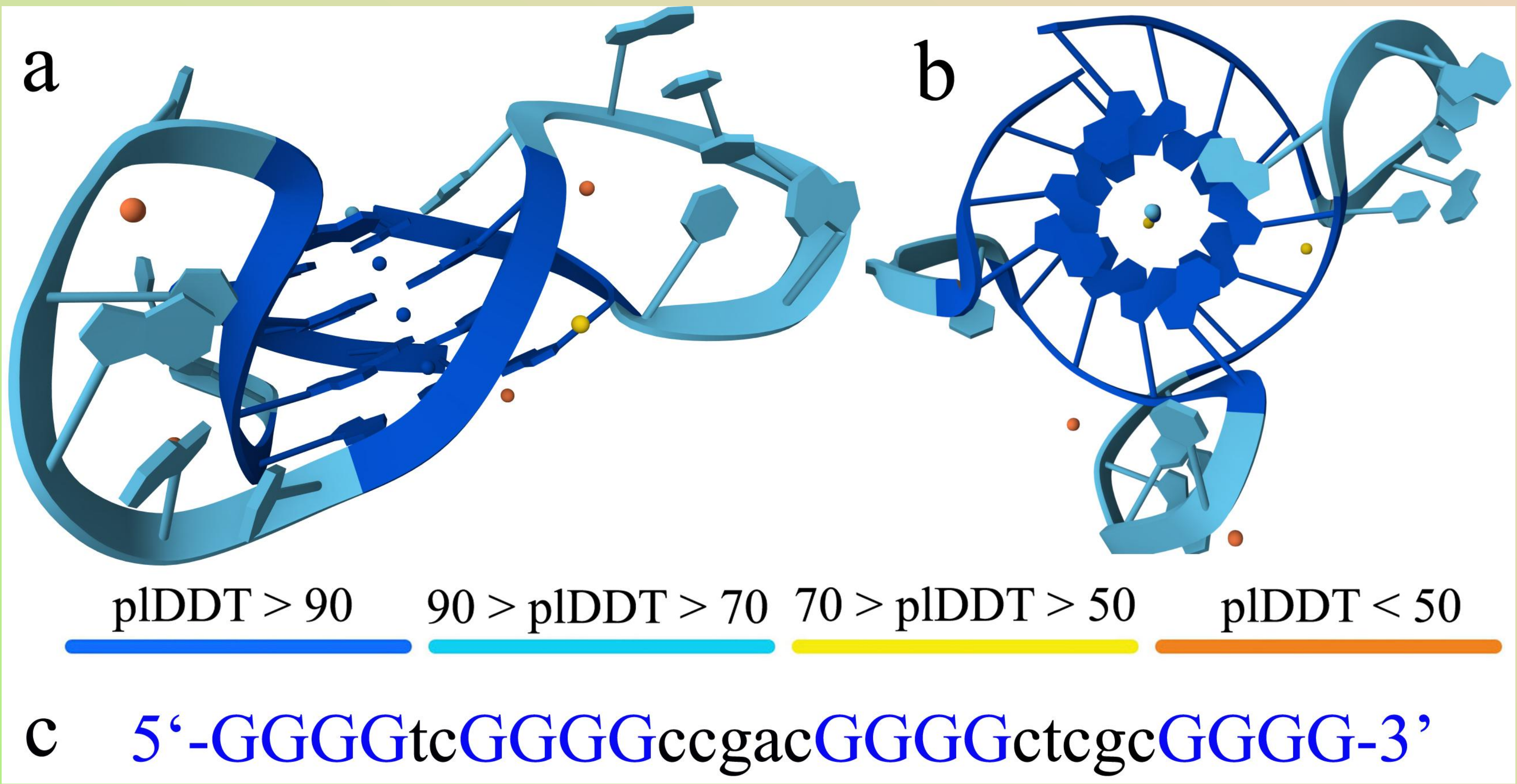

Fig. 2. G-quadruplex (pIDDT>90) is formed by four G-tetrads at position 1091957-1091984 of the *rpsl* gene of DNA sense strand of the *Mycobacterium tuberculosis* H37Rv isolate: side view (a) and top view (b); conservative motif of the putative G-quadruplex (c).

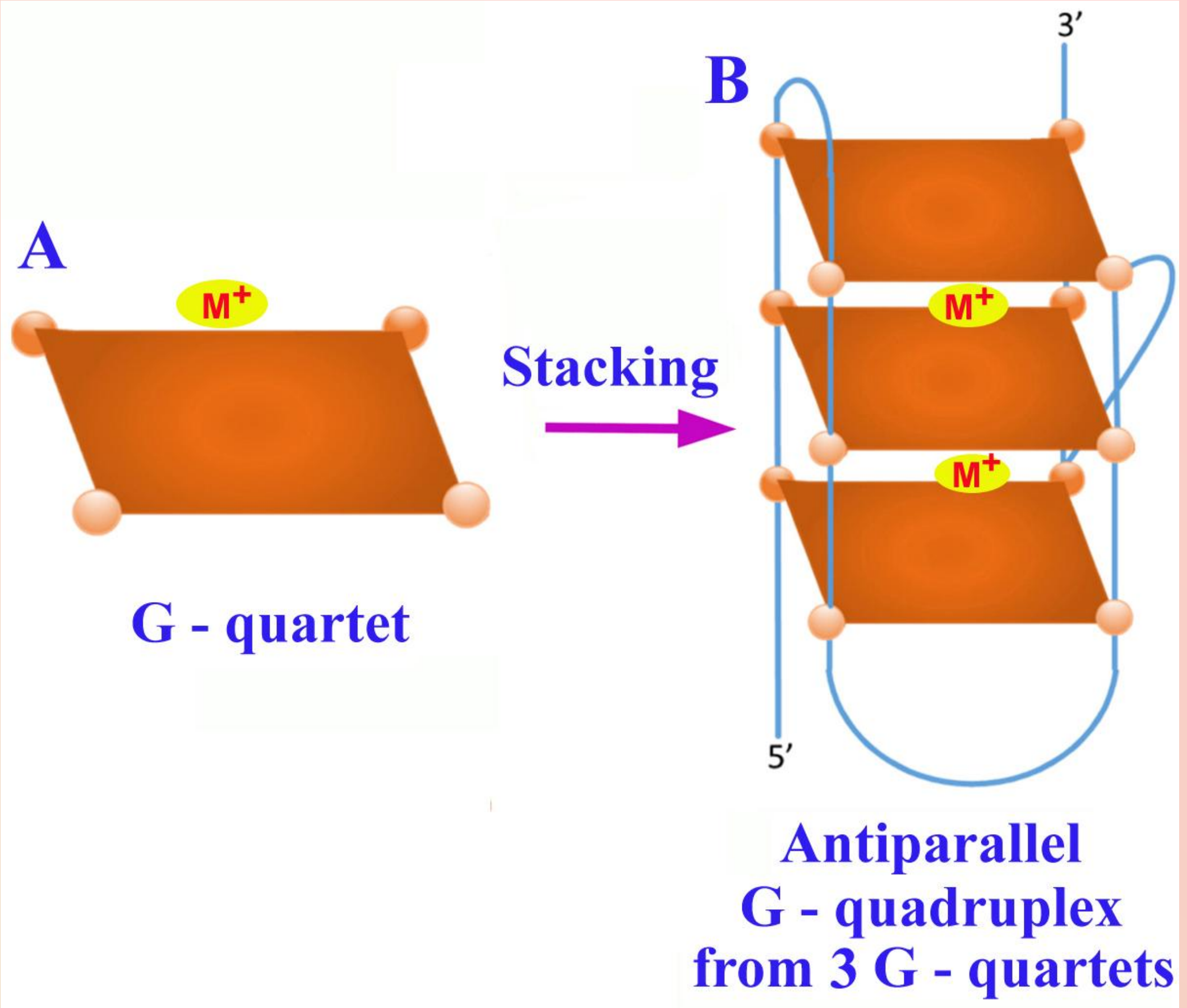

Fig. 1. Intramolecular antiparallel G-quadruplex structure (B) formed by stacking of two three G-quartets (A) and stabilized by monovalent cations.

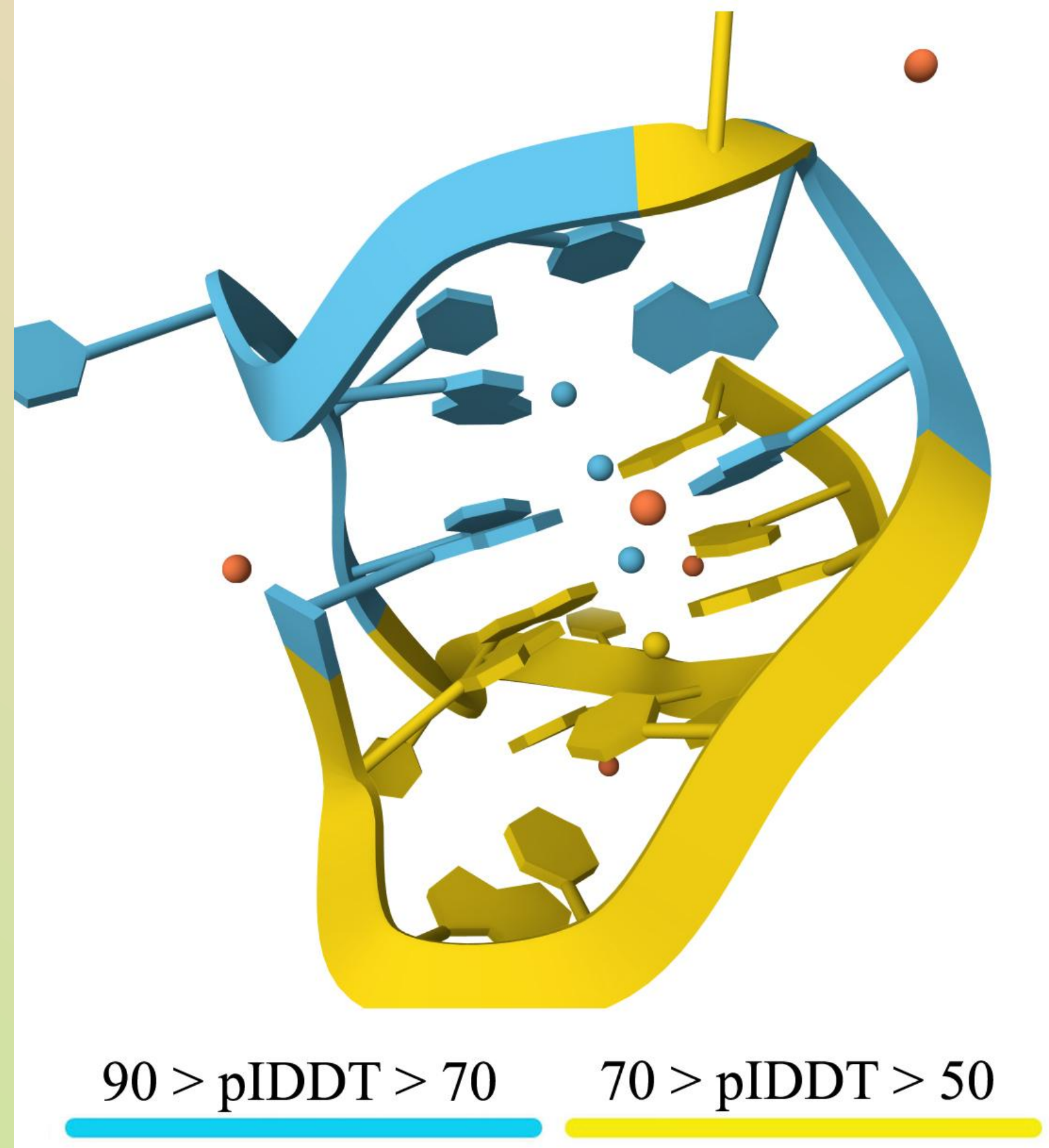

Fig. 3. G-quadruplex (70 > pIDDT > 50) in the *mexC* gene of the gram-positive bacterium *Pseudomonas aeruginosa* was used as a control of the confidence of AlphaFold 3 models. This G4 was experimentally confirmed by NMR spectroscopy (Evans et al, 2023).

Table 1. Conservative sequences of putative G-quadruplexes in DNA sense strand of the *Mycobacterium tuberculosis* H37Rv isolate. G-rich fragments of sequences for G4s are underlined.

Position	G-quadruplex motif	G-score	pIDDT	Number of MTB isolates with 100% identity by BLAST
86591	GGGGCCCGCAAGCTGTGGGGGTTACGGGGTTATGCGGAGGTGGGG	101	70 > pIDDT > 50	727
149893	GGGGCGCCGGGGACTGGGCAGGCCGGCGGGGCTGGGGGG	97	70 > pIDDT > 50	Not found by BLAST
308373	GGGGCTGGGGCTGGAGTTGGGGTGGGGG	102	70 > pIDDT > 50	729
410379	GGGGTCGGCGGGGCAACCGCGGCCGGGGTCGGCGGGG	102	pIDDT < 50,	709
833353	GGGGCGCCGGGGACCGGGGCCGCCGGCGGGG	103	90 > pIDDT > 70	648
926930	GGGGCTGGCGGGGAGGGCGGGGCCGGCGGGG	108	70 > pIDDT > 50	507
1091957	GGGGTCGGGGCCGACGGGGGTCGCGGGG (Fig. 2)	105	pIDDT > 90	629
3861910	GGGGCCGCCGGGGTTTGCGGGGCTGGGG	105	pIDDT > 90	632
3928213	GGGGCCGGCGGGGCCGGCGGGGCCGGCGGGG	108	70 > pIDDT > 50	Not found by BLAST
4149169	GGGGCCCATGCCTGGGGGAGGGGGTCCGCCGGGG	102	90 > pIDDT > 70	719
4359136	GGGGTTCCCGGGGTGATCGGGGTTCCCGGGG	108	70 > pIDDT > 50	696

Conclusion

AlphaFold 3 is a powerfull technique for determining biopolymer structure with precise accuracy of experimental methods.