An Automated Pipeline for 3D Structural Modelling of DNA Probes for Pathogen Detection

Beatriz Pereira[1,2,3], Romeu Fernandes[4], Pedro Soares[5,6], Diogo Pratas[7,8,9,10], Sérgio F. Sousa[11],João Carneiro[3]

CATÓLICA
CBQF · CENTRE FOR BIOTECHNOLOGY AND FINE CHEMISTRY ASSOCIATE LABORATORY CBQF
PORTO

[1]Department of Computer Science, Faculty of Sciences, University of Porto, Porto, Portugal; [2] Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal; [3]Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Rua Diogo Botelho 1327, 4169-005 Porto, Portugal; [4]Informatics Department, University of Minho, 4710-057 Braga, Portugal; [5]Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Braga, Portugal; [6]Institute of Science and Innovation for Bio-Sustainability, University of Minho, Braga, Portugal; [7]Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal; [8]LASI - Intelligent System Associate Laboratory, Portugal; [9]Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal; [10]Department of Virology, University of Helsinki, Helsinki, Finland; [11]LAQV/REQUIMTEBioSIM - Department of Biomedicine, Faculty of Medicine, University of Porto, Porto, Portugal;

Background - Why Structural Probe Modelling?

Respiratory and systemic pathogens remain a major global health challenge. DNA probes offer a **rapid** and sequence **specific** diagnostic alternative, but their performance depends strongly on probe **three-dimensional** structure.

Pathogen Detection — Virus and Bacteria → **Diagnostic Methods** — → Slow → **Antimicrobial Resistance** → **SOLUTION:** **DNA Aptamers** — Rapid and Specific

Stable stem-loop or hairpin conformations can hide the binding region and reduce target accessibility. Accurate 3D modelling is therefore essential for rational probe design and *in silico* optimization before experimental validation.

Objectives

› Develop an automated pipeline for 3D structure prediction of ssDNA probes targeting **bacterial** and **viral** pathogens;

› Predict stable secondary structures;

› Generate 3D conformations via knowledge-based fragment assembly (NSP toolkit);

› Perform conformational **clustering** and **energy minimization** to produce representative, refined 3D models;

› Enable future **DNA:DNA docking** studies using HADDOCK3 against genomic targets.

Methodology - Workflow

01 Input — Probe sequence — Excel / Csv

02 Secondary Str. — RNAFold + DNA params — Mathews 2004 MFE
- Dot-bracket notation
- (((...))) stems
- → .ct connectivity file

03 3D Generation — NSP toolkit — Fragment assembly x5
- Coarse-grained PDB
- Fragment library match
- → 5 conformations

04 Clustering — Conformational grouping — Representative structures
- RMSD-based grouping
- Structural variability
- → centroid per cluster

05 Energy Min. — All-atom reconstruction — Amber OL15 force field
- LEaP all-atom H-add
- Sander minimization
- → ranked by Epot

***_minenerg.pdb** — Lowest-energy 3D structure

final_structures/[1-5].pdb — All refined conformations

quality_report.xlsx — Structural quality metrics

[Future] HADDOCK3 DNA:DNA docking — ssDNA probe vs genomic target

Results - 3D Structure Predictions

A test sample consisting of five probes for five species was used, based on previous studies.

Target Pathogen	MFE kcal/mol	R. Gyration (Å)	RMSD spread (Å)	Clash score
S.aureus	0,0	21,56	43,868	0,696
HN1	-0,2	17,90	11,664	9,142
A. fumigatus	-1,0	27,33	0,903	8,040
H. Influenzae	-0,3	13,11	3,391	38,606
P. aeruginosa	-3,1	14,10	130,498	29,924

The variability observed in the higher RMSD values reflects the sampling of distinct yet energetically accessible conformational states, rather than structural instability, as the molecule may adopt multiple stable folds in solution.

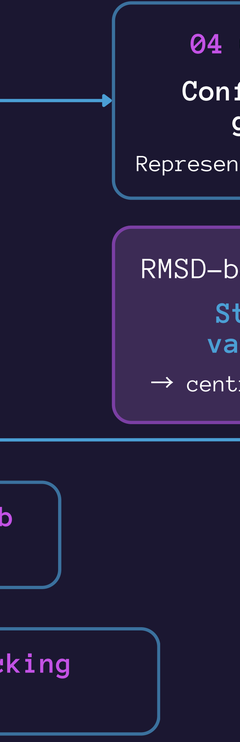

Fig.1 - Aspergilus fumigatus probe minenergy 3D prediction

MFE (kcal/mol)
2D structure stability. more negative = more stable

RMSD spread (Å)
Max pairwise RMSD across 5 predictions. Measures sampling convergence and prediction confidence. ideal: < 3 Å

Radius Gyration (Å)
Mass-weighted RMSD of all atoms from the centre of mass, measures compactness. higher → extended; lower → compact

Clash Score
Steric overlaps per 1000 atoms. High values indicate unresolved atomic conflicts in the model. ideal: < 20

Conclusions

✓ The pipeline successfully generates realistic, energy minimized 3D conformations of ssDNA probes from sequence input;

✓ Five independent predictions per probe, followed by clustering and energy ranking, ensure sampling of the conformational diversity;

✓ The descriptors (MFE, radius gyration, RMSD spread, and clash score) constitute a structured feature set that reflects probe foldability and reliability, providing a foundation for probe selection;

→ Final probe structures will be used as input for *ab initio* DNA:DNA docking against genomic targets.

Acknowledgements

This work is carried out within the scope of contract ref. 2023.15056.TENURE.043, funded by national funds through the FCT - Foundation for Science and Technology, I.P. This work is also funded by national funds through the FCT - Foundation for Science and Technology within the scope of UID/50016/2025 and LA/P/0076/2020 (https://doi.org/10.54499/LA/P/0076/2020).

fct Fundação para a Ciência e a Tecnologia · ciimar · IJUP · U.PORTO · UP Investigação Jovem da Universidade do Porto