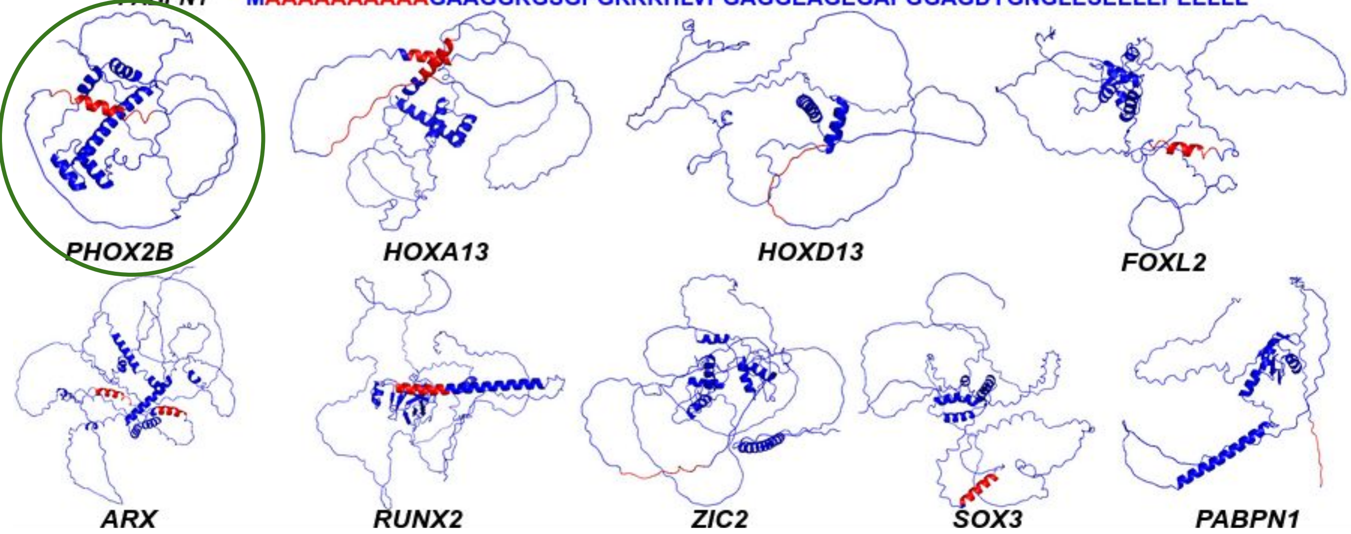
Deciphering the Structural and Functional Roles of Poly-Alanine Repeats: Insights from the Combined Application of NMR and Molecular Simulations

Santiago Rodriguez, Carlos A. Elena-Real, Ons Dakhlaoui, Amin Sagar, Pau Bernadó
Centre de Biologie Structurale (CBS), Montpellier, France

E-mail:srodriguez@cbs.cnrs.fr

PolyA containing proteins

PHOX2B	AGAPGAAGPGGPGGEPGKGGAAAAAAAAAAAAAAAAAAAAGGLAAAGGPGQGWAPGPGPI
HOXA13	GAAAAAAAAAAAAAAGA...FSVAAAAAAAAAAAANQ...PPSAAAAAAAAAAAAAAAAAASSS
HOXD13	GQCRGFLSAPVFAGTHSGRAAAAAAAAAAAAAAASGFAYPGTSERTGSSSSSSSSAVVAARPE
FOXL2	PPKYLQSGFLNNSWPLPQPPSPMPYASCQMAAAAAAAAAAAAAAGPGSPGAAAVVKGLAGPA
ARX	GAAAAAAAAAAAAAAAATATAGPRGEAPPPPPPTARPGERPDGAGAAAAAAAAAAAAWDTLK
RUNX2	AAQQQQQQQQQQQQQQQQQQQQQQQEAAAAAAAAAAAAAAAAAVPRLRPPHDNRTMVEIIA
ZIC2	VSPSAEPQSSSNLSPAAAAAAAAAAAAAAAVSAVHRGGGSGSGGAGGGSGGGSGSGGGGGG
SOX3	RRKTKTLLKKDKYSLPSGLLPPGAAAAAAAAAAAAAAASSPVGVGQRLDTYTHVNGWANGAYS
PABPN1	MAAAAAAAAAAGAAGGRGSGPGRRRHLVPGAGGEAGEGAPGGAGDYGNGLESEELEPEELLL

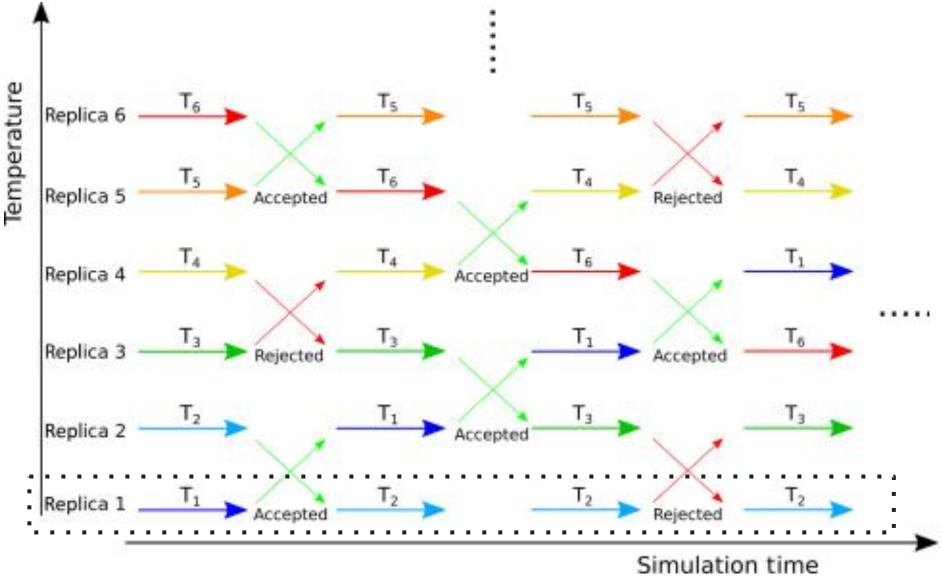


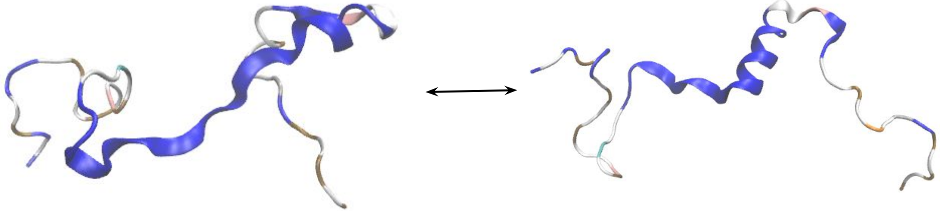
Proteins involved in different biological processes

Expansions in the number of alanines in the polyA region triggers disease

To study the structural behaviour of polyA in PHOX2B with different number of alanines and flanking regions

Conformational sampling of proteins





$$p_{\text{accept}} = min(1, e^{-(E_2-E_1)(\beta_2-\beta_1)}) > \text{random}[0,1)$$

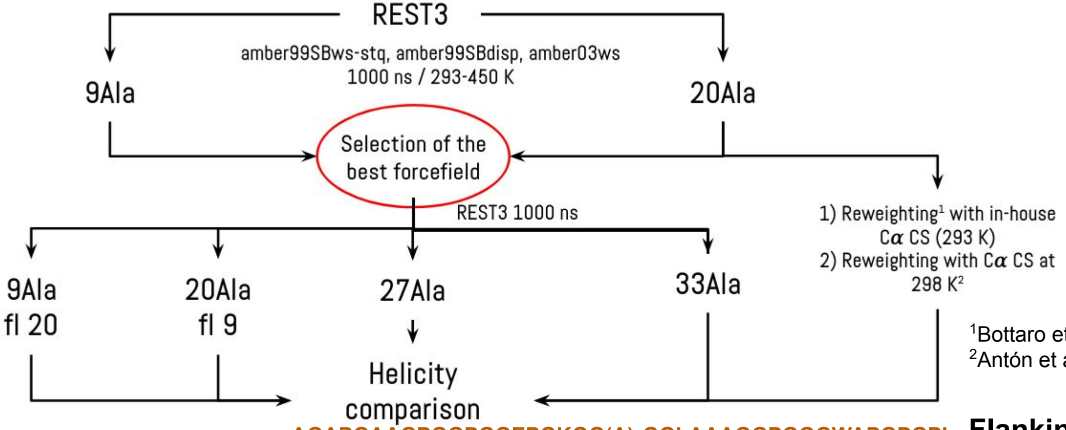
$$\text{where} \quad \beta = \frac{1}{k_B T} \quad \& \quad T_2 > T_1$$

Libraries: Gromacs 2022.5 + Plumed 2.9, compiled with both GPU and MPI library; MKL for FFT speed increment.

Resources: 1 node, 24 CPU cores, 4 GPU cards (HPC Olympe, CALMIP, Toulouse).

Performance: ~50 ns/day

Workflow and results



REST3

amber99SBws-stq, amber99SBdisp, amber03ws
1000 ns / 293-450 K

9Ala

20Ala

Selection of the best forcefield

REST3 1000 ns

1) Reweighting[1] with in-house Cα CS (293 K)
2) Reweighting with Cα CS at 298 K[2]

9Ala fl 20

20Ala fl 9

27Ala

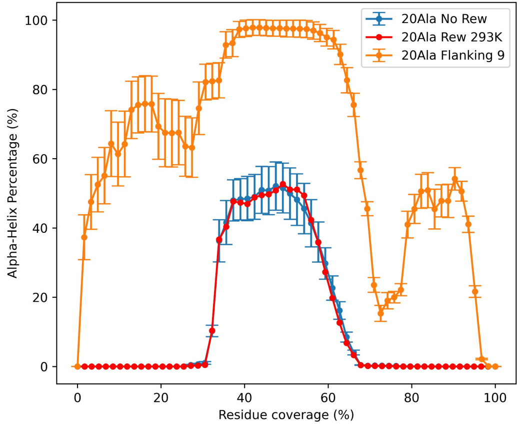
33Ala

Helicity comparison

[1]Bottaro et al., (2020)
[2]Antón et al., (2024)

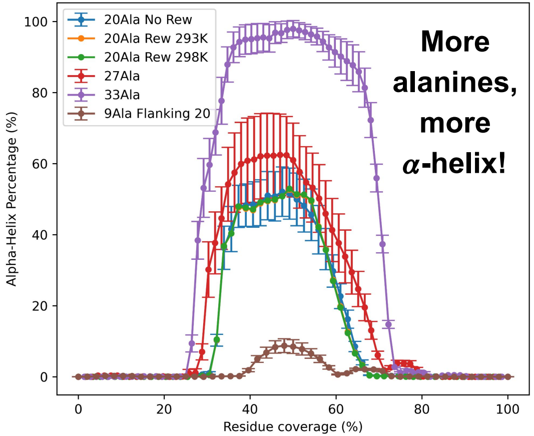
AGAPGAAGPGGPGGEPGKGG$(A)_9$GGLAAAGGPGQGWAPGPGPI
AGAPGAAGPGGPGGEPGKGG$(A)_{20}$GGLAAAGGPGQGWAPGPGPI
AGAPGAAGPGGPGGEPGKGG$(A)_{27}$GGLAAAGGPGQGWAPGPGPI
AGAPGAAGPGGPGGEPGKGG$(A)_{33}$GGLAAAGGPGQGWAPGPGPI

AGAPGAAGPGGPGGEPGKGG$(A)_{20}$GGLAAAGGPGQGWAPGPGPI
DLTEARVQVWFQNRRAKFRKQER$(A)_{20}$KNGSSGKKSDSSRDDESKEA

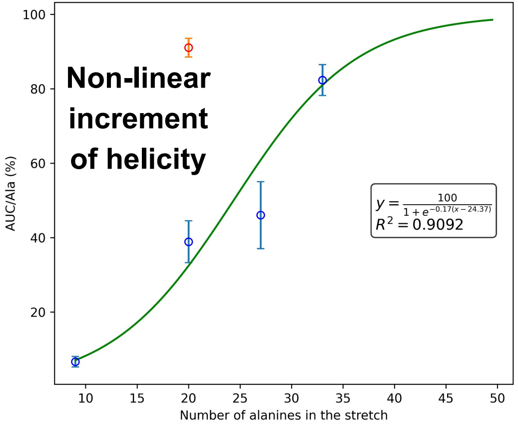


Flanking regions determines the level of helicity!

DLTEARVQVWFQNRRAKFRKQER$(A)_9$KNGSSGKKSDSSRDDESKEA
AGAPGAAGPGGPGGEPGKGG$(A)_9$GGLAAAGGPGQGWAPGPGPI

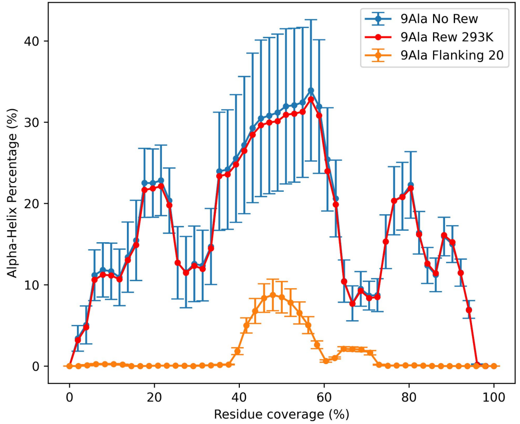


More alanines, more α-helix!



$$y = \frac{100}{1+e^{-0.17(x-24.37)}}$$
$$R^2 = 0.9092$$

Non-linear increment of helicity



Conclusions

✓ We tested three different optimized force fields for disordered proteins, and amber03ws provides the closest results to experimental data.

✓ Similar trends in terms of helicity were obtained when reweighting is done using in-house data (293K) or data from ref 2 (298K).

✓ Helicity is correlated with the number of alanines of the tract.

✓ Flanking regions play a role in the helicity of the poly-A region, suggesting that structure-disrupting residues, like glycines and/or prolines, limit the helicity expansion of both N- and C- terminal tails of the 20Ala stretch.

✓ REST3 simulations allow us to structurally characterize aberrant poly-A expansions, for which the experimental characterization remains extremely challenging.

SCAN ME!

Acknowledgements

CBS

calmip

UNIVERSITÉ MONTPELLIER

REFRACT