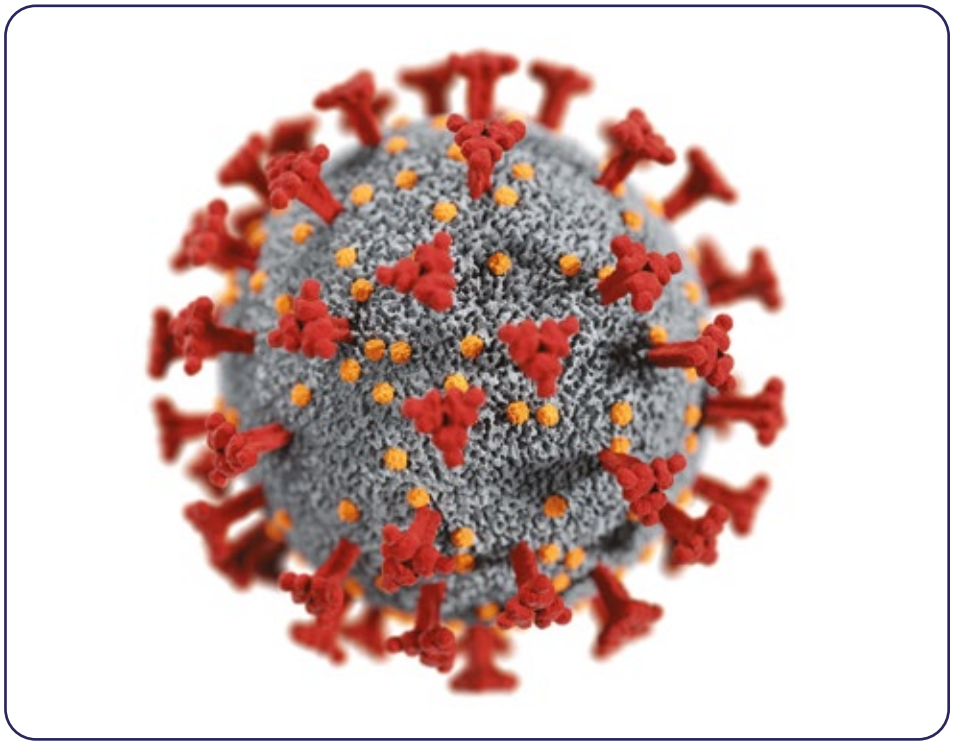
Simulating a flexible infectious viral protein

Disordered proteins play important roles in various diseases, including cancer and certain viral infections. These proteins can change shape depending on their biological context, making them challenging to study. Through an ARCHER2 Pioneer Project, researchers from University College London, King's College London, and the Francis Crick Institute conducted extensive, computationally demanding simulations to characterise a specific region of a protein from SARS-CoV-2, the virus responsible for causing COVID-19. These simulations provided the team with detailed insights into this protein, and the findings may also contribute to a better understanding of other disordered proteins and their roles in various diseases. Ultimately, this research could pave the way for new therapeutic strategies to treat such diseases.



SARS-CoV-2 virion: The virus responsible for COVID-19, which carries the instructions to produce the ORF6 protein that helps it evade the immune system. *Image Credit: BlenderTimer/ Pixabay*.

www.archer2.ac.uk

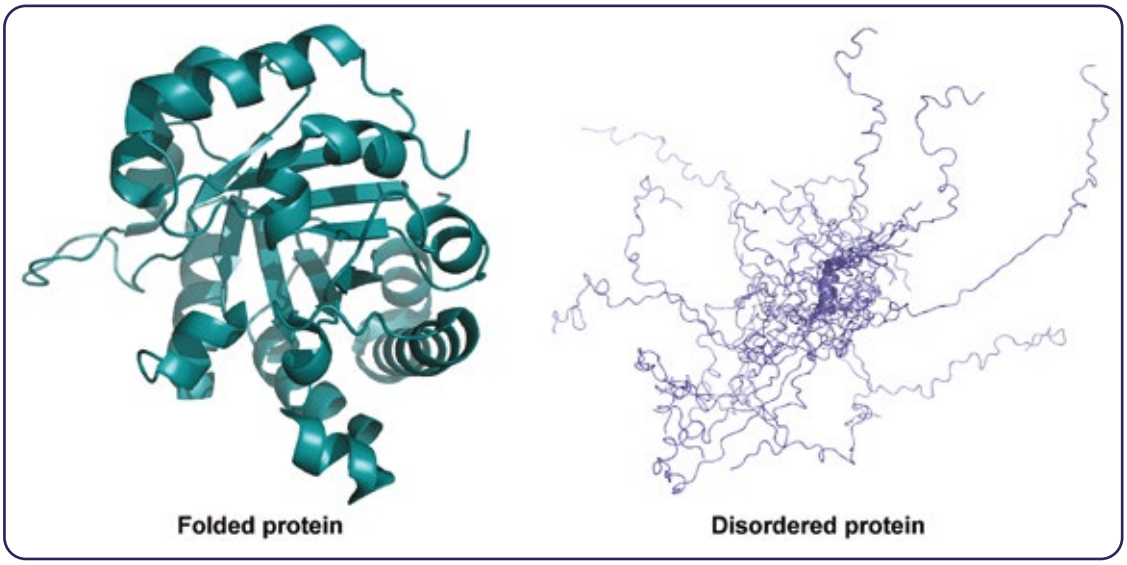


Figure 1: Example structures of a folded protein[1] (left) and a disordered protein ensemble[2] (right).

Intrinsically disordered proteins ('disordered proteins') play important functional roles in many biological processes and are involved in various diseases, including cancer and viruses such as SARS-CoV-2. Unlike folded proteins that generally have a well-defined single structure, disordered proteins are flexible and can change shape depending on their biological context, making them challenging to study (Fig 1).

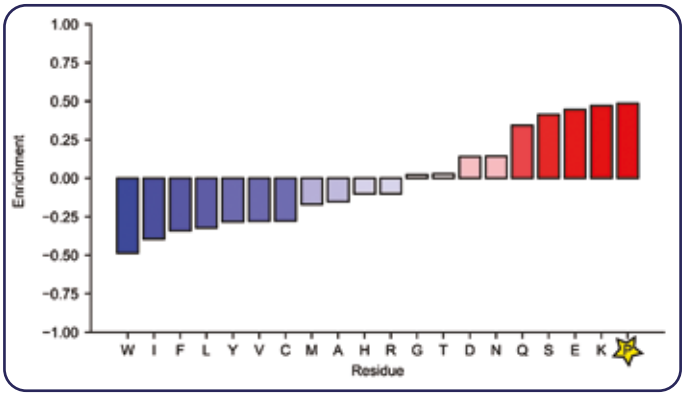


Figure 2: Prolines (P) have an almost two-fold higher abundance in disordered proteins than in folded proteins. Figure adapted from[3].

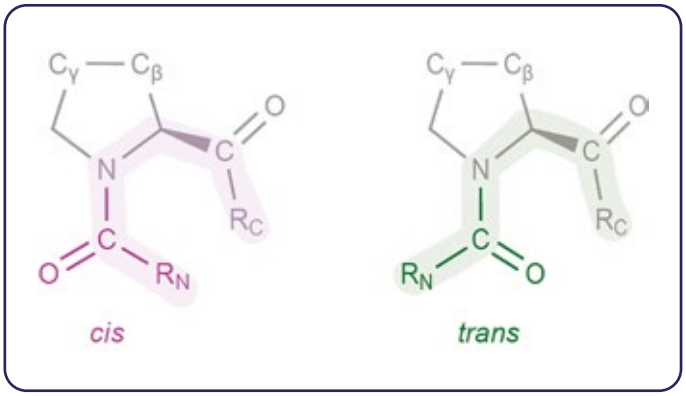
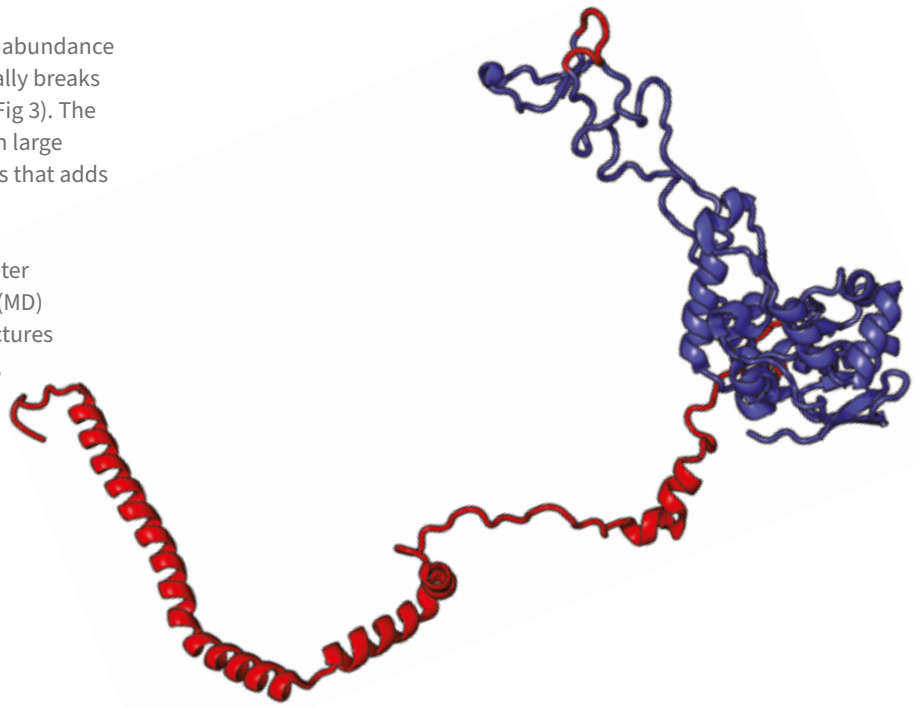


Figure 3: *cis* and *trans* forms of proline.

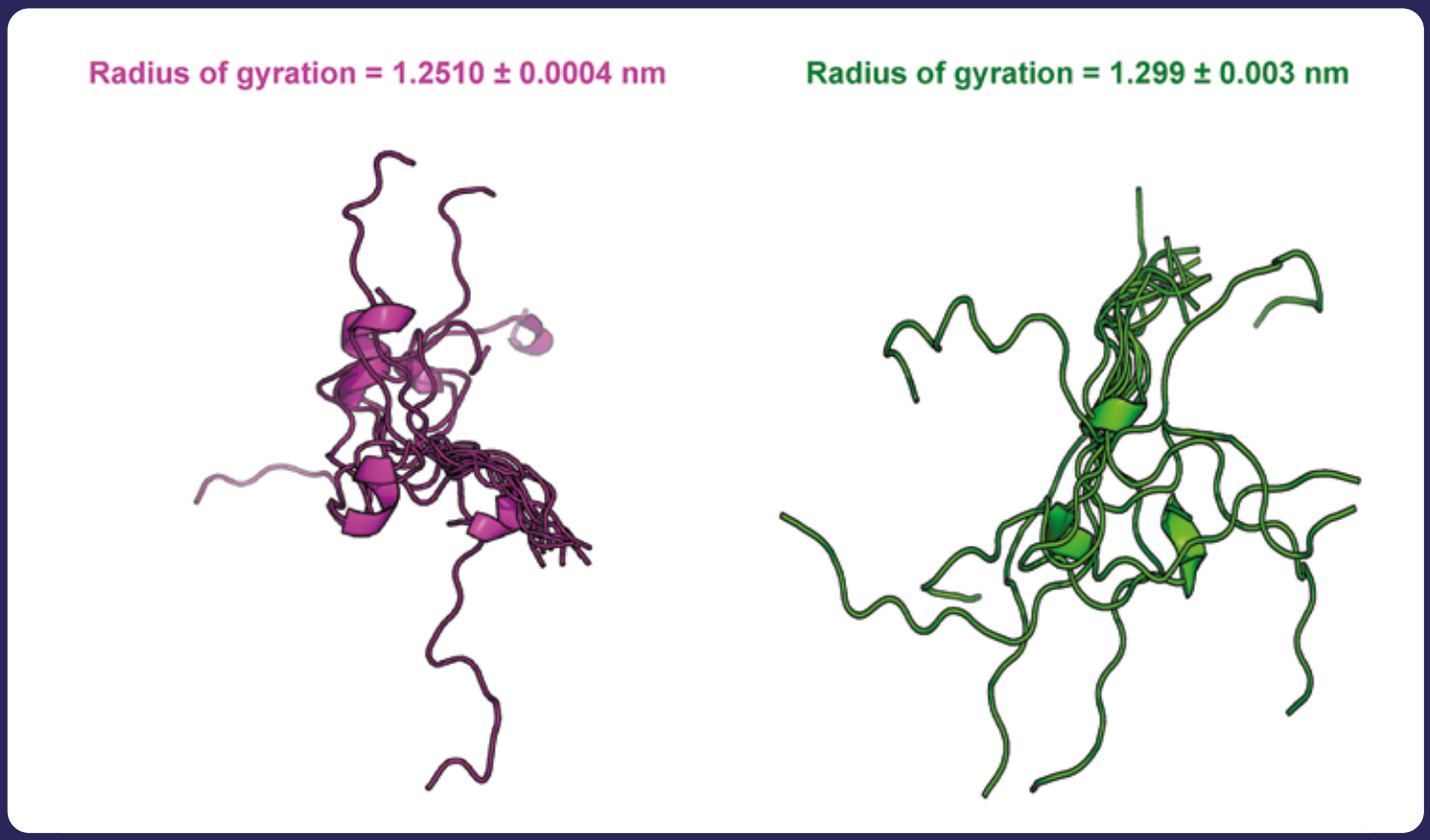
One important feature of disordered proteins is the high abundance of a constituent amino acid, proline (Fig 2), which generally breaks structures, and exists in two forms, called *cis* and *trans* (Fig 3). The switching between these *cis* and *trans* states can result in large conformational changes of proteins and is a slow process that adds complexity to understanding disordered proteins[4].

To study disordered proteins, scientists often use computer simulations. In particular, all-atom molecular dynamics (MD) simulations are regularly employed to visualise the structures of disordered proteins at an atomic resolution. However, because the switching process of proline is slow, simulating both the *cis* and *trans* states can be difficult to achieve using standard approaches.



In our research, we focused on understanding a specific part of ORF6, a disordered protein that is made in cells infected by the SARS-CoV-2 virus. The C-terminal region of ORF6 ($ORF6_{CTR}$) contains a single proline (P57), which might influence the virus's interactions with human proteins and its role in disease[5]. In order to allow us to study both *cis* and *trans* states of the proline in $ORF6_{CTR}$, an enhanced sampling technique called metadynamics was used[6]. By combining metadynamic MD simulations with experimental data, we were able to accurately describe this part of the protein. Our findings revealed that the *cis* form of the proline results in a slightly more compact conformation of the protein than the *trans* form, although both forms are highly dynamic and flexible (Fig 4).

Figure 4: Characterisation of the $ORF6_{CTR}$ *cis* (pink) and *trans* (green) forms by simulation. Both the *cis* and *trans* conformations are highly flexible and lack secondary structure.



Having access to ARCHER2 over the course of two years through the Pioneer Project made it possible to gain detailed insights into $ORF6_{CTR}$ with high accuracy and precision. The significant CPU allocation provided by ARCHER2 and its advanced architecture enabled us to run extensive metadynamic simulations on $ORF6_{CTR}$. These simulations demand substantial computational power due to the complexity and slow timescales of proline conversions in $ORF6_{CTR}$.

We used the GROMACS package[7] to perform our MD simulations, patched with the open-source, community developed PLUMED library[8]. The input and analysis scripts used for our metadynamic MD simulations can be found at https://github.com/hansenlab-ucl/orf6-ctr_cis_trans_conformers/tree/master.

The insights gained from using ARCHER2 are valuable not only for understanding this specific flexible protein from SARS-CoV-2 but may also be important for studying other disordered proteins, proline isomerisation, and their potential roles in various diseases. Our research may help scientists uncover important information about how proline-containing disordered proteins contribute to disease, potentially leading to new therapeutic strategies.

Reference:

1. Roland, B. *et al*. Triosephosphate isomerase I170V alters catalytic site, enhances stability and induces pathology in a Drosophila model of TPI deficiency. *Biochimica et Biophysica Acta*. **1852**, 61-69. (2015).
2. Song, J. *et al*. Micelle-Induced Folding of Spinach Thylakoid Soluble Phosphoprotein of 9 kDa and Its Functional Implications. *Biochemistry*. **45**, 15633-15643. (2006).
3. Quaglia, F. *et al*. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Research*. **50**, D480. (2022).
4. Grathwohl, C. & Wüthrich, K. NMR studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* **20**, 2623–2633 (1981).
5. Gao, X. *et al*. Structural basis for Sarbecovirus ORF6 mediated blockage of nucleocytoplasmic transport. *Nat Commun* **13**, 4782 (2022).
6. Pfaendtner, J. & Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J Chem Theory Comput* **11**, 5062–5067 (2015).
7. Abraham, M. J. *et al*. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
8. Bonomi, M. *et al*. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods* **16**, 670–673 (2019).

Collaborators:

Alice J. Pettitt (University College London)

Gabriella T. Heller (University College London)

Christian D. Lorenz (King's College London)

D. Flemming Hansen (University College London and the Francis Crick Institute)

Journal articles:

Alice J. Pettitt, Vaibhav Kumar Shukla, Angelo Miguel Figueiredo, Lydia S. Newton, Stephen McCarthy, Alethea B. Tabor, Gabriella T. Heller, Christian D. Lorenz, D. Flemming Hansen. An integrative characterisation of proline *cis* and *trans* conformers in a disordered peptide. *Biophysical Journal*. (2024). ISSN 0006-3495. **https://doi.org/10.1016/j.bpj.2024.09.028**

Contact Details:

Hansen Lab **https://www.ucl.ac.uk/hansen-lab/**

Lorenz Lab **https://nms.kcl.ac.uk/lorenz.lab/wp/**

PI: D. Flemming Hansen **d.hansen@ucl.ac.uk**

PI: Gabriella T. Heller **g.heller@ucl.ac.uk**

PI: Christian D. Lorenz **chris.lorenz@kcl.ac.uk**

Lead author: Alice J. Pettitt **alice.pettitt.16@ucl.ac.uk**

About ARCHER2

ARCHER2 is the UK's National Supercomputing Service, a world class advanced computing resource for UK researchers. ARCHER2 is provided by UKRI, EPCC, HPE and the University of Edinburgh. ARCHER2 is the latest in a series of National Supercomputing Services provided to UK researchers.

More ARCHER2 case studies can be found at:
https://www.archer2.ac.uk/research/case-studies/

This project was an **ARCHER2 Pioneer Project**, awarded by UKRI to conduct computationally intensive modelling, simulation and calculations to deliver ambitious and pioneering projects.

epcc | Hewlett Packard Enterprise | UKRI Engineering and Physical Sciences Research Council | Natural Environment Research Council | THE UNIVERSITY of EDINBURGH | www.archer2.ac.uk