From electrons to proteins to data

Societal impact

Structural Biology has had a profound societal impact on medicine, particularly in the context of vaccine development for infectious diseases. The relatively recent Resolution Revolution in cryo-electron microscopy (cryo-EM) has enabled researchers to investigate the intricate 3D structures of biological molecules, including viruses and their constituent proteins.

With the outbreak of the COVID-19 pandemic caused by the SARS-CoV-2 virus, cryo-EM played a crucial role in understanding the structure of the virus, specifically the spike protein responsible for viral entry into host cells. High-resolution structures obtained through cryo-EM allowed researchers to identify key regions on the spike protein that are crucial for interaction with human receptors, which subsequently facilitated the design of vaccines targeting these specific regions [1]. In the past, determining the structure of biological molecules was a time-consuming and technically challenging task. However, cryo-EM has significantly reduced the time required to obtain high-quality structures, allowing researchers to quickly characterise new strains of the virus and adapt existing vaccines accordingly.

The work in this thesis will have contributed to the continuous improvement of cryo-EM as a technology. An illustrative example of this progress is the integration of graphene as a substrate material for sample carriers (Chapter 4). This finding may effectively address the challenges of electrostatic sample charging encountered by researchers, offering the potential to enhance the overall operational efficiency of cryo-EM. Furthermore, investigation of the use of the Timepix3 HPD chip (Chapter 2 and Chapter 3) holds the promise of democratising access to expensive cryo-EM techniques in the coming years. Microscopes designed for 100 kV are a lot more affordable, making it more accessible for smaller labs to obtain such a device. Furthermore, the use of 100 kV has shown great promise, but is still hindered by an effective detector for this energy range. A role that the Timepix3 or future HPD chips could fulfil (Chapter 7). This is further underlined by the recent market introduction of Thermo Fischer's less expensive 100 kV Tundra electron microscope.

In general, public funding of expensive lab techniques, such as electron microscopes, has been a topic of debate. For example, is it feasible for multiple laboratories in the Netherlands to operate a cryo-EM facility? As an alternative to public funding, researchers are encouraged or even mandated to engage in public-private partnerships. E.g. this thesis was (partially) funded by a grant supporting a public-private partnership, in this case, to develop and characterise the Timepix3 chip as detector for electron microscopy. This successful partnership

led to the publication of a patent that was licenced by the private partner [2]. However, it is imperative to acknowledge that the engagement of public-private partnerships to secure funding for these endeavours carries certain inherent complexities and potential risks. One prominent concern revolves around the potential influence that private entities may wield over the research trajectory, potentially skewing scientific pursuits toward commercial interests. The use of patents, while offering a route for private partners to commercialise scientific breakthroughs, further introduces a layer of potential conflict that necessitates vigilance.

The second aspect of this thesis is the description, storage, and management of life science research data and the organisation thereof to help make research data more findable, accessible, interoperable, and reusable (FAIR). The Open Science and FAIR movement have been instrumental in promoting data reuse and addressing the issue of irreproducibility. The reuse of research data has a profound societal impact that extends far beyond the academic community. When researchers share their data for others to reuse and validate, it fosters a culture of transparency, collaboration, and open dissemination of knowledge. Embracing FAIR principles ensures that research data are well organised, described with standardised metadata, and made available through easily accessible repositories. One of the most significant benefits of data reuse is the potential to build on existing research, saving time, resources, and reducing redundancy. This, in turn, can lead to accelerated scientific progress and the development of more robust and reliable conclusions.

On the flip side, the costs of irreproducible science can be substantial and detrimental to society. When research results cannot be independently verified due to a lack of access to the underlying data or flawed methodologies, it can lead to misleading or even harmful applications in real-world contexts. The consequences can range from wasted funding and resources to public trust erosion in scientific findings. Furthermore, in fields where research results impact public policy or medical decisions, irreproducibility can have severe consequences on human well-being.

However, we should be careful not to make the sharing of data a goal on its own. I am careful not to advocate for RDM and data sharing for their own sake. In all cases, the effective use to which research data can be put should be at the heart of our activities. In some cases, a subset of processed data can serve (nearly) the same purpose as the full raw data archive. For example, when the full raw data archive makes the data unwieldy to use. In other cases, this may mean that it is better to store a sample in a freezer than to store (all) data on a server. For example, reanalysis of the sample may be cheaper and may even be done with improved technology in the future. So, whilst (expensive) digital infrastructures and specialised personnel can be part of the complicated data

management practises I am encouraging, they should never replace a pragmatic focus on improving the efficiency and inclusivity of the research process and helping new research questions be answered.

References

1. Wigge, C., Stefanovic, A. & Radjainia, M. The rapidly evolving role of cryo-EM in drug design. *Drug Discovery Today: Technologies* **38,** 91–102. doi:10.1016/j.ddtec.2020.12.003 (2020).
2. **Van Schayck, J. P.** & Ravelli, R. B. G. https://data.epo.org/gpi/EP3525229A1(2018).