Recombinant Protein Production for Drug Discovery at AstraZeneca

Samiyah Ahmed, Lizzi Underwood, Sean Gray, Katherine Peters, Hazel Mak, Alex Edwards

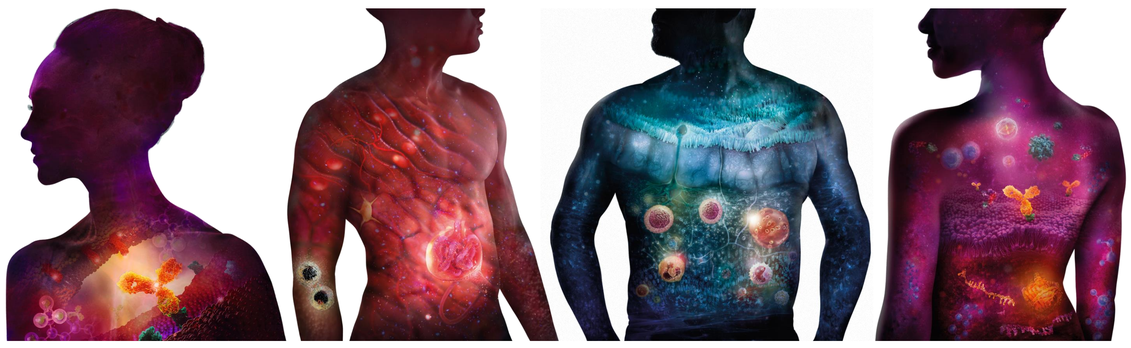
Abstract

Discovery Sciences at AstraZeneca is responsible for identifying molecules active against novel targets in a range of therapeutic areas. For most drug discovery projects in this early phase it is integral to have high quality recombinant protein to drive screening and lead optimisation activities.

Our platform and processes for recombinant protein production at AZ are comprised of a number of stages: Construct design and molecular biology, protein expression, protein purification and quality control.
We are constantly looking for ways to increase throughput and streamline our methods to improve efficiency as we support an increasingly complex project portfolio and are stretched to reduce timelines.

Introduction

At AZ we believe that there are many mechanisms that are yet to be discovered to aid the treatment for many diseases. Our focus is on the following areas:



Oncology | **Cardiovascular, Renal & Metabolism** | **Respiratory & Immunology** | **Rare Diseases**

We produce proteins for the very earliest part of drug discovery (Figure 1). Relative to the later stages, such as clinical development, these are inexpensive activities and as such we work on many projects at any time.

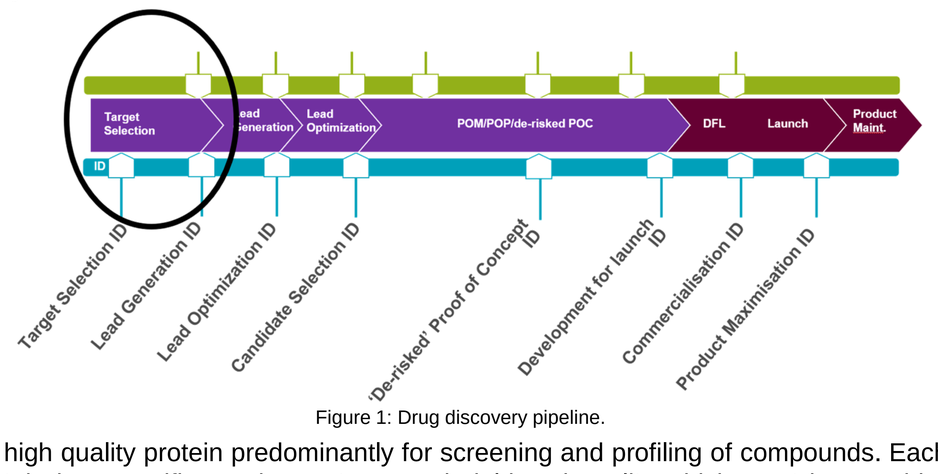


Figure 1: Drug discovery pipeline.

We produce high quality protein predominantly for screening and profiling of compounds. Each end-use of a protein has specific requirements around yield and quality which must be considered at every stage. Before any screening effort can start (Table 1), we must produce the reagent – this can put us under significant time pressure and drives innovation.

Table 1: Drug discovery approaches

Hit-finding approach	Protein Science Requirements
Fragment based ligand generation	• Labelled/immobilised proteins for primary screening • Constructs for iterative crystallography
DNA Encoded Libraries	• High purity, non-aggregated, functional protein • Multiple proteins/variants per screen
High Throughput Screening	• Full length target and substrates including large complex proteins • Multi-mg scale
Fast follower	• Proteins for assay and crystallography in short timelines

Project teams shown in Figure 2, are all motivated towards a common goal. This makes for a very collaborative and interesting environment to work and learn in.

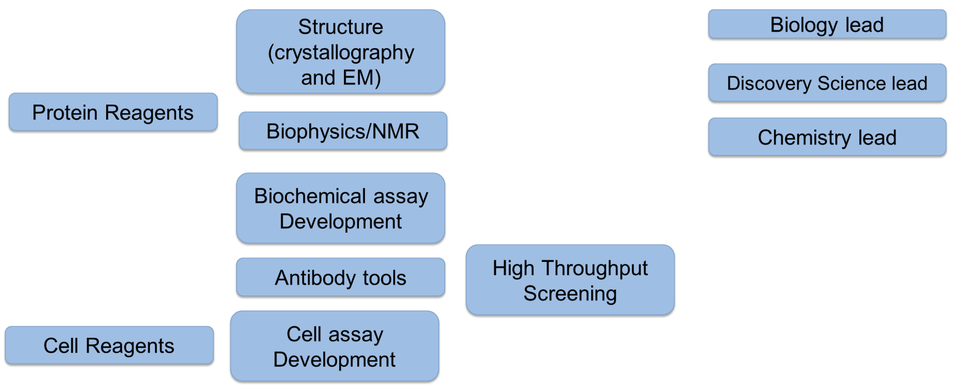


Figure 2: Drug discovery teams.

Assessing Target Feasibility

Prior to initiating any work on a new protein, we must assess its feasibility to generate realistic timelines for its production (Table 2).

Table 2: Target feasibility

	Green	Amber	Red
Commercial availability	Full-length protein commercially available. Published QC data including biological activity	Domain of interest and full-length commercially available but only from cell free system	Not available commercially
Structural data availability	Structure in PDB	Structure of full-length homologue/species variant or structure of domain/truncate in PDB	No structural data reported
Expression precedent	Literature or in house precedence for expression/ purification of protein or similar in suitable form (length/ tags) with QC data including activity. Scale sufficient for an LG campaign	Protein produced but not in suitable form (length/ tags). Questionable QC. Would need work to optimise expression	No internal precedent Little/ no literature

Gene to Protein

Construct Production

Design (1-2 weeks)
- Literature and structural precedent
- Commercial proteins
- Bioinformatic tools

Molecular Biology (2-3 weeks)
- In house

Targets are selected by disease- linkage which requires engagement of all disciplines. Protein engineering techniques, tags, leader sequences and vectors are evaluated.

Expression Testing & Virus Production

E. Coli: pET based vectors with BL21 (DE3) strains as standard.

Insect: Based on the Bac-2-bac system, utilising SF21 insect cells.

Mammalian: CHO/HEK293 cells.

PhyNexus: Robust protocols allow multiple constructs to be run in small scale, parallel, protein purification where comparative expression data is generated.

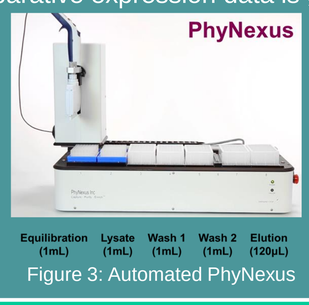


Figure 3: Automated PhyNexus

Large Scale Expression

E. coli expression

In the UK we have capability for up to 60L/week.

Fermenter option is available too if required.

Insect cell expression

In the UK we have capability for up to 35L/week.

Grow volume dependent on expression levels and protein requirement.

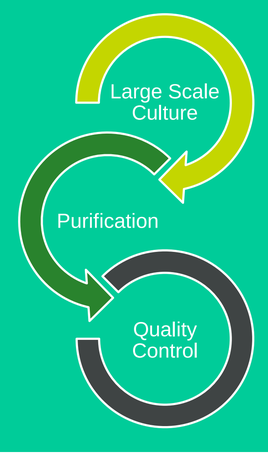
Protein Purification

Protein purification is the least process driven stage of our workflow.

Utilise ÄKTA pure systems with multi-column, multi-loop and sample pumps to allow parallel and streamlined purifications.

Follow a traditional process of:
Affinity Chromatography --> Tag Cleavage/Reverse Affinity/ IEX --> SEC.

Our portfolio requires significant protocol optimisation at this stage as we are challenged to produce larger protein complexes, unprecedented, unstable, uncharacterised proteins and native-like full length proteins



Quality Control of Purified Protein

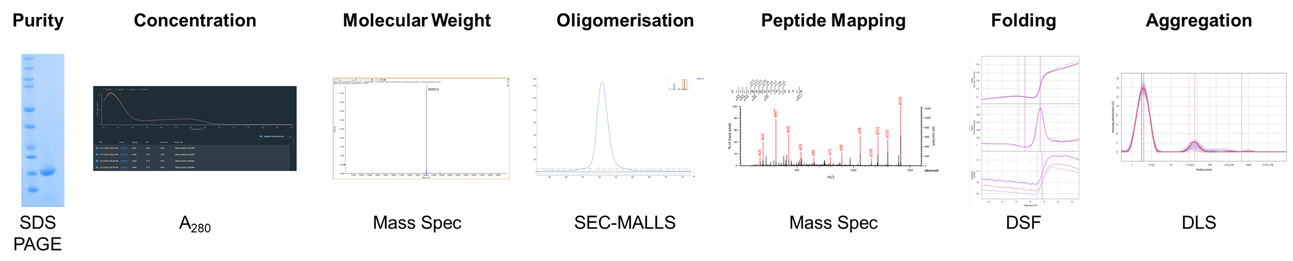


Figure 4: Quality control of purified protein, best practice recommendations.

End Uses in Drug Discovery

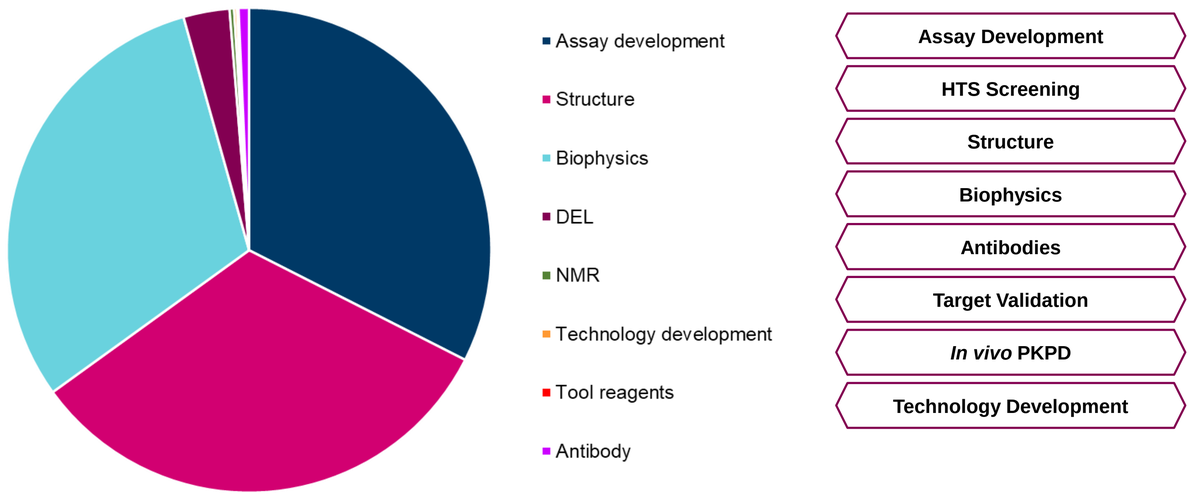


Figure 5: Downstream applications of protein reagents (UK 2019).

Whilst there is approximately an even split of protein batches ultimately used for biophysics, structural biology and assay development/HTS this does not reflect the underlying picture in which often many constructs are ordered/screened for production of protein for X-ray crystallography.

Future Challenges & Areas of Focus

- Can we use machine learning to better predict well expressing & behaved protein constructs? What data do we need to collect to achieve this goal?
- More unprecedented and challenging targets coming to us – how can we work efficiently and innovatively to support these projects?
- Can we improve our QC reagents with “endogenous” qualities?

Acknowledgements

Protein Science Team, Cambridge, UK

Supported by AstraZeneca