Protein characterisation services incorporating Combinatorial Domain Hunting (CDH)
Combinatorial Domain Hunting (CDH)

Introduction
Are you experiencing challenges in expressing your protein domain(s) of interest to sufficient yield, solubility and crystallinity?

Do your proteins have poorly understood domain organisation or are they unrelated to known structures?

If so, Domainex may be able to provide you with a solution.

We provide a comprehensive protein characterisation service. At its heart is our patented Combinatorial Domain Hunting (CDH) technology which has been applied to over 50 targets from a range of cytoplasmic protein classes (Table 1) with an overall success rate in excess of 90%. CDH couples random DNA gene fragmentation to efficient screening of resultant protein fragments to identify soluble protein domains for a range of applications.

Features and advantages of CDH
- Screens thousands of target gene variants in parallel to reduce timelines
- Expression, solubility and biophysical properties of variants compared to select the best ones
- Variant libraries are unbiased to capture diverse protein space
- No prior structural or bioinformatics data required hence allows for quick start-up
- Fragment lengths can be varied so offering full flexibility

CDH enables efficient drug discovery
CDH provides potentially novel protein supply to support a number of key processes in drug discovery (for examples, see Table 2), as follows:
- X-ray crystallography
- Bioassay development
- Biophysical characterisation, e.g. with known ligands or co-factors
- Antibody validation and epitope mapping
- Fragment-based drug discovery, e.g. Target Immobilised NMR Screening (via ZoBio)

CDH deliverables
- Generation of screening library containing ≥20,000 clones of any target gene
- At least 5 clones, expressing soluble, folded domain(s) of interest
- Purified protein to >90% in at least low mg amounts
- Regular updates with our experts to allow a rapid response to your project’s needs
- Optimised expression and purification protocols in a final report

Table 1. Breadth to date of target class application of CDH (as disclosable)

<table>
<thead>
<tr>
<th>Target Class</th>
<th>Number of CDH Projects Performed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinases</td>
<td>22</td>
</tr>
<tr>
<td>Transcription factors</td>
<td>1</td>
</tr>
<tr>
<td>Lysine methyl transferases</td>
<td>3</td>
</tr>
<tr>
<td>Arginine methyl transferases</td>
<td>1</td>
</tr>
<tr>
<td>Metalloproteases</td>
<td>1</td>
</tr>
<tr>
<td>Histone deacetylases</td>
<td>3</td>
</tr>
<tr>
<td>Ubiquitin peptidases</td>
<td>1</td>
</tr>
<tr>
<td>Chaperones</td>
<td>1</td>
</tr>
<tr>
<td>Carboxylases</td>
<td>2</td>
</tr>
<tr>
<td>Lyases</td>
<td>1</td>
</tr>
<tr>
<td>Plakins</td>
<td>1</td>
</tr>
<tr>
<td>Polymerases</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Exemplars of successful application of CDH

<table>
<thead>
<tr>
<th>Target</th>
<th>Target Class</th>
<th>Domain(s) Expressed</th>
<th>Ultimate Use for Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsp90</td>
<td>Chaperone</td>
<td>N-terminal and middle</td>
<td>Protein crystallography</td>
</tr>
<tr>
<td>MEK1</td>
<td>Thr Tyr kinase</td>
<td>Kinase</td>
<td>Protein crystallography</td>
</tr>
<tr>
<td>Human PI3K delta</td>
<td>Phosphoinositide 3 kinase</td>
<td>Kinase</td>
<td>Protein crystallography</td>
</tr>
<tr>
<td>BMX</td>
<td>Non-receptor Tyr kinase</td>
<td>Kinase</td>
<td>Protein crystallography</td>
</tr>
<tr>
<td>MLL4</td>
<td>Lys methyl transferase</td>
<td>SET and post SET</td>
<td>Protein crystallography</td>
</tr>
<tr>
<td>ADAMT5S</td>
<td>Metalloprotease</td>
<td>Spacer</td>
<td>Characterisation of antibody</td>
</tr>
<tr>
<td>PIKFyve</td>
<td>Phosphoinositide kinase</td>
<td>Kinase</td>
<td>Undisclosed</td>
</tr>
<tr>
<td>Maize/Pea aphid ACCase</td>
<td>Acetyl CoA carboxylase</td>
<td>Transferase</td>
<td>Bioassay</td>
</tr>
</tbody>
</table>
Amino acid sequence

Enables expressible domains of proteins to be identified quickly.
Provides protein for structural studies, assay development, etc.
Extremely valuable for novel but challenging drug targets!

Gene fragmentation
Random enzymatic fragmentation of DNA

Gene recoding and synthesis
Sequence optimised for the CDH process and for expression in E. coli

Cloning
Proprietary vectors cover all possible reading frames
Generation of fragment libraries of 20-100,000 clones

Expression and colony selection
Our method ensures selection of fragments inserted in the correct reading frame.
~750 positive colonies are selected for small-scale culture

Final hit selection
Constructs encoding well expressed, soluble protein are ready for scale-up expression in E. coli or transfer to alternative expression systems

Ni²⁺ capture
Recombinant His-tagged proteins are captured on Ni²⁺ resin and analysed

Small-scale culture
Each colony is individually grown and expressed at 4ml scale

Sequencing
Positive clones are sequenced and aligned against the original template
Case Study 1 – MEK-1

MEK is a member of the RAS-RAF-MEK-ERK signal transduction pathway that controls a range of responses including cell proliferation and survival. MEK inhibitors are being approved to treat patients with cancer, e.g. Trametinib to treat BRAF-mutated melanoma. However, this enzyme had proved challenging to generate X-ray crystal structures.

Previously UCB Pharmaceuticals applied traditional protein expression approaches to generating the core enzyme domain (residues 61-393 of the 393 amino acid protein). This resulted in good quality (95% pure) but low yielding (<1mg/L of E. coli culture) protein that tended to oligomerise and did not crystallise.

UCB approached Domainex to undertake a CDH campaign to see if a MEK core domain could be identified that could aid structure-based drug design. Of 24,000 E. coli colonies screened, 15 soluble fragments were identified. One fragment, F11, when complexed with AMPPNP-Mg²⁺ and inhibitors, gave diffraction quality crystals to 2.3Å resolution (Figure 1).

**Characteristics of fragment 11**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>45-392 (Δ264-307 replaced with a linker)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield</td>
<td>&gt;10mg/L in E. coli and up to 50mg/L with insect cells</td>
</tr>
<tr>
<td>Form</td>
<td>monodisperse, no aggregation observed</td>
</tr>
<tr>
<td>Purity</td>
<td>&gt;95%</td>
</tr>
</tbody>
</table>

Figure 1 (A) 2.3Å resolution crystal structure of MEK-1 showing a typical kinase fold consisting of an N-terminal lobe (blue), a C-terminal lobe (orange), an ADPNP co-factor (pink) and an allosteric ligand (green). Water molecules are shown as red spheres. (B), A close-up of the allosteric binding site which is defined by both the N- and C-terminal lobes, where the iodoaryl group of the allosteric ligand is buried deep within the binding site.

Case Study 2: KMT2D (MLL4)

In collaboration with Dr. Wilson of the Francis Crick Institute, CDH was used successfully to identify a KMT2D (MLL4) SET domain construct suitable for crystallography.

Previous efforts to produce MLL4 constructs did not yield crystals despite exhaustive attempts. Within three months a CDH library of 157,000 clones was generated, 25,000 colonies were screened and 18 unique, well-expressed, soluble constructs were identified covering both the SET and postSET domains. One of these CDH constructs was successfully used to produce a 2.2 Å crystal structure in complex with cofactor. This structure has revealed a mechanism for SET domain activation based on the structural differences between MLL1 and MLL4 (Zhang et al., (2015) Structure 23, 1–13).

Figure 2: (A) SDS-PAGE gel of a MLL4 CDH construct purified to near homogeneity from E.coli. (B) Structure of MLL4 at a resolution of 2.2Å (PDB entry: 4Z4P). The protein is shown in cartoon representation with the C-terminal 6xHis tag shown in yellow. Attempts to remove the tag produced active protein that did not crystallise. The cofactor product, S-adenosyl homocysteine, is shown in stick representation, and the single coordinated Zn²⁺ ion as a sphere. The substrate-binding channel, inferred from the structure of MLL1, is indicated in grey.
**Protein expression capabilities**

Domainex provides a wealth of experience in molecular biology, protein expression and purification. We progress projects from construct design to protein expression in as little as 5 weeks. Once purified, we have the expertise to provide full protein characterisation.

The key features of our protein expression services are:

- **Speed** – from construct to initial protein in as little as 5 weeks
- **Customisation** – for instance in choice of expression systems
- **Scale** – with 30L of expression culturing and analytical scale chromatography, we can deliver multi-mg quantities
- **Quality** – e.g. we use a wide range of analytical chromatography for purification

**Case Study – Protein**

We utilised our platform to produce several mgs of full-length, pure G9a lysine methyltransferase and were able to show there were no post-translational modifications (A and B). Differential Scanning Fluorimetry (DSF) was deployed to show that the protein exhibited functional binding to both its cofactor and substrate (C).
About Domainex

Domainex is a fully integrated drug discovery service company based in Cambridge, UK serving pharmaceutical, biotechnology, academic and patient foundations globally. The company also has established research programmes to develop drug candidates in two clearly defined areas: 1) inhibitors of kinases TBK1/IKK epsilon to treat COPD/inflammation and 2) lysine methyltransferases, an emerging target class implicated in oncology. Domainex’s drug discovery service business was established in 2001 and since that time has continued to expand to serve a wider range of clients across the world, including UCB, FORMA Therapeutics, St George’s University, The Institute of Cancer Research and Auspherix. Our expertise and commitment to providing high quality services have resulted in a strong success record in drug discovery, delivering for our clients on average one candidate drug every year for the past six years.

HOW CAN DOMAINEX HELP YOUR DRUG DISCOVERY PROJECT?

Domainex’s highly experienced molecular biologists, assay biologists, medicinal, computational and analytical chemists can be leveraged through our CRO services. Domainex’s focus is on providing highly efficient and well considered scientific solutions to enable successful drug discovery programmes against a wide range of drug targets. Whether your project is at an early stage of drug discovery or has already identified chemical matter, our processes have been shown to result in 30% time-saving compared to industry standards and use less resource, allowing prudent management of your budget.

CONTACTS

If you would like to know more about Domainex’s discovery pipeline, or speak to us regarding your own drug discovery needs, please contact us at: enquiries@domainex.co.uk or call us at +44 (0) 1223 743170

Alternatively we can be contacted directly as follows:

Dr. Thomas Mander MBA
Chief Operating Officer
tom.mander@domainex.co.uk
Tel: +44 (0)1223 743174
Mob: +44 (0)7584 578024

Domainex
162 Cambridge Science Park
Milton Road
Cambridge
CB4 0GH. UK

Publications


