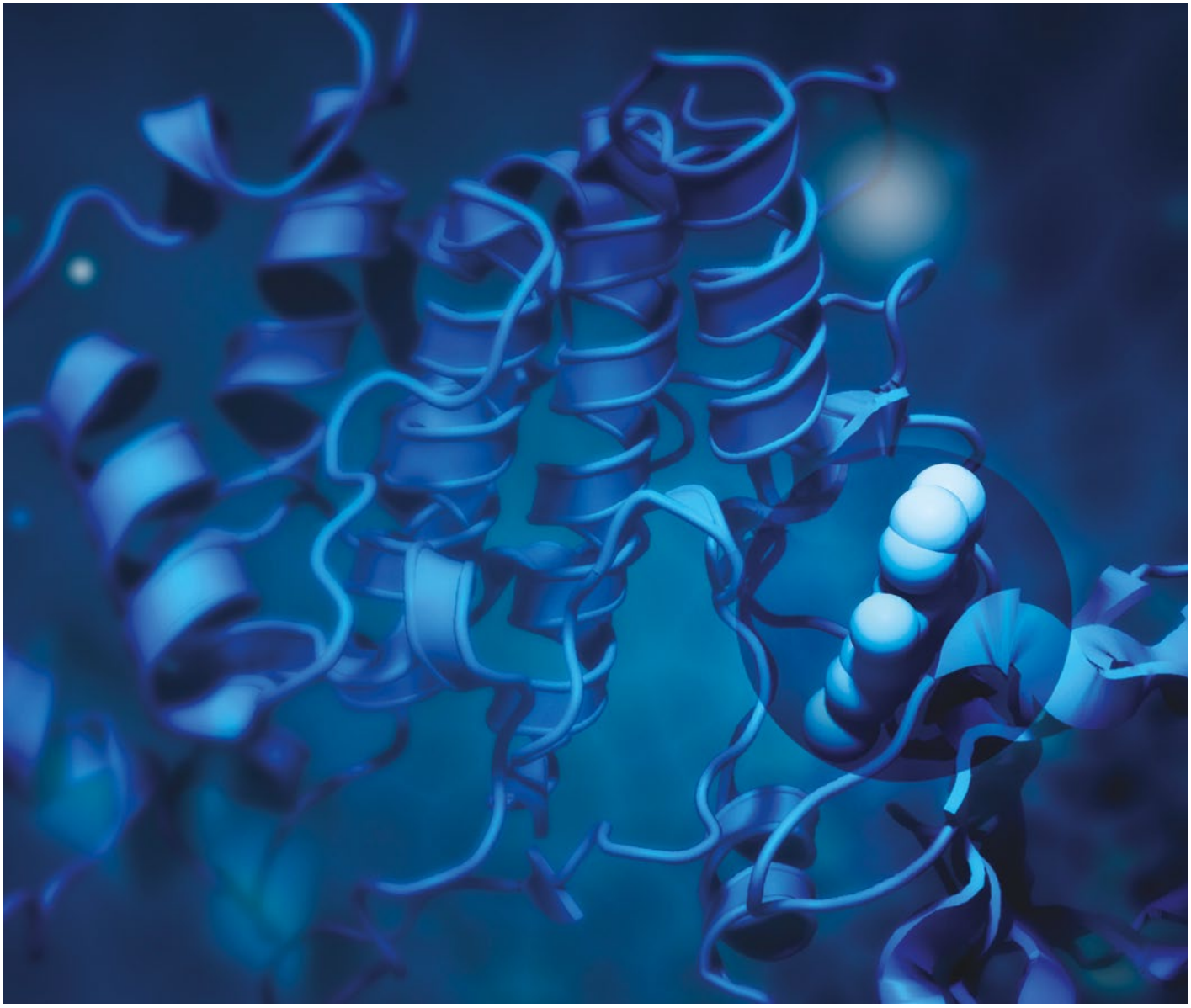
Combinatorial Domain Hunting

Proprietary technology for the rapid generation of novel protein constructs



domainex
the faster route to drug discovery

www.domainex.co.uk

Combinatorial Domain Hunting

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 of tens of 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), many of which can be provided as additional services from Domainex:

- X-ray crystallography
- Bioassay development
- Biophysical characterisation, e.g. with known ligands or co-factors
- Antibody generation and validation
- Fragment-based drug discovery via *FragmentBuilder*, the leading platform with MicroScale Thermophoresis (MST) at its core

CDH deliverables

- Generation of screening library containing 20,000-100,000 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
- Confirmed antibody binding activity or oligomeric state
- Confirmation of small molecule binding eg. by DSF or MST
- 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)

Target Class	Number of CDH Projects Performed
Kinases	25
Transcription factors	3
Lysine methyl transferases	3
Arginine methyl transferases	1
Metalloproteases	1
Histone deacetylases	3
Ubiquitin peptidases	1
Chaperones	1
Carboxylases	2
Lyases	1
Plakins	1
Polymerases	1

Table 2. Exemplars of successful application of CDH

Target	Target Class	Domain(s) Expressed	Ultimate Use for Protein
Hsp90	Chaperone	N-terminal and middle	Protein crystallography
MEK1	Thr Tyr kinase	Kinase	Protein crystallography
Human PI3K delta	Phosphoinositide 3 kinase	Kinase	Protein crystallography
BMX	Non-receptor Tyr kinase	Kinase	Protein crystallography
MLL4	Lys methyl transferase	SET and post SET	Protein crystallography
ADAMTS5	Metalloprotease	Spacer	Characterisation of antibody
PIKfyve	Phosphoinositide kinase	Kinase	Undisclosed
Maize/Pea aphid ACCase	Acetyl CoA carboxylase	Transferase	Bioassay

CDH Process Flow Chart

Amino acid sequence

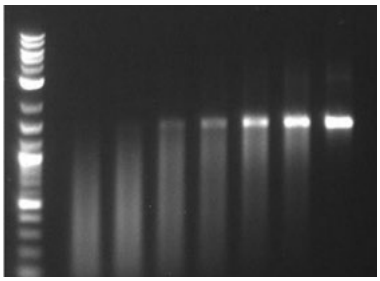
ARTKQTARKSTGGKA
PRKQLATKAARKSAP
ATGGVKKPHRYRPGT
VALREIRRYQKSTEL
LIRKLPFQRLVREIA
QDFKTDLRFOSSAVM
ALQEACEAYLVGLFE

Gene recoding and synthesis

Sequence optimised for the CDH process and for expression in *E. coli*

Gene fragmentation

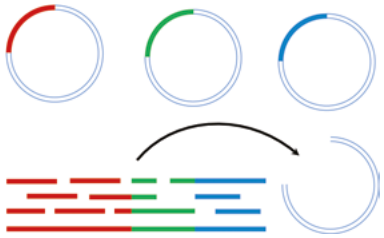
Random enzymatic fragmentation of DNA



Cloning

Proprietary vectors cover all possible reading frames

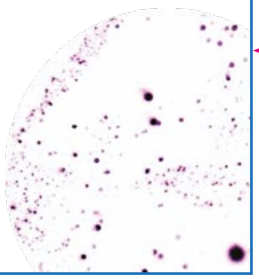
Generation of fragment libraries of 20-100,000 clones



Expression and colony selection

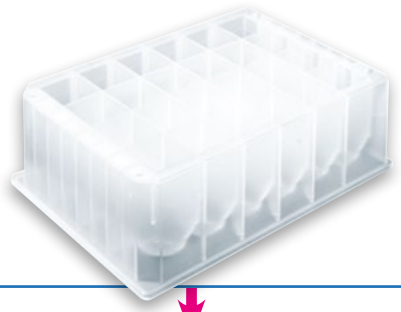
Our method ensures selection of fragments inserted in the correct reading frame.

~750 positive colonies are selected for smallscale culture



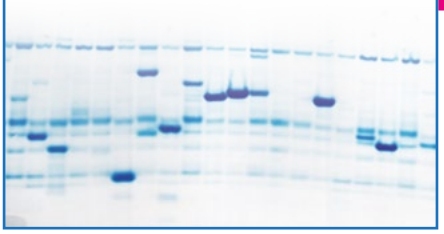
Small-scale culture

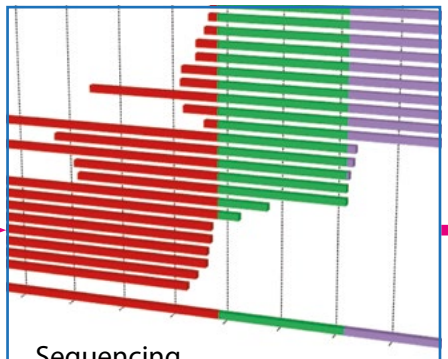
Each colony is individually grown and expressed at 4ml scale



Ni^{2+} capture

Recombinant His-tagged proteins are captured on Ni^{2+} resin and analysed



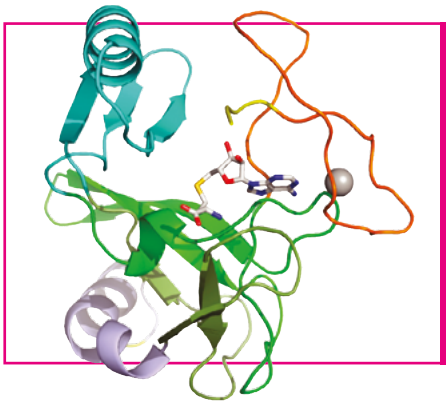


Sequencing

Positive clones are sequenced and aligned against the original template

Final hit selection

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



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.

Combinatorial Domain Hunting

Case Studies

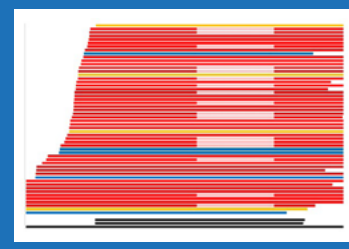
Case Study 1: Deubiquitinylating Enzyme (DUB)

CDH is well suited to a wide range of drug target classes, including kinases and other enzymes. The following data are taken from a client CDH programme performed on a deubiquitinylating enzyme (DUB). In this instance, for an additional fee the 100,000 clone library was split 50:50 between the wild-type and a mutant wherein a flexible loop was replaced with a short linker. The library was produced and screened within 3 months. During the CDH screen, 54 clones were identified that produced soluble protein in good yields (see Figures 1A and 1B).

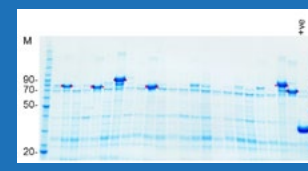
Together with the client, five constructs were chosen for scale up and protein purification. For all clones between 1 and 10 mg of homogenous protein at concentrations between 3-8mg/ml and a purity >95% (see Figure 1C) were delivered to the client for preliminary structural studies.

The client was able to get a high-resolution X-ray crystal structure for one of the constructs, which enabled their SBDD programme for a target which had hitherto been refractory to crystallography.

A



B



C

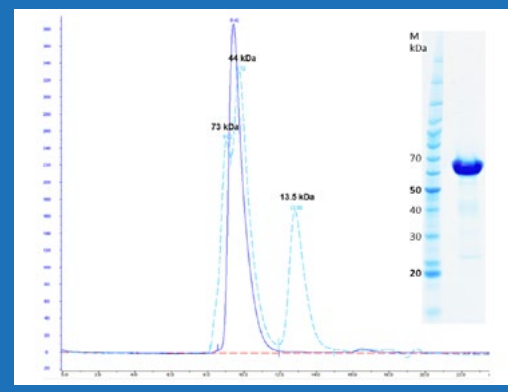


Figure 1: (A) The graph represents soluble protein fragments as identified by CDH and their relative location compared to the parental gene. The bottom bars (black) represent the CDH template used for the process (longer bar) and the catalytic domain as predicted by bioinformatics (shorter bars). The coloured bars represent the clones as identified during the CDH screening process, which yield high (red), medium (yellow) and lower (blue) amounts of soluble protein. The pale red regions represent the sequence variation (loop excision) introduced at the template stage. (B) Partially Ni-IMAC purified CDH clones. SDS-PAGE gel analysis of a representative set of expressed proteins. Each lane contains partially Ni purified protein from approximately 1 ml of cell culture during the screening process. (C) Analysis of 1 mg/ml conc. sample by size-exclusion chromatography: Superdex 75 (10/300). The elution profile shows one homogenous peak of the expected size. (Inset) SDS-PAGE analysis of purified protein (10 µg).

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).

A

B

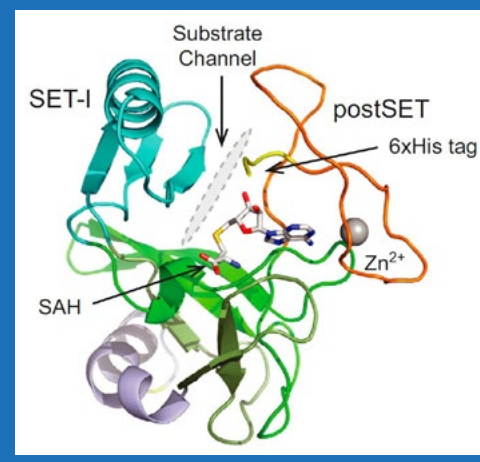


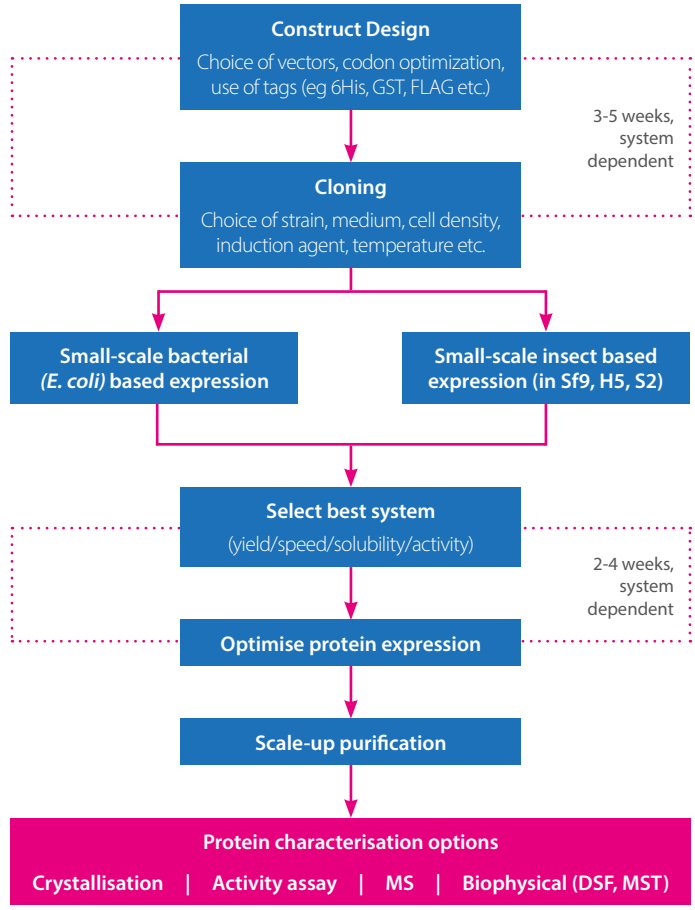
Figure 2: (A) SDS-PAGE gel of a MLL4 CDH construct purified to near homogeneity from *E.coli*. (B) Structure of MLL4 at a resolution of 2.2Å (PDB entry: 4Z4P). The protein is shown in cartoon representation with the C-terminal 6xHis tag shown in yellow. Attempts to remove the tag produced active protein that did not crystallise. The cofactor product, S-adenosyl homocysteine, is shown in stick representation, and the single coordinated Zn^{2+} 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:

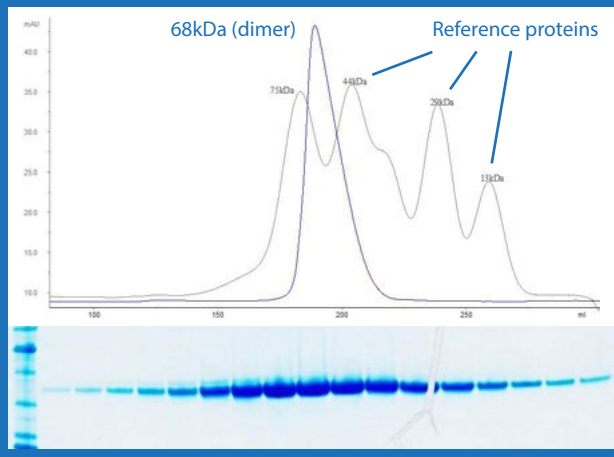
- **Speed**
  - From amino acid sequence to protein in as little as 5 weeks
- **Customisation**
  - Choice of tags
  - Choice of expression systems
  - Choice of protein chromatography columns (affinity, SEC, IEX)
- **Scale**
  - Up to 30L expression cultures
  - We can deliver multi-mg quantities of purified protein
- **Quality**
  - Range of analytical techniques (DSF, MST, SEC, UV, MS)
  - We deliver high quality protein for x-ray crystallography, NMR, bioassay use, fragment screening, bioanalytics of antibodies.



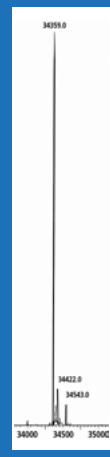
Case Study 3: Protein

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).

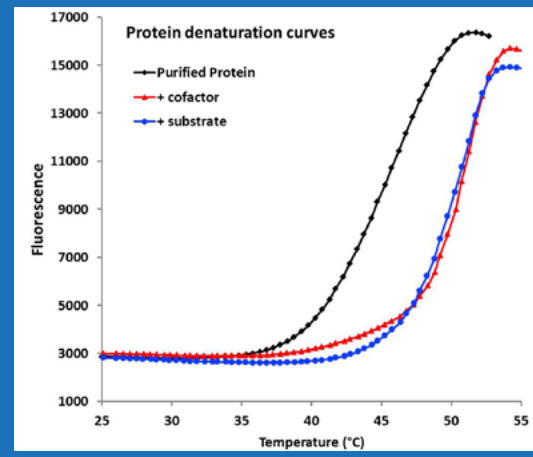
A



B



C



About Domainex

Domainex is a fully integrated drug discovery service company based near Cambridge, UK serving pharmaceutical, biotechnology, academic and patient foundations globally. 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 has resulted in a strong success record in drug discovery, delivering an average of 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 provides 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 a 30% time-saving compared to industry standards and use less resource, allowing prudent management of your own budget.

Contacts

If you would like to know more about Domainex's discovery services, or speak to us regarding your own drug discovery needs, please contact us at: enquiries@domainex.co.uk

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
Chesterford Research Park
Little Chesterford
Cambridge
CB10 1XL, UK

The London BioScience
Innovation Centre
2 Royal College St
London
NW1 0NH, UK

Publications

Reich *et al.*, (2006) Combinatorial Domain Hunting: An effective approach for the identification of soluble protein domains adaptable to high-throughput applications. *Protein Sci.* Oct;15(10):2356-65

Maclagan *et al.*, (2011) A combinatorial method to enable detailed investigation of protein-protein interactions. *Future Med Chem.* Mar;3(3):271-82

Meier *et al.*, (2012) Engineering human MEK-1 for structural studies: A case study of combinatorial domain hunting. *J Struct Biol.* Feb;177(2):329-34

Zhang *et al.*, (2015) Evolving Catalytic Properties of the MLL Family SET Domain. *Structure.* Oct; 23: 1–13

McAlister M, *et al.* (2003) Method for Producing and Identifying Soluble Protein Domains. WO 03/040391