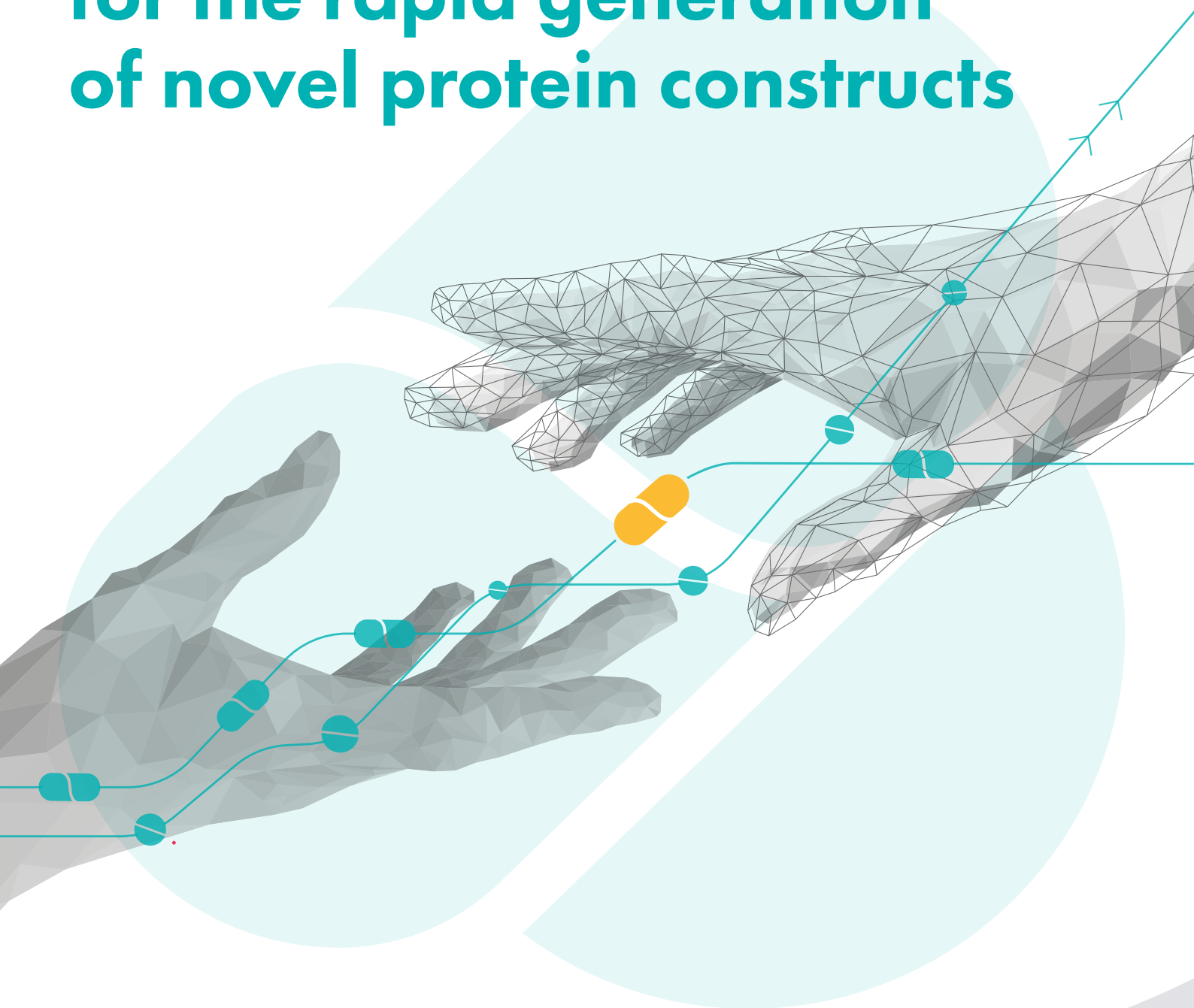
Combinatorial Domain Hunting

Unlocking Challenging Proteins

Proprietary technology for the rapid generation of novel protein constructs



Domainex

domainex.co.uk

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 a range of targets, including DUBs, epigenetic targets (incl. lysine, arginine and RNA methyltransferases), metalloproteases, polymerases, transcription factors and biologics, 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 2–3 months.
- 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
- Fragment lengths can be varied so offering full flexibility
- A core region can be fixed and variation introduced to both N- and C-termini (Figure 1A) or the whole sequence used (Figure 1B)
- Variations can also be introduced to internal loops, connecting two or more fixed cores (Figure 1C)
- A variant of CDH for exploring protein-protein interactions known as CDH2 is available.
- CDH has the potential for competitive advantage if your target has hitherto been difficult to fold or crystallise

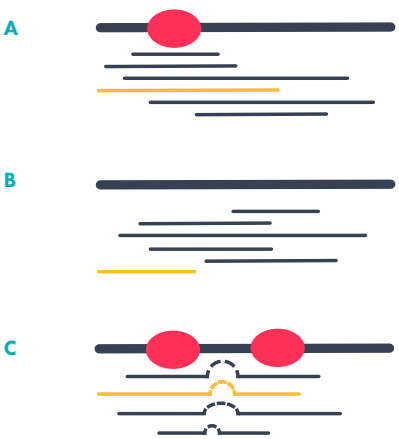


Figure 1 DNA library options. (A) Fixed core, (B) random fragmentation, (C) two fixed cores with variations of the loop.

CDH enables efficient drug discovery

CDH provides novel protein constructs to support a range of functions, e.g. assay development, fragment screening and X-ray crystallography. CDH is offered as part of a suite of enabling services that Domainex provides:

- 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 up to 100,000 clones of any target gene
- Small-scale culture of hundreds of variants
- Up to 20 clones expressing soluble, folded domains partially sequenced
- Parallel scale-up of best constructs to maximise success rate
- 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 2. Exemplars of successful application of CDH

Target	Target Class	Domain(s) Expressed
Hsp90	Chaperone	N-terminal and middle
MEK1	Thr Tyr kinase	Kinase
Human PI3K delta	Phosphoinositide 3 kinase	Kinase
BMX	Non-receptor Tyr kinase	Kinase
MLL4	Lys methyl transferase	SET and post SET
ADAMTS5	Metalloprotease	Spacer
PIKfyve	Phosphoinositide kinase	Kinase
Maize/Pea aphid ACCase	Acetyl CoA carboxylase	Transferase
Huntingtin	PPI	N-terminal
Usp28	Deubiquitinase	USP

CDH Process Flow Chart

Amino acid sequence

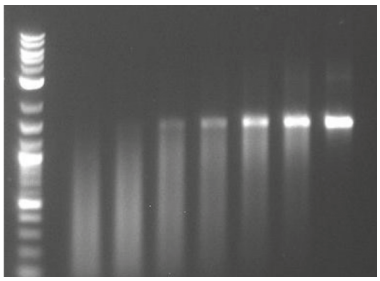
ARTKQTARKSTGGKA
PRKQLATKAARKSAP
ATGGVKKPHRYRPGT
VALREIRRYQKSTEL
LIRKLPFQRLVREIA
QDFKTDLRFOSSAVM
ALQEACEAYLVGLFE

Gene recoding and synthesis

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

Gene fragmentation

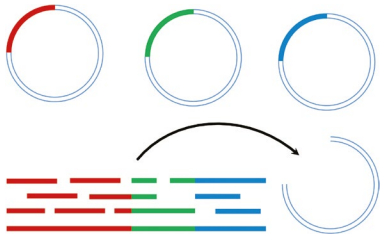
Random enzymatic fragmentation of DNA



Cloning

Proprietary vectors cover all possible reading frames

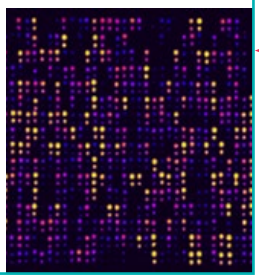
Generation of fragment libraries of 20-100,000 clones



Expression and colony selection

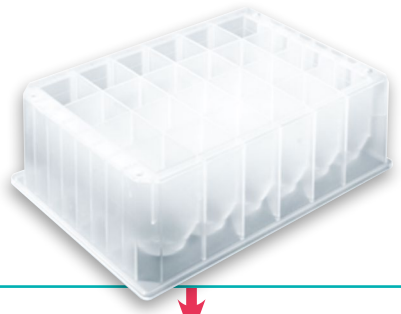
Our method ensures selection of fragments inserted in the correct reading frame.

~750 positive colonies are selected for small-scale culture



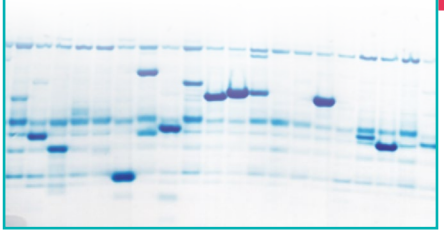
Small-scale culture

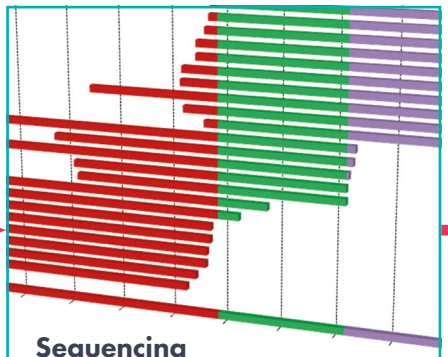
Each colony is individually grown and expressed at 4ml scale



Affinity capture

Recombinant proteins are captured on affinity resin and analysed



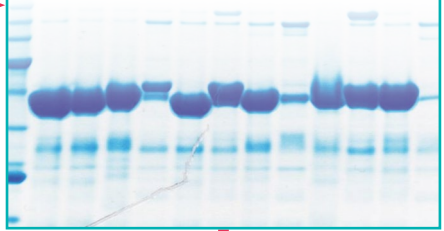


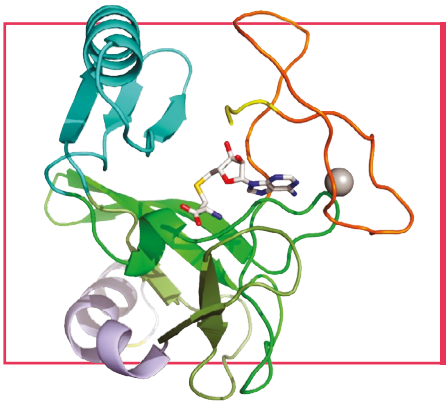
Sequencing

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.

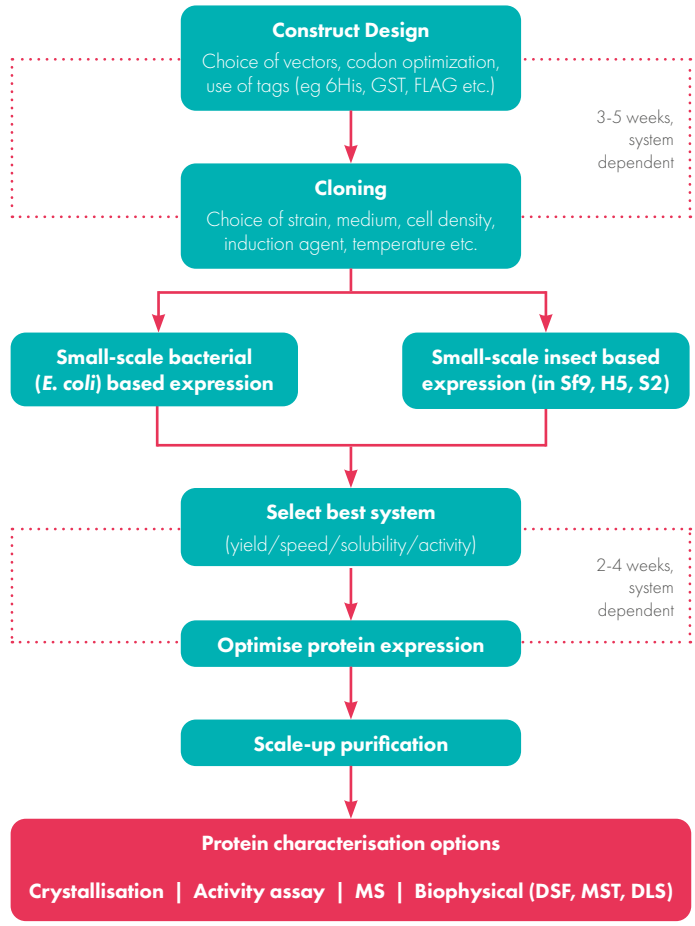
Extremely valuable for novel but challenging drug targets.

Protein expression capabilities

Domainex provides a wealth of experience in molecular biology, protein expression and purification. We progress projects from construct design to purified protein in as little as 2 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 2 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, DLS)
  - We deliver high quality protein for X-ray crystallography, NMR, bioassay use, fragment screening, bioanalytics of antibodies.



Case Study 1: 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 revealed a mechanism for SET domain activation based on the structural differences between MLL1 and MLL4 (Zhang *et al.*, (2015) Structure 23, 1–13).

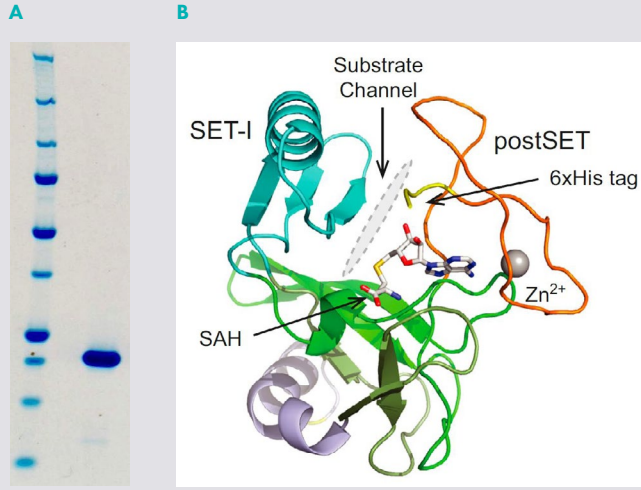


Figure 1 (A) SDS-PAGE gel of a MLL4 CDH construct purified to near homogeneity from *E.coli*. (B) Structure of MLL4 at a resolution of 2.2Å (PDB entry: 4Z4P). The protein is shown in cartoon representation with the C-terminal 6xHis tag shown in yellow. Attempts to remove the tag produced active protein 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.

Combinatorial Domain Hunting

Unlocking Challenging Proteins

Case study 2: Huntingtin

Huntington's Disease (HD) is caused by a mutation in the Huntingtin (HTT) gene, resulting in a pathogenic expansion of a polyglutamine (polyQ) repeat at the N-terminus of Htt. However, only limited structural information on Htt was available at the start of this study. Htt is a large (348 kDa) protein, essential for embryonic development and involved in diverse cellular activities such as vesicular transport, endocytosis, autophagy and transcription regulation. It is suggested that Htt serves as a protein-protein interaction hub. The goal of the project was to utilise CDH to identify and produce soluble and stable protein constructs from the N-terminal region of the Huntingtin (HTT) gene.

A diverse library of over ~90k clones was tested for expression of recombinant, His-tagged protein constructs. Ultimately, two CDH-derived N-terminal HTT constructs were expressed at 7 mg/L and 2.2 mg/L, respectively, at ~90% purity after two columns. Both proteins show evidence of folding by DSF and are ready for further investigations.

To our knowledge, this is the first time that N-terminal HTT protein constructs have been produced at such high yield and purity. This will enable further structure/function analysis of this important region of the Huntingtin protein.

This work was funded by the CHDI Foundation — a biomedical research organisation devoted to accelerating therapeutic development for Huntington's disease.

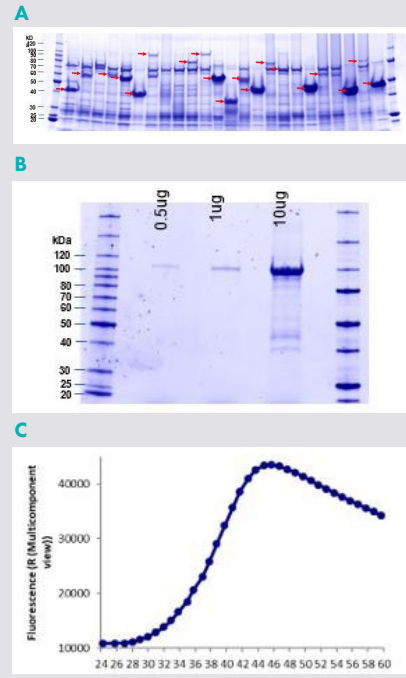


Figure 1: (A) Representative SDS-PAGE analysis from CDH screen (B) SDS-PAGE and (C) DSF analysis of one CDH derived N-terminal Htt protein construct.

Case study 3: CD73 – Protein production

CD73 (also known as ecto-50-nucleotidase, e5NT) is a eukaryotic extracellular glycoprotein with potential applications in the treatment of cancer and inflammation. CD73 catalyses the hydrolysis of extracellular AMP to adenosine and plays a pivotal role in switching on adenosine signalling via the P1 receptors of the purinergic signalling pathway.

Domainex was tasked with establishing a CD73 biochemical assay and a ligand binding assay using Microscale Thermophoresis (MST) as part of a small molecule drug discovery programme. Commercially available CD73 was found to be of insufficient quality for the MST studies, so Domainex was also tasked with producing high-quality CD73 protein to support these assays.

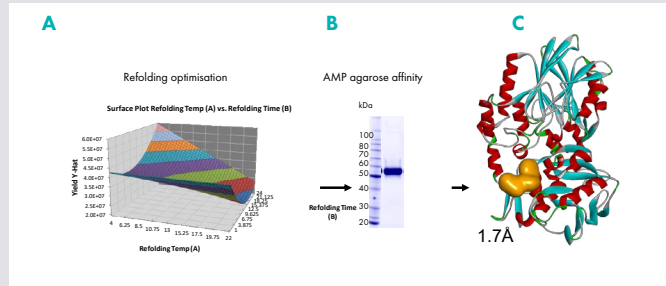


Figure 1: (A) The refolding procedure was optimised using a Design-of-Experiment approach, varying the refolding temperature, the refolding time as well as the ratio of protein dilution. (B) Homogenous and pure protein was purified in multi-milligram quantities: SDS-PAGE analysis. (C) X-ray crystal structure of CD73 bound to adenosine 5'-(α,β methylene)diphosphate (AMPCP), a CD73 inhibitor (dataset resolution 1.7Å). The structure overlays well with the Apo open conformation structure which was also solved.

Recombinant CD73 was successfully isolated from *E.coli* inclusion bodies, refolded and purified to homogeneity in multi-milligram quantities. The resulting protein was not only used to screen the output from a *LeadBuilder* virtual screen by both biochemical and MST assays, but was of sufficiently high yield and quality to enable additional in-house STD-NMR and X-ray crystallography studies.

About Domainex

Domainex is a fully integrated drug discovery service company based at Cambridge, UK. We serve pharmaceutical, biotechnology, academic organisations and patient foundations globally. We have ambitious growth plans and are expecting to reach 110 biologists and chemists in the near future. We provide integrated services, from disease target selection to candidate drug nomination. We have a very strong reputation for contributing innovative ideas, undertaking high-quality experiments and for generating intellectual property on behalf of our clients. We strive to build strong, dynamic relationships. In 2021 we served over 60 clients from the UK, Europe, the United States, Japan and Australia and had a project renewal rate of over 80%.

How Can Domainex Help Your Drug Discovery Project?

Our highly experienced, multi-disciplined scientists – molecular biologists, protein biochemists, assay biologists, structural biologists, medicinal, computational and bio/analytical chemists, *in vitro* pharmacologists and ADME scientists – will support you to advance your drug discovery projects towards drug development effectively and efficiently. We provide customised programmes to address your specific needs at each stage of drug discovery. We draw from a wealth of expertise built up over the last 20 years across a wide range of drug targets and therapeutic areas. From our sites within Europe's leading bioscience hub at Cambridge, UK and with access to the very latest cutting-edge technologies, we are able to help you realise your goals and enrich your discovery pipeline.

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 Executive Officer
tom.mander@domainex.co.uk
Tel: +44 (0) 1223 743174
Mob: +44 (0)7584 578024

Social

Domainex

@Domainex_UK

Chemistry & Bioanalytical Centre of Excellence
Churchill Building
Chesterford Research Park
Little Chesterford
Saffron Walden
CB10 1XL, UK

Biology Centre of Excellence
Iconix 2, Unity Campus
London Road
Pampisford
Cambridge
CB22 3EG, 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