An open library of human kinase domain constructs for automated bacterial expression

Steven K. Albanese[1,2†], Daniel L. Parton[2†], Sonya M. Hanson[2], Lucelenie Rodríguez-Laureano[2], Mehtap Işık[2,3], Julie M. Behr[2,7], Scott Gradia[4§], Chris Jeans[4], Nicholas M. Levinson[6], Markus A. Seeliger[5], John D. Chodera[*2]

***For correspondence:** john.chodera@choderalab.org (John D. Chodera)

[†]These authors contributed equally to this work

Present address: [§]Caribou Biosciences, Berkeley, CA 94720

[1]Gerstner Sloan Kettering Graduate School, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [2]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [3]Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065; [4]QB3 MacroLab, University of California, Berkeley, CA 94720; [5]Department of Pharmacological Sciences, Stony Brook University Medical School, Stony Brook, NY 11794; [6]Department of Pharmacology, University of Minnesota, Minneapolis, MN 55455; [7]Tri-Institutional Program in Computation Biology and Medicine, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065

Abstract Kinases play a critical role in many cellular signaling pathways and are dysregulated in a number of diseases, such as cancer, diabetes, and neurodegeneration. Since the FDA approval of imatinib in 2001, therapeutics targeting kinases now account for roughly 50% of current cancer drug discovery efforts. The ability to explore human kinase biochemistry, biophysics, and structural biology in the laboratory is essential to making rapid progress in understanding kinase regulation, designing selective inhibitors, and studying the emergence of drug resistance. While insect and mammalian expression systems are frequently used for the expression of human kinases, bacterial expression systems are superior in terms of simplicity and cost-effectiveness but have historically struggled with human kinase expression. Following the discovery that phosphatase coexpression could produce high yields of Src and Abl kinase domains in bacterial expression systems, we have generated a library of 52 His-tagged human kinase domain constructs that express above 2 μg/mL culture in a simple automated bacterial expression system utilizing phosphatase coexpression (YopH for Tyr kinases, Lambda for Ser/Thr kinases). Here, we report a structural bioinformatics approach to identify kinase domain constructs likely to express in bacteria, experiments demonstrating our simple construct selection strategy selects constructs with good expression yields in a test of 84 potential kinase domain boundaries for Abl, and yields from a high-throughput expression screen of 96 human kinase constructs. We also demonstrate how the resulting expressing constructs can be used for the biophysical and biochemical study of clinical mutations by engineering a panel of 48 Src mutations and 46 Abl mutations via single-primer mutagenesis and screening the resulting library for expression yields. The wild-type kinase construct library is available publicly via addgene, and should prove to be of high utility for experiments focused on drug discovery and the emergence of drug resistance.

Introduction

Kinases play a critical role in cellular signaling pathways, controlling a number of key biological processes that often include growth and proliferation. There are over 500 kinases in the human genome (***Manning et al., 2002***; ***Hunter and Manning, 2015***), many of which are of therapeutic interest. Perturbations due to mutation, translocation, or upregulation can cause one or more kinases to become dysregulated, often with disastrous consequences (***Knight et al., 2010***). Kinase dysregulation has been linked to a number of diseases, such as cancer, diabetes, and inflammation. Cancer alone is the second leading cause of death in the United States, accounting for nearly 25% of all deaths; in 2015, over 1.7 million new cases were diagnosed, with over 580,000 deaths (***American Cancer Society, 2015***). Nearly 50% of cancer drug development is targeted at kinases, accounting for perhaps 30% of *all* drug development effort globally (***Cohen and Tcherpakov, 2010***; ***Santos et al., 2016***).

The discovery of imatinib, an inhibitor that targets the Abelson tyrosine kinase (Abl) dysregulated in chronic myelogenous leukemia (CML) patients, was transformative in revealing the enormous therapeutic potential of selective kinase inhibitors, kindling hope that this remarkable success could be recapitulated for other cancers and diseases (***Stegmeier et al., 2010***). While there are now 37 FDA-approved selective kinase small molecules inhibitors, these molecules were approved for targeting only 13 out of ∼500 human kinases (***Wu et al., 2015***; ***Roskoski Jr., 2017***), with the vast majority developed to target just a handful of kinases (***Volkamer et al., 2015***). The discovery of therapeutically effective inhibitors for other kinases has proven remarkably challenging.

While these inhibitors have found success in the clinic, many patients cease to respond to treatment due to resistance caused by mutations in the targeted kinase (***Pao et al., 2005***), activation of downstream kinases (***Knight et al., 2010***), or relief of feedback inhibition in signaling pathways (***Chandarlapaty et al., 2011***). These challenges have spurred the development of a new generation of inhibitors aimed at overcoming resistance (***Jia et al., 2016***; ***Politi et al., 2015***), as well as mutant-specific inhibitors that target kinases bearing a missense mutation that confers resistance to an earlier generation inhibitor (***Song et al., 2015***). The ability to easily engineer and express mutant kinase domains of interest would be of enormous aid in the development of mutant-selective inhibitors, offering an advantage over current high throughput platforms (***Karaman et al., 2008***; ***Davis et al., 2011***; ***Médard et al., 2015***), which typically include few clinically-observed mutant kinases.

Probing human kinase biochemistry, biophysics, and structural biology in the laboratory is essential to making rapid progress in understanding kinase regulation, engineering selective inhibitors, and studying the biophysical driving forces underlying mutational mechanisms of drug resistance. While human kinase expression in baculovirus-infected insect cells can achieve high success rates (***Chambers et al., 2004***; ***Wang et al., 2008***), it cannot compete in cost, convenience, or speed with bacterial expression. *E. coli* expression enables production of kinases without unwanted post-translational modifications, allowing for greater control of the system. A survey of 62 full-length non-receptor human kinases found that over 50% express well in *E. coli* (***Chambers et al., 2004***), but often expressing only the soluble kinase domains are sufficient, since these are the molecular targets of therapy for targeted kinase inhibitors and could be studied even for receptor-type kinases. While removal of regulatory domains can negatively impact expression and solubility, coexpression with phosphatase was shown to greatly enhance bacterial kinase expression in Src and Abl tyrosine kinases, presumably by ensuring that kinases remain in an unphosphorylated inactive form where they can cause minimal damage to cellular machinery (***Seeliger et al., 2005***).

The protein databank (PDB) now contains over 100 human kinases that—according to PDB header data—were expressed in bacteria. Many of these kinases were expressed and crystallized as part of the highly successful Structural Genomics Consortium (SGC) effort to increase structural coverage of the human kinome (***The Structural Genomics Consortium, 2012***). Since bacterial expression is often complicated by the need to tailor construct boundaries, solubility-promoting tags,

and expression and purification protocols individually for each protein expressed, we wondered whether a simple, uniform, automatable expression and purification protocol could be used to identify tractable kinases, select construct boundaries, and express a large number of human kinases and their mutant forms to produce a convenient bacterial expression library to facilitate kinase research and selective inhibitor development. To this end, we developed a structural informatics pipeline to use available kinase structural data and associated metadata to select constructs from available human kinase libraries to clone into a standard set of vectors intended for phosphatase coexpression under a simple automatable expression and purification protocol. Using an expression screen for multiple construct domain boundaries of Abl, we found that transferring construct boundaries from available structural data can produce constructs with useful expression levels, enabling simple identification of construct domain boundaries. We then completed an automated expression screen in Rosetta2 cells of 96 different kinases and found that 52 human kinase domains express with yields greater than 2 μg/mL culture, which should be adequate for biochemical, biophysical, screening, and structural biology studies. To demonstrate the utility of these constructs for probing the effect of clinical mutations on kinase structure and ligand binding, we subsequently screened 48 Src and 46 Abl mutations, finding that many clinically-derived mutant kinase domains can be expressed with useful yields in this uniform automated expression and purification protocol.

All source code, data, and wild type kinase plasmids associated with this project are freely available online:

- **Source code and data**: https://github.com/choderalab/kinase-ecoli-expression-panel
- **Interactive table of expression data**: http://choderalab.org/kinome-expression
- **Plasmids**: https://www.addgene.org/kits/chodera-kinase-domains

Results

Construct boundary choice impacts Abl kinase domain expression

To understand how alternative choices of expression construct boundaries can modulate bacterial expression of a human kinase domain, we carried out an expression screen of 84 unique construct boundaries encompassing the kinase domain of the tyrosine protein kinase ABL1.

Three constructs known to express in bacteria were chosen from the literature and used as controls, spanning Uniprot residues 229–500 (PDBID: 3CS9) (***Weisberg et al., 2005***), 229–512 (PDBID: 2G2H) (***Levinson et al., 2006***) and 229–515 (PDBID: 2E2B) (***Horio et al., 2007***). 81 constructs were generated combinatorially by selecting nine different N-terminal boundaries spanning residues 228–243 and nine different C-terminal boundaries spanning residues 490–515, chosen to be near the start and end points for the control constructs (Figure 1A). Each of the three control constructs included six replicates to provide an estimate of the typical standard error in expression readout for the experimental constructs, which was found to be between 0.42–1.5 μg/mL (Figure 1A, green constructs).

Briefly, the impact of construct boundary choice on Abl kinase domain expression was tested as follows (see Methods for full details). His10-TEV N-terminally tagged wild-type Abl constructs[1] were coexpressed with YopH phosphatase in a 96-well format with control replicates distributed randomly throughout the plate. His-tagged protein constructs were recovered via a single nickel affinity chromatography step, and construct yields were quantified using microfluidic capillary electrophoresis following thermal denaturation. Expression yields are summarized in Figure 1A, and a synthetic gel image from the constructs with detectable expression is shown in Figure 2. Abl construct bands are present at sizes between 29 and 35 kDa (due to the variation in construct boundaries), and YopH phosphatase (which is not His-tagged but has substantial affinity for the nickel beads) is present in all samples at its expected size of 50 kDa. Strikingly, despite the fact that N-terminal and C-terminal construct boundaries only varied over 15–25 residues, only a small number

[1] Parent plasmid is a pET His10 TEV LIC cloning vector and is available on Addgene (Plasmid #78173).

of constructs produced detectable expression (Figure 1B). As highlighted in Figure 1C (left), the best N-terminal boundaries (residues 228, 229, 230) are located on a disordered strand distant from any secondary structure; N-terminal boundaries closer to the beta sheet of the N-lobe gave poor or no detectable expression (Figure 1B). The best C-terminal construct boundaries (residues 511 and 512) occur in an α-helix (Figure 1C, right). Of note, this α-helix is not resolved in PDBID:2E2B (***Horio et al., 2007***), suggesting this structural element may only be weakly thermodynamically stable in the absence of additional domains. In other work, this alpha helix is shown to undergo a dramatic conformational change which introduces a kink at residue 516, splitting the helix into two (***Nagar et al., 2003***), indicating a high potential for flexibility in this region.

Surprisingly, two of the control constructs identified as constructs appearing in crystal structures in the PDB (which differ in construct boundary by only one or two residues) were in the top six expressing constructs (Figure 1A), and were in fact within 60% of the maximum observed expression yield. We concluded that transferring construct boundaries from existing kinase domain structural data would be sufficient to at least bias our constructs toward useful expression levels for a large-scale screen of multiple kinases.

Screen of 96 different kinases finds 52 with useful levels of expression in *E. coli*

To begin exploring which human kinase domains can achieve useful expression in *E. coli* using a simple automatable expression and purification protocol, a panel of kinase domain constructs for 96 kinases, for which bacterial expression has been previously demonstrated (determined from

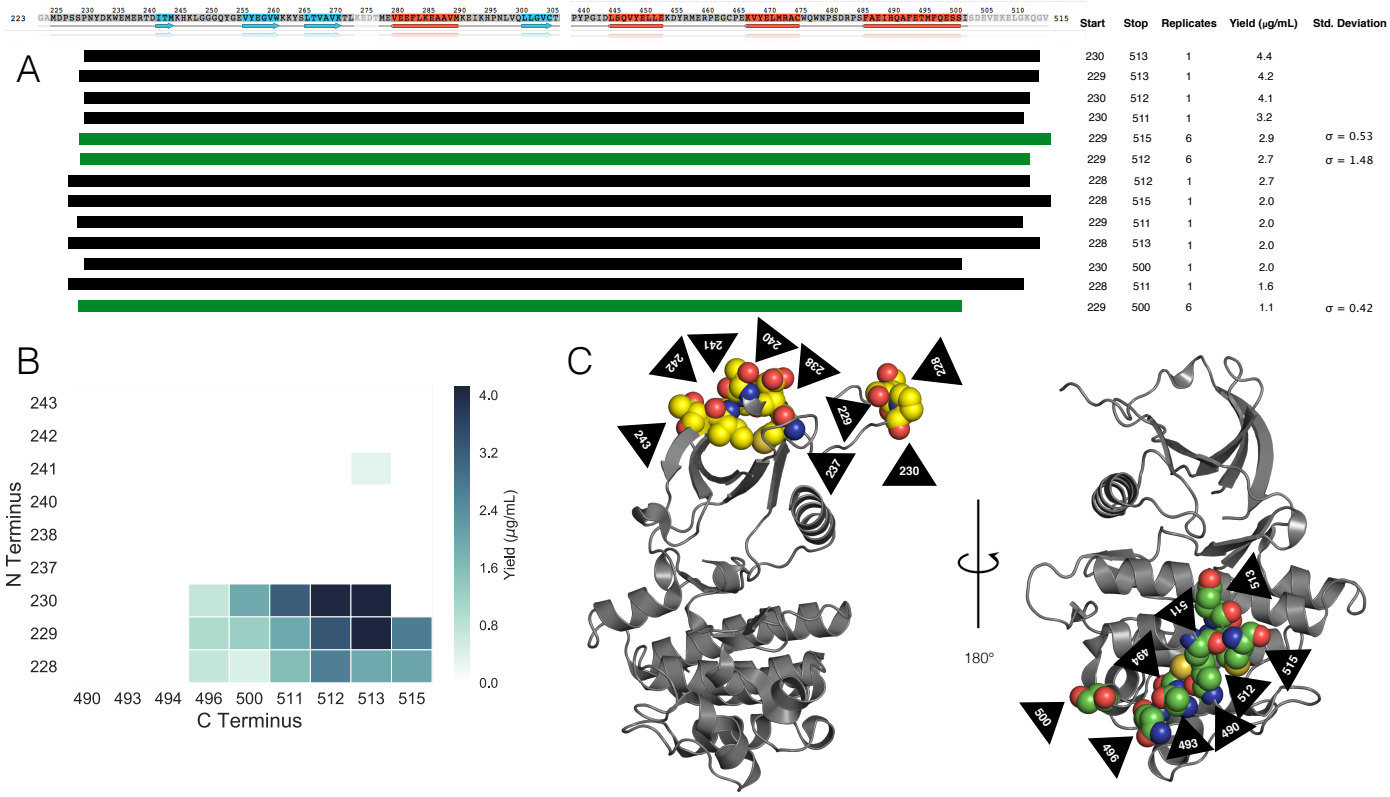

Figure 1. Abl kinase domain construct expression screen illustrates high sensitivity to construct boundaries. (**A**) Abl kinase domain construct boundaries with highest expression yields. Standard deviations of the yield are listed for control constructs for which six replicates were performed to give an indication of the uncertainty in experimental constructs. Secondary structure is indicated on the sequence. Beta sheets are colored blue and alpha helices are colored orange. (**B**) Heatmap showing average yields for constructs (in μg/mL culture) with detectable expression as a function of N- and C-terminal construct boundaries. (**C**) *left*: PDBID: 2E2B with the nine N-terminal construct boundary amino acids shown as yellow spheres. *right*: PDBID: 4XEY with the nine C-terminal construct boundary amino acids shown as green spheres. Black arrows indicate residue numbers.

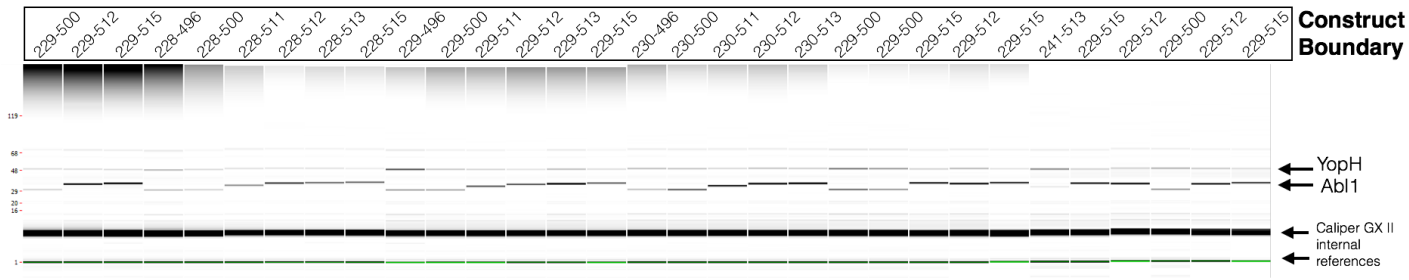

Figure 2. Expression yields of Abl kinase domain constructs for all constructs with detectable expression. A synthetic gel image rendering generated from Caliper GX II microfluidic gel electrophoresis data following Ni-affinity purification and thermal denaturation for all Abl constructs with detectable expression. Each well is marked with the Abl kinase domain construct residue boundaries (Uniprot canonical isoform numbering). Bands for YopH164 phosphatase (50 kDA) and Abl kinsase domain constructs (28–35 kDA) are labeled.

PDB header `EXPRESSION_SYSTEM` records) were selected using a semi-automated bioinformatics pipeline. Briefly, a database was built by querying Uniprot (***UniProt Consortium, 2017***) for human protein kinase domains that were both active and not truncated. This query returned a set of target sequences that were then matched to their relevant PDB constructs and filtered for expression system, discarding kinases that did not have any PDB entries with bacterial expression. As a final filtering step, the kinases were compared to three purchased kinase plasmid libraries (described in the Methods section), discarding kinases without a match. Construct boundaries were selected from PDB constructs and the SGC plasmid library, both of which have experimental evidence for *E. coli* expression, and subcloned from a plasmid in a purchased library. This process is explained in more detail in the Methods section. Selecting the kinases and their constructs for this expression trial in this method rested on the basis of expected success: these specific kinase constructs were bacterially expressed and purified to a degree that a crystal structure could be solved. While the expression protocols used to produce protein for crystallographic studies were likely tailored to maximize expression for individual proteins, we considered these kinases to have a high chance of expressing in our semi-automated pipeline where the *same* protocol is utilized for all kinases. Statistics of the number of kinases obtained from the PDB mining procedure are shown in Figure 3A. Surprisingly, the most highly sampled family was the CAMK family, suggesting that other researchers may have found this family particularly amenable to bacterial expression. Based on the results of the previous experiment scanning Abl constructs for expression, we decided to use construct boundaries that were reported in the literature for each kinase. This process resulted in a set of 96 plasmid constructs distributed across kinase families (Figure 3B).

From these, a set of 96 His10-TEV N-terminally tagged kinase domain constructs were generated, coexpressed with a phosphatase in *E. coli*, purified via nickel bead pulldown, and quantified using microfluidic gel electrophoresis. The 96 kinases were coexpressed with either Lambda phosphatase (for Ser/Thr kinases) or a truncated form of YopH phosphatase[2] (for Tyr kinases).

Instead of eluting with imidazole, purified kinase was cleaved off nickel beads by the addition of 10% TEV protease to minimize phosphatase contamination in the resulting eluate in order to assess whether resulting yields would be sufficient (and sufficiently free of phosphatase) to permit activity assays. While the initial panel of 96 kinases was well distributed amongst the kinase families (Figure 3B), the most highly expressing kinases (yield of more than 12 μg kinase/mL culture) were not evenly distributed (Figure 3D). 52 kinases demonstrated a useful level of soluble protein expression, here defined as greater than 2 μg/mL, naïvely expected to scale up to better than 2 mg/L culture (Table 1). Some kinases (shaded green in Table 1 demonstrated very high levels of expression, while others (shaded orange in Table 1 would likely benefit from further rounds of construct boundary optimization or solubility tags to boost soluble expression. The 17 most highly

[2]Yoph164 phosphatase, engineered to minimize intrinsic affinity for nickel purification resin by the QB3 MacroLab based on parent plasmid pCDFDuet1-YOPH, a gift from the Kuriyan Lab.

expressing kinases showed relatively high purity after elution, though we note that eluting via TEV site cleavage results in a quantity of TEV protease in the eluate (Figure 3C), but does not cause the elution of the His-tagged phosphatases.

Constructs with expression yields above 2 μg/mL have been made available via **Addgene**: https://www.addgene.org/kits/chodera-kinase-domains

Expressing clinically-derived Src and Abl mutants

The advent of next-generation sequencing technology has enabled generation of massive datasets rich with missense alterations in kinases observed directly in the clinic (***Varghese and Berger, 2014***; ***Zehir et al., 2017***; ***Garraway, 2013***), and has been particularly transformative in the field of oncology. To determine how well our kinase domain constructs and automated expression/purification protocol support the expression of clinically-identified kinase domain missense mutants for biophysical, biochemical, and structural characterization, we attempted to express 96 missense mutations mined from sequencing studies of cancer patients, both from publicly available sources as well a private clinical tumor sequencing dataset from Memorial Sloan Kettering (***Zehir et al., 2017***) sequenced as part of the MSK-IMPACT panel (***Cheng et al., 2015***), as identified via the MSK cBioPortal (***Cerami et al., 2012a***).

Similar to the expression screen above, a database was built focusing on the kinases we found to be expressible in *E. coli*. To add the mutation data, we retrieved public datasets from cBioPortal (***Cerami et al., 2012b***; ***Gao et al., 2013***) along with annotations from Oncotator (***Ramos et al., 2015***) through their respective web service APIs. We then added mutations and annotations

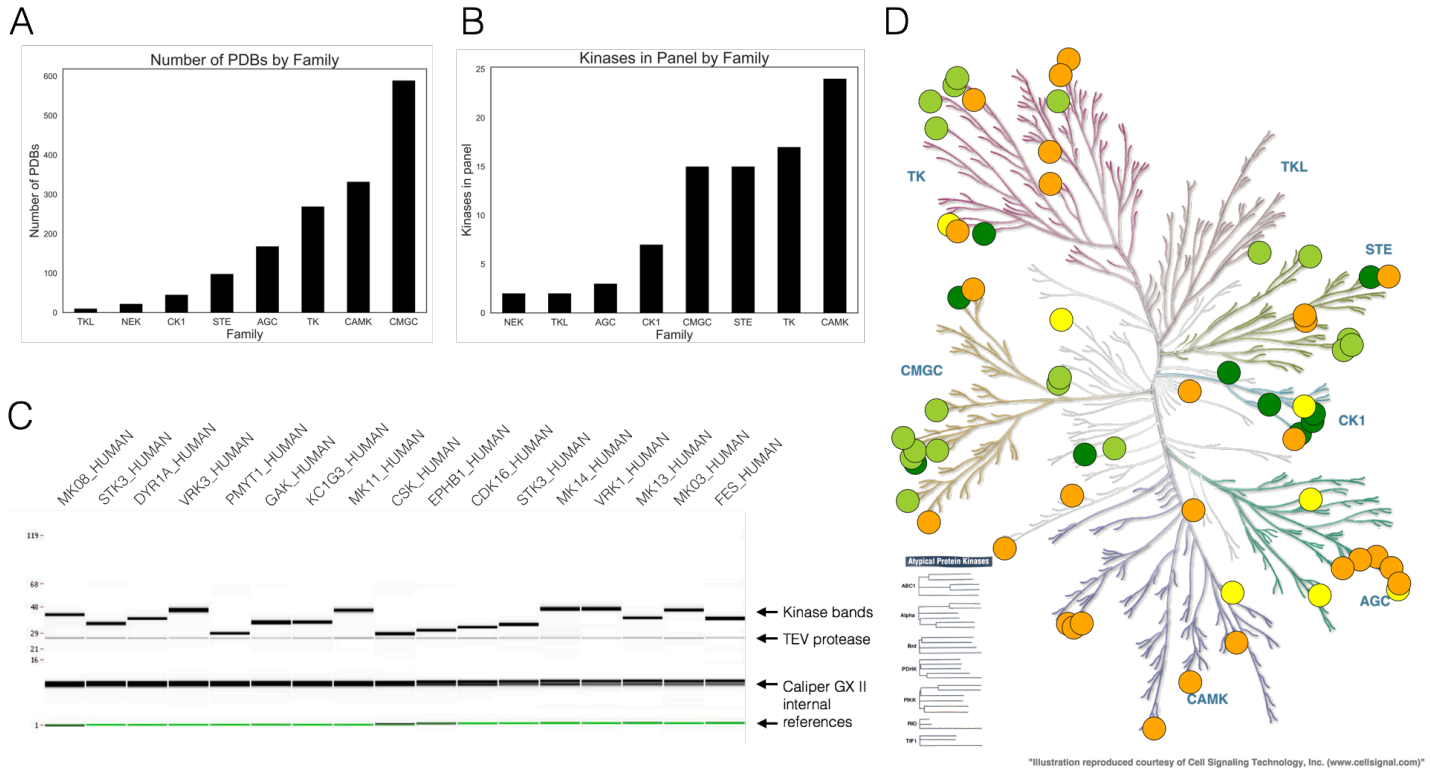

Figure 3. Kinome wide search for expressible kinases. (**A**) The number of PDB structures per kinase family, from the database built to select kinases for expression. (**B**) The distribution among familes of candidate kinases in our expression screen. (**C**) Caliper GX II synthetic gel image rendering of the highest expressing kinases, quantified using microfluidic capillary electrophoresis. (**D**) Kinome distribution of expression based on our 96 kinase screen. Dark green circles represent kinases with expression above 50 μg/mL culture yield. Light green circles represent kinases with expression between 50 and 12 μg/mL yield. Yellow circles represent kinases with expression between 12 and 7 μg/mL yield. Orange circles represent kinases with any expression (even below 2 μg/mL) up to 7 μg/mL yield. Image made with KinMap: http://www.kinhub.org/kinmap.

Table 1. Kinase domain constructs with yields >2 μg/mL culture for 96-kinase expression screen. Kinases are listed by Uniprot designation and whether they were co-expressed with Lambda or truncated YopH164 phosphatase. Yield (determined by Caliper GX II quantitation of the expected size band) reported in μg/mL culture, where total eluate volume was 120 μL from 900 μL bacterial culture. Yields are shaded green (yield > 12 μg/mL), yellow (12 > yield > 7 μg/mL) and orange (yield <7 μg/mL); kinase domain constructs with yields that were undetectable or < 2 μg/mL are not listed. ‡ denotes that the second kinase domain of KS6A1_HUMAN was expressed; all other kinases were the first or only kinase domain occurring in the ORF. Construct boundaries are listed in UniProt residue numbering for the UniProt canonical isoform. An interactive table of expression yields and corresponding constructs is available at http://choderalab.org/kinome-expression

Kinase	Construct Boundary	Plasmid Source and ID	Phosphatase	Yield (μg/mL)
MK14_HUMAN	1–360	Addgene 23865	Lambda	**70.7**
VRK3_HUMAN	24–352	SGC Oxford VRK3A-c016	Lambda	**67.5**
GAK_HUMAN	24–359	SGC Oxford GAKA-c006	Lambda	**64.7**
CSK_HUMAN	186–450	Addgene 23941	YopH	**62.5**
VRK1_HUMAN	3–364	Addgene 23496	Lambda	**62.3**
KC1G3_HUMAN	24–351	SGC Oxford CSNK1G3A-c002	Lambda	**56.3**
FES_HUMAN	448–822	Addgene 23876	YopH	**44.0**
PMYT1_HUMAN	24–311	SGC Oxford PKMYT1A-c004	Lambda	**38.0**
MK03_HUMAN	1–379	Addgene 23509	Lambda	**36.4**
STK3_HUMAN	16–313	Addgene 23818	Lambda	**34.3**
DYR1A_HUMAN	24–382	SGC Oxford DYRK1AA-c004	Lambda	**34.1**
KC1G1_HUMAN	24–331	SGC Oxford CSNK1G1A-c013	Lambda	**34.1**
MK11_HUMAN	24–369	SGC Oxford MAPK11A-c007	Lambda	**31.7**
MK13_HUMAN	1–352	Addgene 23739	Lambda	**31.7**
EPHB1_HUMAN	602–896	Addgene 23930	YopH	**28.9**
MK08_HUMAN	1–363	HIP pJP1520 HsCD00038084	Lambda	**28.5**
CDK16_HUMAN	163–478	Addgene 23754	Lambda	**26.9**
EPHB2_HUMAN	604–898	HIP pJP1520 HsCD00038588	YopH	**25.1**
PAK4_HUMAN	291–591	addgene 23713	Lambda	**23.9**
CDKL1_HUMAN	2–304	SGC Oxford CDKL1A-c024	Lambda	**23.2**
SRC_HUMAN	254–536	Addgene 23934	YopH	**22.0**
STK16_HUMAN	24–316	SGC Oxford STK16A-c002	Lambda	**20.7**
MAPK3_HUMAN	33–349	Addgene 23790	Lambda	**18.8**
PAK6_HUMAN	383–681	Addgene 23833	Lambda	**18.0**
CSK22_HUMAN	1–334	HIP pJP1520 HsCD00037966	Lambda	**17.9**
MERTK_HUMAN	570–864	Addgene 23900	YopH	**16.8**
PAK7_HUMAN	24–318	SGC Oxford PAK5A-c011	Lambda	**14.7**
CSK21_HUMAN	1–335	Addgene 23678	Lambda	**14.5**
EPHA3_HUMAN	606–947	Addgene 23911	YopH	**14.1**
BMPR2_HUMAN	1–329	SGC Oxford BMPR2A-c019	Lambda	**14.1**
M3K5_HUMAN	659–951	HIP pJP1520 HsCD00038752	Lambda	**14.0**
KCC2G_HUMAN	24–334	SGC Oxford CAMK2GA-c006	Lambda	**13.3**
E2AK2_HUMAN	254–551	HIP pJP1520 HsCD00038350	Lambda	**11.6**
MK01_HUMAN	1–360	HIP pJP1520 HsCD00038281	Lambda	**11.2**
CSKP_HUMAN	1–340	HIP pJP1520 HsCD00038384	Lambda	**10.1**
CHK2_HUMAN	210–531	Addgene 23843	Lambda	**8.1**
KC1G2_HUMAN	4–312	SGC Oxford CSNK1G2A-c002	Lambda	**7.6**
DMPK_HUMAN	2 4–433	SGC Oxford DMPK1A-c026	Lambda	**7.6**
KCC2B_HUMAN	11–303	Addgene 23820	Lambda	**7.1**
FGFR1_HUMAN	456–763	Addgene 23922	YopH	**6.1**
KS6A1_HUMAN‡	413–735	SGC Oxford RPS6KA1A-c036	Lambda	**5.7**
DAPK3_HUMAN	9–289	Addgene 23436	Lambda	**4.0**
STK10_HUMAN	18–317	HIP pJP1520 HsCD00038077	Lambda	**3.7**
KC1D_HUMAN	1–294	Addgene 23796	Lambda	**3.7**
KC1E_HUMAN	1–294	Addgene 23797	Lambda	**3.5**
NEK1_HUMAN	23–350	SGC Oxford NEK1A-c011	Lambda	**3.3**
CDK2_HUMAN	1–297	Addgene 23777	Lambda	**3.1**
ABL1_HUMAN	229–512	HIP pJP1520 HsCD00038619	YopH	**2.5**
DAPK1_HUMAN	2–285	HIP pJP1520 HsCD00038376	Lambda	**2.4**
DYRK2_HUMAN	23–417	SGC Oxford DYRK2A-c023	Lambda	**2.4**
HASP_HUMAN	24–357	SGC Oxford GSG2A-c009	Lambda	**2.3**
FGFR3_HUMAN	449–759	Addgene 23933	YopH	**2.3**

from the private dataset at MSKCC by extracting the mutations from a local copy of the dataset and retrieving annotations from Oncotator. The annotated mutations were filtered for mutations that occurred within the construct boundaries of the kinase domains for our expressible kinases. We found 63 clinical mutations appearing within our kinase domain construct boundaries in Abl and 61 in Src. We subsequently selected 48 mutants for Abl and 46 for Src to express, aiming for a panel of mutants that were distributed throughout the entire kinase domain (Figure 4A). Wild type Src

and Abl were included as controls. The mutants were engineered into the Src and Abl plasmids from the 96 kinase expression screen using mutagenesis, and coexpressed in *E. coli* with YopH164 phosphatase. Mutants with expression yields above 2μg kinase/mL culture are shown in Table 2, while the 3D distribution for all of the mutants is shown in Figure 4A and colored by expression level. A synthetic gel image produced from microfluidic gel electrophoresis of thermally denatured protein eluate for all of the mutants is shown in Figure 4B, giving an indication of relative purity after eluting with TEV protease. While the vast majority of the Src mutants expressed at a usable level, many of the Abl mutants expressed below the 2μg/mL threshold. This can primarily be attributed to the low level of expression for wild-type Abl construct (Table 1). In instances where kinase activity is not required, this yield could be increased via the introduction of inactivating mutations (***Seeliger et al., 2005***) or further tailoring of expression and purification protocols.

Methods

Semi-automated selection of kinase construct sequences for *E. coli* expression

Selection of human protein kinase domain targets

Human protein kinases were selected by querying the UniProt API (query date 30 May 2014) for any human protein with a domain containing the string "protein kinase", and which was manually annotated and reviewed (i.e. a Swiss-Prot entry). The query string used was:

```
taxonomy:"Homo sapiens (Human) [9606]" AND domain:"protein kinase" AND reviewed:yes
```

Data was returned by the UniProt API in XML format and contained protein sequences and relevant PDB structures, along with many other types of genomic and functional information. To select active protein kinase domains, the UniProt domain annotations were searched using the regular expression `^Protein kinase(?!; truncated)(?!; inactive)`, which excludes certain domains annotated "Protein kinase; truncated" and "Protein kinase; inactive". Sequences for the selected domains were then stored. The sequences were derived from the canonical isoform as determined by UniProt.

Matching target sequences with relevant PDB constructs

Each target kinase gene was matched with the same gene in any other species where present, and UniProt data was downloaded for those genes also. The UniProt data included a list of PDB structures which contain the protein, as well as their sequence spans in the coordinates of the UniProt canonical isoform. This information was used to filter out PDB structures which did not include the protein kinase domain; structures were kept if they included the protein kinase domain sequence and truncated fewer than 30 residues at each end. PDB coordinate files were then downloaded for each PDB entry. The coordinate files contain various metadata, including an `EXPRESSION_SYSTEM` annotation, which was used to filter PDB entries to keep only those which include the phrase "ESCHERICHIA COLI". The majority of PDB entries returned had an `EXPRESSION_SYSTEM` tag of "ESCHERICHIA COLI", while a small number had "ESCHERICHIA COLI BL21" or "ESCHERICHIA COLI BL21(DE3)".

The PDB coordinate files also contain SEQRES records, which should contain the protein sequence used in the crystallography or NMR experiment. According to the PDB File Format FAQ (http://deposit.rcsb.org/format-faq-v1.html), "All residues in the crystal or in solution, including residues not present in the model (i.e., disordered, lacking electron density, cloning artifacts, HIS tags) are included in the SEQRES records." However, we found that these records are very often misannotated, instead representing only the crystallographically resolved residues. Since expression levels can be greatly affected by insertions or deletions of only one or a few residues at either terminus (***Klock et al., 2008***), it is important to know the full experimental sequence, and we thus needed a way to measure the authenticity of a given SEQRES record. We developed a crude measure by hypothesizing that a) most crystal structures would be likely to have at least one or a few unresolved residues at one or both termini and b) the presence of an expression tag (which is typically not crystallographically resolved) would indicate an authentic SEQRES record. To achieve

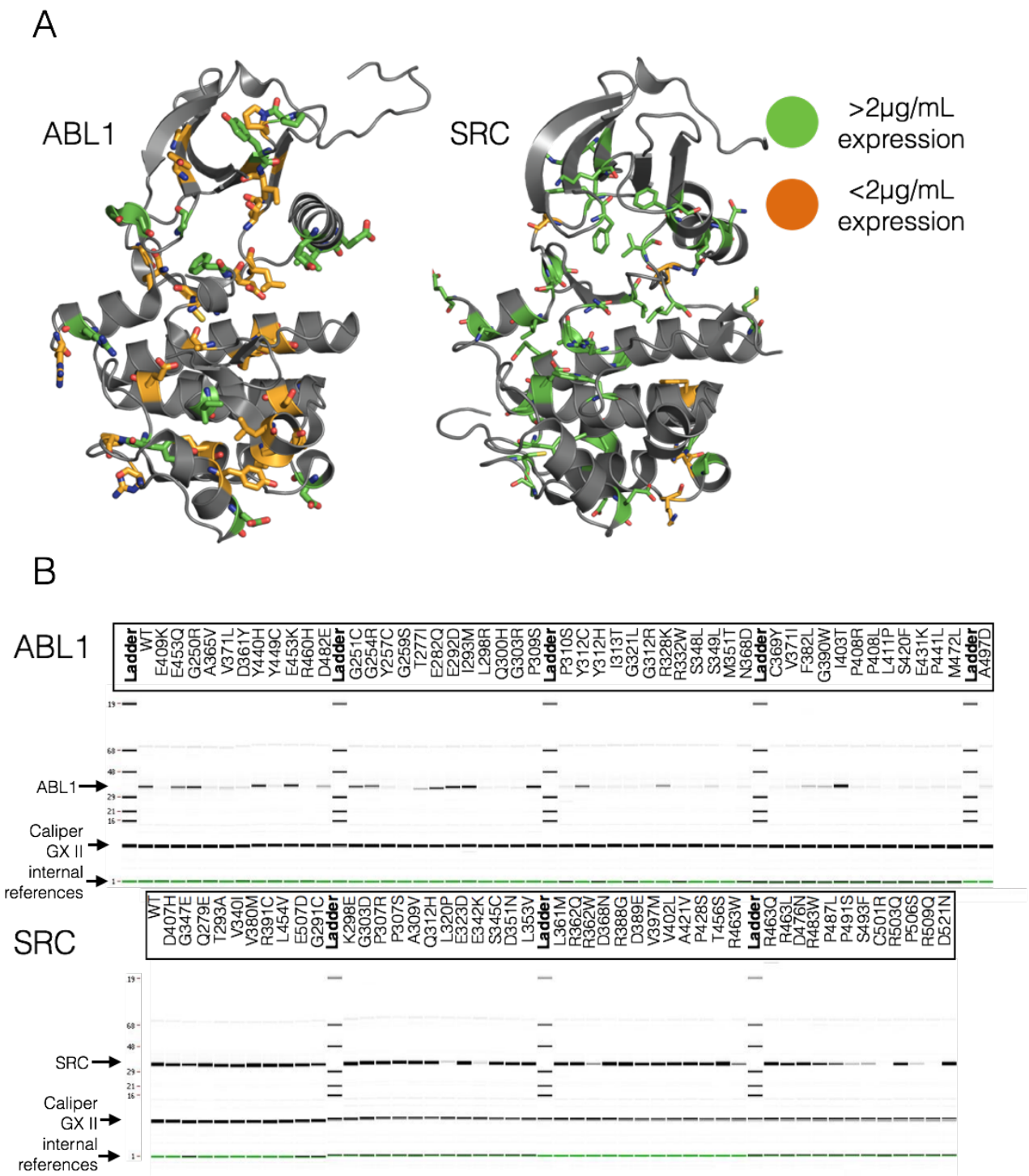

Figure 4. Expression yields for engineered clinically-derived Src and Abl missense mutants. (**A**) All Abl and Src clinically-identified mutants assessed in the expression screen are displayed as sticks. Mutants with expression yields >2μg/mL are colored green, while those with yields <2μg/mL are colored orange. Rendered structures are Abl (PDBID: 2E2B) and Src (PDBID: 4MXO) (***Levinson and Boxer, 2014***). (**B**) Synthetic gel images showing ABl (*top*) or Src (*bottom*) expression, with wells labeled by missense mutation. Yield was determined by Caliper GX II quantitation of the expected size band and reported in μg/mL culture, where total eluate volume was 120 μL following nickel bead pulldown purification from 900 μL bacterial culture. Residue mutations use numbering for the Uniprot canonical isoform.

Table 2. Expression yields for engineered clinical missense mutants of Src and Abl kinase domains with yields > 2 μg/mL culture. Src and Abl kinase domain constructs with engineered clinical mutations with expression yields > 2μg/mL culture are listed, sorted by yield. Yield was determined by Caliper GX II quantitation of the expected size band and reported in μg/mL culture, where total eluate volume was 80 μL purified from 900 μL bacterial culture. Wild-type (WT) controls for both Src and Abl (here, a single well for each) are shown as the first entry for each gene.

Abl1 (229–512)	Mutation[1]	Functional Impact Score[2]	yield (μg/mL)	% of WT expression
	WT	–	5.1	–
	I403T	Low	17.8	350
	I293M	Low	9.8	193
	P309S	Neutral	7.8	153
	E453K	Low	7.3	144
	Y440H	Medium	7.1	140
	E292D	Low	6.9	135
	G251C	High	5.2	102
	E282Q	Neutral	5.1	102
	G250R	Neutral	5.1	100
	G254R	High	5.0	98
	Y312C	Neutral	4.7	93
	E453Q	Low	3.7	73
	R328K	Low	3.5	69
	D482E	Neutral	2.5	49
	F382L	Medium	2.1	41
	G390W	Medium	2.1	41
Src (254–536)	**Mutation[1]**	**Functional Impact Score[2]**	**yield (μg/mL)**	**% of WT expression**
	WT	–	35.7	–
	T456S	Neutral	80.9	227
	R388G	Medium	61.5	172
	K298E	High	54.5	153
	V380M	Neutral	51.7	145
	D368N	Neutral	49.9	140
	D521N	Low	42.8	120
	R463Q	Neutral	38.4	108
	R391C	Neutral	37.5	105
	E323D	Low	37.2	104
	A309V	Low	35.q	98
	G303D	Neutral	34.1	96
	R362Q	Neutral	33.6	94
	L361M	Medium	31.7	89
	A421V	Neutral	30.7	86
	V402L	Neutral	30.6	86
	V397M	Medium	29.8	84
	Q278E	Neutral	29.6	83
	Q312H	Low	29.5	83
	L353V	Medium	29.0	81
	L454V	Neutral	29.0	81
	P307R	Neutral	28.6	80
	V340I	Low	28.0	78
	P307S	Neutral	24.2	68
	D476N	Neutral	23.3	65
	D351N	Neutral	22.9	64
	T293A	Neutral	22.2	62
	S345C	Low	22.2	62
	P428S	Medium	22.2	62
	E507D	Neutral	20.7	58
	D389E	High	20.0	56
	R503Q	Neutral	17.3	49
	D407H	High	15.9	45
	R463L	Neutral	14.9	42
	G291C	Medium	11.9	33
	G347E	Medium	10.2	29
	R483W	High	9.8	27
	P487L	Medium	6.0	17
	R463W	Medium	5.2	15
	R362W	Low	3.9	11
	S493F	Low	3.0	8
	P491S	Low	2.2	6

[1] Uniprot amino acid sequence numbering of primary isoform
[2] MutationAssesor Score (***Reva et al., 2007, 2011***), which predicts functional impact via conservation

this, unresolved residues were first defined by comparing the SEQRES sequence to the resolved sequence, using the SIFTS service to determine which residues were not present in the canonical

isoform sequence (***Velankar et al., 2013***). Regular expression pattern matching was used to detect common expression tags at the N- or C-termini. Sequences with a detected expression tag were given a score of 2, while those with any unresolved sequence at the termini were given a score of 1, and the remainder were given a score of 0. This data was not used to filter out PDB structures at this stage, but was stored to allow for subsequent selection of PDB constructs based on likely authenticity. Also stored for each PDB sequence was the number of residues extraneous to the target kinase domain, and the number of residue conflicts with the UniProt canonical isoform within that domain span.

Plasmid libraries

As a source of kinase DNA sequences, we purchased three kinase plasmid libraries: the addgene Human Kinase ORF kit , a kinase library from the Structural Genomics Consortium (SGC), Oxford (http://www.thesgc.org), and a kinase library from the PlasmID Repository maintained by the Dana-Farber/Harvard Cancer Center. The aim was to subclone the chosen sequence constructs from these plasmids, though we did not use the same vectors. Annotated data for the kinases in each library was used to match them against the human protein kinases selected for this project. The plasmid open reading frames (ORFs) were translated into protein sequences and aligned against the target kinase domain sequences from UniProt. Also calculated were the number of extraneous protein residues in the ORF, relative to the target kinase domain sequence, and the number of residue conflicts with the UniProt sequence.

Selection of sequence constructs for expression

Of the kinase domain targets selected from UniProt, we filtered out those with no matching plasmids from our available plasmid libraries and/or no suitable PDB construct sequences. For this purpose, a suitable PDB construct sequence was defined as any with an authenticity score greater than zero, i.e., those derived from SEQRES records with no residues outside the span of the resolved structure. Plasmid sequences and PDB constructs were aligned against each target domain sequence, and various approaches were then considered for selecting (a) the sequence construct to use for each target, and (b) the plasmid to subclone it from. Candidate sequence constructs were drawn from two sources: PDB constructs and the SGC plasmid library. The latter sequences were included because the SGC plasmid library was the only one of the three libraries which had been successfully tested for *E. coli* expression.

For most of the kinase domain targets, multiple candidate sequence constructs were available. To select the most appropriate sequence construct, we sorted them first by authenticity score, then by the number of conflicts relative to the UniProt domain sequence, then by the number of residues extraneous to the UniProt domain sequence span. The top-ranked construct was then chosen. In cases where multiple plasmids were available, these were sorted first by the number of conflicts relative to the UniProt domain sequence, then by the number of residues extraneous to the UniProt domain sequence span, and the top-ranked plasmid was chosen. This process resulted in a set of 96 kinase domain constructs, which (by serendipity) matched the 96-well plate format we planned to use for parallel expression testing. We therefore selected these construct sequences for expression testing.

An interactive table of the selected plasmids, constructs, and aligned PDB files can be viewed at http://choderalab.org/kinome-expression.

Automation of the construct selection process

While much of this process was performed programmatically, many steps required manual supervision and intervention to correct for exceptional cases. While these exceptions were encoded programmatically as overrides to ensure the scheme could be reproduced from existing data, we hope to eventually develop a fully automated software package for the selection of expression construct sequences for a given protein family, but this was not possible within the scope of this work.

Mutagenesis protocol

Point mutations were introduced with a single-primer QuikChange reaction. Primers were designed to anneal at 55°C both upstream and downstream of the point mutation, and with a total length of approximately 40 bases. At the codon to be modified, the fewest possible number of bases was changed. Plasmid template (160 ng) was mixed with 1 μM primer in 1x PfuUltra reaction buffer, with 0.8 mM dNTPs (0.2 mM each) and 1 U PfuUltra High-Fidelity DNA polymerase (Agilent), in a total volume of 20 μL. Thermocycler settings were 2 minutes at 95°C, followed by 18 cycles of 20s at 95°C, 1 min at 53°C, 12 minutes at 68°C (2min/kb), then 1 minute at 68°C. After cooling to room temperature, 4 μL of the PCR reaction was added to 16 μL CutSmart Buffer (NEB) containing 10 U DpnI (NEB). After incubation for 2.5 hours at 37°C, 6 μL of this mixture was used to directly transform XL1-Blue chemically competent cells (Agilent) according to the manufacturer's protocol. Transformants were picked for plasmid mini-preps and the presence of the point mutations was confirmed by sequencing.

Expression testing

For each target, the selected construct sequence was subcloned from the selected DNA plasmid. Expression testing was performed at the QB3 MacroLab (QB3 MacroLab, University of California, Berkeley, CA 94720) [http://qb3.berkeley.edu/macrolab/], a core facility offering automated gene cloning and recombinant protein expression and purification services.

Each kinase domain was tagged with a N-terminal His10-TEV and coexpressed with either the truncated YopH164 for Tyr kinases or lambda phosphatase for Ser/Thr kinases. All construct sequences were cloned into the 2BT10 plasmid, an AMP resistant ColE1 plasmid with a T7 promoter, using ligation-independent cloning (LIC). The inserts were generated by PCR using the LICv1 forward (TACTTCCAATCCAATGCA) and reverse (TTATCCACTTCCAATGTTATTA) tags on the primers. Gel purified PCR products were LIC treated with dCTP. Plasmid was linearized, gel purified, and LIC-treated with dGTP. LIC-treated plasmid and insert were mixed together and transformed into XL1-Blues for plasmid preps.

Expression was performed in Rosetta2 cells (Novagen) grown with Magic Media (Invitrogen autoinducing medium), 100 μg/mL of carbenicillin and 100 μg/mL of spectinomycin. Single colonies of transformants were cultivated with 900 μL of MagicMedia into a gas permeable sealed 96-well block. The cultures were incubated at 37°C for 4 hours and then at 16°C for 40 hours while shaking. Next, cells were centrifuged and the pellets were frozen at -80°C overnight. Cells were lysed on a rotating platform at room temperature for an hour using 700 μL of SoluLyse (Genlantis) supplemented with 400 mM NaCl, 20 mM imidazole, 1 μg/mL pepstatin, 1 μg/mL leupeptin and 0.5 mM PMSF.

For protein purification, 500 μL of the soluble lysate was added to a 25 μL Protino Ni-NTA (Machery-Nagel) agarose resin in a 96-well filter plate. Nickel Buffer A (25 mM HEPES pH 7.5, 5% glycerol, 400 mM NaCl, 20 mM imidazole, 1 mM BME) was added and the plate was shaken for 30 minutes at room temperature. The resin was washed with 2 mL of Nickel Buffer A. Target proteins were eluted by a 2 hour incubation at room temperature with 10 μg of TEV protease in 80 μL of Nickel Buffer A per well and a subsequent wash with 40 μL of Nickel Buffer A to maximize protein release. Nickel Buffer B (25 mM HEPES pH 7.5, 5% glycerol, 400 mM NaCl, 400 mM imidazole, 1 mM BME) was used to elute TEV resistant material remaining on the resin. Untagged protein eluted with TEV protease was run on a LabChip GX II Microfluidic system to analyze the major protein species present.

Discussion

We have demonstrated that a simple, uniform, automatable protocol is able to achieve useful bacterial expression yields for a variety of kinase domain constructs. While yields could likely be improved further by a variety of methods—such as tailoring the protocol to individual kinases, the addition of solubility-promoting tags, construct domain boundary and codon optimization,

or mutations to improve the solubility or ablate catalytic activity—the simplicity of this approach suggests widespread utility of automated bacterial expression for biophysical, biochemical, and structural biology work for the further study of human kinase domains.

Our expression test of 81 different construct boundaries of the Abl kinase domain demonstrated a surprising sensitivity of expression yields to the precise choice of boundary. This sensitivity may be related to where the construct is truncated with respect to the secondary structure of the protein. Disrupting secondary structure could cause the protein to improperly fold, leading to low soluble protein yield even when total expression is high. Of note, the best C-terminal boundaries for Abl were residues 511 and 512. These residues fall in the regulatory alpha helix αI (***Nagar et al., 2003***). This helix has been shown to undergo a dramatic conformational change upon binding to the myristoylated N-terminal cap, which introduces a sharp "kink" in residues 516–519. These residues may lead to higher levels of soluble expression by truncating an secondary structural element that is unusually flexible. Indeed, this helix is not resolved in some X-ray structures (PDBID:2E2B) (***Horio et al., 2007***), further suggesting that this helix is less thermodynamically stable than expected. Control replicates of three constructs indicate good repeatability of expression yields in the high-throughput format. This suggests that optimization of construct boundaries could potentially further greatly increase yields of kinase domains found to express > 2 μg/mL culture. Codon optimization for bacterial expression could also increase expression for kinase domains with low yield due to codon bias (***Sørensen and Mortensen, 2005***).

The tolerance of these bacterial constructs to many engineered clinical missense mutations suggests a promising route to the high-throughput biophysical characterization of the effect of clinical mutants on anticancer therapeutics. Mutations that did not express well may destabilize the protein, or may increase the specific activity of the kinase. A higher specific activity would require more phosphatase activity, wasting ATP to prevent high levels of phosphorylation that have been hypothesized to cause difficulty expressing kinases without a coexpressed phosphatase in bacteria (***Seeliger et al., 2005***). Mutations that are destabilizing may show improved expression if coexpressed with more elaborate chaperones such as GroEL and Trigger factor (***Haacke et al., 2009***). Mutations that increase the specific activity of the kinase might also express better when combined with an inactivating mutation.

High-throughput automated kinase expression could be combined with enzymatic or biophysical techniques for characterizing the potency of a variety of clinical kinase inhibitors to assess which mutations confer resistance or sensitivity. While the process of engineering, expressing, purifying, and assaying mutants currently takes approximately two weeks, it is possible that new techniques for cell-free bacterial expression (***Kim and Swartz, 1999***; ***Sawasaki et al., 2002***) may reduce this time to a matter of days or hours in a manner that might be compatible with clinical time frames to impact therapeutic decision-making.

We hope that other laboratories find these resources useful in their own work.

Acknowledgments

DLP, SMH, LRL, SKA, MI, and JDC acknowledge support from the Sloan Kettering Institute. This work was funded in part by the Marie-Josée and Henry R. Kravis Center for Molecular Oncology, the National Institutes of Health (NIH grant R01 GM121505 and National Cancer Institute Cancer Center Core grant P30 CA008748), and a Louis V. Gerstner Young Investigator Award. MAS acknowledges funding support by NIH grant R35 GM119437. The authors are grateful to Gregory Ross (MSKCC) for assistance in preparing the computational infrastructure for selecting clinical point mutants, and to Sarah E. Boyce (current address: Schrödinger, New York, NY) for assistance with multiple stages of this project. We gratefully acknowledge the members of the MSKCC Molecular Diagnostics Service in the Department of Pathology for their efforts in collecting and compiling mutations for Abl and Src kinases used here. The authors are grateful to addgene for their help in making the plasmids generated by this work available to the research community at minimal cost.

References

American Cancer Society, Cancer Facts & Figures 2015. American Cancer Society; 2015.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. Cancer Discovery. 2012 May; 2(5):401–404. doi: 10.1158/2159-8290.CD-12-0095.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery. 2012 May; 2(5):401–404.

Chambers SP, Austen DA, Fulghum JR, Kim WM. High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells. Protein Expression and Purification. 2004; 36:40–47. doi: 10.1016/j.pep.2004.03.003.

Chandarlapaty S, Sawai A, Scaltriti M, Rodrik-Outmezguine V, Grbovic-Huezo O, Serra V, Majumder PK, Baselga J, Rosen N. AKT Inhibition Relieves Feedback Suppression of Receptor Tyrosine Kinase Expression and Activity. Cancer Cell. 2011 Jan; 19(1):58–71. http://linkinghub.elsevier.com/retrieve/pii/S1535610810004332, doi: 10.1016/j.ccr.2010.10.031.

Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT). The Journal of Molecular Diagnostics. 2015 May; 17(3):251–264. doi: 10.1016/j.jmoldx.2014.12.006.

Cohen P, Tcherpakov M. Will the Ubiquitin System Furnish as Many Drug Targets as Protein Kinases? Cell. 2010 Nov; 143(5):686–693. http://linkinghub.elsevier.com/retrieve/pii/S0092867410012961, doi: 10.1016/j.cell.2010.11.016.

Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. Nature biotechnology. 2011 Nov; 29(11):1046–1051.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Science signaling. 2013 Apr; 6(269):pl1–pl1.

Garraway LA. Genomics-driven oncology: framework for an emerging paradigm. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013 May; 31(15):1806–1814.

Haacke A, Fendrich G, Ramage P, Geiser M. Chaperone Over-Expression in Escherichia Coli: Apparent Increased Yields of Soluble Recombinant Protein Kinases Are Due Mainly to Soluble Aggregates. Protein Expr Purif. 2009 Apr; 64(2):185–193. doi: 10.1016/j.pep.2008.10.022.

Horio T, Hamasaki T, Inoue T, Wakayama T. Structural factors contributing to the Abl/Lyn dual inhibitory activity of 3-substituted benzamide derivatives. Bioorganic & medicinal . . . . 2007; .

Hunter T, Manning G. The Eukaryotic Protein Kinase Superfamily and the Emergence of Receptor Tyrosine Kinases. In: Wheeler DL, Yarden Y, editors. *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Disease* New York, NY: Springer New York; 2015.p. 1–15.

Jia Y, Yun CH, Park E, Ercan D, Manuia M, Juarez J, Xu C, Rhee K, Chen T, Zhang H, Palakurthi S, Jang J, Lelais G, DiDonato M, Bursulaya B, Michellys PY, Epple R, Marsilje TH, McNeill M, Lu W, et al. Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. Nature. 2016 Jun; 534(7605):129–132.

Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, Chan KW, Ciceri P, Davis MI, Edeen PT, Faraoni R, Floyd M, Hunt JP, Lockhart DJ, Milanov ZV, Morrison MJ, Pallares G, Patel HK, Pritchard S, Wodicka LM, et al. A quantitative analysis of kinase inhibitor selectivity. Nature Biotech. 2008; 26:127–132.

Kim DM, Swartz JR. Prolonging Cell-Free Protein Synthesis with a Novel ATP Regeneration System. Biotechnol Bioeng. 1999; 66(3):180–188. doi: 10.1002/(SICI)1097-0290(1999)66:3<180::AID-BIT6>3.0.CO;2-S.

Klock HE, Koesema EJ, Knuth MW, Lesley SA. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins: Structure, Function, and Bioinformatics. 2008 May; 71(2):982–994. doi: 10.1002/prot.21786.

Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. Nature Reviews Cancer. 2010 Feb; 10(2):130–137. http://www.nature.com.proxy.library.cornell.edu/nrc/journal/v10/n2/full/nrc2787.html, doi: 10.1038/nrc2787.

Levinson NM, Boxer SG. A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. Nature chemical biology. 2014 Feb; 10(2):127–132.

Levinson NM, Kuchment O, Shen K, Young MA, Koldobskly M, Karplus M, Cole PA, Kuriyan J. A Src-like inactive conformation in the Abl tyrosine kinase domain. Plos Biology. 2006; 4:e144.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002; 298:1912–1934.

Médard G, Pachl F, Ruprecht B, Klaeger S, Heinzlmeir S, Helm D, Qiao H, Ku X, Wilhelm M, Kuehne T, Wu Z, Dittmann A, Hopf C, Kramer K, Kuster B. Optimized chemical proteomics assay for kinase inhibitor profiling. Journal of proteome research. 2015 Mar; 14(3):1574–1586.

Nagar B, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W, Clarkson B, Superti-Furga G, Kuriyan J. Structural basis for the autoinhibition of c-Abl tyrosine kinase. Cell. 2003 Mar; 112(6):859–871.

Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MG Maureen F nd Kris, Varmus H. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS medicine. 2005 Mar; 2(3):e73.

Politi K, Ayeni D, Lynch T. The Next Wave of EGFR Tyrosine Kinase Inhibitors Enter the Clinic. Cancer Cell. 2015 Jun; 27(6):751–753.

Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, Meyerson M, Getz G. Oncotator: cancer variant annotation tool. Human mutation. 2015 Apr; 36(4):E2423–9.

Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. . 2007; 8(11):R232. http://genomebiology.biomedcentral.com/articles/10.1186/gb-2007-8-11-r232, doi: 10.1186/gb-2007-8-11-r232.

Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Research. 2011; 39(17):e118. http://dx.doi.org/10.1093/nar/gkr407, doi: 10.1093/nar/gkr407.

Roskoski Jr R, USFDA approved protein kinase inhibitors; 2017. http://www.brimr.org/PKI/PKIs.htm, updated 3 May 2017.

Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP. A Comprehensive Map of Molecular Drug Targets. Nat Rev Drug Discov. 2016 Dec; 16(1):19–34. doi: 10.1038/nrd.2016.230.

Sawasaki T, Ogasawara T, Morishita R, Endo Y. A Cell-Free Protein Synthesis System for High-Throughput Proteomics. Proc Natl Acad Sci. 2002 Nov; 99(23):14652–14657. doi: 10.1073/pnas.232580399.

Seeliger MA, Young M, Henderson MN, Pellicena P, King DS, Falick AM, Kuriyan J. High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. Protein Sci. 2005; 14:3135–3139.

Song Z, Wang M, Zhang A. Alectinib: a novel second generation anaplastic lymphoma kinase (ALK) inhibitor for overcoming clinically-acquired resistance. Acta Pharmaceutica Sinica B. 2015 Jan; 5(1):34–37.

Stegmeier F, Warmuth M, Sellers WR, Dorsch M. Targeted cancer therapies in the twenty-first century: Lessons from imatinib. Clin Pharm & Therap. 2010; 87:543–552. doi: 10.1038/clpt.2009.297.

Sørensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in Escherichia coli. Journal of Biotechnology. 2005; 115(2):113 – 128. http://www.sciencedirect.com/science/article/pii/S0168165604004560, doi: http://dx.doi.org/10.1016/j.jbiotec.2004.08.004.

The Structural Genomics Consortium, SGC: Progress in protein kinase structural biology; 2012. http://www.thesgc.org/scientists/resources/kinases, updated 18 Jul 2012.

UniProt Consortium. UniProt: The universal protein knowledgebase. Nucl Acids Res. 2017; 45:D158–D169. doi: 10.1093/nar/gkh131.

Varghese AM, Berger MF. Advancing clinical oncology through genome biology and technology. Genome biology. 2014 Aug; 15(8):427.

Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, Oldfield TJ, O'Donovan C, Martin MJ, Kleywegt GJ. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. Nucleic Acids Research. 2013; 41(D1):D483. http://dx.doi.org/10.1093/nar/gks1258, doi: 10.1093/nar/gks1258.

Volkamer A, Eid S, Turk S, Jaeger S, Rippmann F, Fulle S. Pocketome of Human Kinases: Prioritizing the ATP Binding Sites of (Yet) Untapped Protein Kinases for Drug Discovery. Journal of Chemical Information and Modeling. 2015; 55(3):538–549. http://dx.doi.org/10.1021/ci500624s, doi: 10.1021/ci500624s, pMID: 25557645.

Wang L, Foster M, Zhang Y, Tschantz WR, Yang L, Worrall J, Loh C, Xu X. High yield expression of non-phosphorylated protein tyrosine kinases in insect cells. Protein Express Pur. 2008; 61:204–211. doi: 10.1016/j.pep.2008.05.017.

Weisberg E, Manley PW, Breitenstein W, Brüggen J, Cowan-Jacob SW, Ray A, Huntly B, Fabbro D, Fendrich G, Hall-Meyers E, Kung AL, Mestan J, Daley GQ, Callahan L, Catley L, Cavazza C, Azam M, Mohammed A, Neuberg D, Wright RD, et al. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. Cancer Cell. 2005 Feb; 7(2):129–141.

Wu P, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. Trends in Pharmacological Sciences. 2015 Jul; 36(7):422–439. http://linkinghub.elsevier.com/retrieve/pii/S0165614715000772, doi: 10.1016/j.tips.2015.04.005.

Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, Hellmann MD, Barron DA, Schram AM, Hameed M, Dogan S, Ross DS, Hechtman JF, DeLair DF, Yao J, Mandelker DL, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nature medicine. 2017 May; 159:676.