Title

Applications of Disulfide Tethering to Fragment Discovery and Protein Dynamics in Protein-Protein Interactions

Author

Hallenbeck, Kenneth K

Publication Date

2018

Peer reviewed|Thesis/dissertation

Applications of Disulfide Tethering to Fragment Discovery and Protein Dynamics in Protein-Protein Interactions

by

Kenneth K. Hallenbeck

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

[redacted]

Dedication and Acknowledgements

To my wife Elyse.

The completion of the work described herein would have been impossible without the help and guidance of many colleagues, collaborators, and friends. Of first importance is my advisor and mentor Dr. Michelle Arkin, to whom I owe the chance to work on such a fascinating and well-resourced topic. The remainder of my thesis committee, James Fraser and Bill Degrado, regularly provided thoughtful direction. Other UCSF and Gladstone faculty were actively engaged in this work, including Katerina Akkassoglou, John Gross, Jim Wells, and Deanna Kroetz.

My coauthors and collaborators are noted for their contributions at the beginning of each chapter. I would also like to thank the many Arkin Lab members and UCSF colleagues who provided experimental and emotional support, including Kyle Defreese, Ben Spangler, Ryan Muir, Marcus Chin, Yuliya Birman, Stacie Bulfer, Yinyan Tang, Yao Fan, Daniel Medina-Cleghorn, Kenny Ang, Marja Tarr, Kazuko Olsen, Allison Doak, Ninwe Maraha and others who help along the way.

ABSTRACT

The study of protein protein-interactions (PPIs) can be approached with many lenses. Interaction networks can be mapped genetically or with proteomic methods. Biophysical characterization of specific partner interactions can clarify molecular mechanism of partner interaction. Screening methods can to identify novel interaction modulators, and site-directed chemical probes can used to demonstrate the importance of protein partner interaction in disease or biology. In this work I begin with an overview of the chemical biology toolbox for targeting non-catalytic cysteines residues. I apply one such tool, disulfide tethering, to two PPIs with the goal of broadening the PPIs which are targetable with covalent chemical probes. I then describe the use of cysteine-reactive molecules in understanding protein ligandability and its connection to protein dynamics. Finally, I study a promiscuous PPI receptor (Mac-1) to understand if inhibitors could be developed for specific ligands. Taken together, these data demonstrate the applicability of cysteine reactivity generally, and disulfide tethering specifically, to the study of PPIs.

Table of Contents

List of Tables

List of Figures

Chapter 1

COVALENT STRATEGIES AND METHODS FOR TARGETING NON-CATALYTIC CYSTIENE RESIDUES

Contributing Authors:
DAVID M. TURNER, JULIA L. DAVIES, CONNIE MERRON, PIERCE OGDEN, ELINE SIJBESMA, CHRISTIAN OTTMANN, ADAM R. RENSLO, CHRISTOPHER WILSON, MICHELLE R. ARKIN

INTRODUCTION

Cysteine is an underrepresented residue in protein sequence (3.3% frequency[1]) but is disproportionately involved in protein function, with >50% of cysteine residues being solvent exposed and implicated in a myriad of biochemical processes[2]. Cysteine serves as the reactive nucleophile in many hydrolases (such as cysteine proteases) and can mediate redox reactions (e.g., protein disulfide isomerase). Oxidized forms of cysteine with sulfenic acid or nitrosothiol functionality are increasingly appreciated as playing a role in cellular signaling, and this suggests the possibility of targeting such oxidized forms with specific small molecules[3-4]. Finally, disulfide bond formation between two cysteine residues has been long recognized as contributing to protein tertiary structure. Taken together, these features make cysteine an attractive target for modification by small molecules.

The concept behind covalent modification of cysteine residues is schematized in Figure 1. An initial non-covalent complex (E*I) positions the electrophilic group within range of the nucleophilic thiol moiety and facilitates bond formation (k2). For truly irreversible inhibitors, the resulting covalent complex (E-I) remains intact; however, for reversible electrophiles (e.g. disulfides), the ligand-bound complex dissociates (k-2) over time to reform the initial non-covalent complex (E*I). Thus, cysteine-modifying drugs rely on two binding interactions –

covalent and non-covalent – that can be independently and iteratively optimized to obtain the necessary selectivity and potency to be useful chemical probes or drug leads.

Cysteine-modifying compounds have been directed at both catalytic and non-catalytic residues. Modifying catalytic cysteines, such as those found in deubiquitinases and caspases, has an obvious impact on enzyme function. However, catalytic residues in enzyme active sites are generally highly conserved within families, and isoform selectivity can be difficult to achieve. Non-catalytic cysteines are generally less conserved, making them attractive for selective target modulation. Chemical proteomic studies employing activity-based probes has identified various reactive, functional, and non-catalytic cysteine residues whose functions could be probed and modulated with drug-like covalent molecules. These studies have revealed that inherent thiol reactivity spans six orders of magnitude[5], an observation that is germane in any effort to develop highly selective cysteine-targeted compounds.

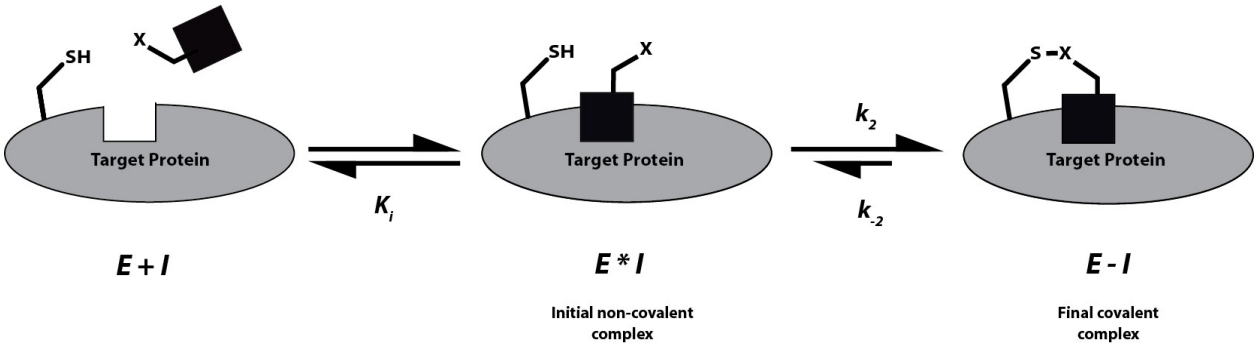


Figure 1: **Schematic of Covalent Enzyme Inhibition**. An initial non-covalent binding event brings the cysteine sulfhydryl group in proximity to the warhead X, driving covalent bond formation. The covalent bond can be irreversible (e.g vinyl sulfonamide), where k_{-2} is zero, reversible (e.g. another sulfhydryl group) where k_{-2} depends on reaction conditions or reversible covalent (e.g. cyanoacrylamide), where k_{-2} depends on warhead reactivity.

As appreciation for the targetable nature of cysteine residues has grown, covalent approaches to drug discovery are resurgent. However, covalent pharmacology inevitably raises the concern that reactive drugs or drug metabolites can induce organ damage or evoke an immune response through off-target protein binding[6-7]. A related concern is that an

electrophilic drug can be rapidly inactivated and eliminated via reaction with native nucleophiles (e.g. GSH)[6, 8]. These arguments are often countered by noting that many safe and well tolerated drugs in use for decades, such as aspirin and penicillin, act via covalent modification of their targets[9-10] and that the intentional targeting of nucleophilic sites with appropriately tuned electrophiles can mitigate the risk of covalent pharmacology. Giving appropriate attention to these potential issues is likely a factor in the recent clinical successes of covalent drug candidates[11].

Taking the advantages and challenges into account, the design of selective cysteine-modifying molecules as drug leads or as chemical probes for biomolecules has proven an attractive approach for the following distinct applications:

(1) Lead optimization. Covalent pharmacology can enhance potency and selectivity of lead compounds, most compellingly by increasing target residence time. Covalent pharmacology is typically associated with 'durable' target inhibition in vivo, and often very different pharmacokinetic/pharmacodynamics (PK/PD) relationships as compared to drugs exhibiting reversible, fast-off inhibition kinetics[6]. Selectivity for protein isoforms containing the targeted residue is another benefit of this approach for lead optimization. Nearly any cysteine proximal to a known drug-binding site is a potential candidate. Knowledge of the structure, to support computational-guided design of the cysteine-reactive analog, is also highly desirable.

(2) Chemical handles to identify new lead scaffolds. In targets where structural data indicates a binding pocket is available near a solvent-exposed cysteine, screening libraries of diverse molecules containing a cysteine-reactive moiety is an effective strategy for lead discovery. Computational approaches help triage promising target sites and identify scaffolds around which to build electrophile libraries for screening. One particularly interesting application

for new drug discovery is targeting cysteine mutations found in oncogenic proteins; cysteine reactive molecules could first validate the function of these mutations in disease, then serve as lead compounds for therapeutic development.

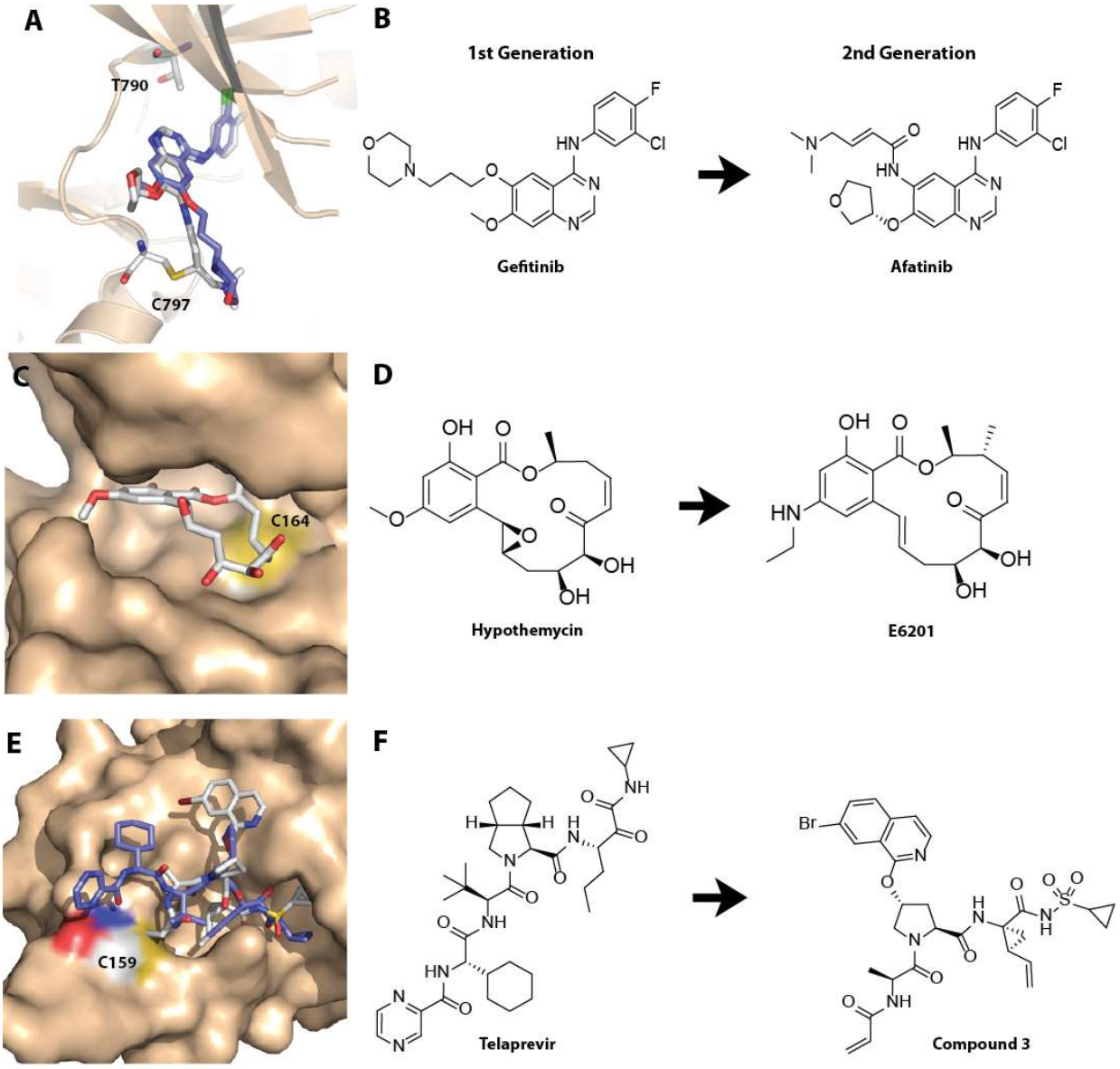


Figure 2: **Deriving Covalent Inhibitors from Known Scaffolds**. A-B) Non-covalent 1st generation EGFR inhibitor Gefitinib (purple) overlayed with covalent 2rd generation inhibitor Afatinib (white). T790, the eventual site of resistancemutation, is drawn. B-C) Telaprevir (purple) overlayed with covalent derivative Compound 3 (white) which binds to conserved Cys159. E-F) Natural product Hypothemycin bound to ERK2 and synthetic drug candidate E6201.

(3) Site-specific study of protein allostery, dynamics, and structure-function relationships. Native or engineered surface cysteines can be used to find molecules that bind at known sites of allosteric regulation, or can uncover previously undetected ('cryptic') binding pockets.

1. Improving drug properties for known scaffolds

Kinases

It is startling to recall that twenty-five years ago, kinases were considered 'undruggable' targets. The central importance of kinases and the high structural homology within the family suggested that imperfect selectivity would lead to unacceptable levels of toxicity. Through the creative efforts of many laboratories, kinase inhibitors are now a well-established class of cancer therapeutics. However, the development of drug resistance during kinases inhibitor therapy is also common. First-generation inhibitors of epidermal growth factor receptor (EGFR) were effective in treating certain subtypes of lung carcinoma, but a mutation in the gatekeeper residue (T790M) resulted in a steric clash in the binding site (Figure 2A) ultimately leading to clinical relapse[12-13]. Walter and colleagues demonstrated that EGFR Cys797, which sits at the edge of the ATP-binding pocket and is present in just 2% of kinases, could be targeted to improve potency and recover function in the presence of T790M[14]. Selectively targeting a rare cysteine to increase drug residence time was expected to result in an improved clinical outcome (Figure 2A, B). However, the effectiveness of second-generation EGFR inhibitors was limited by on-target toxicity, since the drugs inhibited both mutant and WT EGFR[15-16]. To selectively target oncogenic T790M EGFR, the 1st and 2nd generation quinazoline scaffold was replaced by other heterocyclic scaffolds into which an acrylamide electrophile could be readily introduced[17]. Several pyrimidine-based molecules were identified that exhibited 30 to 100-fold selectivity for T790M over WT EGFR. Two of these pyrimidines, Osimertinib (Phase 1/2)[18] and Rociletinib

(Phase 2/3) [19], are now in advanced clinical trials for treatment of non-small-cell lung cancer and are showing promising effects in patients[20-21]. The search for the next generation of covalent EGFR inhibitors continues, guided by rational design and lessons learned from the clinic[22].

As these examples illustrate, the Michael reaction between cysteine thiol as nucleophile and an alpha-beta unsaturated carbonyl (e.g. acrylamide) has figured prominently in the design of covalent kinase inhibitors. Michael 'acceptors' are attractive for these applications because their reactivity can be tuned by changing the nature of the electron-withdrawing carbonyl (or related) function and/or by altering the steric environment surrounding the electrophilic beta carbon atom.

The early success of covalent EGFR inhibitors motivated use of the approach in many other kinases. Ibrutinib, which received a breakthrough drug designation in 2013 for mantle cell lymphoma, del17p chronic lymphocytic leukemia and Waldenström's macroglobulinemia[23], contains an acrylamide warhead that irreversibly modifies Cys481 in Bruton's tyrosine kinase (BTK). Ibrutinib's scaffold was identified in a screen and was prioritized because it showed selectivity for a small group of Tec and Src-family kinases. Sequence comparison and structural homology modeling suggested that BTK contained a nucleophilic cysteine in the position analogous to Cys797 in EGFR[24]. This observation motivated a structure-guided medicinal chemistry effort that sampled three potential Michael acceptors – propiolamide, vinyl sulfonamide, and acrylamide – and found the last had the best activity (0.5 nM against BTK) and selectivity profile.

The Janus kinase (JAK) family have >80% amino acid identity in their ATP-binding site, exemplifying the kinase selectivity problem. Because of their importance to cytokine signaling

pathways and potential as autoimmune disease therapeutics, many pan-JAK inhibitors have been described[25]. However, no reversible, isoform-specific inhibitors exist. JAK3 is the only JAK that contains a cysteine (Cys909) at the EGFR and BTK site, and was therefore targeted with tricyclic JAK inhibitors that included a terminal electrophile designed to irreversibly react with JAK3 Cys909[26]. These compounds inhibit JAK3 with <100 nM potency in cells and selectively ablate JAK3-dependent signaling pathways, with little to no JAK2 activity up to 50uM. However, Goedken and colleagues report poor pharmacokinetic profiles for these JAK3 compounds[26], and more optimization is necessary before JAK3 inhibitors catch up to their EGFR or BTK counterparts.

As it happens, covalent kinase inhibition is not solely a product of human ingenuity. The fungal natural product hypothemycin[27] and related macrocycles[28] are known to covalently inhibit a subset of human kinases with a cysteine preceding the kinase DXG motif (Figure 2C,D). These so-called CDXG kinases comprise 48 of 518 human kinases and include important cancer drug targets such as MEK, ERK, PDGFR, VEGFR2, and FLT3[29-30]. The macrocyclic structure of these compounds contains a cis-enone that serves as the cysteine-reactive moiety. The epoxide present in some family members is remarkably unreactive due to a macrocyclic conformation that blocks the approach of would-be nucleophiles. The wholly synthetic drug candidate E6201, described as a dual MEK1 and FLT3 inhibitor, is in early clinical trials for advanced hematologic malignancies with documented FLT3 mutation[31].The pharmacokinetics of E6201 in preclinical species and in humans is characterized by moderate to high distribution but rapid clearance[32], which might be regarded as an ideal PK profile for a drug exhibiting covalent pharmacology.

Beyond Kinases

These kinase examples illustrate a general principle: if a non-covalent scaffold for a target pocket is already available, adding a suitably positioned electrophilic group to form a covalent bond with a nearby cysteine residue provides large gains in potency. To achieve this, analogs of the inhibitor are prepared in which an electrophilic function, typically a Michael acceptor, is placed in a position and orientation informed by structural information about ligand binding, if available. Recent studies have extended this approach to non-kinase enzymes. For example, in 2010 Avila Therapeutics reported selective peptidomimetic inhibitors targeting a cysteine in Hepatitis C virus (HCV) protease[33]. Using sequence and structural alignment, they identified non-catalytic Cys159, which does not occur in human proteases but was conserved across all 919 HCV NS3 sequences known at the time. An acrylamide was appended to a scaffold based on the protease inhibitor telaprevir using a structure-guided approach, realizing gains in IC50 from 2,500 nM for telaprevir to 2 nM for the electrophilic analog (Figure 2E,F). The addition of the acrylamide also improved selectivity vs off-target mammalian proteases. Whereas covalent protease inhibitors targeting active-site nucleophiles are common, to our knowledge, HCV NS3 is the only published example where a non-catalytic cysteine was leveraged to develop an irreversible inhibitor.

Daniel et. al, recently reported a similar strategy for inhibiting histone deacetylases (HDACs). The HDAC inhibitor suberoylanilide hydroxamic acid (SAHA) was modified to include a covalent moiety that reacted with a conserved cysteine 5.6Å from the enzyme active site[34]. This dual-action inhibitor did not achieve isoform selectivity or increased potency, but represented an interesting example of using structure-guided design and cysteine modification to develop a novel drug-targeting scheme.

Identifying New Scaffolds

Incorporating cysteine reactivity into de novo drug (and probe) discovery efforts is a newer concept that builds on the successes described above. This hit-discovery strategy might be adopted because non-covalent approaches failed to give validated chemical starting points, as is often the case for challenging targets like protein-protein interactions and some classes of proteases, or because the targeted cysteine is hypothesized to be important for the protein's pathological function. For instance, a recent survey of oncogenic mutations found that mutations to cysteine were 2.6-fold more likely than would be expected by chance; of the fifteen most commonly found cysteine mutants, fourteen were surface exposed and several were known to affect protein function[35]. Oncogenic cysteine mutations in the extracellular domain of fibroblast growth factor receptor (FGFR3), for instance, are found in 88% of mutated FGFR3, and have been shown to induce ligand-independent activation of the receptor through disulfide bond formation. Conceptually, these mutations could be strong candidates for a cysteine-binding inhibitor approach.

Of surface-available, non-catalytic cysteines, only a subset is proximal to a site capable of interacting favorably with a small molecule. When structural data are available, computational methods for detecting potential binding pockets are therefore useful for assessing potential druggability. Several programs, including FTmap[36] and WaterMap[37], exist for predicting binding pockets on a protein surface. As an example, FTmap is a simple, DOCKING-based algorithm that scans a protein surface for propensity to bind very small organic molecules. Clusters of molecules indicate hotspots for ligand binding and suggest whether a tractable pocket is available[36]. FTMap has been applied to regions of FGFR3 near two of the known cysteine mutations, and the hotspot analysis suggests that residue has more of a pocket than the other[35].

Neither pocket was high scoring, but it is important to note that proteins have regions of structural flexibility, and even small changes in side chain orientation or secondary structure can create cryptic pockets not seen in the unbound protein structure[38].

Computational Library Design

Of equal importance to the selection of a druggable cysteine/binding site, is the selection of a suitable compound library to screen. As in traditional high-throughput screening, there are three general approaches: targeted libraries designed to bind conserved features of target class (e.g., kinases), diversity libraries meant for screening a wide range of targets, and virtual screens.

Computational methods for in silico screening of covalent small molecules have recently been described[39-40]. For example, virtual screening using DOCKovalent, an adaptation of DOCK3.6, led to experimentally validated inhibitors of the β-lactamase AmpC and kinases RSK2 and MSK13a[40]. In this method, a large virtual library based on commercially available compounds built with a range of electrophiles was evaluated and the top 1% were manually prioritized for experimental validation. For AmpC, a library of boronic acids was screened for covalent modifiers of catalytic Ser64. Six diverse hits were tested for AmpC inhibition and three had a Ki < 1uM. The most potent inhibitor (Ki = 40 nM) was crystalized to confirm the predicted docking pose. Comparisons with the original screen motivated the purchase of 7 additional compounds, ultimately yielding a 10 nM inhibitor with a similar binding pose. This compound represented a novel AmpC inhibitor with good selectivity (>104-fold) over common serine proteases that bind boronic acids.

Experimental Library Design

Assembly of cysteine-reactive small molecule libraries for experimental screening have tended to use a fragment-based philosophy[41-42]. Fragment-based drug discovery (FBDD)

seeks to identify low molecular weight fragments (typically <300 Da) that bind with high ligand efficiency to sub-pockets within a binding site. An attractive feature of FBDD is the ability to efficiently sample chemical space with a relatively small number of compounds (often ≤ 2000 fragments)[43]. Once fragments are identified, hits are further evolved into more complex and optimized leads[44-45]. Since a cysteine-reactive library generally needs to be synthesized from scratch, the focus on small libraries of fragment-sized molecules is appealing. Furthermore, the use of covalent elements in the fragment library serves to increase the initial potency, making fragments easier to find in a binding- or activity-based screen. Finally, the use of engineered or native cysteine residues makes these methods site-directed, allowing the chemical biologist to evaluate the ligandability of a given site on a protein.

In designing a cysteine-reactive fragment library, both the nature/reactivity of the electrophile and the structure of the non-covalent 'diversity' elements must be considered. The distance between the electrophilic warhead and the diversity element is also important, since an effective hit must make productive interactions with the protein whilst also displaying the electrophile to react with the cysteine residue. Thus, linker lengths and geometries provide another opportunity for diversification. The composition of the linker can also have important effects on the chemical reactivity of the electrophile, as described below.

In selecting diversity elements for a library, one can take cues from a large literature on fragment library design[46-47]. For instance, researchers at Astex proposed guidelines for constructing fragment libraries with desirable physiochemical properties[48]. This rule of thumb, dubbed the "Rule of three"[49] recommends a molecular weight <300 Da, number of H-bond acceptors ≤3, number H-bond donors ≤3, and cLogP ≤3. Further considerations include limiting the number of rotatable bonds to ≤3 and the polar surface area to ≤60 Å2 [50]. Another recent

trend favors selecting fragments with greater shape diversity, including more sp3-rich structures to complement the generally 'flat' aromatic heterocycles commonly included in fragment libraries. While such fragments are underrepresented in commercial libraries, they can be accessed through bespoke synthesis or diversity-oriented synthesis[51-52].

Finally, electrophilic warheads must be selected. Electrophiles come in three flavors, irreversible, reversible, and reversible-covalent, which will be considered separately.

Irreversible Warheads

Traditional cysteine-reactive compounds are irreversible electrophiles. Given the wide range of cysteine nucleophilicity in proteins, it is desirable to include a range of electrophilic warheads with differing reactivity during library construction. This concept was illustrated by Flanagan and coworkers at Pfizer using LCMS and NMR based kinetic studies to measure thiol-reactivity for a number of irreversible electrophiles. This study revealed a 450-fold range in reaction rates, from minutes to days[53]. To further guide warhead design, computational methods are available to identify those electrophiles likely to exhibit undesirable non-selective protein reactivity[54-55]. As seen above, acrylamides are commonly used warheads, since they often show low rates of reaction in solution. Non-covalent affinity for the binding site, however, increases the local concentration and residency time enough to allow reaction with the cysteine-of-interest.

Identifying the targets and off-targets of irreversible-covalent inhibitors can be significantly facilitated by the covalent nature of inhibition. In a recent example, the natural product hypothemycin (Figure 2F) was found to inhibit growth of the eukaryotic parasite T. brucei, the causative agent of Human African Trypanosomiasis (sleeping sickness). Though the parasite target was unknown, Choy et al suspected a CDXG kinase, given the known reactivity of

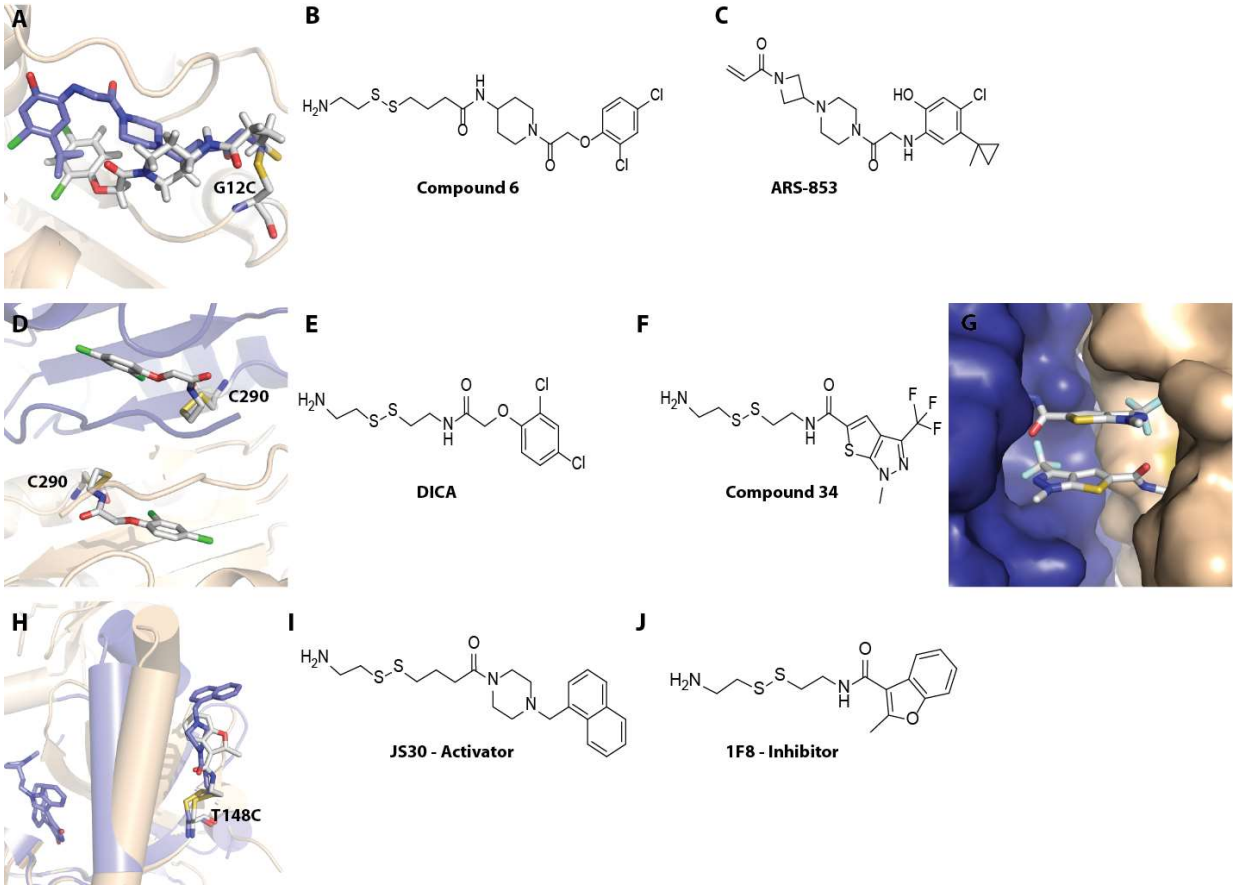


Figure 3: **Applications of Disulfide Tethering**. A) K-Ras G12C mutation targeted by an optimized disulfide tethering hit (white) B) and an irreversible inhibitor (purple) C) with nM potency. D) Caspase-1 zymogen dimer (monomers colored tan/purple) bound to two DICA molecules E) at Cys290. G) Caspase-7 dimer (monomer surface colored tan/purple) bound to two interacting copies of Compound 34 F). H) PDK1 bound to activator JS30 I) and inhibitor 1F8 J) at the PIF-pocket. The activator shifts the regulatory C-helix down toward the active site where GDP is bound (purple).

hypothemycin with mammalian CDXG kinases. They used the X-ray structure of hypothemycin bound to ERK2 (Figure 2D) to design a propargyl analog of hypothemycin for labeling and pull-down applications[56]. Application of this probe to T. brucei lysates, followed by 'click' conjugation of a fluorescent dye, allowed putative protein targets to be visualized by SDS-PAGE. Specific, 'saturable' binding interactions could be distinguished from non-specific labeling by co-incubation with hypothemycin, which competes with the probe for labeling of saturable (specific) targets. The same probe was also employed for pull-down and quantitative MS analysis, leading to the identification of TbGSK3short and TbCLK1/2 as bona fide, saturable

targets of hypothemycin in T. brucei. The probe was then employed to demonstrate that hypothemycin treatment at cytotoxic concentrations fully inhibited TbCLK1 but only marginally inhibited TbGSK3short, suggesting that TbCLK1 was the central target While hypothemycin proved to be a relatively selective electrophile in the parasite proteome, the same approach can be applied to identify off-targets for less selective covalent inhibitors[57].

Reversible Warheads

In contrast to irreversible fragment library screening, where compounds are selected through a combination of kinetic trapping and binding thermodynamics, disulfide trapping (Tethering) uses readily reversible disulfide bonds to screen for fragments based primarily on thermodynamic stabilization[58].

In disulfide tethering, a library of disulfide-containing fragments is assayed against a cysteine-containing protein under reducing conditions. Mass spectrometry or functional assays are used to screen for fragments that form disulfide bonds with the desired cysteine thiolate[59]. Using reversible disulfide-exchange chemistry allows the screening assay to reach thermodynamic equilibrium; this equilibrium (and the stringency of the screen) is also controlled by the reduction potential of the buffer, which is varied based on the chemical reactivity of the target cysteine[60]. The library design also favors hit-selection based on non-covalent binding interactions; the disulfide moieties are separated from the diversity elements by 2-3 carbon aliphatic linkers, which serve to separate the diversity element from the reacting thiol and lend similar intrinsic nucleophilicity to each library member. Disulfide hits are then developed into leads through replacement of the disulfide bond with electrophiles such as acrylamide, or they can be converted to non-covalent ligands through removal of the thiol and structure-guided optimization. For example, in developing inhibitors of the interleukin-2 (IL-2)/IL-2 receptor

interactions, disulfide trapping identified a pharmacophore that bound in a site proximal to that of a known inhibitor; linking the two compounds provided a 30-fold enhancement in affinity[61].

Converting Reversible to Irreversible Warheads

The GTPase K-Ras (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) is an oncogene with many cancer-associated activating mutations[62]. Pharmacological inhibition of K-Ras has been the goal of many drug discovery efforts but the target has proved mostly intractable[63]. One K-Ras mutation, G12C, introduces a cysteine that sits proximal to the Switch-I and -II regions involved in K-Ras nucleotide binding. A disulfide tethering screen of 480 fragments identified two fragments that labeled the G12C mutant[64]. Crystal structures of K-Ras bound to an optimized screening hit (compound 6; Figure 3A,B) informed the replacement of the reversible thiol with an irreversible acrylamide. This acrylamide-containing compound bound to a previously unidentified allosteric pocket that formed under the switch-II pocket, stabilizing the inactive GDP-bound state and also disrupting the binding of effector proteins. Despite this novel mechanism of inhibition, the compound was ultimately incapable of engaging intracellular K-Ras, motivating efforts to find more potent G12C-specific molecules. Wellspring Biosciences recently reported ARS-853 (Figure 3A,C), which was optimized through iterative structure-guided design and crystallography[65]. ARS-853 bound to the same switch-II pocket and was able to ablate downstream signaling at low micromolar potency in cells. ARS-853 is the first reported K-Ras inhibitor to reach the potency range of a quality drug lead.

The K-Ras G12C mutation is also close enough to the nucleotide binding site that it can be used in conjunction with a GDP analog to covalently inhibit at the active site. This approach yielded SML-8-73-1, a nucleotide analog that forms a stable thioether with G12C and competes for binding with GDP and GTP[66-67]. Though SML-8-73-1 bears two ionizable functions and

is cell impermeable, cell-permeable analogs were cell permeable and showed EC50 values of 25-45 uM against three K-Ras G12C cell lines. Achieving sub-micromolar cell-based activity with a GDP analog will be challenging and may require a permeable prodrug strategy. Nevertheless, the general strategy of targeting disease-specific cysteine mutants is a novel and highly attractive one that might well be applied to other targets with surface-exposed cysteine mutations[35].

Reversible-Covalent Warheads

The most recent development in cysteine targeting has been the development of slowly reversible electrophiles. These warheads offer the benefits of irreversible and rapidly reversible chemistries, while mitigating their limitations. Since slowly reversible inhibitors can be selected based on thermodynamic binding, it should be possible to optimize their non-covalent interactions with the binding site. Additionally, reversibly covalent ligands could provide the high affinity and long residency time of covalent warheads while reducing the risks of immunogenicity arising from truly irreversible binding. Modulating the residency time of the E*I complex (Figure 1) can enable on-target action in vivo long after free drug has been eliminated from circulation[68].

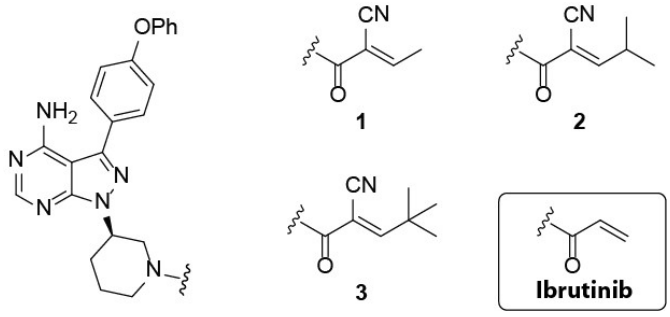


Figure 4: **BTK Reversible-Covalent Cyanoacrylamide Inhibitors.** The core Ibrutinib scaffold with three systematically substituted cyanoacrylamide warheads: **1** Me, **2** i-Pr, and **3** t-Bu.

The concept of slowly reversible electrophiles was described by Taunton and coworkers in 2012[69]. In this work, a Michael acceptor based on the β-cyanoacrylamide function was

found to react with cysteine thiols in a slowly reversible reaction. The labeling of C436 on the kinase RSK2 was evaluated with a series of congeneric pyrrolopyrimidine inhibitors bearing various electrophilic side chains positioned to react with C436. Acrylate and acrylamide-based inhibitors showed irreversible inhibition, while the cyanoacrylamide analog exhibited reversible inhibition. Interestingly, the kinetically stable covalent bond between C436 and the cyanoacrylamide inhibitor was rapidly broken upon unfolding of the kinase domain of RSK2 with detergent. This result revealed that non-covalent interactions between the inhibitor and surrounding residues of the binding site served to stabilize the covalent adduct.

The reversible-covalent reactivity of cyanoacrylamides motivated a subsequent effort by the same group to explore tuning of target residence time via systematic modifications to the cyanoacylamide warhead[70]. In this work, Bradshaw et al. targeted C481 in BTK, the same cysteine targeted by the irreversible acrylamide Ibrutinib. Unlike C436 in RSK2, C481 in BTK lies outside the immediate confines of the ATP binding site. Accessing this surface-exposed residue thus necessitated the introduction of a piperidine spacer between the hinge-binding pyrazolopyrimidine core and the cyanoacrylamide. Additionally, the cyanoacrylamide warhead in the BTK inhibitor was 'reversed', with the linkage forged via the amide function of the cyanoacrylamide rather than at the beta carbon as in the RSK2 inhibitors. Thus, structure-based design was crucial in successfully engaging cysteines C436 and C481 in RSK2 and BTK, respectively.

In addition to optimally presenting the cyanoacrylamide for reaction with C481 in BTK, reversal of the warhead allowed the kinetics of cysteine engagement to be altered via introduction of various substituents at the electrophilic beta carbon. Increasing the steric demands of this substituent correlated with increasing target residence time; 20 hours after

removing unbound compound, ~60% of the t-Bu compound was still bound to BTK (Figure 4, compound 3), compared to ~30% for i-Pr (2) and ~10% for the Me-containing compound (1). Further exploration of diverse beta substituents led to a series of inhibitors exhibiting a range of target residence times, from minutes to several days. Importantly, the intrinsic potency of these inhibitors in biochemical assays was not correlated with the durability of target engagement, highlighting the potential of this approach to enable separate optimization of potency and target residence times to meet the requirements of a specific therapeutic application.

One of the most promising BTK inhibitors exhibited a residence time of >1 week in vitro was tested for kinase selectivity across a panel of kinases. Only 6 of 254 tested kinases were inhibited >90% at 1 □M, and each of these sensitive kinases possessed the analogous C481. Importantly, other C481 kinases, including EGFR and JAK3 that were regarded as possible off-targets, were relatively unaffected (IC50 > 3 uM). Finally, this compound was evaluated for BTK inhibition in vivo and was found to retain target engagement >24 hours after oral dosing, by which time free inhibitor had been cleared from circulation. Overall, the tunability of reversible-covalent warheads and the corresponding benefits in terms of selectivity and in vivo PD properties make a convincing case for the wider application of this approach in drug discovery[70].

Exploring Allostery

Cysteine-targeted agents have also served as chemical-biology tools to explore protein conformation and allostery. Conformational changes, e.g., caused by posttranslational modification, protein-protein interactions, and the binding of signaling molecules, play a critical role in protein function. Designing molecules that affect a specific protein state could make for highly selective drugs. However, compounds that act at allosteric sites on proteins are often found serendipitously. Allosteric sites can also be cryptic sites, in that they are not apparent in

crystal structures of the unbound protein and are therefore difficult to discover and model computationally. On the other hand, some cryptic sites might not bind endogenous ligands, yet might have nascent allosteric potential that synthetic ligands could harness. For these reasons, predicting and evaluating the functional relevance of cryptic sites are active areas of research, and the potential to pro-actively identify and target allosteric and/or cryptic sites remains an important challenge for structure-based drug discovery.

Native Cysteine Residues

In the case of caspases, disulfide trapping identified a previously unknown allosteric site. Caspase-7 is a dimeric cysteine protease and a potent effector of cell death by apoptosis. The enzyme is expressed as an inactive dimeric zymogen that is cleaved under apoptotic conditions to the active form; a series of loop movements causes a significant change in the overall conformation of the active sites and the dimer interface. Disulfide Tethering identified two compounds called FICA and DICA (Figure 3E,G) that bound selectively to the Cys290 residue at the dimer interface and stabilized a zymogen-like, inactive conformation of the 'active' caspase-7[71]. Caspase-1, a caspase involved in pro-inflammatory processes, was also found to have a cysteine residue (Cys331) at the dimer interface that could be targeted by a disulfide-containing compound[72]. Each of these disulfide-trapped compounds bound with two molecules tethered to symmetry-related cysteines at the interface, but the compounds bound with different orientations; for instance, compound 34 made compound-compound interactions in the caspase-1 site, while 2 DICA molecules did not interact (Figure 3D). For both proteins, analysis of the structures of the compound-trapped inactive state and the active conformation uncovered an allosteric network over the 15A between the allosteric site and the enzyme active site[73-74]. In

an interesting and potentially generalizable application, compound 34-bound caspase-1 was used to select anti-caspase-1 antibodies that preferentially bound to the inactive conformation[75].

3.2. Engineered Cysteine Residues

Naturally, many allosteric or cryptic sites will not have native cysteine residues nearby. In these cases, engineering cysteine residues and probing them with cysteine-reactive compounds can provide deep insight into protein structure and dynamics. The surface of the protein hormone IL-2 was probed with a series of cysteine residues followed by disulfide trapping. IL-2 is a four-helix bundle that binds to a trimeric receptor. Eleven cysteine mutations were made, one-at-a-time, along the surface of the alpha-chain binding site of IL-2 [38]. This face of IL-2 was found to be amphiphilic, with one side being hydrophilic, flat, and structurally stable, while the other side was more hydrophobic and structurally dynamic. Disulfide screening identified many more compounds that bound to the dynamic portion of the interface, in some cases trapping conformations not seen in structures of the apo protein, including conformations induced by the binding of ligands at other locations on the surface of IL-2. Thus, protein-protein interfaces can have regions of structural adaptivity, which might have a role in binding multiple protein partners[76] and/or may be exploited by small-molecule ligands[38, 77].

Protein-protein interactions can also allosterically regulate enzyme activity, as is the case with different classes of kinases. The AGC kinases, for instance, have a common allosteric site in the N-lobe of the kinase domain, where substrate proteins or regulatory domains of the kinase itself can bind[78]. Binding to this allosteric site co-localizes the substrate and enzyme and also allosterically activates catalysis by positioning the regulatory ‘C helix’ into an active position. The AGC kinase PDK1 is a well-studied example of this allosteric regulation, and multiple small-molecule modulators that bind to the allosteric site have been designed. Cysteine

mutagenesis followed by disulfide trapping identified several compounds that allosterically activated or inhibited kinase activity[79]. Importantly, both inhibitors and activators could be selected at the same cysteine residue; hence, the details of the non-covalent binding interactions and molecular shape determined allosteric outcome, not structural changes in the protein due to cysteine mutagenesis per se. X-ray structures of an activator and inhibitor bound to PDK1 (Figure 3H) highlighted the mechanism of allostery; the smaller compound (Figure 3J) stabilized a conformation in which the regulatory C helix was pulled away from the active site, while the larger compound (Figure 3I) pushed the C-helix down and into the active conformation[79]. These studies underscore the subtlety between binding and allostery that makes small-molecule design of allosteric modulators both fascinating and complicated.

Bishop and coworkers have taken a protein engineering approach to develop selective allosteric inhibitors of protein tyrosine phosphatase (PTP) enzymes[80]. This important class of signaling enzymes has encountered significant challenges for drug discovery, given the similarity of the active site across the family and the challenge with obtaining cell-permeable active-site inhibitors. The Bishop lab found that the PTP Shp2 is sensitive to inhibition by the cysteine-reactive dye FLAsH, which binds to 2-, 3-, or 4 cysteine residues. They identified two nearby cysteine residues, Cys333 and Cys367 that are buried in the apo-structure of Shp2, but become surface accessible in the presence of FLAsH [80]. Cys333 is unique to Shp2 among PTPs, potentially providing a novel therapeutic strategy for targeting Shp2. However, FLAsH itself did not bind tightly enough to wild type Shp2 to bind selectively in cell lysates, so the investigators engineered an additional Cys368 to provide trivalent coordination of FLAsH[81]. Intriguingly, they were able to create similar allosteric sites by engineering cysteine residues at the analogous

position in several PTPs. Thus, they discovered a novel, cryptic allosteric site that can be used to probe the functions of specific PTPs in cells.

Cysteine mutagenesis/reactivity has also been used to experimentally validate potential hidden/cryptic allosteric sites identified computationally. Bowman, et al. utilized a Markov state model that evaluates protein structural changes on the microsecond to millisecond timescale[82]. Using a drug-resistant mutant of TEM-1 beta-lactamase as a model system, they collected an ensemble of protein structures and looked for transient pockets that a) were fragment-to-lead sized, b) correlated with motions at the active site, and c) included residues that changed from buried to surface-exposed upon pocket formation. They then mutated these residues to cysteine and used the thiol-detection reagent 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) to demonstrate that they could be labeled. The rate of DTNB reaction with cysteine mutants supported the hypothesis that pockets opened and closed transiently. In three cases, TNB-labeled enzyme had a reduced catalytic efficiency, suggesting that trapping these pockets did have an allosteric effect on the active site. The authors envisaged a pipeline in which cryptic pockets would be identified computationally, then validated through cysteine mutation and screening with cysteine-reactive libraries. Structure-guided design could then be used to optimize the compounds so that they allosterically inhibit the non-cysteine containing protein.

CONCLUSION

The post-translational modification (PTM) of reactive amino acid side chains is well recognized as an essential mechanism by which biology regulates cell signaling, protein structure, and the epigenetic control of gene expression. The various examples and approaches to cysteine modification described herein might be considered as examples of unnatural PTM

leveraging the vastly greater access to chemical space enabled by synthetic organic chemistry. While chemical biologists and drug discovery scientists have yet to equal the exquisite selectivity of biochemical PTM, structure-guided design and ever improving computational tools for predicting chemical reactivity and ligand binding portend a bright future for cysteine-reactive small molecules in chemical biology and drug discovery.

REFERENCES

[1] King, J. L.; Jukes, T. H., Non-Darwinian evolution. *Science (New York, N.Y.)* **1969,** *164* (3881), 788-98.

[2] Requejo, R.; Hurd, T. R.; Costa, N. J.; Murphy, M. P., Cysteine residues exposed on protein surfaces are the dominant intramitochondrial thiol and may protect against oxidative damage. *The Febs Journal* **2010,** *277* (6), 1465-1480.

[3] Leonard, S. E.; Reddie, K. G.; Carroll, K. S., Mining the Thiol Proteome for Sulfenic Acid Modifications Reveals New Targets for Oxidation in Cells. *ACS Chemical Biology* **2009,** *4* (9), 783-799.

[4] Majmudar, J. D.; Martin, B. R., Strategies for profiling native S-nitrosylation. *Biopolymers* **2014,** *101* (2), 173-179.

[5] Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B. D.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F., Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010,** *468* (7325), 790-795.

[6] Johnson, D. S.; Weerapana, E.; Cravatt, B. F., Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future medicinal chemistry* **2010,** *2* (6), 949-64.

[7] Mah, R.; Thomas, J. R.; Shafer, C. M., Drug discovery considerations in the development of covalent inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2014,** *24* (1), 33-39.

[8] David-Cordonnier, M.-H.; Laine, W.; Joubert, A.; Tardy, C.; Goossens, J.-F.; Kouach, M.; Briand, G.; Thi Mai, H. D.; Michel, S.; Tillequin, F.; Koch, M.; Leonce, S.; Pierre, A.; Bailly, C., Covalent binding to glutathione of the DNA-alkylating antitumor agent, S23906-1. *European Journal of Biochemistry* **2003,** *270* (13), 2848-2859.

[9] Bauer, R. A., Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discovery Today* **2015,** *20* (9), 1061-1073.

[10] Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A., The resurgence of covalent drugs. *Nat Rev Drug Discov* **2011,** *10* (4), 307-317.

[11] Moghaddam, M. F.; Tang, Y.; O'Brien, Z.; Richardson, S. J.; Bacolod, M.; Chaturedi, P.; Apuy, J.; Kulkarni, A., A Proposed Screening Paradigm for Discovery of Covalent Inhibitor Drugs. *Drug Metabolism Letters* **2014,** *8* (1), 19-30.

[12] Pao, W.; Miller, V. A.; Politi, K. A.; Riely, G. J.; Somwar, R.; Zakowski, M. F.; Kris, M. G.; 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,** *2* (3), e73.

[13] Yun, C.-H.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K.-K.; Meyerson, M.; Eck, M. J., The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (6), 2070-2075.

[14] Walter, A. O.; Sjin, R. T. T.; Haringsma, H. J.; Ohashi, K.; Sun, J.; Lee, K.; Dubrovskiy, A.; Labenski, M.; Zhu, Z.; Wang, Z.; Sheets, M.; Martin, T. S.; Karp, R.; van Kalken, D.; Chaturvedi, P.; Niu, D.; Nacht, M.; Petter, R. C.; Westlin, W.; Lin, K.; Jaw-Tsai, S.; Raponi, M.; Dyke, T. V.; Etter, J.; Weaver, Z.; Pao, W.; Singh, J.; Simmons, A. D.; Harding, T. C.; Allen, A., Discovery of a mutant-selective covalent inhibitor of EGFR that overcomes T790M-mediated resistance in NSCLC. *Cancer discovery* **2013,** *3* (12), 1404-1415.

[15] Katakami, N.; Atagi, S.; Goto, K.; Hida, T.; Horai, T.; Inoue, A.; Ichinose, Y.; Koboyashi, K.; Takeda, K.; Kiura, K.; Nishio, K.; Seki, Y.; Ebisawa, R.; Shahidi, M.; Yamamoto, N., LUX-Lung 4: A Phase II Trial of Afatinib in Patients With Advanced Non–Small-Cell Lung Cancer Who Progressed During Prior Treatment With Erlotinib, Gefitinib, or Both. *Journal of Clinical Oncology* **2013,** *31* (27), 3335-3341.

[16] Sequist, L. V.; Besse, B.; Lynch, T. J.; Miller, V. A.; Wong, K. K.; Gitlitz, B.; Eaton, K.; Zacharchuk, C.; Freyman, A.; Powell, C.; Ananthakrishnan, R.; Quinn, S.; Soria, J.-C., Neratinib, an Irreversible Pan-ErbB Receptor Tyrosine Kinase Inhibitor: Results of a Phase II Trial in Patients With Advanced Non–Small-Cell Lung Cancer. *Journal of Clinical Oncology* **2010,** *28* (18), 3076-3083.

[17] Zhou, W.; Ercan, D.; Chen, L.; Yun, C.-H.; Li, D.; Capelletti, M.; Cortot, A. B.; Chirieac, L.; Iacob, R. E.; Padera, R.; Engen, J. R.; Wong, K.-K.; Eck, M. J.; Gray, N. S.; Janne, P. A., Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* **2009,** *462* (7276), 1070-1074.

[18] ClinicalTrials.gov Search: Osimertinib. https://www.clinicaltrials.gov/ct2/results?term=Osimertinib.

[19] ClinicalTrials.gov Search: Rociletinib. https://www.clinicaltrials.gov/ct2/results?term=Rociletinib.

[20] Jänne, P. A.; Yang, J. C.-H.; Kim, D.-W.; Planchard, D.; Ohe, Y.; Ramalingam, S. S.; Ahn, M.-J.; Kim, S.-W.; Su, W.-C.; Horn, L.; Haggstrom, D.; Felip, E.; Kim, J.-H.; Frewer, P.; Cantarini, M.; Brown, K. H.; Dickinson, P. A.; Ghiorghiu, S.; Ranson, M., AZD9291 in EGFR Inhibitor–Resistant Non–Small-Cell Lung Cancer. *New England Journal of Medicine* **2015,** *372* (18), 1689-1699.

[21] Sequist, L. V.; Soria, J.-C.; Goldman, J. W.; Wakelee, H. A.; Gadgeel, S. M.; Varga, A.; Papadimitrakopoulou, V.; Solomon, B. J.; Oxnard, G. R.; Dziadziuszko, R.; Aisner, D. L.; Doebele, R. C.; Galasso, C.; Garon, E. B.; Heist, R. S.; Logan, J.; Neal, J. W.; Mendenhall, M. A.; Nichols, S.; Piotrowska, Z.; Wozniak, A. J.; Raponi, M.; Karlovich, C. A.; Jaw-Tsai, S.; Isaacson, J.; Despain, D.; Matheny, S. L.; Rolfe, L.; Allen, A. R.; Camidge, D. R., Rociletinib in EGFR-Mutated Non–Small-Cell Lung Cancer. *New England Journal of Medicine* **2015,** *372* (18), 1700-1709.

[22] Engel, J.; Richters, A.; Getlik, M.; Tomassi, S.; Keul, M.; Termathe, M.; Lategahn, J.; Becker, C.; Mayer-Wrangowski, S.; Grütter, C.; Uhlenbrock, N.; Krüll, J.; Schaumann, N.; Eppmann, S.; Kibies, P.; Hoffgaard, F.; Heil, J.; Menninger, S.; Ortiz-Cuaran, S.;

Heuckmann, J. M.; Tinnefeld, V.; Zahedi, R. P.; Sos, M. L.; Schultz-Fademrecht, C.; Thomas, R. K.; Kast, S. M.; Rauh, D., Targeting Drug Resistance in EGFR with Covalent Inhibitors: A Structure-Based Design Approach. *Journal of Medicinal Chemistry* **2015,** *58* (17), 6844-6863.

[23] Gayko, U.; Fung, M.; Clow, F.; Sun, S.; Faust, E.; Price, S.; James, D.; Doyle, M.; Bari, S.; Zhuang, S. H., Development of the Bruton's tyrosine kinase inhibitor ibrutinib for B cell malignancies. *Annals of the New York Academy of Sciences* **2015,** *1358* (1), 82-94.

[24] Pan, Z.; Scheerens, H.; Li, S.-J.; Schultz, B. E.; Sprengeler, P. A.; Burrill, L. C.; Mendonca, R. V.; Sweeney, M. D.; Scott, K. C. K.; Grothaus, P. G.; Jeffery, D. A.; Spoerke, J. M.; Honigberg, L. A.; Young, P. R.; Dalrymple, S. A.; Palmer, J. T., Discovery of Selective Irreversible Inhibitors for Bruton's Tyrosine Kinase. *ChemMedChem* **2007,** *2* (1), 58-61.

[25] O'Shea, J. J.; Kontzias, A.; Yamaoka, K.; Tanaka, Y.; Laurence, A., Janus kinase inhibitors in autoimmune diseases. *Annals of the Rheumatic Diseases* **2013,** *72* (suppl 2), ii111-ii115.

[26] Goedken, E. R.; Argiriadi, M. A.; Banach, D. L.; Fiamengo, B. A.; Foley, S. E.; Frank, K. E.; George, J. S.; Harris, C. M.; Hobson, A. D.; Ihle, D. C.; Marcotte, D.; Merta, P. J.; Michalak, M. E.; Murdock, S. E.; Tomlinson, M. J.; Voss, J. W., Tricyclic Covalent Inhibitors Selectively Target Jak3 through an Active Site Thiol. *Journal of Biological Chemistry* **2015,** *290* (8), 4573-4589.

[27] Tanaka, H.; Nishida, K.; Sugita, K.; Yoshioka, T., Antitumor efficacy of hypothemycin, a new Ras-signaling inhibitor. *Japanese journal of cancer research : Gann* **1999,** *90* (10), 1139-45.

[28] Barluenga, S.; Jogireddy, R.; Koripelly, G. K.; Winssinger, N., In Vivo Efficacy of Natural Product-Inspired Irreversible Kinase Inhibitors. *ChemBioChem* **2010,** *11* (12), 1692-1699.

[29] Winssinger, N.; Barluenga, S., Chemistry and biology of resorcylic acid lactones. *Chemical Communications* **2007,** (1), 22-36.

[30] Schirmer, A.; Kennedy, J.; Murli, S.; Reid, R.; Santi, D. V., Targeted covalent inactivation of protein kinases by resorcylic acid lactone polyketides. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (11), 4234-4239.

[31] ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT02418000.

[32] Kumar, V.; Schuck, E. L.; Pelletier, R. D.; Farah, N.; Condon, K. B.; Ye, M.; Rowbottom, C.; King, B. M.; Zhang, Z.-Y.; Saxton, P. L.; Wong, Y. N., Pharmacokinetic characterization of a natural product–inspired novel MEK1 inhibitor E6201 in preclinical species. *Cancer Chemotherapy and Pharmacology* **2011,** *69* (1), 229-237.

[33] Hagel, M.; Niu, D.; St Martin, T.; Sheets, M. P.; Qiao, L.; Bernard, H.; Karp, R. M.; Zhu, Z.; Labenski, M. T.; Chaturvedi, P.; Nacht, M.; Westlin, W. F.; Petter, R. C.; Singh, J., Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat Chem Biol* **2011,** *7* (1), 22-24.

[34] Daniel, K. B.; Sullivan, E. D.; Chen, Y.; Chan, J. C.; Jennings, P. A.; Fierke, C. A.; Cohen, S. M., Dual-Mode HDAC Prodrug for Covalent Modification and Subsequent Inhibitor Release. *Journal of Medicinal Chemistry* **2015,** *58* (11), 4812-4821.

[35] Visscher, M.; Arkin, M. R.; Dansen, T. B., Covalent targeting of acquired cysteines in cancer. *Current Opinion in Chemical Biology* **2016,** *30*, 61-67.

[36] Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S., The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protocols* **2015,** *10* (5), 733-755.
[37] Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W., Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins: Structure, Function, and Bioinformatics* **2012,** *80* (3), 871-883.
[38] Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C., Binding of small molecules to an adaptive protein–protein interface. *Proceedings of the National Academy of Sciences* **2003,** *100* (4), 1603-1608.
[39] Dong, G. Q.; Calhoun, S.; Fan, H.; Kalyanaraman, C.; Branch, M. C.; Mashiyama, S. T.; London, N.; Jacobson, M. P.; Babbitt, P. C.; Shoichet, B. K.; Armstrong, R. N.; Sali, A., Prediction of Substrates for Glutathione Transferases by Covalent Docking. *Journal of Chemical Information and Modeling* **2014,** *54* (6), 1687-1699.
[40] London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermančič, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J., Covalent docking of large libraries for the discovery of chemical probes. *Nat Chem Biol* **2014,** *10* (12), 1066-1072.
[41] Kathman, S. G.; Xu, Z.; Statsyuk, A. V., A Fragment-Based Method to Discover Irreversible Covalent Inhibitors of Cysteine Proteases. *Journal of Medicinal Chemistry* **2014,** *57* (11), 4969-4974.
[42] Miller, R. M.; Paavilainen, V. O.; Krishnan, S.; Serafimova, I. M.; Taunton, J., Electrophilic Fragment-Based Design of Reversible Covalent Kinase Inhibitors. *Journal of the American Chemical Society* **2013,** *135* (14), 5298-5301.
[43] Hopkins, A. L.; Groom, C. R.; Alex, A., Ligand efficiency: a useful metric for lead selection. In *Drug Discov Today*, England, 2004; Vol. 9, pp 430-1.
[44] Murray, C. W.; Rees, D. C., The rise of fragment-based drug discovery. *Nat Chem* **2009,** *1* (3), 187-192.
[45] Kumar, A.; Zhang, A. V. a. K. Y. J., Fragment Based Drug Design: From Experimental to Computational Approaches. *Current Medicinal Chemistry* **2012,** *19* (30), 5128-5147.
[46] Joseph-McCarthy, D.; Campbell, A. J.; Kern, G.; Moustakas, D., Fragment-Based Lead Discovery and Design. *Journal of Chemical Information and Modeling* **2014,** *54* (3), 693-704.
[47] Doak, B. C.; Morton, C. J.; Simpson, J. S.; Scanlon, M. J., Design and Evaluation of the Performance of an NMR Screening Fragment Library. *Australian Journal of Chemistry* **2013,** *66* (12), 1465-1472.
[48] Francis, C. L.; Kenny, P. W.; Dolezal, O.; Saubern, S.; Kruger, M.; Savage, G. P.; Peat, T. S.; Ryan, J. H., Construction of the CSIRO Fragment Library. *Australian Journal of Chemistry* **2013,** *66* (12), 1473-1482.
[49] Rahman, A., *Frontiers in Drug Design and Discovery*. Bentham Science Publishers: 2007.
[50] Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W., The 'rule of three' for fragment-based drug discovery: where are we now? *Nat Rev Drug Discov* **2013,** *12* (8), 644-644.
[51] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H., A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* **2003,** *8* (19), 876-877.

[52] Hajduk, P. J.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **2007,** *6* (3), 211-219.
[53] Hung, A. W.; Ramek, A.; Wang, Y.; Kaya, T.; Wilson, J. A.; Clemons, P. A.; Young, D. W., Route to three-dimensional fragments using diversity-oriented synthesis. *Proceedings of the National Academy of Sciences* **2011,** *108* (17), 6799-6804.
[54] Dahlin, J. L.; Nissink, J. W. M.; Strasser, J. M.; Francis, S.; Higgins, L.; Zhou, H.; Zhang, Z.; Walters, M. A., PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *Journal of Medicinal Chemistry* **2015,** *58* (5), 2091-2113.
[55] Baell, J.; Walters, M. A., Chemistry: Chemical con artists foil drug discovery. *Nature* **2014,** *513* (7519), 481-3.
[56] McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K., A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *Journal of Medicinal Chemistry* **2002,** *45* (8), 1712-1722.
[57] Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D., Is There a Difference between Leads and Drugs? A Historical Perspective. *Journal of Chemical Information and Computer Sciences* **2001,** *41* (5), 1308-1315.
[58] Baell, J. B.; Holloway, G. A., New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry* **2010,** *53* (7), 2719-2740.
[59] Kwarcinski, F. E.; Steffey, M. E.; Fox, C. C.; Soellner, M. B., Discovery of Bivalent Kinase Inhibitors via Enzyme-Templated Fragment Elaboration. *ACS medicinal chemistry letters* **2015,** *6* (8), 898-901.
[60] Allen, C. E.; Curran, P. R.; Brearley, A. S.; Boissel, V.; Sviridenko, L.; Press, N. J.; Stonehouse, J. P.; Armstrong, A., Efficient and facile synthesis of acrylamide libraries for protein-guided tethering. *Organic letters* **2015,** *17* (3), 458-60.
[61] Flanagan, M. E.; Abramite, J. A.; Anderson, D. P.; Aulabaugh, A.; Dahal, U. P.; Gilbert, A. M.; Li, C.; Montgomery, J.; Oppenheimer, S. R.; Ryder, T.; Schuff, B. P.; Uccello, D. P.; Walker, G. S.; Wu, Y.; Brown, M. F.; Chen, J. M.; Hayward, M. M.; Noe, M. C.; Obach, R. S.; Philippe, L.; Shanmugasundaram, V.; Shapiro, M. J.; Starr, J.; Stroh, J.; Che, Y., Chemical and Computational Methods for the Characterization of Covalent Reactive Groups for the Prospective Design of Irreversible Inhibitors. *Journal of Medicinal Chemistry* **2014,** *57* (23), 10072-10079.
[62] Schwöbel, J. A. H.; Koleva, Y. K.; Enoch, S. J.; Bajot, F.; Hewitt, M.; Madden, J. C.; Roberts, D. W.; Schultz, T. W.; Cronin, M. T. D., Measurement and Estimation of Electrophilic Reactivity for Predictive Toxicology. *Chemical Reviews* **2011,** *111* (4), 2562-2596.
[63] Enoch, S. J.; Ellison, C. M.; Schultz, T. W.; Cronin, M. T. D., A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Critical Reviews in Toxicology* **2011,** *41* (9), 783-802.
[64] Nishino, M.; Choy, J. W.; Gushwa, N. N.; Oses-Prieto, J. A.; Koupparis, K.; Burlingame, A. L.; Renslo, A. R.; McKerrow, J. H.; Taunton, J., Hypothemicin, a fungal natural product, identifies therapeutic targets in Trypanosoma brucei. *eLife* **2013,** *2*, e00712.
[65] Choy, J. W.; Bryant, C.; Calvet, C. M.; Doyle, P. S.; Gunatilleke, S. S.; Leung, S. S. F.; Ang, K. K. H.; Chen, S.; Gut, J.; Oses-Prieto, J. A.; Johnston, J. B.; Arkin, M. R.; Burlingame, A. L.; Taunton, J.; Jacobson, M. P.; McKerrow, J. M.; Podust, L. M.;

Renslo, A. R., Chemical–biological characterization of a cruzain inhibitor reveals a second target and a mammalian off-target. *Beilstein Journal of Organic Chemistry* **2013,** *9*, 15-25.

[66] Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A., Site-directed ligand discovery. *Proceedings of the National Academy of Sciences* **2000,** *97* (17), 9367-9372.

[67] Lodge, J. M.; Justin Rettenmaier, T.; Wells, J. A.; Pomerantz, W. C.; Mapp, A. K., FP tethering: a screening technique to rapidly identify compounds that disrupt protein-protein interactions. *MedChemComm* **2014,** *5* (3), 370-375.

[68] Erlanson, D. A.; Lam, J. W.; Wiesmann, C.; Luong, T. N.; Simmons, R. L.; DeLano, W. L.; Choong, I. C.; Burdett, M. T.; Flanagan, W. M.; Lee, D.; Gordon, E. M.; O'Brien, T., In situ assembly of enzyme inhibitors using extended tethering. *Nat Biotech* **2003,** *21* (3), 308-314.

[69] Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M., K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **2013,** *503* (7477), 548-551.

[70] Erlanson, D. A.; Arndt, J. W.; Cancilla, M. T.; Cao, K.; Elling, R. A.; English, N.; Friedman, J.; Hansen, S. K.; Hession, C.; Joseph, I.; Kumaravel, G.; Lee, W.-C.; Lind, K. E.; McDowell, R. S.; Miatkowski, K.; Nguyen, C.; Nguyen, T. B.; Park, S.; Pathan, N.; Penny, D. M.; Romanowski, M. J.; Scott, D.; Silvian, L.; Simmons, R. L.; Tangonan, B. T.; Yang, W.; Sun, L., Discovery of a potent and highly selective PDK1 inhibitor via fragment-based drug discovery. *Bioorganic & Medicinal Chemistry Letters* **2011,** *21* (10), 3078-3083.

[71] Raimundo, B. C.; Oslob, J. D.; Braisted, A. C.; Hyde, J.; McDowell, R. S.; Randal, M.; Waal, N. D.; Wilkinson, J.; Yu, C. H.; Arkin, M. R., Integrating Fragment Assembly and Biophysical Methods in the Chemical Advancement of Small-Molecule Antagonists of IL-2: An Approach for Inhibiting Protein−Protein Interactions†. *Journal of Medicinal Chemistry* **2004,** *47* (12), 3111-3130.

[72] Adjei, A. A., Blocking Oncogenic Ras Signaling for Cancer Therapy. *Journal of the National Cancer Institute* **2001,** *93* (14), 1062-1074.

[73] Wang, Y.; Kaiser, C. E.; Frett, B.; Li, H.-y., Targeting Mutant KRAS for Anticancer Therapeutics: A Review of Novel Small Molecule Modulators. *Journal of medicinal chemistry* **2013,** *56* (13), 5219-5230.

[74] Patricelli, M. P.; Janes, M. R.; Li, L.-S.; Hansen, R.; Peters, U.; Kessler, L. V.; Chen, Y.; Kucharski, J. M.; Feng, J.; Ely, T.; Chen, J. H.; Firdaus, S. J.; Babbar, A.; Ren, P.; Liu, Y., Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. *Cancer Discovery* **2016**.

[75] Hunter, J. C.; Gurbani, D.; Ficarro, S. B.; Carrasco, M. A.; Lim, S. M.; Choi, H. G.; Xie, T.; Marto, J. A.; Chen, Z.; Gray, N. S.; Westover, K. D., In situ selectivity profiling and crystal structure of SML-8-73-1, an active site inhibitor of oncogenic K-Ras G12C. *Proceedings of the National Academy of Sciences* **2014,** *111* (24), 8895-8900.

[76] Lim, S. M.; Westover, K. D.; Ficarro, S. B.; Harrison, R. A.; Choi, H. G.; Pacold, M. E.; Carrasco, M.; Hunter, J.; Kim, N. D.; Xie, T.; Sim, T.; Jänne, P. A.; Meyerson, M.; Marto, J. A.; Engen, J. R.; Gray, N. S., Therapeutic Targeting of Oncogenic K-Ras by a Covalent Catalytic Site Inhibitor. *Angewandte Chemie International Edition* **2014,** *53* (1), 199-204.

[77] Copeland, R. A.; Pompliano, D. L.; Meek, T. D., Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov* **2006,** *5* (9), 730-739.
[78] Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frodin, M.; Taunton, J., Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat Chem Biol* **2012,** *8* (5), 471-6.
[79] Bradshaw, J. M.; McFarland, J. M.; Paavilainen, V. O.; Bisconte, A.; Tam, D.; Phan, V. T.; Romanov, S.; Finkle, D.; Shu, J.; Patel, V.; Ton, T.; Li, X.; Loughhead, D. G.; Nunn, P. A.; Karr, D. E.; Gerritsen, M. E.; Funk, J. O.; Owens, T. D.; Verner, E.; Brameld, K. A.; Hill, R. J.; Goldstein, D. M.; Taunton, J., Prolonged and tunable residence time using reversible covalent kinase inhibitors. *Nat Chem Biol* **2015,** *11* (7), 525-531.
[80] Hardy, J. A.; Lam, J.; Nguyen, J. T.; O'Brien, T.; Wells, J. A., Discovery of an allosteric site in the caspases. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (34), 12461-12466.
[81] Scheer, J. M.; Romanowski, M. J.; Wells, J. A., A common allosteric site and mechanism in caspases. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (20), 7595-7600.
[82] Datta, D.; Scheer, J. M.; Romanowski, M. J.; Wells, J. A., An Allosteric Circuit in Caspase-1. *Journal of molecular biology* **2008,** *381* (5), 1157-1167.
[83] Hardy, J. A.; Wells, J. A., Dissecting an Allosteric Switch in Caspase-7 Using Chemical and Mutational Probes. *The Journal of Biological Chemistry* **2009,** *284* (38), 26063-26069.
[84] Gao, J.; Sidhu, S. S.; Wells, J. A., Two-state selection of conformation-specific antibodies. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106* (9), 3071-3076.
[85] DeLano, W. L.; Ultsch, M. H.; de, A. M.; Vos; Wells, J. A., Convergent Solutions to Binding at a Protein-Protein Interface. *Science (New York, N.Y.)* **2000,** *287* (5456), 1279-1283.
[86] Wells, J. A.; McClendon, C. L., Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007,** *450* (7172), 1001-9.
[87] Gold, M. G.; Barford, D.; Komander, D., Lining the pockets of kinases and phosphatases. *Current Opinion in Structural Biology* **2006,** *16* (6), 693-701.
[88] Sadowsky, J. D.; Burlingame, M. A.; Wolan, D. W.; McClendon, C. L.; Jacobson, M. P.; Wells, J. A., Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proceedings of the National Academy of Sciences of the United States of America* **2011,** *108* (15), 6056-6061.
[89] Chio, C. M.; Lim, C. S.; Bishop, A. C., Targeting a Cryptic Allosteric Site for Selective Inhibition of the Oncogenic Protein Tyrosine Phosphatase Shp2. *Biochemistry* **2015,** *54* (2), 497-504.
[90] Chio, C. M.; Yu, X.; Bishop, A. C., Rational design of allosteric-inhibition sites in classical protein tyrosine phosphatases. *Bioorganic & Medicinal Chemistry* **2015,** *23* (12), 2828-2838.
[91] Bowman, G. R.; Bolin, E. R.; Hart, K. M.; Maguire, B. C.; Marqusee, S., Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proceedings of the National Academy of Sciences* **2015,** *112* (9), 2734-2739.

Chapter 2

A LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY METHOD FOR SCREENING DISULFIDE TETHERING FRAGMENTS

Contributing Authors:
JULIA L. DAVIES, CONNIE MERRON, PIERCE OGDEN, ELINE SIJBESMA, CHRISTIAN OTTMANN, ADAM R. RENSLO, CHRISTOPHER WILSON, MICHELLE R. ARKIN

INTRODUCTION

The last decade has seen an increase in the development of covalent inhibitors as potential therapeutic agents. This interest has been driven by an appreciation of the advantages of covalent mechanisms of inhibition[1,2], including the ability to overcome resistance, such as in EGFR gatekeeping mutations[3], the opportunity to increase affinity for otherwise 'undruggable' targets, and distinct pharmacokinetic properties due to very long target-residency times[4-5]. A barrier to the pursuit of such compounds has been the perception that electrophilic drugs present greater risk due to nonspecific binding to off-targets, formation of reactive metabolites, or rapid inactivation by reaction with glutathione or other endogenous nucleophiles[6-8]. However, the design and synthesis of covalent inhibitors, particularly targeting cysteine residues, has proven an effective discovery approach for select targets and therapeutic areas[9]. Furthermore, covalent inhibitors have been used as chemical probes of proteins with native or engineered cysteine residues. These success stories have utilized reversible adduct formation, such as disulfides[10] and cyanoacrylamides [11], or irreversible electrophiles [12-14].

As interest in covalent drug discovery has grown, so have analytical techniques to screen for adduct formation, as well as chemical methodologies to prepare disulfide-based and electrophilic compound libraries [12, 15-16]. Despite these improvements, the largest reported

screen of an electrophile library involved just 1000 compounds [13], similar in size to the UCSF Small Molecule Discovery Center (SMDC) 1600-member disulfide-fragment library (Appendix Fig 1). Library sizes reflect several challenges inherent to the goal of discovering selective covalent inhibitors. Adduct-forming libraries are generally custom synthesized to normalize chemical reactivity and optimize structural diversity[12, 14-16]. Ideally, covalent ligand binding involves initial non-covalent recognition of the protein surface, followed by reaction with a proximal nucleophilic residue on the protein. If a compound is too reactive, binding is dominated by the energetics of covalent bond formation and is insensitive to molecular recognition (such chemotypes are unfortunately ubiquitous in many HTS libraries, and act as "pan assay interference compounds", or PAINS) [17]. At the same time, small changes to compound structure can impact chemical reactivity through electronic or steric effects, obscuring underlying structure-activity relationships that derive from molecular recognition of the target. Well-designed libraries therefore seek to normalize reactivity, either by selecting electrophiles with lower functional-group sensitivity [14] or by separating the diverse structure elements from the reactive group using linkers [10]. The design of covalent compound libraries and the development of effective covalent screening conditions must therefore control for the differing reactivity of screening compounds, and/or include counter-screens to establish selectivity [2].

When identifying covalent ligands is the goal, it is reasonable for the primary screen to detect the formation of a covalent bond, with secondary screens for biochemical and cellular activity. Methods for measuring covalent protein modification are usually based on liquid-chromatography mass spectrometry (LC/MS), analyzing either intact protein or proteolytic peptides (LC/MS/MS). The chromatographic step in tandem MS generally takes > 10 minutes

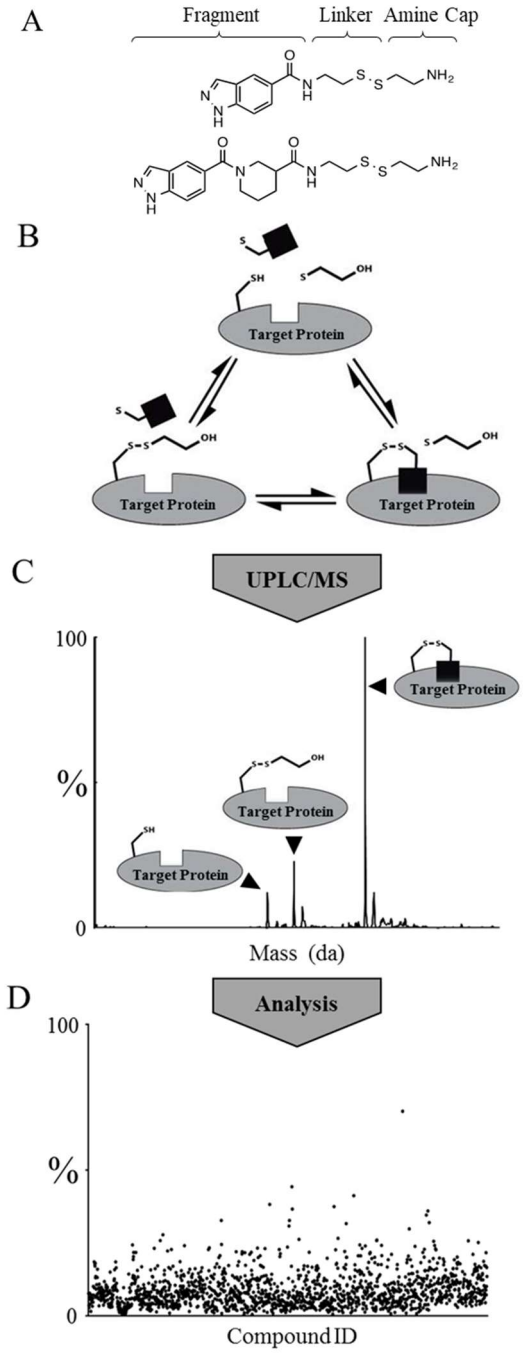


Figure 1. Liquid chromatography/mass spectrometry (LC/MS) screening workflow. (A) Examples of structures from the tethering library.15,16 (B) Labeling reaction scheme. Target protein, β-mercaptoethanol (βME), and various fragments (black square) are mixed in individual wells of a 384-well plate and incubated until equilibrium. (C) Rapid ultra-pressure liquid chromatography (UPLC) desalting, time-of-flight (TOF) detection, and m/z deconvolution identify unlabeled, βME-capped, and fragment-bound protein species. (D) Detected species are checked for expected fragment adduct formation and plotted as a percentage of protein that is fragment bound. Results are checked for data quality and uploaded to an internal database where hits are selected for follow-up.

and is therefore incompatible with demands of high-throughput screening (HTS), where seconds per sample is ideal. Intact protein detection has been reported at ~3min/sample in LC formats that take advantage of Ultra-Pressure Liquid Chromatography (UPLC) [18] and as quickly as 1.5 min/sample at high concentrations (> 10 uM) with flow-injection analysis [19]. Solid-phase extraction MS (SPE-MS) has been shown to be a viable alternative to LC/MS with reported

speeds of 20s/sample [13]. While fast, SPE-MS does not allow fractionation of complex samples through chromatography. Typically, only the expected masses – rather than a full spectrum – are recorded, which can lead to false positives for noisy spectra and loss of information about multiple adduct formation [13]. Finally, SPE-MS is a relatively insensitive MS method, using high ng/low μg amounts of protein/injection; screening is therefore done with micromolar concentrations of protein, limiting the ability to distinguish high-affinity binders and measure apparent binding affinities.

In 2013, the intact protein LC/MS methods in the SMDC functioned in a 96-well format with a 360s/sample chromatographic cycle time. The library was formatted into 20 plates, requiring 10 overnight runs to complete data collection. Semi-automated processing and manual QC added additional researcher time, for a total of 135 hours of instrument time and 10 hours of data inspection time. To address the LC/MS limitations to collecting high-throughput tethering screens, the SMDC purchased an Acquity/Xevo G2S (Waters). Here, we report the optimization of our intact protein LC/MS method for the rapid (84 sec/sample) screening of covalent small molecules using a custom 1600 compound library of disulfide–bearing fragments. We achieve a totally screening time of 40 hours (3.5-fold faster) and incorporate automated data processing, QC, and database storage. While our method remains 4-fold slower than available SPE-MS methods [13], our approach takes advantage of efficient UPLC desalting to inject less sample. For an example 19.5 kDa protein, Campuzano and colleagues' SPE-MS method has detection limits of 40 ng and screening injections of 400 ng (10 μL of 2 μM). Across 31 proteins of various molecular weights (MW), our method has detection limits of 0.2-20 ng, with screening injections of 12-120 ng (6μL of 100-500 nM). This enables screening of our library, including assay

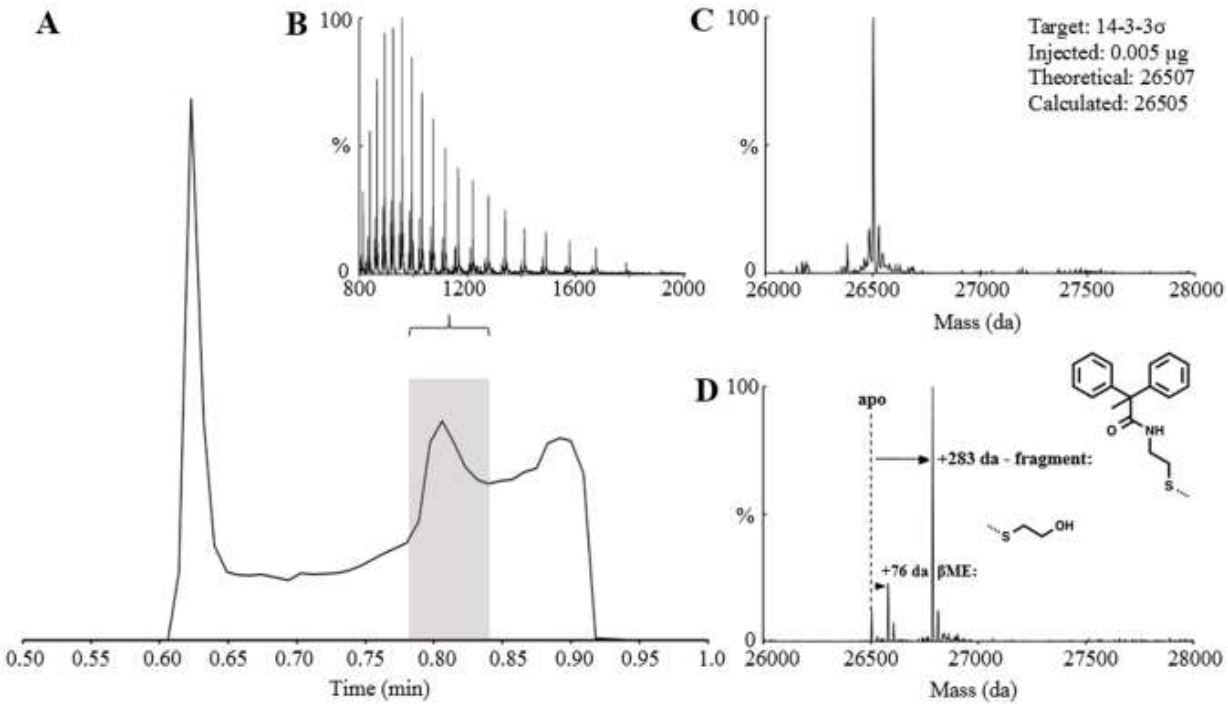


Figure 2. Liquid chromatography/mass spectrometry (LC/MS) data and processing. (A) Total ion count trace of liquid chromatography step. Flow before 0.6 min and after 0.9 min is diverted to waste with Xevo G2S fluidics. (B) The peak corresponding to protein ions (0.78–0.84 min) is combined, background subtracted, and reported as m/z. (C) MaxEnt (maximum entropy) deconvolution of the m/z charge spectrum identifies the masses present in a sample containing unlabeled protein. (D) MaxEnt spectrum deconvoluted from m/z shown in (B) of a reaction containing β-mercaptoethanol (βME) and screening compound, noting adduct formations.

development, with as little as 20 µg of purified protein. We have applied the method to a range of protein classes, collecting high-quality spectra at a speed capable of sustainably screening 1000 compounds/day. Custom pipelines facilitate data processing and analysis. Since this method uses commonly available equipment, has low protein consumption, and is analyzed with publicly available computational tools, it can be readily adopted in other laboratories.

Results & Discussion

Tethering screening technology

Figure 1 describes the Tethering screening methodology. Library compounds are built from structurally diverse fragment moieties (commonly < 200 Da), joined via amides, 1,2,3-triazoles, or other more extended linkers to a common aliphatic disulfide terminated with a basic amine to afford good solubility (Fig 1A). The common aliphatic disulfide moiety roughly normalizes library members' intrinsic reactivity in disulfide exchange reactions. Fragments are mixed with proteins containing native or engineered disulfides under conditions (pH, reduction potential) that favor thiolate-disulfide exchange. Once equilibrium is reached, the reaction mixture is injected onto a UPLC/MS system; UPLC offers partial purification and ESI-TOF mass spectrometry allows determination of protein and protein+adduct masses. Sample data are provided in Figure 2.

Method Optimization

The UPLC step was optimized for speed, signal/noise, and consistency by varying solvent flow rate (0.2-1.0 mL/min), column chemistry (C4, C8, C18), and elution strategy. A 0.4 mL/min flow over a 50 mm C4 column with a rapid (10s) gradient provided the fastest desalting which still afforded separation of proteins from post-elution noise (Fig 2A). A second 'wash' elution immediately followed the detected gradient to reduce carry-over of compounds and proteins on the C4 column (Appendix Fig 2). Flow diversion to waste before 0.3 min and after 0.9 min minimized contamination of the Xevo ion source.

We optimized the Xevo G2 LC/MS ionization conditions for detection of various proteins between 500-5000 m/z. Varying cone voltage (80-200 V), desolvation temperature (350-650 °C), the source capillary proximity to the cone, and angle toward the cone led us to the settings described in the Materials & Methods. We then performed a limit of detection (LOD) test on a series of proteins with varying molecular weight, without modifying the experimental or analysis

parameters (Fig 3A). LOD was defined as the lowest concentration at which a given sample could be successfully processed in the data analysis pipeline; LOD values varied from 5-10 nM (ca. 1-5 ng per 6 ul injection; 12 proteins) to 250 nM (5 proteins). Representative chromatograms, m/z spectra and deconvoluted masses from a range of protein classes and MW are shown in Fig 3B-E.

Assay Development

Assay development for screens followed a 3-step process. First, protein concentration was selected to be 2-fold LOD. For example, various cysteine mutants of adapter protein 14-3-3σ have detection limits of 10-50 nM (0.2-2.5 ng; Fig 3), and we selected a screening concentration of 100 nM. Second, tethering constructs were probed for reactivity with a titration of β-mercaptoethanol (βME), a thiol capable of forming a disulfide with an available cysteine thiolate, to confirm solvent accessibility and chemical reactivity of the target cysteine [10]. Screens were run from 100-1000 uM βME, and screening conditions were selected where a minor βME peak (ca. 20%) was present. Higher βME concentration resulted in a more stringent screen by providing competitor and increasing reduction potential of the mixture; selecting an appropriate screening concentration allowed tuning of the signal/noise and hit-rate. Notably, some cysteines showed no βME labeling during assay development but resulted in normal screening datasets.

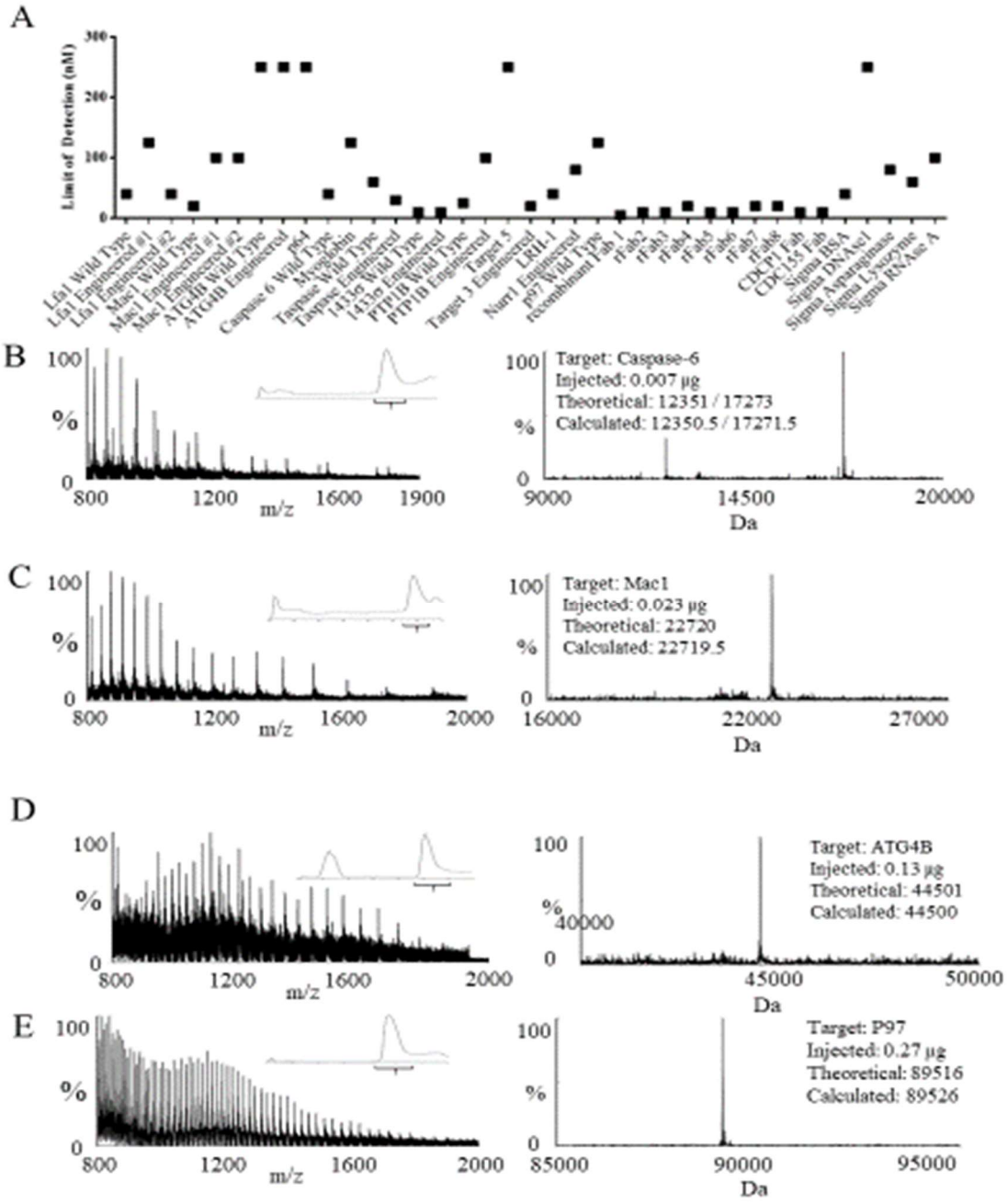


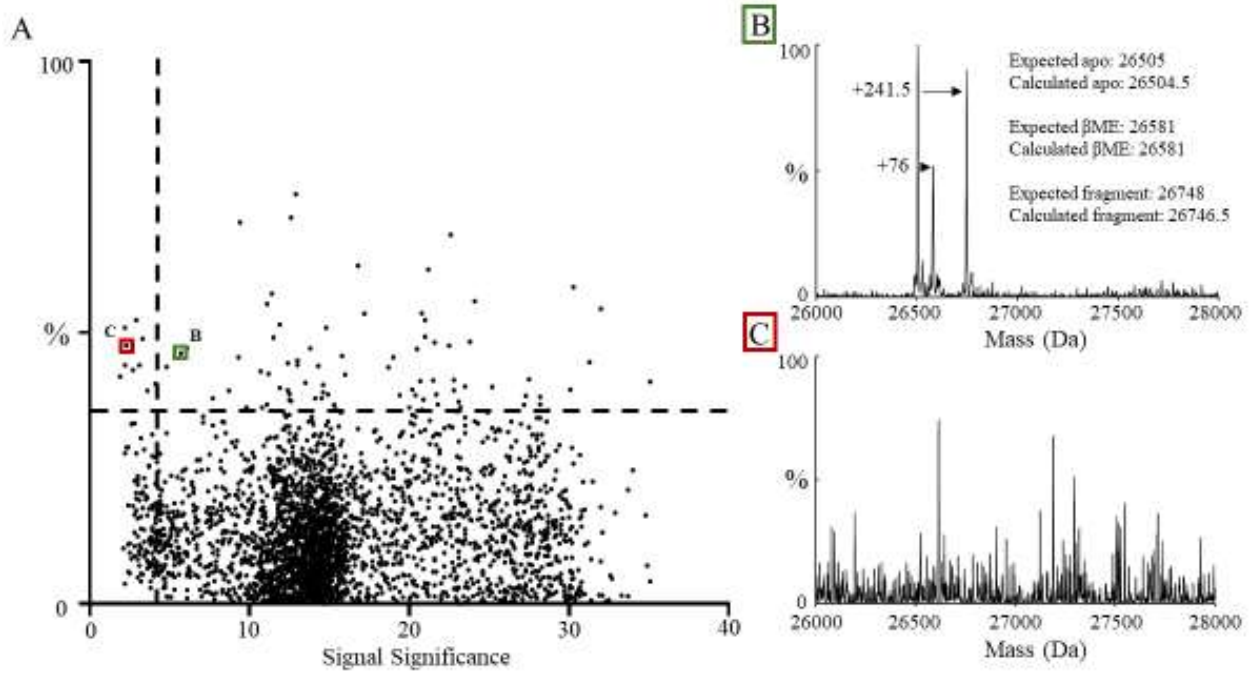
Figure 3. Liquid chromatography/mass spectrometry (LC/MS) data across molecular weight (MW) and class. (A) Limit of detection studies. Using a 5-µL injection for a wide range of protein samples, the limit of detection ranged from 5 to 250 nM for our recombinant samples and a suite of controls (Sigma). (B) A low MW target, caspase-6, is a tetramer containing small and large subunits, which are resolved by MaxEnt1. (C) An intermediate MW target, the I-domain of Mac1 is a 22.7-kDa monomeric ligand-binding domain. (D) A higher MW target, ATG4B, is a monomeric cysteine protease. (E) A high MW target, P97, is a hexameric AAA+ ATPase that ionizes as an 89.5-kDa monomer. For each example protein, the left panel m/z spectrum is combined from the inset LC chromatogram, and the corresponding MaxEnt1 deconvolution is shown in the right panel, along with the amount injected and the theoretical and calculated mass.

As previously shown [10], disulfide labeling assays are thermodynamically (vs kinetically) controlled, balancing chemical reactivity with specific small-molecule/protein interactions. One benefit of directly injecting the biochemical reaction (vs methods requiring sample pre-processing) is access to time-course and washout experiments to test labeling equilibrium and reversibility. For example, by repeated injection of 2 µL from the same well containing a 100 µL reaction of 100 µM compound, 100 nM 14-3-3σ and buffer, the reaction reached equilibrium in 5 minutes and remained stable for 45 minutes (Appendix Fig 6). An aliquot was then diluted 1000-fold with reaction buffer and assayed to confirm reversibility. Typically, the stability of the target at selected protein and βME concentration was tested by incubation at room temperature for 1-3 hours before analysis. A time was selected where the signal intensity was stable and no change in signal or % βME labeling was observed, indicating thermodynamic equilibrium.

Primary Screening

The library of 1,600 disulfide fragments was stored in 384-well format in DMSO at 50 mM. 30 nL of the compound library was pinned into a reaction mixture of protein diluted into 20 mM TRIS or Ammonium Acetate pH ≥8.0, the high pH chosen to increase the concentration of thiolate and therefore facilitate thiolate/disulfide exchange. The exchange reaction was incubated until reaching equilibrium (1-3 hours) before beginning analysis.

The Acquity UPLC was equilibrated at initial conditions until ΔPSI over one minute was ≤1% of total system PSI (2-3 minutes) before beginning injections. Two plates of 320 compounds were queued simultaneously, with water in the first two and last two columns. Four dummy injections of HPLC-grade H_2O were included to remove impurities in the UPLC before injections. The experiment cycle time was 84 seconds, a rate that allowed us to complete two 384-well plates overnight (15 hours) and was sustainable over long periods of use. Initial tests



achieved tractable MS datasets with cycle times as fast as 50 seconds/sample but suffered from

Figure 4. Data set analysis. (A) A typical data set with each of the 1600 screening compounds plotted to compare signal significance of each sample versus its calculated percent bound. The horizontal dotted line is drawn at 3 standard deviations above the mean percent bound. The horizontal line is drawn at an arbitrary cutoff for low-quality samples determined by manual inspecting the data. (B) MaxEnt spectrum for a sample (green box) with medium signal significance (<10), where adduct formation and calculated percent bound are well correlated. (C) MaxEnt spectrum for a sample (red box) with low signal significance (<5) where high noise has artificially inflated the percent bound value.

column pressure buildup and salt residue deposits on the ion source, leading us to extend cycle time to decrease maintenance requirements. The speed of the LC step relies on minimizing the amount of labeling reaction injected per sample. Injecting more than 2-fold the LOD of a target protein leads to detectable carry-over between samples. Conversely, screening too close to the LOD results in low signal/noise and increases the false positive rate from compound ion-suppression. We find that 2-fold LOD allows for rapid, sustainable LC desalting. In 24 months of operation at 84 seconds/sample, we performed 184,301 injections over 6317 hours of experimental time, consuming 134 L of mobile phase. Including idle time, regular maintenance, and intermittent instrument repair, these values translated to 8.75 hours, 251 experiments and 0.18 L of solvent per day for two years. During this time, we performed screens of several target

proteins; representative screens are shown in Table 1. The method was broadly applicable and agnostic to target class or construct size. While we have not attempted to screen a protein >50 kDa, the method detected proteins ranging from 8-90 kDa (Fig 3).

Table 1. Screening Outcomes Across Representative Targets

Target Protein	*Protein Class*	*Protein Mass (kda)*	*Engineered/Native*	*Hit rate (>3sigma)*
ATG4B	Protease	44.5	Native	0.1 %
			Native[a]	1.4 %
			Engineered	1.5 %
Lfa-1	Integrin, I-domain	21.0	Engineered[b]	1.8 %
			Engineered[b]	1.6 %
Mac-1	Integrin, I-domain	22.8	Engineered[b]	2.6 %
			Engineered[b]	2.3 %
LRH-1[c,d]	Nuclear Receptor	28.3	Native	0.7 %
Target 1	Ubiquitin Ligase	8.78	Native	1.4 %
Target 2	Kinase	19.3	Native	0.4 %
Target 3[c]	Kinase	37.3	Engineered	0.6 %
14-3-3σ	Adapter Protein	26.5	Native	1.8 %
			Engineered[b]	2.8 %
			Engineered[b]	2.7 %

[a]Screened in the presence of a protein partner[b]Cys-mutants targeting the same pocket on respective target
[c]Screened 1280/1600 compounds
[d]Reference [25]

Data Processing

Raw screening data were processed with Waters OpenLynx program, software designed to apply a single Waters algorithm across large datasets. M/z data were combined across the total ion count (TIC) peak, subtracted, and analyzed with MaxEnt 1, a maximum entropy algorithm

for deconvoluting intact protein mass (Fig 2, 3). These data were reported as mass vs %, in .rpt format.

Due to the volume of data and the varying quality of individual spectra, we developed a high-throughput analysis algorithm to quantify adduct formation. OpenLynx output files were read and processed using a custom Pipeline Pilot (BIOVIA) protocol to quantify binding and indicate the quality of each experiment (Appendix Fig 5). Spectra were divided into small mass bins surrounding the expected masses for free protein, βME-capped protein, and protein bound to adduct, as well as one large bin for unexpected masses (Appendix Table 1). "Expected mass" bins included +/- 5 amu from the expected mass to accommodate resolution fluctuations due to signal/noise or drift of mass lock. The bin width could be varied from screen-to-screen to match sample quality, from +/- 2 to +/- 5 amu from target peaks. If bin overlap occurred due to a larger bin size (possible in lower quality data) or a similarity in mass between the adduct and the reductant (possible for small fragments), bins were adjusted by dividing the difference between the cap and adduct mass by 2, rounding down to the nearest integer. Within each bin, the intensities were summed and used to calculate the percent bound as in Eq. (1),

Eq. (1) $$\% \, bound = \frac{\sum i_{adduct} + \sum i_{double\ adduct}}{\sum i_{protein} + \sum i_{adduct} + \sum i_{double\ adduct}}$$

where the % of βME-protein adduct is included with 'protein'. The protocol also checked for double-adduct formation in constructs that had alternative nucleophilic residues, e.g., two exposed cysteine residues near compound-binding sites. The algorithm additionally identified unanticipated species and adducts by reporting a maximum intensity found outside of the expected mass ranges as a secondary peak. In fact, these data were used in one study to identify and correct incorrectly drawn structures in the database.

Screening hits could be identified by plotting % bound vs compound number, e.g., as shown in Fig 1D. However, this strategy was sensitive to false positives; during the +/- 5 amu binning step, experiments with low signal/noise could report high % labeling. To provide indicators of data quality, a "signal significance number", analogous to a signal to noise ratio, was generated by calculating the percentage of the sum of intensities in meaningful bins versus the sum of all intensities Eq. (2),

$$\text{Eq. (2)} \qquad signal\ significance = 100\ x\ \left(\frac{\sum i_{protein}+\sum i_{adduct}+\sum i_{double\ adduct}+\sum i_{secondary}}{\sum i_{protein}+\sum i_{adduct}+\sum i_{double\ adduct}+\sum i_{secondary}+\sum i_{noise}}\right).$$

False positives with high % bound but low signal significance were readily identified by plotting the results of Eq. (1) vs. the results of Eq. (2) (Fig 4). Importantly, wells with high labeling sometimes also reported low signal significance; thus, manual inspection of hits from the lowest 5% of the signal significance range was found to be necessary (Fig 4B-C). The protocol code for the analysis step has been uploaded with an example dataset to the publicly accessible ScienceCloud Protocol Exchange (Biovia)[21] as "Read and Analyze HTS LCMS RPT File". Additionally, the module code is included as text in the Supplemental Information. The outputs from Eq (1) and Eq (2) were then loaded into the SMDC's MySQL database for further analysis in a custom web application, HiTS [22].

In conclusion, we report an optimized LC/MS method for screening intact protein for covalent adduct formation, using a library of disulfide-capped fragments. By taking advantage of advances in UPLC and ESI-TOF technology, we developed an LC method capable of more rapid (<90s) and sustainable injections than previously reported [18]. The method can detect proteins across range of molecular weights and with varying amenability to electrospray ionization (Figs 3, 4). Additionally, the labeling reaction is directly injected, facilitating kinetic studies

(Appendix Fig 6). While our approach remains slower than extraction-based methods, it benefits from a LC desalting step to increase MS data quality, requiring less than ten ng of material per injection and 20-200 ug of protein for a full screen. The sensitivity allows the screening of low expressing and/or poorly ionizing proteins, and ability to characterize binding events over a wide affinity range. Finally, full MS spectra are collected and analyzed for unexpected adducts and for acceptable signal/noise (signal significance), allowing post-hoc inspection of data quality.

A throughput of 1000 compounds per day represents an advance in LC/MS-based screening which shifts the limiting factor in screening covalent compounds to the size of available libraries. We routinely screen and analyze our library of 1600 compounds in 3 days. Further increasing the throughput of LC/MS methods or screening compounds in mixtures will become attractive as larger libraries of electrophilic compounds become available.

Though our method is applicable to multiple target classes (Fig 3, Table 1), some targets are intractable due to protein instability at ≥10 °C or in low salt, highly reducing conditions. These limitations represent inherent facets of this approach, and targets not amenable to UPLC desalting would require a re-imagining of our screening conditions. In rare cases where the target protein is excessively hydrophobic and requires more robust chromatography, we have extended the elution gradient step from 15 seconds to 120-180 seconds, keeping all other parameters identical. While successful, the resulting screening time of 5 days could motivate the use of higher-throughput and higher-consumption methods such as SPE or Matrix-assisted laser desorption/ionization.

Applications of this and other LC/MS screening of covalent molecules extends beyond drug discovery. Adduct formation is a complex reaction, where reaction rate and equilibrium report on availability and reactivity of the nucleophile and the affinity of the probe molecule for

the local environment [2]. Experiments that control for compound reactivity and affinity can probe surface ligandability [23]. Screens can be run in the presence and absence of a PPI partner or an active-site ligand to identify or confirm active-site binding or allosteric regulation [24]. Combining the control of site-directed technologies with the sampling size of high-throughput experiments generates compelling data about a target protein and the molecules that bind to it.

Materials and Methods

Protein Expression and Purification

Desired WT sequences of target proteins were cloned from their respective cDNA into a pET15b plasmid containing a 6xHis affinity tag followed by a TEV protease cleavage site at the N-terminus. Cysteine mutations were made via Megawhop PCR [20] or QuikChange™ Site-Directed Mutagenesis Kit (Agilent). All constructs were verified by DNA sequencing.

Recombinant protein expression protocols for targets in Table 1 varied to obtain optimal yield. For example, Lfa1, Mac1 and 14-3-3σ were grown in *E. coli* Rosetta 2(DE3) at 37 °C until OD_{600} reached 0.3. The temperature was reduced to 25 °C and at OD_{600} = 0.6 expression was induced with 0.25 mM IPTG followed by overnight culture. Cells were harvested by centrifugation, resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM MgCl, 0.25 mM TCEP, 10 mM imidazole and 5% w/v glycerol, and lysed by microfluidization (Microfluidics). The soluble lysate fraction was incubated with HisPur™ Cobalt resin (Thermo), washed and eluted by gravity flow in lysis buffer containing 150 mM imidazole. To remove the 6xHis affinity tag, purified protein was incubated overnight at 4 °C with 0.5 mg recombinant TEV protease with its own 6xHis affinity tag and dialyzed with an excess of 20 mM HEPES pH 7.5,

250 mM NaCl, 10 mM MgCl, 0.25 mM TCEP and 5% w/v glycerol. TEV protease and uncleaved protein were removed by re-pass over a HisPur™ Cobalt resin column equilibrated in lysis buffer. Cleaved and re-passed protein was further purified by size exclusion chromatography on a Superdex 75 16/600 column (GE Healthcare) in 20 mM HEPES pH 7.5, 250 mM NaCl, 10 mM MgCl, and 5% w/v glycerol. Protein purity was confirmed via SDS-PAGE. WT protein identity and cysteine mutation presence were confirmed by intact protein LC/MS on a Xevo G2-S (Waters). Pure protein was concentrated to >5 mg/mL, flash frozen in LN_2 and stored at -80 °C.

Compound Library

A custom library of 1600 disulfide exchangeable compounds available at the UCSF Small Molecule Discovery Center (SMDC) was synthesized using parallel methods as previously described[15-16]. For screening, the compounds were arrayed in 384w plates as 50 mM solutions in DMSO.

Disulfide Tethering

Protein constructs containing target cysteines were diluted to screening concentration (Table 1) in 20 mM Tris pH 8.0. 15 μL of the dilute protein was plated into columns 3-22 of a 384-well Low Volume V-Well Greiner Bio plate, with water in rows 1-2 and 23-24. 30 nL of disulfide-capped fragments were pinned into the 320 wells containing protein with a Biomek FX (Beckman), and the reaction mixture was incubated for 1-3 hours at RT (depending on experimental determination of time-to-equilibrium). Two plates of compounds were prepared simultaneously for overnight data collection.

Liquid Chromatography

UPLC used an I-Class Acquity UPLC (Waters) using a BEH C4, 300 Å, 1.7 µm x 2.1 mm x 50 mm column. A flow rate of 0.4 mL/min was used with the gradient scheme outlined in Appendix Fig 2, operating at pressures 8000-10,000 psi. Mobile phase A was H_2O + 0.5% formic acid and B was acetonitrile + 0.5% formic acid. 6 µL of sample was drawn from 384-well low-volume plates and injected, a 12 s process. Post-injection wash of 50:50 MeOH:H_2O added 6 s to yield a total experiment time of 84 s. The UPLC was diverted to waste from time = 0 to 0.30 min, and again after 0.90 min; eluent from 0.30 to 0.90 min was routed to the mass spectrometer for detection. UV absorbance at 280 nM was collected for troubleshooting purposes during the experiment time of 0.30 min to 0.90 min.

Mass Spectrometry

Mass Spectrometry data was acquired on a Xevo G2-XS Quadropole Time of Flight mass spectrometer with a ZSpray ion source (Waters). Electrospray ionization (ESI) conditions were optimized for m/z signal intensity of a Leucine Enkacephalin dimer (LeuEnk) (Waters) peak at 1111.6 Da by direct infusion of 200 pg/µL solution MeOH:H_2O with 0.1% FA. The dimer peak was used because it falls in the typical m/z range of analyzed protein charge envelopes (1000-2000 Da). Two ng/µL LeuEnk was additionally used as a detector control with the ZSpray LockSpray system. Screening experiments were done at a capillary voltage 3.20 kV, cone voltage 40 V, source temperature 150 °C, desolvation temperature 650 °C, cone gas 50 L/hr, desolvation gas 1200 L/hr. Data was collected at 1 spectra/second from 50-5000 m/z.

Limit of Detection Experiments

Limit of Detection (LOD) experiments were run using the LC/MS conditions reported above. Protein samples were 2-fold serial diluted from 500-5 nM in 10mM TRIS pH 8.0 using Optima LC/MS-grade water (Fisher). Injections of increasing concentration were monitored by

manual inspection of the chromatogram until a protein peak began to appear (between 0.75-0.90 min). The LOD was defined as the first concentration at which the processing parameters below yielded the expected deconvoluted mass.

Data Processing

Raw LC/MS data files were batch processed with Waters OpenLynx within a MassLynx v4.1 environment. A maximum entropy algorithm for mass deconvolution, MaxEnt1, was used on background subtracted m/z spectra from the portion of the LC chromatogram containing protein signal. Peak picking of the chromatogram was performed with parameters noted in Appendix Fig 3 and always fell between 0.75-0.90 min. As noted in previous work [13], rare peak-picking errors in noisy data can be manually inspected and combined prior to deconvolution. The OpenLynx processing parameters subtracted m/z background between 750-2000 Da, with background defined as ≤1% maximum intensity. The 750-2000 Da range for m/z subtraction and deconvolution was chosen for general application to a range of target MW, but can be varied to match a target protein charge envelope. Deconvolution was performed with a range of +/- 6000 Da around the target's expected mass, a target resolution of 0.5 Da, with 20 iterations of MaxEnt1 (Appendix Fig 4). 384-well plates were batch-processed into one large .rpt file at an analysis rate of ~30s per sample.

The resulting .rpt text file was inspected for data quality within MassLynx. The expected highest abundance monoisotopic adduct masses were calculated for all compounds using Pipeline Pilot (BIOVIA) via a systematic transformation using a defined virtual reaction (Appendix Fig 4A). Once the expected adduct structure was verified, the highest abundance monoisotopic masses were registered using an adduct mass registration system through the Pipeline Pilot WebPort into a MySQL database. The protocol code for the adduct mass

registration system has been uploaded with an example compound set to the publicly accessible ScienceCloud Protocol Exchange (Biovia) as "Adduct Highest Abundance Monoisotopic Mass Registration". The mass of the protein-βME conjugate (cap) was calculated analogously. Protein and cap masses were registered via HiTS, a custom web application, into a MySQL database. Finally, a separate Pipeline Pilot algorithm used Eq. (1) (see Results and Discussion) to report adduct formation and Eq. (2) to provide a measure of data quality; the output was recorded in a MySQL database (Appendix Fig 4B).

REFERENCES

[1] Bauer, R. A., Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discovery Today* **2015,** *20* (9), 1061-1073.

[2] Hallenbeck, K.; Turner, D.; Renslo, A.; Arkin, M., Targeting Non-Catalytic Cysteine Residues Through Structure-Guided Drug Discovery. *Current Topics in Medicinal Chemistry* **2017,** *17* (1), 4-15.

[3] Engel, J.; Richters, A.; Getlik, M.; Tomassi, S.; Keul, M.; Termathe, M.; Lategahn, J.; Becker, C.; Mayer-Wrangowski, S.; Grütter, C.; Uhlenbrock, N.; Krüll, J.; Schaumann, N.; Eppmann, S.; Kibies, P.; Hoffgaard, F.; Heil, J.; Menninger, S.; Ortiz-Cuaran, S.; Heuckmann, J. M.; Tinnefeld, V.; Zahedi, R. P.; Sos, M. L.; Schultz-Fademrecht, C.; Thomas, R. K.; Kast, S. M.; Rauh, D., Targeting Drug Resistance in EGFR with Covalent Inhibitors: A Structure-Based Design Approach. *Journal of Medicinal Chemistry* **2015,** *58* (17), 6844-6863.

[4] Lu, H.; Tonge, P. J., Drug–target residence time: critical information for lead optimization. *Current Opinion in Chemical Biology* **2010,** *14* (4), 467-474.

[5] Copeland, R. A., The drug-target residence time model: a 10-year retrospective. *Nature reviews. Drug discovery* **2016,** *15* (2), 87-95.

[6] David-Cordonnier, M.-H.; Laine, W.; Joubert, A.; Tardy, C.; Goossens, J.-F.; Kouach, M.; Briand, G.; Thi Mai, H. D.; Michel, S.; Tillequin, F.; Koch, M.; Leonce, S.; Pierre, A.; Bailly, C., Covalent binding to glutathione of the DNA-alkylating antitumor agent, S23906-1. *European Journal of Biochemistry* **2003,** *270* (13), 2848-2859.

[7] Johnson, D. S.; Weerapana, E.; Cravatt, B. F., Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future medicinal chemistry* **2010,** *2* (6), 949-64.

[8] Mah, R.; Thomas, J. R.; Shafer, C. M., Drug discovery considerations in the development of covalent inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2014,** *24* (1), 33-39.

[9] Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A., The resurgence of covalent drugs. *Nature reviews. Drug discovery* **2011,** *10* (4), 307-317.

[10] Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A., Site-directed ligand discovery. *Proceedings of the National Academy of Sciences* **2000,** *97* (17), 9367-9372.

[11] Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frodin, M.; Taunton, J., Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nature chemical biology* **2012,** *8* (5), 471-6.
[12] Allen, C. E.; Curran, P. R.; Brearley, A. S.; Boissel, V.; Sviridenko, L.; Press, N. J.; Stonehouse, J. P.; Armstrong, A., Efficient and Facile Synthesis of Acrylamide Libraries for Protein-Guided Tethering. *Organic Letters* **2015,** *17* (3), 458-460.
[13] Campuzano, I. D.; San Miguel, T.; Rowe, T.; Onea, D.; Cee, V. J.; Arvedson, T.; McCarter, J. D., High-Throughput Mass Spectrometric Analysis of Covalent Protein-Inhibitor Adducts for the Discovery of Irreversible Inhibitors: A Complete Workflow. *Journal of biomolecular screening* **2016,** *21* (2), 136-44.
[14] Kathman, S. G.; Xu, Z.; Statsyuk, A. V., A Fragment-Based Method to Discover Irreversible Covalent Inhibitors of Cysteine Proteases. *Journal of Medicinal Chemistry* **2014,** *57* (11), 4969-4974.
[15] Burlingame, M. A.; Tom, C. T. M. B.; Renslo, A. R., Simple One-Pot Synthesis of Disulfide Fragments for Use in Disulfide-Exchange Screening. *ACS Combinatorial Science* **2011,** *13* (3), 205-208.
[16] Turner, D. M.; Tom, C. T. M. B.; Renslo, A. R., Simple Plate-Based, Parallel Synthesis of Disulfide Fragments using the CuAAC Click Reaction. *ACS Combinatorial Science* **2014,** *16* (12), 661-664.
[17] Dahlin, J. L.; Nissink, J. W. M.; Strasser, J. M.; Francis, S.; Higgins, L.; Zhou, H.; Zhang, Z.; Walters, M. A., PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *Journal of Medicinal Chemistry* **2015,** *58* (5), 2091-2113.
[18] Everley, R. A.; Croley, T. R., Ultra-performance liquid chromatography/mass spectrometry of intact proteins. *Journal of chromatography. A* **2008,** *1192* (2), 239-47.
[19] Helmich, F.; van Dongen, J. L. J.; Kuijper, P. H. M.; Scharnhorst, V.; Brunsveld, L.; Broeren, M. A. C., Rapid phenotype hemoglobin screening by high-resolution mass spectrometry on intact proteins. *Clinica Chimica Acta* **2016,** *460*, 220-226.
[20] Miyazaki, K., MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. *Methods in enzymology* **2011,** *498*, 399-406.
[21] ScienceCloud Exchange. https://exchange.sciencecloud.com/exchange/ (accessed 7/XX).
[22] Michelle, R. A.; Kenny, K. H. A.; Steven, C.; Julia, D.; Connie, M.; Yinyan, T.; Christopher, G. M. W. a. A. R. R., UCSF Small Molecule Discovery Center: Innovation, Collaboration and Chemical Biology in the Bay Area. *Combinatorial Chemistry & High Throughput Screening* **2014,** *17* (4), 333-342.
[23] Hajduk, P. J.; Huth, J. R.; Fesik, S. W., Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* **2005,** *48* (7), 2518-25.
[24] Bowman, G. R.; Bolin, E. R.; Hart, K. M.; Maguire, B. C.; Marqusee, S., Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc Natl Acad Sci U S A* **2015,** *112* (9), 2734-9.
[25] de Jesus Cortez, F.; Suzawa, M.; Irvy, S.; Bruning, J. M.; Sablin, E.; Jacobson, M. P.; Fletterick, R. J.; Ingraham, H. A.; England, P. M., Disulfide-Trapping Identifies a New, Effective Chemical Probe for Activating the Nuclear Receptor Human LRH-1 (NR5A2). *PLOS ONE* **2016,** *11* (7), e0159316.

Chapter 3

A FRAGMENT-BASED SCREENING PARADIGM FOR THE DISCOVERY OF PROTEIN-PROTEIN INTERACTION STABILIZERS

Contributing Authors:
ELINE SIJBESMA, SEPPE LEYSEN, PIM J DE VINK, JORIS M C ADRIAANS, JOY K PETRICK, LUKASZ SKORA, WOLFGANG JAHNKE, LUC BRUNSVELD, CHRISTIAN OTTMANN, MICHELLE R ARKIN

INTRODUCTION

Disulfide tethering has been used in the discovery of fragments targeting orthosteric sites of defined function, for identification and targeting of allosteric sites, and for the general study of surface ligandability, as reviewed in Chapter 1. In the case of Caspase-1 and -7, disulfide fragments were targeted to a homo-dimer interface and stabilized the inactive form of the enzyme. We imagined a similar approach could be used to in the case of a hetero-dimer. Stabilization of a specific PPI by targeting a small-molecule to the interface could prove a general strategy to modulating PPIs for drug discovery and biological inquiry.

The family of eukaryotic hub proteins known as 14-3-3 recognize and dock onto specific, mainly phosphorylated, binding motifs, thereby influencing the enzymatic activity, stability, subcellular localization, folding or dimerization behavior of their partner proteins. With a wide range of cellular interaction partners (>300), 14-3-3 proteins have a central regulating role in many essential cellular signaling pathways and dysregulation of their PPIs are implicated in numerous pathologies[1,2]. Each monomer in a 14-3-3 dimer contains an independent phospho-accepting channel, referred to as the 14-3-3 primary interaction site. These ligand-binding grooves are arranged in an antiparallel orientation and enable a dimer to interact with two motifs simultaneously, either from a single protein or from two different binding partners [3,4] With a

central position in phosphorylation-dependent signaling pathways and strongly disease-associated interaction partners, including Raf kinases5, heat shock proteins (HSP)[6] and tumor suppressors (p53)[7], 14-3-3 proteins, and more specifically, their protein-protein interactions, are arising as attractive drug targets[8,9].

Protein-protein interactions have only been successfully targeted in drug discovery efforts in the last 10-15 years[10]. PPIs were considered 'undruggable' mostly because of the challenges posed by the typically flat and expanded nature of their interface area, in contrast to the well-defined cavities and pockets typically observed for native ligand binding sites on proteins[11]. This view has dramatically changed, due to an increased appreciation that some interfaces are more amenable to small-molecule binding than others, based on the structure and dynamics of the interfaces. The past two decades have also seen the development of computational, chemical and biophysical technologies suitable for tackling challenges related to PPI druggability[12,13]. However, whereas in recent years many inhibitors of PPIs have been published, the opposite strategy of PPI stabilization has not been systematically pursued, despite the clear biological potential[14].

In considering 14-3-3 PPI inhibition or stabilization, it is important to note that 14-3-3 binding can have either agonizing or antagonizing effects on its substrates. For instance, 14-3-3 binding to the ion channels TASK3 and CFTR increases membrane trafficking, leading to higher levels of functional channels (reference). By contrast, binding to estrogen receptor alpha (ERα) inhibits formation of ER dimers and nuclear translocation; stabilizing this PPI would therefore antagonize ERα function. Both inhibition and stabilization for 14-3-3 PPIs have been reported[15]. However, even though various approaches have resulted in the identification of novel small molecule and peptidic inhibitors, thus far the only compounds that are known to act

as stabilizers of 14-3-3 PPIs are natural products. Most studied and described are the fusicoccanes. Fusicoccin-A (FC-A) is a fungal toxin, found to stabilize the binary complex between 14-3-3 and the plasma membrane H+-ATPase (PMA) by enhancing their interaction affinity, even though the compound itself has a low binding affinity for the apo protein[16]. This mode of action is found for more natural compounds and supports the rationale of systematically screening for PPI stabilizers versus inhibitors.

A major challenge in the development of 14-3-3 PPI modulators is related to the cellular selectivity and specificity of the compounds. Inhibitors targeting the primary ligand-binding groove may affect a large number of 14-3-3 partner proteins resulting in off-target effects. An added benefit of looking for PPI stabilizing compounds therefore is an increase in client-specific selectivity. Furthermore, there are seven isoforms of 14-3-3 in humans; achieving specific binding to one isoform would further increase selectivity in stabilizing 14-3-3/client interactions.

We set out to develop a platform to screen for 14-3-3 PPI stabilizers targeting a druggable pocket that was previously validated with FC-A[17]. We envisioned taking advantage of the structural knowledge of 14-3-3/client complexes by applying a fragment-based drug discovery method known as 'Disulfide Tethering', which makes use of a cysteine on the protein as a selectivity handle for a disulfide-containing library, combined with X-Ray crystallography and NMR. Fragment-based drug discovery has to deal with the great difficulty of detecting low affinity binders, but has strong advantages over conventional high-throughput-screening, including the requirement for smaller library sizes and more efficient lead optimization[18,19]. Disulfide tethering is a site-directed approach, pioneered by Wells and Erlanson and co-workers[20,21]. This strategy discovers binding of low molecular weight ligands to a specific protein region through the formation of an intermediary disulfide tether. Reversible binding of

fragments via a disulfide bond increases their affinity at least 10-100 fold and hits can easily be detected by mass spectrometry [20,21]. Using this approach, small molecule inhibitors have been developed that bind to a gain-of-function cysteine mutant K-Ras (G12C) irreversibly and with high selectivity over the wild-type which furthermore resulted in the identification of an unknown allosteric pocket on K-Ras [22]. Additionally, a second small molecule binding site was identified by applying disulfide tethering to explore an adaptive region of interleukin 2 (IL-2), showing cooperative binding of ligands to two pockets where the first strongly influences the affinity of the second [23]. These results illustrate the impact of disulfide tethering on drug discovery projects with difficult targets and how powerful even fragments can be for exploring protein function. Whereas functional inhibition and PPI antagonists have been the main focus in previous examples, we are interested in cooperative binding by disulfide tethering applied to PPI stabilization. By translating the knowledge from these previous results to the aims of this field, we envisioned to screen for cooperative binding of a ligand and the binding epitope of a protein partner, thereby selecting for ligands that bind to 14-3-3 only in the presence of the 14-3-3 binding phospho-motif of a client protein, forming a more preferred, stable ternary complex.

The 14-3-3 binding partner we selected as a suitable and medically relevant test case is the Estrogen Receptor α (ERα), an important breast cancer target. ERα is a member of the superfamily of Nuclear Receptors (NRs), which act as ligand-dependent transcription-factors. Upon ligand binding, NRs undergo conformational changes and translocate to the nucleus where they can bind specific recognition elements in the DNA as homo- or heterodimers or monomers. Their function is strongly dependent on cofactor proteins and therefore many of these can be targeted to regulate NR transcriptional activity. The current modulation of ERα by small molecules is focused on the pocket of the ligand binding domain (LBD), thereby disrupting the

interaction of ERα and cofactor proteins necessary for its transcriptional activity. These compounds, like the partial agonist 4-hydroxy-tamoxifen, compete for binding with the endogenous ligand [17] β-estradiol (E2), resulting in complications including gain-of resistance in recurrent disease.

It was reported by de Vries-van Leeuwen et. al. that phosphorylation on ERα-Thr594 is essential for the interaction between ERα and the sigma isoform of 14-3-3 (14-3-3σ) [17]. Furthermore, they showed that this interaction inhibits the E2-dependent transcriptional activity of ERα by interfering with its dimerization, and that this negative regulation can be enhanced by stabilizing the interaction with 14-3-3 by the natural product Fusicoccin-A (FC-A) [17]. As illustrated in figure 1A, FC-A occupies a pocket in the 14-3-3 binding groove, in close proximity to the C-terminal valine of ERα, in a very similar fashion as the above mentioned interaction between 14-3-3 and PMA. Motivated by the physiological relevance for the stabilization of the interaction between ERα and 14-3-3 and with FC-A as a proof-of-principle, we aimed for the identification of small molecule stabilizers of this interaction by disulfide tethering.

Here, we report the first successful application of the disulfide tethering approach to identify 14-3-3 PPI stabilizers. We screened disulfide fragments to identify those that bound cooperatively to 1433 in the presence of an ERα phospho-peptide. The screening results indicate the position of the cysteine in the FC-A binding pocket has a large effect on the cooperativity of the fragment hits. We validated the most promising hits and studied the binding mechanism to explore the mechanism of cooperativity. Binding of the ligand increases the apparent affinity of the phospho-peptide for the protein by ~30-fold. Furthermore, the resulting fragments are selective towards the type of 14-3-3 client binding motif that they were selected for over other

interaction motifs. Tethering at protein-protein interfaces is a valid and promising drug discovery tool for facing the challenges of PPI stabilization.

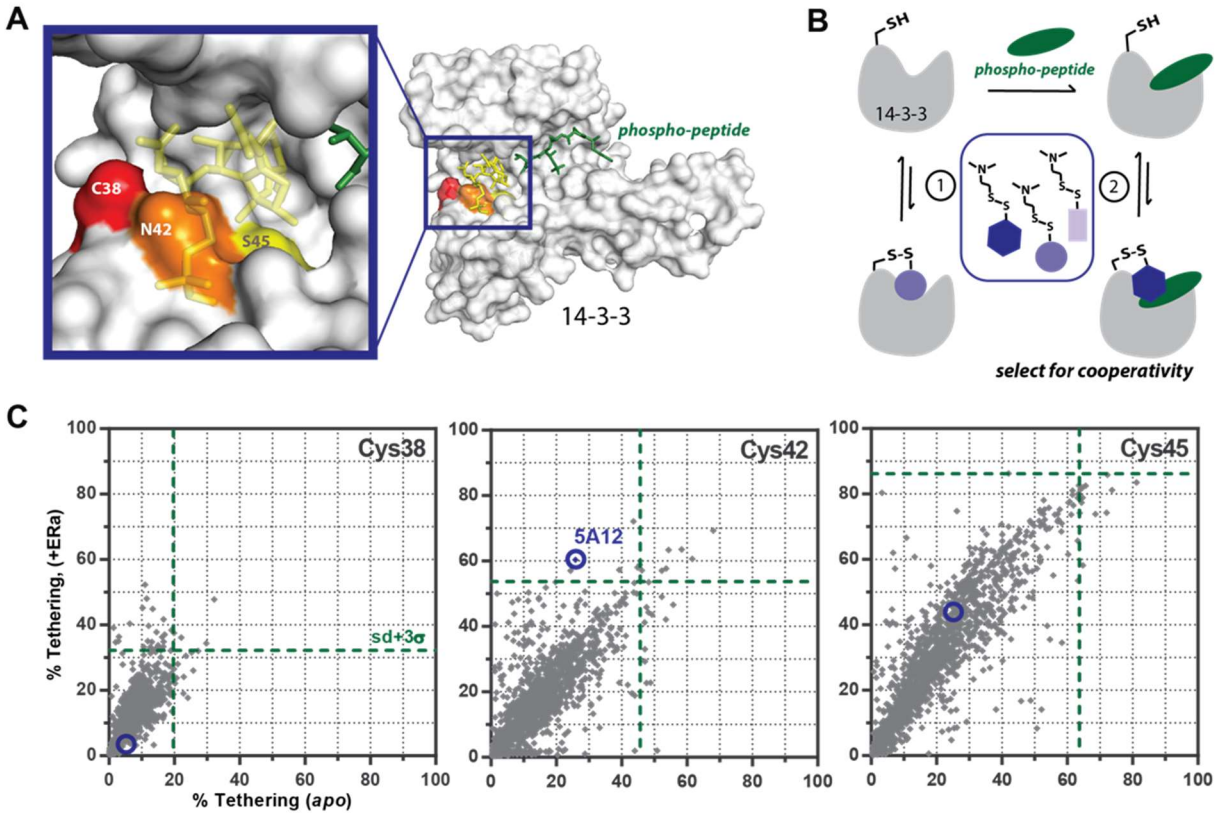


Figure 1: Disulfide tethering for cooperativity: overview of screening design and results. A) The target pocket for stabilizing the interaction of 14-3-3 (surface representation, white) and Estrogen Receptor α (green sticks) bound by FC-A (yellow sticks) provides the rationale for the site-directed approach. (PDB: 4JDD). Next to the native cysteine in 14-3-3σ on the 38-position (red surface), two constructs containing introduced cysteines at an alpha-helix turn +1 and +2 are included in the tethering screen (N42C and S45C; orange and yellow surface area). B) Schematic overview of the approach selecting for PPI stabilizers by disulfide tethering. 1: the cysteine-containing 14-3-3 is incubated with a disulfide library under reducing conditions. The equilibrium is shifted to the conjugated state only when the monophore has an inherent affinity for the protein binding pocket, thereby selecting for pocket binders. 2: the same screen is repeated with the cysteine-containing 14-3-3 protein bound to its binding partner-derived phospho-peptide. C) The overview of screening results in 2D-plots illustrate the correlation between percent tethering for the *apo* 14-3-3 and the 14-3-3/ERα complex for individual disulfide molecules (each represented by grey dot). The library was screened against 14-3-3σ containing native Cys38 (left); introduced Cys42 (middle) or Cys45 (right). In each graph 959996 is indicated by the purple circle.

RESULTS

A tethering screen for cooperativity

The interaction between ERα and the adaptor protein 14-3-3 comprises an interesting drug target and provides a possible alternative modulation of ERα transcriptional activity. 14-3-3σ is the only isoform that contains a native cysteine at the edge of the FC-A binding pocket (Cys38) (Fig1A). It was unclear if tethering fragments bound to Cys38 would be able to reach the far side of the FC-A pocket to make cooperative contacts with the C-terminus of ERα. We therefore designed two additional protein constructs, with introduced cysteines at positions 42 and 45 in 14-3-3, one and two α-helix turns closer towards the C-terminal valine of ERα, respectively. To identify fragments with preferential binding to 14-3-3/ERα complex, the three 14-3-3 constructs were each screened against a 1600-member disulfide library twice: once in apo form and once after incubation with an ERα-derived 14-3-3 binding phospho-peptide. In both cases, 1mM beta-mercaptoethanol (βME) was used as a reductant. Figure 1B illustrates the design of the disulfide-tethering screen. The protein or protein/ERα complex was incubated with disulfides fragments in a low-volume 384-well format and conjugate formation analyzed by intact protein LC/MS. For most screening samples, three peaks were observed in MS spectra, corresponding to the mass of apo 14-3-3, BME-capped 14-3-3, or the disulfide protein-conjugate (Figure 2A-B). The intensity of the protein-conjugate complex over the total of all protein peaks was determined for all samples using an automated custom pipeline as previously described (Chapter 2) [24].

Figure 1C illustrates the screening results for each protein construct, in presence and absence of the ERα phospho-peptide. In both contexts, the overall tethering rate of the fragment library increased as the target cysteine moved further into the FC-A binding pocket (Cys45 > Cys42 > Cys38 in Fig 1C), confirming the FC-A pocket was available for fragment binding and is an attractive site for identifying novel chemical matter. The correlation between tethering for

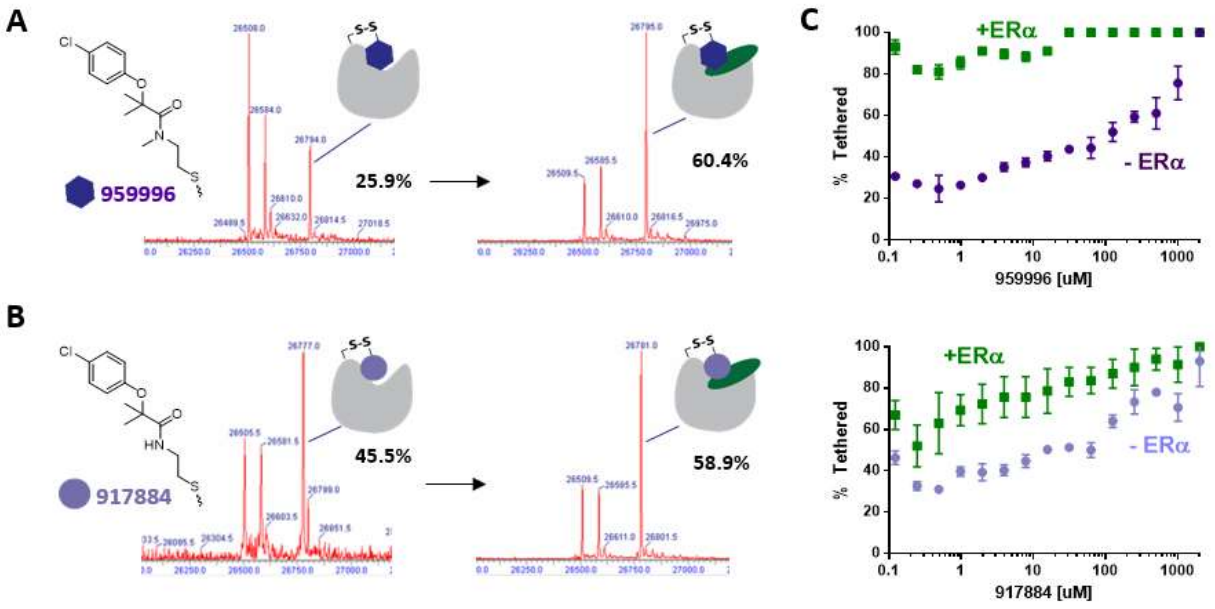


Figure 2: Validation of cooperativity of 14-3-3/ERα tethering hits. A) Deconvoluted LC/MS spectra of tethering screen results for 959996 conjugated to 14-3-3(Cys42) in *apo* (left) or ERα bound (right) state, resulting in 25.9% and 60.4% tethering, respectively. 14-3-3 expected mass: 26509 Da, BME capped mass: 26585 Da, protein-disulfide conjugate mass: 26795 Da. B) LC/MS spectra of tethering screen results for 917884 conjugated to 14-3-3 (Cys42) *apo* (45.5%) or in complex with ERα (58.9%). C) LC/MS dose-response curves for 959996 (top) and 917884 (bottom). Percentage of fragment-protein conjugate formation for titrations of disulfides to 14-3-3 (Cys42) *apo* (purple) and in complex with ERα (green). Experiments were performed at 100nM 14-3-3, 1mM TRIS pH 8.0, 1mM βME.

each compound in the 14-3-3 apo and 14-3-3/ERα complex were plotted in 2D graphs (Fig 1C). For each screen, hits were selected based on a cut-off for percentage tethering (3 standard deviations from the mean) and categorized into disulfides that were: (1) active in only the apo screen, indicating a preference for binding to 14-3-3 in absence of ERα-peptide; (2) bound preferentially to the 14-3-3/ERα complex, indicating an increase in binding affinity in presence of the phospho-peptide, and therefore demonstrating cooperative binding; or (3) equally bound in absence or presence of ERα.

Cys38 showed more hits for the 14-3-3/ ERα complex than for apo 14-3-3, but the percent tethering of hits was low (<55% bound), suggesting that fragments bound to Cys38 may not optimally access the FC-A pocket (Figure 1C, left). Conversely, the cysteine deepest in the

FC-A pocket, Cys45, yielded more hits for the apo screen, all of which had a high % tethering. Cys45-bound fragments may access both the FC-A pocket and parts of the ERα phospho-peptide pocket, therefore preferring the apo protein (Figure 1C, right). Satisfyingly, the intermediate Cys42 yielded hit fragments for both apo and phospho-peptide bound 14-3-3 (Fig 1C, middle). Together, these screening data show the positioning of the cysteine directly influences the available ligand binding pocket and the resulting hit compounds. Structurally described PPI interfaces such as the 14-3-3/ ERα complex allow the design of screening campaigns which result both in hits with a desired modulation: inhibitors and stabilizers are both accessible.

One advantage of tethering is that it detects covalent adduct formation that is stable only when the fragment has an inherent affinity for the binding pocket. An increase in % tethering therefore reports directly on an increase in affinity of the fragment for the target protein. This allows the identification of hits whose apparent affinity for 14-3-3 is increased in the presence of saturating ERα phospho-peptide by simple comparison of screening data. Indeed, some fragments demonstrated greatly increased binding in the presence of ERα; for example, tethering of compound 5A12 on Cys42 increased 2.3-fold from 25,9% for apo to 60,4% for the complex (Fig 1C). To validate screening hits with increased % tethering in the 14-3-3/ERα screen, we selected 13 structurally diverse fragment hits from the Cys42 and Cys45 screens for validation (Appendix Table 2).

Hit validation and quantification of cooperativity

Selected hits were validated in tethering dose-response experiments by collecting intact protein LC/MS spectra for titrations of hit fragment with 14-3-3 in the presence and absence of ERα phospho-peptide. A clear cooperative effect was observed for a number of hits, represented by a shift in EC_{50} of % tethering. Of these hits, two highly similar fragments selective for Cys42

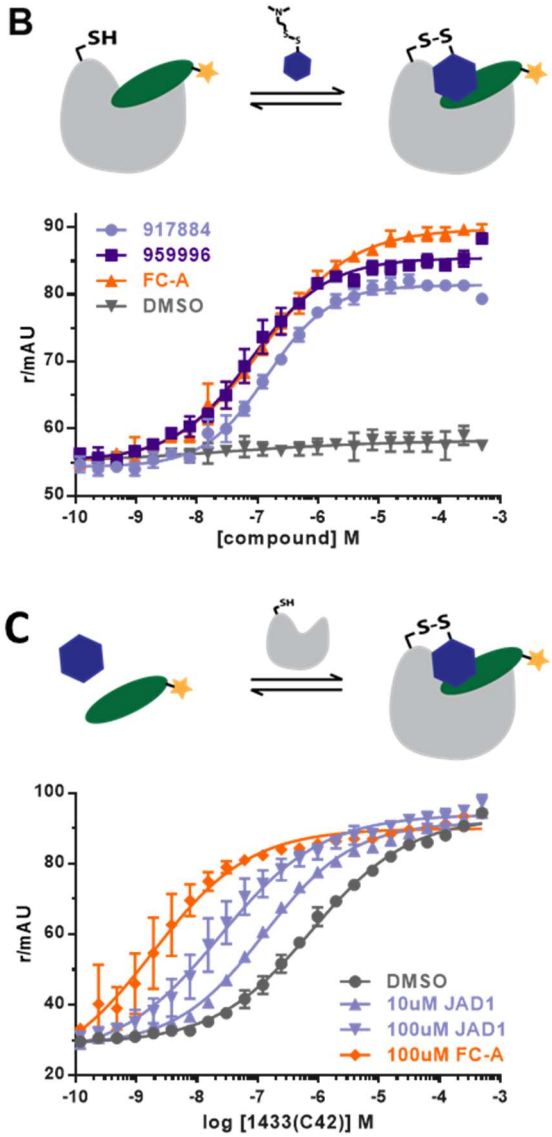


Figure 3: Fluorescence anisotropy experiments to quantify 14-3-3/ERα stabilization. A) Compound titration to 2μM 14-3-3 and 100nM FAM-ERα resulting in EC_{50} values for 917884 (142.6 ± 10.2 nM), 959996 (77.1 ± 7.6 nM) and FC-A (129.7 ± 10.6 nM). B) Protein titration to 100nM FAM-ERα and saturating concentration of 917884 or FC-A (100μM) or DMSO control. Dissociation constants (K_d^{ap}) were calculated from non-linear fitting of the data, resulting in 524 ± 33.8 nM for DMSO, 16.2 ± 3.5 nM for 917884 and 0.85 ± 0.88 nM for FC-A.

demonstrated strong preference for binding to the 14-3-3/ERα complex over the 14-3-3 protein alone (Fig 2C). Compound 917488 showed an ~200-fold increase in EC_{50} from ~1 mM for apo 14-3-3 to 3 μM for the 14-3-3/ ERα complex. Compound 959996, an N-methylated version of 917884, had an EC_{50} of ~100 μM for apo 14-3-3 but remained >80% tethered to the 14-3-3/ ERα complex down to 100 nM in 1 mM βME. Binding and cooperativity are two separate parameters in evaluating allosteric modulators; when tested against Cys45, 959996 and 917884 retained their 14-3-3 affinities but lost cooperativity with ERα, confirming that the precise placement of the fragment within the binding site determined its cooperativity.

To confirm that disulfide fragments also stabilize the binding of 14-3-3 to ERα, we determined the effect of 959996 and 917884 on the affinity of the ERα/14-3-3 complex. Using a fluorescein-labeled ERα peptide, we monitored ERα binding to 14-3-3 by fluorescence anisotropy. We first replicated our tethering results by performing a compound titration. The disulfides 959996 and 917884 and the positive control FC-A were titrated into Cys42 14-3-3 preincubated with FAM-labeled ERα phospho-peptide (the concentration at which 20% ERα was bound; Fig 3B). All three compounds increased the anisotropy of the FAM-ERα peptide, indicating an increased binding affinity for the peptide upon compound addition. As anticipated by the tethering titrations, 959996 was effective at slightly lower doses than 917884 (EC_{50} = 77.1 ± 7.6 nM nM and 142.6 ± 10.2 nM nM respectively). Notably, both had a similar EC_{50} as FC-A, a much larger but non-covalent molecule.

To quantify the increase in ERα affinity for 14-3-3 upon fragment binding, we reversed the anisotropy experiment. Increasing concentrations of 14-3-3 were titrated into a saturating concentration of FAM-ERα phospho-peptide in the presence of 10 or 100 μM of 917884, 100 μM FC-A, or DMSO (Fig 3C). The affinity of the 14-3-3/ERα complex was ~1 μM in DMSO and increased 500-fold to 2 nM in the presence of the positive control FC-A. Increasing concentrations of 917884 also strongly increased the affinity of ERα phospho-peptide (K_d = 100 nM in the presence of 10 μM 917884, 20 nM in the presence of 100 μM of 917884). These data confirmed the tethering screen hit 917884 stabilizes of the 14-3-3/ERα PPI.

Structural mechanism of 14-3-3/ERα PPI Stabilization

To understand the SAR and structural mechanism of 14-3-3/ERα PPI stabilization, we collected crystal structures of the fragment/14-3-3/ERα ternary complex for a series of disulfide fragments. Crystals of the binary N42C 14-3-3/ERα complex were grown under previously

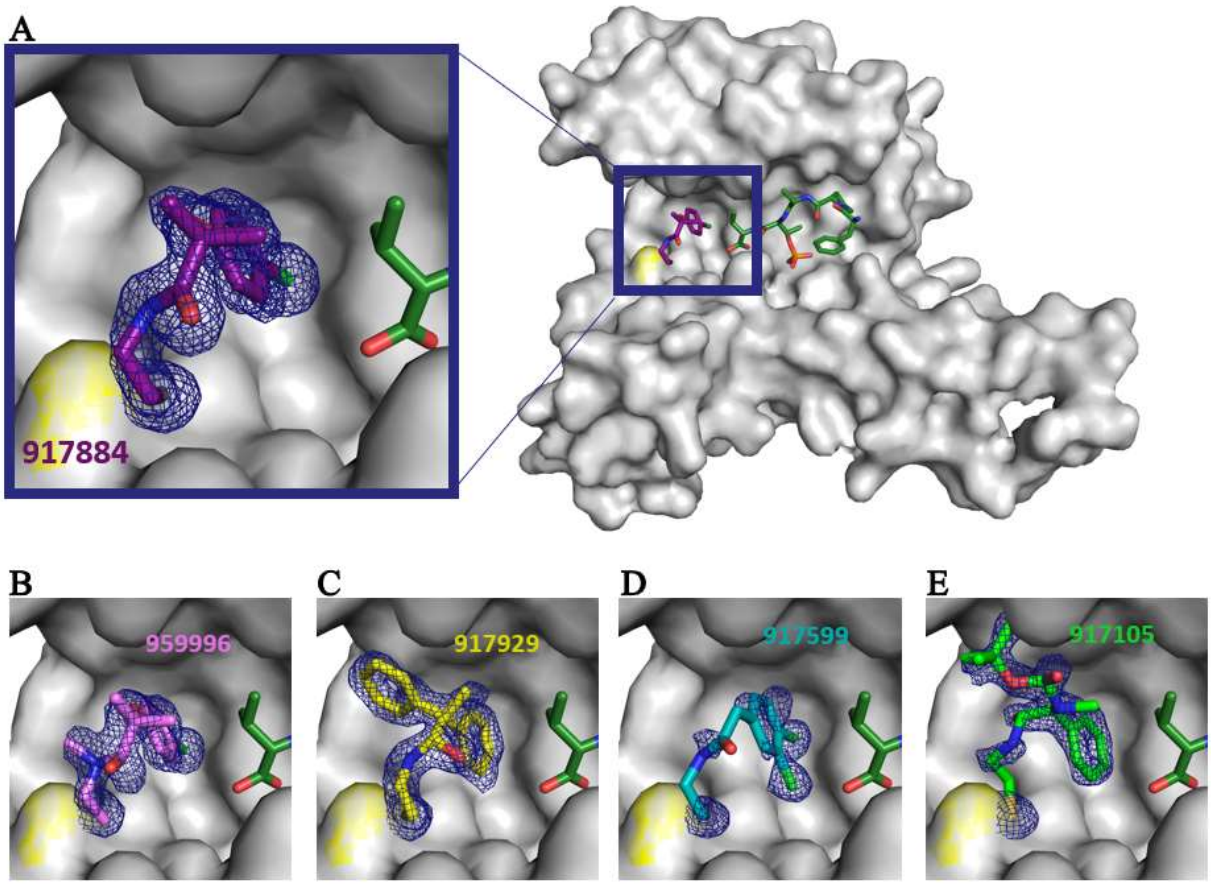


Figure 4: Structural Basis for N42C 14-3-3/ERα PPI Stabilization. Ternary crystal structures of 14-3-3, ERα phospho-peptide, and A) 917884 B) 959996 C) 917929 D) 917599 and E) 917105. Each fragment contains an aromatic ring which points toward the back of the PPI pocket and contacts 14-3-3 on one side and the C-terminal Valine of the ERα phospho-peptide on the other.

reported conditions and soaked with 1 mM of 959996, 917884, and a selection of analogs and other screening hits. After data collection and model refinement, electron density representing covalently-bound compound was visible for five of the twenty compounds targeting N42C (Fig 4). All five of the compounds contained an aromatic ring pointed into the back of the FC-A pocket, rotated to make a hydrophobic contact with the C-terminal V595 of ERα. The chloro-phenyl substitution in 917884 and 959996 is fully buried in the pocket, explaining the loss of affinity observed in analogs where the para-chloro-phenyl was modified. Remarkably, the additional contacts made by 917929, 917599 and 917105 correlated with decreased cooperativity with ERα (Appendix Fig 7). Comparison of the N42C hits with S45C hits reveals several close

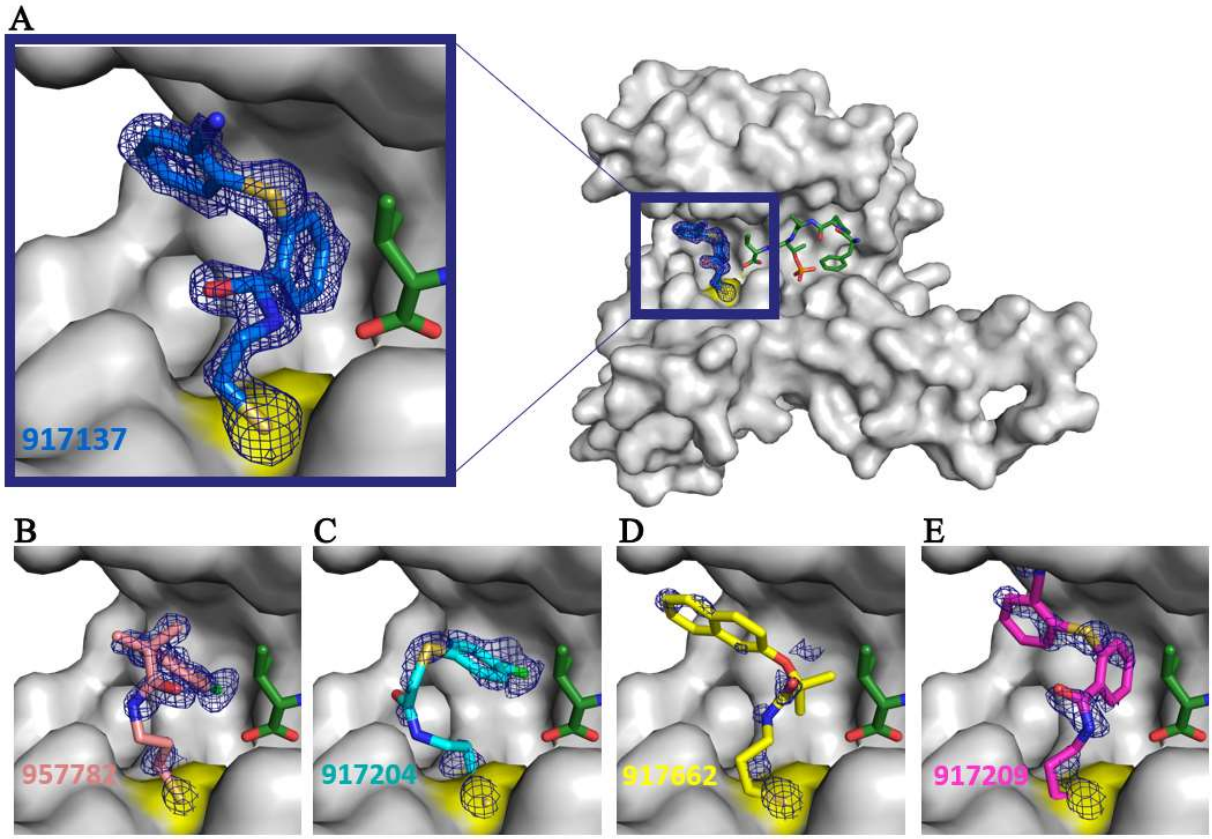


Figure 5: Structural Basis for S45C 14-3-3/ERα PPI Stabilization. Ternary crystal structures of 14-3-3, ERα phospho-peptide, and A) 917137 B) 957782 C) 917204 D) 917662 and E) 917209. Fragment density for 917204, 917662 and 917209 is limited and the displayed models are likely binding poses which fit the available density.

analogs of 917884, such as 957782, which contains the identical fragment moiety but has one more carbon in the aliphatic linker to the disulfide-forming thiol. To test whether 957782 and similar S45C hits bound to the 14-3-3/ERα complex with a similar mechanism as the N42C, we solved an additional five structures of compounds bound to S45C-14-3-3/ERα. The density observed for 917137 and 957782 allowed the unambiguous assignment of the fragment binding pose (Figure 5A-B). In both fragments, the aromatic ring occupied the same pocket as in the N42C hits, though in 957782 the chloro-phenyl ring is rotated to maintain the orientation of the chloro substitution, confirming the importance of the ion for binding at the interface (Appendix Fig 8). Other compounds in the S45C series show fragmented or minimal electron density, reducing confidence in the modeled binding pose (Fig 5C-E). However, density is visible in the

essential aromatic ring pocket for all three datasets, suggesting these fragments may also bind via a similar mechanism. This binding mechanism may explain which 917662 displays such poor electron density – the double ring is too bulky to fit in the chloro-phenyl pocket.

Stabilization Selectivity of 917884

14-3-3 binds to many phosphorylated proteins in the ERα binding pocket. The convergent binding mode of the N42C and S45C hits suggests the cooperativity of the tethering stabilizers with ERα could also occur for 14-3-3 partners which have C-terminal valines capable of making similar contacts. Conversely, the fragments may serve as antagonists to 14-3-3 binding proteins which do not adopt a pose similar to ERα, either due to extension of the peptide into the FC-A/fragment pocket, or a sequence change at the C-terminus.

To understand if the stabilization of our tethering hits through contacts with the C-terminal valine of ERα conveyed any selectivity for other 14-3-3 binding partners, we selected a panel of proteins with diverse C-terminal tail length and sequence for which binary 14-3-3/phospho-peptide structures have been reported. These proteins fall into two classes: ones which ought to form similar contacts with the tethering hits and therefore be cooperative, and ones which should sterically clash with the tethering pocket and therefore be antagonists. We tested 917884, given its excellent cooperativity profile in tethering and polarization experiments (Fig 3A-B) and strong electron density in the ternary compound/14-3-3/ERα complex (Fig 4A). Of the binding partners, TASK3 phosho-peptide has the most similar pose toERα in the phospho-peptide pocket and contains a C-terminal valine. Therefore, we predicted that the 14-3-3/TASK3 interaction would have been stabilized by 917884. Lastly, the EXOS phospho-peptide extends into the 917884 binding site, and therefore ought to act as an inhibitor of 917884 binding. To test

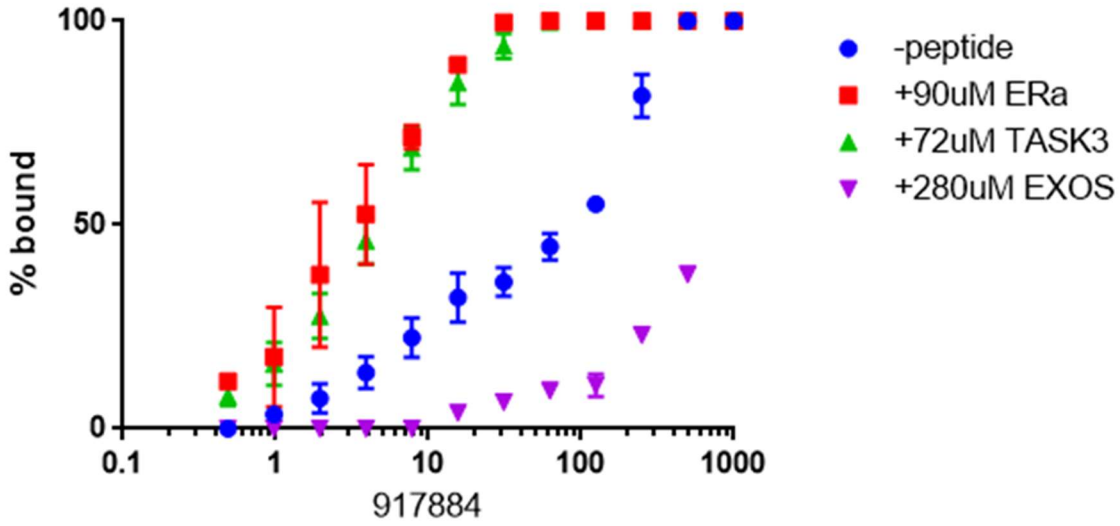


Figure 6: Selectivity of 917884 for 14-3-3 PPI Partner Peptides. Disulfide tethering titration of 917884 in the absence (blue) or presence of saturating concentrations for 14-3-3 binding peptides ERα (red), TASK3 (green), and EXOS (purple).

our predictions, we performed tethering dose-response experiments with 917884 in the presence and absence of saturating concentrations (10-fold the reported K_d) of ERα, TASK3, and EXOS (Fig 6). As expected, TASK3 increased the tethering efficiency of 917884 as well as ERα did (from EC_{50} = 100 μM to 3 μM). EXOS is indeed an antagonist of 917884 tethering, shifting the % tethering EC_{50} from 100 μM to ~500 μM. These data suggest 917884 could serve as a stabilizer for any phospho-peptide containing a C-terminal Valine which adopts a pose similar to ERα. However, additional screening campaigns modeled after the efforts reported here could target peptides with divergent C-termini, such as TAU, and yield class-selective stabilizers of 14-3-3 PPIs.

Discussion

The rationale and biological relevance for small molecule protein-protein interaction modulation comes from nature itself. Many fruitful drug discovery efforts have resulted in the development of PPI inhibitors, whereas the discovery of PPI stabilizers is lagging despite the physiologic benefit illustrated by many natural product stabilizers [14,15]. A possible explanation for this is that the rational design of small-molecule stabilizers faces different challenges for PPI stabilizers vs inhibitors. Inspiration for PPI inhibitors in many cases comes from targeting the small 'hot spots' of an interaction responsible for most of the binding energy and aiming to mimic one of the binding partners at the interface. This has, for example, resulted in stapled peptides and small molecule α-helical PPI mimetics [25,26]. On the contrary, stabilizers generally do not have starting points in nature; furthermore, stabilizers must not only bind to the protein(s) but also cause a difficult-to-predict allosteric effect. Here, we have developed and validated the use of disulfide tethering to select for a compound that acts as a stabilizer towards 14-3-3/ ERα. Such a compound would bind more stably to the protein/peptide complex but also have some (weaker) affinity for the protein alone.

We show that the interaction between 14-3-3 and ERα can be stabilized by increasing the apparent K_d of this interaction over 30-fold. The fragment hits resulting from these screens are chemical starting points for optimization towards potent and selective PPI stabilizers. Further study of why analogs of 917884 which made additional contacts with 14-3-3 had decreased ERα cooperativity is necessary. The ablated cooperativity may be due to strain on the phenyl-chloro binding pose introduced by coordinating an additional protein surface, in which case a larger molecule may accommodate ideal contacts with the phospho-peptide and other 14-3-3 surface residues.

One unexplored benefit of our approach is that a single screening campaign could yield both inhibitors and stabilizers. One major concern in targeting 14-3-3 PPIs is the client selectivity, related to the wide range of 14-3-3 binding partners. It is highly valuable to explore different approaches in circumventing the non-specific targeting of the primary binding groove of 14-3-3. The approach presented here was found to yield small molecules showing a client selectivity for ER□-like (called "mode III") peptides over mode I or II in a straightforward selection procedure. This approach might be expanded by implementing a screening step over all known 14-3-3 interaction motifs and picking up hits that show high selectivity, directly from the disulfide library. Alternatively, one could make use of the growing evidence for secondary interaction sites between 14-3-3 and its client proteins. These secondary sites occur outside of the phospho-accepting groove based on structural information of binary complexes of 14-3-3 bound to client protein domains. For instance, to Ottmann lab previously discovered secondary-site fragment binders on outer surfaces of the 14-3-3 protein [27]. This opens up new possibilities in another direction for selectivity in modulation of 14-3-3 PPIs.

Even though the main aim has firstly been to develop and validate a platform for the identification of PPI stabilizers, the next challenge is the development of potent, selective non-covalent PPI stabilizers based on the main hits from these screening campaigns. As 14-3-3σ contains a native cysteine, the hits resulting from this screen might be further developed into covalent stabilizers, with the added benefit of an isoform selective compound. The hits from the other cysteine positions, however, result from 14-3-3 constructs with introduced cysteines. To optimize these fragments, we have identified fragments that bind to nearby sites. Though these fragments have no stabilizing activity, they are within linking distance of the stabilizing disulfide

fragments; linking these two fragment classes could provide increased affinity while maintaining stabilization activity.

The results presented here demonstrate the utility of a site-directed tethering approach to finding PPI modulators. By leveraging structural data available from binary complexes, a rationally designed screen allowed the identification of fragment hits with high cooperativity for the 14-3-3/ERα. Characterization of these hits revealed a convergent binding pose, well-defined SAR, and selectivity for the ERα-like phospho-peptides. Of the hundreds of 14-3-3 partner proteins, fewer than 10 have ERα-like C-terminal binding poses, and we expect our compounds would selectively stabilize <2% of 14-3-3 interactions. Though our effort began from a defined natural product binding pocket, similar screening campaigns could be conceived against any PPI interface with a structural description. Applications then extend from finding small molecule stabilizers of a PPI to identifying fragments for use in bi-functional targeting molecules such a Proteolysis Targeting Chimeras (PROTACs). PROTACs have strict binding site and orientation requirements to enable target degradation. Tethering approaches allow control of these variables, and specifically recruiting cellular degradation machinery to a PPI complex, rather than to the individual component of the complex, could be a powerful chemical biology tool.

Methods

Protein expression and purification

The 14-3-3 σ isoform with a C-terminal His-tag was expressed in NiCo21 (DE3) competent E.coli (New England Biolabs) from a pPROEX HTb expression vector. Site-directed mutagenesis to obtain C38N-N42C and C38N-S45C was performed using QuickChange

Lightening site-directed mutagenesis kit (Agilent Technologies) following manufacturer's instructions. Correct constructs were confirmed by DNA sequencing. The wild-type and Cys-mutant proteins were first purified by Ni-affinity chromatography (HisTrap HP column, GE), followed by His-tag cleavage by TEV protease during dialysis overnight. The flow-through of a second HisTrap column was subjected to final purification by size-exclusion chromatography (Superdex75). The protein was concentrated to ~60 mg/ml, analyzed for purity by SDS-PAGE and Q-Tof LC/MS and aliquots flash-frozen for storage at -80°C.

Disulfide Tethering

The primary screening was performed on a 384-well plate format. A custom library of 1600 disulfide-containing fragments of the UCSF Small Molecule Discovery Center (SMDC), synthesized as previously reported, was available as 50 mM stock solutions in DMSO [28,29]. For screening, 14-3-3 wild-type and Cys-mutants were diluted to 100 nM in buffer (10 mM Tris, 100 μM β-mercaptoethanol (BME), pH 8.0) and plated in 384-well plates (15 μl/well). Of the disulfides stocks, 30 nl was pinned into the protein samples using a Biomek FX (Beckman). The reaction mixtures were incubated for 3 hours at RT before subjected to LC/MS (I-class Acquity UPLC / Xevo G2-XS Quadrupole Time of Flight mass spectrometer, Waters) data collection and automated processing using a custom pipeline, as previously described in detail [24]. Dose-Response titrations of hit compounds were collected at identical LC/MS conditions and 14-3-3 concentrations. The screening concentration of 100 μM βME was increased to 1 mM to raise the stringency of the labeling reaction after complete labeling was observed at <1μM concentrations of top hits.

X-Ray Crystallography

14-3-3 crystals were grown as previously described, with slight modification by E. Sijbesma [27]. Crystals containing ERα phospho-peptide were soaked with 1 mM disulfide hits form 50 mM stocks in DMSO. Diffraction data were collected at PETRA III, beamline P11, DESY, Hamburg, Germany. Resulting datasets were indexed and integrated with XDS [30], scaled with SCALA [31], and models were determined via molecular replacement and refinement in Phenix [32]. Appendix Table 3 reports the data collection, processing, refinement and validation statistics.

Fluorescence Anisotropy

Fluorescence polarization experiments were performed by E. Sijbesma as previously reported in work from C. Ottmann's lab [33].

REFERENCES

[1] Morrison, D. K. The 14-3-3 proteins: integrators of diverse signaling cues that impact cell fate and cancer development. Trends Cell Biol. 19, 16–23 (2009).

[2] Aitken, A. 14-3-3 proteins: A historic overview. Semin. Cancer Biol. 16, 162–172 (2006).

[3] Yang, X. et al. Structural basis for protein-protein interactions in the 14-3-3 protein family. Proc. Natl. Acad. Sci. U. S. A. 103, 17237–17242 (2006).

[4] Obsil, T. & Obsilova, V. Structural basis of 14-3-3 protein functions. Semin. Cell Dev. Biol. 22, 663–672 (2011).

[5] Tzivion, G., Luo, Z. & Avruch, J. A dimeric 14-3-3 protein is an essential cofactor for Raf kinase activity. Nature 394, 88 (1998).

[6] Sluchanko, N. N. et al. Structural basis for the interaction of a human small heat shock protein with the 14-3-3 universal signaling regulator. Struct. Lond. Engl. 1993 25, 305–316 (2017).

[7] Schumacher, B., Mondry, J., Thiel, P., Weyand, M. & Ottmann, C. Structure of the p53 C-terminus bound to 14-3-3: Implications for stabilization of the p53 tetramer. FEBS Lett. 584, 1443–1448 (2010).

[8] Zhao, J., Meyerkord, C. L., Du, Y., Khuri, F. R. & Fu, H. 14-3-3 proteins as potential therapeutic targets. Semin. Cell Dev. Biol. 22, 705–712 (2011).

[9] Aghazadeh, Y. & Papadopoulos, V. The role of the 14-3-3 protein family in health, disease, and drug development. Drug Discov. Today 21, 278–287 (2016).

[10] Berg, T. Modulation of protein-protein interactions with small organic molecules. Angew. Chem. Int. Ed Engl. 42, 2462–81 (2003).

[11] Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. Nat. Rev. Drug Discov. 3, 301–17 (2004).

[12] Yin, H. & Hamilton, A. D. Strategies for Targeting Protein–Protein Interactions With Synthetic Agents. Angew. Chem. Int. Ed. 44, 4130–4163 (2005).
[13] Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450, 1001–9 (2007).
[14] Andrei, S. A. et al. Stabilization of protein-protein interactions in drug discovery. Expert Opin. Drug Discov. 12, 925–940 (2017).
[15] Stevers, L. M. et al. Modulators of 14-3-3 Protein–Protein Interactions. J. Med. Chem. (2017). doi:10.1021/acs.jmedchem.7b00574
[16] Ottmann, C. et al. Structure of a 14-3-3 Coordinated Hexamer of the Plant Plasma Membrane H+-ATPase by Combining X-Ray Crystallography and Electron Cryomicroscopy. Mol. Cell 25, 427–440 (2007).
[17] De Vries-van Leeuwen, I. J. et al. Interaction of 14-3-3 proteins with the estrogen receptor alpha F domain provides a drug target interface. Proc. Natl. Acad. Sci. U. S. A. 110, 8894–9 (2013).
[18] Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. Nat. Chem. 1, 187–192 (2009).
[19] Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. Nat. Rev. Drug Discov. 15, 605–619 (2016).
[20] Erlanson, D. et al. Site-directed ligand discovery. Proc. Natl. Acad. Sci. U. S. A. 97, 9367–9372 (2000).
[21] Erlanson, D. a, Wells, J. a & Braisted, A. C. Tethering: fragment-based drug discovery. Annu. Rev. Biophys. Biomol. Struct. 33, 199–223 (2004).
[22] Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. a & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature 503, 548–51 (2013).
[23] Hyde, J., Braisted, A. C., Randal, M. & Arkin, M. R. Discovery and characterization of cooperative ligand binding in the adaptive region of interleukin-2. Biochemistry (Mosc.) 42, 6475–6483 (2003).
[24] Hallenbeck, K. K. et al. A Liquid Chromatography/Mass Spectrometry Method for Screening Disulfide Tethering Fragments. SLAS Discov. Adv. Life Sci. R D 2472555217732072 (2017). doi:10.1177/2472555217732072
[25] Dewal, M. B. & Firestine, S. M. Non-peptidic α-helical mimetics as protein-protein interaction inhibitors. Curr. Med. Chem. 18, 2420–2428 (2011).
[26] Arkin, M. R., Tang, Y. & Wells, J. A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. Chem. Biol. 21, 1102–1114 (2014).
[27] Sijbesma, E. et al. Identification of Two Secondary Ligand Binding Sites in 14-3-3 Proteins Using Fragment Screening. Biochemistry (Mosc.) 56, 3972–3982 (2017).
[28] Burlingame, M. A., Tom, C. T. M. B. & Renslo, A. R. Simple One-Pot Synthesis of Disulfide Fragments for Use in Disulfide-Exchange Screening. ACS Comb. Sci. 13, 205–208 (2011).
[29] Turner, D. M., Tom, C. T. M. B. & Renslo, A. R. Simple Plate-Based, Parallel Synthesis of Disulfide Fragments using the CuAAC Click Reaction. ACS Comb. Sci. 16, 661–664 (2014).

[30] Kabsch W. (2010) XDS. Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 125–13210.1107/S0907444909047337.
[31] Evans P. (2006) Scaling and assessment of data quality. Acta Crystallogr., Sect. D: Biol. Crystallogr. 62, 72–8210.1107/S0907444905036693.
[32] Adams P. D.; Afonine P. V.; Bunkóczi G.; Chen V. B.; Davis I. W.; Echols N.; Headd J. J.; Hung L.-W.; Kapral G. J.; Grosse-Kunstleve R. W.; McCoy A. J.; Moriarty N. W.; Oeffner R.; Read R. J.; Richardson D. C.; Richardson J. S.; Terwilliger T. C.; Zwart P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 213–22110.1107/S0907444909052925.
[33] Bier D, Mittal S, Bravo-Rodriguez K, et al. The Molecular Tweezer CLR01 Stabilizes a Disordered Protein–Protein Interface. *Journal of the American Chemical Society*. 2017;139(45):16256-16263. doi:10.1021/jacs.7b07939.

Chapter 3

DIVERGENT CONFORMATIONAL DYNAMICS CONTROLS ALLOSTERIC LIGAND ACCESSIBILITY ACROSS EVOLUTIONARILY RELATED I-DOMAIN-CONTAINING INTEGRINS

Contributing Authors:
RAHEL A. WOLDEYES, SAMUEL J. PFAFF, GREGORY LEE, SAUL V. CORTEZ, MARK J. KELLY, KATERINA AKASSOGLOU, JAMES S. FRASER, MICHELLE R. ARKIN

INTRODUCTION

Allostery, the process by which binding at one site can influence the conformational ensemble at distant sites, is an important property in the regulation of proteins [1]. Proteins within a family are generally thought to share conserved allosteric mechanisms. For example, binding of ligands to GPCRs is accompanied by an 11 Å displacement of a helix in the cytoplasmic region about 40 Å away from the orthosteric site [2,3]. Similarly, many kinases use protein-protein interactions to stabilize a catalytically active state [4]. The evolutionary origins of allostery may lie in the adaptability of the conformational ensemble [5], suggesting that the ease of reaching an allosteric state may vary between family members.

The β2 family of integrins are heterodimers containing the CD18 β chain and one of four α chains (CD11a/Lfa-1, CD11b/Mac-1, CD11c, or CD11d). β2 integrins sit on the surface of immune cells in an inactive conformation; activation can occur via intracellular signals that change the integrin conformation or via binding of extracellular proteins such as ICAM-1 [7] or Complement fragment iC3dg [8]. These proteins bind to a Metal Ion-Dependent Adhesion Site (MIDAS) located on the inserted (I)-domain of the α chain (Fig 1). X-ray crystallography and Nuclear Magnetic Resonance (NMR) studies of isolated I-domains show they constitutively

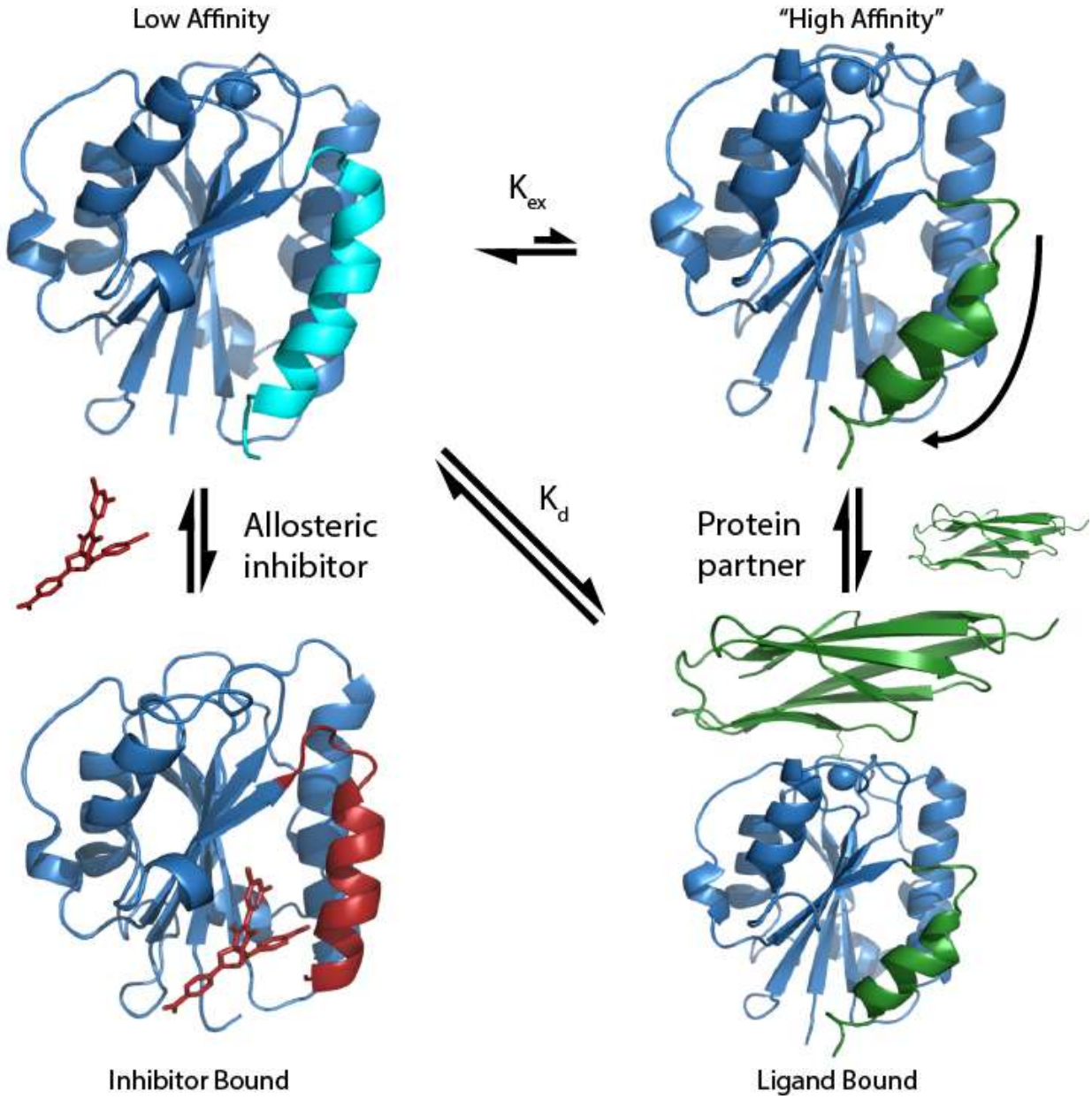


Figure 1: The classical thermodynamic cycle of Lfa-1 and Mac-1 I-domains. I-domains can adopt low or high affinity states as marked by the position of the α7 helix (colored cyan or green). Ligand binding to the MIDAS stabilizes the high affinity state. Allosteric ligand binding to Lfa-1 stabilizes the helix in a low affinity-like state (red).

adopt the un-bound, 'low affinity' conformation. Binding to the MIDAS site leads to a ~10A displacement of the α7 helix; this 'high affinity' conformation can be mimicked with disulfide locks, point mutations, and truncations of the α7 helix [7,9]. For Lfa-1, diverse allosteric, small-molecule inhibitors have been found to bind to a pocket occupied by the α7 helix in the high

affinity state (the 'α7 pocket'), thereby stabilizing an inactive-like state [7,10-11]. While the I-domain structures are very similar, and the proposed conformational changes leading to integrin activation are thought to be conserved, it is not clear whether the other β2 integrins have a small-molecule binding site analogous to the α7 pocket seen in Lfa-1. For example, Mac-1 is a target for autoimmune diseases such as multiple sclerosis [12,13] but there are no structurally characterized molecules that bind to the Mac-1 α7 pocket. Indeed, although statin drugs interact with the allosteric pocket of Lfa-1 [14,15], an X-ray structure of Mac-1 bound to simvastatin shows the statin's carboxylate moiety coordinating the MIDAS metal ion. This divergent mechanism is noteworthy, given that the Lfa-1 and Mac-1 I-domains adopt nearly identical overall folds and have ~30% sequence identity in the putative α7 pocket.

Here, we have examined the link between conformational heterogeneity, evolutionary conservation, and "ligandability" of the α7 allosteric site for LFA-1 and Mac-1 I-domains. Using site-directed, disulfide-trapping of molecular fragments, we find that Lfa-1 captures more ligands than the closely related Mac-1. The α7 helix of Lfa-1 is also more conformationally heterogeneous than the equivalent region in Mac-1, suggesting that increased dynamics could explain the differences in ligandability. We also characterized how allosteric ligand binding rearranges the MIDAS of Lfa-1 I-domain and use these data to propose mutations to make Mac-1 more closely resemble Lfa-1. Despite the more restrictive conformational landscape of Mac-1, a small number of molecules can be identified that bind to the allosteric site and alter the Mac-1 MIDAS. Collectively, these results paint a picture of broadly conserved conformational minima, but idiosyncratic energy landscapes connecting them. LFA-1 samples a broad ensemble that

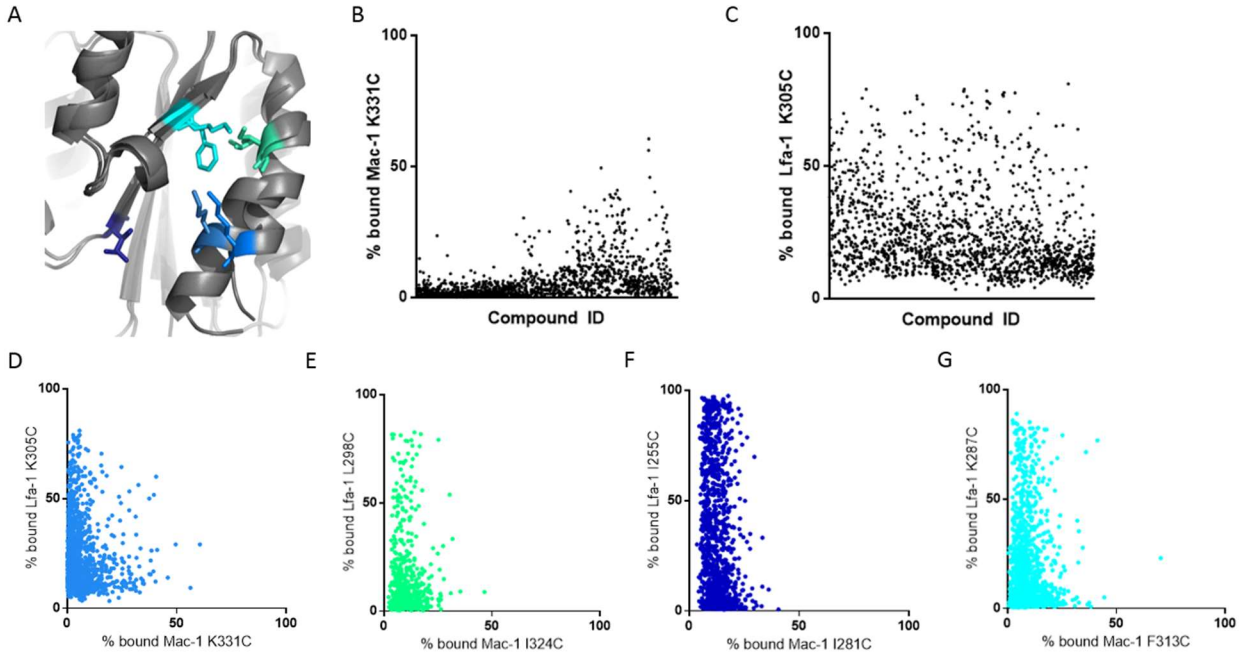


Figure 2: Disulfide trapping highlights greater ability of I-domain of Lfa-1 to capture small-molecule fragments than Mac-1. A) Homologous sites of mutation to cysteine for tethering, color-coded as in D-G. B) Disulfide-containing molecules vs % binding to Mac-1 K331C, as determined by mass spectrometry (MS). C) Disulfide-containing molecules vs % binding to Lfa-1 K305C, as determined by MS. D-G) Correlation plot reflecting the % labeling by each molecule to analogous positions on Mac-1 (x-axis) vs Lfa-1 (y-axis). D) Mac-1 K331C vs Lfa-1 K305C; E) Mac-1 I324C vs Lfa-1 L298C; F) Mac-1 I281C vs Lfa-1 L255C; G) Mac-1 F313C vs Lfa-1 K287C.

enables promiscuous allosteric ligand binding whereas Mac-1 transitions between two sharply defined minima and has more stringent ligand binding criteria.

Results

Disulfide-trapping indicates differential ligandablilty between Mac-1 and Lfa-1 I-domains

As part of our efforts to identify allosteric inhibitors of Mac-1 and Lfa-1, we developed a series of disulfide-trapping screens. 'Fragments' are molecules approximately half the size of a drug (i.e., < 250 Da); disulfide tethering is a reversible, covalent labeling reaction using a custom library of disulfide-capped fragments [14]. Under reducing conditions, fragments undergo

disulfide-thiolate exchange with solvent-exposed cysteine residues positioned near a potential binding site. At thermodynamic equilibrium, fragments that bind in the pocket with the proper orientation stabilize the disulfide bond. The fragment/protein adduct and residual unbound protein are then detected by liquid chromatography/mass spectrometry LC/MS (Fig 2B-C) [16]. Disulfide trapping has been highly successful at identifying chemical probes and drug leads for allosteric sites and protein-protein interactions [17-20].

Because neither Lfa-1 or Mac-1 I-domains contain native cysteine residues, we introduced matching pairs of cysteines to four sites around the putative allosteric pocket. We mutated residues on both the α7 helix itself and the opposing β-strand to remove bias based on individual cysteine orientation (Fig 2A). Next, we screened each cysteine mutation against a custom library of 1600 disulfide fragments, using identical buffer, chromatography, and mass spectrometry conditions. For each pair of mutations, we found that Lfa-1 captured many more disulfide fragments than Mac-1 (Fig 2D-G). The differences between the two proteins could also be seen in the labeling efficiency of the top hits; for Lfa-1 mutants, the top 10% of fragments labeled residues K305C, L298C, I255C, and K287C by at least 47.0, 40.3, 84.8, and 51.5 %, respectively, whereas the top 10% of fragments for the analogous Mac-1 residues K331C, I324C, I281C and F313C labeled by at least 13.1, 16.5, 18.1, and 16.0 %, respectively. Furthermore, the most labeled Mac-1 mutation was K331C, located on the α7 helix; in contrast, the highest labeling Lfa-1 mutation was I255C, located deep in the α7 pocket (Fig 2F). The greater accessibility of the buried cysteine residue in Lfa-1 further suggested that the pocket was more open than in Mac-1. Thus, the high hit-rate, high %-binding of tethering hits, and the location of

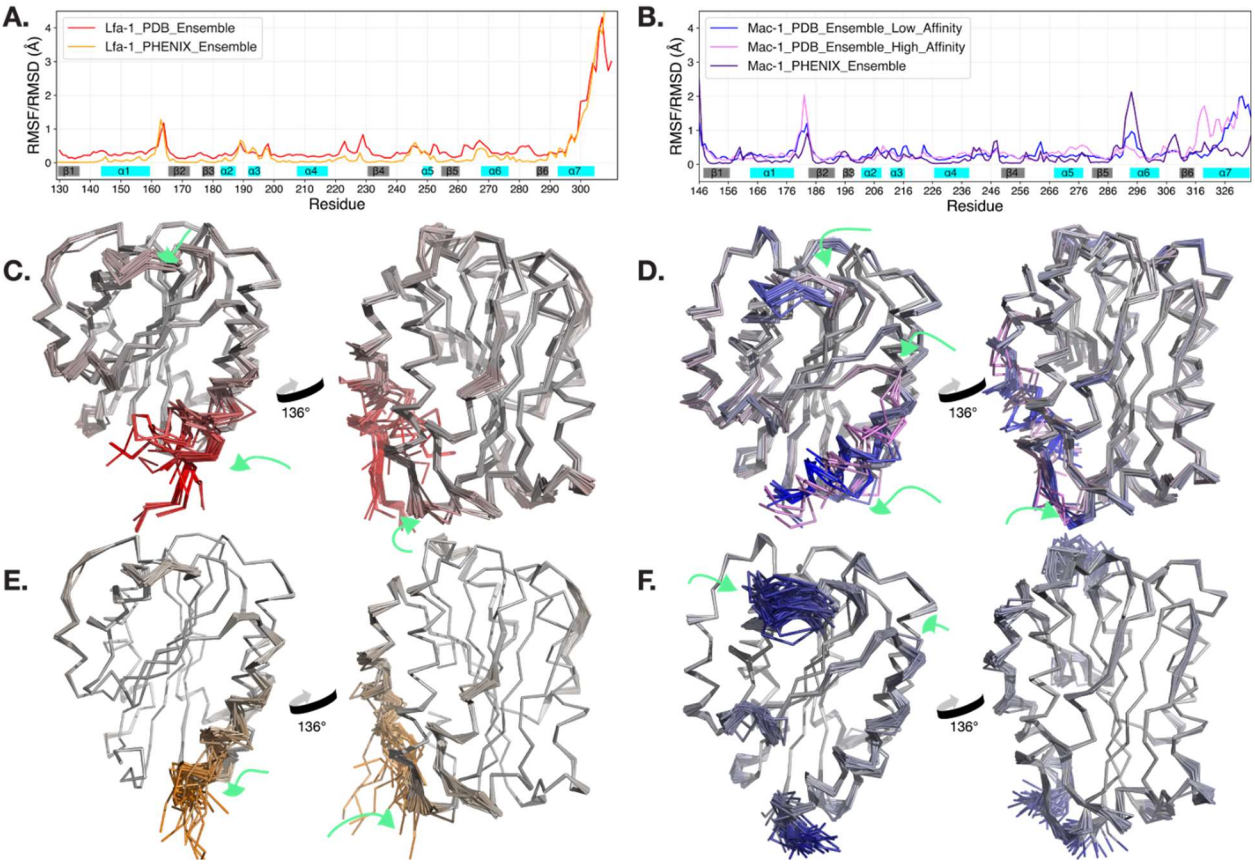


Figure 3. Ensemble-based analyses of I-domain crystal structures. A) Overlaid plots of amino acid sequence vs c-α root mean square difference (RMSD) for all Lfa-1structures in the low affinity state deposited in the protein data bank (PDB; red) and sequence vs root mean square fluctuation (RMSF) for an ensemble refinement of a 1.8 Å x-ray dataset collected at room temperature (orange). B) Overlaid plots of amino acid sequence vs c-α RMSD for all PDB Mac-1 structures in the low affinity state (blue) and high affinity state (pink) compared to RMSF for an ensemble refinement of a 1.7 Å x-ray dataset collected at room temperature (purple). C) Structural overlay of the PDB ensemble for Lfa-1, colored by RMSD values in a gradient from 0 Å (gray) to 4.3 Å (red). D) Structural overlay of Mac-1 PDB ensemble, colored by RMSD from 0 Å (grey) to 2.0 Å (blue, low affinity state) or 2.0 Å (pink, high affinity state). E) The final time-averaged ensemble for room-temperature x-ray structure of Lfa-1, colored by RMSF from 0 Å (grey) to 4.8 Å (orange). F) The final time-averaged ensemble for room-temperature x-ray structure of Mac-1, colored by RMSF from 0 Å (grey) to 2.5 Å (purple). Green arrows show areas with high conformational heterogeneity. Lfa-1 shows greatest mobility in the α7 helix and the α1-β2 loop, whereas Mac-1 shows greatest mobility in the β5-α6 loop and lower variability in the α7 helix.

the highest-labeling residues implied that, surprisingly, the Mac-1 allosteric site was much less ligandable than the analogous site on Lfa-1.

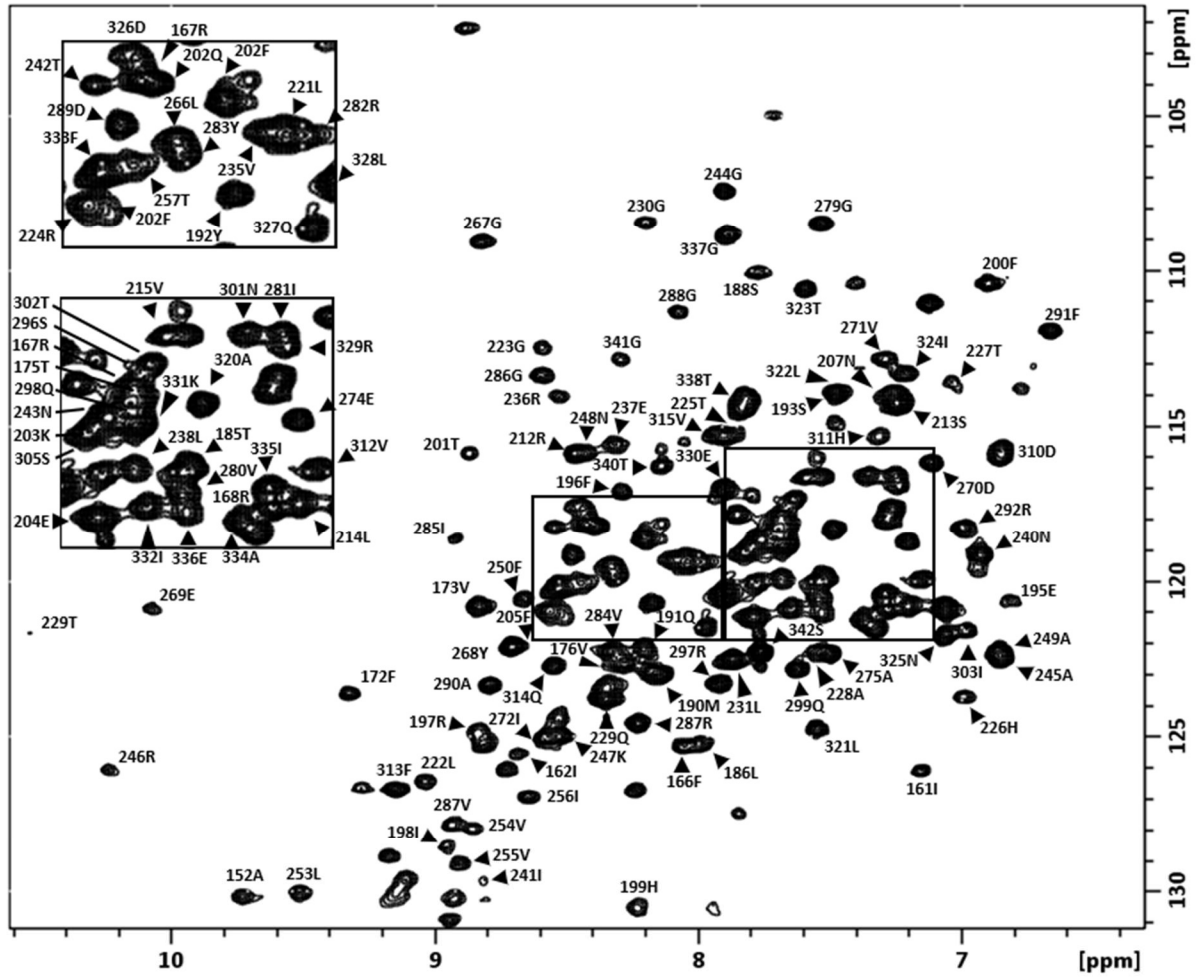
X-ray crystallography suggests differences in flexibility surrounding α7 pocket

To explore the structural basis for the differential ligandability of the Lfa-1 and Mac-1 allosteric pockets, we analyzed all X-ray and NMR structures deposited in the protein data bank

(PDB), including 25 for Lfa-1 (excluding disulfide-bond stabilized mutants) and 13 for Mac-1. Alignment of the structures demonstrated that conformational heterogeneity was greatest in the α1- β2 loop and the α7 helix of Lfa-1, with a root mean square difference (RMSD) of 0.30 - 1.18 Å for c-α backbone atoms for the α1- β2 loop and 0.29 - 2.96 Å for the α7 helix of Lfa-1 (Fig 3A,C). Alignment of Mac-1 I-domain structures also showed heterogeneity in the α1- β2 loop of Mac-1 (0.37 -1.21 Å closed, 0.50 -2.04 Å open; Fig 3B,D). However, the two integrins differed in the α7 helix. Mac-1 could be segregated into two different classes, each of which had limited heterogeneity in the α7 helix compared to Lfa-1 (RMSD < 2 Å); in contrast, and in agreement with multistate dynamics proposed by Kukic, et al, the α7 helix of Lfa-1 could not be neatly clustered into two populations unless a ligand was bound to the MIDAS [21].

To test the idea that the α7 helix of Lfa-1 was variable even within a single crystal, we collected high-resolution, room-temperature crystal structures of the Mac-1 and Lfa-1 I-domains (Appendix Table 4). These structures aligned well with previously reported structures that were collected on cryo-cooled crystals, but collection at room temperature increased our ability to visualize alternate conformations. For Lfa-1, the α7 helix had no density after residue 308, whereas backbone density extended to residue 322 in Mac-1, five more residues than for Lfa-1 and just four residues short of the construct's C-terminus (Appendix Fig 9). Time-averaged ensemble refinement (PHENIX) of Lfa-1 revealed that the most dynamic areas were the α1- β2 loop, with root mean square fluctuations (RMSF) up to 1.28 Å, and the α7 helix (RMSF = 0.18 - 3.42 Å) (Fig 3A,E). The Mac-1 time-averaged ensemble identified heterogeneity in the α1- β2 loop (RMSF = 0.18 -0.92 Å) and low heterogeneity for the α7 helix (Fig 3B,F). Thus, for both

A.



B.

GHMQED**SDIAF**LI**DGS**GS**II**P**H****DFRR**MKE**FV**S**TV**MEQLKKSK**T**
LFSL**MQYS**E**EFRIHFTFKEFQNN**P**N**P**RSLVK**P**ITQ**LLGRTHTA**
TGIRK**VVREL**F**NITNGARKNAFK**I**LVVIT**DGEKFGD**P**LGYEDV**
IP**EADREGVIRYVIGVGDAFR**SEK**SRQE**L**NTIAS**K**P**P**R**DHVFQ**
VNNFEALKTIQNQLREKIFAIEGTQTGS

Figure 4. Backbone resonance assignment of Mac-1 I-domain. (A) Annotated 15N-TROSY spectrum on 1.5 mM U-15N Mac-1 I-domain. (B) Expressed Mac-1 sequence, with assigned resonance in bold, unassigned in plain text, and prolines in red.

Lfa-1 and Mac-1, the heterogeneity observed between PDB-deposited structures was also observed within the single, room-temperature structures. From these analyses, we concluded that Lfa-1 sampled a larger range of near-isoenergetic conformations than did Mac-1.

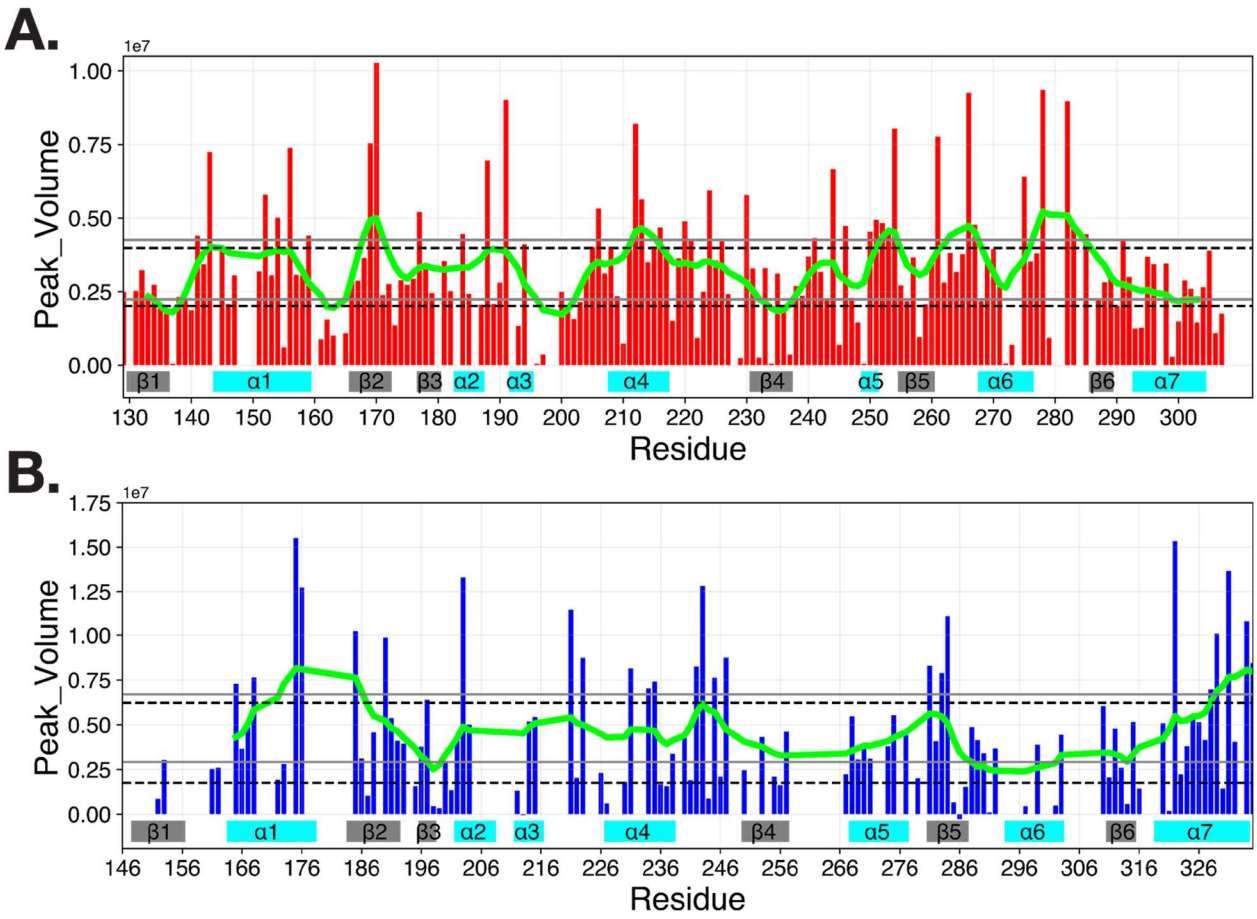


Figure 5. Analysis of 15N-1H HSQC NMR peak intensities for I-domains. A) Peak intensities for assigned residues in Lfa-1 (red bars; 159 residues) with the rolling average (green; with a rolling window of 5 residues). Residues along α7 helix are generally below +/- 1 standard deviation from the mean(grey solid lines) and +/- the iqr value from the median(black dotted lines). The rolling average is depressed to the bottom 25 percentile values for residues on the α7 helix. B) A similar analysis for Mac-1 peak intensities (blue bars; 112 residues) indicates that α7 helix is more rigid in Mac-1. The α7 helix residues have sharpened peaks with peak intensity values that are mostly greater than +/- 1 standard deviation from the mean

NMR confirms differential α7 flexibility between Lfa-1 and Mac-1

To test whether the conformational heterogeneity observed by our room-temperature crystal structures were present in solution, we turned to NMR. We performed 15N-1H HSQC experiments on both I-domains and analyzed peak broadening using peak intensities. When combined with knowledge of secondary structure from crystallography, smaller peak intensities suggest regions of protein structure that are exchanging between multiple conformations on the μs-to-ms timescale.

For Lfa-1, we observed significant peak broadening around α1- β2 loop and the α7 helix, consistent with x-ray crystallography analysis (Fig 4A). This observation confirms previous findings by Legge et al., who have shown the flexibility of the helix using heteronuclear NOE and deuterium exchange experiments, and aligns well with the reduced restraints on the α7 helix reported in the Lfa-1 NMR structure [22]. Peak broadening/lower peak intensity is also observed around the α3-α4 loop, which contains one of the five residues that make up the MIDAS.

To evaluate the dynamics of Mac-1 I-domain by NMR, we assigned the backbone resonances, achieving 72% sequence coverage, including the entire α7 helix and allosteric pocket. Peak-broadening analysis showed that, in contrast to the Lfa-1 α7-helix resonances, the 15N-1H HSQC NMR peaks corresponding to the Mac-1 α7 helix remained sharp to the C-terminus (Fig 4B). To expand upon this observation, we collected Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion experiments on the Mac-1 I-domain. $^{1}H/^{15}N$ CPMG experiments interrogate the relaxation of H-N bonds on the μs-ms timescale, precluding fast protein motions such as bind vibration and rotamer flips, but observing slow conformation changes such as those undergone by the I-domain α7 helix. Data was collected at 800 Mhz, and R_{ex} was calculated in MatLab by fitting the relative intensities of each backbone resonance. Surprisingly, residues on the Mac-1 α7 helix exhibit strong μs-ms movement, suggesting the Mac-1 α7 helix may be sampling the 'high-affinity' conformation (Fig 6).

From these data, we concluded that conformational transitions of the α7 helix of Mac-1 I-domain occur on a timescale slower than 15N/1H signal relaxation. Collectively, the NMR and X-ray results indicated that the differences in ligandability were paralleled by the differences in conformational heterogeneity and dynamics between Lfa-1 and Mac-1.

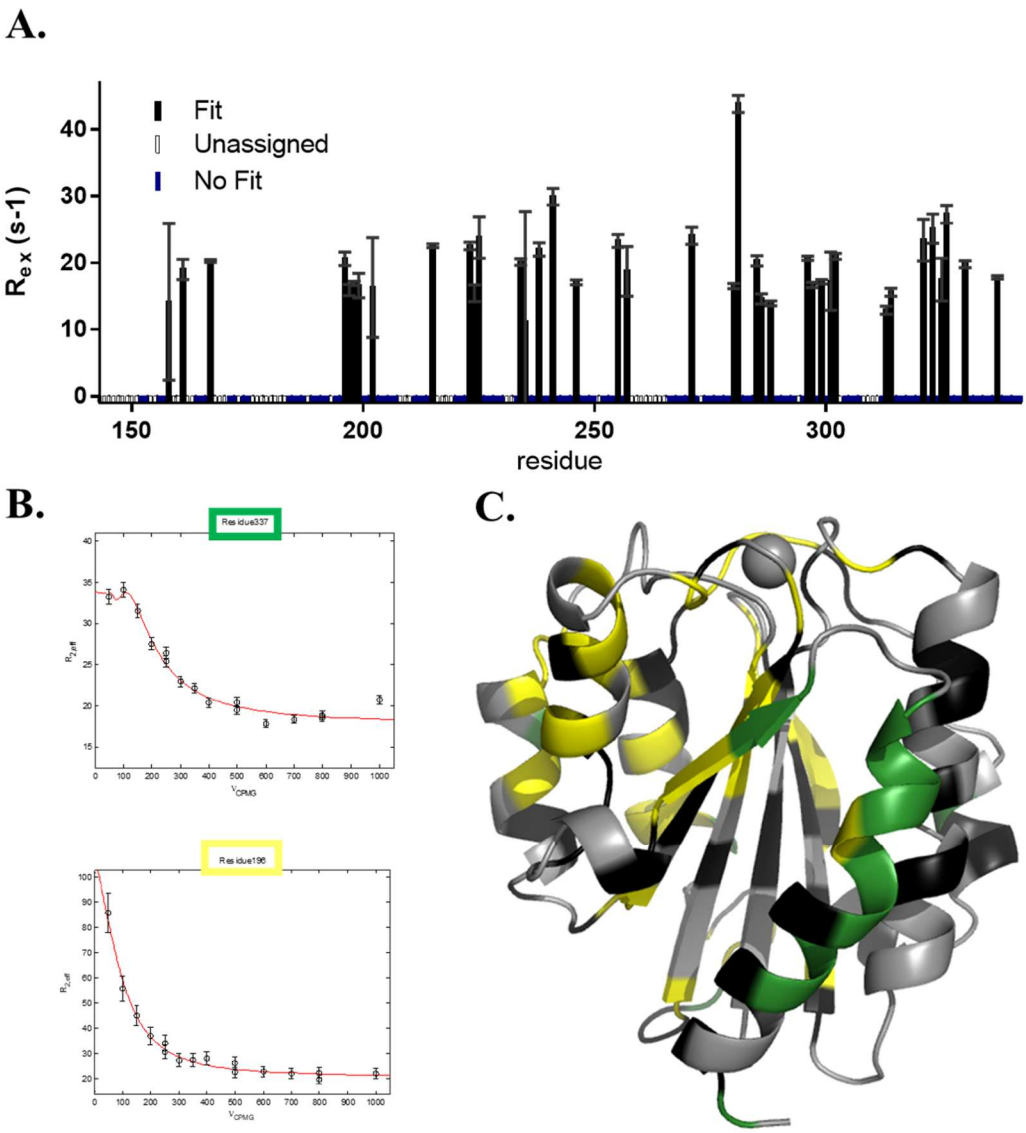


Figure 6. Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion of Mac-1 I-domain. (A) Relaxation coefficient (R_{ex}, y-axis) plotted for each residue in the Mac-1 I-domain (x-axis). (B) Representative R_{2eff} vs. ν_{CPMG} curves for residues demonstrating µs-ms relaxation (top, green) and faster than µs relaxation (bottom, yellow). (C) Overlay of fast (yellow), intermediate (green), no (black), or un-fit (gray) residues onto the crystal structure of the Mac-1 I-domain.

We hypothesized the ligandability of the Mac-1 α7 pocket could be modulated with a mutation designed to decrease the stability of the α7 helix. To identify candidate sites for mutagenesis, we aligned the sequence of the human I-domain homologs and inspected the crystal

structure of Mac-1 and Lfa-1 sites where the amino-acid identity Mac-1 and Lfa-1 diverged from the broader family. These analyses identified Mac-1 residue Q325 (analogous to Lfa-1 F299), which sat on the solvent-exposed side of the α7 helix and packed against the nearby α1 helix (Fig 7A). Reasoning that the solvent-exposed hydrophobic phenylalanine might destabilize the α7 helix, we prepared the Mac-1 mutant Q325F.

Destabilizing mutation in Mac-1 does not alter α7 pocket ligandability

Mac-1 Q325F expressed and purified similarly to wild type. To test whether the mutation had a global effect on stability, we performed thermal denaturation melts and monitored the effect by circular dichroism spectroscopy. These assays revealed that Q325F was thermally destabilized by 3.4°C relative to wild type (Fig 7B). Using surface plasmon resonance, we confirmed the mutant protein had similar affinity as wild type Mac-1 for the protein ligand C3dg (Kd = 4.0 μM and 2.1 μM, respectively; Appendix Fig 10). To test whether Q325F increased conformational heterogeneity in the Mac-1 α7 helix, we performed peak-broadening analysis on the 1H-15N-HSQC spectrum of 15N-labeled K331C/Q325F, (Fig 7D). Indeed, the peak intensities for K331C/Q325F for the α7 residues N-terminal to the mutation were suppressed; the moving average only rose above the IQR at the very C-terminus. Peak intensities were lower for the mutant compared to wild type Mac-1, where the α7 helix showed the largest peak intensities in the HSQC spectrum (Fig 4B); however, peaks were not as broadened as those observed in the homologous region of Lfa-1 (Fig 4A).

Next, we tested whether this variant increased the ligandability of the α7 pocket. Surprisingly, we found no significant impact on the ligandability of the Mac-1 pocket compared to Mac-1 K331C (Fig 7C). The failure of Q325F to promote the open conformations of the

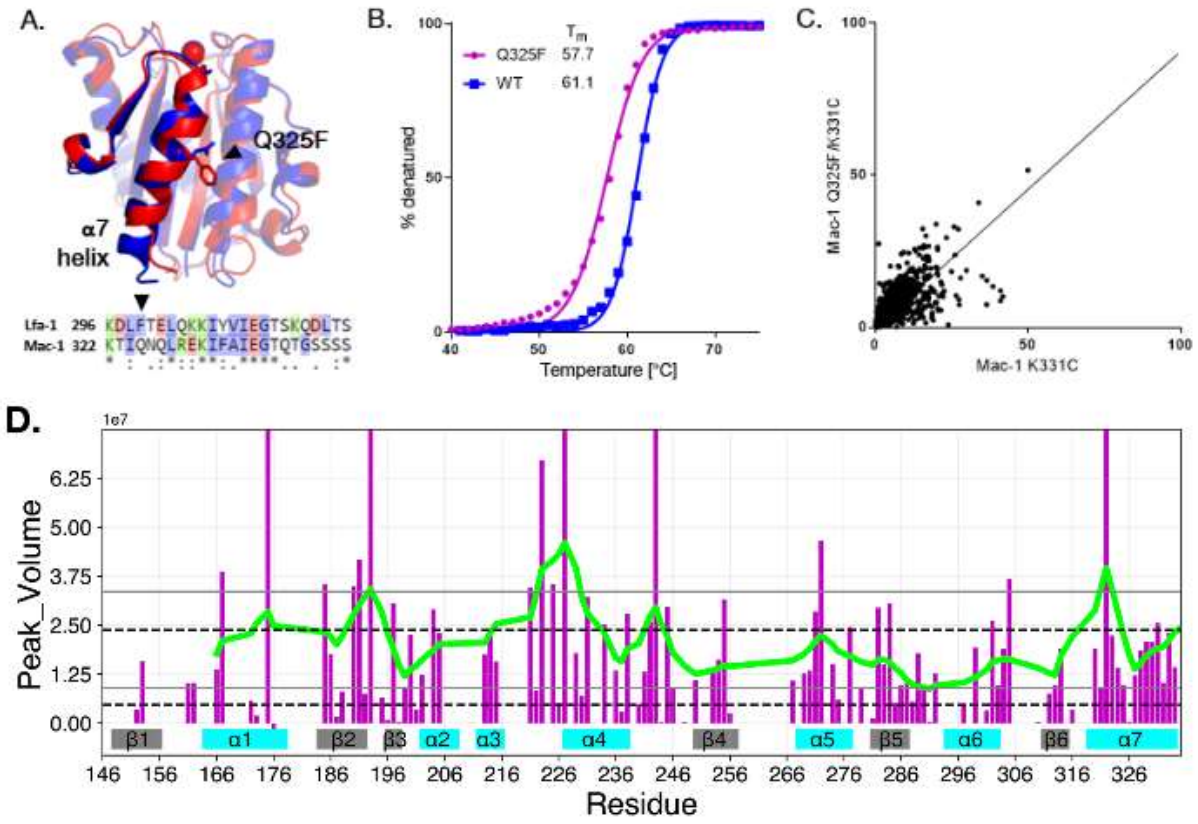


Figure 7: Characterization of Mac-1 Q325F. A) Selection of Q325F, based on a comparison of the Mac-1 (blue) and Lfa-1 (red) α7 helix sequences. B) Melting curves for wild type Mac-1 (blue) and Mac-1 Q325F (pink). C) Disulfide tethering screen comparing K331C (x-axis) to Mac-1 K331C/Q325F (y-axis). D) 15N-HSQC peak intensities of Mac-1 K331C/Q325F mutant (purple bars) and moving average (green line). Despite lower protein stability and increased α7 dynamics, ligandability is not increased.

allosteric pocket despite the 3.4 °C shift in melting temperature and the perturbation of the α7 helix via NMR relaxation suggests further study is necessary to understand the conformational trajectory which results in the accessibility of the I-domain allosteric pocket.

Discussion

Protein ligandability is a function of both the chemistry at the binding site and the accessibility of the site [23]. While Mac-1 and Lfa-1 I-domains both share a hydrophobic surface

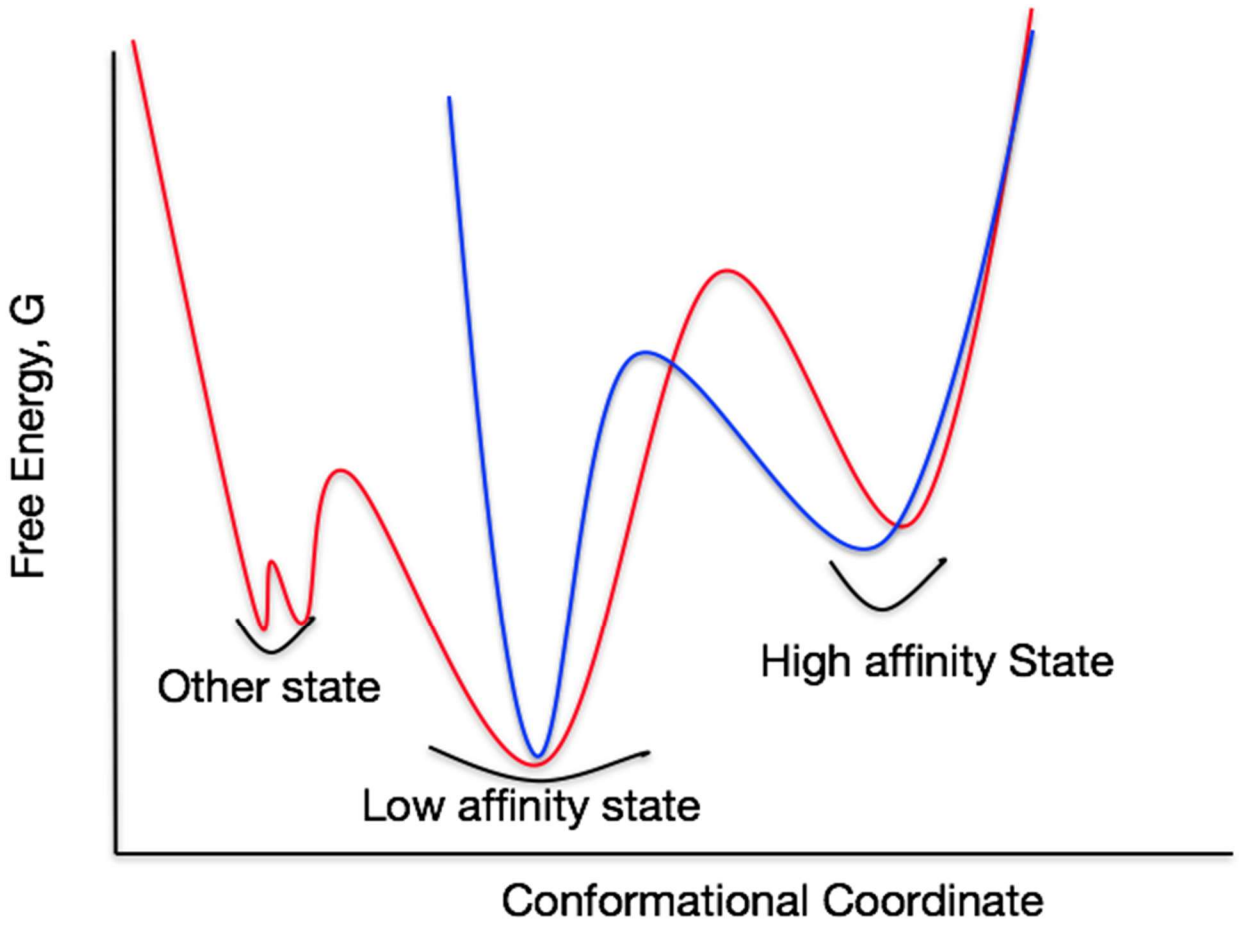


Figure 8. Proposed energy landscapes of Lfa-1(Red) and Mac-1(blue) I-domains. Lfa-1 samples multiple conformations in the ligand unbound state while Mac-1 mostly samples around the low affinity state and adopts the high affinity conformation when the energy penalty is paid by introducing mutations or truncations.

underneath the dynamic α7 helix, and similar allosteric mechanisms of activation have been proposed for both proteins, small-molecule allosteric ligands have only previously been discovered for Lfa-1. Our studies demonstrate that this difference in known ligands is no accident - the two family members also differ in their ability to be covalently labeled by disulfide-containing fragments when homologous residues are mutated to cysteine (Fig 2). Since the chemistry of the two sites are quite similar, with X% amino-acid identity (or homology), we hypothesize that a difference in accessibility of the allosteric site is responsible for the differences in ligandability.

Our data point to differences in the dynamics of the α7 helix and the MIDAS as the primary determinants of ligandability. Two separate ensemble analyses based on x-ray crystallography data indicate a broad distribution of α7 conformations for Lfa-1, while the same helix in Mac-1 shows two distinct conformations (Fig 3). NMR peak-broadening and a destabilizing mutation in Mac-1 further support the notion that the Lfa-1 helix is more dynamic than the same region of Mac-1 (Figs 4, 7). These data are consistent with previous studies that suggest that the Lfa-1 I-domain samples a continuum of α7 conformations in its unliganded state and that allosteric inhibitors function by reordering the MIDAS. A mutation in Mac-1 that destabilized the helix did not improve ligandability of the α7 site, further supporting the notion that additional features - perhaps at the MIDAS - are required to attain a ligandable conformation.

Together, the data suggests a model in which, unlike Lfa-1, Mac-1 samples a defined two-state landscape that is probably more iso-energetic between the extremes, with a higher energetic penalty for the excursions that lead to formation of the allosteric pocket (Figure 8). This energy landscape allows Lfa-1 to readily create a ligandable, allosteric pocket adjacent to the helix, while for Mac-1, there is a high energetic cost to forming an analogous ligandable pocket. Future studies which clarify the timescales on which the Mac-1 and Lfa-1 I- domains sample their various conformations will add clarify to the model of accessibility we propose here.

For discovery of small-molecule probes and drug leads, the quality of hits is more important than obtaining a high hit rate, per se. Finding the rare allosteric ligand is more difficult when the conformational ensemble is narrow (as in Mac-1). Additionally, the energy landscape of the α7 helix that controls accessibility of the allosteric site could be markedly different in the

context of the full-length integrin. Nevertheless, this study demonstrates a link between the breadth of the conformational ensemble and the ability to identify ligands that bind at cryptic allosteric sites. These results are therefore significant in the context of defining the fundamental principles of ligandability and allosteric networks within protein families.

METHODS

Protein expression and purification

Wild-type Lfa-1 I-domain (E124-D316) was expressed in BL21(DE3) cells from a pET15b vector with an N-terminal hexahistidine (His6)- tag followed by a TEV protease cleavage site. Cells were grown in LB media (for crystallography) or M9 media (for NMR) and expression was induced with 1mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 4hrs at 37℃.

For co-crystallization experiments, we expressed a smaller construct of wild-type Lfa-1 I-domain (S125-I309) in Rosetta 2(DE3) cells using a pET15b vector containing an N-terminal hexahistidine (His6)- tag followed by a TEV protease cleavage site. The cells were grown in LB or in M9 minimal media containing 1 g/mL of ^{15}N NH_4Cl (Sigma), until an OD_{600} of 0.5 - 0.7 at 37℃. Protein expression was induced using 200 uM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 18℃ overnight. The cells were then pelleted and flash frozen in liquid N2.

For crystallization and NMR experiments, Lfa-1 I-domain was purified from inclusion bodies as previously described with some modifications [22]. The pellet of cells were dissolved in unfolding buffer containing 6 M Gdn-HCl, 20 mM Phosphate buffer, 10 mM Imidazole, 500 mM NaCl, and 1 mM TCEP. The cells were lysed using a sonicator and the inclusion bodies were solubilized for 2hrs at room-temp. The cell debris was pelleted and HisTrap HP column

(GE Healthcare) was used to affinity purify the I-domain from the soluble fraction. The protein was refolded using the rapid dilution method previously described in a buffer with 50 mM Tris pH 8.5, 1 mM MgSO4, 1mM TCEP and 5% (v/v) glycerol. The refolded I-domain was then concentrated using a HisTrap HP column. We added TEV protease to cleave the His6 tag and dialyzed the sample overnight at 4℃ into the crystallization buffer (50 mM Tris pH 8.5, 1 mM MgSO4 and 1mM TCEP). We removed the cleaved tag and uncleaved products by running the sample through an additional HisTrap step. The untagged flow through was loaded onto Superdex 75 gel filtration column equilibrated with crystallization buffer for a final purification step. For NMR experiments, we buffer exchanged into an NMR buffer (10 mM Sodium phosphate pH 7.2, 10 mM $MgSO_4$, 150 mM NaCl and 0.05 % NaN_3) and added 10% D_2O.

For co-crystallization experiments, the pelleted Rosetta-2 cells were dissolved in a buffer containing 50 mM Tris-HCl pH 7.4, 3 mM Imidazole, 500 mM NaCl, 10 mM MgSO4, 1 mM TCEP, and 5% Glycerol in the presence of cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Roche). The cells were lysed using sonication and the soluble fraction was loaded onto a HisTrap HP column (GE Healthcare) for affinity purification. The eluted fractions were dialyzed overnight into a buffer containing 50 mM Tris pH 8.5, 10 mM MgSO4, 1 mM TCEP and 5% (v/v) glycerol while removing the His-tag from the protein using TEV protease. The untagged protein was ran through a Superdex 75 gel filtration column using the co-crystallization buffer (50 mM Tris pH 8.5, 10 mM MgSO4 and 1 mM TCEP).

The Mac-1 I-domain sequence coding for residues 146-342 was codon-optimized for expression in *Escherichia coli* by DNA 2.0 and subcloned into a pET15b plasmid containing a 6xHis affinity tag followed by a TEV protease cleavage site at the *N*-terminus. Point

mutations were made via Megawhop PCR or QuikChange SiteDirected Mutagenesis Kit (Agilent). All constructs were verified by DNA sequencing.

Recombinant Mac-1 I-domain was expressed and purified essentially as described [8] without an ion exchange step. Rosetta 2(DE3) *Escherichia coli* were grown in 2XYT (for crystallography and tethering) or M9 (for NMR) at 37 °C until OD_{600} reached 0.3. The temperature was reduced to 25 °C, and at OD_{600} = 0.5-0.7, expression was induced with 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by culture for 16-20 hours. Cells were harvested by centrifugation; resuspended in 50 mL buffer per L of culture of 50 mM HEPES (pH 7.5), 500 mM NaCl, 10 mM MgCl, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 10 mM imidazole, and 5% w/v glycerol. Cells were lysed by sonication while on ice, and the insoluble lysate fraction was removed by centrifugation. The soluble lysate fraction was collected and incubated with HisPur Cobalt resin (Thermo) for 1-2 hours at 4°C, washed, and eluted by gravity flow in lysis buffer containing 150 mM imidazole. To remove the 6xHis affinity tag, purified protein was incubated and dialyzed overnight at 4 °C with 0.5 mg recombinant Tobacco etch virus (TEV) protease, engineered to contain its own 6xHis affinity tag. Cleavage/dialysis buffer was 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl, 0.25 mM TCEP, and 5% w/v glycerol. TEV protease, contaminants, and uncleaved protein were removed by repass over a HisPur Cobalt resin column equilibrated in lysis buffer. Cleaved and re-passed I-domain was further purified by size exclusion chromatography (SEC) on a Superdex 75 16/600 column (GE Healthcare). For disulfide tethering mutations, SEC was done with 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl, and 5% w/v glycerol. For crystallography, SEC was done with 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl. For NMR, SEC was done with 10 mM HEPES, 150 mM NaCl, 10 mM MgCl,

0.05% (w/v) NaN_3 and 5% D_2O to avoid further buffer exchange. Protein purity of SEC fractions was examined via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and fractions representing >95% purity were combined and concentrated to >5 mg/mL Preparations containing 5% glycerol were flash frozen in LN2 and stored at -80 °C. Preparations without 5% glycerol were stored at 4°C. WT protein identity and cysteine mutation presence were confirmed by intact protein (LC/MS) on a Xevo G2-S (Waters). Yield of pure protein varied across constructs but was 4.2 mg/L for codon-optimized WT Mac-1 I-domain.

Crystallization

Lfa-1 crystals were obtained via hanging drop vapor diffusion method by mixing 1μL of 20-22 mg/ml of protein and 1μL reservoir solution containing 30% (w/v) PEG 3350 and 0.1 M potassium phosphate dibasic. Crystals were formed in 2-5 days at 295K.

Mac-1 crystallization was induced by hanging drop vapor diffusion at room temperature. 2 μL of 10 mg/mL Mac-1 I-domain was mixed with an equal volume of well solution: 10 mM $MgCl_2$, 1.3 M $(NH_4)_2SO_4$ and 1.6 M NaCl. Crystals formed overnight and grew to full size in 4-5 days.

X-ray Data Collection, Data processing and Modeling

The X-ray datasets were collected at the Advanced Light Source (ALS) beamline 8.3.1 on crystals maintained at 273K (RT). The Lfa-1 and Mac-1 I-domain diffraction data was processed and scaled with HKL2000 [24]. The co-crystal structures were processed using DIALS [25]. Space group determination and scaling were done using POINTLESS, AIMLESS and CTRUNCATE in CCP4 [26-28].

The structure was solved by molecular replacement using the program Phaser in PHENIX [29]. The structure of Wild-type LFA-1 I-domain (PDB 3F74; residues G128-Y307) was used as a search model. Six N-terminal residues and the α7–helix were manually rebuilt using Coot [30] and the model was refined using PHENIX [31]. For the co-crystal structures we used the new cryogenic Lfa-1 structure as a search model and used the same pipeline to solve the structures.

X-ray structural heterogeneity Modeling

Multi-conformer models were built using the electron density sampling algorithm qFit 2.0 [32,33]. qFit first builds in up to 4 different conformations as well as occupancies per residue into the electron density and then connects the conformations to give a multi-conformer model. We did a grid search over backbone amplitude ranges 0.05 to 0.35 and sigma value ranges 0.01-0.05 and selected the multi-conformer model with the lowest Rfree value.

The CONTACT algorithm was used to discover network of residues undergoing collective conformational exchange [34]. The algorithm identifies network of residues that can switch between alternate conformations to relieve clashes caused by overlapping van der Waals radii (vdW). A threshold value for clashes (Tstress) was set to the worst 25% of vdW overlaps and the maximum vdW overlap allowed after relief was 10%. We used contactApp which uses Cytoscape and Chimera to visualize the networks (generously supplied by Scooter Morris) [35,36]. We made the CONTACT network figures using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Ensembles of conformations were built using phenix.ensemble_refinement [37]. We did a grid search over pTLS (0.1-1), wXRAY (2.5-10) and tX (0.5-1.5) refinement parameters to obtain optimal values. We selected the ensemble with the lowest Rfree value. Root mean square fluctuation (RMSF) were calculated using the CA atoms of each residue in the ensemble. The

ensemble structures are colored based on RMSF values using the Spectrumany function in PyMOL.

PDB Ensemble generation

The Lfa-1 PDB ensemble was created using PDB ID and CHAIN: 5E6S AND CHAIN A, 5E6S AND CHAIN C, 5E6S AND CHAIN E, 5E6U AND CHAIN A, 5E6R AND CHAIN A, 1RD4 AND CHAIN A, 1RD4 AND CHAIN B, 1RD4 AND CHAIN C, 1RD4 AND CHAIN D, 1ZOP AND CHAIN A, 1ZOP AND CHAIN B, 1ZOO AND CHAIN A, 1ZOO AND CHAIN B, 1ZON AND CHAIN A, 1LFA AND CHAIN A, 1LFA AND CHAIN B, 4IXD AND CHAIN A, 1XUO AND CHAIN A, 1XUO AND CHAIN B, 1XDG AND CHAIN A, 1XDG AND CHAIN B, 1XDD AND CHAIN A, 1XDD AND CHAIN B, 3E2M AND CHAIN A, 3E2M AND CHAIN B, 3EOB AND CHAIN I, 3EOB AND CHAIN J, 3EOA AND CHAIN I, 3EOA AND CHAIN J, 3BQN AND CHAIN B, 3BQN AND CHAIN C, 3BQM AND CHAIN B, 3BQM AND CHAIN C, 1CQP AND CHAIN A, 1CQP AND CHAIN B, 3M6F AND CHAIN A, 2O7N AND CHAIN A, 2ICA AND CHAIN A, 3F78 AND CHAIN A, 3F78 AND CHAIN B, 3F78 AND CHAIN C, 3F74 AND CHAIN A, 3F74 AND CHAIN B, 3F74 AND CHAIN C, 1ZON CHAIN A, and 1DGQ AND CHAIN A (the first chain from the NMR ensemble). We aligned these structures to the cryogenic Lfa-1 structure we have solved using residue ranges 120-310 and calculated root mean square deviation (RMSD) values using the CA atoms of each residue. The ligand stabilized high affinity structure with a native α7–helix (3TCX) was not used in the RMSD calculations and is not shown on the image of the ensemble (pink). The PDB ensembles are colored by RMSD values.The ensemble included all the structures currently (Oct. 2017) found in the PDB except for the disulfide bond stabilized structures and a structure that had the α7–helix pointing the opposite direction compared to other structures in the PDB. We didn't

include the disulfide bond stabilized structures because we wanted to measure the conformational heterogeneity of Lfa-1 close to a native condition.

The PDB ensemble for Mac-1 I-domain was generated using PDB ID and CHAIN: 1BHO AND CHAIN 1, 1BHO AND CHAIN 2, 1BHQ AND CHAIN 1, 1BHQ AND CHAIN 2, 1IDN AND CHAIN 1, 1IDN AND CHAIN 2, 1IDO AND CHAIN A, 1JLM AND CHAIN A, 1M1U AND CHAIN A, 1MF7 AND CHAIN A, 1N9Z AND CHAIN A, 1NA5 AND CHAIN A, 3Q3G AND CHAIN E, 3Q3G AND CHAIN G, 3Q3G AND CHAIN I, 3Q3G AND CHAIN L, 3QA3 AND CHAIN E, 3QA3 AND CHAIN G, 3QA3 AND CHAIN I, 3QA3 AND CHAIN L, 4M76 AND CHAIN B, and 4XW2 AND CHAIN A. We used residue ranges 130-319. We split the ensembles to open and closed state ensembles. The open state is defined as structures with an overall CA RMSD values greater than 1 Å and included PDB ID and CHAIN: 1N9Z_A 4M76_B 1M1U_A 1IDO_A and 4XW2_A. We aligned and calculated RMSD values of these structures as described for the Lfa-1 PDB ensembles.

NMR Data Collection

All Lfa-1 NMR experiments were conducted on a Bruker 800 MHz spectrometer at 300 K. We conducted a 2D ^{1}H-^{15}N HSQC TROSY experiments as implemented in trosyetf3gpsi pulse sequence; avance-version (12/01/11) [38-43]. For Lfa-1_BMS-688521 titration experiments, 413 μM of uniformly ^{15}N-labeled protein in the NMR buffer (10 mM Sodium phosphate pH 7.2, 10 mM MgSO4, 150 mM NaCl and 0.05 % NaN3) was used to collect spectra at 0 μM, 10 μM, 50 μM, 100 μM, 150 μM and 500 μM concentrations of BMS-68852. For the Lovastatin experiments, we titrated in 0 μM, 25 μM, 125 μM, 200 μM and 1000 μM of lovastatin into 187 μM of the ^{15}N-labeled protein in the NMR buffer.

Mac-1 NMR samples were purified in or buffer exchanged into 5% $^{2}H_2O$, 10 mM phosphate buffer pH 7.2 with 150 mM NaCl, 10 mM $MgSO_4$ and 0.05% w/v NaN_3. Several samples were used to acquire backbone amide resonance assignments, a uniformly $^{13}C/^{15}N/^{2}H$-labeled sample at 1.3 mM/0.5 mL and a uniformly $^{13}C/^{15}N$-labeled and ~70% ^{2}H-labeled sample at 0.7 mM/0.5 mL. A suite of three-dimensional triple-resonance experiments were collected on each: HNCA, HNcoCA and ^{15}N-NOESY-HSQC on the 70% ^{2}H sample, and HNCO, HNcoCACB and HNCACB on the U-$^{13}C/^{15}N/^{2}H$-labeled sample. Additionally, $^{15}N/^{1}H$ HSQC spectra were collected on five additional samples containing either ^{15}N-glycine, leucine, alanine, isoleucine or valine to achieve single-residue $^{15}N/^{1}H$-labeling (Appendix Figure 11). A Transverse relaxation optimized spectroscopy (TROSY) experiment was used to deconvolute overlapped sidechain resonances. Chemical shifts with unambiguous assignments are reported in Appendix Table 5.

All data were collected at 298 K on Bruker Avance 500 MHz and 800 MHz instruments equipped with cryogenic probes. Data were processed and analyzed in NMRPipe [44], CCPNMR [45] and SPARKY [46].

NMR Spectral Analysis for Titration experiments

The Lfa-1 NMR titration spectra were processed using NMRPipe [44]. We automatically processed the data series using autoProc.tcl script and used nmrDraw to find peaks from the series. The backbone chemical shift assignments for Lfa-1 I-domain were obtained using both BMRB entry 18941 and 4553 [47-48]. We transferred the assignments to the titration series using ipap.tcl script in NMRpipe. To export the chemical shift values for the Lfa-1 our series, we used view2D.tcl. Chemical shift values for Mac-1 were recorded manually by overlay and comparison processed spectra in CcpNmr Analysis. The chemical shift perturbation (CSP) values

for both Lfa-1 and Mac-1 were calculated using modelTitr.tcl script with coefficient values of 1.0 and 0.2 for HN and 15N respectively (CSP = Sqrt[1.0*(X shift)^2 + 0.2*(Y shift)^2]). CSP values are displayed in the figures using a putty representation and a color gradient from grey to crimson/ruby. A wider tube and a higher intensity color indicate high CSP values. Full and per peak spectral images were made using CcpNmr Analysis using a macro generously provided by Wayne Boucher [45].

NMR Intensity Calculations

Peak intensities were calculated using nmrDraw. Intensities are calculated by averaging the intensities from multiple points around the peak. In the figures, the calculated intensities are shown as bar graphs with rolling average values shown as a green line graphs.

Tethering Experiments

I-domain constructs containing target cysteines were diluted to 500 nM in 20 mM Tris (pH 8.0) containing 500 μM β-mecaptoethanol (βME). 25 μL of the reaction mixture was dispensed into 384-well low-volume V-shape plate (Grenier Bio). Using a Biomek FX (Beckman), 50 nL of tethering fragment was pinned into each well from a 384-well source plate containing 50 mM fragments in DMSO. The reaction mixture was incubated for 3 hours at room temperature, and covalent adduct formation was detected via intact protein LC/MS as previously described [16].

SPR Mac-1/Q325F I-domain binding C3dg

Recombinant C3dg (Complement 3 residues 955-1296) was expressed and purified identically to the Mac-1 I-domain (above). SPR data were collected on a Biacore 4000 (GE

Healthcare). 50ug/mL of C3dg, diluted in 10 mM HEPES pH 7.5, 0.1% Tween 20 , was immobilized to an activated CM5 chip surface via EDC/NHS random amine coupling. WT and Q325F Mac-1 were tested on a 60s immobilization with surface RU of 339. Sensograms were collected on increasing concentrations of WT of Q325F Mac-1 with a 450s association step followed by a 1000s dissociation step. Reference subtraction, dissociation constant determination, and sensogram plotting were performed in Biacore 4000 Analysis Software (GE Healthcare).

REFERENCES

[1] Motlagh, H.N., Wrabl, J.O., Li, J., and Hilser, V.J. (2014). The ensemble nature of allostery. Nature *508*, 331–339.

[2] Changeux, J.-P., and Christopoulos, A. (2016). Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation. Cell *166*, 1084–1102.

[3] Rasmussen, S.G.F., Choi, H.-J., Fung, J.J., Pardon, E., Casarosa, P., Chae, P.S., DeVree, B.T., Rosenbaum, D.M., Thian, F.S., Kobilka, T.S., et al. (2011). Structure of a nanobody-stabilized active state of the β2 adrenoceptor. Nature *469*, 175–180.

[4] Jura, N., Zhang, X., Endres, N.F., Seeliger, M.A., Schindler, T., and Kuriyan, J. (2011). Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. Mol. Cell *42*, 9–22.

[5] Raman, A.S., White, K.I., and Ranganathan, R. (2016). Origins of Allostery and Evolvability in Proteins: A Case Study. Cell *166*, 468–480.

[6] Kang, S., Kim, C.U., Gu, X., Owens, R.M., van Rijn, S.J., Boonyaleepun, V., Mao, Y., Springer, T.A., and Jin, M.M. (2012). Complex structure of engineered modular domains defining molecular interaction between ICAM-1 and integrin LFA-1. PLoS One *7*, e44124.

[7] Shimaoka, M., and Springer, T.A. (2003). Therapeutic antagonists and conformational regulation of integrin function. Nat. Rev. Drug Discov. *2*, 703–716.

[8] Bajic, G., Yatime, L., Sim, R.B., Vorup-Jensen, T., and Andersen, G.R. (2013). Structural insight on the recognition of surface-bound opsonins by the integrin I domain of complement receptor 3. Proc. Natl. Acad. Sci. U. S. A. *110*, 16426–16431.

[9] Xiong, J.P., Li, R., Essafi, M., Stehle, T., and Arnaout, M.A. (2000). An isoleucine-based allosteric switch controls affinity and shape shifting in integrin CD11b A-domain. J. Biol. Chem. *275*, 38762–38767.

[10] Watterson, S.H., Xiao, Z., Dodd, D.S., Tortolani, D.R., Vaccaro, W., Potin, D., Launay, M., Stetsko, D.K., Skala, S., Davis, P.M., et al. (2010). Small molecule antagonist of leukocyte function associated antigen-1 (LFA-1): structure-activity relationships leading to the identification of 6-((5S,9R)-9-(4-cyanophenyl)-3-(3,5-dichlorophenyl)-1-methyl-2,4-dioxo-1,3,7-triazaspiro[4.4]nonan-7-yl)nicotinic acid (BMS-688521). J. Med. Chem. *53*, 3814–3830.

[11] Weitz-Schmidt, G., Welzenbach, K., Brinkmann, V., Kamata, T., Kallen, J., Bruns, C., Cottens, S., Takada, Y., and Hommel, U. (2001). Statins selectively inhibit leukocyte function antigen-1 by binding to a novel regulatory integrin site. Nat. Med. *7*, 687–692.

[12] Ley, K., Rivera-Nieves, J., Sandborn, W.J., and Shattil, S. (2016). Integrin-based therapeutics: biological basis, clinical use and new drugs. Nat. Rev. Drug Discov. *15*, 173–183.

[13] Ryu, J.K., Petersen, M.A., Murray, S.G., Baeten, K.M., Meyer-Franke, A., Chan, J.P., Vagena, E., Bedard, C., Machado, M.R., Rios Coronado, P.E., et al. (2015). Blood coagulation protein fibrinogen promotes autoimmunity and demyelination via chemokine release and antigen presentation. Nat. Commun. *6*, 8164.

[14] Jensen, M.R., Bajic, G., Zhang, X., Laustsen, A.K., Koldsø, H., Skeby, K.K., Schiøtt, B., Andersen, G.R., and Vorup-Jensen, T. (2016). Structural Basis for Simvastatin Competitive Antagonism of Complement Receptor 3. J. Biol. Chem. *291*, 16963–16976.
[15] Kallen, J., Welzenbach, K., Ramage, P., Geyl, D., Kriwacki, R., Legge, G., Cottens, S., Weitz-Schmidt, G., and Hommel, U. (1999). Structural basis for LFA-1 inhibition upon lovastatin binding to the CD11a I-domain. J. Mol. Biol. *292*, 1–9.
[16] Hallenbeck, K.K., Davies, J.L., Merron, C., Ogden, P., Sijbesma, E., Ottmann, C., Renslo, A.R., Wilson, C., and Arkin, M.R. (2017). A Liquid Chromatography/Mass Spectrometry Method for Screening Disulfide Tethering Fragments. SLAS Discov 2472555217732072.
[17] Fahr, B.T., O'Brien, T., Pham, P., Waal, N.D., Baskaran, S., Raimundo, B.C., Lam, J.W., Sopko, M.M., Purkey, H.E., and Romanowski, M.J. (2006). Tethering identifies fragment that yields potent inhibitors of human caspase-1. Bioorg. Med. Chem. Lett. *16*, 559–562.
[18] de Jesus Cortez, F., Suzawa, M., Irvy, S., Bruning, J.M., Sablin, E., Jacobson, M.P., Fletterick, R.J., Ingraham, H.A., and England, P.M. (2016). Disulfide-Trapping Identifies a New, Effective Chemical Probe for Activating the Nuclear Receptor Human LRH-1 (NR5A2). PLoS One *11*, e0159316.
[19] Ostrem, J.M., Peters, U., Sos, M.L., Wells, J.A., and Shokat, K.M. (2013). K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature *503*, 548–551.
[20] Raimundo, B.C., Oslob, J.D., Braisted, A.C., Hyde, J., McDowell, R.S., Randal, M., Waal, N.D., Wilkinson, J., Yu, C.H., and Arkin, M.R. (2004). Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of IL-2: an approach for inhibiting protein-protein interactions. J. Med. Chem. *47*, 3111–3130.
[21] Kukic, P., Alvin Leung, H.T., Bemporad, F., Aprile, F.A., Kumita, J.R., De Simone, A., Camilloni, C., and Vendruscolo, M. (2015). Structure and dynamics of the integrin LFA-1 I-domain in the inactive state underlie its inside-out/outside-in signaling and allosteric mechanisms. Structure *23*, 745–753.
[22] Legge, G.B., Kriwacki, R.W., Chung, J., Hommel, U., Ramage, P., Case, D.A., Dyson, H.J., and Wright, P.E. (2000). NMR solution structure of the inserted domain of human leukocyte function associated antigen-1. J. Mol. Biol. *295*, 1251–1264.
[23] Surade, S., and Blundell, T.L. (2012). Structural biology and drug discovery of difficult targets: the limits of ligandability. Chem. Biol. *19*, 42–50.
[24] Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
[25] Waterman, D. G. et al. The DIALS framework for integration software. CCP4 Newsletter on Protein Crystallography 49, 16–19 (2013).
[26] Evans, P. (2005). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. *62*, 72–82.
[27] Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D Biol. Crystallogr. *67*, 282–292.
[28] Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of theCCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.
[29] Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.
[30] Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.
[31] Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2012). PHENIX: a comprehensive Python-based system for macromolecular structure solution. In International Tables for Crystallography, pp. 539–547.
[32] van den Bedem, H., Dhanik, A., Latombe, J.C., and Deacon, A.M. (2009). Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. Acta Crystallogr. D Biol. Crystallogr. *65*, 1107–1117.
[33] Keedy, D.A., Fraser, J.S., and van den Bedem, H. (2015). Exposing Hidden Alternative Backbone Conformations in X-ray Crystallography Using qFit. PLoS Comput. Biol. *11*, e1004507.
[34] van den Bedem, H., Bhabha, G., Yang, K., Wright, P.E., and Fraser, J.S. (2013). Automated identification of functional dynamic contact networks from X-ray crystallography. Nat. Methods *10*, 896–902.
[35] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

[36] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

[37] Burnley, B.T., Afonine, P.V., Adams, P.D., and Gros, P. (2012). Modelling dynamics in protein crystal structures by ensemble refinement. Elife *1*, e00311.

[38] Czisch, M., and Boelens, R. (1998). Sensitivity enhancement in the TROSY experiment. J. Magn. Reson. *134*, 158–160.

[39] Meissner, A., Schulte-Herbrüggen, T., Briand, J., and Sørensen, O.W. (1998). Double spin-state-selective coherence transfer. Application for two-dimensional selection of multiplet components with long transverse relaxation times. Mol. Phys. *95*, 1137–1142.

[40] Pervushin, K.V., Wider, G., and Wüthrich, K. (1998). Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. J. Biomol. NMR *12*, 345–348.

[41] Rance, M., Loria, J.P., and Palmer AG3rd (1999). Sensitivity improvement of transverse relaxation-optimized spectroscopy. J. Magn. Reson. *136*, 92–101.

[42] Weigelt, J. (1998). Single Scan, Sensitivity- and Gradient-Enhanced TROSY for Multidimensional NMR ExperimentsJ. Am. Chem. Soc.1998,120, 10778−10779. J. Am. Chem. Soc. *120*, 12706–12706.

[43] Zhu, G., Kong, X.M., and Sze, K.H. (1999). Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. J. Biomol. NMR *13*, 77–81.

[44] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

[45] Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

[46] Goddard TD, Kneller DG. Sparky. Vol. 3. University of California; San Francisco: 1999

[47] Kriwacki, R.W., Legge, G.B., Hommel, U., Ramage, P., Chung, J., Tennant, L.L., Wright, P.E., and Dyson, H.J. (2000). Assignment of 1H, 13C and 15N resonances of the I-domain of human leukocyte function associated antigen-1. J. Biomol. NMR *16*, 271–272.

[48] Zimmerman, T., Oyarzabal, J., Sebastián, E.S., Majumdar, S., Tejo, B.A., Siahaan, T.J., and Blanco, F.J. (2007). ICAM-1 peptide inhibitors of T-cell adhesion bind to the allosteric site of LFA-1. An NMR characterization. Chem. Biol. Drug Des. *70*, 347–353.

Chapter 5

CHARACTERIZING AND INHIBITING THE PROTEIN-PROTEIN INTERACTIONS OF THE MAC-1 I-DOMAIN

Contributing Authors:
SAMUEL J. PFAFF, MENGQI ZHONG, JAE K. RYU, ANKE MEYER-FRANKE, KATERINA AKASSOGLOU, MICHELLE R. ARKIN

INTRODUCTION

Recent re-appreciation for the importance of innate immunity has led to a deeper understanding of host defense and autoimmune disease, opening the door to novel therapeutic strategies targeting specific innate immune receptors. Canonically, innate immune cells patrol healthy tissues using many redundant receptors to detect injury or infection and initiate the recruitment of specialized B-cells and T-lymphocytes [1]. In this paradigm, inhibiting innate immune receptors is an unattractive therapeutic strategy, leading to either broad immunosuppression, or dampening by compensatory pathways. However, work on Multiple Sclerosis pathogenesis found that disrupting the protein-protein interaction (PPI) between blood-clotting protein fibrin(ogen) and microglial (and macrophage) integrin Mac1 ablated neurodegeneration in mice [2]. Mac1 is an integrin expressed on the surface of microglia, neutrophils, and monocytes involved in cell adhesion and chemotaxis. Binding of Mac1 by some extracellular PPI partners initiates inflammation, and loss of functional Mac1 results in immunocompromise [3]. Genetically replacing the fibrin epitope for Mac1 with alanine is enough to increase survival in mice with experimental autoimmune encephalomyelitis, a model disease for Multiple Sclerosis in mice. Notably, these mice retain all hemostatic functions

associated with fibrin clotting, suggesting that the fibrin/Mac1 interaction is independent of fibrinogen's platelet signaling activity or fibrin's clotting function.

While the biological validation data are compelling, the fibrin/Mac1 molecular interaction is complex. Fibrinogen is proteolyzed and cross-linked to form heterogeneous fibrin clots at sites of injury; it is the fibrin form that appears to attract Mac1-binding cells. Through a poorly understood mechanism, fibrin displays a 'cryptic epitope' to immune cell receptors that is not observed with soluble, full-length fibrinogen [4]. This cryptic epitope, called P2, is in the C-terminal γ-chain of fibrinogen. P2 is recognized by the Mac1 ligand binding domain, an apical 200-residue Inserted (I)-domain [5]. Immobilizing an 18-mer peptide containing the P2 sequence on a surface such as an ELISA plate is sufficient to cause adhesion of cells expressing Mac1 and to detect binding of purified Mac1 I-domain [6]. Interestingly, the P2 binding site on the I-domain is thought to be different from the canonical ligand-binding site for well-characterized complexes such as the complement protein iC3b6. Despite the biological relevance of the interaction, the structural basis for fibrinogen γ-chain binding to Mac1 I-domain remains unclear.

The unresolved nature of the γ-chain/I-domain interaction is not unique to fibrinogen. Mac1 has 30+ reported protein-protein interaction partners, and rules delimiting Mac1 PPI partners similarly lacks a clear structural basis. Previous work on the mechanism behind Mac1's PPI promiscuity centered on ligand structure and sequence. Podolnikova et al. examined peptide arrays of reported Mac1 ligands and identified a generic recognition motif of a basic residue flanked by one or more hydrophobic residues (H-B-H) [9]. However, a similar analysis by the same group observed the Mac1 I-domain was capable of binding to multiple fibrinogen-derived peptides, not just P2, even though some of the sequences do not fit the proposed H-B-H binding

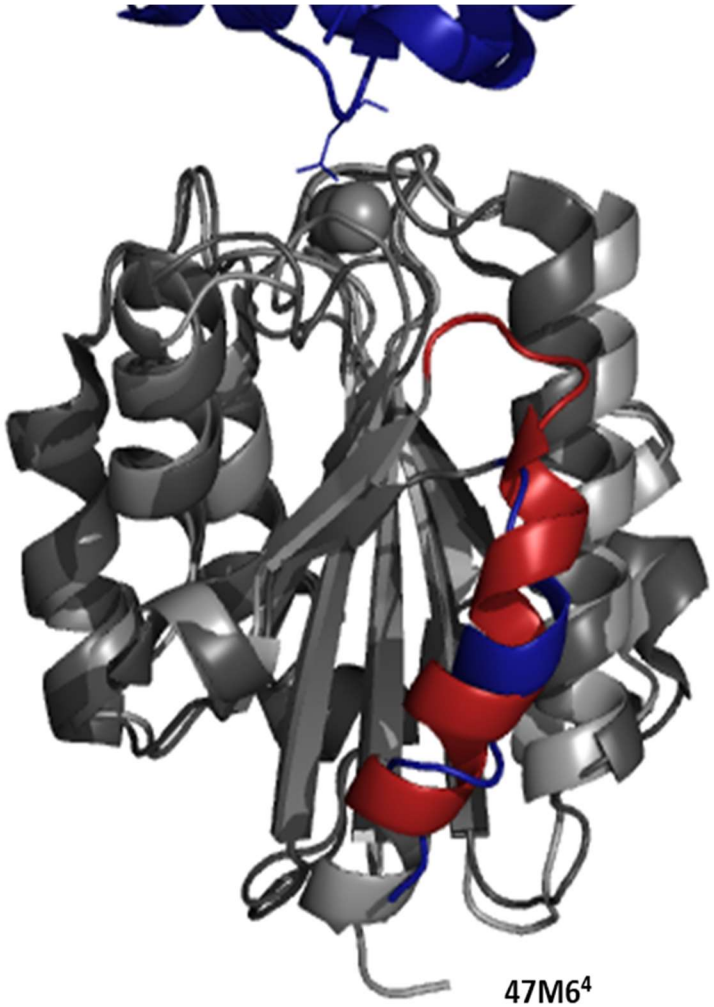


Figure 1. Mac-1 Undergoes a conformation switch upon ligand binding. The Mac-1 I-domain is constitutively in a low-affinity conformation (red, see also Chapter 3 Fig 1) where the C-terminal α7 helix rest in an up position. Upon complex formation with C3d, the α7 shifts downwards ~10Å to transmit an allosteric signal to the proximal domain in the integrin.

motif [10]. I hypothesized that Mac-1's propensity to interact with diverse ligands was encoded in the structure of the I-domain rather than in its PPI partners.

Five of the eight integrin I-domains are structurally characterized (αX, αM/Mac-1, αL/Lfa-1, α1 and α2) and they all adopt nearly identical Rossman-like folds, with a series of seven α-helices surrounding six β-sheets to form a cylinder. The interaction of Mac-1 with iC3b and its fragments is a canonical I-domain/ligand interaction (Fig 1). As discussed in Chapter 4, trapping of the allosteric α7 helix into the low-affinity conformation is known to reduce binding

of Lfa-1 and Mac-1 to PPI partners. Therefore, identifying small molecules that trap the Mac-1 α7 helix in a low-affinity form is attractive (a) as a potential inhibitor of inflammatory signaling in innate immunity, and (b) as a chemical probe to test whether the multiple Mac-1 I-domain ligands converge on a common allosteric signaling pathway.

Here I report work toward such a model for the Mac-1 I-domain. I characterized the binding interaction between the fibrinogen γ-chain and the Mac1 I-domain across a variety of complimentary binding assays, using the canonical MIDAS-binding ligand iC3b as a control. I collected 1H/15N HSQC ligand titrations on U-15N labeled Mac-1 to compare the molecular rearrangements in the I-domain upon ligand binding. Attempts to add resolution to these data by co-crystallizing Mac-1 and the fibrinogen γ-chain to determine a binary complex structure were unsuccessful. Finally, a disulfide screening campaign for small molecule fragments targeting the α7-helix pocket yielded the first reported allosteric small molecule for the Mac-1 I-domain. This hit can serve as a starting point for a comparative chemical biology approach to understanding the Mac-1/fibrinogen interaction and the structural basis for Mac-1 PPI promiscuity.

RESULTS

The physiological interaction of Mac-1 with its protein partners occurs in the context of the heterogeneous extracellular matrix. Often, the partner protein is deposited on a surface such as a fibrin clot (in the case of fibrinogen) or an opsonized cell membrane (in the case of Complement 3 and its fragments). Surface-based binding assays are a natural choice to study Mac-1 PPIs *in vitro.* In an Enzyme Linked Immunosorbent Assay (ELISA) study of the Mac-1 I-domain binding to a surface of adsorbed γ-chain or iC3b, Mac-1 binds to γ-chain with an

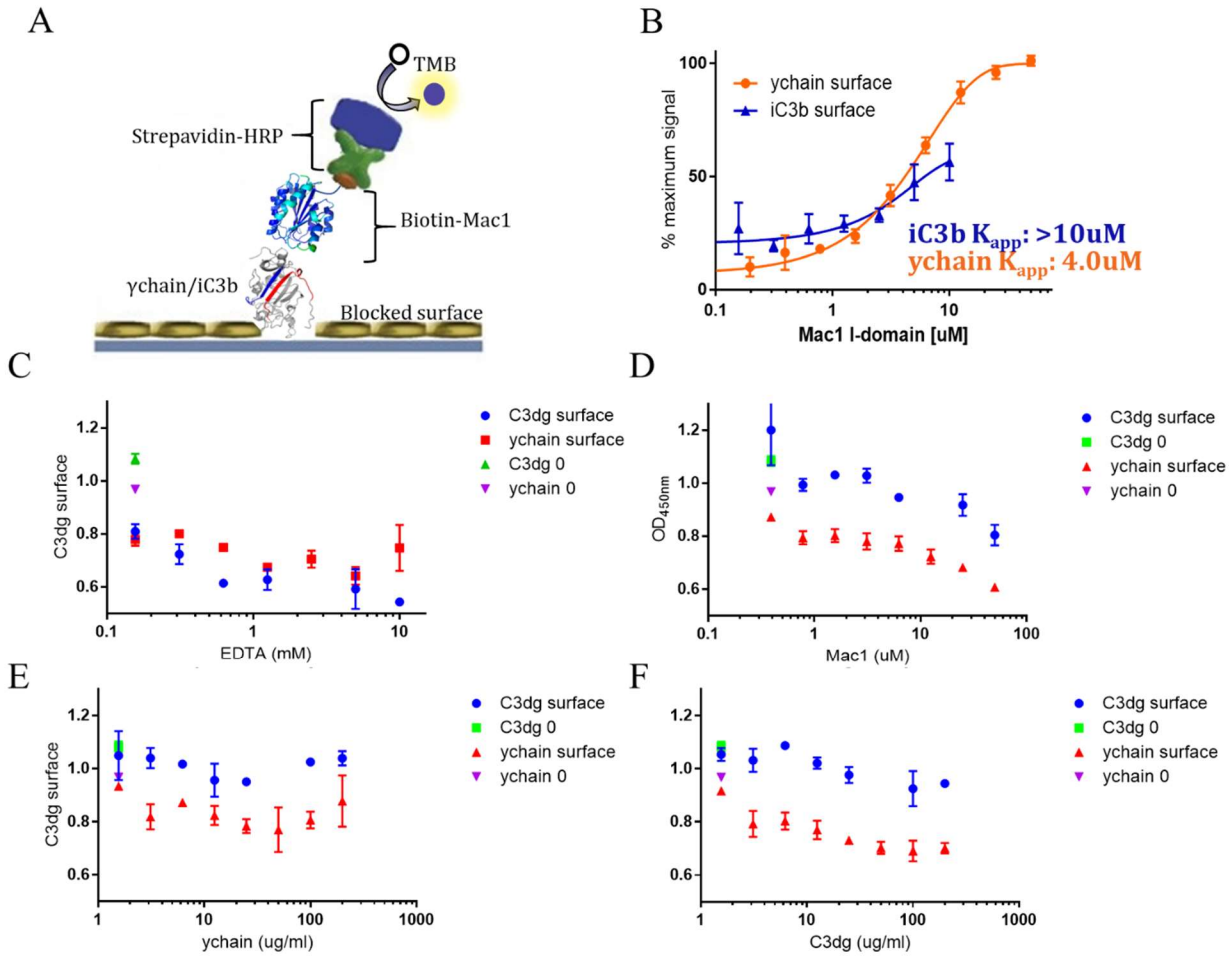


Figure 2. Mac-1 I-domain bind γ-chain and iC3b in a Protein-Protein Interaction ELISA. (A) Assay format for detecting biotin-Mac1 I-domain to an ELISA surface via Strepavidin-HRP. (B) Titration of Mac-1 onto a surface of either γ-chain or iC3b reveals a >2-fold higher affinity for adsorbed γ-chain. Binding of 6uM biotinylated Mac1 I-domain in the presence of increasing EDTA (C), unlabeled Mac1 I-domain (D), γchain (E), and C3dg (F) to wells coated with 45 ug/mL C3dg (Blue) or 45ug/mL γchain (red). A zero-competitor control is plotted above the lowest concentration for comparison, (green = zero on C3dg surface, purple = zero on γchain surface).

apparent K_d of 4 μM, a >2-fold preference over iC3b (Fig 2A-B). However, iC3b is a much larger protein than fibrinogen γ-chain (175 kDa and 29 kDa respectively) and adsorbing at the same surface density deposits fewer iC3b molecules than γ-chain molecules. To correct for this, recombinant C3dg was expressed and purified. C3dg is a smaller C3 fragment (43 kDa) competent to bind the I-domain in a MIDAS-dependent fashion [7]. Mac-1 binds C3dg with an apparent K_d of 2 μM, weaker than the reported K_d of 450 nM determined by Isothermal

Calorimetry [7]. In the ELISA format, C3dg and γ-chain are equally potent ligands for Mac-1. Both interactions were MIDAS dependent and were ablated by chelation of the Mg^{2+} by addition of EDTA (Fig 2C). Unlabeled Mac-1 was able to compete with biotin-Mac-1, as expected (Fig 2D). However, addition of soluble γ-chain did not compete with either a γ-chain or C3dg surface, and soluble C3dg showed only modest inhibition to either surface (Fig 2E-F).

The reported propensity of the Mac-1 I-domain to bind to generic hydrophobic sequences convolutes the interpretation of ELISA binding data, given the disordered adsorption of ligand inherent to an ELISA and the inability of the soluble ligands to compete with identical molecules adsorbed to the assay surface. The observed binding of Mac1 to C3dg and γ-chain could depend on partial or total ligand unfolding driven by adsorption. While this possibility aligns well with the γ-chain cryptic epitope hypothesis, C3dg is known to make an ordered, one-to-one interaction with the I-domain MIDAS. The surprisingly similar 2-4 μM apparent K_d of γ-chain and C3dg led us to study the contribution of surface adsorption to the affinity of Mac-1 for its ligands using a pair of surface-based biophysical techniques, Bio-Layer Interferometry (BLI) and Surface Plasmon Resonance (SPR).

In SPR, a dextran-coated gold surface is functionalized with an immobilization chemistry, allowing the site-specific or random coupling of proteins to the activated surface. Binding is detected through the change in refractive index (RI) at the surface; the RI is proportional to the mass deposited on the activated surface. The affinity and kinetics of interactions between two molecules can therefore be studied using microfluidics-controlled flow of a solution containing one molecule over and SPR surface containing the other molecule. To best imitate a heterogenous fibrin clot, γ-chain was coupled to a GE Healthcare CM5 Biacore chip via Random Amine Coupling (RAC) [8]. RAC covalently links free amines, such as

surface-exposed lysines, to the chip surface. The surface lysines on both γ-chain and the C3dg control are evenly distributed across the protein surface. While some linkages will obscure the putative Mac-1/ligand binding site, the overall surface serves as an approximation of the complex substrates of fibrin clots and C3 opsonins.

Immobilization of C3dg and γ-chain to a CM5 surface through RAC was optimized by varying surface activation time, protein concentration, and contact time. The minimum activation and contact time (60s each) reduced the surface response to <1000 RU for each protein, when injected at 50 ug/mL. Lower concentrations of ligand failed to immobilize. Binding of wild-type Mac-1 I-domain to the immobilized ligand surfaces yielded dose-dependent binding sensograms (Fig 3A-B). Mac-1 bound both C3dg and γ-chain with a fast k_{on}, followed by a slow binding step which did not saturate. These data yielded K_d = 4.0 uM for C3dg and K_d = 33 uM for γ-chain. Both surfaces had <20% theoretical Rmax, indicating a large population of the randomly-coupled ligands were unavailable for Mac-1 binding.

BioLayer Interferometry relies on similar surface reflection principles as SPR. However, in BLI the surface is a disposable biosensor tip that can be manipulated by an automated tip-head to sample various analytes in a plate-based format. To test whether the affinity of Mac-1 for C3dg and γ-chain was influenced by either the microfluidic flow of SPR or the RAC immobilization method, Mac-1 I-domain was genetically tagged with a C-terminal Avi sequence and co-expressed with bacterial biotin ligase BirA. The resulting Mac-1 I-domain was confirmed to be uniformly labeled with biotin via intact protein LC/MS. Biotin-Mac-1 was loaded onto Octet Red (ForteBio) biosensor tips coated in streptavidin and exposed to a solution of either γ-chain or C3dg (Fig 3A-B). In this reversed format, the C3dg/Mac-1 interaction maintained fast

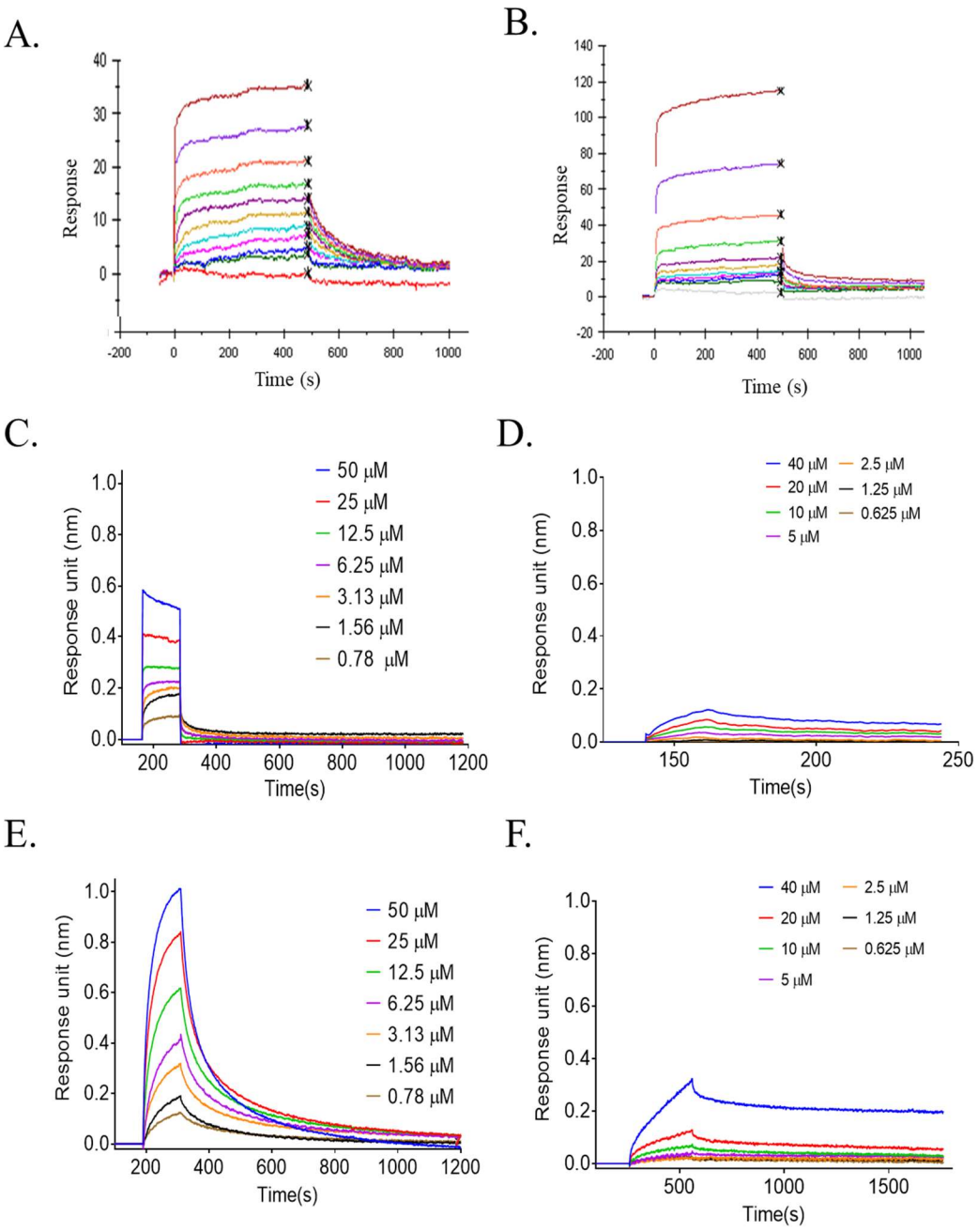


Figure 3. Mac-1 I-domain binds γ-chain and iC3b in a Octet and SPR formats. (A) Increasing concentrations of WT Mac-1 flowing over an SPR chip coated with 339 RU of C3dg yields a K_d = 4.0uM. (B) Increasing concentrations of WT Mac-1 flowing over an SPR chip coated with 996 RU of γ-chain yields a K_d = 33uM (C & E) an Octet tip coated in biotin-Mac-1 WT (C) or I316G (E) dipped into increasing concentrations of C3dg. (D & F) An Octet tip coated in biotin-Mac-1 WT (D) or I316G (F) dipped into increasing concentrations of γ-chain.

on-rate and off-rates as demonstrated by the instantaneous rise to a signal plateau after the Mac-1-coated tip enters and exits the ligand solution at 150s and 250s respectively (Fig 3A). In

contrast, WT Mac-1 binding to γ-chain is mostly ablated (Fig 3D), suggesting the affinity for Mac-1 to the γ-chain observed in SPR is greatly increased by deposition on a surface.

Previous studies of the Mac-1/fibrinogen interaction had identified 'activating' mutations in the I-domain which increased affinity of the I-domain for various ligands by as much as 10-fold [11]. Surprisingly, introducing I316G to the I-domain converts the fast on/off rate of I-domain binding to C3dg to slow-on and slow-off, as well as increasing the overall signal/noise (Fig 3E). Mac-1 I316G binding to γ-chain is increased compared to WT (Fig 3D & F), but the γ-chain that binds to the Mac-1-coated biosensor does not dissociate even with an extensive washing time (>1000s). The slow kinetics suggests the interaction is irreversible and may result from a non-specific mechanism such as partial unfolding of the I-domain or the γ-chain.

From these data and NMR studies described below, I hypothesize that I-domain affinities measured using activating point mutations or truncations do not represent increases in K_d, but rather a propensity for co-aggregation. Whether this co-aggregation is the primary means of Mac-1/fibrin interaction in the context of full-length fibrin and integrin is unclear. Regardless, the preference of Mac-1 to bind surface-adsorbed γ-chain versus soluble γ-chain (Fig 2D, 3E) explains the inability of γ-chain to compete for I-domain binding in the competition ELISA format (Fig 2E). A partial lack of competition was observed for C3dg (Fig 2F), despite the ability of the I-domain to bind soluble C3dg (Fig 3C). Taken together, these data suggest the Mac-1 I-domain interactions with γ-chain and C3dg may share a common binding mechanism when the ligands are adsorbed to a surface, but that C3dg can access a soluble binding mode unavailable to γ-chain.

To understand if γ-chain or the putative P2 binding epitope retain any ability to interact with WT Mac-1 I-domain in a soluble format, I turned to Nuclear Magnetic Resonance (NMR)

spectroscopy, a technique capable of detecting the fine conformational changes in protein structure that define PPIs. Cross peaks in 1H/15N-HSQC are due to individual H-N bonds in the 15N-labeled protein, most common in the peptide-bind amide. Chemical shift perturbation (CSP) of 1H/15N cross peaks upon binding of ligands denote changes in the environment of specific residues. Perturbations can be due to changes in local environment (seen as a shift in the peak and/or broadening of the peak) and/or due to changes in the relaxation rate of the resonance, as when a larger protein complex is formed (seen as broadening of the peak). When paired with x-ray structures, chemical shift perturbation enables the identification of individual residues and broader structural motifs involved in a PPI.

The structure and function of the Mac-1 I-domain are known to be sensitive to the presence of a divalent metal ion in the apical MIDAS [7]. To characterize the regions of the I-domain structurally coupled to the MIDAS, a series of 1H/15N HSQC spectra were collected on U-15N Mac-1 I-domain with increasing concentrations of EDTA (Appendix Fig 12). The resulting CSP were binned into strong, intermediate and weak shifts and mapped to the I-domain crystal structure (red, orange, yellow in Fig 5A). As expected, the strongest CSP occurred at the MIDAS. Another region of CSPs occurred at the α7 helix, aligning with the well-known allosteric coupling of the α7 helix to the MIDAS (see Chapter 4). Finally, a region of intermediate CSP occurred in beta-sheet five, a part of the hydrophobic core connected to a MIDAS-coordinating loop, providing evidence for involvement of the I-domain core in propagating structural rearrangement at the MIDAS.

To compare CSP caused by PPI ligands with those demonstrated by EDTA, similar 1H/15N HSQC experiments were collected on C3dg, γ-chain, and LL-37, a small anti-microbial peptide reported to bind the Mac-1 I-domain [12]. A constant concentration of U-15N Mac-1 I-

domain with increasing concentrations unlabeled protein partner. LL-37 is an ideal candidate for NMR studies because it contains only 37 residues, and binding of unlabeled LL-37 to 15N-labeled Mac-1 can be detected without peak broadening or loss of HSQC signal. Fast-exchange CSP was observed in an 8-step titration of LL-37 up to 1.8 mM, allowing the identification of residues in direct contact with the peptide (Appendix Fig 13). Mapping those residues to the Mac-1 crystal structure revealed a profile similar to EDTA: the most intense shifts occurred at the MIDAS, α7 helix and the β-sheet under the MIDAS loop between the metal biding site and the top of the α7 helix. These data provided strong evidence LL-37 bound to the I-domain in a MIDAS-dependent fashion and caused the canonical allosteric α7 helix shift. Additional structural characterization via the determination of an I-domain/LL-37 co-crystal structure would confirm this hypothesis.

Unlike the LL-37 interaction, formation of a 1:1 Mac-1:γ-chain or Mac-1:C3dg complex would result in a 50-60 kDa heterodimer, more than double the molecular weight of free Mac-1 (22.8 kDa). The slower solution tumbling of the larger complex would lead to broadening and then loss of 15N-HSQC signal if the Mac-1 and protein partner form a complex. As expected, we observed extensive peak broadening upon the addition of C3dg. However, the high-affinity of the complex led to loss of >90% of HSQC signals in one titration step (Appendix Fig 13, 485uM Mac-1, 240uM C3dg). Repeating the Mac-1/C3dg HSQC titration in a finer dose response will allow assignment of residues implicated in the interaction.

Consistent with the hypothesis γ-chain does not access a binding-competent state in solution, the dose-response with γ-chain revealed significantly less signal broadening (Fig 4). In a 1:1 interaction, regions in direct contact with γ-chain would undergo more rapid signal relaxation, allowing the mapping of important residues for the Mac-1/γ-chain interaction (Fig 4,

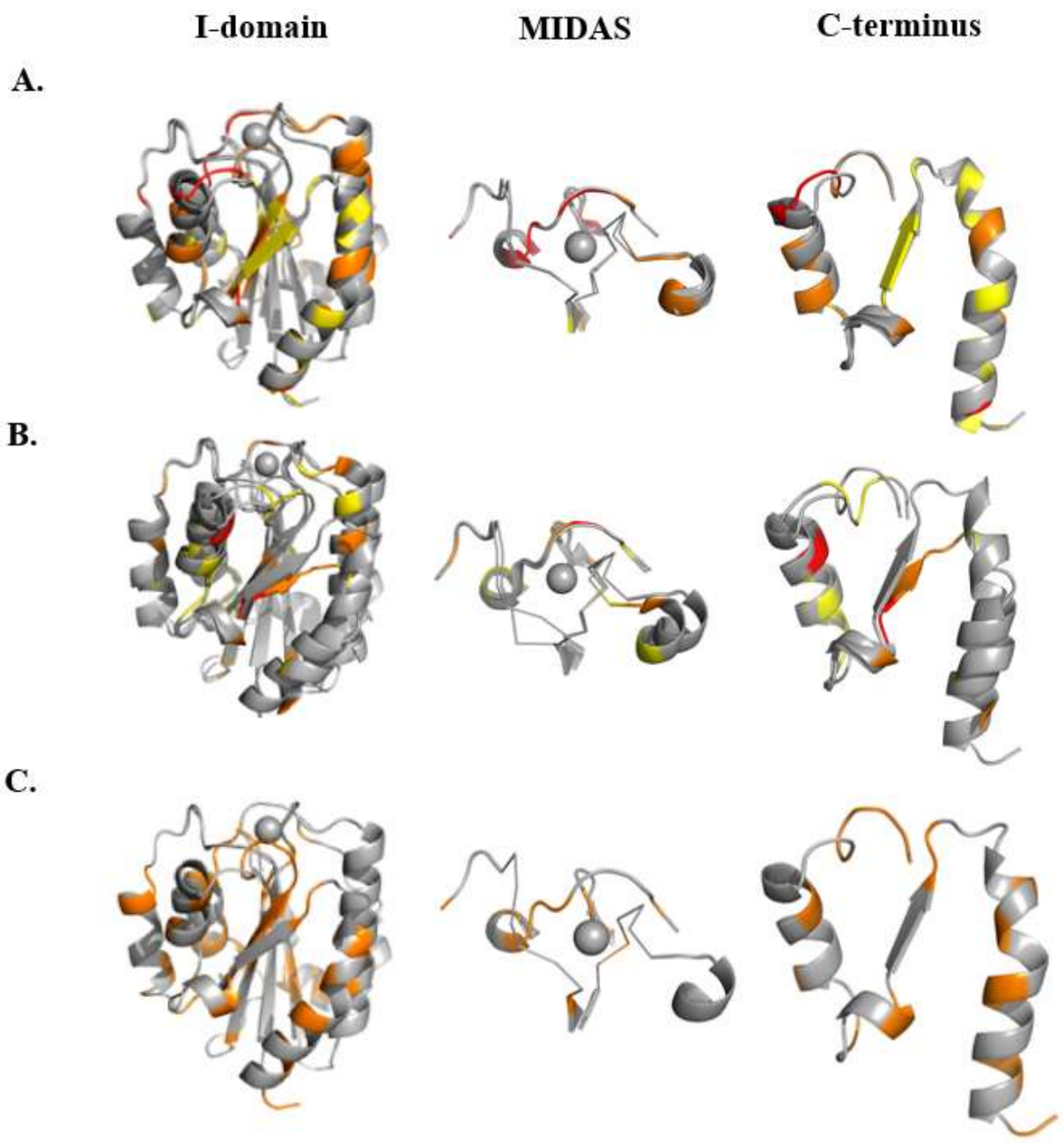


Figure 5: 15N HSQC Titration CSP Mapped to Mac-1 Crystal Structure. WT Mac-1 chemical shift perturbations at saturating concentrations of (A) EDTA or (B) LL-37, or precipitation-limited concentrations of (C) γ-chain.

arrows). Correlating broadened peaks with their location in the crystal structure of the Mac-1 I-domain revealed diffuse impact of γ-chain binding. Residues at the MIDAS, α7 helix, as well as the 'back' of the protein all underwent relaxation (Fig 5C).

Though these NMR data are consistent with a 1:1 Mac-1:γ-chain binding event followed by an overall conformation shift to a 'bound' state, they are also consistent with co-aggregation of receptor and ligand. In the final stage of the 15-N HSQC titration, visible aggregates were observed in the NMR tube. Recovering the sample, centrifuging to collect any aggregates, and resuspending the precipitate allowed SDS-PAGE analysis of the residue's contents. Bands with molecular weights corresponding to both Mac-1 and γ-chain were present in the precipitate, supporting the co-aggregation hypothesis (data not shown). The aggregation and precipitation of the complex is consistent with BLI and ELISA data, which both suggest that soluble γ-chain does not bind soluble Mac-1 I domain in a reversible manner. From these data, we cannot determine whether aggregation-dependent binding is due to truncating the native proteins (I-domain for Mac-1; γ-chain for fibrinogen) or whether it reflects a biologically relevant binding activity.

Co-crystallization of the Mac-1 I-domain with C3d, a truncation of C3dg, was reported in 2015, resulting in a high-resolution crystal structure which elucidated the structural basis for the C3d/Mac-1 interaction at the MIDAS [7]. I therefore explored co-crystallizing the Mac-1 I-domain with γ-chain and its peptide epitope P2. Initial screens centered around previously described I-domain and γ-chain crystallization conditions. I then expanded into 96-well plate-based screening of the JCSG Core I, II, III, IV and V suites. The results are summarized in Table 1. Focused screens were performed on conditions from the 96-well plates with granular precipitate or microcrystals (Appendix Fig 14). Crystals which grew to >50 um were confirmed to be protein and taken to the Advanced Light Source (ALS) 8.3.1 beamline for analysis. All crystals diffracted poorly (>3 Å resolution). Datasets collected from diffracting crystals were indexed, which revealed the unit cell dimensions were too small to contain a Mac-1:γ-chain

complex. Future co-crystallization efforts of the Mac-1:γ-chain would require a redesign of the crystallization constructs, or the introduction of crystallization chaperones such as γ-chain or I-domain specific Fabs.

TABLE 1: FIBRINOGEN/MAC-1 CO-CRYSTALLIZATION TRIALS

DATE (MO, YR)	PROTEINS	NUMBER OF WELLS	CRYSTALLIZATION BUFFER	CRYSTALLIZATION SALTS	CRYSTALLIZATION ADDITIVES	RESULT	SYNCHOTRON DATA
MAY-16	aM Idomain, ychain	24 (12x1uL, 12x2uL)	0.1 M Tris pH 8.0	70 mM CaCl2	15-25% PEG 4000	negative	
MAY-16	aM Idomain, ychain	24 (12x1uL, 12x2uL)	0.1 M MES pH 6.0	70 mM CaCl2	13% PEG 8000	negative	
MAY-16	aM Idomain, P2	4x2uL	pH 5-6	1.3M (NH4)2SO4, 1.55M NaCl, 10 mM MgCl		small crystals	Low res, died in radiation
JUN-16	aM Idomain, P2, Seeding	15x2uL	pH 5-6	1.3-1.8M (NH4)2SO4, 1.56-2.3M NaCl, 10 mM MgCl		negative	
JAN-17	aM Idomain, ychain	24 (12x1uL, 12x2uL)	pH 5-6	1.3-1.8M (NH4)2SO4, 1.56-2.3M NaCl, 10 mM MgCl		small multiform crystals	
FEB-17	aM Idomain, ychain	96x0.2uL	Screening plate	varied	varied	3 hits (next page)	
MAR-17	aM Idomain, ychain	384 (4x96x0.2 uL)	Screening plates	varied	varied	negative	
MAR-17	aM Idomain, ychain	30 (15x1uL, 15x2uL)	0.1M Na Acetate pH 4.6	0.2-0.8 M (NH4)2SO4	15-30% PEG 4000	negative	
MAR-17	aM Idomain, ychain	30 (15x1uL, 15x2uL)	0.1M Na Acetate pH 4.6	0.2-0.8 M (NH4)2SO4	15-30% PEG MME 8000	negative	

Identifying Fragments Binding to Mac-1 Allosteric Pocket

In the α7 helix-targeted screens reported in Chapter 3, the Mac-1 cysteine mutants identified a small number of disulfide fragments capable of tethering to the allosteric site. Of the 6 screened cysteines, the highest number of compounds (42) labelled K331C by more than three standard deviations above the mean (Fig 2D). K331C also selected the highest-labeling fragment (60.6 %-bound at 100 μM fragment, 500 μM βME. We therefore focused on dose-response experiments to identify hits capable of fully labeling K331C. From these experiments, we identified SMDC917607 (Fig 6B), a methyl-hydroxy quinoline capable of fully labeling K331C

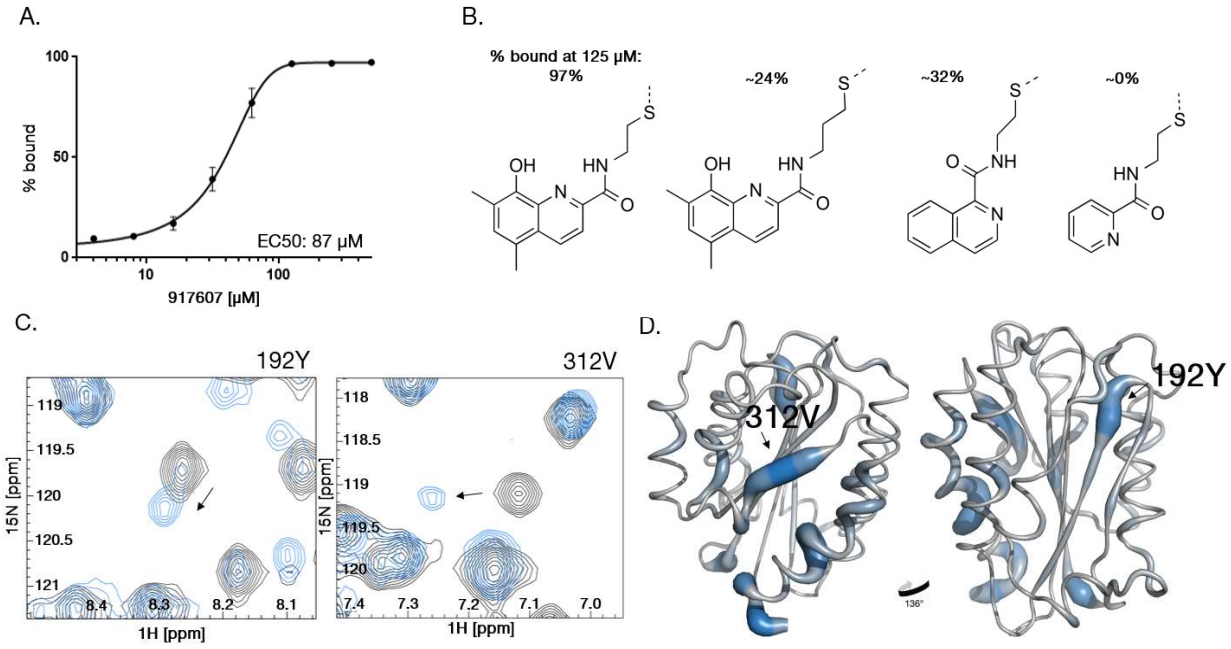


Figure 6: Disulfide fragment binds at α7 allosteric site. **A)** Dose-response for SMDC917607 binding Mac-1 K331C in 10 mM TRIS pH 8.0 and 500 uM βME detected by intact protein LC/MS. **B)** Chemical structure of the fragment SMDC917607 and analogs present in screening library. **C)** 15N-HSQC titration of SMDC917607 into 15N-labeled, K331C Mac-1 I-domain (blue) compared to HSQC of apo K331C Mac-1 (black). **D)** Mapping of CSP, e.g., from panel C, onto the x-ray structure of Mac-1 I-domain, colored from grey (0 ppm) to blue (0.25 ppm).

with a EC50 of 87 µM in 10 mM TRIS pH 8.0 and 500 µM βME (Fig 6A). Notably, several analogs of SMDC917607 contained in our library showed reduced binding to K331C (Fig 6B). Extending the aliphatic linker from 2 to 3 carbons, removing the substitutions to the quinoline ring, or consolidating to a single ring system all diminish compound binding. Increasing the reducing potential of the buffer to 1 mM βME also ablated binding (data not shown), implying a weak interaction driven by non-covalent contributions of the fragment rather than disulfide bond formation.

To determine the binding mechanism of SMDC917607, we collected 15N-HSQC NMR CSP experiments on 0.59 mM K331C Mac-1 I-domain (Fig 6C). The residues demonstrating CSP at 1.1 mM 917607 were mapped to the Mac-1 X-ray structure (Fig 6D). Because the hydroxyquinolone core of SMDC917607 could potentially chelate the Mg2+ present in the Mac-1 MIDAS, we included a 10 mM excess of MgCl2 during data collection. Residues V312 and

F313, near K331C, were two of the largest peak shifts (0.15 ppm and 0.25 ppm respectively), suggesting that these residues made direct contact with SMDC917607 (Fig 6C). Other large chemical shifts included nearby residues A249 and Q314 (0.20 ppm and 0.14 ppm, respectively) and residues in the unstructured C-terminus of the α7 helix (I335 and G337). Y192 was the only residue that shifted significantly in regions distant from the tethering site on the α7 helix (Fig 6C). Notably, Y192 was positioned directly below the MIDAS, providing evidence for allosteric coupling between the α7 helix and the MIDAS.

To determine whether SMDC917607 could bind to the α7 site in the absence of a disulfide bond to K331C, we performed a 1H-15N-HSQC titration of SMDC917607 against wild type Mac-1 I-domain (Appendix Fig 15). We were unable to saturate binding at the solubility limit of SMDC917607 (1100 μM), suggesting a loss in affinity of >10-fold relative to the disulfide-trapped complex (Fig 6); nevertheless, we did observe CSP for 312V, aromatics in the proposed α7 pocket, and 192Y and surrounding residues (Appendix Fig 15B). Thus, both tethered and noncovalently bound fragment were able to bind to the α7 pocket and induce allosteric shifts at the MIDAS.

DISCUSSION

Here we report the comparison of the binding mode of the Mac-1 I-domain with two ligands, Complement 3 fragment C3dg and fibrinogen γ-chain, as well as the discovery of small molecule allosteric binder of the I-domain α7 helix.

C3dg and γ-chain are both protein ligands of the I-domain, which each bind with single- or double-digit μM affinity when they are deposited on an ELISA or SPR surface (Fig 1).

However, C3dg retains binding in solution-phase experiments, whereas the Mac-1/γ-chain interaction is ablated (Fig 2). NMR experiments at >100μM concentrations point to a co-aggregation or partial unfolding mechanism which may also be promoted by surface deposition the I-domain, γ-chain, or both. Future work to test this hypothesis will require confirmation of co-aggregation, perhaps via Dynamic Light Scattering (DLS), Small-angle X-ray Scattering (SAXS), or the generation of I-domain variants resistant binding the γ-chain. DLS and SAXS would clarify the nature of the I-domain/γ-chain complex and whether the aggregates observed here represent a monodisperse particles with defined stoichiometry, or whether the two proteins simply co-precipitate. I-domain mutations which ablate γ-chain binding but not C3dg or LL-37 binding could be studied to clarify whether the γ-chain is accessing a non-canonical I-domain binding mode.

I also investigated the binding of ligands to the I-domain using NMR. 15N-HSQC spectra of the I-domain with EDTA display baseline CSP profile caused by perturbing the MIDAS: intense shifts in the nearby loops and α7 helix. Repeating the experiments with putative PPI partner peptide LL-37 implicate residues similar to those perturbed by EDTA, suggesting LL-37 binds and induces an allosteric signal via the canonical I-domain α7 signaling helix. C3dg 1H/15N HSQC spectra were inconclusive, making it difficult to compare LL-37 CSP to the published Mac-1/C3dg crystal structure. Either more careful C3dg NMR studies, or co-crystallization of the LL-37/I-domain complex are necessary to confirm the mode of LL=37 binding. Aggregation limited the concentration at which the Mac-1/γ-chain interaction can be studied by NMR. The CSP data we did observe is consistent with γ-chain binding the I-domain with an alternative, non-MIDAS, mechanism. However, a screening campaign to identify conditions for co-crystalizing γ-chain the I-domain was unsuccessful, as was soaking I-domain

crystals with γ-chain or its peptide epitope P2 (Fig 5). The non-canonical binding of γ-chain to the I-domain therefore remains unsubstantiated. Future efforts to determine the atomic structure of the γ-chain/I-domain complex could be continued with single-particle electron microscopy techniques, especially if SAXS studies reveal homogeneous particle formation upon co-aggregation.

Identification of an allosteric inhibitor at the cryptic α7 pocket Mac-1 allows study of the I-domain's various PPIs from a different angle. MIDAS-dependent interactions are sensitive to modulation of α7 helix movement (Chapter 4), while some partners such as fibrinogen γ-chain may be resistant to allosteric inhibitors. Here we describe the first small-molecule binder of the Mac-1 I-domain α7 pocket. Starting from the ligandability tethering screen screens reported in Chapter 3, we perform dose response to identify top hits, and NMR chemical shifts perturbations to validate the binding mode (Fig 6). Experiments to determine if 917607 inhibits I-domain binding to C3dg, γ-chain and LL-37 are an area for immediate future work. Even if the fragment hit itself has no or limited activity, co-crystallization of the 917607/I-domain complex would enable medicinal chemistry expansion of the initial hit.

This work identifies differences in Mac-1 I-domain binding to C3dg and γ-chain and sets the stage for experiments designed to describe those differences. Leveraging the α7 helix fragment hit described here to study the binding modes of I-domain PPI partners will allow future work to circumvent the aggregate-prone biochemistry of fibrin/fibrinogen. Site-specific modulation if the I-domain could be extended beyond the α7 helix to scan for other ligandable pockets which could be partner-specific.

METHODS

Protein Expression and Purification

WT Mac-1 I-domain (146-341) was expressed and purified as described in previous chapters (Chp 2 and 4). C3dg (955-1296) was expressed and purified identically to the Mac-1 I-domain. Synthetic LL-37 was purchased from Elim Biopharma (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES).

Recombinant γ-chain (158-420) was with a protocol modified from [13]. Rosetta-2 cells transformed with the pet20b plasmid containing the y-chain construct were grown to saturation overnight at 37 °C in LB containing ampicillin and chloramphenicol. O/N cultures were diluted 1:60 into 2X YT containing amp/chlor and grown for 2−2.5 h at 37 °C.Between OD600 = 0.3 and 0.4 temperature was dropped to 25 °C, and cells were induced at OD600 = 0.6-0.8 with isopropyl 1-thio-β-d-galactopyranoside (750uL of 500mM IPTG). Culture was incubated for 12-16 hours (O/N) before harvesting via centrifugation.

Harvested cells were re-suspended in washing buffer (50 mM Tris-HCl buffer, pH 7.4, containing 0.15 M NaCl, 5 mM EDTA, and 0.5% Triton X-100) and lysed via sonication. Lysed cells were centrifuged for 30min at 13000g and supernatant was collected for analysis. The pellet was washed 5 times with washing buffer for removal of membranes and soluble proteins, collecting the supernatant each time.

The washed pellet was solubilized in 50 mL of 4 M guanidine hydrochloride (GdnHCl), and the concentration of the protein was measured via A280. The protein in 4 M GdnHCl was diluted with 8M urea to a final concentration of 0.2−0.25 mg/mL and dialyzed against a 4-fold volume of 8 M urea at room temperature for several hours.

The concentration of urea in the container was step-wise reduced to 0.125 M by addition of 20 mM Tris buffer, pH 8.0, at 4 °C over a period of 36-48 hr followed by 12-16 hr dialysis of the protein versus the same buffer. Sometimes cloudy white precipitate and a slight pellet was observed. The protein was then dialyzed for 24 hr versus TBS (20 mM Tris buffer, pH 7.4, with 0.15 M NaCl) containing 1 mM Ca2+.

The refolded protein solution was cleared via centrifugation and concentrated in a 10k Amicon concentrator (Amicon), to <2 mL before injection on a 16/600 S75 Gel Filtration column equilibrated in TBS (20 mM Tris pH 7.4, 0.15M NaCl). Fractions with UV signal were collected, analyzed via LC/MS and SDS-PAGE, and those containing soluble and monomeric y-chain were pooled. Pooled fractions were concentrated to ~5 mg/ml, flash-frozen in LN2 and stored at −80 °C.

ELISA

100 µL of 50 µg/mL partner protein was adsorbed onto clear, flat-bottom, 96-well MaxiSorp plates overnight at 4 °C. Wells were washed with 3x with 200 µL PBST, and blocked with 5% non-fat milk for 60-120 minutes at room temperature. Wells were washed with 3x PBST and biotinylated Mac-1 was incubated with the surface for 60 minutes at room temperature. Wells were washed with 3x PBST, and Strepavidin conjugated to HRP (ThermoFisher) was incubated with the surface for 30 minutes. Wells were wash with 5x PBST and peroxidase substrate was added for 5-20 minutes, stopping with 0.1M HCL before saturation. Absorbance was read at 540 nM read in a Flexstation plate reader (Molecular Devices)

SPR

Surface plasmon resonance experiments were conducted on a Biacore 4000 (GE Healthcare) using a CM5 sensor chip (GE Healthcare). Random amine coupling to the chip surface was achieved with EDC/NHS activation/blocking. Target proteins were immobilized to 300-1000 RU by injecting 50 µg/mL for 60s in immobilization buffer (10mM NaOAc pH 5.0, 150mM NaCl, 0.05% Tween 20). After immobilization, experiments were conducted in running buffer (10 mM HEPES, pH 7.5, 150mM NaCl, 1mM MgCl2, 0.05% Tween 20).

BLI

Experiments were performed on an Octet Red (ForteBio). Biotinylated WT and I316 Mac-1 I-domain was loaded onto Streptavidin Dip and Read Biosensors for kinetics (ForteBio) in assay buffer (PBS pH 7.4, 0.2% BSA, 0.05% Tween 20, 1 mM MgCl2. The tips were washed and blocked by dipping into 10 µM biotin. Sample association step (5-10 minutes) was also performed in 10 µM biotin to minimize signal artifacts.

NMR

1H/15N HSQC spectra were collected on an 800 Mhz Bruker magnet equipped with a cryoprobe. Data was processed and analyzed as described in Chaper 4.

REFERENCES

[1] Turvey, S. E. & Broide, D. H. Chapter 2: Innate Immunity. The Journal of allergy and clinical immunology 125, S24-S32, doi:10.1016/j.jaci.2009.07.016 (2010).

[2] Adams, R. A. et al. The fibrin-derived γ(377-395) peptide inhibits microglia activation and suppresses relapsing paralysis in central nervous system autoimmune disease. The Journal of Experimental Medicine 204, 571-582, doi:10.1084/jem.20061931 (2007).

[3] Rosetti, F. & Mayadas, T. N. The many faces of Mac-1 in autoimmune disease.

[4] Medved, L., Tsurupa G Fau - Yakovlev, S. & Yakovlev, S. Conformational changes upon conversion of fibrinogen into fibrin. The mechanisms of exposure of cryptic sites.

[5] Yakubenko, V. P. et al. Identification of the binding site for fibrinogen recognition peptide gamma 383-395 within the alpha(M)I-domain of integrin alpha(M)beta2.

[6] Yakovlev, S., Zhang, L., Ugarova, T. & Medved, L. Interaction of Fibrin(ogen) with Leukocyte Receptor αMβ2 (Mac-1): Further Characterization and Identification of a Novel Binding Region within the Central Domain of the Fibrinogen γ-Module. Biochemistry 44, 617-626, doi:10.1021/bi048266w (2005).

[7] Bajic, G., Yatime, L., Sim, R.B., Vorup-Jensen, T., and Andersen, G.R. (2013). Structural insight on the recognition of surface-bound opsonins by the integrin I domain of complement receptor 3. Proc. Natl. Acad. Sci. U. S. A. *110*, 16426–16431.

[8] O'Shannessy, D. J., et al. (1992). "Immobilization chemistries suitable for use in the BIAcore surface plasmon resonance detector." Anal Biochem 205(1): 132-136.

[9] Podolnikova, N. P., et al. (2015). "Ligand Recognition Specificity of Leukocyte Integrin αMβ2 (Mac-1, CD11b/CD18) and Its Functional Consequences." Biochemistry 54(6): 1408-1420.

[10] Lishko, V. K., et al. (2004). "Multiple Binding Sites in Fibrinogen for Integrin αMβ2 (Mac-1)." Journal of Biological Chemistry 279(43): 44897-44906.

[11] Xiong, J.-P., et al. (2000). "An Isoleucine-based Allosteric Switch Controls Affinity and Shape Shifting in Integrin CD11b A-domain." Journal of Biological Chemistry 275(49): 38762-38767.

[12] Lishko VK, Moreno B, Podolnikova NP, Ugarova TP. Identification of Human Cathelicidin Peptide LL-37 as a Ligand for Macrophage Integrin αMβ2 (Mac-1, CD11b/CD18) that Promotes Phagocytosis by Opsonizing Bacteria. Research and reports in biochemistry. 2016;2016(6):39-55.

[13] Medved, L., et al. (1997). "Domain Structure and Functional Activity of the Recombinant Human Fibrinogen γ-Module (γ148−411)." Biochemistry 36(15): 4685-4693.

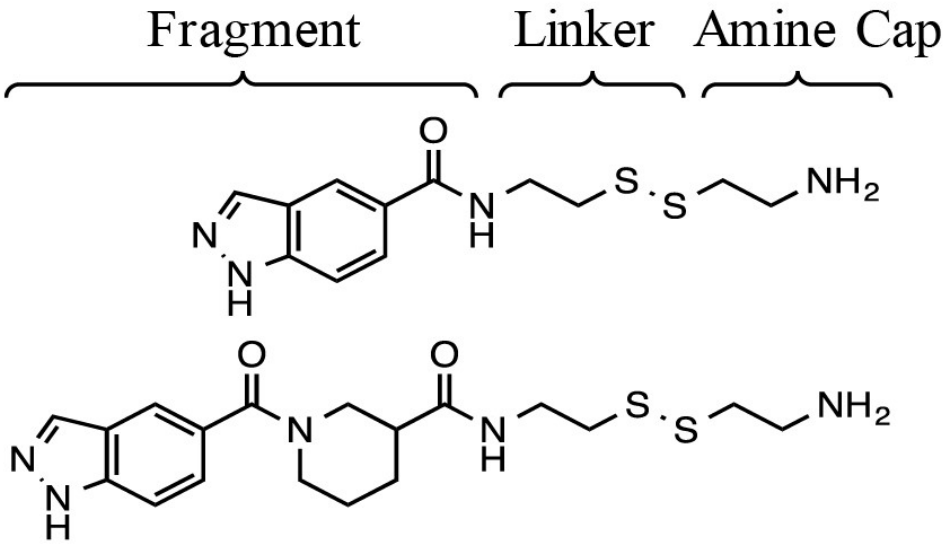
Appendix

Appendix Table 1. % Bound Calculation Bins[a]

Bin Category	*Mass Range*
Protein	$m_{protein} \pm 5$
Protein	$m_{protein} + m_{cap} \pm 5$
Adduct	$m_{protein} + m_{adduct} \pm 5$
Protein	$m_{protein} + 2 \times m_{cap} \pm 5$
Adduct	$m_{protein} + m_{cap} + m_{adduct} \pm 5$
Double-Adduct	$m_{protein} + 2 \times m_{adduct} \pm 5$
Secondary[b]	$m_{\max intensity} \pm 5$
Other	m_{other}

[a]Free protein, capped protein (+βME) and adduct-bound protein (+compound) ranges are each added and binned to calculate a final total and % bound values using Eq. (1). Values outside these bins are placed in 'other' and used to estimate data quality with Eq. (2).

[b]Secondary peaks are reported only if secondary % is greater than % of the expected free, capped or adduct-bound protein.

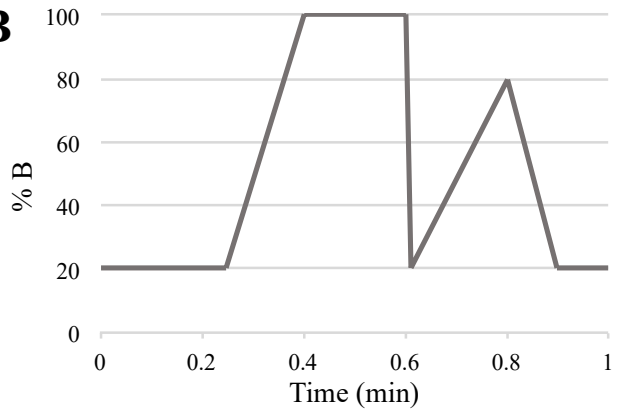


Appendix Figure 1. Disulfide Fragment Scheme. Two example compounds from the SMDC disulfide tethering library. The fragment moiety is connected to a common aliphatic linker terminating in a disulfide capped with a free amine.

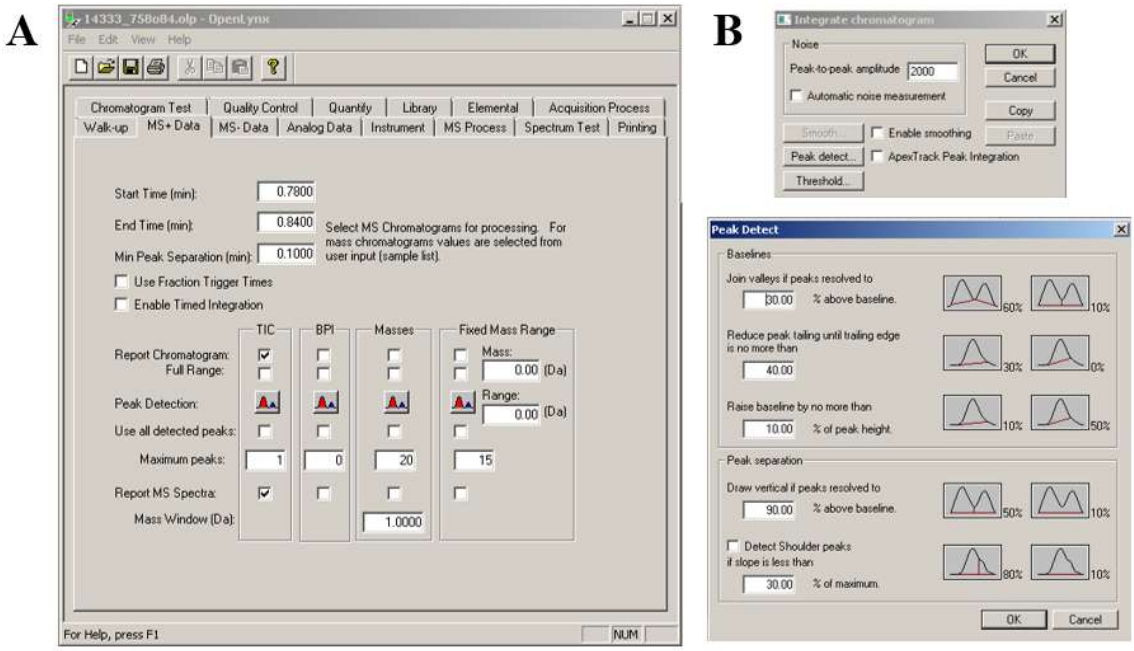
A

Time (min)	% A	%B
0	80	20
0.25	80	20
0.40	0	100
0.60	0	100
0.61	80	20
0.80	20	80
0.90	80	20
1.00	80	20

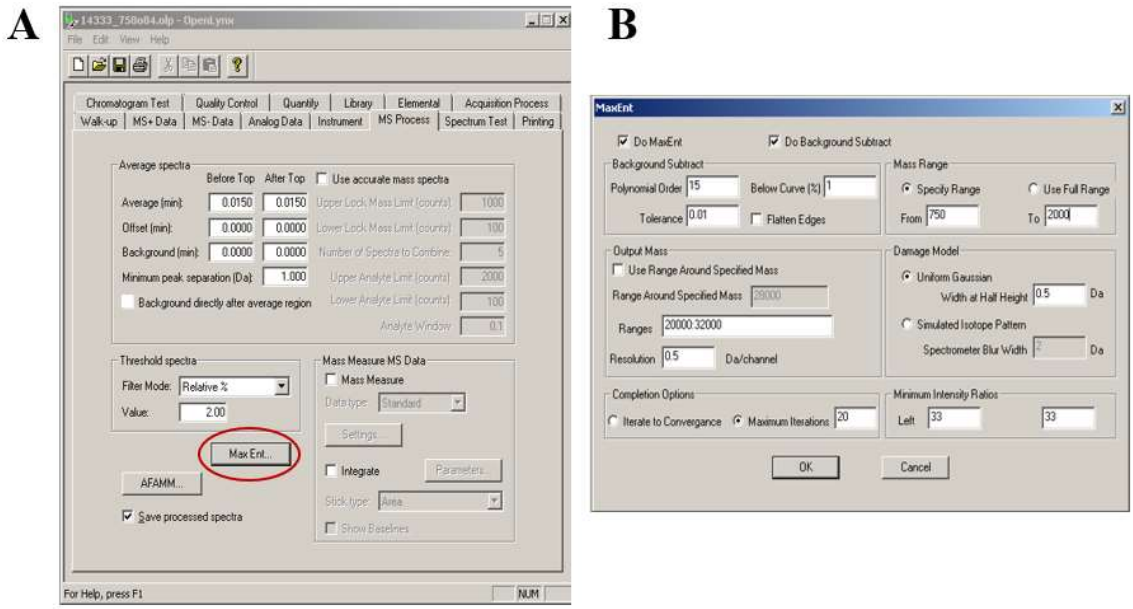
B



Appendix Figure 2. UPLC Elution Strategy. A) Elution scheme for the rapid desalting over a short BEH C4 column. A steady state of 0.25 min is followed by a 0.25 min gradient to %100 B (Acetonitrile). After a 0.20 min hold at %100 B, an instantaneous switch to start conditions is followed by a second 0.20 min 'wash' elution, which is important to minimize sample carry-over. A final re-equilibration at starting conditions prepares for subsequent sample injection. B) a plot of the gradient over time.

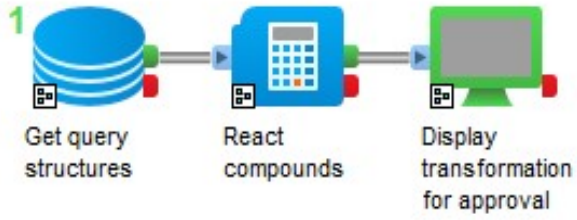


Appendix Figure 3. Openlynx Chromatogram Processing Parameters. A) MS+ input data selection. Start time and end time (min) direct the peak detection to the appropriate portion of the LC chromatogram. B) Peak detection and integration parameters.

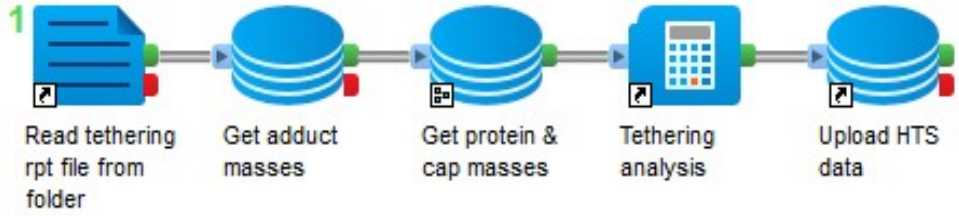


Appendix Figure 4. OpenLynx MaxEnt Deconvolution Parameters. A) MS Process menu for selecting MaxEnt1. B) MaxEnt deconvolution parameters for a representative protein with a MW of 26512 Da, 14-3-3σ (Fig 2).

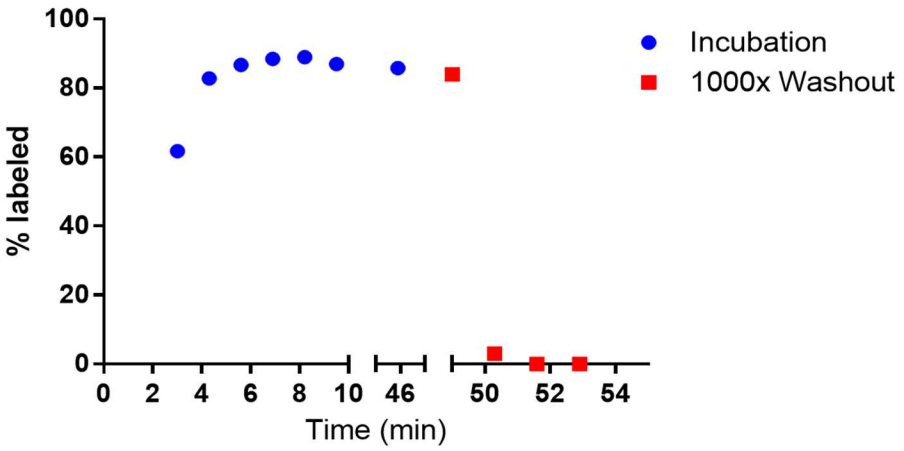
A



B



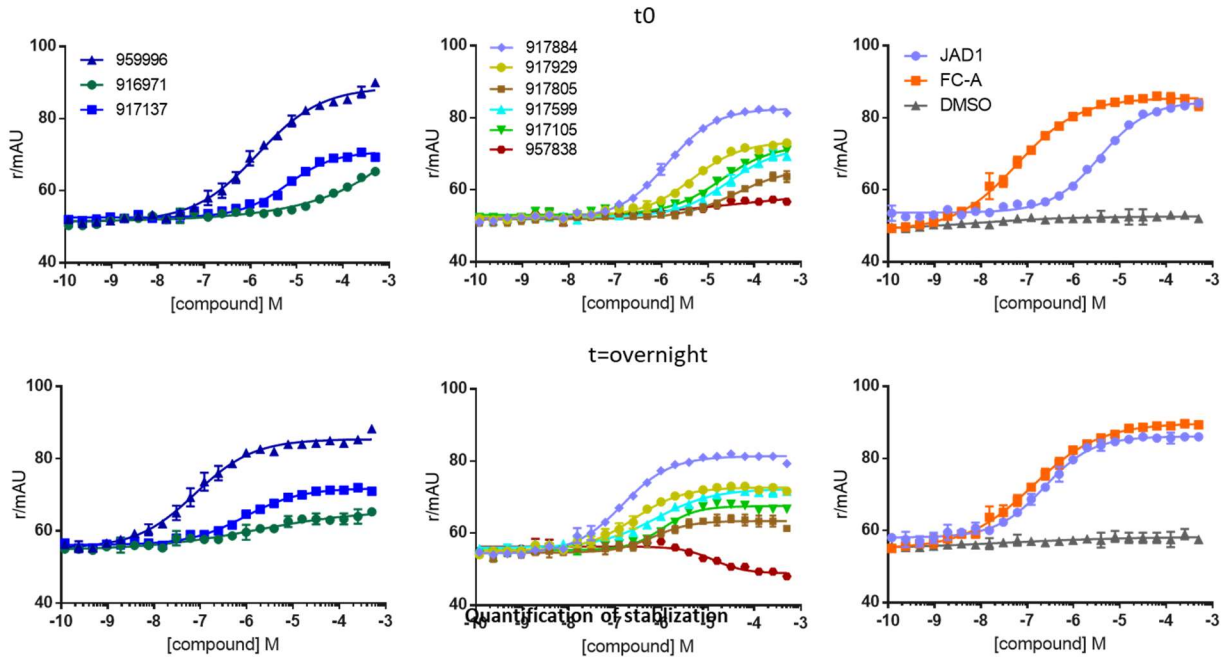
Appendix Figure 5. Pipeline Pilot Workflows. A) The calculation of expected adduct masses was generalized using a three-step algorithm which requires only compound structure. This was essential for developing and adding to the screening library. B) The calculation of % bound values for screening data from the MassLynx output. Eq. (1) is applied in module four and reported below.



Appendix Figure 6. Equilibrium Washout Experiment. Incubation of 14-3-3σ with a lead compound assayed by repeated injection of 2 uL from the same well containing a 100 µL reaction. The disulfide exchange reaches equilibrium in 5 minutes and remains stable for 45 minutes (blue). An aliquot is then diluted 1000-fold with reaction buffer and assayed to confirm reversibility (red).

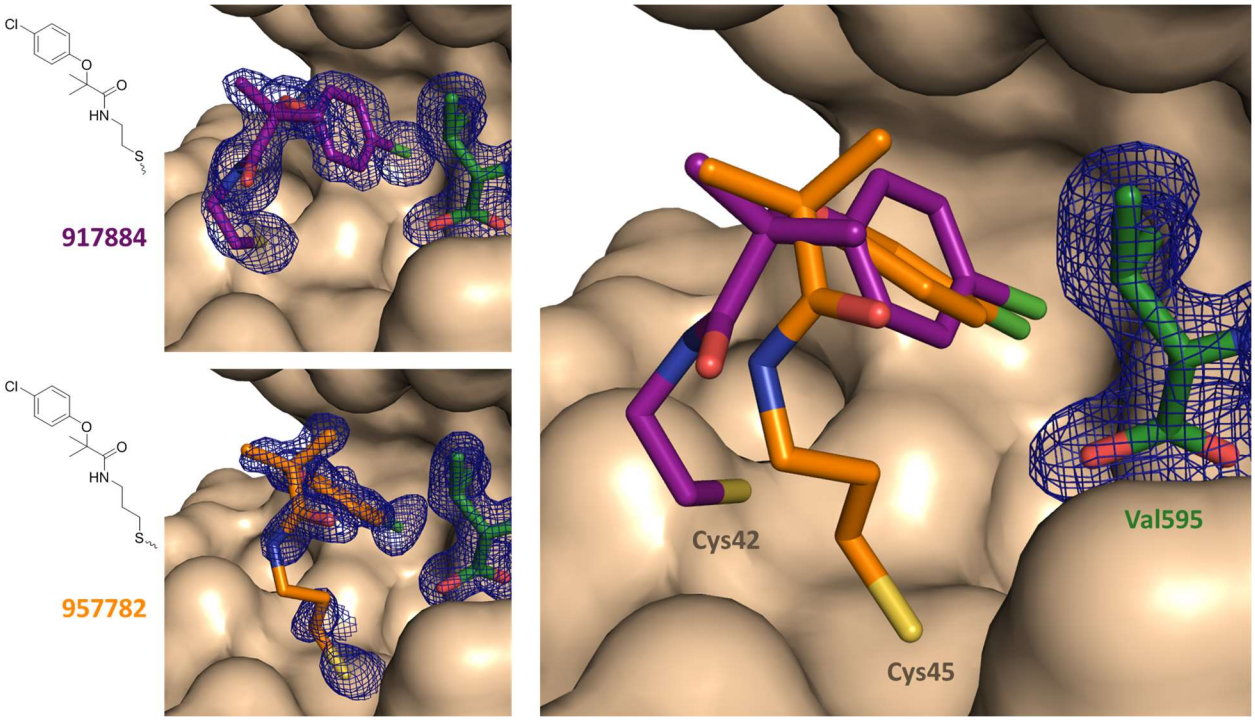
Appendix Table 2: 14-3-3 Tethering hits for follow up experiments

Compound ID	Tethering DR	Cooperative	FP	Xray	Notes
917137	X			Full density	S45C
917929	X			Clear density	N42C
959996	X	X		Full density	N42C
917884	X	X		Full density	N42C
917284	X				S45C antag (negative)
917226	X				S45C antag (negative)
917672	X				S45C antag (negative)
917695	X				S45C antag (negative)
917105	X			Monophore density	N42C
917599	X			Limited density – mostly Cl	N42C
917662	X			No convincing binding mode	S45C
917209	X			No convincing binding mode	S45C
917782	X			Some density – fits well	S45C



	EC_{50} (μM, t0)	EC_{50} (μM, o/n)
959996	1.4 ± 0.2	0.08 ± 0.02
916971	~	1.6 ± 2.9
917137	6.3 ± 1.4	0.93 ± 0.3
917884	1.3 ± 0.1	0.01 ± 0.002
917929	4.5 ± 0.8	0.03 ± 0.005
917805	39.6 ± 19	0.07 ± 0.03
917599	27.4 ± 8.2	0.08 ± 0.02
917105	19.2 ± 5.7	1.1 ± 0.3
957838	~	13.3 ± 6.4
FC-A	0.06 ± 0.01	0.01 ± 0.002
JAD1	3.7 ± 0.5	0.02 ± 0.003

Appendix Figure 7. Fluorescence Polarization Stabilization of 14-3-3/ERα by hit compounds.



Appendix Figure 8: X-Ray crystal structures and overlay of two strongly related fragments in complex with 14-3-3 and ERα. A) Disulfide bound fragment 917884 on Cys42. B) 957782 forming a disulfide bond with Cys45. C) Overlay of the binding modes of 917884 (purple sticks) and 957882 (orange sticks). Nitrogen is depicted in blue, oxygen in red, chlorine in green and sulfur in yellow. 2Fo-Fc electron density maps contoured at 1σ.

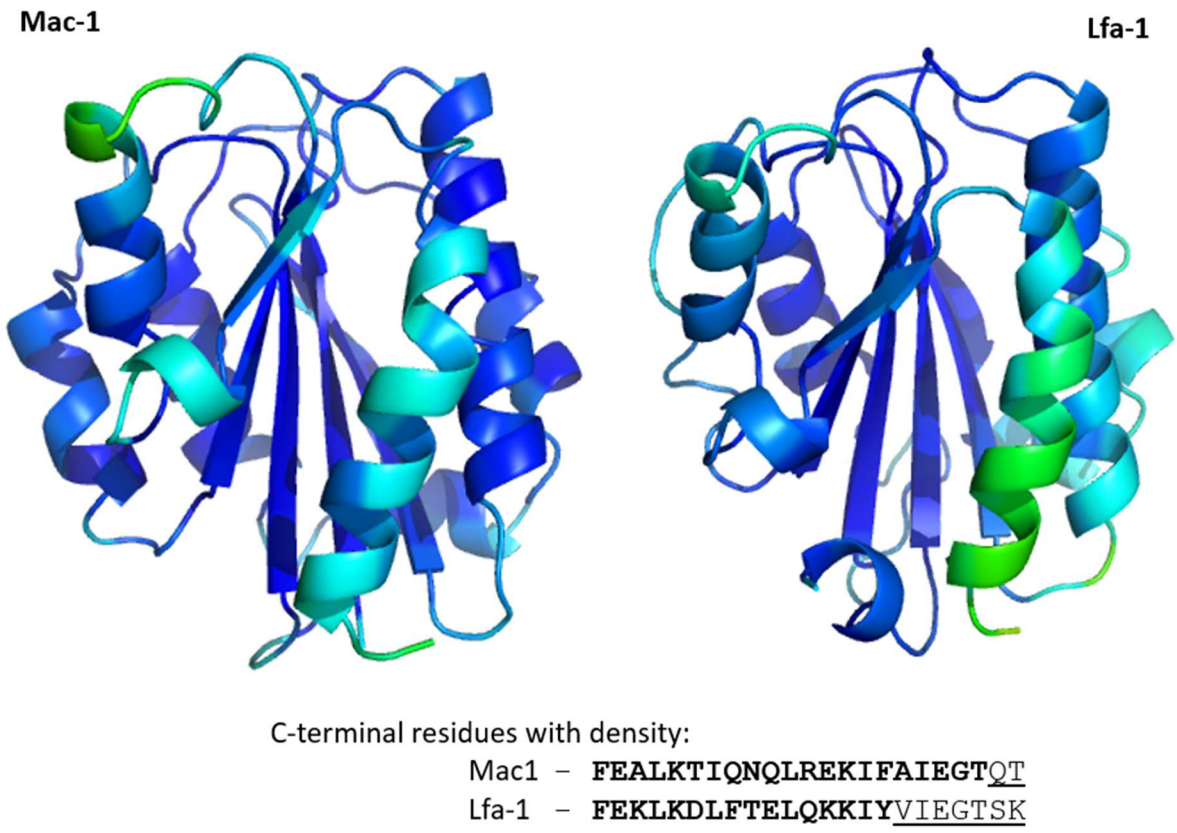
Appendix Table 3: 14-3-3/Fragment co-crystal structure X-ray data statistics

	14-3-3σ/C42-917884	14-3-3σ/C42-959996	14-3-3σ/C42-917929	14-3-3σ/C42-917599	14-3-3σ/C42-917105
Data collection					
Wavelength (Å)	1.54 (SLS)	1.54	1.54	1.54 (SLS)	1.54 (SLS)
Resolution (Å)	66.28-1.70 (1.73-1.70)	45.39-1.80 (1.84-1.80)	45.46-1.80 (1.84-1.80)	66.37-1.70 (1.73-1.70)	66.27-1.70 (1.73-1.70)
Space group	C2221	C2221	C2221	C2221	C2221
Unit cell	82.06 112.42 62.40	81.87 112.21 62.41	82.02 112.53 62.47	82.23 112.39 62.44	82.06 112.36 62.44
Total reflections[a]	148591 (7511)	166665 (6652)	167745 (6546)	151020 (7696)	148994 (7386)
Unique reflections[a]	32032 (1668)	26771 (1423)	26955 (1428)	32176 (1681)	30678 (1530)
Redundancy[a]	4.6 (4.5)	6.2 (4.7)	6.2 (4.6)	4.7 (4.6)	4.9 (4.8)
Completeness (%)[a]	99.8 (100.0)	99.3 (89.4)	99.3 (89.4)	99.9 (100.0)	95.8 (92.1)
Average $I/\sigma_{(I)}$[a]	19.3 (8.0)	30.8 (7.6)	33.6 (8.3)	15.4 (7.5)	17.3 (7.6)
Wilson B-factor	8.2	6.1	7.2	6.3	5.9
$CC_{1/2}$[a,b]	0.998 (0.976)	0.999 (0.972)	0.999 (0.973)	0.995 (0.967)	0.997 (0.973)
R_{sym}[a,c]	0.055 (0.169)	0.048 (0.191)	0.043 (0.176)	0.075 (0.188)	0.066 (0.197)
R_{meas}[a,d]	0.062 (0.192)	0.052 (0.214)	0.047 (0.198)	0.084 (0.212)	0.074 (0.221)
Refinement					
Reflections (refinement)	31975	26753	26935	32128	30631
Reflections (R-free)	1669	1302	1292	1674	1604
Non-hydrogen atoms (protein / solvent)	2225 / 309	2234 / 321	2141 / 226	2256 / 339	2268 / 353
R_{work} (%)	17.7	17.1	18.1	17.7	17.0
R_{free} (%)	21.4	21.5	21.1	20.0	20.7
RMS (bonds) / (angles)	0.007 / 1.43	0.006 / 0.827	0.003 / 0.59	0.006 / 0.883	0.006 / 0.795
Average protein B-factor	12.82	10.94	11.18	10.79	12.19
Ramachandran: favored / outliers (%)	98.3 / 0.0	98.3 / 0.0	98.3 / 0.0	98.3 / 0.0	98.3 / 0.0
Clashscore	0.79	2.12	1.59	1.32	2.11

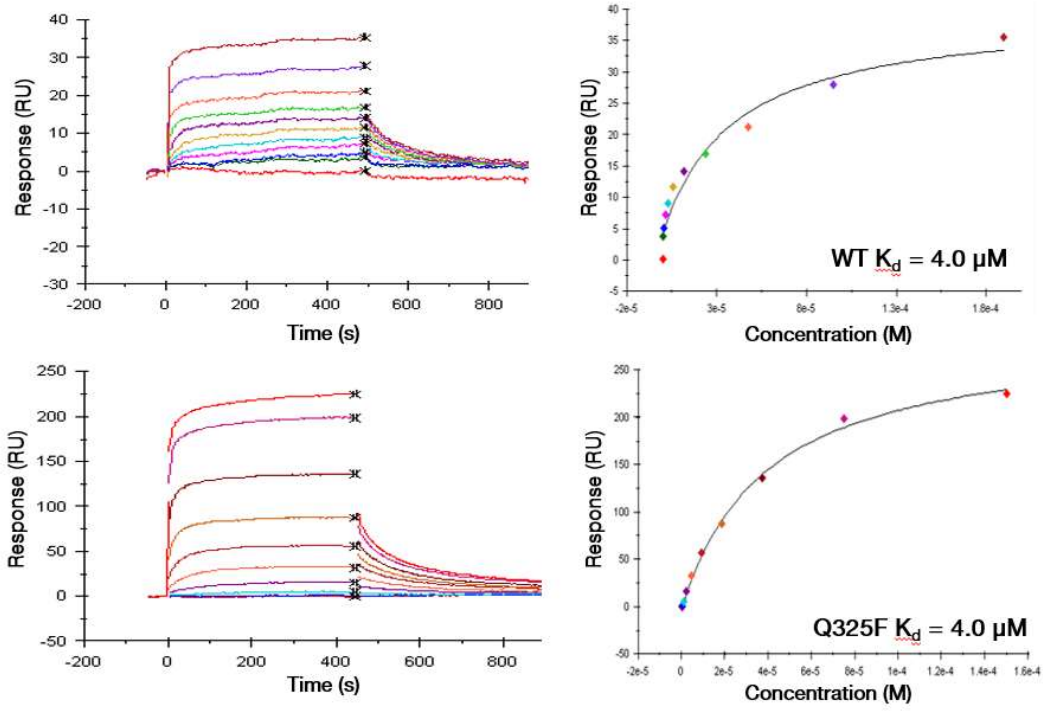
Appendix Table 4: Mac-1/Lfa-1 Room Temperature crystal structure statistics

	Mac-1_RT	Mac-1_RT_Ensemble	Lfa-1_RT	Lfa-1_RT_qFit	Lfa-1_RT_Ensemble
Wavelength	1.116		1.116		
Resolution range	46.04 - 1.71 (1.77 - 1.71)	46.04 - 1.71 (1.77 - 1.71)	44.90 - 1.80 (1.86 - 1.80)	44.90 - 1.80 (1.86 - 1.80)	44.90 - 1.80 (1.86 - 1.80)
Space group	P 21 21 21	P 21 21 21	P 32 2 1	P 32 2 1	P 32 2 1
Unit cell	37.57 51.35 103.92 90 90 90	37.57 51.35 103.92 90 90 90	104.90 104.90 51.64 90 90 120	104.90 104.90 51.64 90 90 120	104.90 104.90 51.64 90 90 120
Total reflections			182242 (13598)		
Unique reflections	22162 (2126)	22162 (2126)	30598 (2961)	30569 (2961)	30569 (2961)
Multiplicity			6.0 (4.5)		
Completeness (%)	98.74 (97.43)	98.74 (97.43)	91.90 (74.72)	91.90 (74.72)	91.89 (74.72)
Mean I/sigma(I)			10.55 (0.64)		
Wilson B-factor	23.34	23.34	22.67	22.67	22.67
R-merge			0.1205 (1.775)		
R-meas			0.1318 (2.007)		
R-pim			0.05274 (0.92)		
CC1/2			0.995 (0.296)		
Reflections used in refinement	22162 (2124)	22162 (2124)	0.999 (0.676)		
Reflections used for R-free	1113 (107)	1113 (107)	30569 (2240)	30569 (2240)	30569 (2240)
R-work	0.1729 (0.3213)	0.1372 (0.2820)	1827 (153)	1827 (153)	1826 (153)
R-free	0.2104 (0.3749)	0.1956 (0.3675)	0.1702 (0.3336)	0.1433 (0.2775)	0.1746 (0.3323)
Number of non-hydrogen atoms	1664	90358	0.2033 (0.3228)	0.2032 (0.3505)	0.2141 (0.3565)
macromolecules	1560	87360	0.960 (0.659)		
ligands	2	112	0.929 (0.712)		
solvent	102	2886	1674	2745	57336
Protein residues	193	193	1511	2582	54396
RMS(bonds)	0.004	0.015	1	1	36
RMS(angles)	0.73	1.69	162	162	2904
Ramachandran favored (%)	97.91	90.05	188	188	188

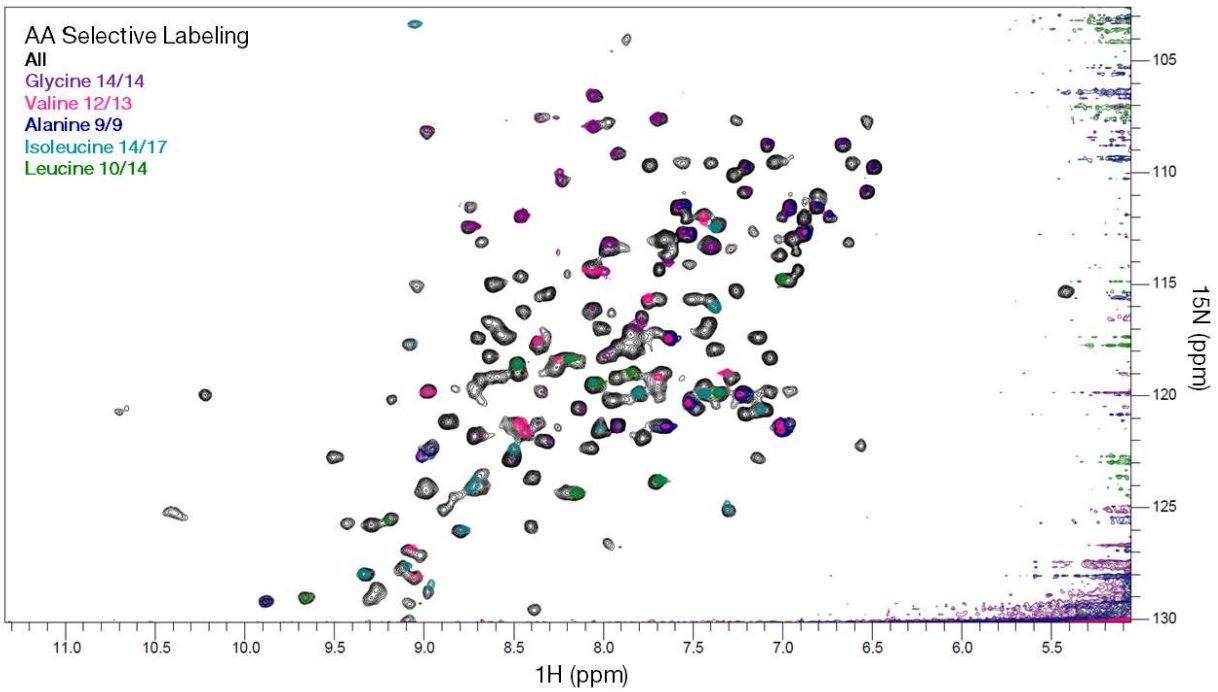
Ramachandran allowed (%)	2.09	6.81	0.011	0.006	0.018
Ramachandran outliers (%)	0.00	3.14	1.29	0.97	1.74
Rotamer outliers (%)	1.16	13.51			
Clashscore	3.51	0.00			
Average B-factor	28.10	22.83			
macromolecules	27.49	22.81			
ligands	38.55	26.08			
solvent	37.20	23.52			



Appendix Figure 9. **X-ray models of the Mac-1 and Lfa-1 I-domains color scaled to β-factor** (low = blue, high = green). Lfa-1 α7 helix is more flexible than Mac-1. structure has an extra five residues resolved at the end of the α7.



Appendix Figure 10. Q325F retains I-domain PPI affinity. WT Mac-1 I-domain (top row) has similar affinity for canonical Mac-1 protein partner C3dg as Q325F I-domain (bottom row), I-domain detected by SPR



Appendic Figure 11. Selective 15N- G/V/A/I/L labeling to aid in resonance assignment. WT Mac-1 was expressed as reported in the Methods, but in one liter of M9 media containing a mix of twenty 14N-amino acids and one 15N-amino acid, either glycine, valine, alanine, isoleucine or leucine. Purified sample concentrations ranged from 0.7 mM to 1.5 mM. HSQC spectra for the selectively labeled samples is overlaid on a U-15N HSQC, and the number of observed versus expected peaks is noted.

Appendix Table 5: Mac-1 1H/15N Backbone Resonance Chemical Shift Assingments

Sequence Number	Amino Acid	H (ppm)	N (ppm)	CA (ppm)	CB (ppm)
149	Ser	7.87	108.82	55.34	63,73
150	Asp	8.61	123.77	58.26	31.35
151	Ile	9.3	127.69	59.9	38.9
152	Ala	9.82	128.74	-	-
162	Ile	8.77	124.28	60.64	35.18
172	Phe	9.44	122.35	56.9	40.86
176	Val	8.39	121.57	55.53	31.73
185	Thr	7.63	118.57	62.80	68.99
186	Leu	8.1	124	53.17	45.32
191	Gln	8.29	121	-	31.04
192	Tyr	8.28	119.52	54.87	41.25
193	Ser	7.59	113.58	65.12	58.64
195	Glu	6.91	119.35	53.41	32.49
196	Phe	8.38	115.95	57.01	-
197	Arg	8.94	123.55	53.79	32.44

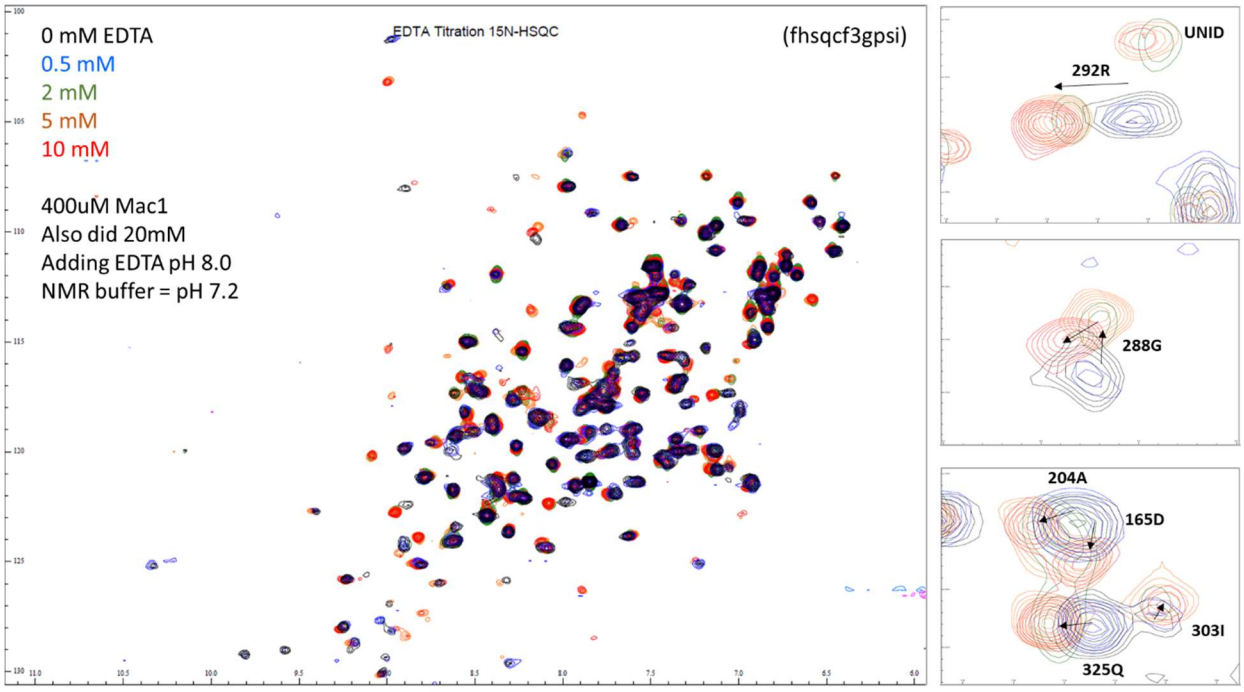
198	Ile	9.04	127.32	61.75	36.03
200	Phe	6.99	109.15	38.56	55.85
201	Thr	-	114.85	59.59	71.31
202	Phe	8.33	125.51	55.54	27.70
213	Ser	7.35	113.11	60.14	-
215	Val	7.68	115.27	63.03	31.59
222	Leu	9.15	125.28	-	-
223	Gly	8.68	111.27	-	
225	Thr	7.97	114.09	-	64.29
227	Thr	7.13	112.37	68.78	-
228	Ala	7.65	121.04	55.65	16.31
230	Gly	8.29	107.19	-	-
231	Ile	7.96	121.33	-	-
236	Arg	8.61	112.85	58.64	30.23
237	Glu	8.4	114.41	56.94	29.55
238	Leu	7.78	118.6	57.6	-
240	Asn	7.03	117.98	52.53	39.98

241	Ile	8.9	128.25	63.74	37.14
242	Thr	8.65	117.08	64.9	68.06
244	Gly	8	106.29	45.53	-
245	Ala	6.97	120.83	51.4	19.84
246	Arg	10.33	124.8	55.73	-
266	Leu	8.43	118.24	52.92	42.15
267	Gly	8.91	107.8	42.77	-
269	Glu	10.14	119.56	58.24	26.83
270	Asp	7.21	114.84	55.76	41.81
271	Val	7.38	111.57	61.76	33.15
272	Ile	8.66	123.71	62.42	-
274	Glu	7.28	117.44	58.71	28.66
275	Ala	7.6	121.07	54.99	16.94
277	Arg	8.32	123.25	58.97	28.97
279	Gly	7.63	107.32	-	-
280	Val	7.63	119.28	58.75	29.44
281	Ile	7.32	115.89	-	37.1

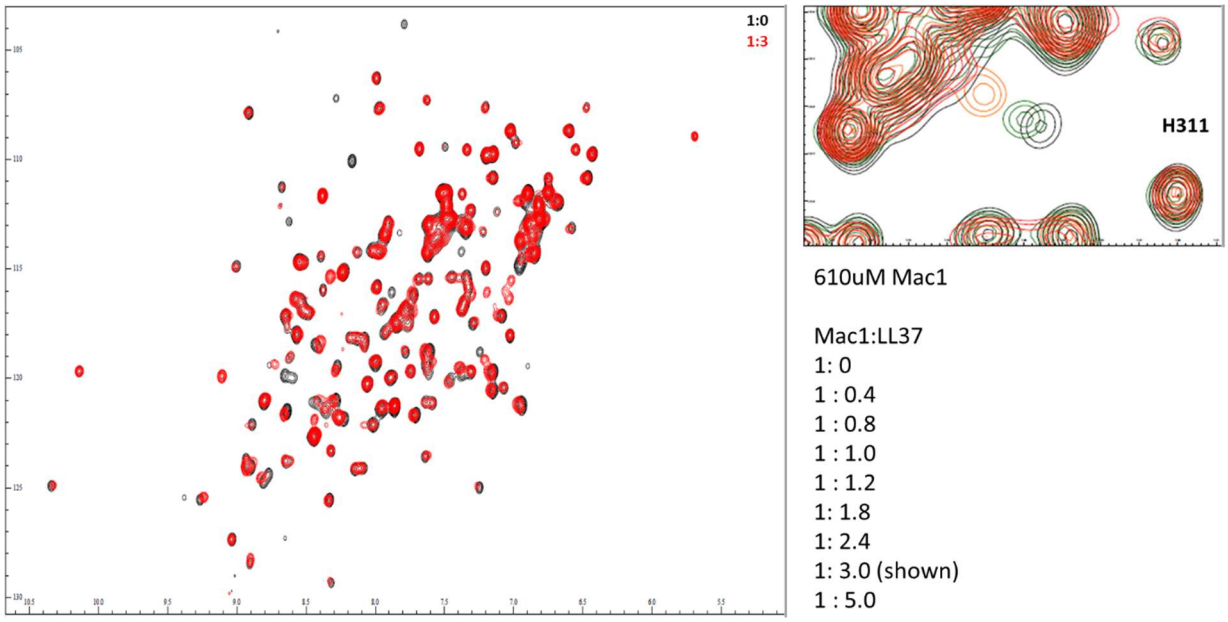
283	Tyr	8.41	118.57	58.83	31.26
284	Val	8.42	121.02	66.87	30.18
285	Ile	9.03	117.52	-	-
286	Gly	8.69	112.03	44.85	-
289	Asp	8.57	118	-	40.10
290	Ala	8.89	122.03	54.07	17.96
291	Phe	6.76	110.82	57.11	37.43
292	Arg	7.09	117.14	-	29.48
301	Asn	7.35	115.25	54.65	36.65
302	Thr	7.72	116.19	65.11	68.11
303	Ile	7.08	120.34	63.68	38.66
304	Ala	7.17	119.64	56.28	20.68
311	His	7.44	113.82	56.71	32.15
312	Val	7.23	118.96	61.21	32.77
313	Phe	9.25	125.41	55.6	41.82
315	Val	8.27	115.15	58.12	-
316	Asn	9.11	119.81	-	-

319	Glu	8.9	123.95	58.9	28.01
320	Ala	7.57	117.2	51.89	18.22
321	Leu	7.64	123.39	58.07	38.92
322	Lys	7.57	112.72	58.41	30.99
323	Thr	7.68	109.49	63.74	69.46
324	Ile	7.31	112.35	60.06	37.74
325	Gln	7.16	120.49	60.05	28.09
326	Asn	8.58	116.23	55.84	36.86
327	Gln	8.05	120.2	58.49	28.34
328	Leu	8	119.05	57.55	40.47
332	Ile	7.75	119.62	64.74	36.1
333	Phe	8.6	118.82	57.37	36.55
334	Ala	7.47	120.11	52.92	17.5
335	Ile	7.39	119.59	62.94	36.92
336	Glu	7.64	119.73	56.57	29.58
337	Gly	7.99	107.73	45	-
338	Thr	7.94	113.34	61.65	69.38

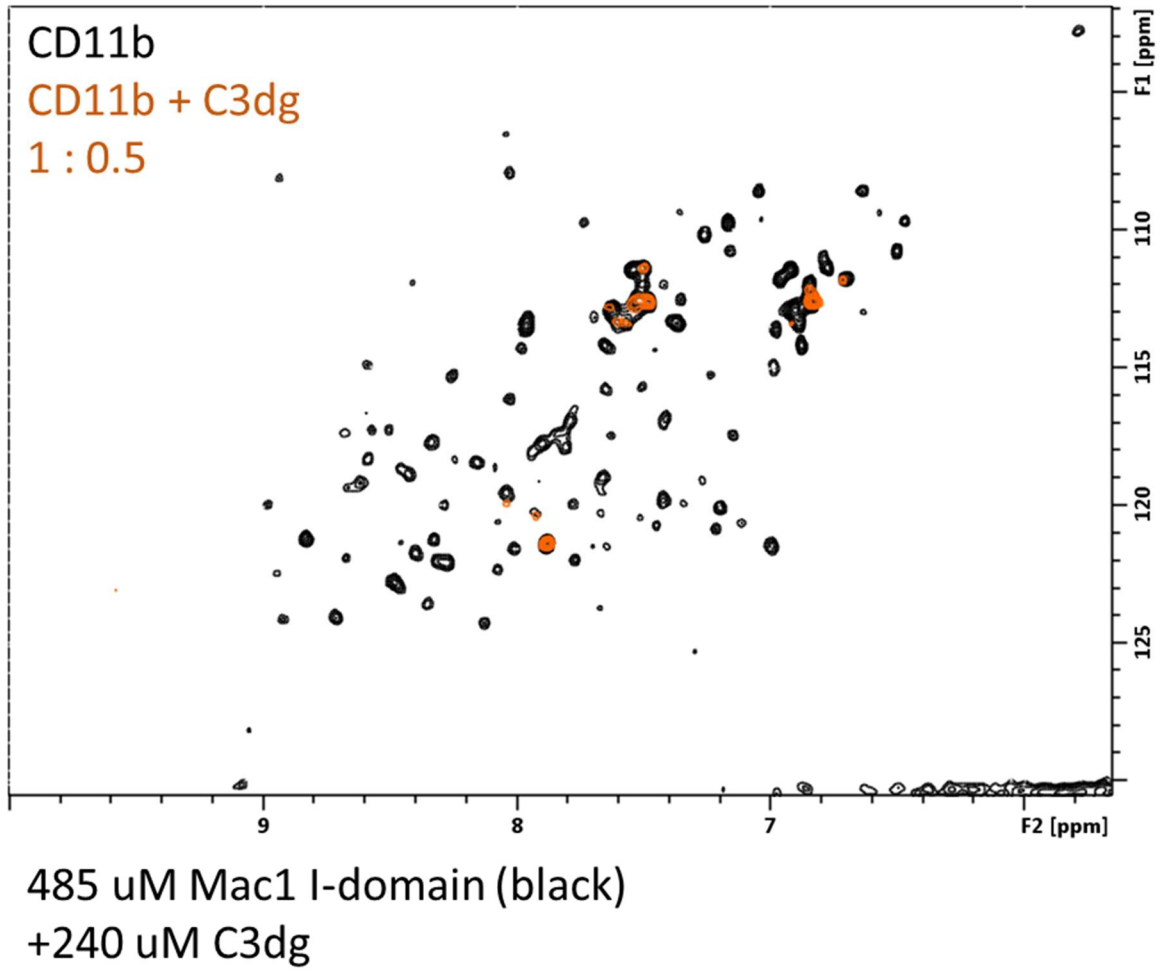
339	Gln	8.44	122.56	55.55	28.41
340	Thr	8.22	115.15	61.63	69.27
341	Gly	8.38	111.62	44.88	-
342	Ser	7.86	121.17	59.53	64.09



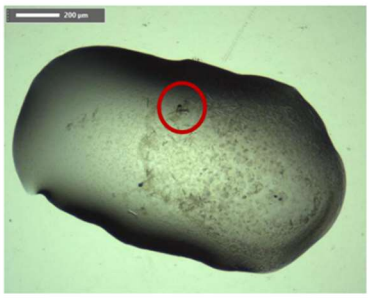
Appendix Figure 12. 1H/15N HSQC of Mac-1 I-domain with increasing concentrations of EDTA.



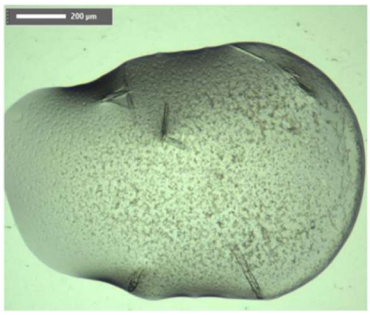
Appendix Figure 12. 1H/15N HSQC of Mac-1 I-domain with increasing concentrations of LL-37. The 8-step titration is noted on the right.



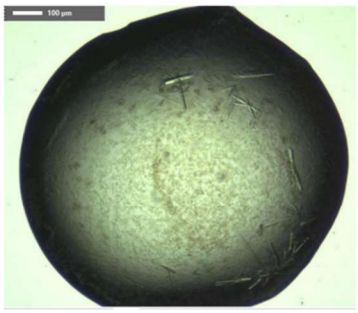
Appendix Figure 13. 1H/15N HSQC of Mac-1 (Cd11b) I-domain before and after addition of a half-equivalent of C3dg.



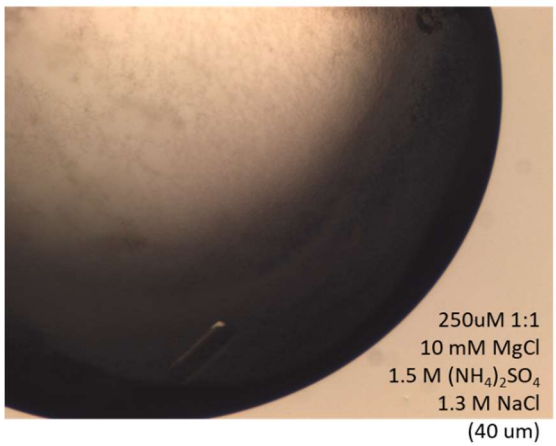
0.2 M (NH4)2SO4
0.1M Na Acetate pH 4.6
25% PEG MME 2000
Growth over 5 days

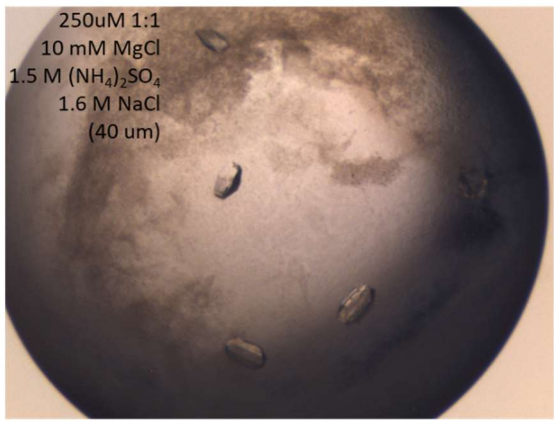


0.1M Na3 Citrate pH 5.5
20% PEG 3000
Growth over 21 days

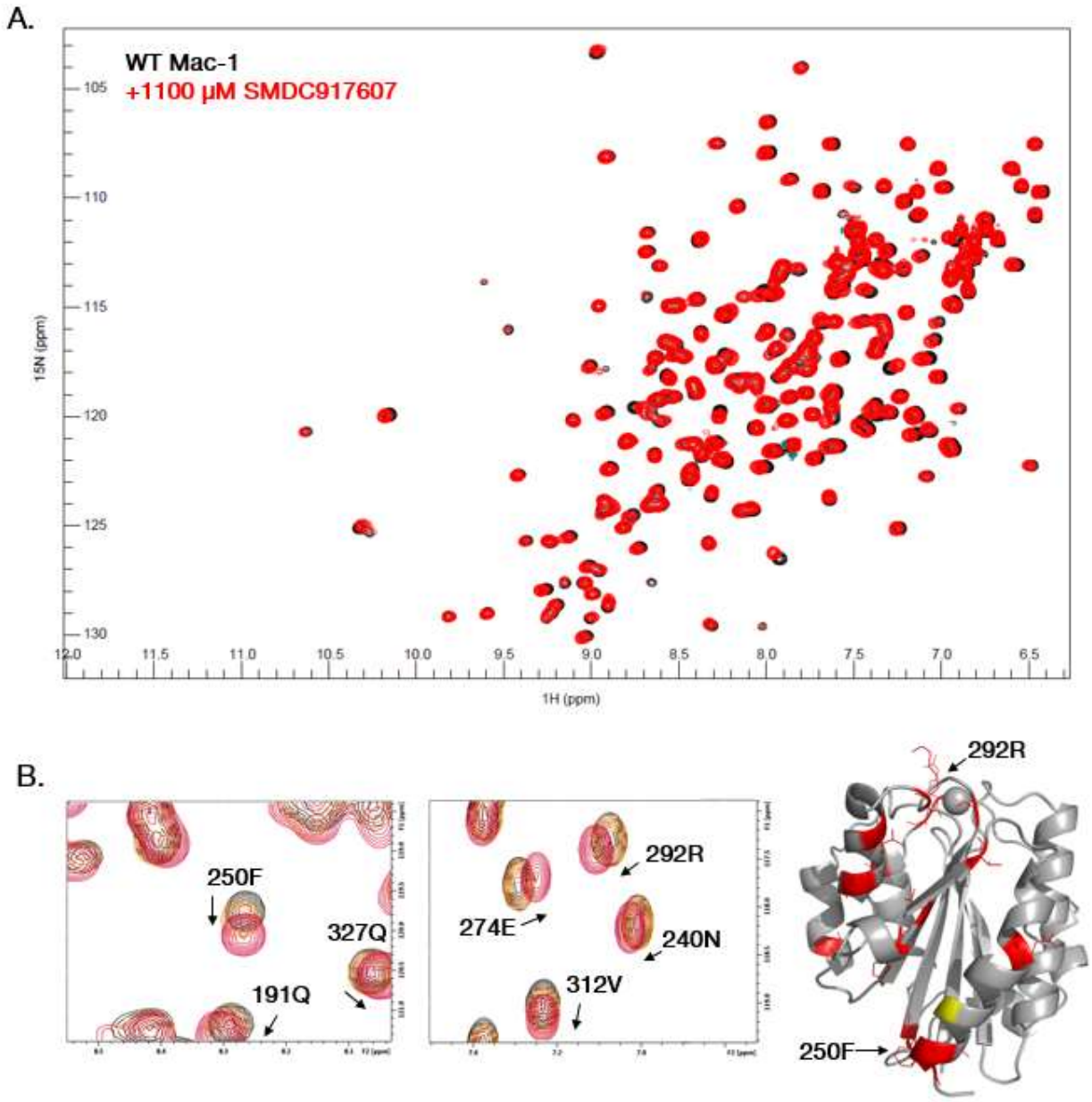


0.2 M (NH4)2SO4
0.1M Na Acetate pH 4.6
25% PEG 4000
Growth over 12 hours





Appendix Figure 14. Hits from Mac-1/γ-chain co-crystallization screening



Appendix Figure 15. WT Mac-1 Interacts with SMDC917607. (A) HSQC spectra WT Mac-1 before (black) and after (red) addition of SMDC917607. (B) Annotated 15N HSQC spectra for residues displaying CSP, mapped to the WT Mac-1 structure.

Publishing Agreement

It is the policy ofthe University to encourage the distribution ofall theses, dissertations, and manuscripts. Copies ofall UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best oftheir abilities, in perpetuity.

Please sign the following statentent:

I hereby grant permission to the Graduate Division ofthe University of California, San Francisco to release copies ofmy thesis, dissertation, or manuscript to the Campus Library to provide access andpreservation, in whole or in part, in perpetuity.

Author Signature

Date 3/29/18