Title

Defining Sec61 client sensitivity using substrate-selective Sec61 modulators

Author

Wenzell, Nicole

Publication Date

2023

Defining Sec61 client sensitivity using substrate-selective Sec61 modulators

by
Nicole Wenzell

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Jack Taunton
0FD9D71201FC43C...

Jack Taunton
Chair

DocuSigned by:
Jason Gestwicki
[redacted]E5...

Jason Gestwicki

DocuSigned by:
Martin Kampmann
03BA4012CC044AF...

Martin Kampmann

Committee Members

Acknowledgements

The successful completion of my doctoral dissertation was only possible through the support and help I had from so many people across my academic and personal spheres. First, I would like to thank my PhD advisor, Jack Taunton. I joined Jack's lab knowing I would grow as a scientist, and I believe the work presented here shows that he made that possible. I learned so much from my time in Jack's lab and I am forever grateful for his dedication to my scientific endeavors and growth. Jack gave me the opportunity to explore a scientific area I previously had no expertise in, and he was there every step along the way to advise, to help troubleshoot problems, and to provide support and encouragement. Again, thank you, Jack.

I was fortunate to work in the Taunton Lab with such supportive lab mates who were always there for advice when I needed it. Thank you especially to those I was closest with during my time in lab – Adolfo Cuesta, Tangpo Yang, Haoyuan Wang, Ying Chen, Keely Oltion, Jackie Weaver, Greg Craven, and Matt Lyons – these people are not only great scientists but are also wonderful people. Haoyuan mentored me as a rotation student and showed me the caliber of science done in the lab. Keely taught me most of what I know about molecular biology, cell culture, flow cytometry (the list could go on), and I don't know how I would have fared without her scientific guidance. Keely is also a great friend; she was there with me for most of my PhD, and we really leaned on each other to get through. Tangpo and Ying were valuable resources for my proteomics experiments and beyond.

I also want to give a thank you to the broader UCSF community who helped me along the way. Thank you to my thesis committee members, Jason Gestwicki and Martin Kampmann, for continual positive encouragement throughout my PhD journey. Their scientific advice and contributions to my project were incredibly helpful. Martin's lab was an important resource for help with next-generation sequencing experiments which, prior to my work, was an expertise outside the Taunton lab. In addition, Luke Gilbert and his lab were instrumental in providing expertise on how to build large pooled-cell libraries. Danny Conrad helped immensely in

teaching me how to run and analysis next-generation sequencing samples. Thank you to the Shokat Lab for sharing scientific ideas and resources, and for being friendly neighbors – our fifth-floor community made this journey that much more fun.

Finally, thank you to my family and friends. My support system outside of UCSF was so important to me throughout my PhD. Thank you to my mom, Karen, and my dad, David, for their unconditional love and support that I have felt my entire life. My parents made it clear to me I could do anything I put my mind to and without them none of this would be possible. Thank you to my sister, Emily, who is so special to me. Knowing you look up to me motivates me so much. Thank you, Mom, Dad, and Em.

Defining Sec61 client sensitivity using substrate-selective Sec61 modulators

Nicole Wenzell

Abstract

One third of the human proteome relies on the Sec61 translocon for proper folding and function. Each Sec61 client has a unique signal peptide or signal anchor, which plays an essential role in promoting its translocation into the endoplasmic reticulum. Cotransins are cyclic depsipeptides that inhibit the biogenesis of a subset of secreted and membrane proteins by preventing cotranslational translocation into the endoplasmic reticulum. Sensitivity to cotransins is determined by the protein's N-terminal signal peptide or signal anchor, which intercalates into the lateral gate of Sec61 to initiate translocation. However, it has been difficult to predict Sec61 client sensitivity to cotransins due to the large diversity of signal peptides and signal anchors and the lack of high throughput assays to effectively survey all predicted human Sec61 clients.

We developed a fluorescence-based reporter system with a library of 3880 human (3212) and mouse (668) signal peptides to interrogate which Sec61 clients are most affected by cotransin analogs. This pooled-cell screening platform allowed us to profile two cotransins with distinct effects on cancer cells, KZR-9873 and KZR-8445. We identified a total of 665 sensitive signal peptides, most of which were previously not known to be sensitive to cotransins. Among several validated targets, we discovered that the oncoprotein HER3, a coreceptor for HER2, is preferentially inhibited by the more selective cotransin, KZR-9873. Comparison of the human and mouse signal peptide paralogs within the library revealed a position-dependent role for Arg and Lys in conferring cotransin sensitivity. Our screening platform revealed distinct signal peptide sequence determinants that confer sensitivity to cotransins.

We used mass spectrometry to profile the effects of KZR-9873 and KZR-8445 on endogenous proteins in three cancer cell lines, CAL27, BxPC3 and SW48 cells. This approach allowed us to interrogate Sec61 client sensitivity beyond SP-containing proteins, since type II

and multi-spanning membrane proteins are understudied with respect to their potential cotransin sensitivity. Of the Sec61 clients identified across the three cell lines, there were 360 KZR-9873-sensitive and 659 KZR-8445-sensitive Sec61 clients. We identified a total of 51 and 81 type II membrane proteins that were sensitive to KZR-9873 and KZR-8445, respectively. In addition, we identified 34 and 45 multi-spanning N-cyt membrane proteins that were sensitive to KZR-9873 and KZR-8445, respectively. Our proteomic approach showed that cotransins can inhibit the biogenesis of proteins that utilize the lateral gate of Sec61 to initiate cotranslational membrane insertion.

Table of Contents

List of Figures

List of Tables

Chapter 1: Recent insights into signal peptide function and inhibition

Introduction

Part I: Sec61 inhibitor pharmacology and client selectivity

Secreted and membrane proteins rely on machinery in the endoplasmic reticulum (ER) for translocation. Most of these proteins are cotranslationally translocated into the ER membrane or ER lumen via insertion into the Sec61 translocon before they reach their final location in the cell[1]. Cotranslationally translocated proteins are directed to the ER by their hydrophobic signal peptide (SP) or transmembrane domain (TMD)[2]. These hydrophobic segments are recognized by Sec61 and insert into the ER membrane via the lateral gate of Sec61 (**Fig. 1.1**)[3,4]. More than 3400 unique SPs are predicted to mediate the translocation of diverse Sec61 clients, including secreted, type I transmembrane, GPI-anchored, and certain multi-transmembrane proteins.

Recent studies have elucidated the mechanism of action of a class of Sec61 inhibitors, cotransins, which are structurally related to the fungal natural product HUN-7293 (**Fig. 1.2**). HUN-7293 was identified in a screen for inhibitors of cell adhesion molecule expression. An unusual feature of HUN-7293 is that it inhibited the induced expression of three adhesion molecules, albeit with distinct IC_{50}s (vascular cell adhesion molecule 1 (VCAM1), intracellular adhesion molecule 1 (ICAM-1) and E-selectin)[5,6]. Initial studies on the mechanism of action showed that cotransins inhibit Sec61-mediated translocation of VCAM-1 in an SP-dependent manner[6,7]. Swapping the SP from VCAM-1 with 12 other SPs revealed that the SP derived from vascular endothelial growth factor (VEGF) was the only one that retained cotransin sensitivity; all other SP swaps conferred resistance[6–9]. Harant et al. systematically mutated the VCAM1 and VEGF SPs and found that single amino acid changes drastically modulate SP sensitivity to this class of Sec61 inhibitors. Mutations that increase hydrophobicity within the H-region of each SP resulted in 30-fold and 20-fold increases in resistance to cotransin for the VCAM-1 and VEGF SPs (**Table 1.1**)[8,9]. The helicity of the H-region is also important, as introduction of a proline (a helix breaking residue) sensitized the VEGF SP by 4-fold. These studies showed that

cotransins discriminate between Sec61 clients based on the specific amino acid sequences of their SPs. Direct evidence for cotransin binding to Sec61 was revealed by the use of a clickable photo-cotransin (CT7), which incorporates a diazirine ring into a leucine side chain[10]. CT7 photo-crosslinked directly to Sec61 and was competed by CT8, a photostable analog. Further mechanistic insights were revealed by Cys-crosslinking experiments between TNFa and Sec61a. These experiments showed that cotransin stabilizes a 'pre-integration' intermediate, characterized by the formation of crosslinks between engineered cysteines in the signal anchor of nascent TNFa and the cytosolic tip of the Sec61 lateral gate[11]. In addition, the cotransin-induced crosslinks occurred along one face of the TNFa transmembrane helix, suggesting that it is trapped in a specific orientation within the cytosolic vestibule of Sec61. Thus, upon cotransin treatment, the TNFa signal anchor cannot bind Sec61 in its preferred orientation – between the lateral gate helices and the lipid bilayer – presumably because this space is occupied by cotransins.

The current model for cotransin-mediated inhibition posits that cotransins compete with signal peptides and transmembrane domains for binding to the lateral gate. In support of this model, mutations in the TNFa signal anchor strongly affected cotransin sensitivity[11]. As with the SPs described above, introducing hydrophobic (leucine) residues in the middle of the TNFa transmembrane segment resulted in a right-shifted dose response curve, whereas a helix-breaking (proline) residue resulted in a left-shifted dose response curve. The cotransin IC_{50} dependence on the hydrophobicity and helicity of the SP/TMD suggests that SP/TMD and cotransin compete for binding to the lateral gate. In addition, cotransin resistance mutations in the Sec61 plug domain, which resides near the lateral gate, suggest that cotransins act as a molecular wedge that prevent polypeptide translocation by blocking displacement of the plug[11]. This competition model is not solely based on equilibrium partitioning of SPs and TMDs into the lipid phase[12], but instead may involve the disruption of dynamic interactions between the SP/TMD and the lateral gate of Sec61 (i.e., kinetic competition). However, the formation of a

cotransin-stabilized SP/Sec61 complex, as discussed above, suggests that other factors, including direct contacts between cotransins and SPs, may also contribute to selectivity. This concept is explored more extensively in Chapter 2.

Cyclotriazadisulonamide (CADA), a synthetic macrocycle which is structurally unrelated to cotransins, was also found to be a client-selective Sec61 inhibitor (**Fig. 1.2**). CADA was first discovered in a screen for antiviral compounds and was subsequently shown to reduce the expression of CD4, the primary entry receptor for HIV in T cells[13,14]. Structure activity relationship campaigns successfully increased the potency of CADA analogs against CD4 from the micromolar to nanomolar range[15]. Analogous to cotransins, the potency of CADA toward CD4 biogenesis strictly depends on the amino acid sequence of the CD4 SP. When the SP from human CD4 was swapped with the SP from mouse CD4, CD4 expression levels were unaffected by CADA treatment[16]. In addition, alanine mutational scanning of the human CD4 SP revealed specific hydrophobic and polar amino acids that determine of CADA sensitivity[17].

Several other natural products have been shown to block secretory and membrane protein biogenesis by targeting Sec61. Most of these compounds lack Sec61 client selectivity. Examples include mycolactone and ipomoeassin F (Ipom-F), which were isolated from mycobacteria and the plant *Ipomoea squamosa*, respectively (**Fig. 1.2**)[18–20]. When tested against a panel of 10 Sec61 clients in an *in vitro* translocation assay, Ipom-F inhibited the translocation of 9, suggesting broad inhibitory activity[21]. Furthermore, Ipom-F is highly cytotoxic, exhibiting IC_{50} values in the low nanomolar range[21]. Mycolactone is produced by the human pathogen *Mycobacterium ulcerans* and is essential for pathogenesis. Similar to Ipom-F, the immunosuppressive effects of mycolactone are attributed to nonselective Sec61 client inhibition which prevents the secretion of cytokines that are important for the immune response[22]. These Sec61 inhibitors are structurally distinct, but shared resistance mutations in the Sec61 plug domain (S82P and R66I) suggest that they bind to a similar site, as discussed in more detail below[11,21,23]. Mycolactone and Ipom-F compete with the photo-cotransin CT7, further suggesting

a shared binding site among selective and nonselective inhibitors[21,23,24]. The nonselective compounds may have a higher intrinsic affinity (faster on-rate, slower off-rate) for apo-Sec61 or ribosome-engaged Sec61 relative to the more selective inhibitors, given their nanomolar inhibitory potency toward essentially all Sec61 clients. However, to our knowledge neither the binding affinities nor kinetics have been reported for any Sec61 inhibitor. The membrane-embedded nature of the heterotrimeric Sec61 complex makes these biophysical measurements difficult. Nevertheless, such experiments would greatly add to our knowledge of the mechanism of Sec61 inhibition by these structurally distinct small molecules.

Part II: Recent structural insights into Sec61 inhibition mechanism

The crystal structure of the archaeal SecYEβ complex provided a major breakthrough in our understanding of the structural basis of translocation[25]. This structure revealed that SecY (homologous to eukaryotic Sec61α) forms a 10 transmembrane helix bundle with two halves resembling a clamshell (**Fig. 1.3a**). The narrowest part of the channel is formed by a pore ring comprising hydrophobic residues, which is occluded by the alpha-helical plug domain. The lateral gate is formed by the interface of transmembrane helices H2/3 and H7/8, which when opened allows the SP to access the lipid bilayer. As an SP inserts into the lateral gate, this event allosterically releases the plug domain and opens the pore to allow translocation of lumenal and extracellular domains.

In addition to the SecY crystal structure which shows a closed channel, a series of recent cryo-EM structures of eurkaryotic Sec61 engaged in different stages of translocation have provided mechanistic insights into SP-dependent gating[26,27]. These structures represent the progression of translocation and show the primed and engaged states of Sec61 (**Fig. 1.3b-c**). In the primed state, the ribosome interacts with the cytosolic loops of Sec61 forcing a conformation that weakens the interactions between the lateral gate helices and shifts the

position of helix h2. However, the pore and plug residues do not shift relative to the closed state. Further along the translocation process, the structure of the engaged state of Sec61 captures the bound pPL signal peptide and further reveals how the SP intercalates between the lateral gate helices. The lateral gate of the SP-bound structure is partially open within the detergent micelle, a conformation that would expose the hydrophobic SP to the lipid environment of the membrane. The opening of the engaged channel is asymmetric; the cytosolic end of the lateral gate remains closed and the lumenal end parts by ~15 Å. The SP has access to the lipid bilayer as space is created between H2 and H7 (**Fig. 1.2d**). The position of the pPL SP in the channel at this stage overlays completely with the position of H2 in the closed state, as if the SP takes the place of H2 (the SP even has the same number of helical turns) (**Fig. 1.2e**). Furthermore, the pore ring residues and the plug domain are displaced, allowing for unobstructed access to the ER lumen. Comparison of the different states of Sec61 during translocation revealed the stepwise mechanism by which hydrophobic SPs insert into Sec61, involving the displacement of the lateral gate, pore, and plug residues.

The binding mechanism of Sec61 inhibitors was recently revealed by cryo-EM structures from Itskanov et al.[28] (7 structures) and Rehan et al. (1 structure)[29]. Collectively, these studies present the structures of 8 distinct inhibitors bound to Sec61. The inhibitors bind to a common pocket formed by the channel with a partially open lateral gate, whereas the plug domain remained closed. The structures enable a comparison of the binding modes of CP2 (cotransin analog), CADA, mycolactone, Ipom-F (**Fig. 1.4a-d**)[28], and a structurally distinct cotransin analog, KZR-8445 (**Fig. 1.4e**)[29]. Each inhibitor is within 4 Å of N300 in the lateral gate H7. Hydrogen bond acceptor and/or donor groups on KZR-8445, CP2, CADA and Ipom-F are within 3.3 Å of the N300 side change amide, suggesting the formation of a hydrogen bond. All but one inhibitor (KZR-8445) also form a hydrogen bond with Q127 of lateral gate H3. The apparent absence of a hydrogen bond to Q127 in the KZR-8445/Sec61 model may be due to the lower resolution of the electron density map in this region, relative to the high-resolution mpa of CP2

bound to Sec61. Nevertheless, mutation of Q127 to alanine had no effect on KZR-8445 sensitivity, whereas this mutation conferred resistance to the related cotransin, CP2. Sec61 residues N300, Q127, and T86 were previously identified as the lateral gate “polar cluster”, which forms a network of interhelix hydrogen bonds that regulate translocation through Sec61[30]. Interactions with these residues, especially N300, may help stabilize the inhibitor-bound states, preventing complete opening of the channel and inhibitor displacement by SPs. The importance of N300 and Q127 in the mechanism of action of Sec61 inhibitors was underscored by mutational analysis. N300A mutations abrogated sensitivity to all inhibitors, whereas Q127A abrogated sensitivity to all inhibitors except KZR-8445. In addition to binding the lateral gate helices, each inhibitor directly contacts the plug domain, which must move or unfold to allow cotranslational translocation. Hence, the inhibitors prevent Sec61 clients from accessing the ER lumen by maintaining the plug domain in a closed conformation.

The cryo-EM structures also provide insights into how cotransins achieve greater selectivity than mycolactone and Ipom-F, despite binding to the same site on Sec61. In each structure, the lateral gate is open to varying degrees depending on the inhibitor, providing further evidence that the Sec61 channel is highly dynamic and conformationally flexible. The differences in Sec61 conformation among the different inhibitors are most pronounced in the position of H2 (**Fig. 1.5a**). Measuring the distance between T86 on H2 and A295 on H7 shows that the width of the lateral gate roughly correlates with the selectivity of the inhibitor. This distance is greatest for the cotransins, CP2 (15.2 Å) and KZR-8445 (15.0 Å), and shortest for Ipom-F (8.5 Å). The more open lateral gate in the cotransin-bound structure suggests a potential mechanism by which Sec61 clients are able to more efficiently promote the dissociation of cotransins relative to Ipom-F and other less selective inhibitors. In this sense, hydrophobic SPs may act as a wedge by engaging the lipid-exposed region of the lateral gate, accelerating the dissociation of the bound inhibitor and release of the lumenal plug.

Compared to the natural product inhibitors, CADA showed a less extensive binding interface with Sec61 overall, which may explain its increased selectivity and lower potency (micromolar) (**Fig. 1.4c, Fig. 1.5b**). Although structurally distinct, CADA and KZR-8445 have a related binding mode, in that one aromatic side chain of CADA overlays partially with the aromatic R5 group of KZR-8445 but extends into the channel to a lesser extent (**Fig. 1.5b**). Conversely, the nonselective inhibitors mycolactone and Ipom-F have extended chains that penetrate deep into the channel toward the cytosolic vestibule and occupy a more substantial part of the channel's cytosolic funnel (**Fig. 1.5c-d**). These extensions may enhance their binding affinity and prevent a greater number of SPs from reaching the lateral gate. For reasons that are unclear, two other published cryo-EM reconstructions reached very different conclusions about the binding sites and orientations of mycolactone[31] and a CADA analog[32], as compared to the high-resolution structures published by Itskanov et al.[28]

Rehan et al. compared the structures of Sec61 bound to KZR-8445 vs. two signal peptides (α-Factor and pPL) and observed that both SPs overlap with the binding site of KZR-8445, but to different extents (**Fig. 1.6a-b**)[29]. Because the α-Factor SP binding site only partially overlapped with KZR-8445 in the vicinity of its large R5 side chain, a distinct cotransin analog, KZR-9508, was designed with a smaller side chain at R5. R5 of KZR-8445 is a *N*-bromobenzyl tryptophan side chain, whereas R5 of KZR-9508 is a smaller *N*-ethyl tryptophan (the only difference between the two cotransin analogs). This reduction in size at R5 increased selectivity against a panel of 5 SPs. KZR-8445 strongly inhibited translocation of 4 SPs in the panel whereas KZR-9508 strongly inhibited translocation of only 1 SP. Rehan et al. did not solve the structure of KZR-9508 bound to Sec61. However, CP2 in the structure solved by Itskantov et al. has an *N*-MeO tryptophan which is similar in size to the *N*-ethyl group of KZR-9508[28,29]. The overlay of KZR-8445 and CP2 bound to Sec61 shows that the *N*-MeO tryptophan does not

extend as far into the cytosolic vestibule and creates more space for a signal peptide to enter and displace the bound cotransin (**Fig. 1.6c**).

Discussion

The recent cryo-EM structures of apo Sec61, engaged Sec61, and inhibited Sec61 advance our understanding of how signal peptides gate the translocation channel, as well as the mechanism of action of structurally distinct Sec61 inhibitors[25–29]. Of great interest to us is the most selective class of Sec61 inhibitors, the cotransins, due to their ability to discriminate among Sec61 clients based on the primary amino acid sequence of their signal peptide (or signal anchor). The structures, along with previous Cys crosslinking experiments, support a model for the mechanism of action of cotransins in which SPs/TMDs and cotransins compete for binding to the lateral gate of Sec61[11]. Cotransins achieve selective inhibition by out-competing SPs that presumably engage Sec61 with lower affinity and/or slower kinetics, due in part to lower hydrophobicity and alpha-helical propensity. Research on the determinants of SP sensitivity has provided insights into how to make a given Sec61 client more sensitive or resistant to cotransins by reducing or increasing hydrophobicity/helicity, respectively[6–9]. However, we still cannot predict the sensitivity of individual Sec61 clients based on the primary amino acid sequence of their SPs. We describe our efforts to address this challenge in Chapter 2.

Figures

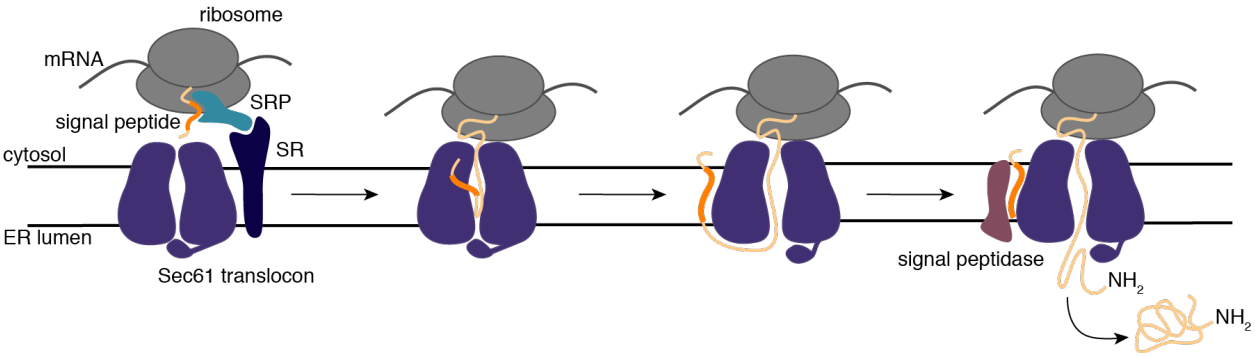


Figure 1.1: Cotranslational translocation via the Sec61 translocon.

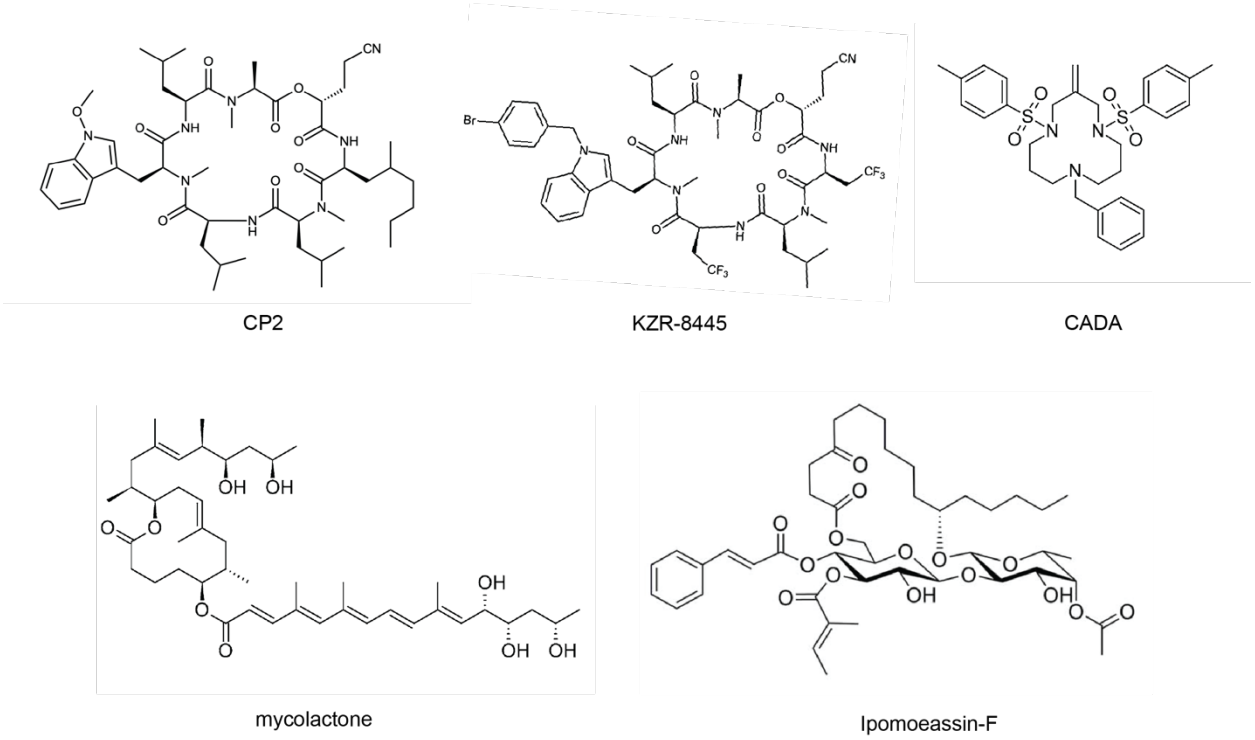


Figure 1.2: Sec61 inhibitor chemical structures.

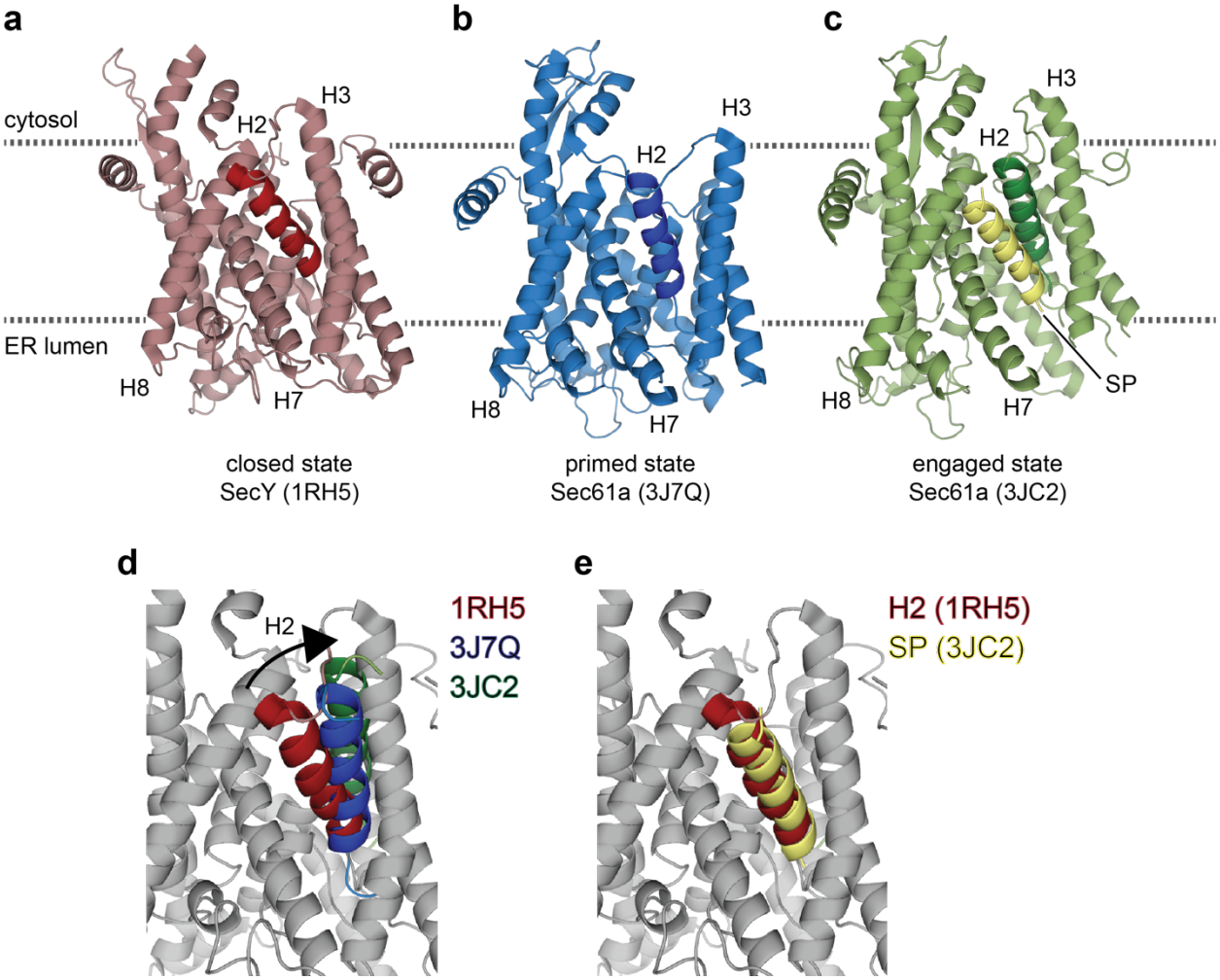


Figure 1.3: Conformational changes in Sec61 during signal peptide translocation. **a**, Overview of SecY in the closed state from the archaeal crystal structure. **b**, Overview of Sec61a in the primed state from the mammalian cryo-EM structure. **c**, Overview of Sec61a in the engaged state with the pPL signal peptide bound from the mammalian cryo-EM structure. **d**, Comparison of the movement of the lateral gate helix, H2, from (**a-c**). **e**, Comparison of the position of H2 in the closed state and the position of the SP in the engaged state of Sec61. PDB codes are indicated.

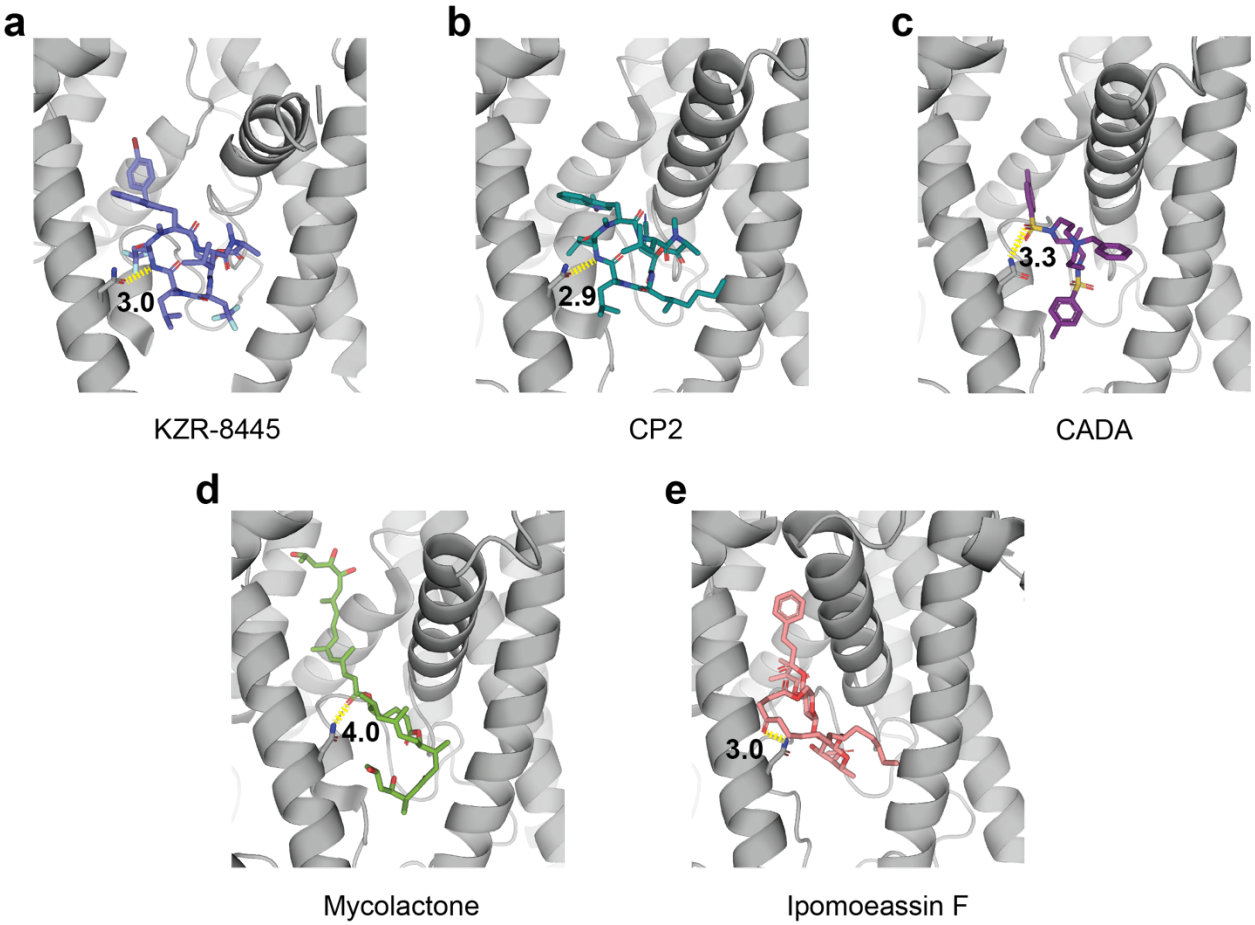


Figure 1.4: Sec61 inhibitors share a common binding site.
a-e, Views of the binding site for the indicated Sec61 inhibitors and the common hydrogen bond formed with N300 of Sec61a. PDB codes are: KZR-8445 (7ZL3), CP2 (8DNX), CADA (8DO2), mycolactone (8DO0), and Ipomoeassin F (8DO1).

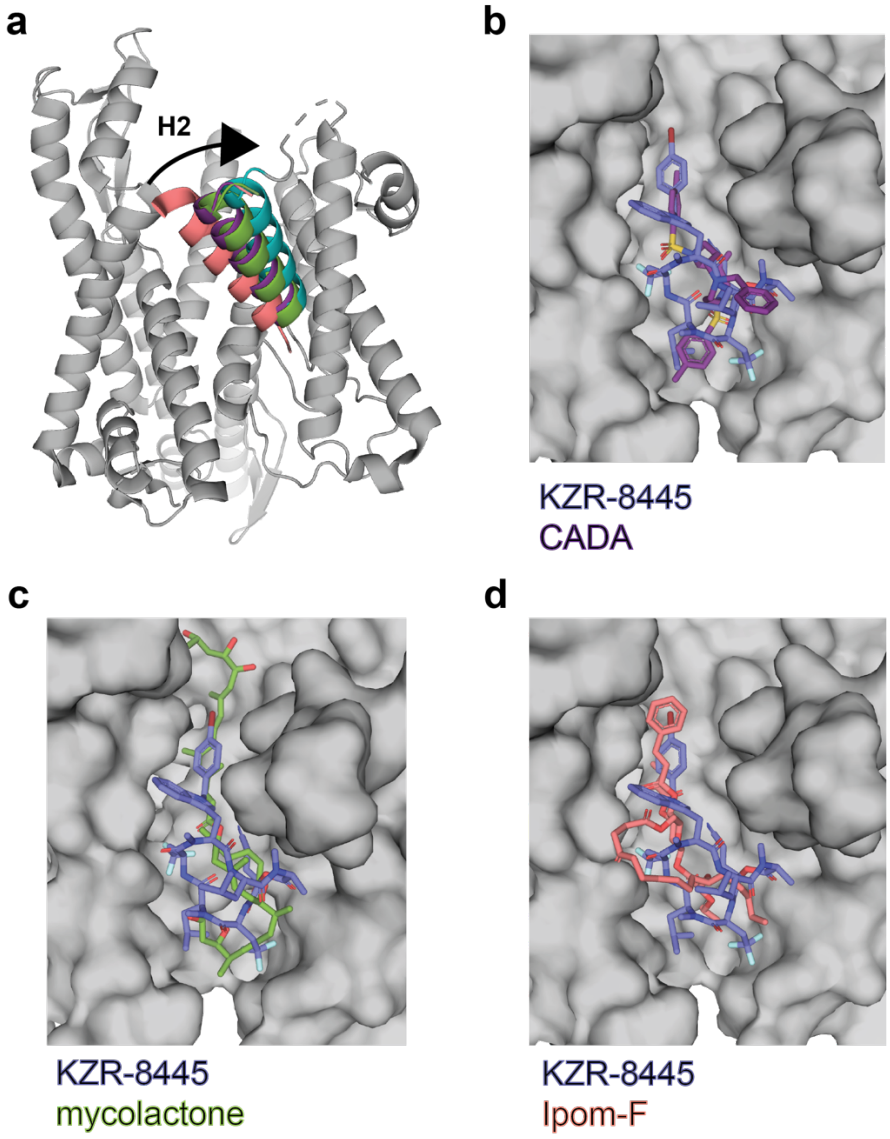


Figure 1.5: Sec61 inhibitors occupy distinct regions of the cytosolic vestibule.
a, Comparison of the Sec61 lateral gate helix, H2, in the structures bound to CP2 (teal), CADA (purple), mycolactone (green) and Ipom-F (salmon). **b-d**, Overlay of the indicated Sec61 inhibitors relative to KZR-8445-bound Sec61.

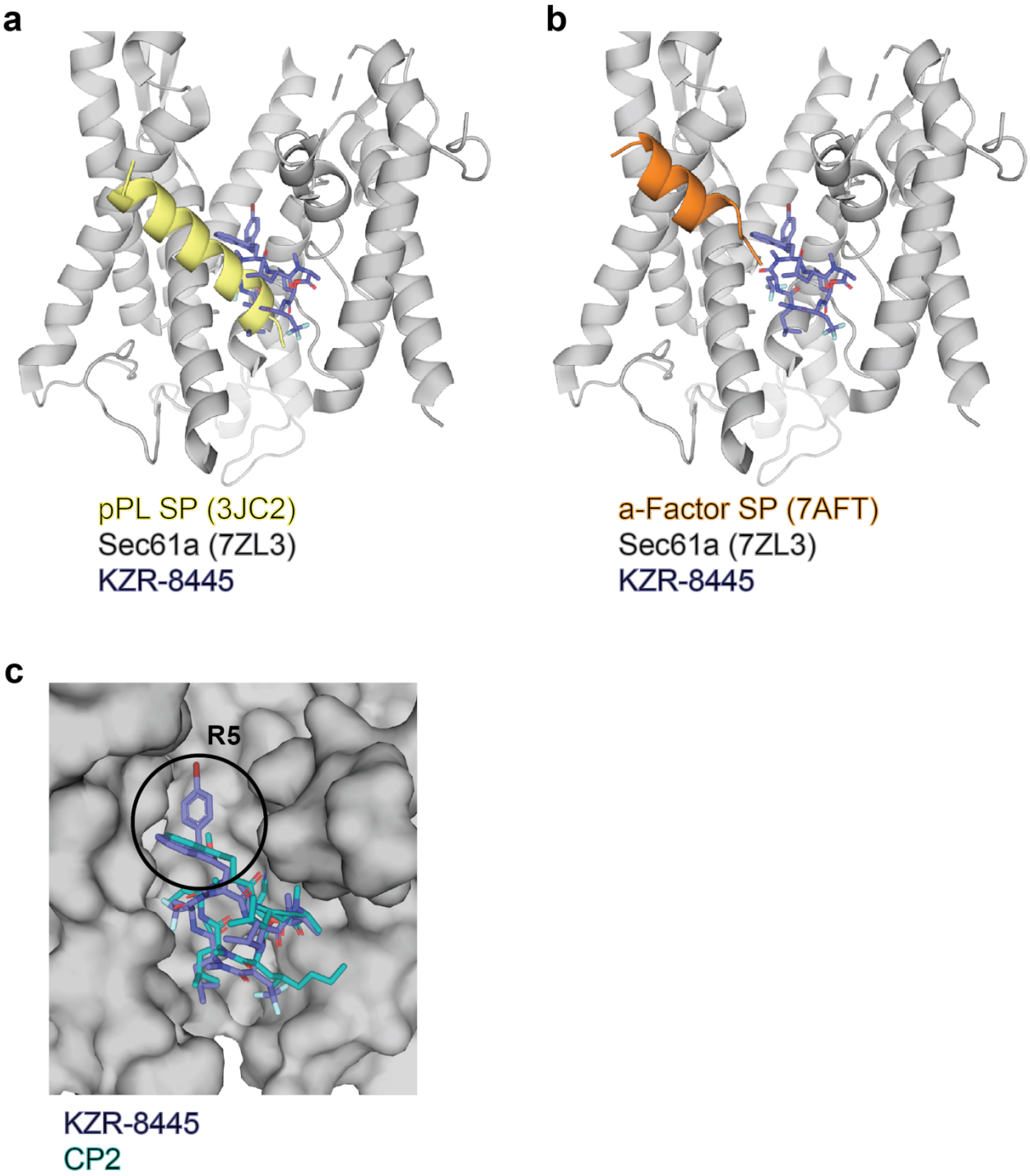


Figure 1.6: The shared Sec61 binding site for cotransins and signal peptides.
a-b, Comparison of the pPL SP (a) and α-Factor SP (b) binding site with the KZR-8445 (light purple) binding site in Sec61a. **c**, Overlay of the two cotransins, KZR-8445 (light purple) and CP2 (teal).

Tables

Table 1.1: The effects of VCAM1 and VEGF SP mutations on the sensitivity to the cotransin CAM-741.

SP name	SP sequence	IC_{50} (nM)
VCAM1 delta2-10 wt	MASNILWIMFAASQA	9
VCAM1 delta2-10 W16L, M18I	MASNIL**L**I**I**FAASQA	272
VEGF WT	MNFLLSWVHWSLALLLYLHHAKWSQA	125
VEGF A13P	MNFLLSWVHWSL**P**LLLYLHHAKWSQA	32
VEGF A13L	MNFLLSWVHWSL**L**LLLYLHHAKWSQA	2670

Chapter 2: Global signal peptide profiling reveals principles of client-selective Sec61 inhibition

Abstract

Cotransins target the Sec61 translocon and inhibit the biogenesis of an undefined subset of secretory and membrane proteins. Remarkably, cotransin inhibition depends on the unique signal peptide of each Sec61 client, which is required for cotranslational translocation into the endoplasmic reticulum. It remains unknown how a signal peptide's amino acid sequence and biophysical properties confer sensitivity to structurally distinct cotransins. Here we describe a fluorescence-based, pooled-cell screening platform to interrogate nearly all human signal peptides in parallel. We profiled two cotransins with distinct effects on cancer cells and discovered a small subset of signal peptides, including the oncoprotein HER3, with increased sensitivity to the more selective cotransin, KZR-9873. By comparing divergent mouse and human orthologs, we unveiled a position-dependent effect of arginine on signal peptide sensitivity. Our multiplexed profiling platform reveals how cotransins can exploit subtle sequence differences to achieve signal peptide discrimination.

Introduction

Proper folding, localization, and function of most secreted and membrane proteins requires translocation into the endoplasmic reticulum (ER)[1]. When an N-terminal signal peptide emerges from the ribosome, it binds the signal recognition particle (SRP) and subsequently engages the multi-subunit Sec61 translocon in the ER membrane[2]. The nascent signal peptide binds directly to the lateral gate of Sec61α, which allosterically releases the lumenal plug helix and opens the translocation pore[3,4]. This signal peptide-induced conformational change in Sec61 results in the translocation of extracellular and lumenal protein domains into the ER lumen and the insertion of transmembrane domains into the ER membrane; meanwhile, the signal peptide is cotranslationally cleaved by signal peptidase[33,34]. There are >6000 predicted Sec61 clients in the human proteome, of which >3400 have predicted signal peptides, with the

remainder engaging Sec61 via one or more transmembrane segments. All classically secreted, 'type I' transmembrane and GPI-anchored proteins, along with a large subset of multi-transmembrane proteins are thought to engage Sec61 and initiate translocation via a cleavable N-terminal signal peptide (SP).

Each SP is unique, usually spanning the first 15-45 residues and comprising diverse amino acid sequences. Despite this sequence diversity, SPs share a common architecture: a polar, often positively charged N-terminal region, a critical hydrophobic region of 6-15 residues, and a polar C-terminal region containing the signal peptidase cleavage site[35]. Based on this architecture, SPs can be predicted from primary amino acid sequences using sophisticated bioinformatic algorithms[36,37]. A cryo-EM structure of mammalian Sec61 engaged by the prolactin SP reveals that the central hydrophobic region (H-region), which plays an essential role in promoting translocation, forms an alpha-helix and intercalates between the lateral gate helices TM2 and TM7[27]. Although all SPs are thought to allosterically release the lumenal plug helix via interactions with the Sec61 lateral gate, the kinetics and mechanistic details of this process are mostly unknown and likely differ among the thousands of diverse SPs. Indeed, sequence variation among SPs can affect the folding, function, and trafficking of Sec61 clients by mechanisms that are poorly understood[38].

Cotransin cyclic heptadepsipeptides[6,7], including KZR-8445[29,39], bind directly to the Sec61 lateral gate and inhibit the translocation of an unknown subset of Sec61 clients[10,40], resulting in their proteasome-dependent degradation in the cytosol. The cotransin sensitivity of a given Sec61 client is determined primarily by its SP13,15. Mutagenesis and SP swapping experiments have revealed a weak inverse correlation between overall hydrophobicity and cotransin sensitivity[8,9,11], consistent with a model in which cotransins compete with SPs for the same lipid-exposed binding site on Sec61[28,29]. Nevertheless, the complete set of cotransin-sensitive SPs, the biophysical and sequence determinants of SP sensitivity, and the mechanisms underlying differential SP inhibition by cotransin variants are all unknown.

In this study, we develop a novel platform for interrogating signal peptides on a global scale. Our platform leverages fluorescence-activated cell sorting (FACS) and massively parallel DNA sequencing to assess the effects of cotransins on a library of nearly all annotated human SPs fused to the mature region of the cell-surface protein PD1. Using this platform, we determined the selectivity profiles of two cotransin variants, KZR-9873 and KZR-8445, which exhibit distinct biological activities. These data revealed that KZR-9873, but not the more promiscuous cotransin KZR-8445, potently inhibits the biogenesis of the oncoprotein HER3. By analyzing a sublibrary of SPs comprising human and mouse orthologs, we discovered that a single Arg or Lys can dramatically enhance cotransin sensitivity. Remarkably, by changing the precise position of an Arg within the hydrophobic region, it is possible to increase sensitivity to KZR-8445 while simultaneously decreasing sensitivity to KZR-9873. These unexpected structure-function relationships argue against a simple competitive binding model and suggest that KZR-8445 and KZR-9873 may differentially trap SPs in a nonproductive configuration while bound to Sec61.

Results

Flow cytometry-based reporter for profiling cotransins

We sought to globally define the signal peptide (SP) selectivity of two cotransins, KZR-8445 and KZR-9873, which differ at three side chain positions (**Fig. 2.1a**). Notably, KZR-9873 has a smaller chlorobenzothiophene as compared to the bromobenzyl tryptophan of KZR-8445, and previous work suggested that a smaller side chain at this position leads to increased selectivity among four tested SPs[29]. We were further motivated to compare KZR-8445 and KZR-9873 based on their distinct antiproliferative effects on a small panel of cancer cell lines derived from diverse tumor types (**Fig. 2.1b**, **Fig. 2.2a**). Whereas KZR-9873 and KZR-8445 inhibited the proliferation of BxPC3, CAL27, and SW48 cells with similar potencies (IC_{50} < 100 nM), KZR-8445 was much more potent than KZR-9873 toward the other cell lines. Based on these results,

we hypothesized that low concentrations of KZR-9873 (below 100 nM) inhibit the expression of one or more Sec61 clients which are essential for growth in a subset of cancer cell lines, whereas KZR-8445 has broader effects on the Sec61 clientome.

To determine the landscape of cotransin-sensitive SPs, we designed a doxycycline-inducible reporter that encodes two proteins amenable to single-cell analysis by flow cytometry (**Fig. 2.1c**): (1) the cell-surface protein, PD1 (aa 24-288, lacking the native SP) fused to the SP of interest (containing the SP's native signal peptidase cleavage site, plus 6 aa of the mature region), and (2) cytosolic iRFP, an internal reference standard translated from the same mRNA via an internal ribosome entry sequence (IRES). Because each functional SP is cleaved cotranslationally by signal peptidase[33,34], the resulting mature form of cell-surface PD1 (plus 6 aa from the mature region of each protein of interest) should be nearly identical for every SP reporter.

We transduced K562-TET3G cells with a reporter bearing the SP from PD1 itself (PD1 SP-reporter) or the SP from prolactin (PRL SP-reporter). Treatment of cells with doxycycline (dox) for 24 h resulted in >100-fold increased expression of cell-surface PD1 (detected with a FITC-conjugated antibody) and cytosolic iRFP (**Fig. 2.1d**). We used flow cytometry to quantify the effect of cotransins on cell-surface PD1 levels relative to the Sec61-independent control, iRFP. Whereas the PRL SP-reporter (PRL-PD1) was unaffected by treatment with either KZR-9873 or KZR-8445, the PD1 SP-reporter (PD1-PD1) was inhibited by both cotransins in a dose-dependent manner. Notably, KZR-8445 was significantly more potent than KZR-9873 toward the PD1 SP-reporter, reducing cell-surface PD1 expression by 40% at 111 nM; by contrast, KZR-9873 had no effect at this concentration (**Fig. 2.1e**, **Fig. 2.2b**). Neither KZR-8445 nor KZR-9873 (up to 1 μM) affected iRFP expression, ruling out general inhibitory effects on transcription or translation (**Fig. 2.2c**). These results show that our cell-surface reporter can distinguish between sensitive and resistant SPs and reveal potency differences between the cotransin variants, KZR-9873 and KZR-8445.

Multiplexed global interrogation of human signal peptides

To expand the SP reporter assay to a massively parallel, multiplexed screening platform (**Fig. 2.3a**), we synthesized an oligonucleotide library encoding 3880 unique SPs. The library includes all 3212 predicted human SPs up to 50 amino acids in length (median: 23 aa), plus six amino acids C-terminal to the signal peptidase cleavage site, as annotated in the Uniprot database. In addition, 668 mouse SPs were included based on the established or likely therapeutic relevance of the corresponding human orthologs (e.g., targets of approved or investigational drugs). We reasoned that pairwise comparison of related, but not identical, human and mouse SPs would provide insights into the sequence determinants of cotransin sensitivity. These oligonucleotides were cloned into the SP reporter (**Fig. 2.1c**), and K562-TET3G cells were transduced with the resulting library of lentiviruses at a low multiplicity of infection. After dox induction, iRFP-positive cells were isolated by FACS and expanded to generate a stably transduced, pooled-cell library (**Fig. 2.3a**).

We developed a screening workflow to determine the sensitivity of each functional SP in the library to KZR-9873 and KZR-8445 (**Fig. 2.3a**). The pooled-cell library was treated for 24 h with dox and either DMSO, KZR-9873 (100 nM), or KZR-8445 (100 nM) in biological duplicates (**Fig. 2.4a**). Each of the six cell populations was sorted into high (top 33%) and low (bottom 33%) fluorescence populations based on the FITC/iRFP ratio (12 sorted fractions total) (**Fig. 2.4b**). After genomic DNA isolation and PCR amplification, SP frequencies in each sorted cell population were determined by Illumina sequencing. Read counts for each SP showed excellent reproducibility between biological replicates across all cell populations (**Fig. 2.4c**). To quantify the effects of KZR-9873 or KZR-8445, we used the DESeq2 package (see Methods) to determine the log2 fold-change (log2FC) and adjusted P-values of the read count distribution for each SP in cotransin-treated cells (ratio of high to low-fluorescence populations), normalized to the read count distribution in DMSO control cells (**Supplementary Table 1**). SPs with a mean

read count of less than 40 (averaged across DMSO-treated samples) were excluded from further analysis, leaving a total of 3666 SPs (94% of the initially synthesized library). We hypothesized that cotransin-sensitive SPs would be enriched in the low-fluorescence population of treated cells to a greater extent than the DMSO control cells, whereas cotransin-resistant SPs would show a similar distribution in treated and control cells.

Using a log2FC cutoff of ≤ -1 and adjusted P-value ≤ 0.05, we identified 152 KZR-9873-sensitive SPs and 637 KZR-8445-sensitive SPs (**Fig. 2.3b-c**). These results suggest that 100 nM KZR-9873 inhibits translocation of a much smaller subset of SPs (4.2-fold) than 100 nM KZR-8445, consistent with its greater selectivity against cancer cell lines (**Fig. 2.1b**). Results from the pooled SP library and Illumina sequencing were consistent with our initial experiments using individual reporters and flow cytometry (**Fig. 2.1e**). The PD1 SP was sensitive to 100 nM KZR-8445 but resistant to KZR-9873, and the PRL SP was resistant to both cotransins (**Supplementary Table 1**). Although most KZR-9873-sensitive SPs were also inhibited by KZR-8445, 22 SPs were uniquely sensitive to KZR-9873 (**Fig. 2.3c**). This unexpected result suggests that KZR-9873 is not merely a less potent version of KZR-8445, but instead selectively inhibits an orthogonal subset of SPs.

We validated the effects of KZR-9873 and KZR-8445, as suggested by the pooled library screen, on 12 additional SPs using the flow cytometry assay described above. For example, we confirmed using individually cloned reporters that HER3 is more sensitive to KZR-9873 than KZR-8445, whereas IL6ST is more sensitive to KZR-8445, and ITGAV is similarly sensitive to both cotransins (**Fig. 2.3d**). In total, we confirmed that 14 individual SP reporters exhibit the same relative sensitivity to KZR-9873 and KZR-8445 as observed in the SP library screen (**Fig. 2.1e**, **Fig. 2.5a**). Moreover, we observed a remarkable correlation between the log2FCs from the SP library – based on a single concentration (100 nM) – and the corresponding log2(IC_{50}) values (7-point dose response) from the individually cloned reporters (**Fig. 2.3e**). These results confirm that KZR-9873 and KZR-8445 preferentially target overlapping yet distinct SPs.

With this rich dataset, we sought to assess whether cotransin-sensitive SPs exhibit specific biophysical features as compared with resistant SPs. As expected based on previous work[11,29], SPs that were sensitive to either KZR-9873 or KZR-8445 are significantly less hydrophobic and have a lower aliphatic index than resistant SPs (**Fig. 2.5b**). We speculate that less hydrophobic SPs generally have a decreased ability to compete with cotransins for binding to the Sec61 lateral gate helices and the lipid bilayer. We also tested whether specific amino acids (or groups of amino acids with similar biophysical properties) are depleted or enriched at specific positions of cotransin-sensitive SPs (**Fig. 2.3f**, **Fig. 2.5c**). For this analysis, we aligned SPs by their predicted C-terminal cleavage sites and used a scanning window of 3 residues to test for significant enrichment or depletion in cotransin-sensitive vs. resistant SPs (see Methods). Whereas Leu is specifically depleted throughout the H-region[33] of KZR-8445-sensitive SPs (-8 to -14 aa relative to the cleavage site, **Fig. 2.3f**), polar amino acids and Pro are significantly enriched within this region, consistent with previous mutagenesis experiments focused on individual SPs[8,9,11]. To our knowledge, this is the first elucidation of position-dependent, sensitivity-conferring features from an unbiased analysis of thousands of SP sequences. Our analysis further revealed that positively charged amino acids (especially Arg) are enriched in a distinct region (-1 to -7 aa) closer to the cleavage site of cotransin-sensitive SPs (**Fig. 2.3f**). The KZR-9873 dataset showed similar position-dependent trends (**Fig. 2.5c**), albeit with lower statistical significance due to fewer sensitive SPs. This is consistent with the observation that most KZR-9873-sensitive SPs are also sensitive to KZR-8445. However, this global analysis of the SP library dataset did not reveal sequence features that are uniquely enriched in KZR-9873-sensitive SPs, likely owing to their low prevalence.

Position-dependent effects of Arg on cotransin inhibition

The sequence determinants underlying KZR-9873 sensitivity are likely distributed throughout the SP in a complex manner, with certain highly degenerate features (e.g., fewer

hydrophobic and more polar residues within the H-region, **Fig. 2.5c**), combined with more specific motifs. To identify specific sequence motifs, we searched the SP library dataset for human/mouse ortholog pairs (668 total) in which one SP was sensitive and the other was relatively resistant to KZR-9873. Of the 28 human/mouse pairs with divergent sensitivity to KZR-9873 (**Supplementary Table 2**), mSdc1 and hSDC1 were initially selected because they differ by a single amino acid, Arg16 vs. Ser16, which is near the C-terminal end of the H-region (as annotated by SignalP 6.0)[36]. We validated the enhanced sensitivity of mSdc1 using an orthogonal assay based on secreted Gaussia luciferase (SP-GLuc)[29,41], which enables higher throughput determination of accurate IC_{50} values across individually tested SPs. We note that the secreted SP-GLuc reporters are generally more sensitive to cotransins than the corresponding SP reporters based on PD1 and provide a greater dynamic range for comparing distinct SP sequences. Using SP-GLuc reporters, we found that mSdc1 is 11-fold more sensitive than hSDC1 to KZR-9873 (IC_{50} 38 nM and 421 nM, respectively; **Fig. 2.6a**), further validating the SP library results and demonstrating that Arg16 plays a critical role in mSdc1 sensitivity.

Because mSdc1 Arg16 is charged and inserted between two hydrophobic residues (Leu15 and Leu17) within the functionally critical H-region, its enhanced sensitivity could potentially be explained by a reduced binding affinity for the lipid-exposed lateral gate of Sec61. If this were true, both positively and negatively charged residues would be expected to confer enhanced sensitivity to KZR-9873. We therefore tested mSdc1 point mutants with polar and charged amino acid substitutions and found that R16K retained KZR-9873 sensitivity (IC_{50} 58 nM), whereas R16Q and R16E were relatively resistant (IC_{50} 229 nM and 422 nM, respectively; **Fig. 2.6b**). Hence, an Arg or Lys – but not Glu, Gln, or Ser – within the H-region of mSdc1 leads to dramatically enhanced sensitivity to KZR-9873. Despite these large differences in KZR-9873 sensitivity conferred by a single amino acid substitution (up to 11-fold), wild-type and mutant mSdc1 SPs were similarly efficient at promoting GLuc secretion (**Fig. 2.7a**). The latter

observation suggests that charged amino acid substitutions within the H-region of mSdc1/hSDC1 have little or no effect on Sec61-mediated translocation and secretion.

We further defined a specific sensitizing role for Arg and Lys in the SP of human EPHA7, which differs from the mouse ortholog at three positions, including Arg20 (Gly20 in mEpha7). hEPHA7 was more sensitive to KZR-9873 than mEpha7 in the SP library screen, and we confirmed that Arg20 is necessary and sufficient to account for this differential sensitivity (**Fig. 2.6c**). Similar to Arg16 in mSdc1, Arg20 in hEPHA7 is embedded between two hydrophobic residues (Leu19 and Phe21) near the C-terminal end of the H-region. Mutation of Arg20 to Gln or Glu decreased KZR-9873 sensitivity, whereas sensitivity was retained in the Lys mutant (**Fig. 2.6d**). Lastly, the SP of human F8 differs from the mouse ortholog at 7 out of 19 residues, including Arg15 (Asn15 in mF8). Similar to the above examples, Arg15 in hF8 is required for increased sensitivity to KZR-9873, since its mutation to Ala increased the IC_{50} by more than 50-fold (**Fig. 2.7b**). We note that hF8 Glu4, which resides between two hydrophobic residues (Ile3 and Leu5) at the opposite end of the H-region, is the only other charged residue in the SP. In contrast to Arg15, mutation of Glu4 to Ala had no effect on KZR-9873 sensitivity (**Fig. 2.7b**). Collectively, these sequence-function relationships – validated across three unrelated SPs – reveal a previously uncharacterized Arg/Lys motif near the C-terminal end of the H-region, which confers increased sensitivity to KZR-9873.

Although an Arg-bearing H-region may indeed compete less effectively than an uncharged H-region for binding to the Sec61 lateral gate, we considered an alternative, non-mutually exclusive explanation for the observed increase in KZR-9873 sensitivity: the Arg/Lys motif directly or allosterically stabilizes a ternary complex containing both KZR-9873 and the nascent SP bound to Sec61. This hypothesis makes two testable predictions: (1) the effect on KZR-9873 potency should depend on the precise position of the Arg within the SP; (2) most importantly, the magnitude or direction of the Arg-induced potency shift should be different for structurally distinct cotransins. We tested both predictions starting with mSdc1 (**Fig. 2.6a**) and

hEPHA7 (**Fig. 2.6c**), both of which have an Arg near the C-terminal end of the H-region. Remarkably, moving the Arg toward the center of the H-region by three residues caused both SPs to become *less* sensitive to KZR-9873, while simultaneously *increasing* their sensitivity to KZR-8445 (**Fig. 2.6e-f**). A cotransin/SP competition model cannot completely account for these divergent effects, since the potency shift for KZR-9873 and KZR-8445 would be similar if moving the Arg simply led to an SP with lower affinity for the lateral gate. Rather, these results are most consistent with the formation of a functionally inhibited ternary complex comprising Sec61, KZR-9873 and the SP (see Discussion).

HER3 is exquisitely sensitive to KZR-9873

Based on the SP library screen, human epidermal growth factor receptor 3 (HER3, gene name: *ERBB3*) was one of only 22 SPs that was sensitive to KZR-9873, but not the 'pan-selective' cotransin KZR-8445 (**Fig. 2.3b-e**). Because KZR-9873 is much less potent than KZR-8445 toward most SPs, it is difficult to rationalize the increased sensitivity of HER3 to KZR-9873 in terms of a competitive binding model. Moreover, it remained unclear which biophysical properties or sequence features, other than decreased hydrophobicity, distinguish the HER3 SP from other SPs. A clue to this puzzle was revealed by comparing the human and mouse orthologs (hHER3 vs. mHer3) in our SP library dataset (**Supplementary Table 2**). Human HER3 and mouse Her3 SPs differ at 6 out of 19 positions, primarily near the N-terminus. Based on the SP library dataset, these sequence differences were sufficient to confer increased sensitivity of hHER3 to KZR-9873; conversely, mHer3 was more sensitive to KZR-8445 (**Fig. 2.8a**). These trends were also confirmed in the SP-GLuc assay. KZR-9873 was 70-fold more potent against hHER3 (IC_{50} 2 nM) than mHer3 (IC_{50} 146 nM). By contrast, KZR-8445 showed the converse species preference, with decreased potency against hHER3 (IC_{50} 43 nM) relative to mHer3 (IC_{50} 14 nM) (**Fig. 2.8a**).

Strikingly, a single 'humanizing' point mutation in mHer3 – Ser2Arg (S2R) – increased

sensitivity to KZR-9873 (IC_{50} 17 nM), yet decreased sensitivity to KZR-8445 (IC_{50} 45 nM) (**Fig. 2.8b-c**). The reciprocal mutation of hHER3 Arg2 to either Ser or Ala (R2S, R2A) led to increased resistance to KZR-9873 (**Fig. 2.9**). A second humanizing mutation (S2R/I4N) further sensitized mHer3 to KZR-9873 (IC_{50} 4 nM), similar to the level of wild-type hHER3 (**Fig. 2.8b-c**). Thus, two substitutions near the N-terminus of mHer3 are sufficient to increase the potency of KZR-9873 by 40-fold, while simultaneously decreasing the potency of KZR-8445. Although the sensitizing Arg is near the N-terminus of HER3 rather than the C-terminal end of the H-region, these divergent effects are reminiscent of the H-region Arg/Lys motif described above (**Fig. 2.6e-f**) and are similarly inconsistent with a simple competitive binding model.

HER3 is a cell-surface receptor that plays a critical role in multiple cancer types by amplifying signals from the oncogenic receptor tyrosine kinase HER2[42,43]. Although HER3 was previously shown to be sensitive to distinct cotransins[29,41], these compounds are neither as potent nor as selective as KZR-9873. We reasoned that KZR-9873-mediated inhibition of HER3 biogenesis could potentially account for its selective antiproliferative effects on certain cancer cell lines, including CAL27 (**Fig. 2.1b**, **Fig. 2.2a**). Consistent with this hypothesis, treatment of CAL27 cells with KZR-9873 for 24 h reduced endogenous HER3 levels at concentrations as low as 11 nM; by contrast, a 30-fold higher concentration of KZR-8445 was required to achieve similar levels of inhibition (**Fig. 2.10a**). The effects of KZR-9873 on endogenous HER3 were paralleled by a dose-dependent reduction in phosphorylated AKT, in line with previous work showing that HER3 is an upstream activator of the PI3K/AKT pathway[44,45]. Phosphorylated AKT was unaffected by KZR-8445 treatment, suggesting that KZR-9873 and KZR-8445 block cell proliferation by distinct mechanisms. In further support of this conclusion, forced overexpression of HER3 reduced the sensitivity of CAL27 cells to KZR-9873 (**Fig. 2.10b-c**), whereas sensitivity to KZR-8445 was unaffected. Finally, we confirmed that genetic knockdown of HER3 is sufficient to reduce CAL27 cell proliferation (**Fig. 2.11a-b**), as reported previously[45,46].

Collectively, our results suggest that reduction of HER3 levels by KZR-9873 is

necessary and sufficient to account for its potent antiproliferative effects on CAL27 cells (**Fig. 2.10b-c**, **Fig. 2.11a-b**). KZR-8445, which does not affect HER3 expression at relevant concentrations, presumably blocks cell proliferation by targeting a distinct subset of Sec61 clients, as suggested by the SP library screen (**Fig. 2.3b**). Such a pan-selective inhibition profile exhibited by KZR-8445 – which nevertheless shows greater discrimination than the nonselective Sec61 inhibitors mycolactone and apratoxin[29] – could deprive cells of multiple Sec61 clients required for growth and/or induce proteotoxic stress due to mislocalization and misfolding of newly synthesized Sec61 clients.

Discussion

Cotransin-family cyclic depsipeptides selectively inhibit the biogenesis of Sec61 clients in a signal peptide-dependent manner. Identifying which SPs are affected by cotransins, among the thousands of unique SPs predicted to mediate secretory and membrane protein biogenesis, is the major unsolved challenge motivating our study. We additionally sought to define previously unknown relationships between the SP amino acid sequence and cotransin sensitivity. While individual SP reporter assays and immunoassays have been useful for assessing cotransin potency and selectivity, only a tiny fraction of Sec61-dependent SPs have been tested using these methods. Hence, the primary goal of our study was to develop a massively parallel screening platform for profiling cotransins against the entire signal peptidome. Toward this goal, our first-generation pooled-cell reporter library includes >90% of the predicted human SPs based on Uniprot annotation. We also included a subset of orthologous mouse SPs, a novel feature of our library design that enabled the discovery of nonconserved SP sequences required for cotransin sensitivity.

Of the 3666 SP reporters quantified in our pooled screen, 637 were sensitive to the pan-selective cotransin KZR-8445 (at 100 nM). A direct comparison with the structural variant KZR-9873 revealed a dramatic difference in SP selectivity. Despite inhibiting only 4% of the SPs

quantified in the pooled reporter screen (compared to 17% inhibited by KZR-8445), KZR-9873 blocked the proliferation of certain cancer cell lines with equal potency to KZR-8445. Strikingly, 22 out of 152 KZR-9873-sensitive SPs, including HER3, were insensitive to KZR-8445, demonstrating for the first time the possibility of targeting orthogonal subsets of Sec61 clients with structurally distinct cotransins. Although poor solubility and pharmacokinetic properties render it unsuitable for in vivo studies, KZR-9873 nevertheless represents a potential early lead for the development of small-molecule therapeutics targeting HER3-driven cancer, including *NRG1*-rearranged solid tumors[47,48].

A mechanistic understanding of how cotransins discriminate among different SPs has focused primarily on a competitive inhibition model[9,11,28,29], in which relatively resistant SPs bind to the Sec61 lateral gate with higher affinity (and/or faster kinetics) than sensitive SPs. Although it has not been technically feasible to measure the relative binding affinities of different SPs for Sec61 (especially in the context of a productive interaction with a ribosome-nascent chain complex), multiple lines of evidence support the competitive inhibition model. KZR-8445 and the prolactin SP bind to a common lipid-facing site between the Sec61 lateral gate helices TM2 and TM7, as revealed by cryo-EM structures[27,29]. In addition, stepwise mutation of polar to hydrophobic residues (e.g., Leu) within the H-region of individual SPs generally causes an additive decrease in cotransin sensitivity, whereas helix-breaking Pro mutations generally have the opposite effect[8,9,11]. We note that our unbiased, position-dependent analysis of hundreds of KZR-8445-sensitive and resistant SP sequences confirms and extends the latter results (**Fig. 2.3f**).

Nevertheless, a simple competitive inhibition model predicts that the difference in potency between KZR-9873 and KZR-8445 should be similar across all SPs, which is not what we observed. We speculate that the increased potency of KZR-9873 toward certain SPs (relative to KZR-8445) derives from cooperative interactions between KZR-9873, Sec61, and the SP. Here, KZR-9873 engages the lateral gate and lumenal plug helix (as observed in the

cryo-EM structure of KZR-8445), while the SP binds to the cytosolic vestibule. What is the role of the H-region Arg/Lys motif in stabilizing this ternary complex? Although a detailed structural understanding will require high-resolution cryo-EM analysis of biochemically defined complexes, we consider two speculative models. In the first model, the positively charged Arg or Lys forms a cation-pi interaction with the chlorobenzothiophene side chain of KZR-9873. Assuming that KZR-9873 binds Sec61 in a similar orientation as KZR-8445[29], the chlorobenzothiophene would be near the center of the cytosolic vestibule, poised to form direct contacts with a nascent SP (e.g., H-region of mSdc1, hEPHA7, and hF8) below the ribosome exit tunnel. By contrast, the corresponding bromobenzyl indole side chain of KZR-8445 may be too bulky to directly engage the Arg/Lys motif, and KZR-8445 may instead rely primarily on a competitive inhibition mechanism. In the second model, the H-region Arg/Lys motif interacts with Sec61 in a manner that allosterically enhances the binding affinity of KZR-9873 to a greater extent than KZR-8445. Arg2 in the HER3 SP could potentially play a similar role in either model, albeit with the HER3 SP adopting an inverted orientation within the cytosolic vestibule. Indeed, previous studies have shown that certain SPs can initially engage Sec61 in a "head-first" orientation prior to inverting to adopt the classical "looped" orientation required for SP cleavage[8,9,49].

The SP library screening dataset reported in this study provides an unprecedented resource of cotransin-sensitive SPs, and we envision that our SP library platform will be useful for profiling client-selective Sec61 inhibitors from different chemical classes, including cyclotriazadisulfonamides (e.g., CADA)[50]. Despite the broad utility of our screening platform, our study has certain limitations. First, the SP screening platform does not reveal the effects of cotransins on full-length Sec61 clients expressed at endogenous levels, which may exhibit quantitative differences compared to the corresponding overexpressed SP reporters. While quantitative proteomics could be a useful alternative approach, it has certain drawbacks, including the need for pulse-labeling to measure the biogenesis of long-lived membrane proteins and the cell type-restricted expression of many secretory and membrane proteins (e.g.,

cytokines, growth factors, immune receptors). Second, the current version of our screening platform cannot be used to assess the large fraction of Sec61 clients that lack an SP, including all type II single-spanning and most multi-spanning transmembrane proteins. We note that the biogenesis of many multi-spanning transmembrane proteins is initiated by the ER membrane complex (EMC) and is resistant to Sec61 inhibitors[51]. Lastly, many KZR-9873-sensitive SPs lack an Arg/Lys motif near the C-terminal end of the H-region. Hence, there are clearly additional sensitivity-conferring sequence motifs that remain to be discovered. Larger libraries containing expanded sets of evolutionarily related yet sequence-divergent SPs (e.g., orthologous SPs from several mammalian or vertebrate species) will likely facilitate the discovery of additional specific sequence motifs. Major goals for the future include the elucidation of novel sensitizing motifs for KZR-9873 and other client-selective Sec61 inhibitors, as well as high-resolution cryo-EM structures of KZR-9873/Sec61 complexes assembled in the presence of selectively sensitive ribosome-nascent chain complexes (e.g., HER3, Sdc1, EPHA7). Such structures promise to inspire the design of Sec61 modulators with even greater levels of client orthogonality.

Experimental Methods

Cell culture

Cell lines were obtained from ATCC (unless stated otherwise) and were used without further authentication. All cell lines tested negative for mycoplasma contamination. BxPC-3 (ATCC, Cat# CRL-1687), SW48 (ATCC, Cat# CCL-231), H929 (ATCC, Cat# CVCL-1600), and HCC95 cells (ATCC, Cat# CRL-2868) were cultured in RPMI-1640 with GlutaMAX (Thermo Fisher Scientific, Cat# 61870036). CAL27 (ATCC, Cat# CRL-2095), HeLa (ATCC, Cat# CCL-2) and HEK293 T-rex FRT cells (Thermo Fisher Scientific, Cat# R78007) were cultured in DMEM with GlutaMAX (Thermo Fisher Scientific, Cat# 10569069). HCT116 cells (ATCC, Cat# CCL-247) were cultured in McCoy's 5A medium with GlutaMAX (Thermo Fisher Scientific, Cat# 36600021). K562-TET3G cells were obtained from Martin Kampmann (UCSF) and were cultured in RPMI-1640 with GlutaMAX. All growth media contained 10% fetal bovine serum (Axenia Biologix, Cat# F001) and 1% penicillin-streptomycin (Gibco, Cat# 15140122).

KZR-9873

KZR-9873 was synthesized using a previously described solid-phase peptide synthesis route[29]. Purity: 99%. ^{1}H NMR (400 MHz, acetone-d6) δ 8.59 (d, *J* = 10 Hz, 1H), 8.22 (d, *J* = 2.4 Hz, 1H), 7.99 (d, *J* = 9.6 Hz, 1H), 7.82 (d, *J* = 8.8 Hz, 1H), 7.77 (d, *J* = 6.8 Hz, 1H), 7.53 (s, 1H), 7.30 (dd, *J* = 2.4 Hz, 8.8 Hz, 1H), 7.23-7.18 (m, 1H), 6.94 - 6.89 (m, 1H), 6.44-6.40 (m, 2H), 5.34 (s, 1H), 5.31-5.28 (m, 1H), 5.06-5.01 (m, 2H), 4.91-4.85 (m, 1H), 4.67-4.64 (m, 1H), 4.28-4.24 (m, 1H), 3.97-3.92 (m, 1H), 3.36-3.31 (m, 1H), 3.20 (s, 3H), 3.10-3.01 (m, 2H), 2.94 (s, 3H), 2.72- 2.63 (m, 2H), 2.54 (s, 3H), 2.50-2.43 (m, 2H), 2.34-2.27 (m, 2H), 2.23-2.19 (m, 2H), 2.17-2.13 (m, 1H), 1.88-1.82 (m, 1H), 1.78-1.75 (m, 2H), 1.70-1.57 (m, 2H), 1.53-1.43 (m, 7H), 1.35-1.30 (m, 1H), 0.98-0.93 (m, 9H), 0.88 (d, J = 6.4 Hz, 3H). HRMS (ESI): Calculated for $C_{51}H_{65}ClFN_7O_8S^+$ $[M + H]^+$, 990.4361; Found 990.4354.

Proliferation assay

Adherent cells were trypsinized, resuspended, and seeded (3,000 – 5,000 cells) in 100 µL complete growth medium per well in 96-well clear-bottom plates. Suspension cells were seeded at 10,000 cells in 100 µL of complete growth medium per well in 96-well clear-bottom plates. Cells were left to adhere or grow overnight. Cells were treated with 25 µL of x5 drug stocks (0.1% DMSO final) and incubated for 72 h. AlamarBlue was used to assess cell viability. After a 1.5-h incubation in alamarBlue at 37 °C, fluorescence intensity was measured using a SPARK microplate reader (Tecan). IC_{50} curves were generated by first normalizing fluorescence intensity in each well to the DMSO-treated plate average. Normalized fluorescence intensity was plotted in GraphPad Prism and IC_{50} values were calculated from nonlinear regression curves.

Plasmid generation

The dox-inducible bicistronic SP reporter encoding full-length PD1_IRES-iRFP was constructed by Gibson assembly of these sequences into the pTRE3G vector. EcoR1 and BamH1 restriction sites were inserted into the 5' and 3' ends of the PD1 SP using site-directed mutagenesis; this enabled excision and replacement of the PD1 SP with oligonucleotide cassettes encoding SPs from any protein of interest. SP reporters were constructed by ligating oligonucleotides encoding the SP (plus 6 aa from the mature region) [SP+6 aa] of a protein of interest into the indicated restriction sites. The pcDNA5 FRT DEST/TO eGLuc2 vector[29,41] ("SP-GLuc" reporter) was modified by the addition of EcoR1 and BamH1 restriction sites to facilitate ligation of [SP+6 aa] oligonucleotide cassettes in frame with the mature region of eGLuc2 (aa 22-387). Full-length HER3 was cloned into the pHR vector containing IRES-AcGFP via Gibson assembly (HER3_IRES-AcGFP). Plasmids encoding sgRNAs were constructed by ligating complementary

oligonucleotides into the BlpI and BstXI restriction sites of pCRISPRia-v2[52]. All plasmids were verified by Sanger sequencing.

Lentivirus generation

Lentivirus was generated by transfecting HEK-293T cells seeded at approximately 60% confluency in 6-well plates in 3 mL of complete DMEM. Using Mirus TransIT-LT1 transfection reagent, cells were transfected with 1.5 µg of the lentiviral plasmid of interest, 1.35 µg of pCMV-dR8.91, and 165 ng of pMD2-G diluted in OPTI-MEM and treated with Alstem ViralBoost. After two days, an additional 1 mL of complete DMEM was added to cells. On the third day, virus-containing medium was harvested and filtered through 0.45 µm sterile SFCA syringe filters (Thermo Scientific). The supernatant was either used immediately or stored at -20 °C.

SP reporter (cell-surface PD1) analysis by flow cytometry

K562-TET3G cells were transduced with a SP reporter by adding growth medium containing 8 µg/mL polybrene and the SP reporter-encoding lentivirus. After centrifugation at 2,500 rpm at 30 °C for 2 h, lentiviral medium was replaced with complete medium. Two days later, cells were treated with 250 ng/mL doxycycline (dox) for 24 h, harvested and analyzed via flow cytometry to determine transduction efficiency. Transduced cells were passaged in complete growth medium.

K562-TET3G cells stably expressing a SP reporter were seeded in 24-well plates at 200,000 cells/mL in 500 µL complete growth medium. Cells were treated with 250 ng/mL dox and the indicated concentrations of KZR-9873 or KZR-8445 for 24 h. Cells were transferred to 96-well V-bottom plates and centrifuged at 2,100 *g* for 3 min at 4 °C. Cells were washed by resuspending in FACS buffer (2% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin, and 2 mM EDTA in PBS lacking Ca^{+2}/Mg^{+2}) and centrifuging. Cells were resuspended in 50 µL of FACS

buffer plus 0.5 µL of FITC-conjugated PD1 mAb (BD Biosciences, Cat# 557860) and incubated, while covered, on ice for 30 min. Cells were washed x3 and resuspended in FACS buffer for flow cytometry analysis using a BD CytoFLEX S to quantify FITC and iRFP intensities. Using FlowJo (BD) software, single cells were analyzed, and debris (FSC-H vs. SSC-A) and doublets (FSC-H vs. FSC-W) were excluded. Single cells were gated on iRFP expression, and cells within this gate were used to calculate the ratio of FITC to iRFP fluorescence intensity. Flow cytometry data (FITC:iRFP ratios) were quantified using the mean fluorescence intensity (MFI) across the single cell populations and plotted as the percentage of the DMSO controls.

Generation of the SP library

A custom SP library was designed to include nearly all human SPs (3212) (based on Uniprot annotation). In addition, the mouse orthologs of therapeutically relevant human SP-containing proteins (targets of approved or investigational drugs based on the CHEMBL26 database) were included as a sublibrary (668 mouse SPs). Full-length cDNA sequences encoding the relevant human and mouse proteins (3880 SP-containing Sec61 clients) were downloaded from Ensembl (Release 99) and used to design oligonucleotide cassettes encoding the SP plus 6 aa from the mature region. Proceeding from the 5' to 3' direction, each oligonucleotide cassette (250 nt total length) contained the following sequences (**Fig. 2.12**): (1) EcoR1 site, (2) filler DNA sequence to compensate for variable SP lengths, (3) T7 (18 nt), (4) unique barcode for each SP (7 nt), (5) Kozak sequence (9 nt), (6) SP+6 aa, (7) BamH1 site. The oligonucleotide pool was synthesized by Twist Biosciences. The oligo pool was PCR amplified, digested with EcoR1/BamH1, ligated into the pTRE3G SP reporter vector (SP-PD1_IRES-iRFP, **Fig. 2.1c**), and transformed into chemically competent E. coli (Takara, Cat# 636763) in solution by shaking for 16 h at 37 °C. At least 100-fold representation of the library was maintained up to this step (assessed by streaking serial dilutions of the ligation/transformation mixture onto LB/carbenicillin plates). Lentivirus was generated by transfecting HEK-293T cells using Mirus TransIT-LT1

transfection reagent. Cells were seeded at approximately 60% confluency in a 10 cm dish in 8 mL of complete DMEM and transfected with 9 µg of SP library plasmid, 8 µg of pCMV-dR8.91, and 1 µg of pMD2-G lentiviral packaging plasmids. $12x10^6$ K562-TET3G cells were transduced with the lentivirus SP library by centrifugation at 2,500 rpm at 30 °C for 2 h. Cells were transduced at an MOI = 0.4 to maintain at least 1000-fold representation of the library. After 2 d, transduced cells were treated with 250 ng/mL dox for 24 h, harvested and sorted based on iRFP intensity using a FACS Aria II (BD Biosciences) to generate the SP library population.

Signal peptide library screen

K562-TET3G signal peptide library cells were maintained at ≥ 500-fold representation (at least $2x10^6$ library cells after passaging). $8x10^6$ library cells were seeded at 400,000 cells/mL (20 mL total volume) and treated with 250 ng/mL dox and either DMSO, 100 nM KZR-9873, or 100 nM KZR-8445 in biological duplicate (6 populations total). After 24 h, the 6 cell populations were harvested and stained with 12 µL of FITC-conjugated PD1 mAb (BD Biosciences, Cat# 557860) in 2 mL FACS buffer (2% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin, and 2 mM EDTA in PBS lacking Ca^{+2}/Mg^{+2}) for 30 min while rotating and covered at 4 °C. Stained cells were then 3X washed in 2 mL of FACS buffer, resuspended a final time at $10x10^6$ cells/mL and sorted on a FACS Aria II (BD Biosciences). Single, iRFP-positive cells were separated into high (top 33%) and low (bottom 33%) populations based on the calculated FITC to iRFP ratio. Approximately $2x10^6$ cells per fraction were collected, centrifuged, and stored at -80 °C.

Genomic DNA was isolated from sorted cell fractions (12 total) with a NucleoSpin midi kit, according to the manufacturer's instructions (Macherey-Nagel, Cat# 740954.20). The region of the oligonucleotide cassette that contains the barcode was amplified by PCR using a forward primer that anneals to the T7 sequence and a reverse primer that anneals to the mature domain of PD1 (**Table 2.1**) (Q5 Hot Start PCR Master Mix, Cat# NEB M094). The unique forward

primers for each sorted fraction contain the Illumina P7 sequence and an index at the 5' end. The common reverse primer contains the Illumina P5 sequence at 3' end. In total, 1 µg of genomic DNA was amplified in 10-16 separate 50 µL-scale PCR reactions (depending on the final concentration of isolated genomic DNA in each sample). PCR products were size-selected using SPRI-select beads to remove PCR primers and genomic DNA (Beckman Coulter, Cat# B23317). Purified samples were pooled and sequenced with a custom sequencing primer (**Table 2.1**) using an Illumina MiSeq instrument.

Table 2.1: Primers for Illumina sequencing of signal peptide library.

Primer	**Primer sequence**
Common reverse	AATGATACGGCGACCACCGAGATCTACACCTCTCCGATGTG TTG
Forward index-1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGT TCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-7	CAAGCAGAAGACGGCATACGAGATCAGATCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGT TCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-9	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGT TCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-10	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-11	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGT TCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Forward index-12	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAG
Sequencing	GAATTCCCAGGATGGTTCTTAGACTCCCCAGACA

Bioinformatics for the signal peptide library screen

Raw Illumina reads derived from each high (top 33%) and low (bottom 33%) sorted population were first trimmed of constant sequences from the SP reporter vector backbone using Cutadapt[53]. The sequence was further trimmed to include only the barcode nucleotides, which were mapped to the reference input library of barcodes using Bowtie2[54]. Count tables were generated from mapped reads that aligned perfectly to the reference sequence (no mismatches or errors allowed). Differences between drug- and DMSO-treated cells were assessed using the R package DESeq2, which is commonly used for differential gene expression analysis[55]. Unlike the typical gene expression experiment in which one group of samples is compared to another as a ratio, here we compared four groups of samples using a ratio of ratios to estimate how much signal peptide enrichment differs with drug treatment relative to the DMSO control: (drug-treated high counts / drug-treated low counts) / (DMSO-treated high counts / DMSO-treated low counts). The significance of that difference was calculated with a likelihood ratio test comparing a full model with an interaction term (treatment : sort population) to a reduced model without the interaction term. The log2[fold change] (log2FC) and Benjamini-Hochberg adjusted P-value (P-adj) are shown in **Supplementary Table 1** for SPs with average DMSO read counts ≥ 40 (3666 signal peptides total). For **Supplementary Table 2**, the criteria for selecting orthologous human and mouse SPs with divergent KZR-9873 sensitivity were as follows: SP1 log2FC ≤ -1, P-adj ≤ 0.05; SP2 log2FC >-1; $|\Delta$log2FC$|$ ≥ 1 (SP1-SP2).

Physicochemical features of sensitive and resistant SPs

Physicochemical properties of SP sequences (aliphatic index and hydrophobicity index) were calculated for all SPs in the library using the Peptides package in R (https://cran.r-project.org/web/packages/Peptides/index.html). dGsub, the apparent free energy for sequence insertion into the endoplasmic reticulum (ER) via Sec61, was calculated using an online tool

(https://dgpred.cbr.su.se/index.php?p=TMpred; subsequence allowed, without length correction)[56,57]. Differences in the distribution of these scores for sensitive and resistant SPs were assessed using a two-sided Wilcoxon rank sum test. For these comparisons, as well as the statistical enrichment analysis described in the next section, we used the following criteria for "sensitive" vs. "resistant" SPs. Sensitive: Log2FC < 0; P-adj ≤ 0.05; mean read count (DMSO samples) ≥ 100. Resistant: Log2FC > 0; mean read count (DMSO samples) ≥ 100.

Analysis of sensitive and resistant SPs for position-dependent amino acid enrichment

All SPs were aligned by their C-terminal ends (denoted as position -1). The number of times each amino acid (or amino acid group) occurs at each position was counted within the sensitive and resistant SPs (see previous section for criteria), separately. Each such count was divided by the total number of observations at its position (note that this latter number can vary across positions because SPs vary in length), yielding the frequency of the amino acid (or amino acid group) at that position. The direction of enrichment and sign of the enrichment score (positive or negative) was determined by comparing the frequency of an amino acid (or amino acid group) at a position in the sensitive SPs to the frequency in the resistant SPs. If the frequency is higher in the sensitive than in the resistant SPs, the enrichment score is positive (indicating enrichment); conversely, if the frequency is lower in the sensitive than in the resistant SPs, the enrichment score is negative (indicating depletion). The value of the enrichment score indicates the statistical significance of the enrichment (or depletion) and is calculated as the $-\log_{10}$ of the P-value derived from a Fisher's exact test comparing the count of the amino acid (or amino acid group) at a particular position to the count of all other amino acids (or all amino acids not in the group) at that position in the sensitive SPs versus the resistant SPs (i.e., a 2x2 contingency table). Upon inspection of the enrichment matrices derived as described above, we noticed that enrichment of a particular amino acid often occurs across multiple adjacent positions. Therefore, to gain power, we modified the above approach, counting amino acids (or amino acid groups)

not only at a given position, but also at the immediately adjacent positions (i.e., within a window of 3 amino acids). All subsequent steps were as described above. Heat maps in **Fig. 2.3e** and **Fig. 2.5d** depict enrichment scores based on the analysis of 3 aa windows.

Immunoblot analysis

CAL27 cells were seeded in 6-well plates at 500,000 cells per plate in 2 mL complete growth medium. Seeded cells were treated with DMSO and increasing concentrations of KZR-9873 and KZR-8445 for 24 h. Treated cells were washed with ice-cold PBS then lysed using NP-40 lysis buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 1% NP-40, 2x EDTA-free protease inhibitors, 1X phosphatase inhibitors). Lysates were collected by scraping and were cleared by centrifugation at 16,100 *g* for 10 min at 4 °C. Total protein was quantified using the Bradford method and normalized prior to electrophoresis using hand-cast 7.5% polyacrylamide gels. Proteins were transferred to 0.45 µm nitrocellulose membranes using a Bio-Rad Criterion transfer system. Membranes were blocked for 1 h at room temperature using blocking buffer (5% BSA, 0.1% sodium azide in TBS-T). Membranes were incubated with primary antibodies (GAPDH: Santa Cruz Biotechnology, Cat# sc-32233; phospho-AKT: Cell Signaling Technology, Cat# 4060S; AKT: Cell Signaling Technology, Cat# 2920S; ERBB3: Cell Signal Technology, Cat# D22C5; 1:1000 dilution) diluted in blocking buffer for 1 h at room temperature. Membranes were rinsed with TBS-T (3 x 5 min at room temperature) and incubated with secondary antibodies (LI-COR, IRDye 680RD goat anti- mouse IgG, Cat# 926-68070; IRDye 800CW goat anti-rabbit IgG, Cat# 926-32211, 1:10000 dilution) diluted in blocking buffer for 1 h at room temperature. Membranes were rinsed with TBS-T (3 x 5 min at room temperature) and were imaged using a Li-Cor Odyssey system and analyzed with Image Studio Lite (Li-Cor) software.

Generation of CAL27 HER3 overexpressing (HER3-OE) cells

CAL27 cells stably overexpressing HER3 (HER3-OE) were generated by transducing CAL27 cells with lentivirus encoding full length HER3_IRES-AcGFP. CAL27 cells were incubated overnight in medium with 8 µg/mL polybrene and the HER3_IRES-AcGFP lentivirus, after which lentivirus was removed from cells and replaced with complete growth medium. Transduction efficiency was determined two days later via flow cytometry analysis. CAL27 HER3-OE cells were sorted based on GFP intensity using a FACS Aria II (BD Biosciences) and were subsequently passaged in complete growth medium.

Analysis of HER3 surface expression by flow cytometry

CAL27 and CAL27 HER3-OE cells were seeded in 12-well plates (75,000 cells/well) in 1 mL complete growth medium. Cells were treated with the indicated concentrations of KZR-9873 or KZR-8445 for 24 h. Cells were harvested by trypsinization, transferred to 96-well V-bottom plates and centrifuged at 2,100 *g* for 3 min at 4 °C. Cells were washed by resuspending in FACS buffer (2% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin, and 2 mM EDTA in PBS lacking Ca^{+2}/Mg^{+2}) and centrifuging. Cells were resuspended in 50 µL of FACS buffer plus 0.5 µL HER3-APC mAb (VWR, Cat# 324708-BL) and incubated, while covered, on ice for 30 min. Cells were washed x3 and resuspended in FACS buffer for flow cytometry analysis using a BD CytoFLEX S to quantify APC intensity. Using FlowJo (BD) software, single cells were analyzed, and debris (FSC-H vs. SSC-A) and doublets (FSC-H vs. FSC-W) were excluded.

CRISPRi/dCas9 knockdown of HER3

CAL27-dCas9 cells were generated by transducing CAL27 cells with lentivirus encoding dCas9-BFP-KRAB. Transduced cells were sorted based on BFP positivity using a FACS Aria II (BD Biosciences) and expanded in complete growth medium. The CAL27-dCas9 cells were transduced with non-targeting (NT) or HER3 sgRNAs in pCRISPRia-v2 (25-55% transduction

efficiency based on BFP expresssion) (**Table 2.2**). After expanding for 4 d, each cell population was seeded into two separate 12-well plates (75,000 cells per well). On the next day (denoted "day 0"), cells from one plate were harvested and analyzed by flow cytometry using an APC-conjugated HER3 mAb (VWR, Cat# 324707-BL). Cells in the second plate were allowed to grow and were harvested and reseeded (1:3 dilution) every 2 d for 14 d. The harvested cells were analyzed by flow cytometry using a BD CytoFLEX S to quantify APC intensity (HER3 mAb) and the percentage of BFP-positive cells (sgRNA+). Using FlowJo (BD) software, single cells were analyzed, and debris (FSC-H vs. SSC-A) and doublets (FSC-H vs. FSC-W) were excluded. The percentage of cells within a BFP-positive gate was calculated and normalized to the day 0 measurement for each sgRNA.

Table 2.2: Protospacer sequences of sgRNAs used for HER3 knockdown.

Name	Protospacer sequence (5' to 3')
Non-targeting (NT) CRISPRi sgRNA[58,59]	GGACTAAGCGCAAGCACCTA
HER3-1 CRISPRi sgRNA[52]	GCCCCTAGGCTAGGACATCG
HER3-2 CRISPRi sgRNA[52]	GTCGGCCTCAGCCTCTTCTG

SP-GLuc reporter assay

Secreted luciferase (eGLuc2) assays were performed in HEK293 T-rex cells (Life Technologies). Cells were cultured in DMEM with GlutaMAX (Thermo Fisher Scientific) and supplemented with 10% fetal bovine serum. Cells were plated at $5x10^5$ cells per well in 6-well plates. After overnight incubation, cells were transfected with 1 µg of the SP-GLuc reporter plasmid using Lipofectamine 2000 (Invitrogen). After incubating for 4 h, the transfected cells were seeded (10,000 cells) in 100 µL of complete growth medium per well in 96-well clear-bottom plates. After adhering overnight, the cells were treated with 25 µL of x5 doxycycline (1 µg/mL final) and x5 drug stocks (0.1% DMSO final) and incubated for 24 h. Secreted luciferase

in the conditioned media was quantified via luminescence using QUANTI-Luc (InvivoGen) and measured using a SPARK microplate reader (Tecan). IC_{50} curves were generated by first normalizing luminescence intensity in each well to the DMSO-treated plate average. Normalized luminescence intensity was plotted in GraphPad Prism and IC_{50} values were calculated from nonlinear regression curves.

Figures

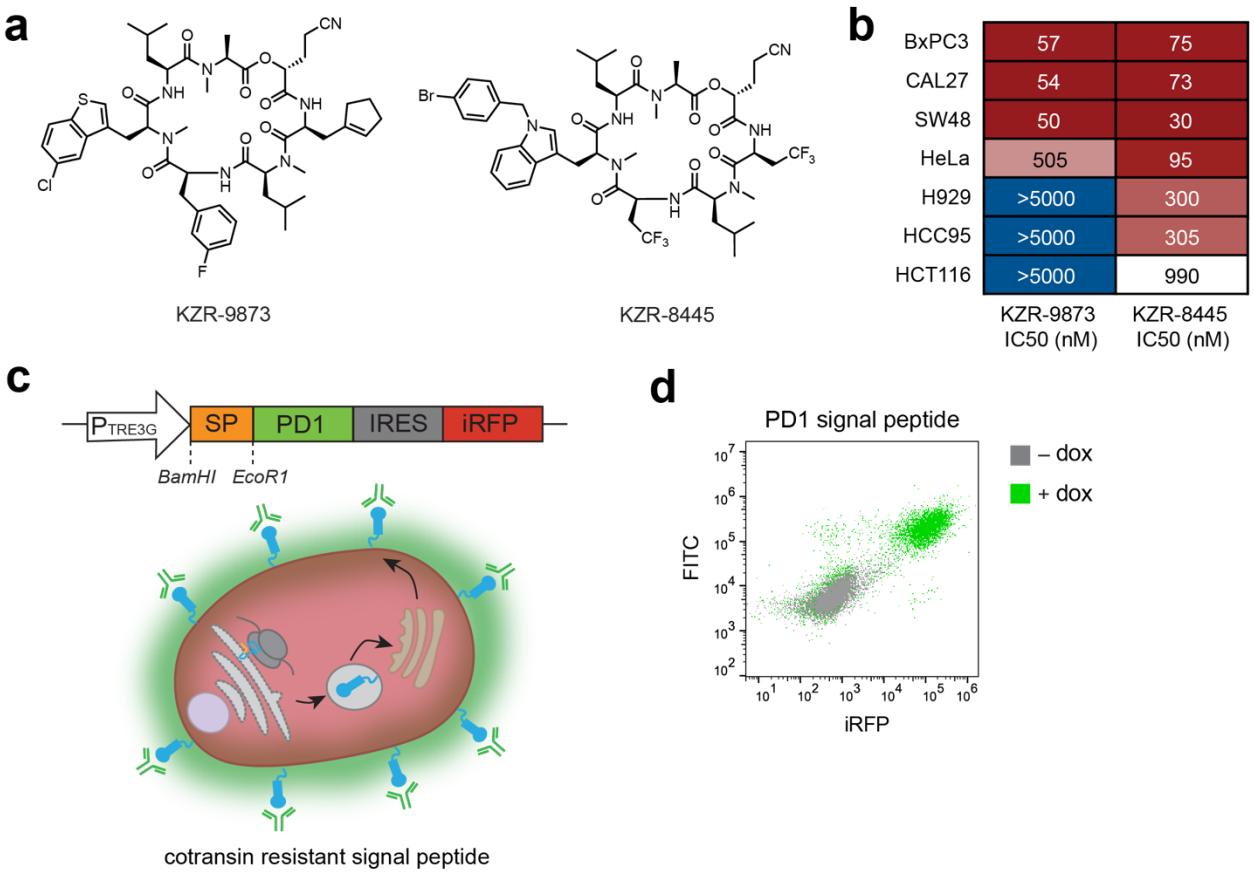


Figure 2.1: SP reporter based on the cell-surface protein, PD1
a, Structures of KZR-9873 and KZR-8445. **b**, Antiproliferative effects of KZR-9873 and KZR-8445 on the indicated cancer cell lines after 72 h treatment. **c**, Schematic representation of the SP-reporter construct. **d**, K562-TET3G cells stably expressing the PD1 SP-reporter (SP derived from PD1) were treated ± doxycycline (dox, 250 ng/mL). After 24 h, cells were stained with a FITC-conjugated PD1 antibody. FITC and iRFP intensities were quantified by flow cytometry. **e**, K562-TET3G cells stably expressing the indicated SP-reporter (SP derived from PD1 or PRL) were treated with dox and increasing concentrations of KZR-9873 and KZR-8445. After 24 h, cells were stained with a FITC-conjugated PD1 antibody and the FITC/iRFP ratio quantified by flow cytometry. Data are presented (Figure caption continued on the next page)

(Figure caption continued from the previous page) as mean ± s.d., n = 3. Statistical significance was determined by unpaired *t* tests, ns = not significant; ***P < 0.0005. The P-value for 111 nM KZR-9873 versus DMSO is 0.2763. The P-value for 111 nM KZR-8445 versus DMSO is 0.0004.

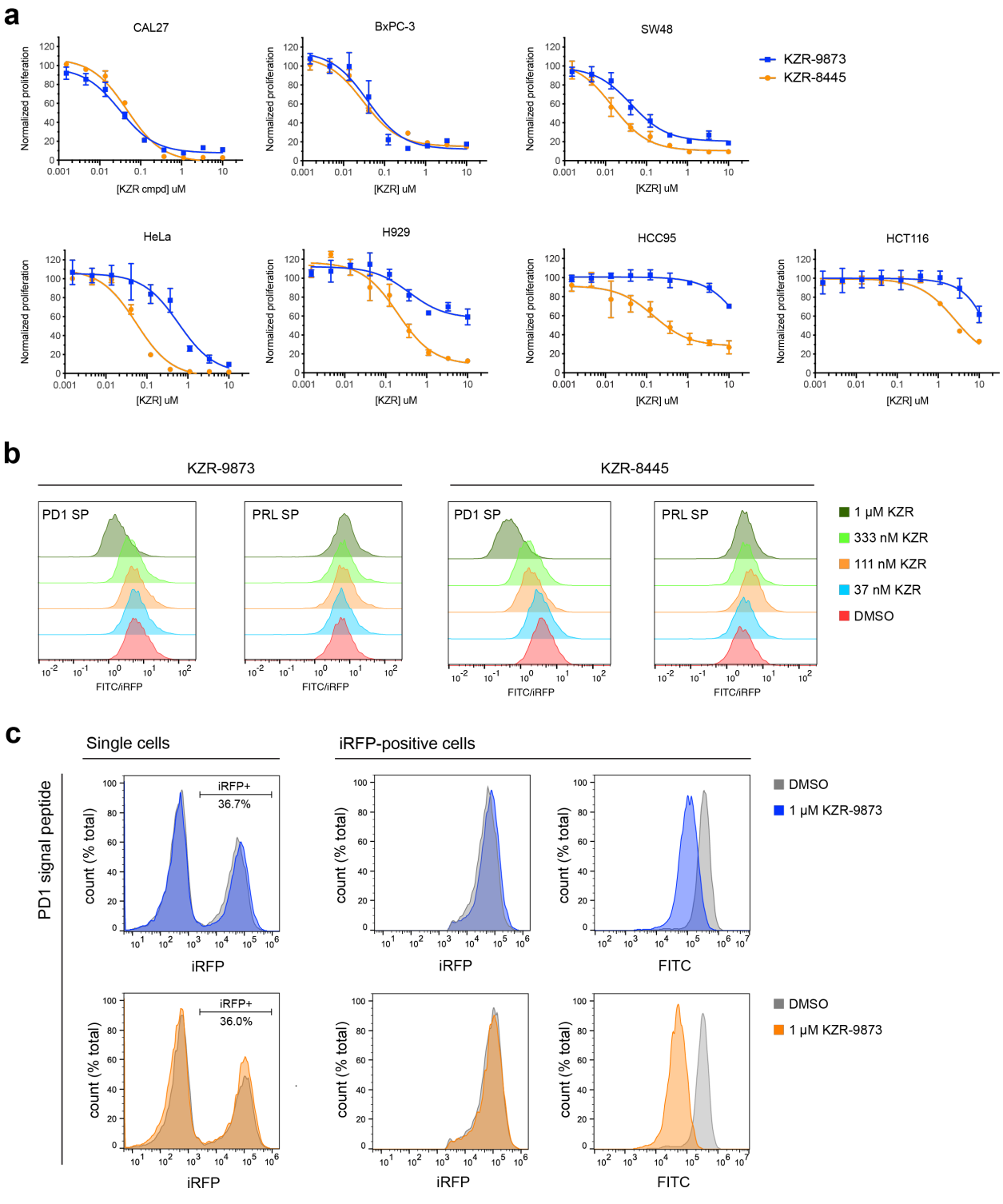


Figure 2.2: Effects of KZR-9873 and KZR-8445 on cancer cell lines and an SP reporter based on the cell-surface protein, PD1.
a, The indicated cell lines were treated with DMSO or increasing concentrations of KZR-9873 and KZR-8445. After 72 h, cell proliferation was quantified using alamarBlue (%DMSO, mean ± s.d., n = 3). **b**, K562-TET3G cells were transduced with SP-reporters (**Fig. 2.1c**) containing PD1 and PRL SPs and treated with dox (250 ng/mL) and DMSO or increasing concentrations of KZR-9873 and KZR-8445. After 24 h, cells were stained with a FITC-conjugated PD1 mAb, and FITC/iRFP intensities were quantified by flow cytometry. iRFP-positive cells are displayed in histograms. Data are representative of three independent experiments. **c**, Flow cytometry histograms (FITC and iRFP) for single cells and iRFP+ cells from the experiment in (b) (PD1 SP, ± 1 μM KZR-9873). Data are representative of three independent experiments.

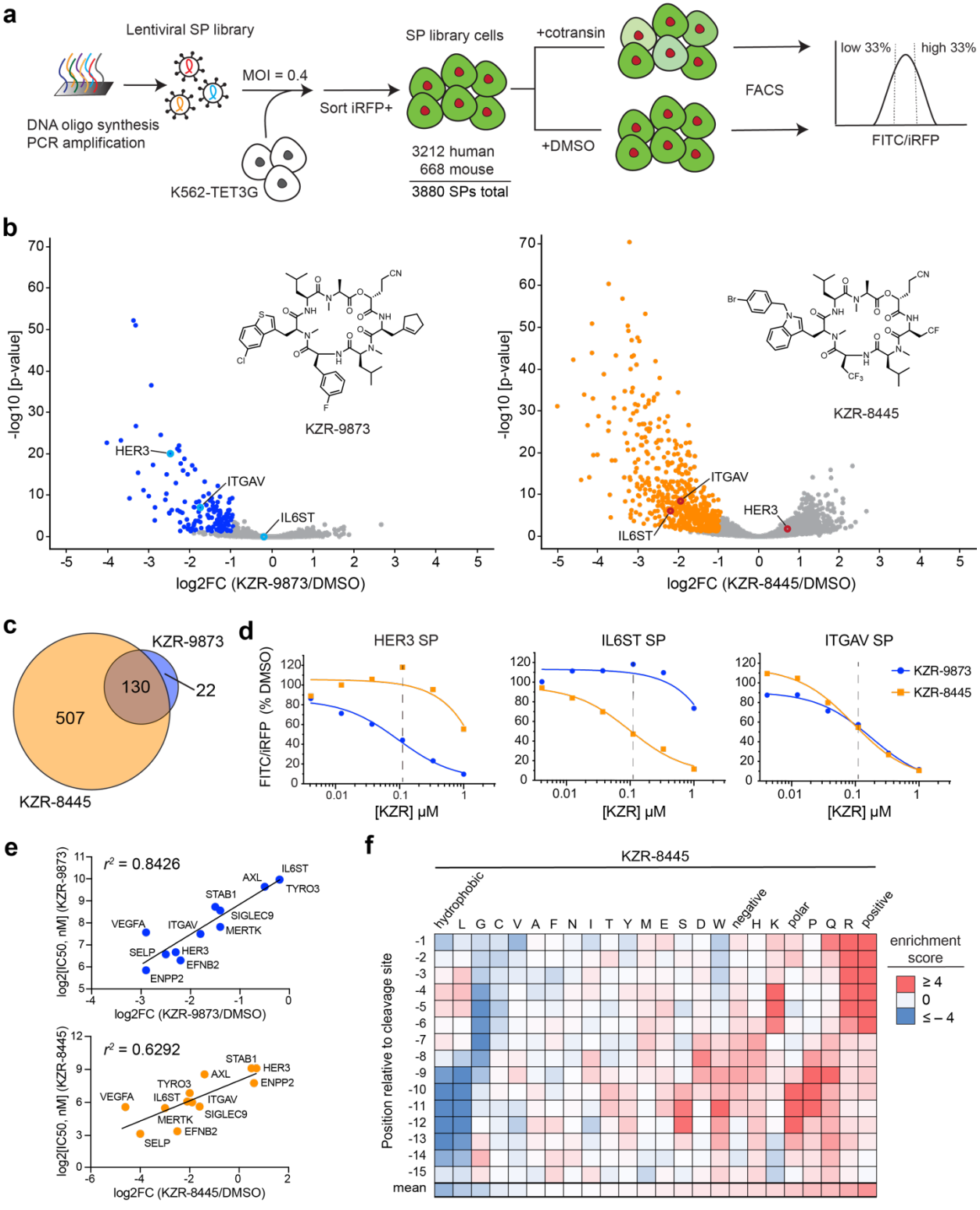


Figure 2.3: Massively parallel library screen identifies cotransin-sensitive SPs.
a, Schematic representation of SP library screen. K562-TET3G cells were transduced with the pooled library of 3880 human and mouse SPs. Stably transduced cells were treated with dox and DMSO, 100 nM KZR-9873, or (Figure caption continued on the next page)

(Figure caption continued from the previous page) 100 nM KZR-8445 for 24 h, then sorted into high and low FITC/iRFP populations (two biological replicates). SP read counts were determined by Illumina sequencing and DESeq2. **b**, Volcano plots showing log2FC relative to DMSO controls versus significance (-log10[P-adj]). Sensitive SPs: dark blue and orange dots. **c**, Venn diagram comparing sensitive SPs (**b**) for each cotransin. **d**, Individually cloned SP-reporter cells were treated with dox and DMSO, KZR-9873, or KZR-8445 for 24 h. FITC/iRFP ratios were quantified by flow cytometry. **e**, Scatter plots comparing log2FC from the SP library screen (**Supplementary Table 1**) versus log2(IC_{50}, nM) from individual SP-reporters (**Fig. 2.5a**). **f**, Heat map showing the statistical significance of amino acid enrichment or depletion at the indicated positions of KZR-8445-sensitive SPs. The enrichment scores shown are equal to -log10(P-value from Fisher's exact test) multiplied by either +1 or -1, depending on whether the aa (or aa category) is enriched or depleted, respectively, among 8445-sensitive SPs relative to 8445-resistant SPs (see Methods). Using the conservative Bonferroni correction, an enrichment score of > 3.86 or < -3.86 is considered statistically significant at $P < 0.05$.

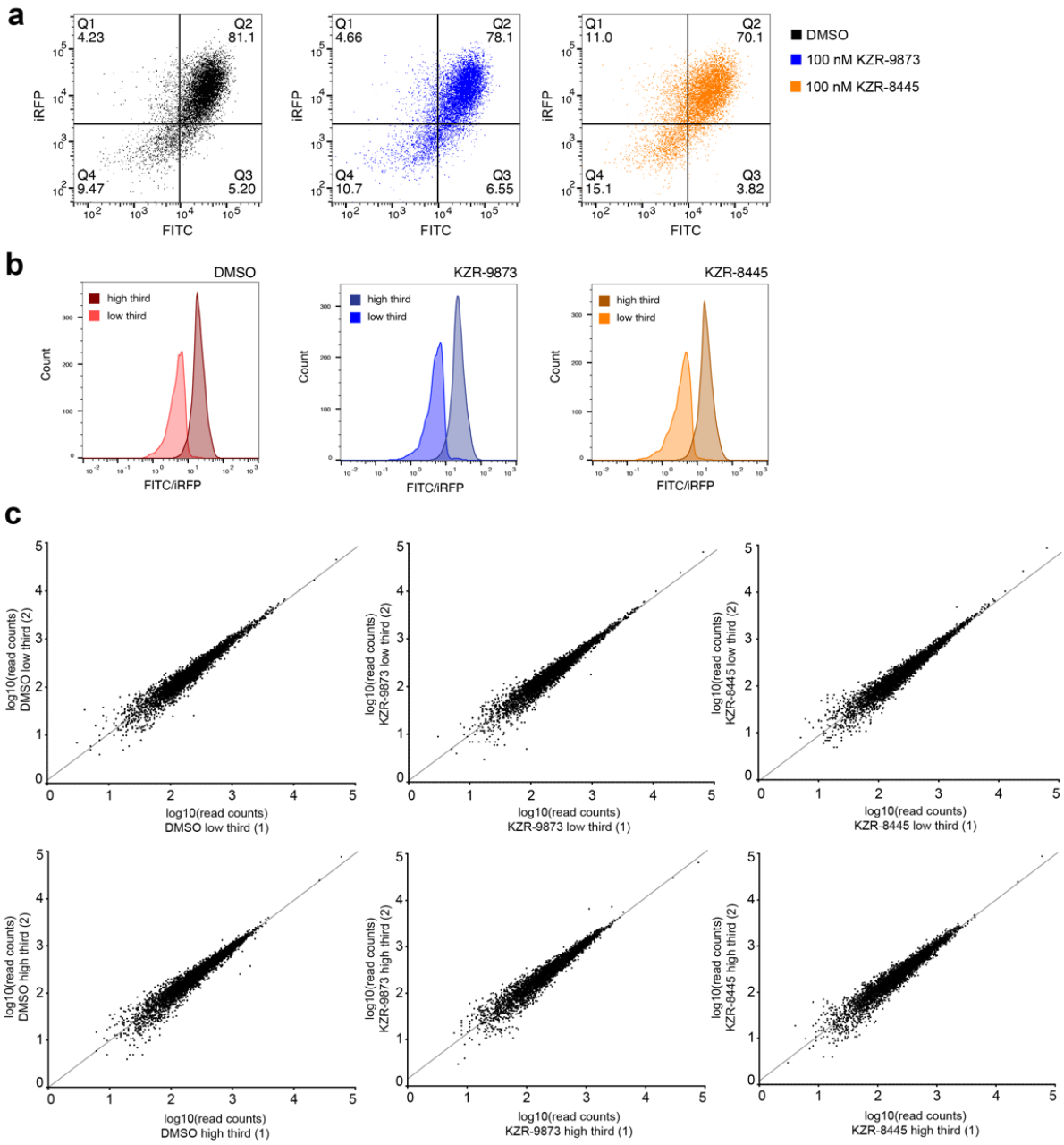
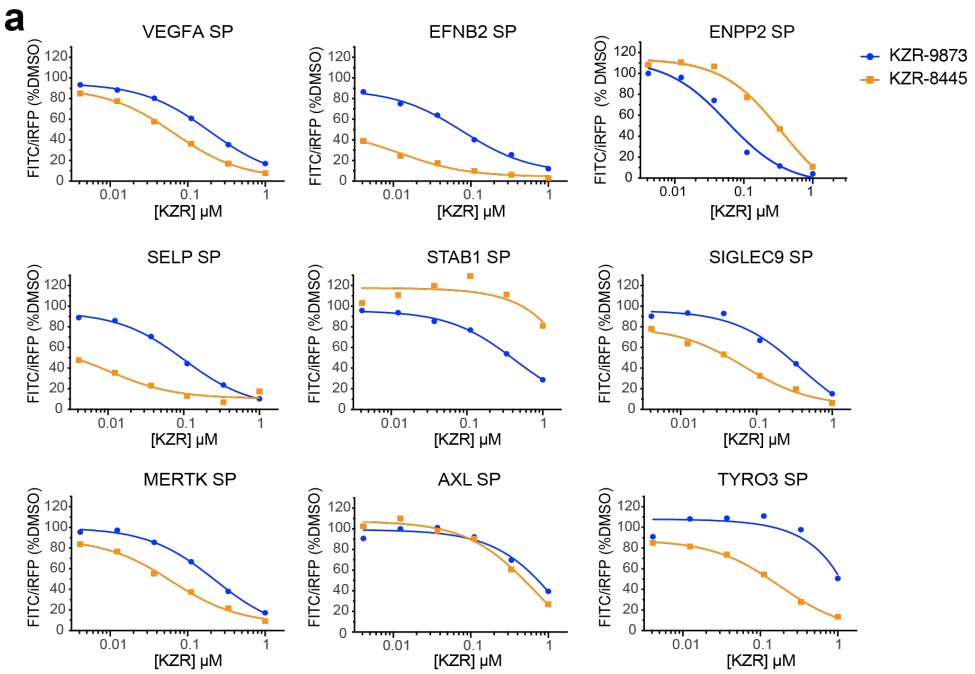
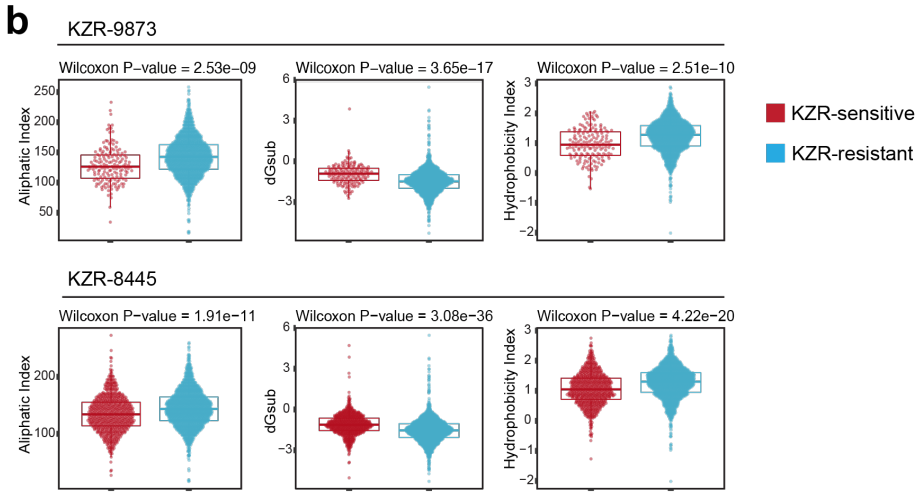


Figure 2.4: SP library screen.
a, Flow cytometry scatter plots depicting FITC and iRFP for the pooled-cell SP library treated with dox (250 ng/mL) and DMSO, 100 nM KZR-8445, or 100 nM KZR-9873 for 24 h. Data are representative of two independent replicates. **b**, Cells from (**a**) were sorted into high (top 33%) and low (bottom 33%) fluorescence populations based on the FITC/iRFP ratio and re-analyzed by flow cytometry. Histograms are representative of two biological replicates. **c**, SP sequencing read counts comparing biological duplicates of the sorted cell populations in (**b**).


a
VEGFA SP
EFNB2 SP
ENPP2 SP
SELP SP
STAB1 SP
SIGLEC9 SP
MERTK SP
AXL SP
TYRO3 SP
FITC/iRFP (%DMSO)
FITC/iRFP (% DMSO)
[KZR] μM
KZR-9873
KZR-8445


b
KZR-9873
Wilcoxon P−value = 2.53e−09
Wilcoxon P−value = 3.65e−17
Wilcoxon P−value = 2.51e−10
Aliphatic Index
dGsub
Hydrophobicity Index
KZR-sensitive
KZR-resistant
KZR-8445
Wilcoxon P−value = 1.91e−11
Wilcoxon P−value = 3.08e−36
Wilcoxon P−value = 4.22e−20

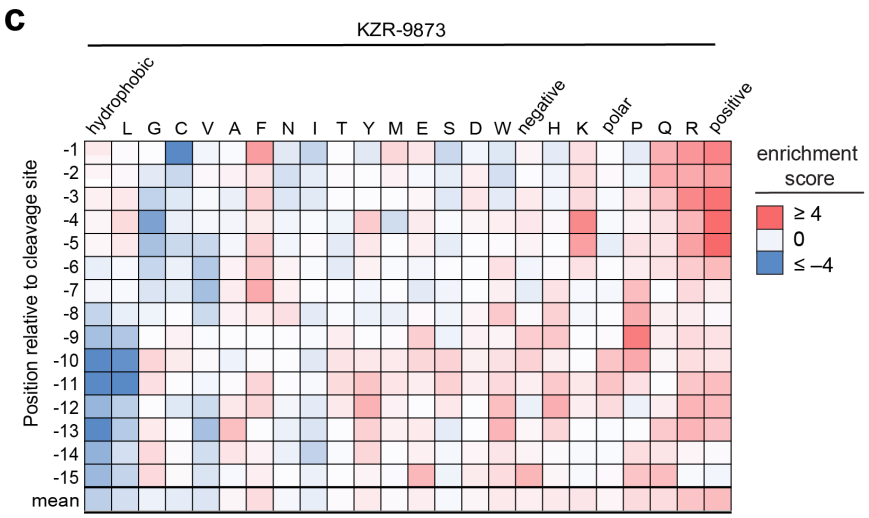

c
KZR-9873
hydrophobic
negative
polar
positive
L G C V A F N I T Y M E S D W H K P Q R
Position relative to cleavage site
-1 -2 -3 -4 -5 -6 -7 -8 -9 -10 -11 -12 -13 -14 -15
mean
enrichment score
≥ 4
0
≤ −4

Figure 2.5: Validation and analysis of the SP-reporter screen.
a, K562-TET3G cells were transduced with the indicated SP-reporters and treated with dox (250 ng/mL) and DMSO, KZR-9873, or KZR-8445. After 24 h, cells were stained with a FITC-conjugated PD1 mAb, and FITC/iRFP intensities were quantified by flow cytometry. Data points are plotted as FITC/iRFP ratios (%DMSO control). **b**, Comparison of biophysical features of sensitive and resistant SPs (see Methods). **c**, Heat map of amino acid enrichment scores at the indicated positions of KZR-9873-sensitive-SPs, calculated as described in **Fig. 2.3f** and Methods.

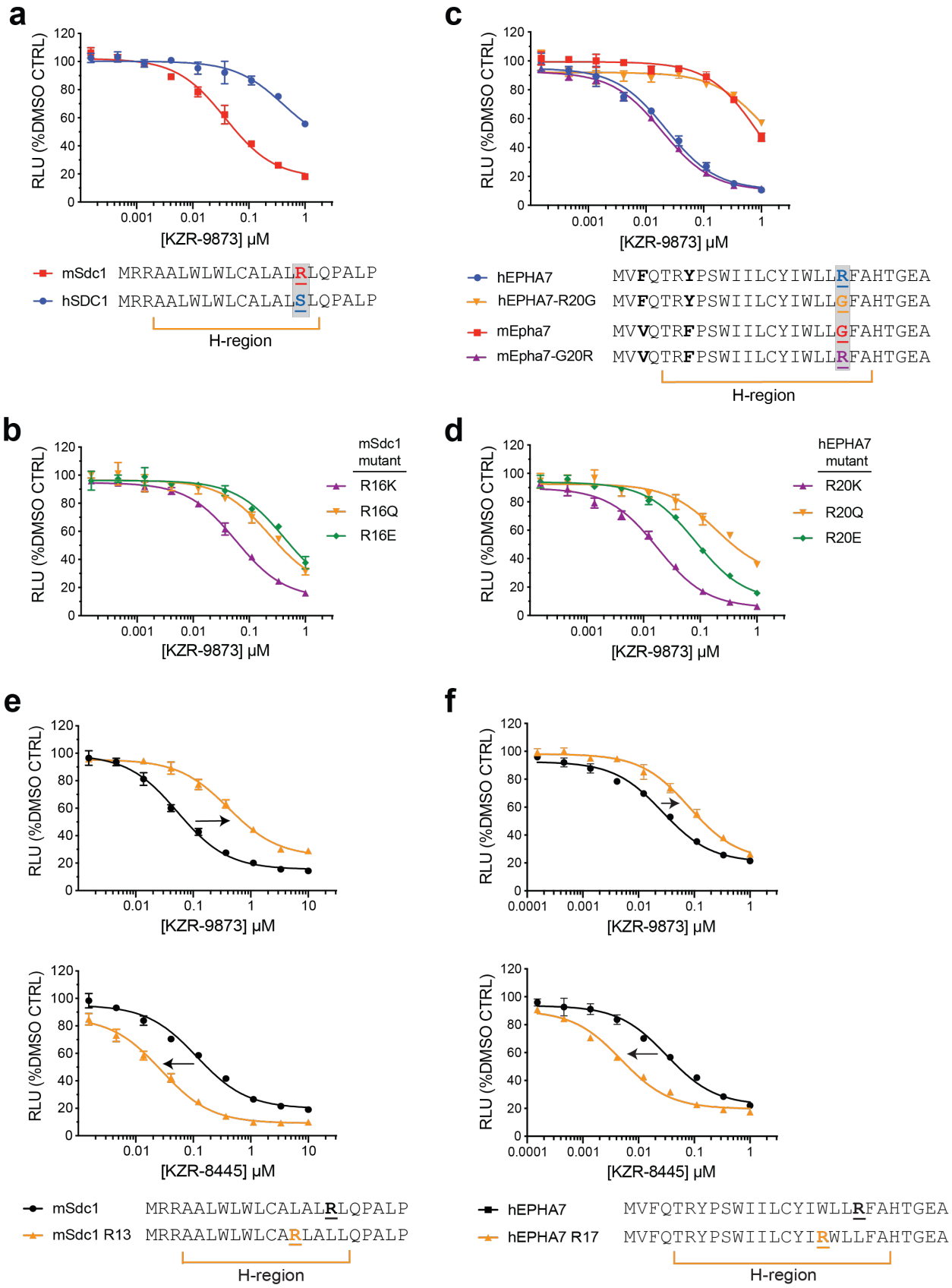

a
RLU (%DMSO CTRL)
[KZR-9873] μM
mSdc1 MRRAALWLWLCALALRLQPALP
hSDC1 MRRAALWLWLCALALSLQPALP
H-region
c
RLU (%DMSO CTRL)
[KZR-9873] μM
hEPHA7 MVFQTRYPSWIILCYIWLLRFAHTGEA
hEPHA7-R20G MVFQTRYPSWIILCYIWLLGFAHTGEA
mEpha7 MVVQTRFPSWIILCYIWLLGFAHTGEA
mEpha7-G20R MVVQTRFPSWIILCYIWLLRFAHTGEA
H-region
b
RLU (%DMSO CTRL)
[KZR-9873] μM
mSdc1 mutant
R16K
R16Q
R16E
d
RLU (%DMSO CTRL)
[KZR-9873] μM
hEPHA7 mutant
R20K
R20Q
R20E
e
RLU (%DMSO CTRL)
[KZR-9873] μM
RLU (%DMSO CTRL)
[KZR-8445] μM
mSdc1 MRRAALWLWLCALALRLQPALP
mSdc1 R13 MRRAALWLWLCARLALLQPALP
H-region
f
RLU (%DMSO CTRL)
[KZR-9873] μM
RLU (%DMSO CTRL)
[KZR-8445] μM
hEPHA7 MVFQTRYPSWIILCYIWLLRFAHTGEA
hEPHA7 R17 MVFQTRYPSWIILCYIRWLLFAHTGEA
H-region

Figure 2.6: Arg/Lys motif in the SP H-region confers sensitivity to KZR-9873.
a-f, HEK293 T-Rex cells were transfected with dox-inducible reporter constructs encoding the indicated SP-GLuc reporters. Transfected cells were treated with DMSO, KZR-9873, or KZR-8445 for 24 h, and secreted GLuc was quantified by luminescence. Data points represent mean RLU (relative luminescence units) as percent DMSO control ± s.d., n = 3. The SP H-region predicted by SignalP 6.0 is shown. All SP-GLuc comparisons of human/mouse pairs and related mutants were performed in the same experiment, and the data are representative of two independent experiments.

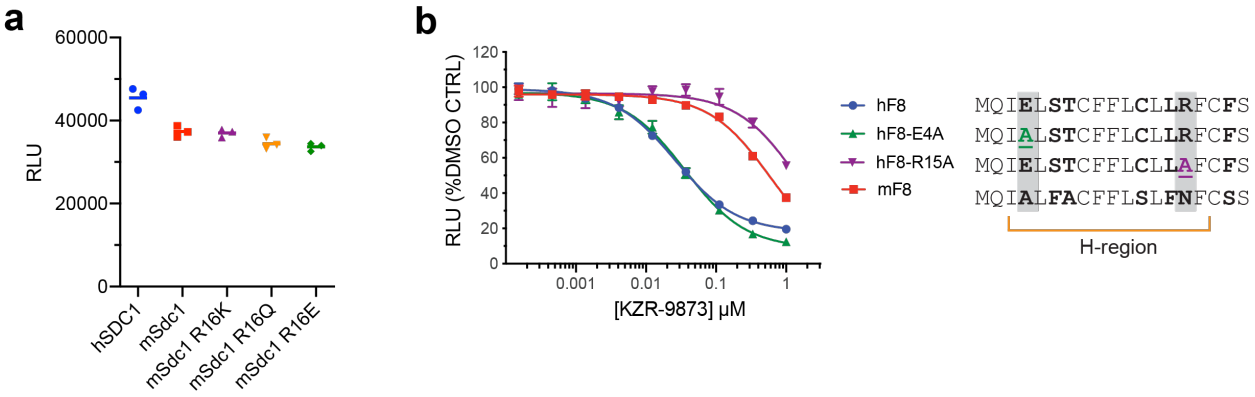


Figure 2.7: Arg/Lys motif in the SP H-region confers sensitivity to KZR-9873.
a, HEK293 T-Rex cells were transfected with the indicated SP-GLuc reporters in triplicate. Transfected cells were treated with dox and DMSO for 24 h, and GLuc secretion was quantified by luminescence. Relative luminescence units (RLU) from all 3 replicates are shown. **b**, HEK293 T-Rex cells were transfected with the indicated SP-GLuc reporters. Transfected cells were treated with dox and KZR-9873 for 24 h, and GLuc secretion was quantified by luminescence. Data points represent mean RLU (%DMSO control) ± s.d., n = 3. All SP-GLuc comparisons of human/mouse pairs and related mutants were performed in the same experiment, and the data are representative of two independent experiments.

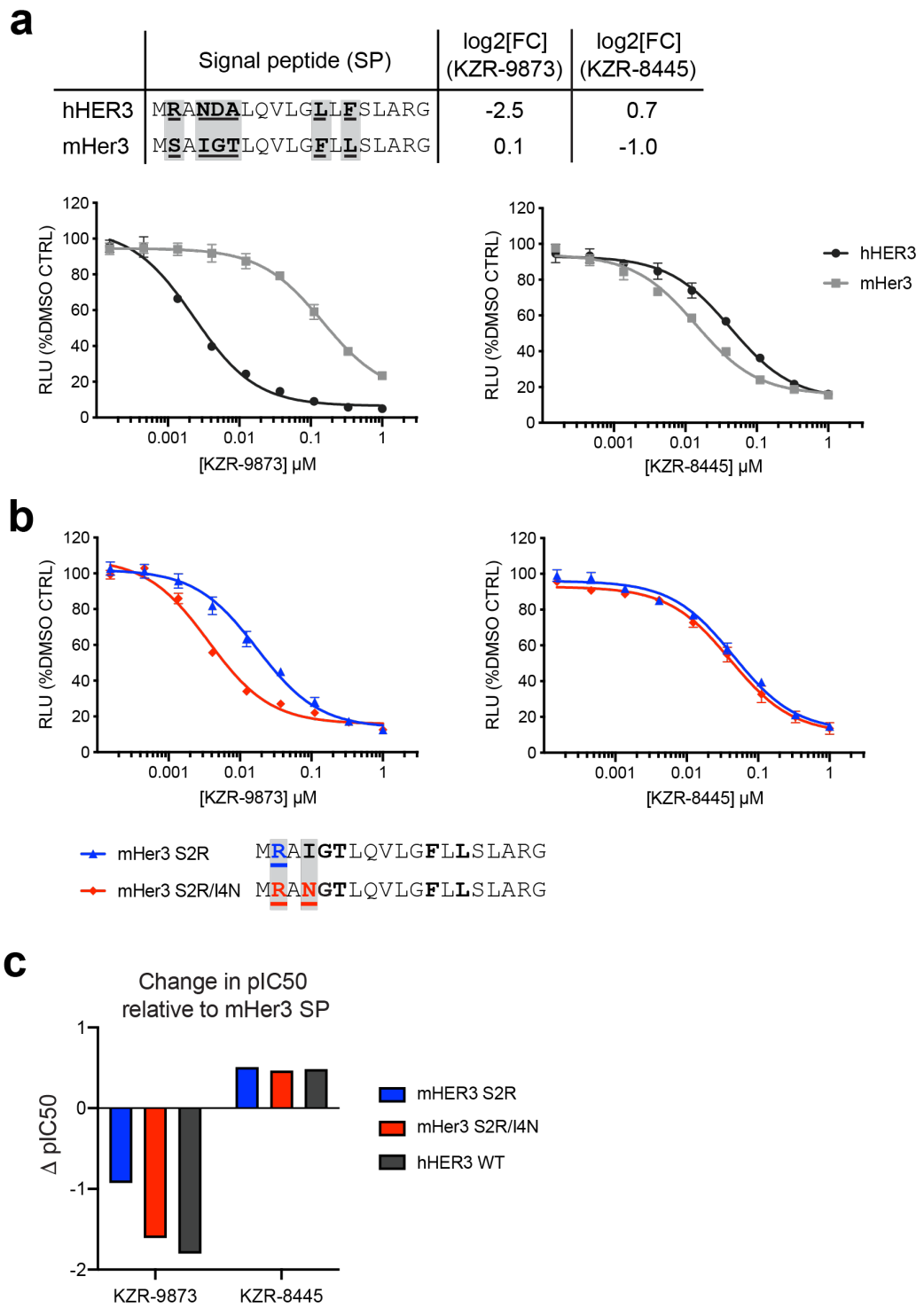


Figure 2.8: Comparison of human and mouse SPs reveals HER3 residues critical for sensitivity to KZR-9873.
a, hHER3 and mHer3 SP amino acid sequences and log2FC from the SP library screen. HEK293 T-Rex cells were transfected with the indicated SP-GLuc reporter constructs. Transfected cells were treated with dox and DMSO or the indicated concentrations of KZR-9873 for 24 h, and secreted GLuc was quantified (Figure caption continued on the next page)

(Figure caption continued from the previous page) by luminescence. **b**, HEK293 T-Rex cells were transfected and treated as in (**a**) with the indicated mutant SP-GLuc constructs. Data points represent mean RLU (%DMSO control) ± s.d., n = 3. **c**, Delta-pIC_{50} values (difference in -log10[IC_{50}]), relative to mHer3 SP, were calculated from dose-response curves in (**a**) and (**b**). All of the SP-GLuc comparisons were performed at the same time, and the data are representative of two independent experiments.

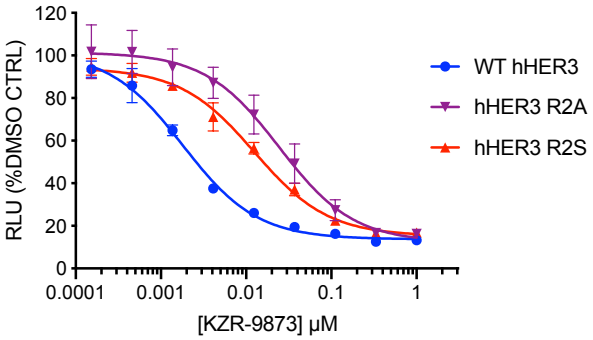


Figure 2.9: hHER3 sensitivity to KZR-9873 requires Arg2.
HEK293 T-Rex cells were transfected with the indicated SP-GLuc reporters. Transfected cells were treated with dox and KZR-9873 for 24 h, and GLuc secretion was quantified by luminescence. Data points represent mean RLU (%DMSO control) ± s.d., n = 3. The SP-GLuc comparisons were performed at the same time, and the data are representative of two independent experiments.

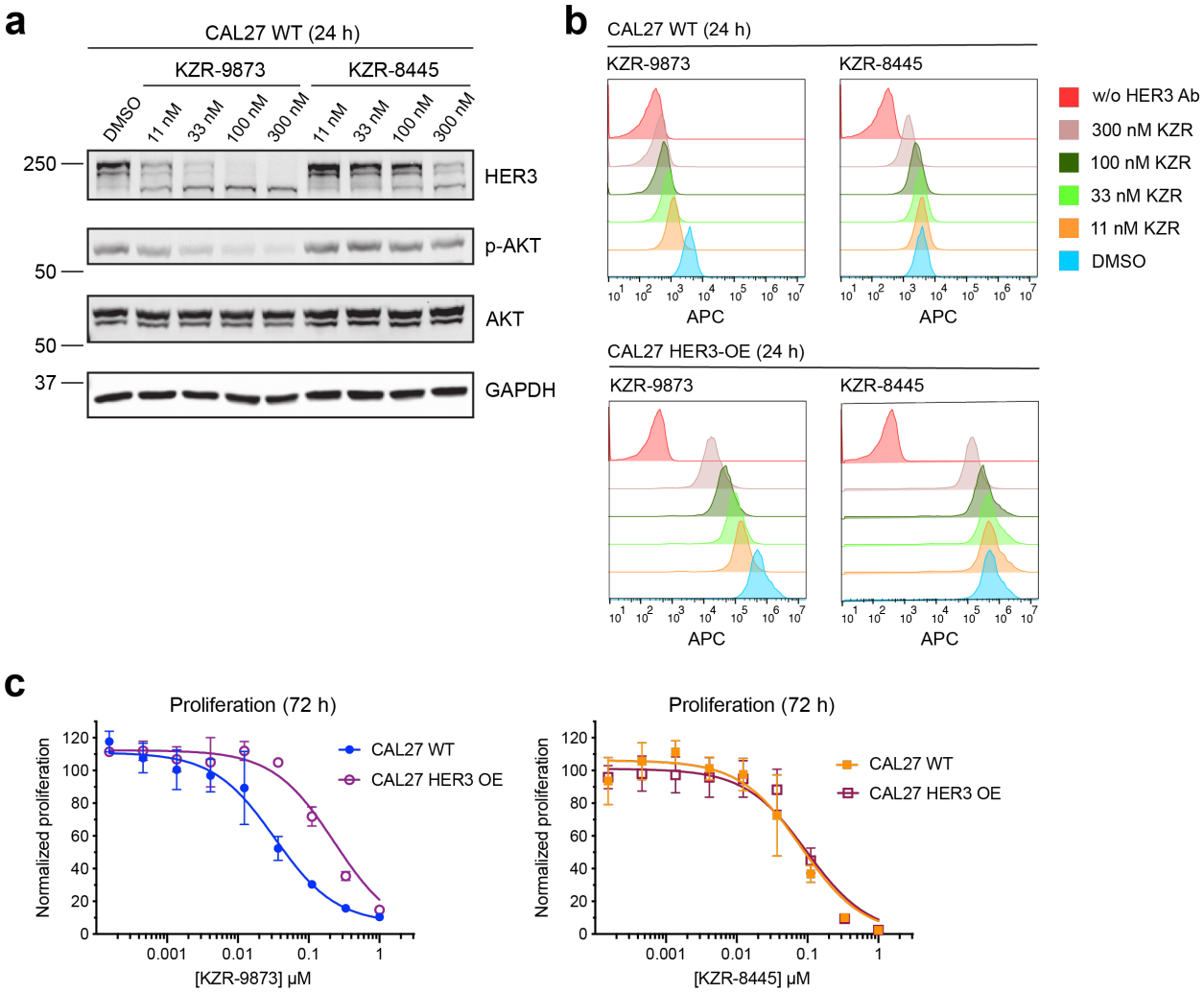


Figure 2.10: KZR-9873 blocks CAL27 cell proliferation via HER3.
a, CAL27 cells were treated with KZR-9873 and KZR-8445 for 24 h and analyzed by immunoblotting. **b**, WT and HER3-OE CAL27 cells were treated with KZR-9873 or KZR-8445 for 24 h. HER3 surface expression was quantified by flow cytometry with an APC-conjugated HER3 antibody. **c**, WT and HER3-overexpressing (HER3-OE) CAL27 cells were treated with KZR-9873 or KZR-8445 for 72 h. Cell proliferation was quantified using alamarBlue and plotted as percent DMSO control. Data points represent mean ± s.d., n = 3. These data (**a-c**) are representative of two independent experiments.

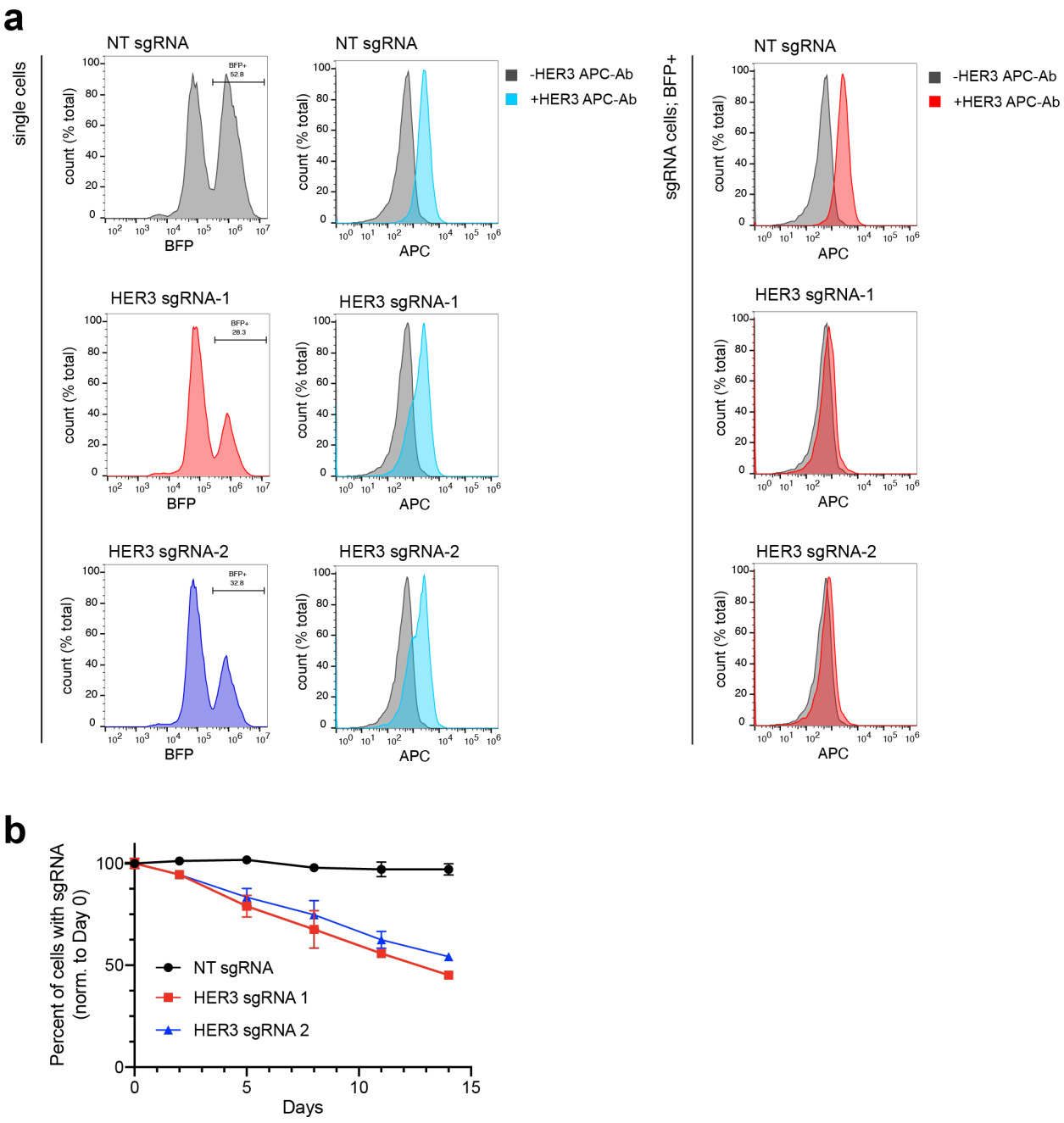


Figure 2.11: Effects of HER3 overexpression or knockdown in CAL27 cells.
a, CAL27-dCas9 cells were transduced with non-targeting (NT) or HER3 sgRNAs in pCRISPRia-v2 and analyzed by flow cytometry on day 0. Histograms depict all single cells or BFP+ (sgRNA containing) populations as indicated. Data are representative of two biological replicates. **b**, CAL27-dCas9 cells transduced with NT or HER3 sgRNAs were monitored by flow cytometry over 14 d. Data points show the mean percentage of BFP+ cells (sgRNA containing) ± s.d., n = 2.

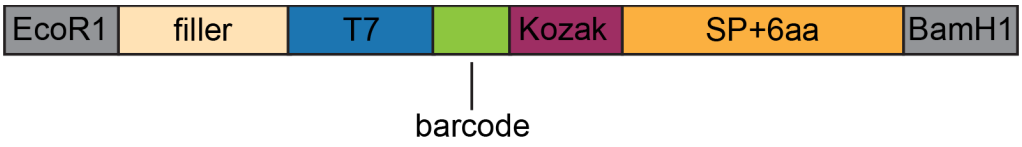


Figure 2.12: SP library oligonucleotide cassette design (250 nt total).

Chapter 3: Characterization of endogenous Sec61 client sensitivity to cotransins

Abstract

Biogenesis of one third of the human proteome, including therapeutically important secretory and membrane proteins, requires the Sec61 translocon. Each Sec61 client has a unique signal peptide or transmembrane domain, which plays an essential role in promoting its translocation into the endoplasmic reticulum. Cotransins are cyclic peptides that selectively inhibit Sec61-mediated translocation. Cotransin sensitivity has primarily been studied in the context of signal peptide-containing Sec61 clients. However, the analysis of cotransin-sensitive type II transmembrane and multi-spanning transmembrane protein is lacking. Here we use mass spectrometry to quantify the effects of KZR-9873 and KZR-8445, two distinct cotransin analogs, on Sec61 clients in three model cancer cell lines. We quantified 2212 Sec61 clients from our datasets and showed that, in addition to signal peptide-containing proteins, subsets of type II and multi-spanning transmembrane proteins are sensitive to cotransins. We also used DepMap dependency scores to provide a short list of candidate proteins, the reduced expression of which could potentially mediate the anti-proliferative effects of KZR-8445 in these three cancer cell lines.

Introduction

One third of the human proteome requires translocation into the endoplasmic reticulum (ER) for proper folding, localization, and function. Most secreted and membrane proteins are translocated via the Sec61 translocon. Nascent secretory and membrane proteins are directed to the ER when an N-terminal signal peptide or transmembrane segment emerges from the ribosome exit tunnel and is recognized by the signal recognition particle (SRP)[2]. The ribosome-nascent chain is then delivered to the Sec61 translocon, a heterotrimeric membrane protein complex, where the signal peptide or first transmembrane segment inserts through the lateral gate of Sec61α to access the lipid bilayer[60]. This mediates (1) translocation of extracellular and

lumenal protein domains into the ER lumen and (2) insertion of transmembrane domains into the ER membrane. There are more than 6000 predicted Sec61 clients in the human proteome, including most secreted proteins, single-spanning membrane proteins, and multi-spanning membrane proteins. However, tail-anchored type IV transmembrane proteins utilize the GET complex to insert into the ER membrane[61–65]. In addition, recent studies have revealed that some multi-spanning transmembrane proteins are translocated by the ER membrane complex (EMC)[51,65–69].

Previously, we and others have described a family of cyclic heptadepsipeptides, termed ‘cotransins’, which inhibit the cotranslational translocation of a subset of secreted and transmembrane proteins by binding directly to Sec61[6,7]. Cotransins prevent the insertion of the nascent protein's signal peptide (or signal anchor) into the lateral gate of Sec61 and thus prevents its translocation, resulting in proteasome-dependent degradation in the cytosol. Cotransin variants with altered side chains exhibit striking differences in signal peptide selectivity[40]. Sensitivity to cotransins is likewise determined by the specific amino acid sequence of the signal peptide, as shown by mutagenesis and signal peptide swapping experiments[6,7,11]. While Chapter 2 presents a comprehensive analysis of signal peptides and their cotransin sensitivity, a similar analysis of transmembrane segments has yet to be performed. Recently, cryo-EM structures of cotransins bound to Sec61 revealed insights into cotransin mechanism of action and showed that these compounds form critical interactions with the lateral gate and plug domain, which stabilize Sec61 in a closed conformation[28,29].

Many Sec61 clients are of therapeutic interest, including receptor tyrosine kinases and their ligands. However, the full range of cotransin-sensitive secreted and transmembrane proteins has not been defined. We previously showed that the cotransin analogs KZR-9873 and KZR-8445 have potent antiproliferative effects against three cancer cell lines, CAL27 (head and neck cancer), BxPC3 (pancreatic cancer), and SW48 (colorectal cancer), despite having distinct signal peptide sensitivity profiles. Therefore, we sought to define the sensitivity of endogenously

expressed Sec61 clients in these cell lines to gain a broader understanding of KZR-9873 and KZR-8445 selectivity profiles. We identified distinct subsets of Sec61 clients that were downregulated by KZR-9873 and/or KZR-8445 across the three cell lines. Finally, we used the DepMap gene dependency scores (https://depmap.org/portal/) to reveal potential mediators of KZR-8445's anti-proliferative effects.

Results

As described in Chapter 2, KZR-9873 and KZR-8445 inhibited the proliferation of BxPC3, CAL27, and SW48 cells with similar potencies (72-hour proliferation IC_{50} 30-75 nM) (**Fig. 3.1a**). Based on results from the signal peptide library, KZR-8445 and KZR-9873 appear to inhibit translocation mediated by overlapping yet distinct subsets of signal peptides. However, the signal peptide library approach does not provide a complete description of the differential pharmacology exhibited by KZR-9873 and KZR-8445. First, a large fraction of Sec61 clients do not have a cleavable N-terminal SP and yet are potentially affected by cotransins. Such clients are not represented in the SP library. Second, the intrinsic potency of KZR-9873 and KZR-8445 toward full-length endogenous SP-containing Sec61 clients may differ from the overexpressed SP reporter constructs. Here, we used 9-plex tandem mass tags (TMT) to quantify the effects of KZR-9873 and KZR-8445 on endogenous Sec61 clients in CAL27, BxPC3, and SW48 cells. Each cell line was separately treated with 100 nM KZR-9873, 100 nM-8445, or DMSO for 24 hours (three biological replicates, nine samples total for each cell line) (**Fig. 3.1b**, **Supplementary Table 3**). We then performed subcellular fractionation to enrich for Sec61 clients (many of which are low-abundance transmembrane proteins) in our samples. Treated cells were first incubated with digitonin buffer to extract cytosolic proteins, followed by complete membrane solubilization with Triton-X buffer. The extracted membrane and lumenal proteins were digested with trypsin and the tryptic peptides from each sample were TMT labeled, combined, and fractionated by preparative high-pH RP-HPLC. Concatenated fractions were

then analyzed by LC-MS/MS to quantify the effects of KZR-9873 and KZR-8445 on protein abundance relative to the DMSO-treated samples. To maximize quantitative accuracy, we used MS3 with real-time search to quantify the intensities of the 9 TMT reporter ions for each identified peptide[70,71].

Using the workflow described above, we quantified 7490, 7778, and 7844 proteins in the Triton-X fraction from CAL27, BxPC3, and SW48 cells, respectively (**Fig. 3.1c**). Of these proteins, 1635 (22%), 1748 (22%), and 1589 (20%) proteins from each respective cell line were classified as Sec61 clients based on their annotated features in the Uniprot database. This includes proteins annotated in two ways: (1) proteins with a predicted SP and (2) proteins with one or more predicted transmembrane, excluding mitochondrial and tail-anchored (type IV) membrane proteins, which do not translocate via Sec61.

Downregulated proteins were defined as those with log2FC of < -0.5 calculated from the ratio of TMT intensities averaged across replicates for the cotransin-treated and DMSO-treated samples. Log2FC values were considered significant with a P-value ≤ 0.05 (**Fig. 3.1d**). Based on these criteria, greater than 68% of sensitive proteins were Sec61 clients for each cotransin and each cell line (the highest percentage being 86%). Downregulated proteins are therefore highly enriched in Sec61 clients, given that they only constitute 20-22% of the total identified proteome in each cell line. This suggests that KZR-9873 and KZR-8445 distinctly alter the proteome, likely by directly inhibiting the cotranslational translocation of most of the downregulated proteins. Across the three cell lines, an average of 16% and 29% of the identified Sec61 clients were sensitive to KZR-9873 and KZR-8445, respectively, confirming that KZR-9873 is more selective.

We divided Sec61 clients and other ER-targeted transmembrane proteins into distinct topological classes based on feature annotation in the Uniprot database. In addition to proteins with annotated SPs (secreted/lumenal, type I single-spanning transmembrane, GPI-anchored, and a subset of multi-spanning transmembrane proteins), we classified transmembrane proteins

without annotated SPs as follows: (1) type II single-spanning transmembrane proteins (N-cyt, cytosolic N-terminus), (2) multi-spanning transmembrane proteins with N-cyt topology for their first transmembrane domain (multi-TM N-cyt), (3) type IV transmembrane proteins (Sec61 independent), (4) multi-spanning transmembrane proteins with N-exo topology for their first transmembrane domain (multi-TM N-exo, extracellular/lumenal N-terminus), and (5) proteins with unknown or unclear topology based on the Uniprot database (**Fig. 3.2a**).

The highest proportion of cotransin-sensitive Sec61 clients were SP-containing proteins, followed by type II and multi-TM N-cyt proteins. SP-containing proteins were highly enriched among cotransin-sensitive Sec61 clients (53-75% of cotransin-sensitive Sec61 clients have SPs vs. an average of 42% of the total Sec61 clients identified in each sample) (**Fig. 3.2a**). We further segmented the type of SP-containing proteins: (1) secreted or lumenal, (2) type I single-spanning transmembrane, and (3) multi-spanning transmembrane (**Fig. 3.2b**). Interestingly, the types of SP-containing Sec61 clients did not vary significantly by proportion between the total identified and cotransin-sensitive proteins.

Among the identified Sec61 clients that do not have an SP, type II TM proteins exhibited the greatest proportional sensitivity to cotransins. Up to ~40% of the identified type II TM proteins were sensitive, depending on the cell line (**Table 3.1**). Across the three cell lines, we identified a total of 51 and 81 type II TM proteins that were sensitive to KZR-9873 and KZR-8445, respectively (47 were sensitive to both cotransins). Moreover, we identified 34 multi-TM N-cyt proteins that were sensitive to KZR-9873 and 45 multi-TM N-cyt proteins that were sensitive to KZR-8445 (26 were sensitive to both cotransins). Only a few of the identified type IV and multi-TM N-exo proteins were found to be cotransin sensitive. This is expected since these TM proteins are not thought to utilize Sec61 for their initial insertion into the ER membrane[51,62,65,69]. The inhibitory effects of KZR-9873 and KZR-8445 on the biogenesis of these proteins may be indirect. Alternatively, cotransin-sensitive multi-TM N-exo proteins may be misannotated in Uniprot (they have an SP, or they may have the N-cyt topology). Finally, it is

possible that cotransins interfere with a poorly understood Sec61-dependent step during the biogenesis of certain multi-TM N-exo proteins (e.g., after insertion of the first TM).

By analyzing three cell lines, we were able to identify a much greater number of cotransin-sensitive and insensitive Sec61 clients than if we had analyzed only one cell line (**Fig. 3.3a**). Each cell line expresses a distinct subset of the total number of Sec61 clients encoded by the human genome. Nevertheless, certain shared Sec61 clients exhibited different cotransin sensitivities depending on the cell line. These differential cotransin sensitivities are likely due to differences in cell growth rates, protein turnover rates, and/or individual protein expression levels. Notably, CAL27 and SW48 cells exhibited a doubling time of 24 hours, whereas the doubling time for BxPC3 cells is 48 hours. We identified fewer sensitive Sec61 clients in BxPC3 cells, which may be attributed to their slower growth rate and hence, a decreased rate of protein biogenesis. The aggregate of sensitive Sec61 clients afforded a broader understanding of how KZR-9873 and KZR-8445 selectively alter the proteome.

The overlap of proteins sensitive to KZR-9873 and KZR-8445 based on our proteomics analysis resembles the overlap observed with the signal peptide library in Chapter 2. In CAL27 and SW48 cells, as with the SP library, most KZR-9873-sensitive Sec61 clients were also sensitive to KZR-8445 (**Fig. 3.3b**). For reasons that are not clear, a lower proportion of KZR-9873-sensitive clients (61 out of 142) were sensitive to KZR-8445 in BxPC3 cells. Nevertheless, the proteomics datasets from all 3 cell lines – based on endogenous, full-length proteins – demonstrate that KZR-9873 is more selective than KZR-8445, similar to the SP library dataset. Given this agreement, we wanted to directly compare our SP library and proteomics results. For SP-containing proteins that were identified in the proteomics datasets and included in the library, we compared their Log2FCs as determined in both assays relative to DMSO controls (**Fig. 3.4**). While there is a positive correlation between Log2FCs in the SP library and proteomics datasets, the correlations are weak (r^2 ~ 0.1, slopes 0.2-0.4). This is not unexpected and can likely be explained by one or more of the following: (1) Translocation of a given

endogenous full-length protein may be more or less *intrinsically* cotransin sensitive, as compared to the corresponding SP-reporter used in Chapter 2; (2) Proteins with slow translation and/or degradation rates are likely to produce small Log2FCs in the proteomics assay, even if they are cotransin sensitive; and (3) KZR-9873 and KZR-8445 at 100 nM strongly reduce CAL27, BxPC3, and SW48 cell proliferation, which is likely to globally reduce protein biogenesis rates and hence lead to "compressed" Log2FC values in the proteomics assay. By contrast, KZR-9873 and KZR-8445 at 100 nM have little or no effect on protein biogenesis in the SP library assay.

We tested whether specific amino acids are depleted or enriched at specific positions of TMs in cotransin-sensitive proteins that lack an SP (**Fig. 3.5a-c**). For this analysis, we aligned all type II TMs and the first TM of multi-TM N-cyt proteins (18-31 aa, median = 22 aa, as annotated in Uniprot) by their predicted C-terminal ends. Based on these alignments, we separately determined sequence logos for cotransin-sensitive and cotransin-resistant TMs. Whereas Leu is enriched throughout the TM in cotransin-resistant proteins, Leu is depleted near the C-terminal end of TMs from cotransin-sensitive proteins (positions 13-16). The C-terminal end of type II TMs and the first TM of multi-TM N-cyt proteins must orient toward the lumenal end of the Sec61 channel to insert into the membrane with the correct topology, similar to the orientation of SPs within Sec61. Based on our analysis in Chapter 2, the amino acids at the C-terminus of an SP play an important role in determining cotransin sensitivity. The TM analysis here is preliminary but it suggests that decreased hydrophobicity at the C-terminal end of TMs also plays a role in cotransin sensitivity. The competition model for cotransin sensitivity suggests that TMs with fewer hydrophobic residues are more likely to be sensitive, and the apparent depletion of Leu in cotransin-sensitive TMs is consistent with this model.

In Chapter 2, we showed that reduced translocation (and hence, expression) of HER3 can explain the antiproliferative effects of KZR-9873 in CAL27 cells. However, we did not uncover the mechanism of KZR-8445's antiproliferative effects. Given that KZR-8445 was a

broader inhibitor of SP translocation in the signal peptide library, we reasoned that KZR-8445 may induce proteotoxic stress in cells and/or inhibit the biogenesis of one or more broadly essential Sec61 clients. The proteomics datasets now provide a resource for identifying such clients. We used the gene dependency scores available from the DepMap database to determine if KZR-9873 or KZR-8445 downregulated proteins are essential in one or more of the cell lines tested (**Supplementary Table 3**). We were interested in genes that were more downregulated by KZR-8445 and had dependency scores greater than 0.9, which indicates reliance on that gene for cell growth.

There were only three Sec61 clients that met these criteria across the three cell lines tested. In CAL27 and SW48 cells, and to a somewhat lesser extent in BxCP3 cells, ATP6AP1 and ATP6AP2 emerged as strong candidates (**Table 3.2**). ATP6AP1, a subunit of the V-type ATPase proton pump that regulates acidification of vesicles in the cell, is an essential protein in eukaryotes[72,73]. ATP6AP2 is involved in the assembly of ATP6AP1 into the proton pump and was recently shown to play a role in the viability of pancreatic cancer cells[74,75]. Further research is needed to determine if cancer cells are selectively dependent on ATP6AP2 functions[76,77]. It is possible that KZR-8445 directly affects only ATP6AP2, but because ATP6AP1 depends on ATP6AP2, they are both downregulated. Another essential protein that showed greater sensitivity to KZR-8445 than KZR-9873 is SCAP (SREBP cleavage activating protein), which regulates the expression of genes involved in cholesterol biosynthesis and uptake (**Table 3.2**)[78,79]. The SREBP signaling pathway has emerged as a potential axis to treat cancer as it might connect lipid metabolism and tumor progression[80,81]. Follow-up experiments are required to determine if the effects revealed by the proteomics datasets are direct. Furthermore, the role of these proteins in mediating the anti-proliferative effects of KZR-8445 must be validated. It is unknown whether the level of downregulation we observed is sufficient to achieve the same effect as in the CRISPR knockout experiments used to determine the DepMap dependency score.

Discussion

In this study, we quantified the sensitivity of endogenous proteins to the cotransin analogs KZR-9873 and KZR-8445 in CAL27, BxPC3, and SW48 cells. Using subcellular fractionation and TMT-based quantification, we identified 2212 Sec61 clients across our three datasets. Of those identified, there were 360 KZR-9873-sensitive and 659 KZR-8445-sensitive Sec61 clients. The majority of downregulated proteins were Sec61 clients (68-86%), suggesting that the cotranslational translocation of most of these proteins is directly inhibited by cotransins. SP-containing and type II membrane proteins were especially sensitive to KZR-9873 and KZR-8445, as expected based on previous work[6,7,11,40]. Our discovery of 53 downregulated multi-TM N-cyt proteins substantially expands this class of cotransin-sensitive Sec61 clients. Based on published mechanistic studies, multi-TM N-cyt proteins engage Sec61 via their first TM in a manner that is similar to SPs and type II TMs. Therefore, cotransins can generally inhibit the biogenesis of proteins that utilize the lateral gate of Sec61 to initiate cotranslational membrane insertion, whereas type IV and multi-TM N-exo TM proteins, which do not insert via Sec61, are generally resistant to cotransins based on our data[51,61,62,69]. In addition, our proteomic analysis revealed candidate Sec61 clients that potentially mediate the anti-proliferative effects of KZR-8445. Comparison of DepMap dependency scores and Log2FCs from our proteomics datasets revealed only a few Sec61 clients that are essential across the three cell lines and downregulated by KZR-8445 to a greater extent than KZR-9873. These targets require follow-up studies to determine whether they are directly affected by KZR-8445 and to test whether their downregulation by KZR-8445 is sufficient to explain its anti-proliferative effects.

Despite the useful resource provided by our proteomic datasets, this study has certain limitations. First, we successfully identified only a third of the Sec61 clients annotated in the human genome, despite our subcellular fractionation protocol designed to enrich for transmembrane and lumenal proteins. Ideally, we would be able to quantify the effects of

cotransins on all Sec61 clients. However, many secreted and membrane proteins are expressed at low levels or in specific cell types and are therefore difficult to detect. Second, downregulated proteins may result from indirect effects; demonstrating direct effects of cotransins on the biogenesis of a given Sec61 client will require rigorous validation, for example by using inducible expression constructs in cells and in vitro translocation assays. Nevertheless, our study provides an unprecedented resource of SP-containing, type II and multi-TM N-cyt proteins that are sensitive to KZR-9873 and KZR-8445. These proteins can be further investigated as potential therapeutic targets.

Experimental Methods

Sample preparation for LC-MS/MS

CAL27 cells were seeded in 6-well plates at 500,000 cells per well in 2 mL complete growth medium. Seeded cells were treated with DMSO or increasing concentrations of KZR-9873 and KZR-8445 for 24 h, in triplicate. Treated cells were washed with ice-cold PBS and then incubated in 500 µL of digitonin lysis buffer (50 mM HEPES, pH 7.4, 100 mM KAc, 2.5 mM MgAc2, 150 µg/mL digitonin, and 2X complete EDTA-free protease inhibitor cocktail (Sigma-Aldrich, Cat # 11873580001)) at 4 °C for 10 minutes. The digitonin lysis buffer was removed and then cells were incubated in Triton X-100 lysis buffer (50 mM HEPES, pH 7.4, 500 mM KAc, 5 mM MgAc2, 1% Triton X-100, and 2X complete EDTA-free protease inhibitor cocktail (Sigma-Aldrich, Cat #: 11873580001)) at 4 °C for 15 minutes. Proteins were clarified by centrifugation (16,000g, 4 °C, 30 min). Disulfide reduction was performed with 10 mM DTT at 55 °C for 30 min, followed by alkylation with 40 mM iodoacetamide at room temperature (RT) for 30 min in the dark. In 15 mL tubes, proteins were precipitated by the addition of MeOH (4 mL), followed by CHCl3 (1 mL), and then H2O (3 mL)[82]. Samples were vortexed after each addition of solvent. The phases were separated by centrifugation (4500g, RT, 30 min). The protein disc formed in between the phases was washed carefully with MeOH (4 mL) and vortexed. The protein precipitate was pelleted by centrifugation (4500g, RT, 15 min), MeOH was removed, and the pellet air dried for 10 min. The pellet was resuspended in 100 µL of 50mM HEPES, 1mM CaCl2, and 2.5 µg trypsin and incubated overnight at 37 °C. Digestion was quenched by adding 1% formic acid. Peptide concentrations were quantified by peptide BCA assay (Thermo Fisher, Cat # 23275) and normalized to 0.5 µg/µL. 25 µg of each sample was carried forward.

TMT labeling and fractionation of tryptic peptides

TMT labeling was performed with the TMT10plex kit (Thermo Fisher, SK257743) according to the manufacturer's instructions with minor modifications. In brief, peptides were reconstituted in

10 µl of 30% MeCN in 200 mM HEPES buffer pH 8.5. TMT reagents were reconstituted in 20 µl of MeCN per vial, and 3 µL of this stock solution was added to each sample for 1 h at room temperature. Reactions were quenched by adding 2 µL of 5% hydroxylamine and incubated at room temperature for 15 min, followed by adding 3 µL of 5% trifluoroacetic acid (TFA) to acidify the solution. TMT-labeled samples were pooled and concentrated by SpeedVac to remove MeCN and desalted using Omix C18 Tips. Peptides were eluted with 50% MeCN, 0.1% TFA, and dried by SpeedVac. TMT-labeled samples were reconstituted in 240 µL of 20 mM ammonium formate, pH 10 and then fractionated via high pH RP chromatography using a Gemini 5 µm C18 110 Å, LC Column (150 x 4.6 mm, Phenomenex). The first 8 and the last 8 peptide-containing fractions were combined via pairwise concatenation to give 8 mixed fractions, while the middle 9 fractions were used as is. 17 total fractions were dried by SpeedVac.

LC-MS/MS analysis

Fractionated TMT-labeled tryptic peptides were reconstituted in 5% MeCN, 0.1% TFA in water, and analyzed on a Orbitrap Eclipse Tribrid Mass Spectrometer connected to an UltiMate 3000 RSLCnano system with 0.1% formic acid as buffer A, and 95% MeCN and 0.1% formic acid as buffer B. Peptides were separated on an EASY-Spray 3 µm, 75 µm × 15 cm C18 column with the following LC settings: flow rate at 0.3 µl min−1, loading samples at 4% B for 20 min, then 4–11% B over 2 min, 11–37% B over 73 min, 37–84% B over 2 min and finally 84% B for 3 min. The RTS-MS3 method was utilized for TMT reporter quantification. MS1 scans were acquired at a resolution of 120 K with an AGC target of 4 × 105, m/z scan range of 400–1,600, a maximum ion injection time of 50 ms, charge states of 2–8 and a 60 s dynamic exclusion time. MS2 spectra were acquired through collision-induced dissociation (CID) at a collision energy of 35%, in the ion trap with an AGC target of 1 × 104, isolation width of 0.7 m/z and maximum ion injection time set to ‘auto’. For real time search, MS2 spectra were searched against the Human

reviewed Swiss-Prot FASTA database with the digestion enzyme set to trypsin. Methionine oxidation was set as a variable modification; carbamidomethylation of cysteine and TMT modification were set as constant modifications. For MS3 acquisition, an SPS of 10 fragments was acquired in the Orbitrap for a maximum ion injection time of 105 ms with an AGC target of 1.5 × 105. MS3 spectra were collected at a resolution of 50 K with HCD collision energy of 65%.

Identification and TMT-based quantification

The LC-MS/MS data were searched using MaxQuant (v.1.6.7.0)[83] against the HUMAN reviewed Swiss-Prot FASTA database. Under “Group-specific parameters”, the “type” was set as “Reporter ion MS3”. The “Isobaric labels” was set as “9plex TMT” and the reporter ion isotopic distributions were incorporated to correct for impurities during synthesis of the TMT reagents according to the manufacturer’s specifications. Methionine oxidation and protein N-terminal acetylation were set as variable modifications, while carbamidomethylation of cysteine was set as a static modification. “Trypsin” was selected as the digestion enzyme with a maximum of 2 missed cleavages. All other parameters were set as default. Relative protein levels were determined by summing TMT reporter ion intensities across all corresponding PSMs (only PSMs corresponding to a unique protein match in the Swiss-Prot FASTA database were considered). TMT intensities in each channel were normalized such that the median TMT intensity values were equivalent across all nine channels. Mean protein intensities from each treatment arm (DMSO, n = 3; KZR-9873, n = 3; KZR-8445, n = 3) were log2-transformed and used to calculate the log2[fold change] between each treatment condition. P values were calculated using Student's t-test (two-tailed, two-sample equal variance).

Figures

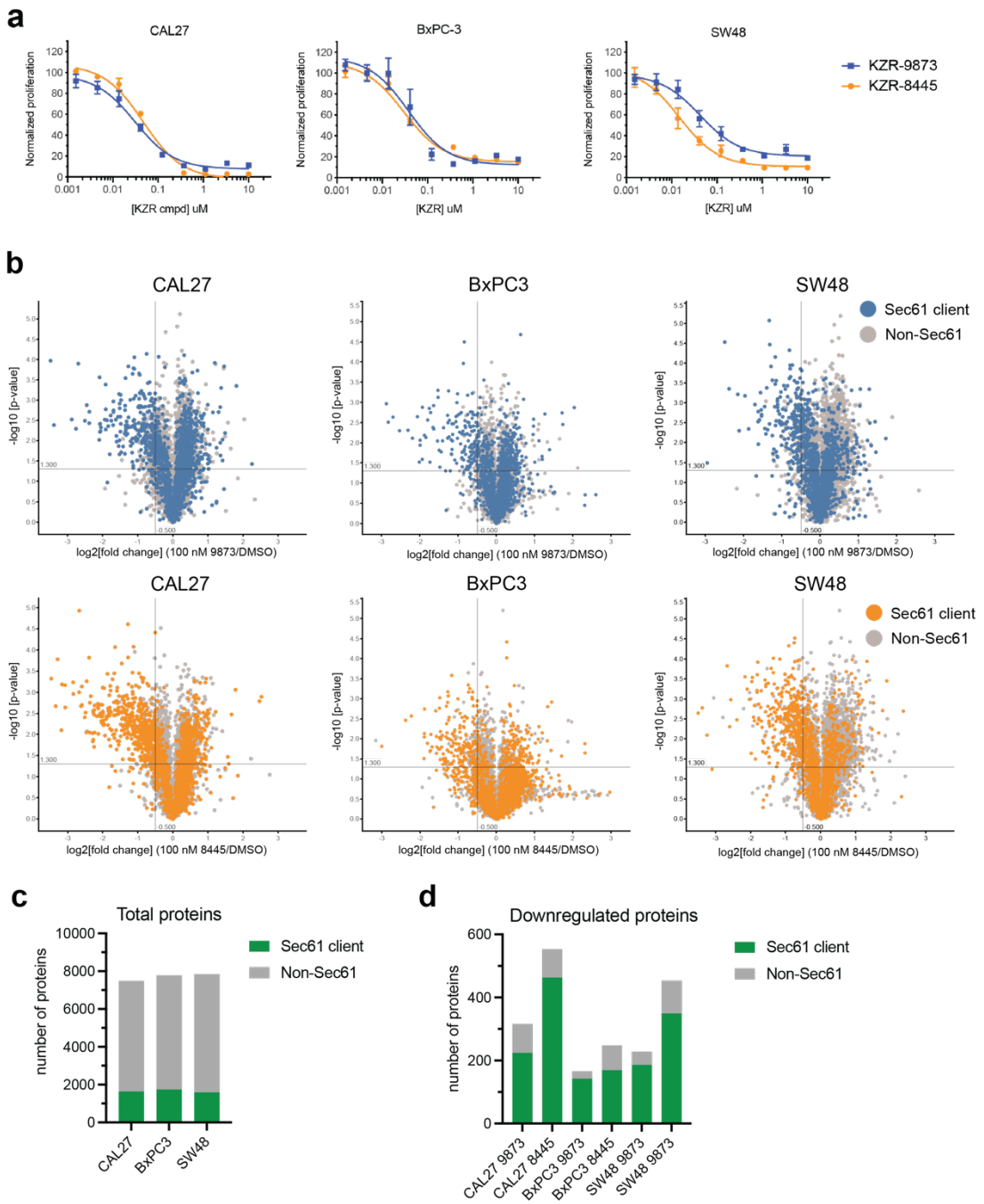


Figure 3.1: TMT proteomic analysis of Sec61 clients depleted in KZR-9873 and KZR-8445 treated cell lines.

a, The indicated cell lines were treated with DMSO or increasing concentrations of KZR-9873 and KZR-8445. After 72 h, cell proliferation was quantified using alamarBlue (%DMSO, mean ± s.d., n = 3). **b**, Volcano plots showing identified proteins with log2 fold change relative to DMSO controls versus significance (-log10[p-value]). P values were determined by Student's t-test (two-tailed, two-sample equal variance). (Figure caption continued on the next page)

(Figure caption continued from the previous page) Sec61 clients are indicated in blue (KZR-9873) or orange (KZR-8445). **c**, Total number of identified proteins. **d**, Number of downregulated proteins for KZR-9873 and KZR-8445 treated conditions (Log2FC < -0.5, P-value ≤ 0.05).

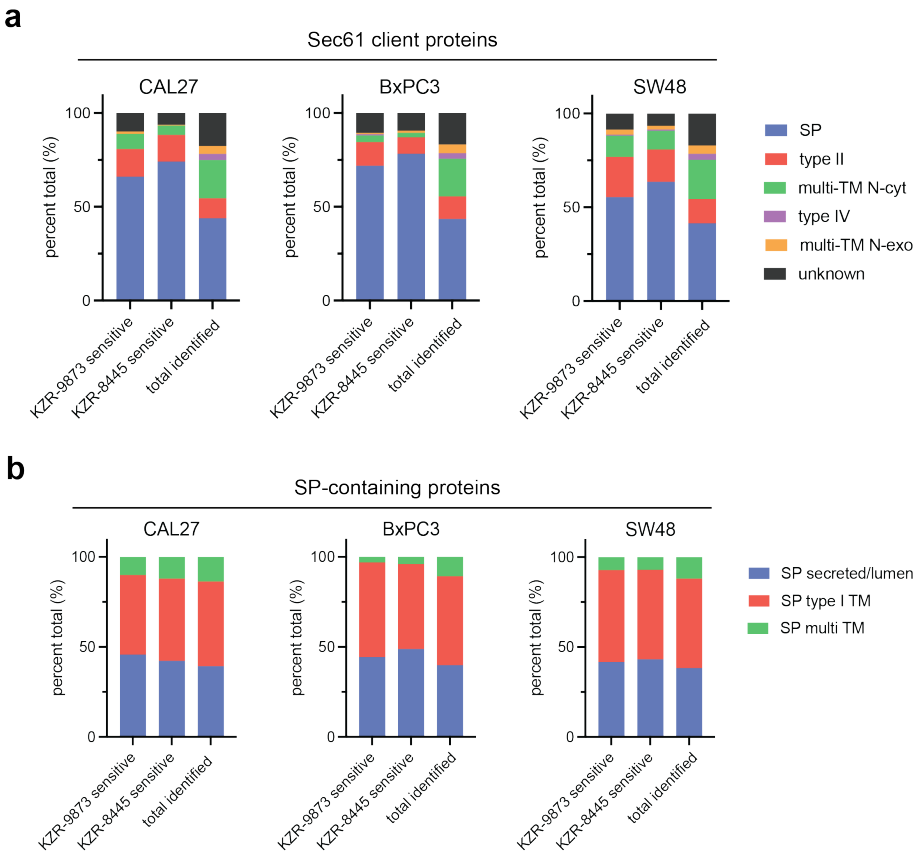


Figure 3.2: KZR-9873 and KZR-8445 sensitive Sec61 clients are enriched in proteins containing signal peptides.
a, Percentage of each Sec61 client type relative to the total number of Sec61 clients identified in the indicated category. **b,** Percentage of each SP-containing Sec61 client type relative to the total number of SP-containing Sec61 clients identified in the indicated category.

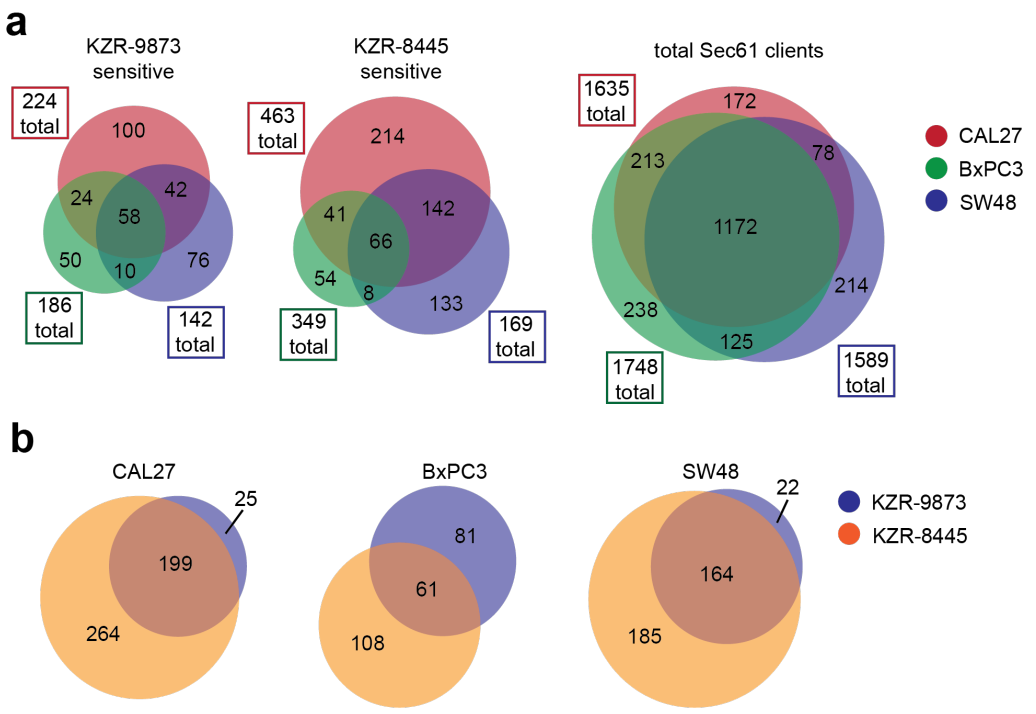


Figure 3.3: TMT proteomic analysis of three cell lines reveals unique sets of KZR-9873 and KZR-8445-senstive Sec61 clients.
a, Number of shared and unique cotransin-sensitive Sec61 clients, and the total identified in each cell line. **b**, Number of Sec61 clients sensitive to KZR-9873 and KZR-8445 in each cell line.

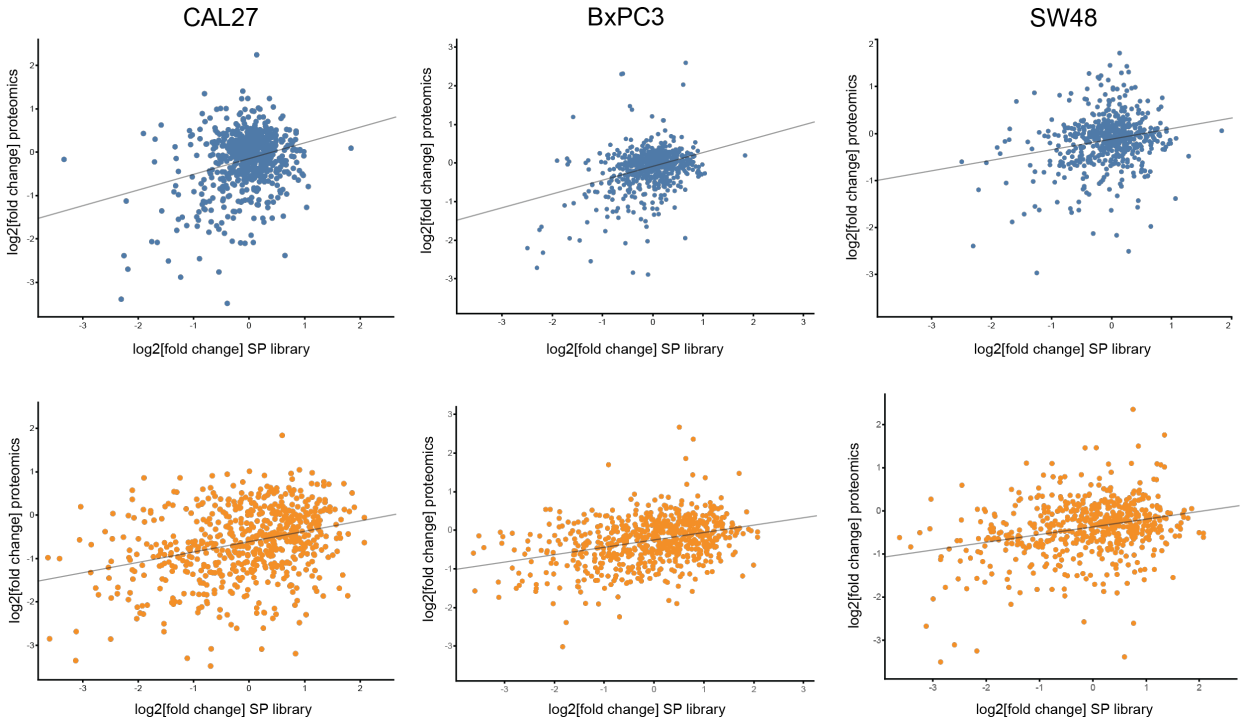


Figure 3.4: Log2[fold change] for SP-containing proteins included in the SP library and identified in the proteomics datasets.
Scatter plots comparing the Log2FCs from the SP library dataset with the SP-containing proteins identified in the proteomic datasets for CAL27, BxPC3 and SW48 cells. There are 649 shared SPs/proteins identified in CAL27 cells, 676 in BxPC3 cells, and 585 in SW48 cells. Top row: KZR-9873. Bottom row: KZR-8445.

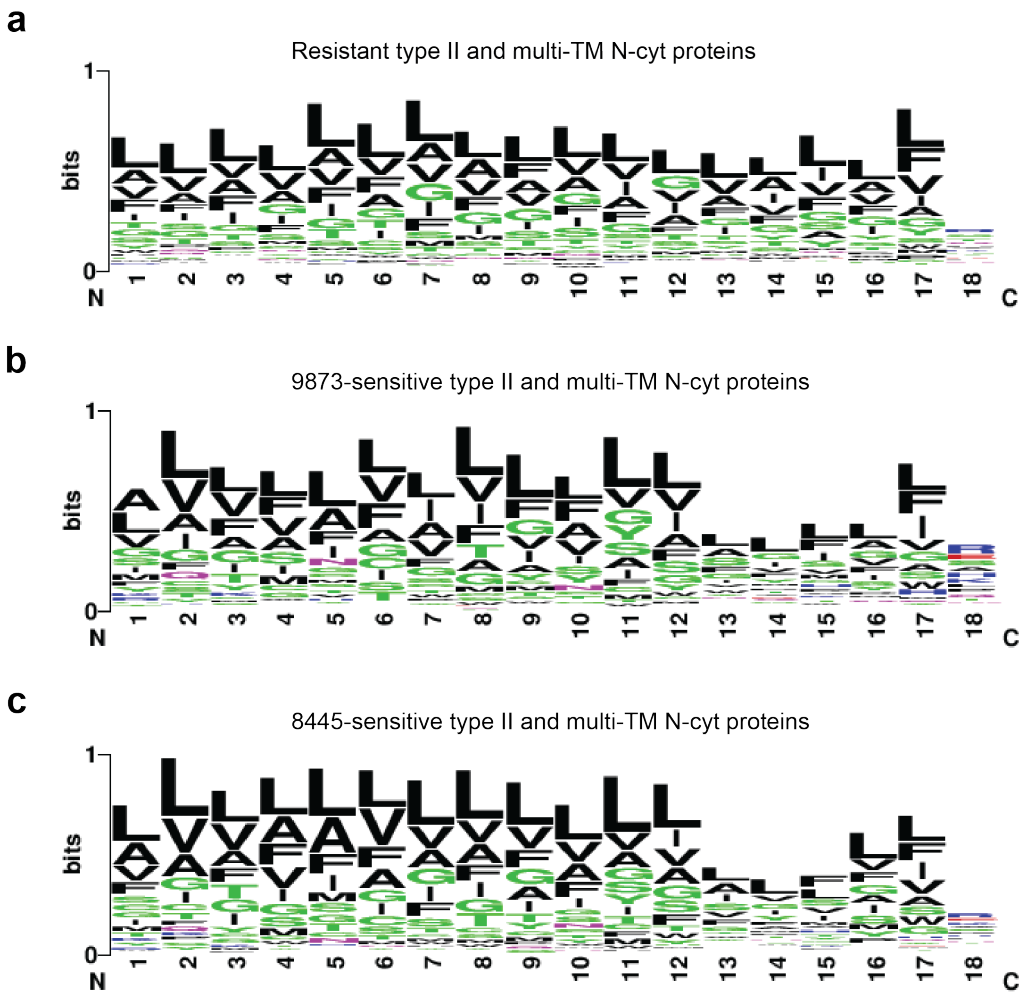


Figure 3.5: Amino acid differences in resistant and sensitive type II TMs and N-cyt multi-TMs.

a-c, Alignment of the first transmembrane domain of type II TMs and N-cyt multi-TMs for the indicated groups. TMs with at least 18 amino acids were included in the analysis, TMs were aligned by the C-terminus. (a) 477 TMs, (b) 85 TMs and (c) 126 TMs. Logo plots generated using https://weblogo.berkeley.edu/logo.cgi.

Tables

Table 3.1: Percentage of proteins identified within each Sec61 client category (SP, type II, etc.) that are sensitive to KZR-9873 or KZR-8445 in CAL27, BxPC3, and SW48 cells (LFC < -0.5, P-value ≤ 0.05).

	SP	Type II	Multi-TM N-cyt	Type IV	Multi-TM N-exo	Unknown
CAL27 KZR-9873 sensitive	21.6% (148)	19.9% (33)	5.6% (18)	0.0% (0)	4.5% (3)	8.0% (22)
CAL27 KZR-8445 sensitive	50.1% (343)	39.8% (66)	7.2% (23)	0.0% (0)	3.0% (2)	10.6% (29)
BxPC3 KZR-9873 sensitive	13.9% (102)	9.0% (18)	1.5% (5)	2.0% (1)	1.2% (1)	5.3% (15)
BxPC3 KZR-8445 sensitive	18.2% (133)	7.5% (15)	1.2% (4)	0.0% (0)	2.5% (2)	5.7% (16)
SW48 KZR-9873 sensitive	16.3% (103)	20.4% (40)	6.6% (21)	2.0% (1)	7.2% (5)	6.2% (16)
SW48 KZR-8445 sensitive	35.1% (222)	30.6% (60)	11.0% (35)	4.1% (2)	10.1% (7)	8.9% (23)

Table 3.2: DepMap dependency scores and Log2FCs (KZR-9873 and KZR-8445) for proteins identified in CAL27, BxPC3, and SW48 proteomics datasets.

	CAL27			BxPC3			SW48		
	DepMap	KZR-9873	KZR-8445	DepMap	KZR-9873	KZR-8445	DepMap	KZR-9873	KZR-8445
ATP6AP1	0.999	-0.3	-1.3	0.999	0.1	-0.3	0.999	-0.2	-1.1
ATP6AP2	0.980	-0.3	-1.4	0.994	0.2	-0.6	0.975	-0.2	-1.4
SCAP	0.998	-0.6	-1.1	0.999	-0.6	-1.0	0.994	-0.7	-1.0

References

1. Shao, S. & Hegde, R. S. Membrane protein insertion at the endoplasmic reticulum. *Annu Rev Cell Dev Biol* **27**, 25–56 (2011).
2. Egea, P. F., Stroud, R. M. & Walter, P. Targeting proteins to membranes: Structure of the signal recognition particle. *Curr Opin Struct Biol* **15**, 213–220 (2005).
3. Voorhees, R. M. & Hegde, R. S. Toward a structural understanding of co-translational protein translocation. *Curr Opin Cell Biol* **41**, 91–99 (2016).
4. Rapoport, T. A., Li, L. & Park, E. Structural and Mechanistic Insights into Protein Translocation. *Annu Rev Cell Dev Biol* **33**, 369–390 (2017).
5. Hommel, U. *et al.* The 3D-structure of a natural inhibitor of cell adhesion molecule expression. *FEBS Lett* **379**, 69–73 (1996).
6. Garrison, J. L., Kunkel, E. J., Hegde, R. S. & Taunton, J. A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum. *Nature* **436**, 285–289 (2005).
7. Besemer, J. *et al.* Selective inhibition of cotranslational translocation of vascular cell adhesion molecule 1. *Nature* **436**, 290–293 (2005).
8. Harant, H. *et al.* The translocation inhibitor CAM741 interferes with vascular cell adhesion molecule 1 signal peptide insertion at the translocon. *Journal of Biological Chemistry* **281**, 30492–30502 (2006).
9. Harant, H. *et al.* Inhibition of vascular endothelial growth factor cotranslational translocation by the cyclopeptolide CAM741. *Mol Pharmacol* **71**, 1657–1665 (2007).
10. MacKinnon, A. L., Garrison, J. L., Hegde, R. S. & Taunton, J. Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. *J Am Chem Soc* **129**, 14560–14561 (2007).
11. Mackinnon, A. L., Paavilainen, V. O., Sharma, A., Hegde, R. S. & Taunton, J. An allosteric Sec61 inhibitor traps nascent transmembrane helices at the lateral gate. *Elife* 1–23 (2014).

12. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381 (2005).
13. Vermeire, K. *et al.* CADA inhibits human immunodeficiency virus and human herpesvirus 7 replication by down-modulation of the cellular CD4 receptor. *Virology* **302**, 342–353 (2002).
14. Vermeire, K. *et al.* The Anti-HIV Potency of Cyclotriazadisulfonamide Analogs Is Directly Correlated with Their Ability to Down-Modulate the CD4 Receptor. *Mol Pharmacol* **63**, 203–210 (2003).
15. Bell, T. W. *et al.* Synthesis and structure-activity relationship studies of CD4 down-modulating cyclotriazadisulfonamide (CADA) analogues. *J Med Chem* **49**, 1291–1312 (2006).
16. Vermeire, K. *et al.* Signal Peptide-Binding Drug as a Selective Inhibitor of Co-Translational Protein Translocation. *PLoS Biol* **12**, (2014).
17. Van Puyenbroeck, V. *et al.* Preprotein signature for full susceptibility to the co-translational translocation inhibitor cyclotriazadisulfonamide. *Traffic* **21**, 250–264 (2020).
18. Luesch, H. & Paavilainen, V. O. Natural products as modulators of eukaryotic protein secretion. *Nat Prod Rep* **37**, 717–736 (2020).
19. Van Puyenbroeck, V. & Vermeire, K. Inhibitors of protein translocation across membranes of the secretory pathway: novel antimicrobial and anticancer agents. *Cellular and Molecular Life Sciences* **75**, 1541–1558 (2018).
20. Cao, S. *et al.* Ipomoeassin F, a new cytotoxic macrocyclic glycoresin from the leaves of Ipomoea squamosa from the Suriname rainforest. *Nat Prod Res* **21**, 872–876 (2007).
21. Zong, G. *et al.* Ipomoeassin F Binds Sec61α to Inhibit Protein Translocation. *J Am Chem Soc* **141**, 8450–8461 (2019).

22. Hall, B. & Simmonds, R. Pleiotropic molecular effects of the Mycobacterium ulcerans virulence factor mycolactone underlying the cell death and immunosuppression seen in Buruli ulcer. *Biochem Soc Trans* **42**, 177–183 (2014).

23. Baron, L. *et al.* Mycolactone subverts immunity by selectively blocking the Sec61 translocon. *J Exp Med* **213**, 2885–2896 (2016).

24. Paatero, A. O. *et al.* Apratoxin Kills Cells by Direct Blockade of the Sec61 Protein Translocation Channel. *Cell Chem Biol* **23**, 561–566 (2016).

25. Van Den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427**, (2004).

26. Voorhees, R. M., Fernández, I. S., Scheres, S. H. W. & Hegde, R. S. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell* **157**, 1632–1643 (2014).

27. Voorhees, R. M. & Hegde, R. S. Structure of the Sec61 channel opened by a signal sequence. *Science (1979)* **351**, 88–91 (2016).

28. Itskanov, S. *et al.* A common mechanism of Sec61 translocon inhibition by small molecules. *Nat Chem Biol* **19**, 1063–1071 (2023).

29. Rehan, S. *et al.* Signal peptide mimicry primes Sec61 for client-selective inhibition. *Nat Chem Biol* **19**, 1054–1062 (2023).

30. Trueman, S. F., Mandon, E. C. & Gilmore, R. A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. *Journal of Cell Biology* **199**, 907–918 (2012).

31. Gérard, S. F. *et al.* Structure of the Inhibited State of the Sec Translocon. *Mol Cell* **79**, 406–415 (2020).

32. Pauwels, E. *et al.* Structural insights into TRAP association with ribosome-Sec61 complex and translocon inhibition by a CADA derivative. *Sci Adv* **9**, (2023).

33. Liaci, A. M. *et al.* Structure of the human signal peptidase complex reveals the determinants for signal peptide cleavage. *Mol Cell* **81**, 3934-3948.e11 (2021).

34. Blobel, G. & Dobberstein, B. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* **67**, 835–851 (1975).

35. Von Heijne, G. Signal Sequences: The Limits of Variation. *J Mol Biol* **184**, 99–105 (1985).

36. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–1025 (2022).

37. Nielsen, H., Tsirigos, K. D., Brunak, S. & von Heijne, G. A Brief History of Protein Sorting Prediction. *Protein Journal* **38**, 200–216 (2019).

38. Hegde, R. S. & Kang, S. W. The concept of translocational regulation. *Journal of Cell Biology* **182**, 225–232 (2008).

39. Meyer, L. K. *et al.* Inhibition of the Sec61 translocon overcomes cytokine-induced glucocorticoid resistance in T-cell acute lymphoblastic leukaemia. *Br J Haematol* **198**, 137–141 (2022).

40. Maifeld, S. V. *et al.* Secretory protein profiling reveals TNF-α inactivation by selective and promiscuous sec61 modulators. *Chem Biol* **18**, 1082–1088 (2011).

41. Ruiz-Saenz, A. *et al.* Targeting HER3 by interfering with its Sec61-mediated cotranslational insertion into the endoplasmic reticulum. *Oncogene* **34**, 5288–5294 (2015).

42. Jura, N., Shan, Y., Cao, X., Shaw, D. E. & Kuriyan, J. Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3. *Proc Natl Acad Sci U S A* **106**, 21608–21613 (2009).

43. Baselga, J. & Swain, S. M. Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. *Nat Rev Cancer* **9**, 463–475 (2009).

44. Soltoff, S. P., Carraway, K. L., Prigent, S. A., Gullick, W. G. & Cantley, L. C. ErbB3 is involved in activation of phosphatidylinositol 3-kinase by epidermal growth factor. *Mol Cell Biol* **14**, 3550–3558 (1994).

45. Wang, Z. *et al.* Disruption of the HER3-PI3K-mTOR oncogenic signaling axis and PD-1 blockade as a multimodal precision immunotherapy in head and neck cancer. *Nat Commun* **12**, 1–13 (2021).

46. Wilson, T. R., Lee, D. Y., Berry, L., Shames, D. S. & Settleman, J. Neuregulin-1-Mediated Autocrine Signaling Underlies Sensitivity to HER2 Kinase Inhibitors in a Subset of Human Cancers. *Cancer Cell* **20**, 158–172 (2011).

47. Drilon, A. *et al.* Response to ERBB3-directed targeted therapy in NRG1 -rearranged cancers. *Cancer Discov* **8**, 686–695 (2018).

48. Gan, H. K. *et al.* A Phase I, First-in-Human Study of GSK2849330, an Anti-HER3 Monoclonal Antibody, in HER3-Expressing Solid Tumors. *Oncologist* **26**, e1844–e1853 (2021).

49. Devaraneni, P. K. *et al.* Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. *Cell* **146**, 134–147 (2011).

50. Vermeire, K. *et al.* Signal Peptide-Binding Drug as a Selective Inhibitor of Co-Translational Protein Translocation. *PLoS Biol* **12**, (2014).

51. Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S. & Hegde, R. S. EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell* **175**, 1507-1519.e16 (2018).

52. Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**, 1–20 (2016).

53. Martin, M. Cutadapt removes adapter sequences from high- throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).

54. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

55. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).

56. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. **433**, 377–381 (2005).

57. Hessa, T. *et al.* Molecular code for transmembrane-helix recognition by the Sec61 translocon. **450**, (2007).

58. Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239-255.e12 (2019).

59. Oltion, K. *et al.* An E3 ligase network engages GCN1 to promote the degradation of translation factors on stalled ribosomes. *Cell* **186**, 346-362.e17 (2023).

60. Gogala, M. *et al.* Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* **506**, 107–110 (2014).

61. Mariappan, M. *et al.* The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **477**, 61–69 (2011).

62. Hegde, R. S. & Keenan, R. J. Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **12**, 787–798 (2011).

63. Schuldiner, M. *et al.* The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane. *Cell* **134**, 634–645 (2008).

64. Guna, A., Volkmar, N., Christianson, J. C. & Hegde, R. S. The ER membrane protein complex is a transmembrane domain insertase. **473**, 470–473 (2018).

65. Hegde, R. S. & Keenan, R. J. The mechanisms of integral membrane protein biogenesis. *Nat Rev Mol Cell Biol* **23**, 107–124 (2022).

66. Hegde, R. S. The Function, Structure, and Origins of the ER Membrane Protein Complex. *Annu Rev Biochem* **91**, 651–678 (2022).

67. Shurtleff, M. J. *et al.* The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *Elife* **7**, (2018).

68. McGilvray, P. T. *et al.* An ER translocon for multi-pass membrane protein biogenesis. *Elife* **9**, 1–43 (2020).

69. Chitwood, P. J. & Hegde, R. S. The Role of EMC during Membrane Protein Biogenesis. *Trends Cell Biol* **29**, 371–384 (2019).

70. Erickson, B. K. *et al.* Active Instrument Engagement Combined with a Real-Time Database Search for Improved Performance of Sample Multiplexing Workflows. *J Proteome Res* **18**, 1299–1306 (2019).

71. Schweppe, D. K. *et al.* Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J Proteome Res* **19**, 2026–2034 (2020).

72. Wang, L., Wu, D., Robinson, C. V., Wu, H. & Fu, T. M. Structures of a Complete Human V-ATPase Reveal Mechanisms of Its Assembly. *Mol Cell* **80**, 501-511.e3 (2020).

73. Jansen, E. J. R. *et al.* ATP6AP1 deficiency causes an immunodeficiency with hepatopathy, cognitive impairment and abnormal protein glycosylation. *Nat Commun* **7**, (2016).

74. Guida, M. C. *et al.* ATP6AP2 functions as a V-ATPase assembly factor in the endoplasmic reticulum. *Mol Biol Cell* **29**, 2156–2164 (2018).

75. Taguchi, T. *et al.* ATP6AP2 is robustly expressed in pancreatic β cells and neuroendocrine tumors, and plays a role in maintaining cellular viability. *Sci Rep* **13**, (2023).

76. Sraer, G. N. *et al.* Pivotal role of the renin/prorenin receptor in angiotensin II production and cellular responses to renin. *J Clin Invest* **109**, 1417–1427 (2002).

77. Jung, Y. S. *et al.* TMEM9 promotes intestinal tumorigenesis through vacuolar-ATPase-activated Wnt/β-catenin signalling. *Nat Cell Biol* **20**, 1421–1433 (2018).

78. Nohturfft, A., Brown, M. S. & Goldstein, J. L. Topology of SREBP cleavage-activating protein, a polytopic membrane protein with a sterol-sensing domain. *Journal of Biological Chemistry* **273**, 17243–17250 (1998).

79. Nohturfft, A., Debose-Boyd, R. A., Scheek, S., Goldstein, J. L. & Brown, M. S. Sterols regulate cycling of SREBP cleavage-activating protein (SCAP) between endoplasmic reticulum and Golgi. *Proceedings of the National Academy of Sciences* **96**, 11235–11240 (1999).

80. Zhao, Q., Lin, X. & Wang, G. Targeting SREBP-1-Mediated Lipogenesis as Potential Strategies for Cancer. *Front Oncol* **12**, (2022).

81. Wen, Y. A. *et al.* Downregulation of SREBP inhibits tumor growth and initiation by altering cellular metabolism in colon cancer article. *Cell Death Dis* **9**, (2018).

82. Seneviratne, U. *et al.* Photoaffinity Labeling and Quantitative Chemical Proteomics Identify LXRβ as the Functional Target of Enhancers of Astrocytic apoE. *Cell Chem Biol* **28**, 148-157.e7 (2021).

83. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).

Supplementary Files

Supplementary Table 1

See supplementary data files.

Supplementary Table 2

See supplementary data files.

Supplementary Table 3

See supplementary data files.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Nicole Wenzell
1463EE9A9E5F49F...
Author Signature

8/29/2023
Date