Discovery of Small-Molecule Ligands of WDR Domain-Containing Proteins using a Deep Convolutional Neural Network

Tigran M. Abramyan[1], Irene Chau[2], Fengling Li[2], Sumera Perveen[2], Albina Bolotokova[2], Levon Halabelian[2], Ashley Hutchinson[2], Peter Loppnau[2], Santha Santhakumar[2], Almagul Seitova[2], Hong Zeng[2], Suzanne Ackloo[2], Niel Henriksen[1], Terry O'Brien[1], Denzil Bernard[1], Jon Sorenson[1], Dalia Barsyte-Lovejoy[2], Peter Brown[2], Masoud Vedadi[2], Cheryl Arrowsmith[2], Matthieu Schapira[2]

1. Atomwise Inc., San Francisco, CA, United States. 2. Structural Genomics Consortium, Toronto, ON, Canada.

Abstract

WD40 repeat (WDR) domains are some of the most abundant protein interaction domains in the human proteome, representing essential subunits of multiprotein complexes involved in a wide variety of cellular processes, ranging from cell signaling, growth, and division to protein degradation **[1,2]**. The domain has a beta-propeller structure with an overall doughnut shape. The central pore of the domain is typically involved in interactions with peptide regions of key interaction partner proteins. While central pores WDR domains can be druggable, often they can also be challenging to target with small drug-like molecules due to their sometimes shallow or hydrophilic nature. Here, using the AtomNet® model **[3]**, a deep convolutional neural network for structure-based drug discovery, we screened billions of compounds against WDR domain-containing proteins implicated in various illnesses such as heart disease and cancer **[4]**. We identified several small drug-like molecules exhibiting micromolar binding affinities validated by differential scanning fluorimetry (DSF), surface plasmon resonance (SPR), or biolayer interferometry (BLI) experiments. This effort will support the Target 2035 initiative of the Structural Genomics Consortium to discover molecular probes for nearly all human proteins (target2035.net).

Background

COP1 E3 ubiquitin-protein ligase

COP1 is an E3 ubiquitin ligase that targets transcription factors for degradation in mammals. COP1 has been implicated as both a tumor suppressor and a cancer promoter. Substrate recognition requires engagement of a binding motif directly (HY5, ETS, and c-Jun protein), or indirectly through an adaptor such as a protein of the Trib family, as for C/EBPα protein. It was demonstrated that COP1 binds its recognition motif on Trib1 protein using the top face of its WDR domain **[5].** Targeting this interface can reduce C/EBPα degradation by COP1 and potentially help treat acute myeloid leukemia **[6].**

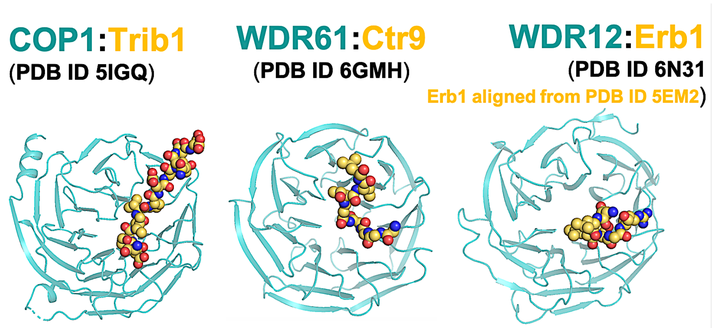
WDR61

WDR61 together with its partner protein, RNA polymerase-associated protein Ctr9, are components of the PAF1 complex, which associates with and regulates RNA polymerase II (Pol II). Small molecules that can inhibit WDR61:CtrR9 interaction may disrupt the transcriptional cycle of Pol II. While some studies linked WDR61 activity to dermatitis, the function of the transcription complex is still being explored **[7]**. Chemical probes could be utilized as tool compounds to further interrogate the function of WDR61 in gene regulation.

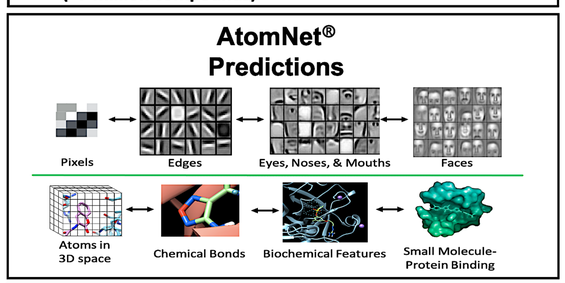
WDR12 ribosome biogenesis protein

WDR12 participates in ribosome biogenesis and cell proliferation. Genome-wide association studies have linked up-regulated WDR12 with early-onset myocardial infarction and coronary artery diseases in humans **[8]**. Additionally, WDR12 represents a potential anti-cancer target for its role in ribosomal biogenesis **[9]**. Structure of a homologous Ytm1 protein of E. coli in complex with Erb1 protein (PDB ID 5EM2) shows that in mammalian cells WDR12 may engage in protein-protein interactions with the top face of its WDR domain, providing a rationale for targeting this interface in drug discovery.

Methods



	COP1	WDR61	WDR12
Enamine REAL Space (~16 billion compounds)	X	X	
Mcule in-stock (~7 million compounds)	X	X	X

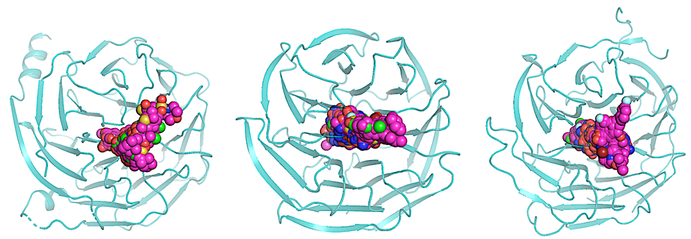


Filtering

1. Remove reactive & unwanted groups & PAINS
2. Lipinski's RO5 (MW, LogP, etc.)
3. Limited ADMET

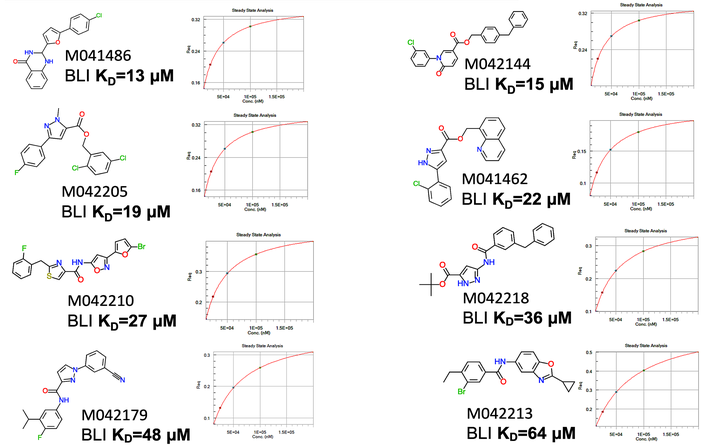
Clustering

450 compounds

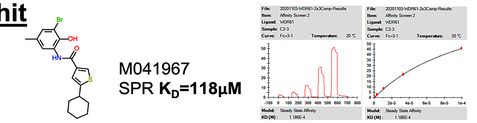


Results and Discussion

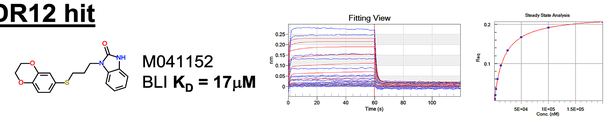
Representative COP1 hits (17 hits in total)



WDR61 hit



WDR12 hit



- The hits for WDR61 and WDR12 were identified in primary DSF assays followed with orthogonal SPR assays. COP1 hit compounds were determined with BLI assay.
- The WDR12 hit was validated with re-purified material in SPR and BLI assays and was advanced to hit expansion.
- Similarly, fresh resupplies of COP1 and WDR61 hits are being re-purified for validation.
- Some of the hits of interest are also being followed up with co-crystallization with the protein target to assist in structure-based hit optimization.

Acknowledgements

The SGC is a registered charity (no: 1097737) that receives funds from AbbVie, Bayer AG, Boehringer Ingelheim, Genentech, Genome Canada through Ontario Genomics Institute [OGI-196], the EU and EFPIA through the Innovative Medicines Initiative 2 Joint Undertaking [EUbOPEN grant 875510], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer, Takeda and the Wellcome Trust [106169/ZZ14/Z].

References

1. PMID: 29026209
2. PMID: 31666732
3. https://arxiv.org/abs/1510.02855
4. PMID: 28956604
5. PMID: 27041596
5. PMID: 23884858
6. PMID: 25915632
7. PMID: 16043514
8. PMID: 30135578