

# **Experimental Evaluation of Halogen-Enriched Fragment Library Reveals Halogen Bonding in Multiple, Novel Binding Motifs**

## **Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät  
der Eberhard Karls Universität Tübingen  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)

vorgelegt von  
Marcel Dammann  
aus Gütersloh

Tübingen  
2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	19.07.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Frank M. Böckler
2. Berichterstatter/-in:	Prof. Dr. Thilo Stehle
3. Berichterstatter/-in:	Prof. Dr. Anna K. H. Hirsch

# Acknowledgments

My first thanks go to **Professor Dr. Frank M. Böckler**, who made it possible for me, first as a master's student, then as a doctoral student, to get to know scientific research and fired my enthusiasm.

I am grateful to **Professor Dr. Thilo Stehle** for his support as my second examiner and for the opportunity to send my crystals to the SLS. My thanks for their willingness to act as my examiners go to **Professor Dr. Anna K. H. Hirsch**, **Professor Dr. Harald Groß**, and **Junior Professor Dr. Matthias Gehringer**.

**Dr. Markus Kramer**, **Priska Kolb**, and **Dominik Brzecki** were of great value for this dissertation, as they not only let me use hours after hours of measuring time at the NMR but always had an eye on my samples and gave helpful advice. A great resource of knowledge was **Dr. Georg Zocher**, who has an answer to every question regarding protein crystallography, data collection, and refinement. Without him, the structures in this study would be of much lesser quality.

If there was a Ph.D. in python scripting, **Dr. Markus Zimmermann** would have earned it years ago. Besides his contribution to the ESP-plots, he earned my gratitude by helping me in all things related to PyMol, MOE, or any other advice needed.

Susanne Hennig provided me with countless liters of buffers, media, and competent cells in the past years. Whenever I was confronted with a tedious task that would take hours, she readily jumped to the rescue. **W. Jason Stahlecker** came to our group for a master's internship and has stayed since. He contributed to the ITC measurements, and without him, it would have taken me months to set up all the crystallization in this dissertation. He had more expertise in protein crystallography and readily shared his knowledge with me. His work ethic and enthusiasm were the reason our results exceeded our expectations.

**Sebastian Vaas**, **Marc Engelhardt**, **Theresa Klett**, and **Martin Schwer** were not the colleagues I was stuck with, but the team I was hoping for. I am proud to have supervised the biochemistry and inorganic chemistry courses with you. The countless hours of hydrogen sulfide fragrance in the student's lab will always stay in my memory. Over the last decade, my fraternity brothers, especially **Juliander Reiner**, were the nearest thing to a family in Tübingen. I had the opportunity to experience and share some of my happiest and saddest moments with these people.

Lastly, I want to thank my girlfriend, **Luisa Kubin**. She had to take a back seat, especially when I was working late again during the last months. Without her, my life would be so much more unimaginative and colorless.

# Publications

## Journal Articles

1. F. Ansideri, M. Dammann, F. M. Boeckler, P. Koch  
Fluorescence polarization-based competition binding assay for c-Jun N-terminal kinases 1 and 2  
*Anal. Biochem.*, **2017**, 532, pp 26-28  
<https://doi.org/10.1016/j.ab.2017.05.022>
2. M. Dammann, M. Kramer, M. O. Zimmermann, F. M. Boeckler  
Quadruple Target Evaluation of Diversity-optimized Halogen-enriched Fragments (HEFLibs) Reveals Substantial Ligand Efficiency for AP2-Associated Protein Kinase 1 (AAK1). *Front. Chem.* **2022**, 9.  
<https://doi.org/10.3389/fchem.2021.815567>
3. M. Dammann, J. Stahlecker, T. Klett, M. Kramer, T. Stehle, F.M. Boeckler  
Screening of Halogen-Enriched Fragment Library Leads to Unconventional Binding Modes.  
*In Revision*
4. J. Stahlecker, T. Klett M. Schwer, S. Jaag, M. Dammann, L.N. Ernst, M. Braun, M.O. Zimmermann, M. Kramer, M. Laemmerhofer, T. Stehle, M. Coles, F.M. Boeckler  
Revisiting the Challenging Binding Site of T-p53C-Y220C by Employing a Diversity-Optimized HEFLib Reveals Diverse Binding Modes  
*Submitted*

## Poster Presentations

1. M. Dammann, M. U. Engelhardt, S. Vaas, M.O. Zimmermann, F. M. Boeckler  
HEFLibs (Halogen-Enriched Fragment Library): Screening and Hit Validation  
4<sup>th</sup> International Symposium on Halogen Bonding, Stellenbosch, South Africa
2. M. U. Engelhardt, M. O. Zimmermann, F.-M. Siemers, M. Dammann, S. Vaas, F. M. Boeckler  
Employing Machine Learning to establish QM-derived Scoring Functions for the Assessment of Halogen- $\pi$ -Interactions  
4<sup>th</sup> International Symposium on Halogen Bonding, Stellenbosch, South Africa

3. S. Vaas, M. O. Zimmermann, T. Klett, M. U. Engelhardt, M. Dammann, B. Drotleff, M. Lämmerhofer, F. M. Boeckler  
Synthesis and Evaluation of Fragments containing Halodifluoro Methyl Groups as Novel XB Donors  
4<sup>th</sup> International Symposium on Halogen Bonding, Stellenbosch, South Africa
4. M. Dammann, W. J. Stahlecker, M.U. Engelhardt, M. O. Zimmermann, F. M. Boeckler  
Halogen Bonding as a Main Contributor to Fragment Binding Modes  
HIPS Annual Symposium 2021, Saarbrücken, Germany

The test of a first-rate intelligence is the ability to hold two opposed ideas in the mind at the same time, and still retain the ability to function. One should, for example, be able to see that things are hopeless and yet be determined to make them otherwise.

F. Scott Fitzgerald, *The Crack-Up*, (1936) *Esquire*

# Table of Contents

<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 SCOPE OF THE STUDY .....	1
1.2 THE HALOGEN BOND (XB) .....	1
1.2.1 APPLICATIONS OF HALOGEN BONDS IN BIOCHEMISTRY .....	6
1.2.2 PROPERTIES AND TUNING OF HALOGEN BONDS .....	8
1.3. FRAGMENT-BASED DRUG DISCOVERY .....	10
1.3.1 HALOGEN-ENRICHED FRAGMENT LIBRARY .....	12
1.4. THREE-STAGE BIOPHYSICAL SCREENING .....	16
1.4.1 SATURATION TRANSFER DIFFERENCE NMR.....	16
1.4.2 ISOTHERMAL TITRATION CALORIMETRY.....	18
1.4.3 PROTEIN CRYSTALLOGRAPHY.....	19
1.5 PHARMACOLOGICAL CONTEXT OF TARGETS .....	21
<b>2. RESULTS AND DISCUSSION</b> .....	<b>27</b>
2.1 TARGET SELECTION .....	27
2.2 PROTEIN PURIFICATION .....	31
2.2.1 AAK1 PURIFICATION.....	31
2.2.2 BIRC5 PURIFICATION .....	34
2.2.3 CAMK1G PURIFICATION.....	35
2.2.4 DOT1L PURIFICATION .....	36
2.2.5 DYRK1A PURIFICATION .....	36
2.2.6 IDO1 PURIFICATION.....	37
2.2.7 JNK2 AND JNK3 PURIFICATION.....	39
2.2.8 GSK3B PURIFICATION .....	40
2.3 CHEMICAL STABILITY OF FRAGMENTS.....	42
2.4 STD-NMR.....	45
2.4.1 STD-NMR OPTIMIZATION .....	45
2.4.2 FRAGMENT SCREENING.....	49
2.4.3 LIBRARY ANALYSIS AND FRAGMENT PRIORITIZATION.....	57
2.4.5 KINASES IN THE SCREENING .....	70
2.4.6 IDO1 SCREENING.....	72
2.4.7 BIRC5 SCREENING .....	73



2.5 HIT CHARACTERIZATION BY ITC .....	74
2.5.1 FRAGMENT VALIDATION OVERVIEW .....	74
2.5.3 JNK2 FRAGMENT VALIDATION .....	78
2.5.4 DYRK1A FRAGMENT VALIDATION .....	80
2.5.5 CAMK1G FRAGMENT VALIDATION .....	82
2.5.6 AAK1 FRAGMENT VALIDATION.....	83
2.5.7 IDO1 FRAGMENT VALIDATION.....	85
2.5.8 PROBLEMATIC FRAGMENT VALIDATION .....	85
2.5.9 RANKING OF FRAGMENTS .....	87
2.5.10 STRUCTURE-AFFINITY RELATIONSHIP OF 9595 WITH AAK1 .....	92
2.5.11 STRUCTURE-AFFINITY RELATIONSHIP OF 0482 WITH DYRK1A.....	97
2.5.12 STRUCTURE-AFFINITY RELATIONSHIP OF 9612 WITH JNK3 AND JNK2.....	100
2.5.13 COMPARISON OF GATEKEEPER MUTANTS OF JNK3 WITH 9612.....	102
2.6. CRYSTALLIZATION .....	105
2.6.1 DYRK1A CRYSTALLIZATION .....	105
2.6.2 KUROM118 CRYSTAL .....	108
2.6.3 FRAGMENT 0482 CRYSTAL .....	110
2.6.4 FRAGMENT CRYSTALLIZATION OF JNK .....	122
2.6.5 DEVELOPMENT OF THE FRAGMENT 9612.....	138
<b>3 CONCLUSION .....</b>	<b>142</b>
3.1. OUTLOOK.....	145
<b>4. MATERIALS AND METHODS .....</b>	<b>147</b>
4.1 MEDIA PREPARATION .....	147
4.2 BUFFER PREPARATION.....	148
4.3 COMPETENT CELL PREPARATION .....	148
4.4 DNA PREPARATION .....	149
4.4.1 HEAT SHOCK TRANSFORMATION .....	149
4.4.2 MUTAGENESIS .....	149
4.4.3 CLONING.....	151
4.5 PROTEIN EXPRESSION AND PURIFICATION .....	152
4.5.1 SDS-PAGE.....	152
4.5.2 TEV .....	153
4.5.3 AAK1 / CAMK1G.....	153

4.5.4 JNK2 .....	154
4.5.5 JNK3 .....	154
4.5.6 DOT1L.....	155
4.5.7 DYRK1A .....	155
4.5.8 IDO1 .....	156
4.5.9 BIRC5.....	157
4.5.10 GSK3.....	157
4.5.11 INCLUSION BODY PREPARATION.....	157
4.5.12 WESTERN BLOT AND ELISA.....	158
4.6 FRAGMENT LIBRARY PREPARATION .....	159
4.7 STD-NMR.....	159
4.8 ITC .....	160
4.9 GSH STABILITY ASSAY .....	160
4.10 CRYSTALLIZATION.....	161
4.10.1 DYRK1A .....	161
4.10.2 JNK3 .....	162
4.10.3 DATA COLLECTION AND PROCESSING.....	162
4.12 FIGURES AND TABLES .....	163
4.13 AUTHOR CONTRIBUTIONS .....	163
<b>5 ZUSAMMENFASSUNG / SUMMARY (GERMAN).....</b>	<b>164</b>
<b>6. LITERATURE.....</b>	<b>170</b>
<b>7 APPENDIX.....</b>	<b>192</b>
APPENDIX A: DNA CONSTRUCTS.....	192
APPENDIX B: STD-NMR SPECTRA .....	197
APPENDIX C: HEFLIB PROPERTIES.....	205
APPENDIX D: PERMISSIONS.....	212

# Table of Figures

FIGURE 1.....	2
FIGURE 2.....	7
FIGURE 3.....	9
FIGURE 4.....	10
FIGURE 5.....	13
FIGURE 6.....	14
FIGURE 7.....	17
FIGURE 8.....	19
FIGURE 9.....	20
FIGURE 10.....	21
FIGURE 11.....	22
FIGURE 12.....	24
FIGURE 13.....	25
FIGURE 14.....	28
FIGURE 15.....	29
FIGURE 16.....	30
FIGURE 17.....	31
FIGURE 18.....	33
FIGURE 19.....	34
FIGURE 20.....	35
FIGURE 21.....	36
FIGURE 22.....	36
FIGURE 23.....	38
FIGURE 24.....	39
FIGURE 25.....	39
FIGURE 26.....	41
FIGURE 27.....	43
FIGURE 28.....	50
FIGURE 29.....	51
FIGURE 30.....	60
FIGURE 31.....	61
FIGURE 32.....	61

FIGURE 33.....	62
FIGURE 34.....	63
FIGURE 35.....	64
FIGURE 36.....	68
FIGURE 37.....	71
FIGURE 38.....	77
FIGURE 39.....	79
FIGURE 40.....	81
FIGURE 41.....	82
FIGURE 42.....	84
FIGURE 43.....	86
FIGURE 44.....	89
FIGURE 45.....	93
FIGURE 46.....	94
FIGURE 47.....	97
FIGURE 48.....	100
FIGURE 49.....	102
FIGURE 50.....	106
FIGURE 51.....	108
FIGURE 52.....	109
FIGURE 53.....	110
FIGURE 54.....	111
FIGURE 55.....	112
FIGURE 56.....	114
FIGURE 57.....	115
FIGURE 58.....	118
FIGURE 59.....	119
FIGURE 60.....	121
FIGURE 61.....	122
FIGURE 62.....	125
FIGURE 63.....	126
FIGURE 64.....	128
FIGURE 65.....	130
FIGURE 66.....	132

FIGURE 67.....	133
FIGURE 68.....	134
FIGURE 69.....	135
FIGURE 70.....	136
FIGURE 71.....	138
FIGURE 72.....	140
APPENDIX B 1.....	197
APPENDIX B 2.....	197
APPENDIX B 3.....	198
APPENDIX B 4.....	198
APPENDIX B 5.....	199
APPENDIX B 6.....	199
APPENDIX B 7.....	200
APPENDIX B 8.....	200
APPENDIX B 9.....	201
APPENDIX B 10.....	201
APPENDIX B 11.....	202
APPENDIX B 12.....	202
APPENDIX B 13.....	203
APPENDIX B 14.....	203
APPENDIX B 15.....	204
APPENDIX B 16.....	204

# List of Tables

TABLE 1 ..... 5  
TABLE 2 ..... 46  
TABLE 3 ..... 52  
TABLE 4 ..... 58  
TABLE 5 ..... 66  
TABLE 6 ..... 67  
TABLE 7 ..... 76  
TABLE 8 ..... 91  
TABLE 9 ..... 95  
TABLE 10 ..... 99  
TABLE 11 ..... 101  
TABLE 12 ..... 103  
TABLE 13 ..... 107  
TABLE 14 ..... 124

## Abbreviations

2xYT	2x yeast extract tryptone broth
5-ALA	5-aminolevulinic acid
AAK1	AP2-associated protein kinase 1
AMP-PCP	$\beta$ , $\gamma$ -methylene adenosine 5'-triphosphate disodium salt
AP2	adaptor complex 2
APS	ammonium persulfate
ASU	asymmetric unit
BB	backbone
Bp	base pairs
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BIRC5	baculoviral IAP repeat-containing protein 5
BisTris	2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
$\beta$ -ME	2-mercaptoethanol
BSA	bovine serum albumin
BSSE	basis set superposition error
CAMK1g	calcium/calmodulin-dependent protein kinase type 1G
CASP7	caspase-7
CC <sub>1/2</sub>	correlation coefficient of the two halves of a data set
CCP4	collaborative Computational Project number 4
COOT	crystallographic Object-Oriented Toolkit
CV	column volume
(k)Da	(kilo) Dalton
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DOT1L	histone-lysine N-methyltransferase, H3 lysine-79 specific
DSF	differential scanning fluorimetry
DTT	dithiothreitol
DYRK1a	dual-specificity tyrosine-phosphorylation regulated kinase 1a
<i>E. coli</i>	<i>Escherichia coli</i>
EDI	EGFR dimerization interface

EGFR	epithelial growth factor receptor
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ESP	electrostatic potential
EWG	electron withdrawing group
FBDD	fragment-based drug design
FBXL7	F-box and leucine-rich repeat protein 7
FP	fluorescence polarization
GSK3	glycogen synthase kinase-3 $\beta$
GSH	glutathione
HA	(non-hydrogen) heavy atoms
HB	hydrogen bond
HEFLib	Halogen-Enriched Fragment Library
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HLT	6xHistag-Lypoyl domain-TEV protease cleavage site
HTS	high throughput screening
IAP	inhibitor of apoptosis protein
IC <sub>50</sub>	50 % maximal inhibitory concentration
IDO1	indoleamine 2,3-dioxygenase 1
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
I/ $\sigma$	signal intensities over noise
ITC	isothermal titration calorimetry
JNK2	c-Jun N-terminal kinase 2
JNK3	c-Jun N-terminal kinase 3
K <sub>D</sub>	dissociation constant
LB	lysogeny broth
LE	ligand efficiency
LLE	lipophilic ligand efficiency
LELP	lipophilicity-corrected ligand efficiency
MES	2-(N-morpholino)ethanesulfonic acid
MOPS	3-(N-morpholino)propanesulfonic acid
MP2	2 <sup>nd</sup> order Møller-Plesset perturbation theory
NBT	nitro blue tetrazolium chloride
NEB	New England Biolabs Incorporated



NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NTA	nitrilotriacetic acid
PAIN	pan assay interference
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethyleneglycol
PMSF	phenylmethylsulfonyl fluoride
PPI	protein-protein interaction
ppm	parts per million
R <sup>2</sup>	coefficient of determination
R <sub>free</sub>	residual factor based on a reflection test set
R <sub>meas</sub>	redundancy independent residual factor
R <sub>work</sub>	residual factor based on the whole data
r.m.s.d	root mean square deviation
RNase	ribonuclease
RO3	rule of three
rpm	revolutions per minute
RT	room temperature
SAR	Structure-affinity relationship
SCF	self consistent field
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SLS	Swiss Light Source
SOC	super optimal broth with catabolite repression
SPR	surface plasmon resonance
STAT3	signal transducer and activator of transcription 3
STD	saturation transfer difference
TBS	TRIS buffered saline
TEV	tobacco etch virus
T <sub>a</sub>	annealing temperature
TAE	TRIS-acetate-EDTA
TCEP	tris(2-carboxyethyl)phosphine

TEMED	tetramethylethane-1,2-diamine
$T_m$	unfolding transition temperature
TPSA	topological polar surface area
TRIS	2-amino-2hydroxymethylpropane-1,3-diol
TZVPP	triple zeta valence with two sets of polarization functions
VdW	Van der Waals
$V_{max}$	most positive electrostatic potential on the $\sigma$ -hole when mapping the ESP onto the isodensity surface at 0.02 au
XB	halogen bond
XDS	X-ray detection software

Standard one-letter nomenclature of amino acids and DNA bases was used. Abbreviations not listed are explained in the text.

# 1. Introduction

## 1.1 Scope of the Study

The work presented here results from efforts to improve the understanding of non-classical protein-ligand interactions. This study combined the unique features of halogen bonding (XB)<sup>1-4</sup> and fragment-based drug discovery (FBDD)<sup>5-13</sup>. The combined approach provides the opportunity to elucidate the influence of halogen bonds on the binding mode to a greater depth than the experimental studies<sup>2, 11, 14-27</sup> have been able to do so far. Simultaneously new binding motifs for drug discovery could be found. The halogen-enriched fragment library (HEFLib)<sup>7, 11, 28</sup> used in the study could help endeavor new pharmacologically active substances, whether in drug discovery projects or pharmacological tool compounds. The necessary first step was to evaluate the physicochemical properties of the HEFLib, and some basic properties had been evaluated in a previous study<sup>7</sup>. The central part of the study consists of a three-step biophysical screening process<sup>29</sup>, which consists of an initial screening of different proteins by saturation transfer difference nuclear magnetization resonance (STD-NMR)<sup>30-33</sup>, followed by a validation phase using isothermal titration calorimetry (ITC)<sup>34-36</sup>. The ITC was comprehensively used to characterize hit fragments and structure-affinity relationships (SAR). Protein crystallization elucidated the binding modes with hit fragments and provided unequivocal evidence of halogen bonding as a primary factor for the binding mode in most cases.

## 1.2 The Halogen Bond (XB)

Proteins can interact with biomolecules such as DNA, RNA, and other proteins through various interactions<sup>37</sup>. The best-known interaction is the hydrogen bond (HB), one of the most prevalent interactions for the specific binding of (bio-) molecules in close cooperation with charged interactions<sup>38</sup>. Polar interactions are considered to stabilize the system by an increase in enthalpy while decreasing entropy<sup>39</sup>. Other interactions increase the system's entropy by displacing weakly bound water molecules<sup>40</sup>; these lipophilic substructures (e.g., methyl, phenyl, or methylene groups) interact by van der Waals forces or dispersion effects with the protein<sup>38</sup>.

In addition, interactions with  $\pi$ -systems are possible, either from cationic residues, other  $\pi$ -systems, or aliphatic groups, which require specific geometries to be attractive but are less dependent on it.

Halogens are extremely rare in proteins or peptides, except in the fungal and bacterial kingdoms<sup>41-43</sup>. Humans' only prominent organic halides are the iodine-containing thyroid hormones<sup>24, 44, 45</sup>. Fluorine and chlorine are typical substituents in medicinal chemistry, whereas bromine and iodine are mainly used as reaction vectors in intermediate compounds for cross-couplings<sup>46, 47</sup>. In textbook organic chemistry, halogens are traditionally regarded as uncharged or partially negatively charged according to the difference in electronegativity to the adjacent carbon. However, halogen contacts in organic crystals and quantum mechanical calculations show adducts of halogens with electron-rich substructures or free valence electrons in a head-on fashion<sup>3, 4, 16, 25, 48-52</sup>. If halogens had isotropic electron density (i.e., uncharged or partially negative), these contacts would purely depend on dispersion effects, resulting in lower than observed adduct formation energies<sup>4</sup>. A model with anisotropic electron distribution of halogens is better suited to explain the behavior of halogens in crystal contacts or the interaction energies derived from quantum mechanical computations (Figure 1)<sup>1</sup>.

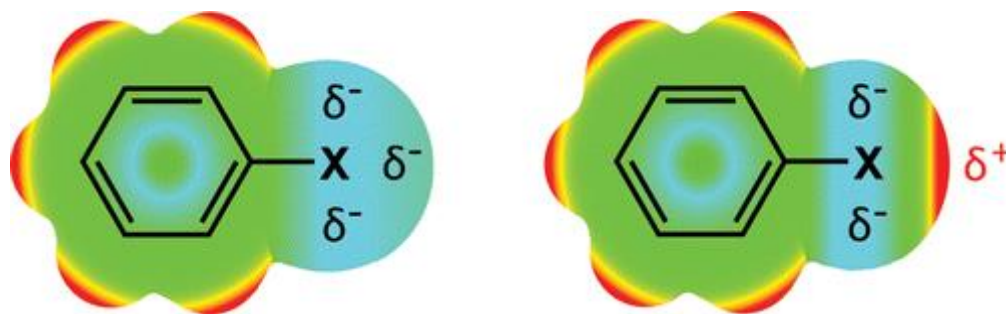


Figure 1. Schematic overview illustrates the changing perception of halogen moieties in organic compounds. The color gradient from cyan to red represents the electrostatic potential mapped onto the electron isodensity surface. Left: Traditional assumption of the halogen as a Lewis base (electron donor) with a predominantly isotropic electron distribution of the halogen. Right: A more realistic description highlights the electron density's anisotropy on the halogen (for fluorine, a positive potential appears only in exceptional cases). The most positive surface potential (including the  $\sigma$ -hole on the halogen) is colored in red, whereas the most negative surface potential is colored in cyan. Copied from Wilcken et al. under the Standard ACS AuthorChoice Agreement (see Appendix D)<sup>1</sup>.

Clark et al. could explain this anisotropy by comparing natural bond order analysis<sup>4, 53</sup>. The resulting orbitals were in an  $s^2p_x^2p_y^2p_z^1$  configuration (with the carbon halogen bond in the z-axis), thus explaining the equatorial belt of higher electron density

combined with a spot or hole of relative electron deficiency in the elongation of the halogen carbon bond. This difference in density is commonly and from now on, referred to as the  $\sigma$ -hole (Figure 1). The electron distribution allows electrostatic interactions at a  $180^\circ$  or  $90^\circ$  angle, with the halogen being the vertex. Database analysis has shown that the interaction at  $90^\circ$  is disfavored compared to the  $180^\circ$  due to the steric hindrance by other substituents or hydrogens in the aromatic ring plane<sup>1, 37, 54</sup>.

The strength of interactions in a biological environment is hard to grasp and even more challenging to quantify, as many interactions overlap and interleave. Synergistic and antagonistic effects of induction, dispersion, or multipolar group contacts<sup>37</sup> with surrounding atoms are a few reasons; the quantification of interactions is only possible in isolated cases<sup>26, 51, 55</sup>. Varying effects of solvation and desolvation of bulk or bound water add to the complexity of the problem. It has to be stressed that the energy values found in the literature are generated from quantum mechanical calculations on different levels of theory and are allotted to isolated interactions in a vacuum<sup>3, 4, 53, 56-60</sup>. Although quantum mechanical calculations do not comprehend the details of extensive biomolecular interactions, the advantages bear great possibilities. The main benefit of such calculations is the overall picture they paint. We are able to compare interactions based on their adduct formation energies in an isolated fashion. All calculated energies stated in this study are calculated on the Møller-Plesset (MP2) level of theory with either a Karlsruhe basis set (TZVPP) or a Pople basis set (6-311+G\*\*). The basis set superposition error (BSSE) was corrected if not otherwise stated. With these precautions in mind, the archetype of a hydrogen bond between two water molecules was calculated to be between 17 to 21 kJ/mol<sup>39, 61</sup>. Compared to hydrogen bonds, CH- $\pi$  bonds are three to four times weaker with an adduct formation energy of methane to benzene of 5-6 kJ/mol<sup>39, 61</sup>. Van der Waals (VdW) interactions are even weaker, and assigning energies to interactions of aliphatic groups becomes more challenging, as the interaction is intertwined in the vague border between fragile hydrogen bond and VdW-based interaction. Theoretical studies assigned <4 kJ/mol values to weakly polarized methyl groups<sup>37, 39</sup>. With the decreasing polar character of the interaction, the dependence on the geometry becomes less crucial with completely non-directional VdW interactions<sup>37</sup>. Although the weakest interaction, the VdW forces can add up significantly due to the abundance of possible interaction partners.

In contrast to VdW interactions, halogen bonds depend entirely on the interaction geometry with an optimal angle at around  $180^\circ$  and distances of 280 to 350 pm<sup>59</sup>. The range of XB adduct formation energies is as widespread as can be seen with the HB adduct formation energies<sup>1, 54</sup>, which is the result of the multitude of XB/HB acceptors multiplied by the three available XB donors (Cl, Br, and I) (Table 1). Adduct formation energies for XBs are in the range of 4-50 kJ/mol<sup>1, 59, 60</sup>. The most accessible and studied system is a halobenzene (PhX) interaction. *N*-methyl acetamide acts as a model for the amino acid backbone (BB), as the backbone is the most common acceptor and most addressed asset in proteins<sup>1, 62</sup>. The adduct formation energy of iodophenyl and the backbone (14.2 kJ/mol) is close to the energies calculated for HBs and comparable to CH- $\pi$  interactions with chlorophenyl (16.1 kJ/mol) (Table 1).

Regarding halogen bonds, some general observations can be made. Firstly, the heavier halides generally form stronger interactions<sup>1, 54</sup>. Second, the least electron-rich XB acceptors form weaker interactions<sup>1</sup>. Thirdly, aromatic structures show adduct-formation energies comparable to more electron-rich moieties (Table 1), which is further raised if desolvation is taken into account. As implied earlier, the solvation effects of interaction partners can eradicate or increase any attractive interactions as the energetic cost of desolvation can be greater than the energy gained from the interaction<sup>37</sup>. Desolvation effects are significant with the charged XB acceptors. The desolvation cost is so high that XBs with charged residues are rarely seen in protein structures. The low net gain in the attraction of 6.5 kJ/mol for bromobenzene with propionate compared to the vacuum adduct formation energy of 41.4 kJ/mol<sup>60</sup>. Desolvation cost must be considered with every protein-ligand interaction but becomes less critical with less polarized residues. As mentioned earlier, calculated energies should not be translated to the impact of binding affinities as complex environments are not incorporated into these calculations. We would assume that the addition of chlorine to an inhibitor should increase affinities only by small increments from the pure energies. A true statement in some cases, but numerous examples have been discovered where such a small change resulted in an extreme affinity increase<sup>1, 15, 63-65</sup> (see section 1.2.1). These examples serve to remind potentially overinterpreted data from highly constructed model calculations.

Table 1. Adduct formation energies of various XB model systems were calculated at the MP2/TZVPP level of theory.

System	Method	$\Delta E$ [kJ/mol]	Distance X-D [pm]	$\sigma$ -hole angle (C <sub>Ar</sub> -X-D) [°]	Reference
PhCl-BB	MP2/TZVPP	5.6	312	171.2	59
PhBr-BB	MP2/TZVPP	9.0	304	177.4	59
PhI-BB	MP2/TZVPP	14.2	302	175.6	59
PhCl-Met	MP2/TZVPP	7.8	347	162.7	14
PhBr-Met	MP2/TZVPP	10.7	343	170.1	14
PhI-Met	MP2/TZVPP	15.2	339	169.5	14
PhCl-Asp/Glu	MP2/TZVPP <sup>b</sup>	15.5	280	179.8	60
PhBr-Asp/Glu	MP2/TZVPP <sup>b</sup>	30.2	260	178.5	60
PhI-Asp/Glu	MP2/TZVPP <sup>b</sup>	56.4	259	180.0	60
PhCl-Asn/Gln	MP2/TZVPP <sup>b</sup>	10.7	327	177.9	60
PhBr-Asn/Gln	MP2/TZVPP <sup>b</sup>	12.9	305	178.5	60
PhI-Asn/Gln	MP2/TZVPP <sup>b</sup>	19.8	305	178.4	60
PhCl-His	MP2/TZVPP	4.3	314	178.4	57
PhBr-His	MP2/TZVPP	8.9	307	178.3	57
PhI-His	MP2/TZVPP	16.0	303	178.0	57
PhCl-Phe	MP2/TZVPP <sup>a</sup>	16.1	500 <sup>c</sup>	130.9 <sup>c</sup>	66
PhBr-Phe	MP2/TZVPP <sup>a</sup>	17	510 <sup>c</sup>	133 <sup>c</sup>	66
PhI-Phe	MP2/TZVPP <sup>a</sup>	18.3	510 <sup>c</sup>	133 <sup>c</sup>	66
PhCl-Tyr	MP2/TZVPP <sup>a</sup>	11.6	330 <sup>c</sup>	166.3 <sup>c</sup>	66
PhBr-Tyr	MP2/TZVPP <sup>a</sup>	14.2	340 <sup>c</sup>	164.5 <sup>c</sup>	66
PhI-Tyr	MP2/TZVPP <sup>a</sup>	20.6	340 <sup>c</sup>	161.9 <sup>c</sup>	66

<sup>a</sup> Distance scan with a fixed angle of the halogen to the aromatic plane at 180°. <sup>b</sup> Based on a Connolly surface distance scan. <sup>c</sup> halogen to the center of mass of the ring

### 1.2.1 Applications of halogen bonds in biochemistry

Initially, the halogen bond was postulated in the 50ies of the last decade, and it could be experimentally proven in the years following in crystal contacts of small molecules with iodine or bromine<sup>3</sup>. Since then, halogen bonds have seen wide applications in the material sciences and catalysis<sup>3, 48, 67</sup>. From the use of transition metal halides to sophisticated organic frameworks, the evolution of halogen bonds as a way of steering stereoselective reactions has come a long way<sup>3, 47, 68, 69</sup>. Similarly, halogens have been successfully used in supramolecular assembly and recognition of anions and specific molecules<sup>3, 49, 70, 71</sup>. The most common instance of a halogen bond in a biological environment is the binding of thyroxine and triiodothyronine to their respective transport and effector proteins<sup>20, 72, 73</sup>.

Regarding the halogen bond, the thyroxine system has been studied intensively to understand the halogen bond in a physiological context<sup>20, 72, 73</sup>. This example and the investigation of known drugs containing halogens lead to the acknowledgment of academia and the pharmaceutical industry of the potential of halogens in a biological environment in the last decade<sup>17, 74-77</sup>. At first, serendipity was the driving element in discovering halogen bonding moieties. As the long-approved drug rivaroxaban shows, chlorine and bromine exerted better affinity than every tested substituent. At this point in time, halogen bonds were not discussed in the literature<sup>78</sup>. Nonetheless, the geometry of the chlorine interaction in rivaroxaban, an inhibitor of Factor Xa, is near perfect with an angle of  $177.5^\circ$  and a distance of  $3.8 \text{ \AA}$  (Figure 2) to the  $\pi$ -system of Y228's sidechain. Since then, more inhibitors and drugs have been found or designed featuring halogens<sup>63, 65, 75, 79, 80</sup>.

In many cases, chlorine atoms are incorporated in drug discovery campaigns, but the heavier halides are rarely introduced. This trend is mirrored by the relative abundance of chlorine in natural products compared to the scarcity of bromine or iodine compounds in nature<sup>81</sup>. As Table 1 showcases, chlorine shows the weakest adduct formation energies with XB acceptors (Cl=5.6 kJ/mol to I=14.2 kJ/mol with the backbone oxygen). Therefore, a great potential for stronger halogen bonds remains by utilizing bromine and iodine as XB donors. The introduction of heavier halides could lead to sharp increases in affinity as observed with chlorine-bearing inhibitors compared to other substituents.



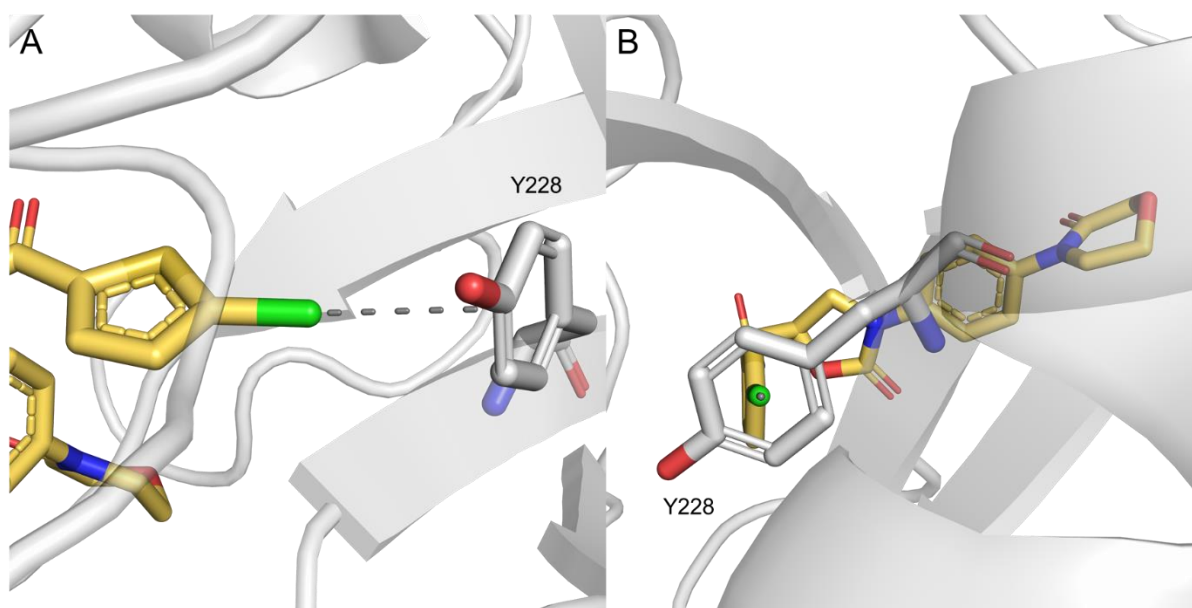


Figure 2. Crystal structure of 2W26<sup>78</sup>. **A**: Side view of the halogen bond of rivaroxaban with Y228, the distance between is 3.8 Å. **B**: Frontal view of the same interaction, the angle is 177.5°. The example illustrates a near-perfect geometry for a halogen bond.

Halogens are mainly introduced in lead optimization stages into inhibitors, but no systematic evaluation of halogens for drug discovery has been undertaken. In previous work of our group, computational methods were applied to investigate the different moieties found in proteins as electron donors for halogen bonds (Table 1)<sup>14, 57, 59, 60</sup>. Combined with the data provided by the PDB<sup>1, 82</sup>, a scoring function of halogen bonds to backbone oxygens was achieved and incorporated into the docking program PLANTS<sup>54, 83</sup>. Quantifying XB in docking lays the foundation for a wider use of halogen bonds in drug discovery, as docking and scoring constitute a significant part of every drug discovery endeavor. Overall, utilizing halogens to facilitate unique drug properties gained traction in drug discovery and development. Nowadays, chlorine and bromine are more common substituents in a hit to lead development. However, in most cases, the halogen bond is not the primary purpose for introducing a halogen, as pharmacokinetic or toxicity reasons are the leading cause for halogen introduction<sup>21</sup>. Besides the effort for new inhibitors, the introduction of novel tRNA-synthetases and DNA/RNA synthesis has made it possible to include halogen-containing nucleotides and amino acids into biomolecules<sup>2, 16, 84</sup>. A development with great potential in protein engineering could stabilize conformations, broaden/reduce substrate specificity, or tailor enzyme activity. It might even be possible to create entirely new functionalities with halogens in biomolecules by producing new versions of existing or designing new ones, as has been done with organic catalysts<sup>52</sup>.

### 1.2.2 Properties and tuning of halogen bonds

Hydrogen and Halogen bonds share a wide range of possible adduct formation energies<sup>1, 61</sup>. This characteristic depends partly on the acceptor of the interaction, which cannot be influenced in most cases. On the other hand, the ligand or inhibitor is subject to the design process to support the interaction. The strength of electrostatic interactions can be increased by lowering the local electron density around the halogen. In the case of the halogen bond, this is explained by an enlarged  $\sigma$ -hole. Utilizing negative inductive and mesomeric effects, the properties of HB and XB can gain significant strength<sup>1, 56</sup>. Electron withdrawing groups (EWG) (with a -I or -M-effect) can be used but require extra space, which is one reason fluorine is a highly successful substituent in many inhibitors. However, increasing the strength of interactions while requiring minimal extra space is a challenging accomplishment. Other effects of fluorine incorporation (metabolic stability and fluorine-based interactions) are omitted to avoid digression. In an optimal case, the introduction of EWG into a scaffold would generate additional interactions with the ligand.

In some cases, no substituent can be added, which leaves the change of the scaffold to improve the compound's attributes<sup>56</sup>. Halogens are mainly used as aryl halides in medicinal chemistry, as many aliphatic halogens, like benzyl halides, are too reactive for noncovalent ligands. However, studies of the recently approved drug, Asciminib (FDA approval in October 2021), showed the possibility of using a chloro-difluoromethoxy group as a stable substituent (Figure 3 A)<sup>75, 85</sup>. Still, in most cases, an aromatic scaffold is present in nearly every ligand, which can be utilized to strengthen a halogen bond<sup>56</sup>. In particular, electron-deficient scaffolds, for example, pyrimidines, triazoles, and pyrazines help increase the  $\sigma$ -hole while maintaining the size<sup>56</sup>. As the complexity of scaffold effects increases with the annulation of ring systems, simple observations of electron density come to their limits. Therefore, we utilized the concept of electrostatic potential (ESP) maps, with the point of the maximum electrostatic potential at the halogen being called  $V_{\max}$ <sup>7, 86</sup>. With this tool in hand, it is easy to assess the size of the  $\sigma$ -hole based on the electrostatics of the compound (see section 1.3.1). As mentioned earlier, we cannot tune the halogen bond by changing the acceptor. This does not mean there is no possibility that the XB-acceptor can tune the halogen bond. As calculations could prove, a positive charge in the vicinity of a halogen can strongly influence the  $\sigma$ -hole (Figure 3 B)<sup>87</sup>.

Depending on the position in the three-dimensional space around the halogen or aromatic scaffold, a positive charge of an arginine or lysine, for example, can increase or decrease the  $\sigma$ -hole of a halogen in a ligand. In a beneficial geometry, the cation can interact with the aromatic scaffold, generating affinity by itself. Additionally, the cation would reduce the electron density through the charge transfer from the aromatic scaffold to the cation and, in turn, reduce the electron density at the halogen, indirectly increasing the  $\sigma$ -hole. These inductive effects of surrounding amino acids in a binding pocket have not been subject to intensive studies, and the exact impact on XB is unclear.

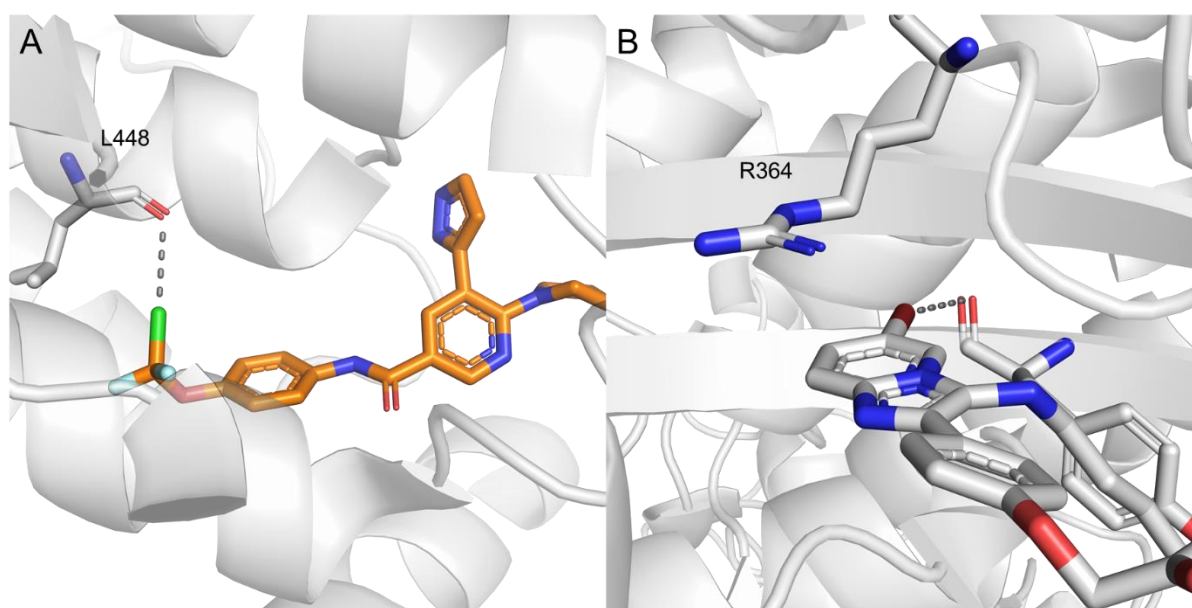


Figure 3. **A:** Structure 5MO4<sup>79</sup> of BCR-ABL with Asciminib, the halogen bond has a distance of 3.3 Å with an angle of 178.3°. The two fluorine further reinforce the interaction in the chlorodifluoromethoxy group. **B:** Structure 4ACF<sup>88</sup> of mycobacterial glutamine synthetase with an inhibitor. The halogen bond (distance 3.7 Å and an angle of 160°) is reinforced by arginine 364 due to the cation- $\pi$  interaction. The ring center is 4.1 Å apart from the guanidium moiety.

## 1.3. Fragment-based drug discovery

Drug discovery is a long-lasting process, with years or decades necessary to develop a new drug. There are many steps to reach this goal. One of the first steps is finding suitable initial binders. These can either function as tool compounds for general pharmacological target validation or could be developed further for potential drug candidates. Suppose minimal information about the target or the binding pocket is available. In that case, the most diverse set of compounds should be tested to generate different binding modes with various characteristics<sup>89, 90</sup>, thus raising the chances of finding a compound portfolio with desirable properties. The chemical space of drug-like compounds is vast, to say the least, and increases exponentially with every additional (non-hydrogen) heavy atom. The most used approach is high throughput screening (HTS) of millions of compounds against a target to find binding molecules. This technology has been steadily developed to a level where automatic screening of thousands of compounds per day is feasible, requiring streamlined assays with extensive automation and instrumental efforts.

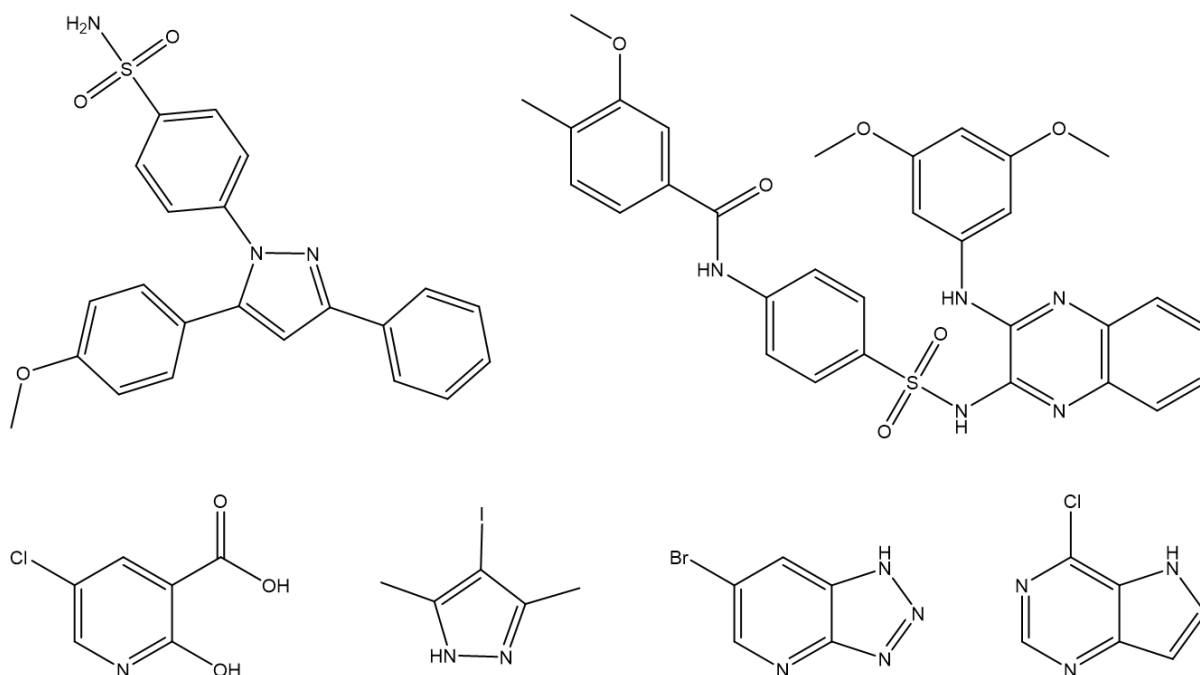


Figure 4. A comparison shows kinase inhibitory HTS compounds from the ChemDiv Protein Kinases Inhibitors Library and the Tocris Kinase Inhibitor Toolbox (top row) with compounds of the HEFLib (bottom row). The HTS compounds are larger, more lipophilic, and the greater number of rotatable bonds increases the entropy loss upon binding.

The last decade saw the rise and incorporation of fragment-based drug discovery (FBDD) into many drug discovery projects. The first key idea of screening smaller molecules with inherently weaker affinities is to cover a more significant part of the chemical space with fewer molecules<sup>8, 28, 91</sup>. Figure 4 illustrates the different sizes and numbers of possible enthalpic interactions. The top row shows typical HTS compounds from a commercial kinase library, and the bottom row includes examples of fragments from the HEFLib. In other words, the reduction of screening efforts in terms of the amount of data point generation necessary for a hit to occur is the first idea behind the concept of FBDD. The second key concept of utilizing small fragments in binding assays is that the binding modes can easier adopt geometries, which feature optimized interactions compared to larger drug-like molecules<sup>5, 8, 10, 92</sup>. These large molecules can adopt very few conformations, of which some may bind to a target but are far away from the best possible conformation. By the large surface of these compounds, the measured affinity of HTS-compounds is relatively strong compared to the much smaller typical fragment. The small size and thus fewer interactions increase the chance of finding fragments with near-optimal binding modes, but the resulting affinity is much weaker than the typical HTS-compound. Due to the generally hard to detect binding affinity, possible other binding modes vanish into the background noise. Larger, drug-like compounds might bind into a pocket, but the best possible overall geometry for every interaction cannot be achieved.

The first fragment-based drug, Vemurafenib, reached FDA approval in 2011<sup>9, 93</sup>; since then, many more pharmacological tool compounds and drugs have evolved from fragments<sup>9</sup>. After the initial hit finding, there are three general ways to develop a fragment into an inhibitor: growing, merging, and linking<sup>92</sup>. Fragment growing starts from an initial hit and subsequent rounds of optimization by adding favorable substituents and scaffolds. Suppose two fragments share part of their binding mode with each other. In that case, they can be merged into one inhibitor with potential synergistic effects on the affinity to the target<sup>8, 28</sup>.

### 1.3.1 Halogen-Enriched Fragment Library

If multiple hits have varying binding modes, linking multiple fragments into one compound can be possible, often resulting in potentiated affinity gains<sup>92</sup>. As the linked fragments have less rigid body entropy losses from binding, a correct linker could form additional interactions, further boosting the affinity<sup>92</sup>. The linker can induce a certain tension on the binding mode of the linked fragments, but if chosen correctly, the strain should be minimal<sup>92</sup>. Additional benefit can come from inhibiting two distinct pockets, changing the compound's binding nature from competitive to non-competitive or vice versa<sup>92, 94</sup>.

There are challenges associated with FBDD, mainly arising from the low binding constants of small molecules. Fragments have to be used in higher (up to millimolar) concentrations in assays, which demands high solubility in aqueous solutions. The screening assays have to be sensitive and robust enough to detect micro- to millimolar affinities. These challenges, mainly the assay problem, have been addressed in the past decades, and nowadays, a broad set of techniques can detect millimolar affinities. The most utilized techniques are NMR-based, crystallography, and surface plasmon resonance (SPR)<sup>9</sup>. The solubility problem remains with each new library generation. However, computational methods in calculating solubility have advanced considerably<sup>95, 96</sup>. There are still compounds with hard to estimate properties and experimental evaluation of solubility is the first step in library evaluation.

### 1.3.1 Halogen-Enriched Fragment Library

As mentioned above, fragment libraries do not have to be of the usual HTS libraries' scale<sup>9, 28, 92</sup>. The reduced cost and effort necessary in acquiring a fragment library make the concept attractive for academic groups. Although HTS libraries have found academic applications, these instances are still scarce compared to the broader variety of fragment libraries developed<sup>6</sup>. Fragment libraries were designed for different purposes in drug discovery, with special libraries inter alia for fluorine NMR screening<sup>97</sup>, target focused libraries, anomalous scattering X-ray crystallography<sup>11, 28</sup>. From the halogen bonding standpoint, fragment libraries are especially interesting. As mentioned in section 1.2.1, halogens have been used in the lead optimization process. The late-stage introduction of halogens can become troublesome due to the higher number of established key interactions. These interactions compete for the best

geometry<sup>28</sup>. The halogen bond is not the most potent interaction combined with the strong dependency on the geometry. It is challenging to generate a favorable halogen bond without disrupting existing interactions<sup>7, 11, 28</sup>. By incorporating halogens into fragments with a limited number of potential hydrogen bonds, it could be easier to find binding modes with attractive halogen bonds in conjunction with the surrounding interaction partners of the compound<sup>7, 28</sup>. The idea arises from the early work of our group with the tumor suppressor p53<sup>11</sup>. An iodine-containing fragment could be identified, stabilizing the thermodynamically unstable Y220C mutant<sup>11, 28, 98</sup>.

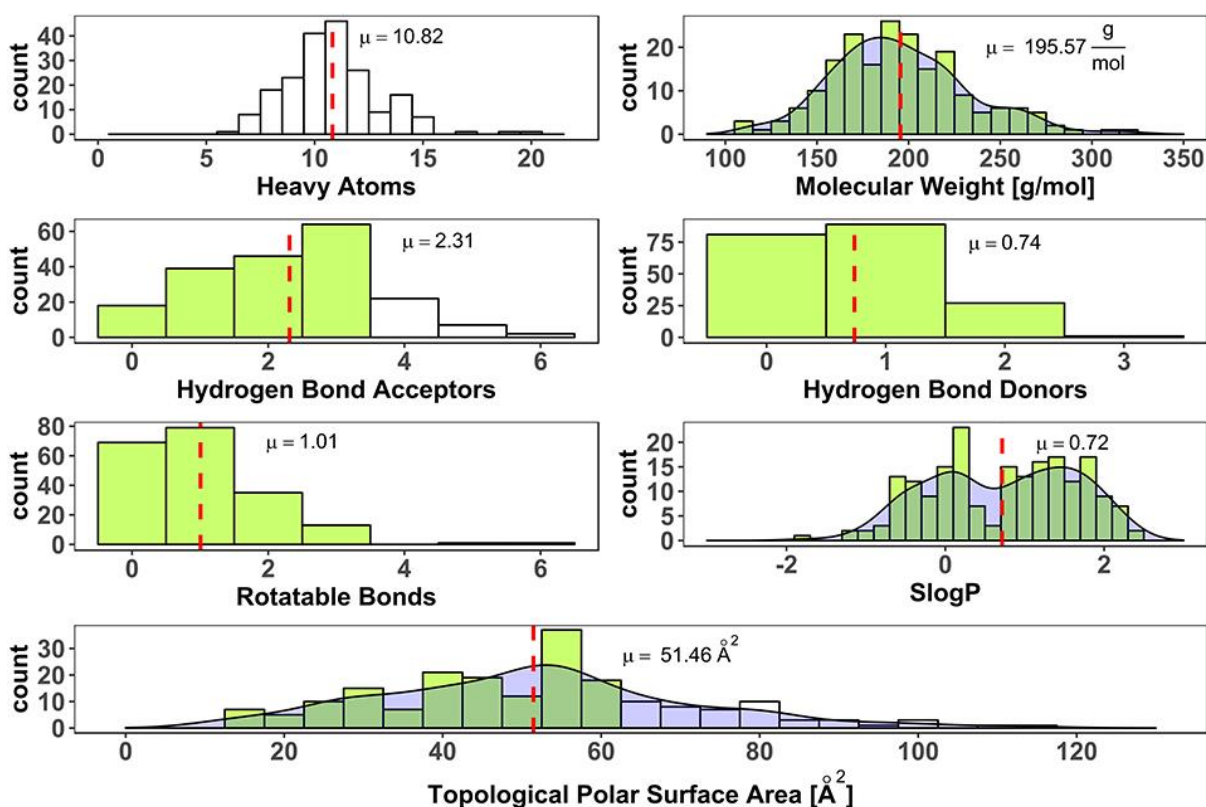


Figure 5. Distribution and mean value for characteristic fragment properties. Green bars represent compliance with the RO3. The red line marks the average value for each criterion, copied from Heidrich et al. under CC BY 4.0 license (see Appendix D)<sup>7</sup>.

The HEFLib was generated with theoretical, practical, and strategical considerations in mind<sup>7</sup>. The key focus in the development of the library was to integrate the less common bromine and iodine into a fragment library, with the main focus on probing protein binding pockets for halogen bonding<sup>7, 28</sup>. In the evaluation process, experimental issues of fragments had to be addressed. A general guideline for library development is the rule of three (RO3)<sup>6, 99</sup>. This ruleset was derived from 40 fragment-hit observations and named after the infamous rule of five<sup>100</sup>. As the authors commented, both “rules” are frameworks or guidelines at best<sup>99-101</sup>. Analog to the rule

### 1.3.1 Halogen-Enriched Fragment Library

of five, fragments should not surpass a molecular mass of 300 Da, a  $\log P < 3$ , not more than three rotatable bonds, no more than three hydrogen bond acceptors or donors, and a polar surface area (PSA)  $< 60 \text{ \AA}^2$ <sup>99</sup>. Although the design process of the library did not focus on the RO3, most fragments in the library adhere to the parameters given (Figure 5). The molecular weight as a metric for halogen-bearing fragments has to be rejected due to the immense weight of bromine and iodine compared to carbon. Weight-wise, one iodine atom is equivalent to ten carbon atoms. The intent of the RO3 authors postulating a weight-based guideline was to limit the size of fragments. Therefore, we adhered more to the number of (non-hydrogen) heavy atoms (HA) below 20<sup>28, 99</sup>.

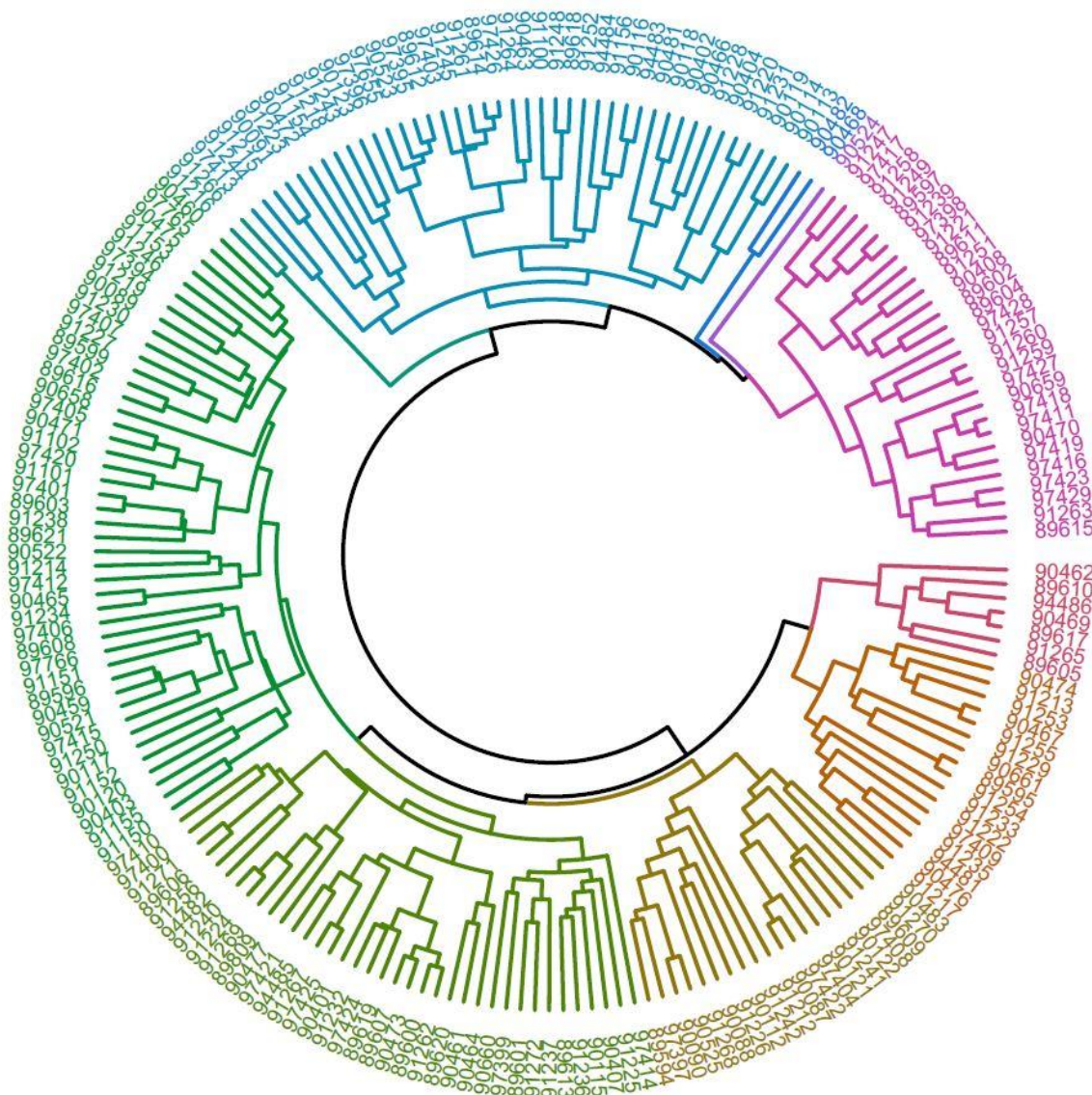


Figure 6. The Dendrogram shows fragment similarity, copied from J. Heidrich<sup>102</sup>. The numbers are the last five digits of the fragment barcode identifier.



Every fragment is cataloged with a unique, nine-digit barcode; the last four digits of this code are sufficient to identify a fragment in the current library (Figure 6). The first five digits are omitted for convenience and readability in this study. The HEFLib used in this study was created by Heidrich et al.<sup>7</sup>, without stringent adherence to the RO3. Nonetheless, most compounds adhere to the earlier mentioned criteria with only minor deviations in the PSA and the number of hydrogen bond acceptors (Figure 5). The molecular weight could be held below the 300 Da mark. In addition, the solubility and thus the SlogP (*RDKit*)<sup>103</sup> value were given greater importance than, for example, the TPSA value or the number of rotatable bonds<sup>7</sup>. As the library was created, it became apparent that the widespread use of iodine and bromine compounds was impossible. The problem is caused by the low commercial availability and the high prices associated with iodine compounds. Thus, an arbitrary price limit of 5 \$ per milligram was used, resulting in 114 chlorine-bearing fragments, 59 bromine, and 14 iodine fragments being acquired<sup>7</sup>.

The main focus in obtaining the library was incorporating diverse halogen bonding motifs. The diversity in the electrostatic potential of fragments can be assessed by the  $V_{\max}$  and ESP-plots as guidelines<sup>86</sup>. Due to tuning and anti-tuning effects, fragments with chlorine or bromine in the library feature  $\sigma$ -holes comparable to the heavier halides. The halogen itself should utilize favorable  $\sigma$ -holes, combined with a versatile electrostatic environment. As the HEFLib compounds are small - even for fragments - with a mean HA count of 10.82 (Figure 5), the XB motif ensures a certain complexity to avoid promiscuous or even unspecific binding<sup>7, 28</sup>. The library's core is made out of fragments with as little similarity as possible (Figure 6). Fragments were included around the diverse core to generate matched molecular pairs or fragments with rudimentary structure-affinity relationships to the core set. Matched molecular pairs enable us to assess the impact on the affinity of the halogen bond<sup>28</sup>. The advantage of some similarity within the HEFLib becomes imminent when a fragment hit has been confirmed, and further evaluation is necessary—enabling quick assessment without acquiring further compounds.

## 1.4. Three-stage biophysical screening

### 1.4. Three-stage biophysical screening

One of the main challenges in FBDD is the weak affinities of small fragments. Not all biochemical techniques are capable of measuring up to millimolar affinities. In recent years the three most prominent techniques for initial screening were NMR (ligand and protein observing methods combined), X-ray crystallography, and SPR<sup>6, 9</sup>. Each technique bears certain advantages and drawbacks. We were looking for a screening protocol for a multitude of possible targets, which makes crystallography challenging, as every target has to be crystallized first, leading to potential bottlenecks. Ligand observing NMR methods, on the other hand, need little optimization in comparison and can reach a relatively high throughput without the extensive facilities necessary. SPR has proven its worth, but only state-of-the-art instruments deliver the sensitivity for small fragments with low affinities. We loosely based our screening protocol on the one published by Mashalidis et al.<sup>29</sup>. After reviewing different methods and pulse protocols, we chose an STD-based initial screening<sup>30, 33</sup>. The hit validation would be performed by ITC and complemented with the crystallization of selected protein-fragment complexes.

#### 1.4.1 Saturation transfer difference NMR

An STD-NMR experiment consists of two interleaved <sup>1</sup>H-NMR experiments, which are subtracted from each other afterwards<sup>30</sup>. In a so-called on-resonance spectrum, a radio pulse with the resonance frequency of methyl groups is used to saturate these groups in the protein (-1 to 1 ppm). As the protein has a large mass, the T<sub>2</sub>-relaxation time is minimal, and the saturation quickly distributes throughout the whole protein. Fragments, on the other hand, have low affinities ( $10^{-3} \frac{\text{mol}}{\text{L}} > K_D > 10^{-8} \frac{\text{mol}}{\text{L}}$ ) with fast exchange rates between the bound and free state. If a fragment binds to the protein, part of the magnetization is transferred to the fragment via the nuclear Overhauser effect (NOE). In general, protons in close contact with the protein receive more saturation than protons with fewer or more distant contacts. As a result, the signals in the on-resonance spectrum are reduced. Another experiment (off-resonance) is performed interleaved with the on-resonance scans. The off-resonance experiment utilizes a saturation pulse at a frequency in the high field (>30ppm) to avoid any

saturation of the protein or compound. A difference of both spectra is calculated, and the protons with protein contacts remain visible in the spectrum (Figure 7). Due to the low residence time compared to the saturation time, multiple fragments can be saturated within a scan. The  $T_2$ -relaxation time of small fragments is prolonged, and they remain saturated within the experiment's time frame. The experiment can be made more efficient by the preparation of fragment mixtures<sup>30</sup>. Only binding fragments remain visible in the difference spectrum, and possible competitive fragments are sorted by their affinity within one sample. In addition, the STD technique provides rudimentary information about the binding mode; as protons with closer contacts give a stronger signal, one can make assumptions about the orientation of the fragment bound to the protein<sup>31, 33</sup>.

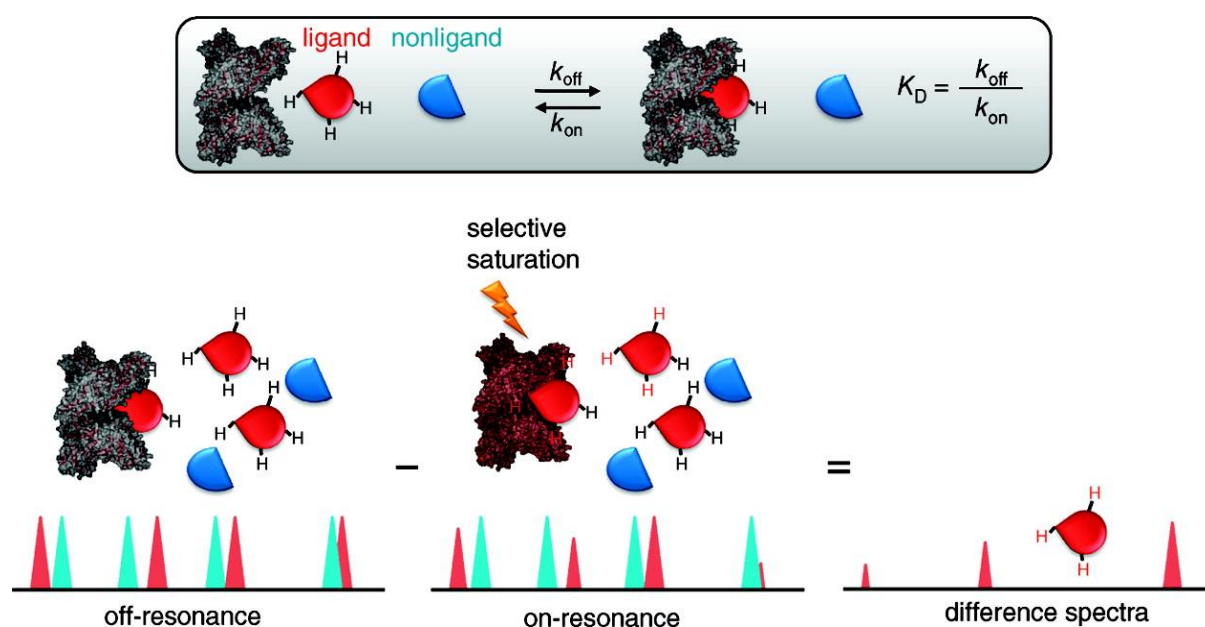


Figure 7. Principle of an STD-NMR experiment. The target protein is saturated with a targeted radio pulse. Upon binding, magnetization is transferred to the fragment, thus reducing the proton signal. The on-resonance spectrum is subtracted from a reference (off-resonance) spectrum, and signals of a binding fragment remain visible, copied from Viegas et al.<sup>104</sup> with permission from ACS (see appendix D).

## 1.4.2 Isothermal titration calorimetry.

The sensitivity of ITC experiments is approximately ten times lower than that of STD experiments. STD-NMR experiments can detect millimolar affinities<sup>105</sup>, whereas the lower limit for the ITC measurements is high micromolar affinities. This limit is imposed by the solubility of either the protein or ligand<sup>34, 106</sup>. Higher concentrations of either ligand or analyte are necessary to measure even lesser affinity. As a result, not every STD hit could be verified or even quantified through ITC. The measurement principle relies on the transportation of heat in a binding event. Protein in the sample cell is heated with a constant heat rate and maintained at a set temperature, and the molecule of interest is added in small steps to the protein solution. The electric current in the heating element to maintain the temperature compared to a reference cell is measured. The resulting peaks in the thermogram are integrated over time (Figure 8). The resulting titration curve can be fitted to a model, which gives the binding event's stoichiometry, affinity, and thermodynamic parameters. The c-value is used to estimate the slope of the curve, defined by Equation 1, to achieve a good resolution and accuracy of the fit.

$$\text{Eq. 1.} \quad c = n \cdot K_a \cdot [R]_t = n \frac{[R]_t}{K_d}$$

$K_a$  and  $K_d$  are the association and dissociation constants,  $n$  is the number of binding sites, and  $[R]_t$  is the total receptor concentration. The higher the c-value, the steeper the titration curves. Traditionally, a c-window for affinity fitting is between 10 and 500<sup>35, 106</sup>. This recommendation dates back to the last century<sup>107</sup>. The last decades have seen tremendous improvement in calorimeter design, which increased the instrument's sensitivity immensely<sup>106</sup>. With newer instruments and higher excess of fragment over protein (up to ten-fold instead of the usual two-fold), titrations with a c-value as low as 0.1 have been proven successful and accurate<sup>34, 106</sup>. However, one caveat must be made, as the stoichiometry at c-values below ten can no longer be fitted. Therefore, the stoichiometry must be known prior, or assumptions must be made, and a fixed value is set (here  $n=1$ ). The resulting curve accurately describes the affinity, but the enthalpy is often fitted with low accuracy and should not be used in data interpretation.

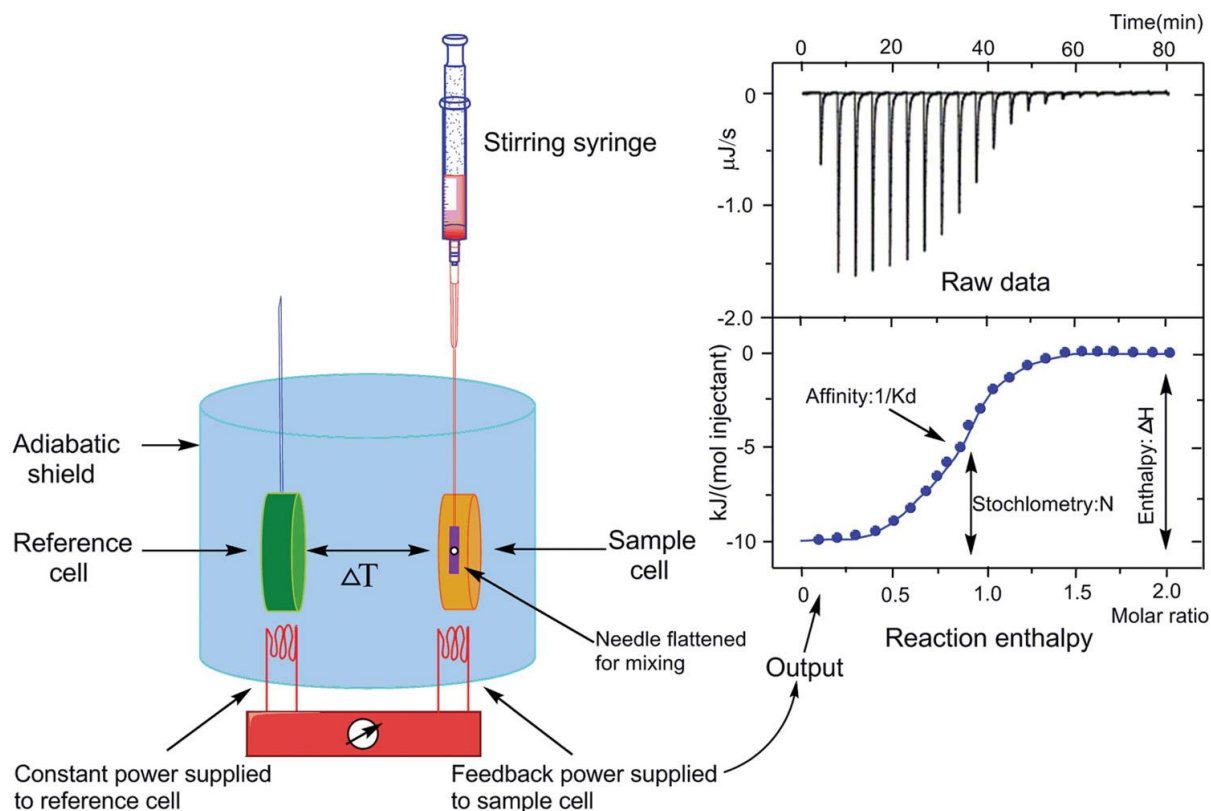


Figure 8. The schematic representation of an ITC with a characteristic thermogram and curve-fitting is on the right. The reference cell is heated to a constant temperature; heat is released upon injection of a compound to the protein in the sample cell. The change in electric current to maintain a stable temperature is measured, copied from Song et al.<sup>108</sup> under the CC BY 4.0 license (see Appendix D).

### 1.4.3 Protein crystallography

The most utilized method in structural biology is the X-ray crystallization of proteins<sup>6</sup>. There are many methods and conditions to produce the necessary crystals. Although significant progress in the generation of crystals has been made, it remains the bottleneck in many cases of structure elucidation. The vapor diffusion method, with its subdivisional hanging and sitting drop variants, is the most widespread within the protein crystallization methods. This study features the sitting drop vapor diffusion method. Figure 9 summarizes an overview of the crystallization process from a phase perspective. A stable protein solution is brought into indirect contact (via the gaseous phase) with a hygroscopic precipitant solution. Water is drawn from the protein drop into the precipitant, slowly supersaturating the protein solution<sup>109</sup>. Either seed crystals or spontaneous nucleation is needed to start the crystallization process. The growing crystals reduce the protein concentration in the solution until a metastable equilibrium

is reached (Figure 9)<sup>109</sup>. The great challenge is to find suitable conditions under which the protein is maximally stable and can form a uniform lattice. Proteins can be co-crystallized with an organic compound, or these compounds are soaked into preformed crystals. The X-ray diffraction pattern is measured at a synchrotron beamline, with special cryo-equipment to prevent the crystals from thawing and reducing radiation damage. Hundreds of diffraction patterns are indexed, integrated, corrected, and combined to generate an electron density map. The proteins examined in this study have been crystallized by other scientists, thus molecular replacement was used to solve the phase problem. Iterative cycles of model-building and refinement are necessary to generate a profound model of the underlying crystal and the molecular structure.

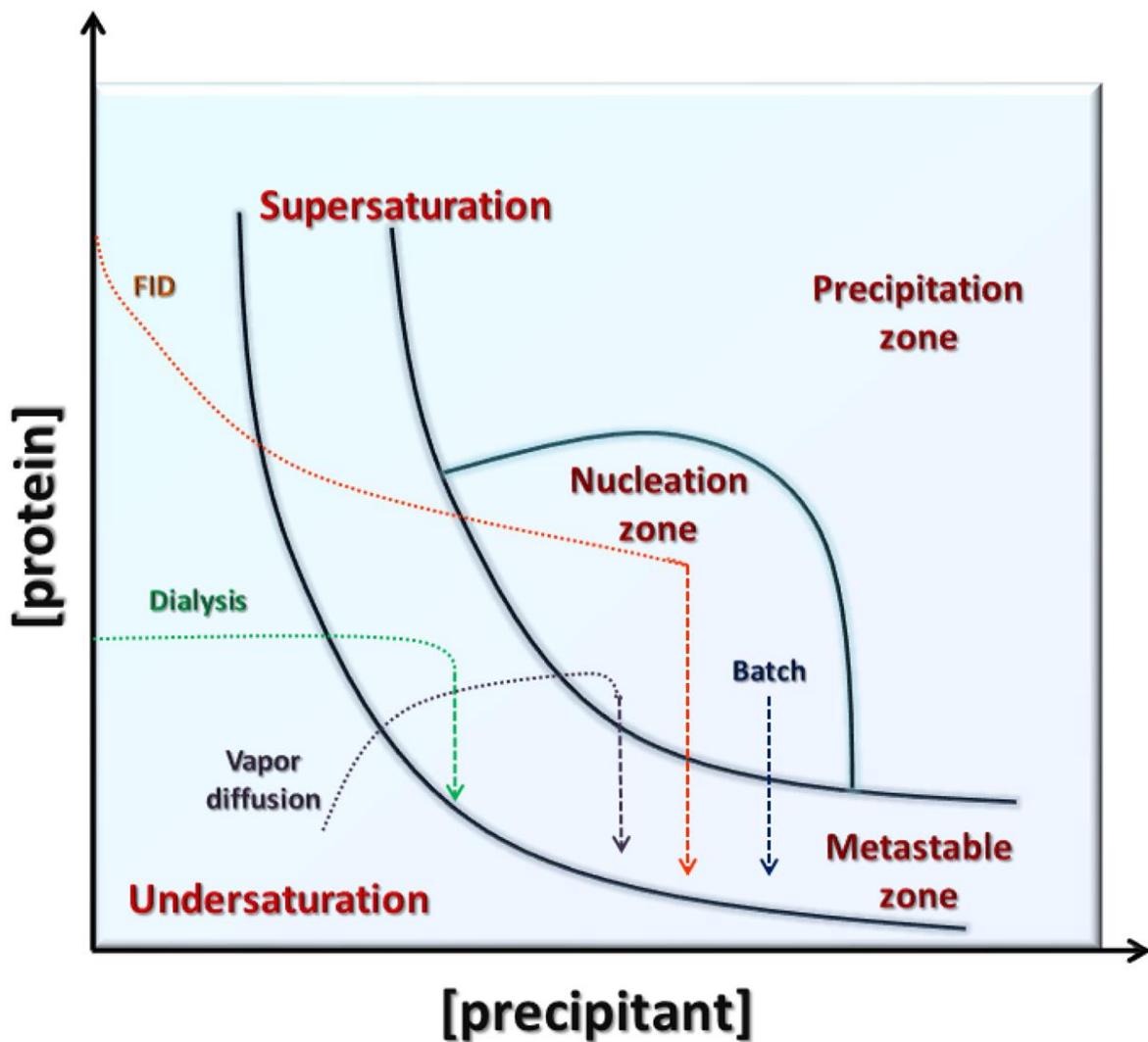


Figure 9. A simplified version of the protein crystallization phase diagram. The transition pathways of four different crystallization techniques are depicted. FID: Free interface diffusion. Figure copied from Krauss et al.<sup>109</sup> under the CC BY 3.0 license (see Appendix D ).

## 1.5 Pharmacological context of Targets

The studied proteins are part of diverse pharmacological processes. Moreover, they are either crucial in the pathophysiology of various diseases or have not been studied extensively enough to estimate the importance of the target in disease processes. Eight proteins with entirely different physiological tasks were studied. Although five of the eight proteins are kinases, their (patho-)physiological purposes widely differ.

The adaptor-associated kinase 1 (AAK1) is a kinase, which is not associated with apoptosis or cell growth processes as most kinases are, but instead seems to play a vital role in clathrin-mediated endocytosis<sup>110, 111</sup>. Clathrin is a vital protein in the endocytosis of a multitude of viruses (e.g., Hepatovirus, Flavivirus,  $\beta$ -Coronavirus)<sup>112</sup> and many other transporting processes in cells. AAK1 has been studied in the context of COVID-19 treatment, but no clinical candidate has been found<sup>113</sup>. Many of the reported inhibitors for AAK1 were initially developed for other kinases as their primary target (Figure 10). Scientific interest in the target has skyrocketed compared to the beginning of this study; many more inhibitors have been reported<sup>114</sup>. They highlight the importance of AAK1 to the scientific community<sup>113</sup>.

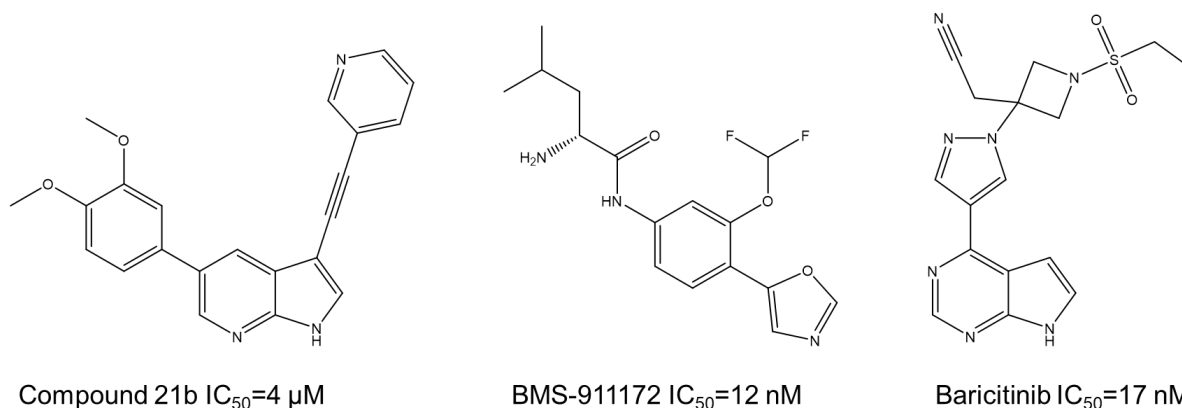


Figure 10. Structures of some inhibitors with AAK1 as an off-target, derived from the IUPHAR database<sup>115</sup>.

The calcium/calmodulin dependent protein kinase (CAMK) is the kinase subfamily of the CAMK1G, which are mainly regarded as so-called dark or orphan kinases, as little to no knowledge about the proteins exists<sup>116</sup>. Evidence for significance in specific pathophysiological processes has yet to be provided. Some data from the Cancer Genome Atlas (TCGA: <https://darkkinome.org/kinase/CAMK1G>) link mutations in the CAMK1G gene to cholangiocarcinoma, but it is unknown whether the mutations are

relevant to cancer development. Obtaining further data on pathological processes in which CAMK1g plays a vital role could consolidate CAMK1G as a vital pathophysiological target. A selective tool compound showcasing the effects of CAMK1G inhibition might be of great help and could alleviate problems with the research around the CAMK1G. No selective inhibitor has been reported so far. A few already established inhibitors and approved drugs have shown a particular affinity towards CAMK1G as an unintentional off-target<sup>117</sup>. The majority of these compounds (e.g., Sunitinib, Midostaurin, and Tamatinib) are known to be highly promiscuous within the kinome (Figure 11).

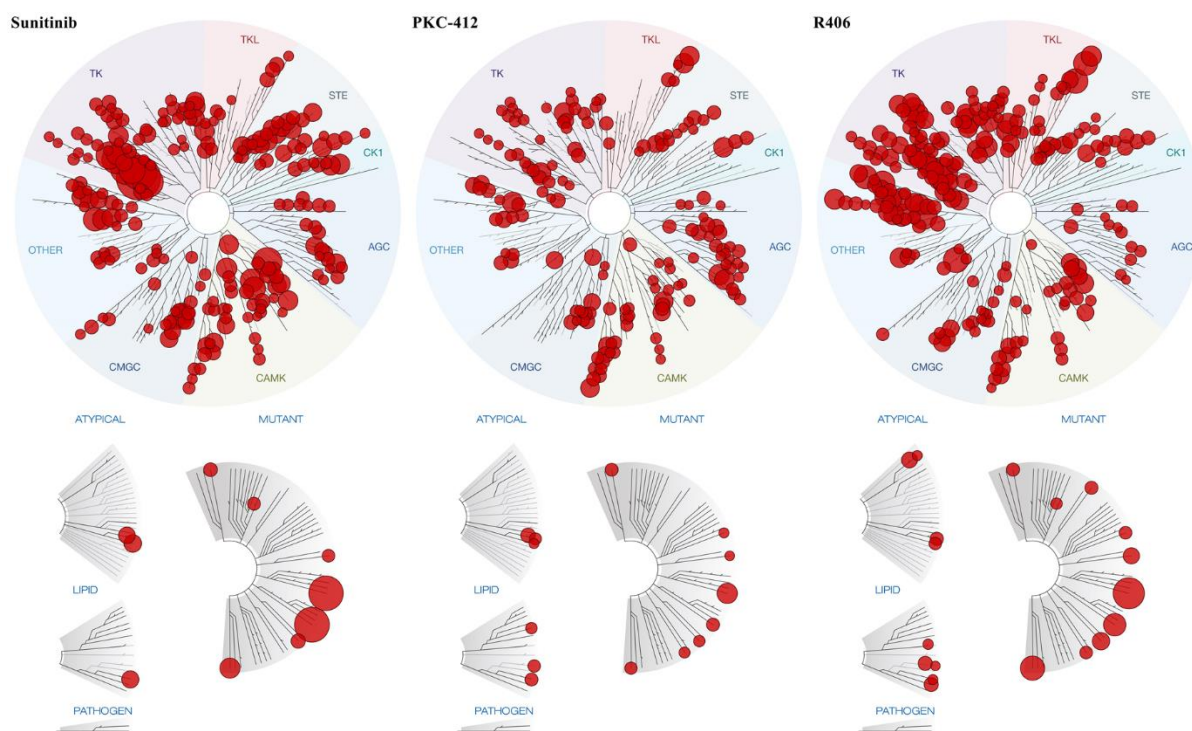


Figure 11. Kinome wide affinity data of Sunitinib, Midostaurin (PKC-412), and Tamatinib (R406). Larger circles represent strong affinities, a cutoff of 3  $\mu\text{M}$  affinity was chosen. Image generated using TREEspot™ Software Tool and reprinted with permission from KINOMEScan®, a division of DiscoverX Corporation, © DISCOVERX CORPORATION 2010.

The dual-specificity tyrosine-phosphorylation-regulated kinase 1A (DYRK1a) is a kinase with much more data available to the scientific community. It belongs to a kinase group (CMGC)<sup>118, 119</sup>, which is generally better understood in its significance in diseases. The DYRK1a syndrome is the primary pathological condition associated with the defunct DYRK1a<sup>120</sup>. Other syndromes with mental retardation have been brought into context with either hyper- or hypoactive DYRK1a (e.g., autistic spectrum disorders and down syndrome). The DYRK1a is expressed in developing nervous systems and



overexpressed in fetal brains of children with trisomy 21<sup>118, 120, 121</sup>. An inhibitor or modulator of DYRK1a might establish a more in-depth understanding of the underlying processes. Many compounds inhibiting the DYRK1a have been reported so far; as with every kinase investigated in this study, none of the inhibitors have reached clinical trials so far<sup>15, 119, 121-125</sup>. As some diseases are associated with hypoactive DYRK1a, a modulator instead of an inhibitor might be beneficial—a way to modulate the kinase directly or indirectly through the binding of upstream or downstream proteins. Over 20 known nuclear proteins interact with DYRK1a, which could hold potential for further development of DYRK1a targeting drugs in a molecular glue or Protein-Protein interactions inhibitor (PPI) fashion<sup>120, 126</sup>. The molecular interactions responsible for the binding of the interactome are unknown, and the discovery of novel binding sites at proteins could lead to further understanding of these pathways.

C-Jun N-terminal kinase 2 (JNK2 or MAPK9) and JNK3 (MAPK10) are two highly similar protein kinases of the MAP-kinase subfamily within the CMGC group. Together with the JNK1 (MAPK8), they form the c-Jun N-terminal kinases (JNK) subfamily. MAP kinases are well known, amongst other things, for their involvement in various cell differentiation and inflammation processes<sup>127-129</sup>. For JNK2, two distinct cancer mutations are known (V13M, K56N)<sup>130</sup>. The knowledge of the JNK3 is significantly greater than that of JNK2, especially if the amount of available crystal structures is considered, with only three structures of JNK2 and over 50 of JNK3. This imbalance of structural data might be the reason inhibitors are designed with JNK3 or JNK1 in mind, but inhibitors for the JNK2 are rarely seen<sup>131</sup>. As the sequence identity within the ATP binding pocket is considerably higher than in the rest of the kinase domain, the selectivity of most JNK3 inhibitors towards the JNK2 is hard to accomplish<sup>128, 132, 133</sup>.

The disrupter of telomere silencing protein 1-like protein (DOT1L) is a methyltransferase, which utilizes S-Adenosyl methionine (SAM) as a cofactor to methylate the L79 of the histone H3 up to three times<sup>134</sup>. The methylation changes the expression pattern and is an essential epigenetic factor<sup>135</sup>. In the context of most mixed-lineage leukemia (MLL) translocations, DOT1L has been studied intensively in the past decade, with many inhibitors found<sup>136-139</sup>. Either they are targeting the SAM binding site or adjacent pockets<sup>140</sup>. The FBDD approach has been especially productive in creating highly potent inhibitors<sup>137-139</sup>. With Pinometostat (EPZ-5676 in Figure 12), a clinical candidate was found in 2019, and the clinical trial is still ongoing<sup>141, 142</sup>.

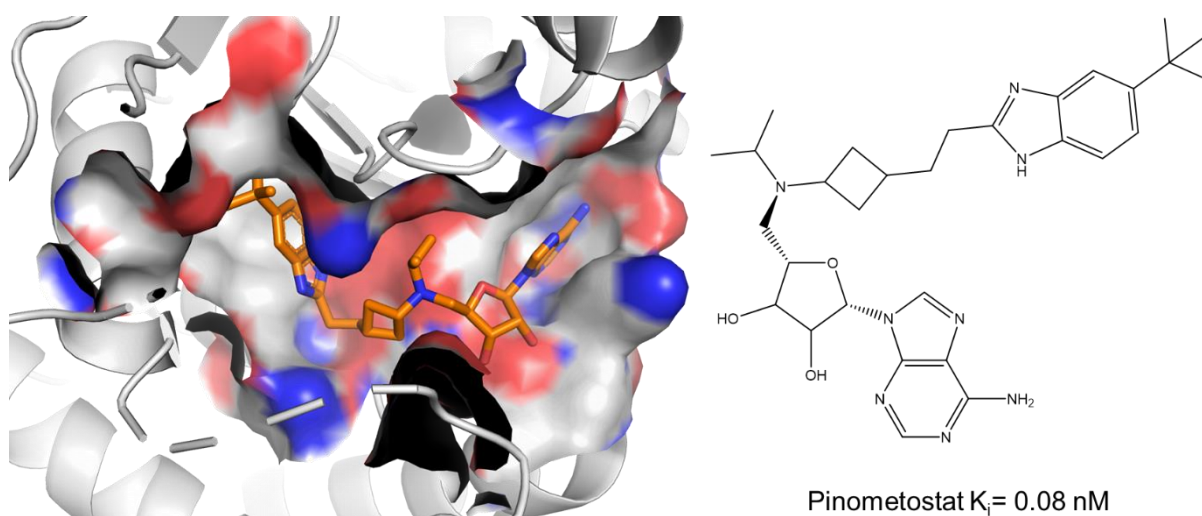


Figure 12. View of Pinometostat (EPZ-5676) in the binding pocket of DOT1L and the two-dimensional structure with the inhibitory constant.

BIRC5 (Baculoviral IAP repeat-containing protein 5), also called Survivin, due to its importance in cell proliferation and anti-apoptotic processes<sup>143</sup>. The underlying mechanisms by which BIRC5 prevents apoptosis are not completely uncovered yet. On a cellular level, BIRC5 is involved in the organization of the spindle apparatus with the chromosomal passenger complex (CPC)<sup>144</sup>. The functions of BIRC5 are further multiplied by repressing other critical proteins in apoptosis, like STAT3, FBXL7, and CASP7<sup>145-147</sup>. BIRC5 is physiologically expressed in replicating cells during mitosis; in contrast, a multitude of cancers have been discovered to use active BIRC5 to evade apoptosis. Many of these cancers utilize mutated versions of BIRC5, which can alter the interaction pattern of BIRC5<sup>145-147</sup>. The protein binds to various DNA sequences via a zinc finger in its monomeric form on a molecular level<sup>143</sup>.

In contrast, the relevance of the dimeric form is debated, as the molecular understanding of the BIRC5 associated network is incomplete<sup>143</sup>. Although a target, mainly expressed in cancer, has a great potential as a drug target, only a few inhibitors with improvable affinities have been reported so far (Figure 13)<sup>143</sup>. Even some compounds have been subjected to preclinical or clinical trials<sup>148</sup>. Most of the compounds discovered are functional inhibitors of BIRC5 and do not directly bind to the protein<sup>148</sup>. The reported compounds often lack the affinity and specificity needed in a tool compound in order to investigate the effects of BIRC5 inhibition<sup>143</sup>.

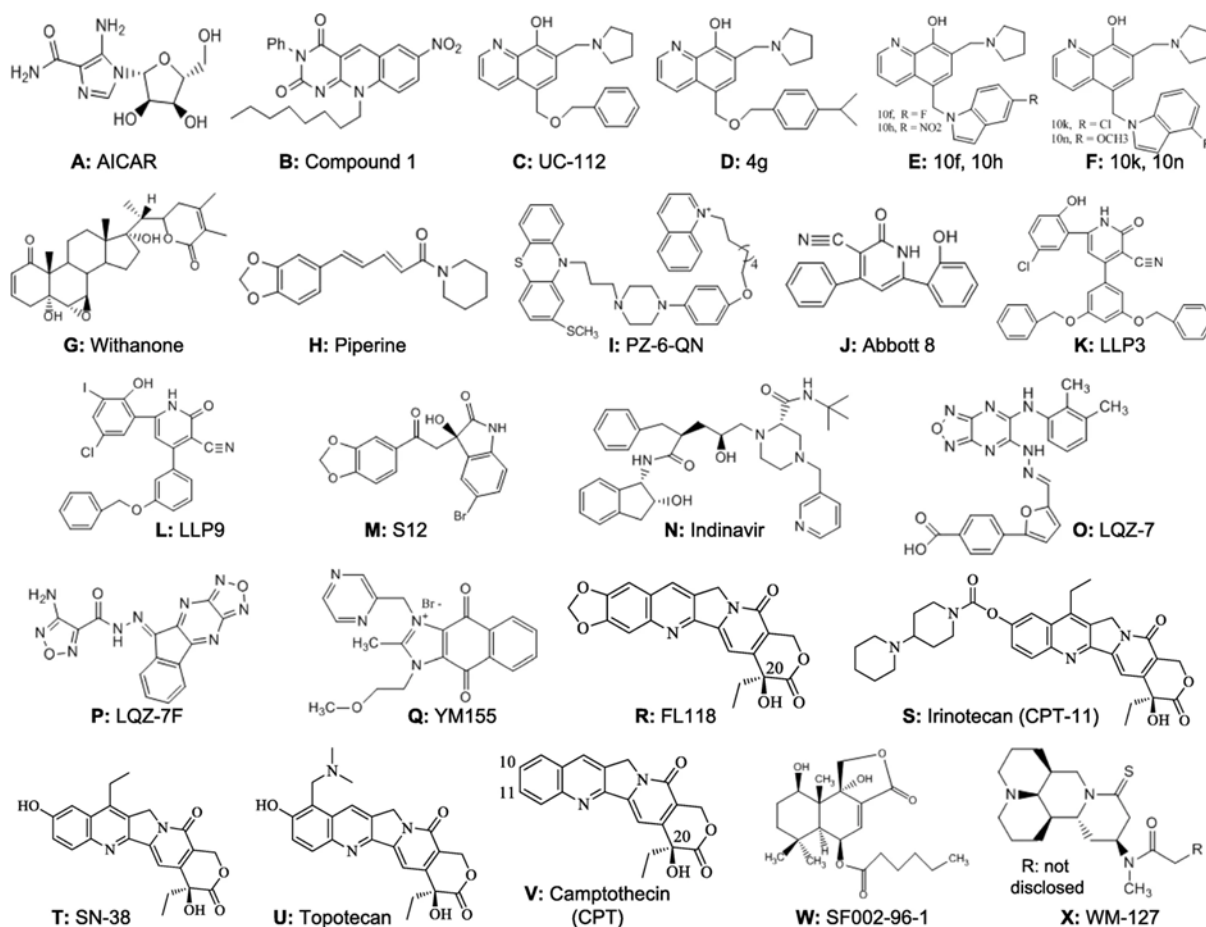


Figure 13. Overview of BIRC5 inhibiting compounds, copied under the CC BY 4.0 license (see Appendix D) from F. Li et al.<sup>143</sup>.

IDO1 (Indoleamine 2,3-dioxygenase 1) is the first and rate-limiting enzyme in tryptophan degradation ( $K_m=21 \mu\text{M}$ )<sup>149, 150</sup>. The concentration of tryptophan and its degradation product N-formyl kynurenine affect immune cells, especially T-cells. The enzyme is involved in the natural immunotolerance of the T-cells towards cells and is especially important in protecting a fetus from its mother's immune system<sup>151</sup>. The overexpression of IDO1 in various cancers has been shown as an indicator of a potentially worse treatment outcome<sup>151-153</sup>. The resulting tryptophane deficiency around a tumor down-regulates the T-cell activity. The IDO1 has been identified as one of the key factors of cancer immune system evasion<sup>151</sup>. As a result, IDO1 has been studied intensively in the past ten years as a target for allergies, autoimmune diseases, immune suppression, and cancer therapy. In 2018 the IDO1 inhibitor Epcadostat entered a phase III study<sup>154</sup>. Since then, enthusiasm for IDO1 has dampened as the clinical efficacy of IDO1 inhibition has often been elusive. Recent publications discuss the dual inhibition of IDO1 and histone deacetylases (HDAC) as means of cancer treatment<sup>155</sup>.

# 2. Results and Discussion

## 2.1 Target Selection

The present thesis showcases the abilities and characteristics of the HEFLib. The overall goal was to establish a biophysical screening routine to find, select, validate, and elucidate the binding of fragments to various proteins. The set of proteins selected for this study was chosen for diverging reasons. Technical aspects such as protein production, purification, stability, and the ease of crystallization were considered, as well as the pharmacological relevance of the targets (see section 1.5). Furthermore, proteins of interest with various binding sites from unrelated protein families were included. Simultaneously, part of the protein set should be well-established in the scientific community, whereas others should be chronically understudied. Lastly, we expected halogen bonding to improve selectivity between similar proteins. As a result, we wanted to include proteins with high similarity to test for potential binding preferences of binding fragments. One target fulfilling every mentioned criterion is impossible due to the inherent oppositions, and in turn, we utilized a set of proteins matching our demands. The diversification of targets with different advantages was chosen to adapt the library and screening process to a range of exemplary targets.

Eight proteins were selected for the screening process; five of these were kinases, and the other three consisted of the heme-dependent Indolamine 2,3-dioxygenase (IDO1), the histone-lysine N-methyltransferase DOT1L, and the inhibitor of apoptosis protein BIRC5 (Figure 14). These three proteins share no similarity neither in the mechanism of action nor general structure and folding pattern. Within the kinases, we selected very close kinases, especially the JNK2 and JNK3 are closely related (84% sequence identity and 88% similarity, calculated by SIAS<sup>156</sup>) and kinases on the different branches of the kinome (Figure 15). Unspecific binding, denaturation of the protein, or other systematic problems of an assay are easier to detect and isolate if multiple unrelated targets are utilized. The varying properties can be utilized to overcome problems with specific compounds, as one compound is unlikely to behave problematically with multiple proteins. However, a class of compounds has been proven to interfere with every or many assays. These compounds are generally referred to as pan assay interference (PAIN) compounds and can interrupt drug discovery projects at an early level<sup>157</sup>.

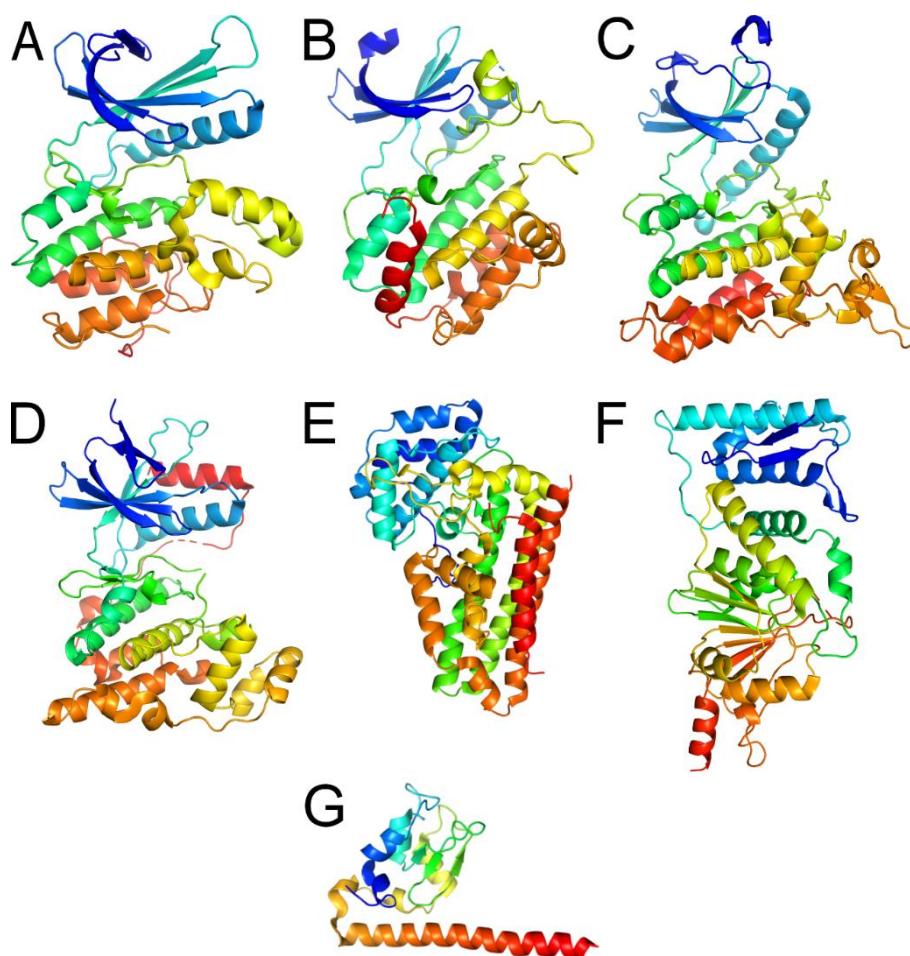


Figure 14. Structural overview of studied proteins, **A** AAK1 (PDB entry 5L4Q<sup>111</sup>), **B** CAMK1G (PDB entry 2JAM), **C** DYRK1a (PDB entry 2VX3<sup>120</sup>), **D** JNK3 (PDB entry 4X21<sup>63</sup>), **E** IDO1 (PDB entry 6E43<sup>158</sup>), **F** DOT1L (PDB entry 3QOW<sup>159</sup>), **G** BIRC5 (PDB entry 6YIH<sup>160</sup>). JNK2 is not depicted due to the high similarity to JNK3.

As the HEFLib has not been used yet, we wanted to account for the possibility of such PAIN compounds being present. By design, the structures in the HEFLib were subjected to a PAIN filter before acquisition<sup>7</sup>. Nonetheless, these filters are based on already identified substructures, and the prediction of PAINs has been painful<sup>161</sup>. Before the proteins were selected, we were curious if halogen bonding could occur in these targets. A script written by Karla Catacora was used to assess if the possibility existed. The binding sites of the selected targets were searched with an iodobenzene probe. The interaction geometries with highly favorable angles (around 180°) and without overlap in the Van der Waals radii of protein atoms and the benzene ring were criteria for possible halogen bonding hotspots. In most cases, sizeable possible interaction surfaces were present at the edge of the binding site (Figure 16 B, C, and D). BIRC5 is not shown in the hotspot analysis as binding sites in the protein are shallow and lead to large surfaces, which could not be called hotspots.

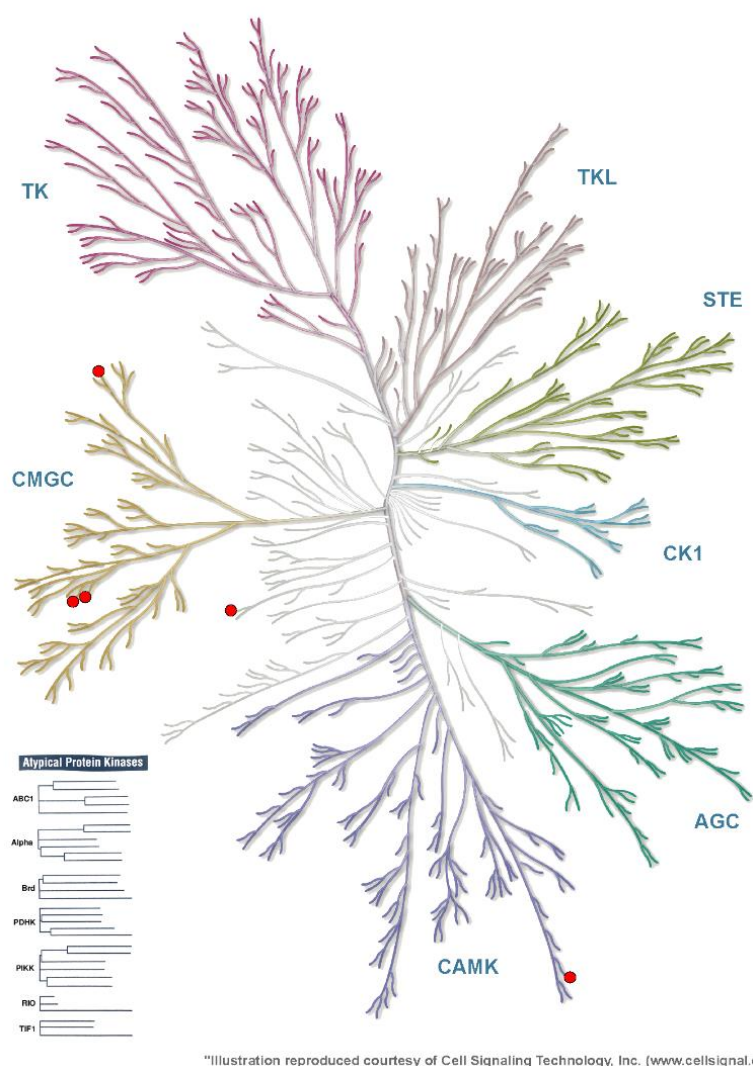


Figure 15. The kinome tree adapted from Manning et al.<sup>162</sup>, the red dots are the selected kinases in this work. Going counterclockwise from the top are the DYRK1a, JNK2, JNK3 (all CMGC family), AAK1, and CAMK1G.

IDO1 has the most enclosed pocket, and due to the heme occupying a large volume of said pocket, the remaining space was very limited for the iodobenzene (Figure 16 G). Perhaps explaining the low number and spots, halogen bonding could occur. Most of the encountered hotspots were backbone oxygen contacts with the occasional oxygen in a sidechain. The utilized search for hotspots is not a definitive measure as many parameters for an in-depth search were omitted (e.g., dynamics, pocket variability, averaging over multiple structures). In addition, the analysis could not include amino acid sidechains without a systematic evaluation of adduct formation energies (i.e., serine, threonine, cysteine).

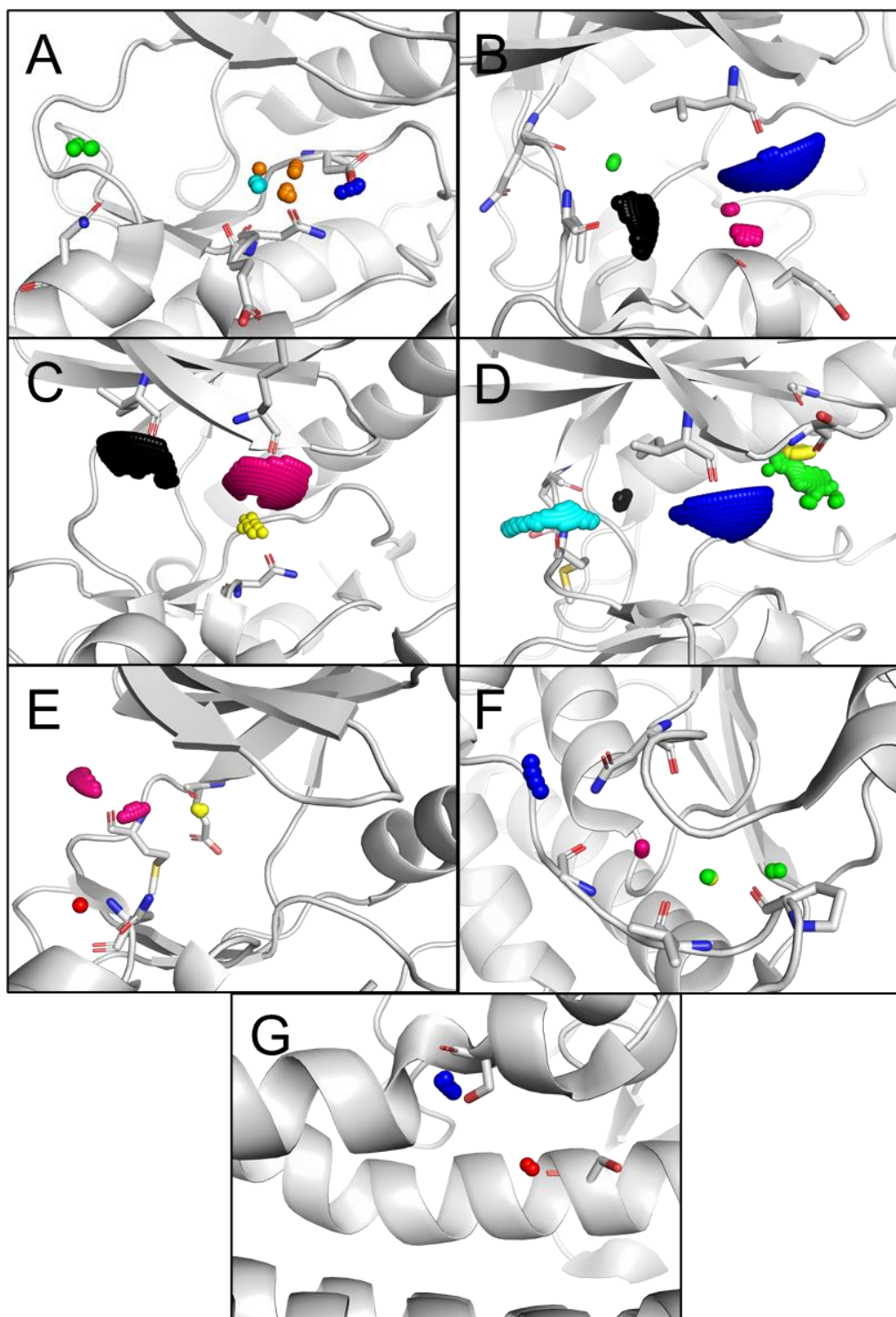


Figure 16. Halogen bond hotspot analysis, the spheres show possible halogen bond geometries with an iodobenzene. A: Binding site of AAK1, showing hotspots with residues N136(green), E180 (cyan), N181 (orange), and D194 (blue). B: Binding site of CAMK1g, showing hotspots with residues L28 (blue), Q99 (green), V101 (black), and E148 (magenta). C: Binding site of DYRK1a, showing hotspots with residues I165 (black), K167 (magenta), and N292 (yellow). D: Binding site of JNK2, showing hotspots with residues I32 (blue), S34 (green), G38 (yellow), E109 (black), and M111 (cyan). E: Binding site of JNK3, showing hotspots with residues E147 (yellow), M149 (magenta), and N152 (red). F: Binding site of DOT1L, showing hotspots with residues P133 (green), V135 (yellow), G137 (magenta), and N211 (blue). G: The Binding site of IDO1 shows hotspots with residues S167 (red) and S267 (blue).



## 2.2 Protein Purification

Panels B-E of Figure 16 feature the screened kinases. The ATP-pocket of all five proteins features potential hot spots with at least three different amino acids. Although the binding pocket of the kinases is highly conserved, some hot spots are found in only one kinase. DOT1L features the smallest hot spots, and it remains unclear if such exact geometries can be found in bound fragments (Figure 16 F).

## 2.2 Protein Purification

For ITC and crystallography, high yields and purities were critical for accurate affinity determination and crystallization, respectively. The expression and purification protocols were adapted and optimized to suit the demands, ensuring high yields with sufficient purity. Generally, a three-step purification protocol with an initial nickel affinity chromatography followed by a TEV protease digest and a reverse affinity chromatography was performed. The final polishing of proteins was done by size exclusion chromatography (SEC) to ensure high purity and incorporate an orthogonal purification method into the protocol.

### 2.2.1 AAK1 Purification

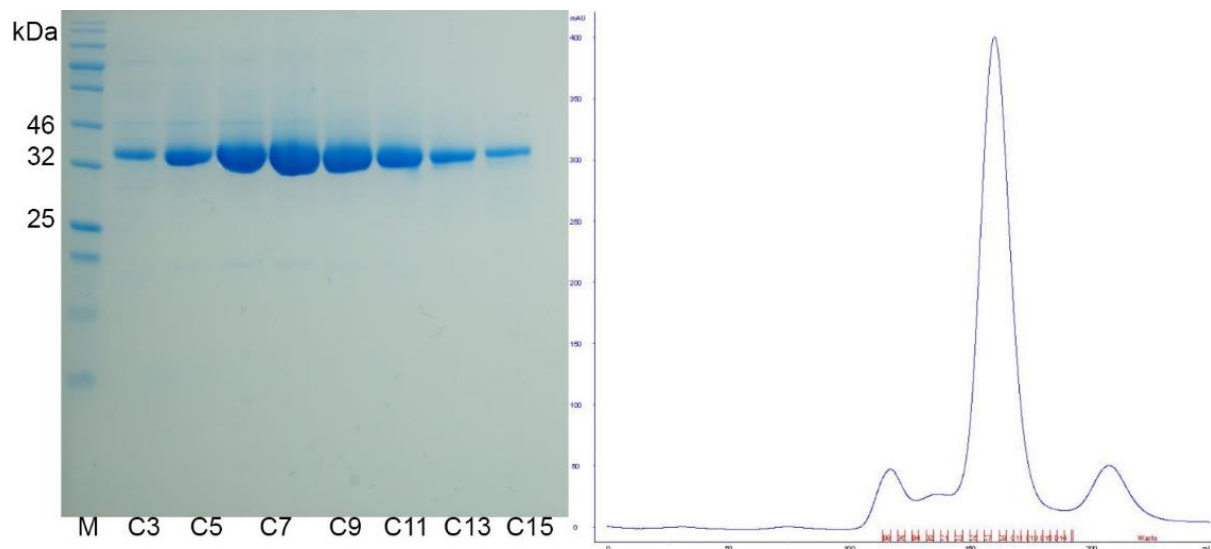


Figure 17. The left panel shows the SDS-PAGE of the pooled fractions after SEC. The left panel shows the corresponding chromatogram. The purity of AAK1 can be estimated to be greater than 95 %.

### 2.2.1 AAK1 Purification

Figure 17 shows that the purification of AAK1 was successful; the purity can be estimated as greater than 95% by the SDS-PAGE. In addition, the AAK1 was expressed in high yields, typically in the range of 20 to 25 mg per liter of cell culture. The AAK1 was co-expressed with  $\lambda$ -Phosphatase to ensure both phosphate groups, which were added by autophosphorylation within the *E. coli*, were hydrolyzed. This construct will be used in the screening of crystallization conditions<sup>163, 164</sup>.

The dephosphorylation was confirmed by ESI-MS (Figure 18, panels A and B). We postulated the dephosphorylation as necessary for crystal packing. All of the published structures show crystal contacts performed by tyrosine 234 to arginine 64 and aspartate 49<sup>111, 164</sup>. This contact becomes impossible if the tyrosine is phosphorylated. Two negative charges would be in close contact (D49 and phosphate group), and there is insufficient space for a phosphate group (Figure 18 C) through a clash with the R46. The mass of the dephosphorylated AAK1 was 39291 Da (theoretical mass of 39292 Da). The not co-expressed AAK1 had a mass of 39374 Da (theoretical mass of 39372 Da). The MS experiment supports the dephosphorylation with a difference of 83 Da ( $\text{PO}_3\text{H}^- = 80 \text{ Da}$ ). Both MS experiments show a good correlation between theoretical mass and measured mass if protonation states are factored in. A phosphate group at Y235 would change the crystal contacts in this particular region, which might have been why all crystallization efforts were unsuccessful (Figure 18). The co-expressed AAK1 has not been crystallized yet, still owning the evidence that the Y235 crystal contact is essential for crystallization.

## 2.2.1 AAK1 Purification

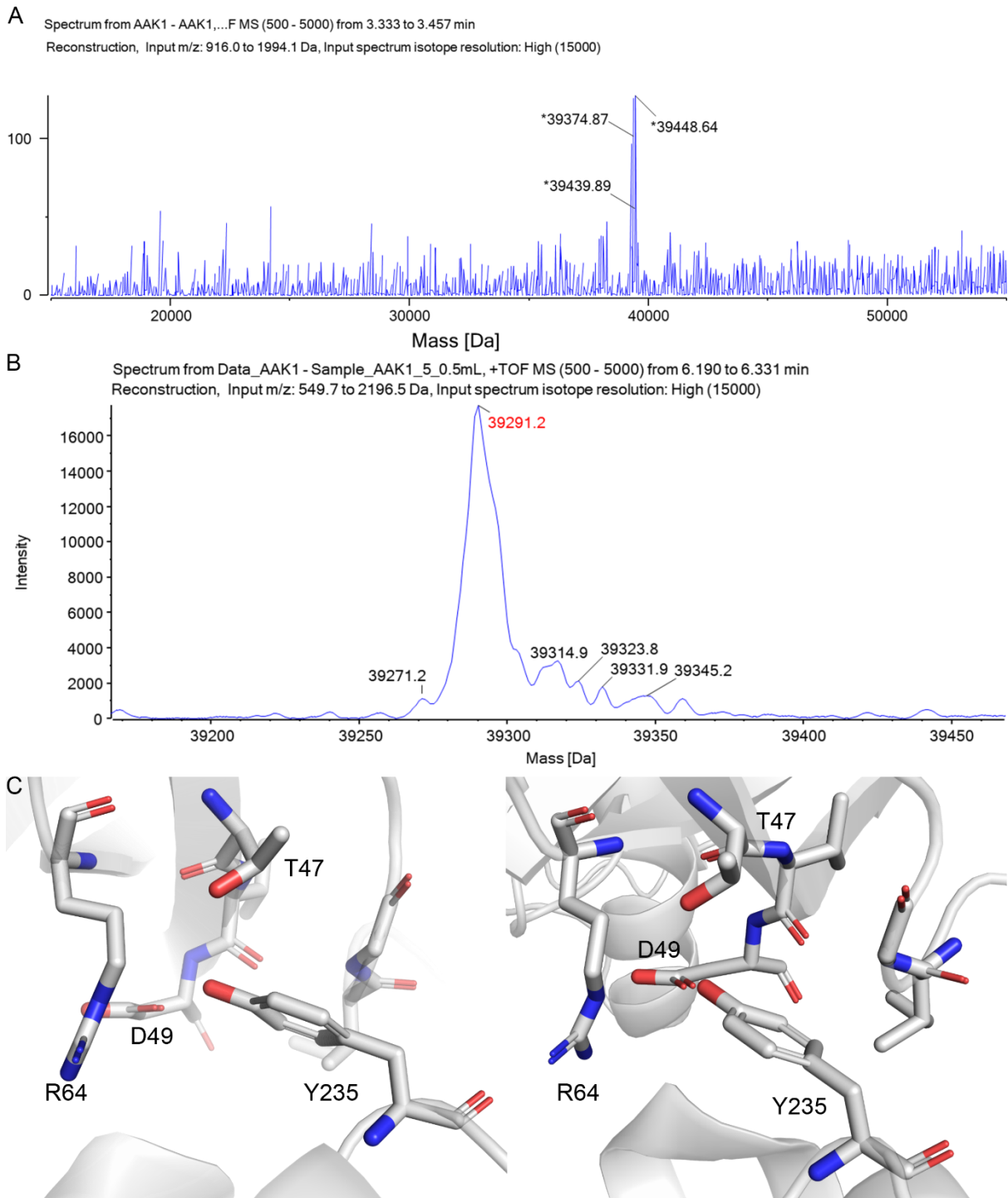


Figure 18. **A**: ESI-MS spectrum of autophosphorylated AAK1, the mass is 39374.87 Da. **B**: ESI-MS spectrum of AAK1 co-expressed with  $\lambda$ -phosphatase with a reduced mass of 39291.2 Da, deviating from the theoretical mass by 1 Da. **C**: two views of the same crystal contact from 5L4Q, the distance from the tyrosine (Y235) oxygen is 3 Å to T49, 2.4 Å to D49, and 3.4 Å to R64. A phosphorylated Y235 could not sustain this crystal contact. As inadequate space and charges are present.

## 2.2.2 BIRC5 Purification

Figure 19 shows the last purification step of BIRC5 (Survivin); the SDS-PAGE shows small residual impurities and a band at 32 kDa, which could be small amounts of dimeric BIRC5. The BIRC5 forms a dimer as the active DNA binding conformation in the physiological buffer solution. In the chromatogram (Figure 19), the smaller second peak (B9) after 190 mL of elution corresponds to the monomer, and the first peak (A7-A15) after 150 mL of elution corresponds to larger proteins of roughly 30-40 kDa. Monomeric BIRC5 has a mass of 16 kDa, too small for an SEC peak after 150 mL of elution. Both peaks were subjected to the ESI-MS, and it was proven that both SEC-peaks contain a protein with a mass of 16606 Da (theoretical mass of 16608 Da). The SDS page confirms the similarity in size of both peaks seen in the SEC, which indicates that most of the BIRC5 was present as the dimer during purification. Overall, purification was sufficient in yields of around 20 mg per liter culture. The purity of BIRC5 was deemed sufficient for NMR and ITC measurements.

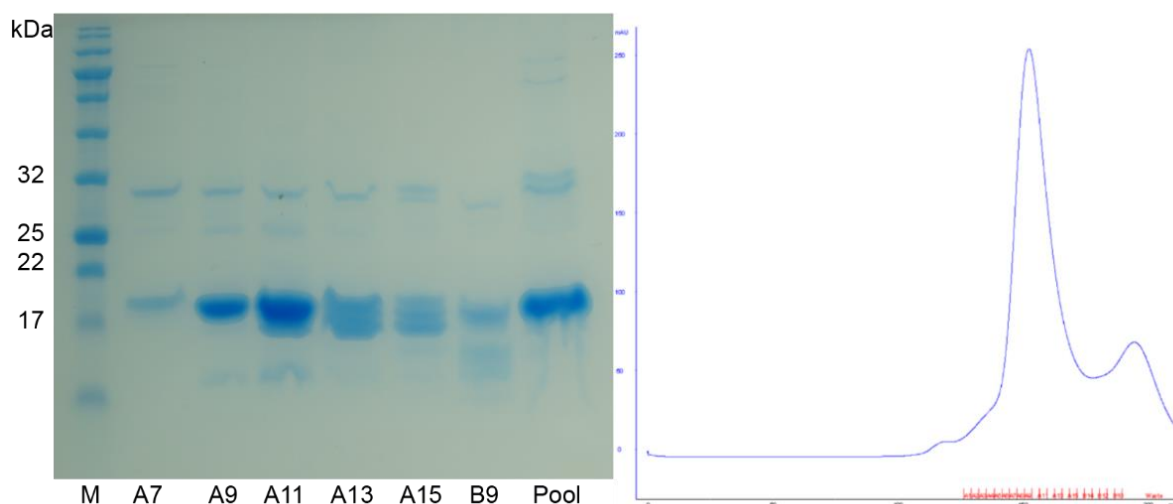


Figure 19. SDS-PAGE and chromatogram of a BIRC5 SEC run. The band at 32 kDa corresponds to small amounts of dimer still present in the SDS buffer, and monomeric BIRC5 can be seen at 17 kDa. In the purification buffer solution, BIRC5 was in dimeric form, which was the reason for the similar retention time in the SEC column (150 mL), comparable to a 30-40 kDa protein like the AAK1. Both SEC peaks have a similar protein composition if SDS-PAGE sample B9 (second peak) is compared to A9 (first peak), for example.

### 2.2.3 CAMK1G Purification

The expression and purification of CAMK1G were more economical than most other proteins. First, the purity was sufficient after the reverse nickel column, and an SEC was unnecessary. Second, the yield of CAMK1G was higher than with AAK1 or most other proteins with 40mg per liter of culture. Figure 20 shows the purity of the flow-through (FT) and wash (W) fractions, which were combined for NMR and ITC measurements.

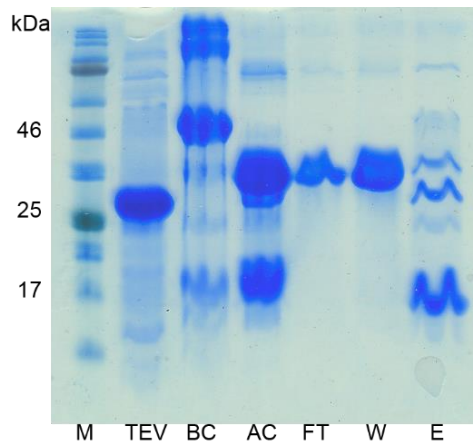


Figure 20. SDS-PAGE after the reverse nickel column from left to right, the fractions are protein standard (M), TEV protease (TEV), CAMK1G before the TEV digest (BC), CAMK1G after digest (AC), the flow-through fraction (FT), wash fraction (W), and elution fraction (E). The SDS-PAGE illustrates the complete digest; uncut protein would remain at 46 kDa (BC), whereas digested CAMK1G lies at 32 kDa (AC). The purity increased immensely, as can be seen by comparing the BC fraction with the W fraction, or E, respectively.

## 2.2.4 DOT1L Purification

### 2.2.4 DOT1L Purification

As shown in Figure 21, the purity by the SEC was exceptional for the DOT1L. The yield for the DOT1L was 16 mg per liter of culture.

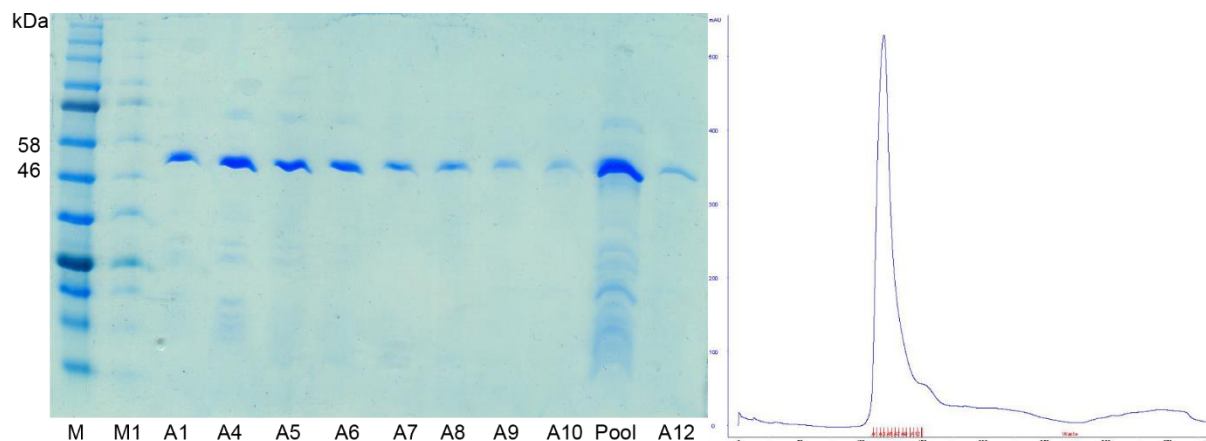


Figure 21. SDS-PAGE and chromatogram DOT1L SEC run. The fractions A1 to A12 are pure and were pooled, and the Pool fraction on the gel was the protein sample before SEC.

### 2.2.5 DYRK1a Purification

DYRK1a was obtained in sufficient yield, in the range of 20 mg per liter culture, and the purity was sufficient for crystallization (see section 2.6.1). From Figure 22, a purity of >95% can be deduced by estimating the gel and chromatogram combined. The shoulder in front of the peak was always present, which would not be collected to ensure high purity for crystallization.

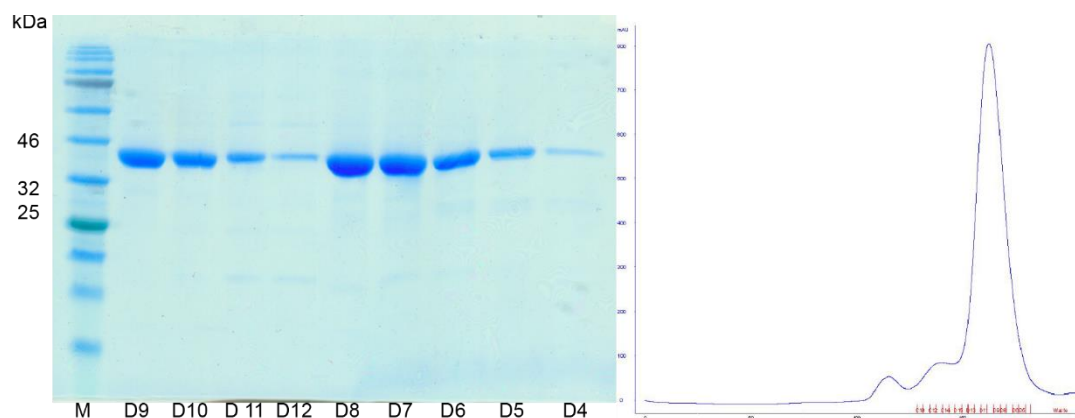


Figure 22. SDS-PAGE and chromatogram of a DYRK1a SEC run. Fractions D12 to D4 were collected, and the shoulder in front of the peak was discarded.

### 2.2.6 IDO1 Purification

While purity and yield are sufficient to characterize the effectiveness of expression and purification of most proteins, the heme incorporation into the IDO1 is essential in obtaining functional IDO1<sup>165</sup>. Therefore, the absorption of the iron atom in the heme cofactor was measured at 405 nm next to the standard 280 nm during purification, as stated by Takikawa et al. and subsequently adopted by many other authors<sup>166</sup>. Takikawa et al. stated that the maximum heme incorporation is achieved at a quotient of  $A_{405} \div A_{280} = 2.2$ <sup>166, 167</sup>. Most publications stated a heme saturation of around 80 % compared to the Takikawa publication<sup>149, 167, 168</sup>. The IDO1 used in this study had a heme saturation of 81 %. It has to be stated that the IDO1 has to be in the inactive ferric state ( $\text{Fe}^{3+}$ ) because the absorption band at 405 nm of heme was roughly doubled in intensity in the ferric compared to the ferrous state ( $\text{Fe}^{2+}$ ).

Takikawa et al. purified the IDO1 from human tissue over one week without any reducing agents present in the buffers<sup>166</sup>. Naturally, the contact with oxygen and the presence of substrate in the human placenta cells would lead to the purification of the ferric IDO1. Most publications call for reducing agents in purification buffers (DTT)<sup>169</sup> to prevent the oxidation and formation of disulfides. These reagents could keep the iron reduced, although activity assays call for additional reducing agents like ascorbic acid and methylene blue<sup>165</sup>. I could observe a substantial increase in absorption at 405 nm after buffer exchange from a DTT containing to a TCEP containing buffer, which resulted in calculated heme incorporation from 30 % to 80 %. Besides the different absorption readings, a distinctive color change could be observed from a light red ( $\text{Fe}^{2+}$ ) of the IDO1 in the DTT containing buffer to the brown ( $\text{Fe}^{3+}$ ) of the IDO1 in the TCEP containing buffer ( $\text{Fe}^{3+}$ ). Both observations indicate the oxidation of the iron. These were further confirmed as the reaction was reversible by adding a thiosulfate solution. The absorption interference of DTT could be ruled out as the same observations were made with  $\beta$ -ME instead of DTT. TCEP could function as a surrogate substrate to tryptophane, or TCEP was incapable of preventing the autooxidation of the heme. Independent of the mechanism, a ferric IDO1 was obtained in the ITC buffer. The NMR buffer never contained any DTT or TCEP and showed no significant absorption increase. Adding TCEP to the NMR buffer also increased the absorption at 405 nm. It has to be assumed that most IDO was in the ferrous state in the NMR samples and the ferric state during ITC measurements.

## 2.2.6 IDO1 Purification

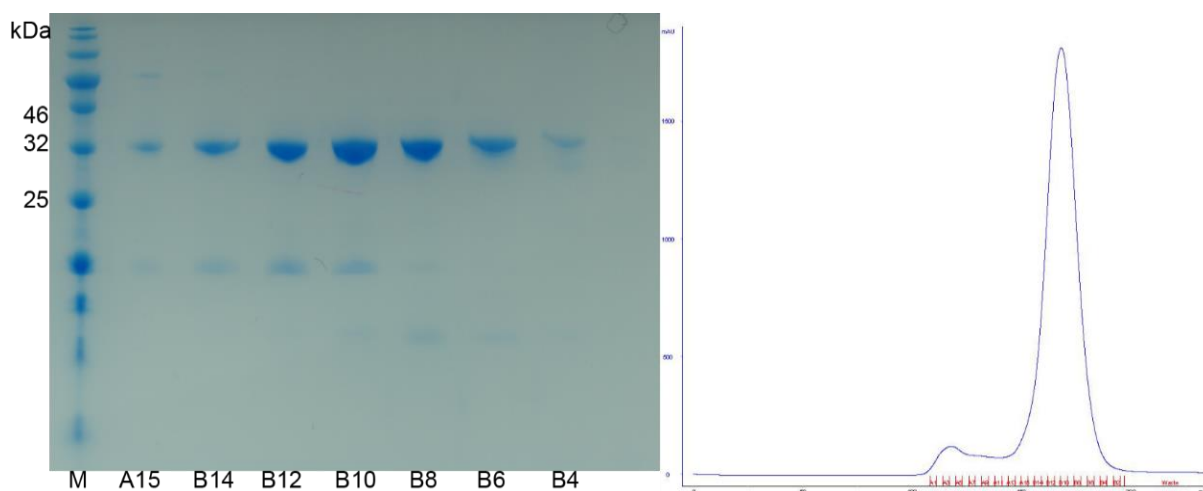


Figure 23. SDS-PAGE and chromatogram of an IDO1 SEC run. Fractions A15 to B4 were pooled. An impurity with a size of 25 kDa was present in small quantities and should not significantly interfere with STD or ITC measurements.

This could affect the binding mode of fragments, as inhibitors selective for one or another oxidation state have been described<sup>149, 150, 165, 169-172</sup>. During this study, Nelp et al. found competitive heme inhibitors for IDO1 and could show that the heme saturation of IDO1 in human cells is incomplete, with only 15% of IDO1 being bound to heme<sup>173</sup>. Therefore, the heme content established by Takikawa et al. might represent only a fraction and not the accurate, complete saturation of IDO1 with heme. Increasing the heme saturation has been tried by further incubation with ferric heme, which is considerably weaker, than the affinity of ferrous heme<sup>173</sup>. Other publications have incubated IDO1 with saturated heme solutions and could not increase the heme saturation above the reported value by Takikawa et al.<sup>166</sup>. On the one hand, this might be a sign of the lability of heme binding to IDO1, or on the other hand, point to the maximum possible heme content found by Takikawa. Besides the unclear situation regarding the heme content, it was possible to generate IDO1 in sufficient yields of 17 mg per liter culture with only a tiny impurity and a heme content of over 80% (Figure 23).



## 2.2.7 JNK2 and JNK3 Purification

Both JNKs were obtained in enormous 50-70 mg per liter culture yields. The purity for NMR or ITC was less critical, and small impurities could be tolerated. For protein crystals, purity was the most crucial factor. Therefore, only fractions without detectable impurities were collected, reducing the yield by 50%. As shown in Figure 24 and Figure 25, the purity of both kinases was over 90% and 95% for JNK2, respectively.

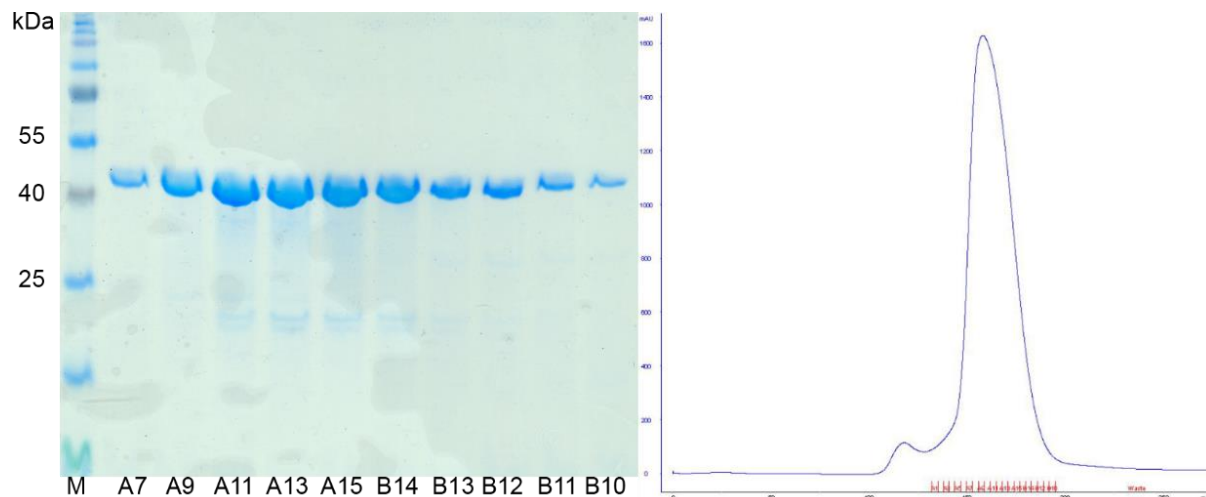


Figure 24. SDS-PAGE and chromatogram of a JNK2 SEC run. Fractions A7 to B10 were pooled. Purity from the gel and chromatogram can be estimated to be greater than 98%.

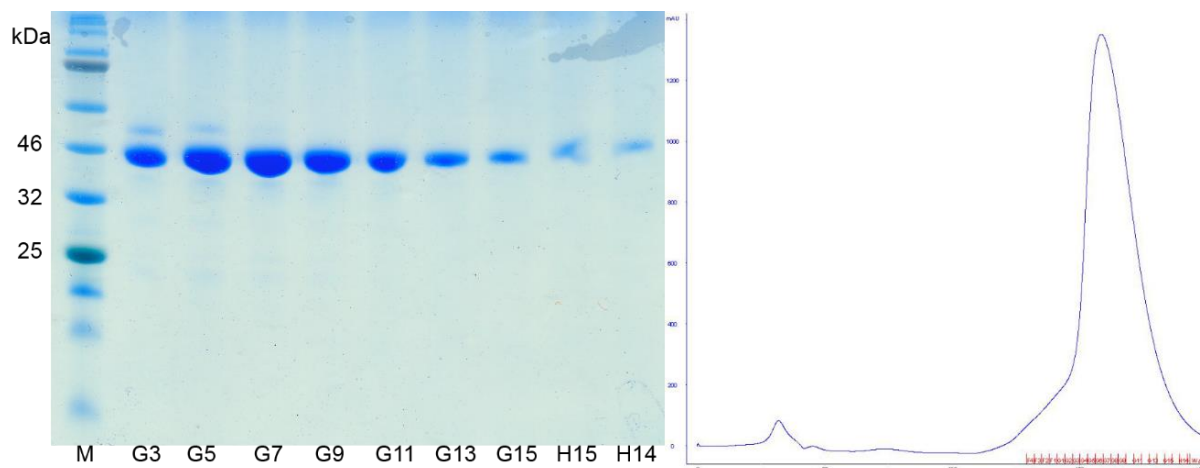


Figure 25. SDS-PAGE and chromatogram of a JNK3 SEC run. Fractions G7 to H14 were pooled. The shoulder in front of the peak was pooled in a second SEC run to increase yield while maintaining over 99% purity.

### 2.2.8 GSK3 $\beta$ Purification

The GSK3 $\beta$  kinase was attempted to be cultivated in *E. coli* cells as published by Stamos et al.<sup>174</sup>. Most previous publications reported expressing GSK3 $\beta$  in insect cells or yeast cultures<sup>175</sup>. The main reason for our interest in the GSK3 $\beta$  was derived from its constitutive activity without upstream activation, which would have allowed us to perform inhibition assays instead of binding assays. Unfortunately, it was impossible to express GSK3 $\beta$ , although various protocols were applied to detect the expressed target protein. These protocols included using different bacterial cell lines, aside from standard BL21 cells; *Vibrio natriegens* VMAX cells<sup>176</sup>, *E. coli* Rosetta2, *E. coli* RIL, and *E. coli* C41 cells were used. The expression conditions were thoroughly screened. Different media (LB, 2xYT, TB) were used in various temperature regimes (expression at 4° up to 37°C) combined with different IPTG concentrations were tested for protein expression every three hours over 24 hours by Western-Blot and ELISA. The cultures were tested to determine whether the formation of inclusion bodies was the cause of the lack of protein<sup>177</sup>. In some cultures, false-positive impurities with the wrong sizes were found. These false positives were mainly observed due to the anti-histidine-tag antibody used in the ELISA, which is inherently prone to unspecific binding. ESI-MS further identified false positives.

A reverse transcription PCR was used, as shown in Figure 26, to check whether the translation or transcription was the cause of the lacking protein expression. Copious numbers of GSK3 $\beta$  mRNA were found, but no protein could be detected in the same culture. Although the DNA was optimized for expression in *E. coli* by the GeneArt service of Thermo Fisher, the RNAfold web server of the University of Vienna predicted a sophisticated secondary structure of the mRNA of the optimized DNA (Figure 26)<sup>178</sup>.<sup>179</sup> DNA obtained by the RT-PCR was subsequently sequenced to verify if primers had worked correctly. No DNA could be detected in sample A. A handling error of the PCR-Kit likely caused the lack of DNA because the rest of the samples could be analyzed as intended. A high amount of DNA copies and, to a lesser extent, RNA of the right size could be detected. Therefore a problem in initializing the translation was assumed.

## 2.2.8 GSK3 $\beta$ Purification

Another GSK3 $\beta$  plasmid was acquired, this time the human cDNA from Addgene (RRID: Addgene\_82126)<sup>180</sup>, to circumvent the problem in the translation, but still, no protein could be detected in the expression, and in turn, efforts were halted. It is still unclear which DNA construct was used by Stamos et al., but the evidence suggests a high dependency on the correct DNA construct to express GSK3 $\beta$  in *E. coli* cells.

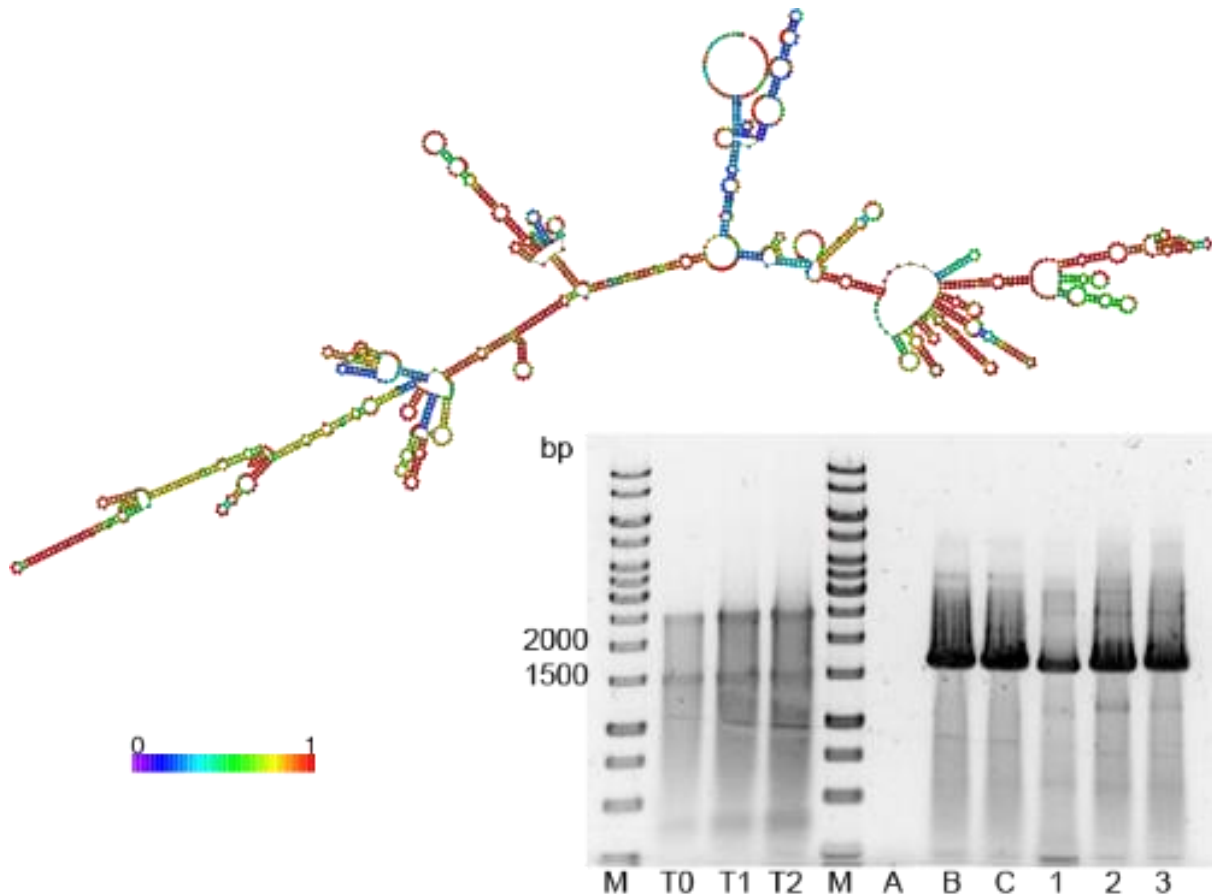


Figure 26. Predicted RNA secondary structure of GSK3 $\beta$  by RNA-fold Webserver<sup>179</sup> (University of Vienna), color scheme indicates the probability of secondary structure, with blue for low and red for high probabilities. RNA of three samples was isolated, at induction (T0), after 5 hours (T1), and after 20 hours (T2) of induction. The three samples were subjected either directly to RT-PCR (1-3) or previously heated to 65°C to destroy the secondary structure (A-C). Samples B, C, and 1 to 3 were verified by sequencing to be GSK3 $\beta$ . GSK3 $\beta$  with its HLT-tag was 1576 bp long.

### 2.3 Chemical stability of fragments

In the library's design, special attention was paid to the magnitude of the  $\sigma$ -hole. An increase in positive electrostatic potential can increase the maximum strength of such a bond. For a tuned halogen in an aromatic scaffold, it is generally helpful to use an electron-deficient scaffold, e.g., pyridines, pyrimidines, or electron-withdrawing groups such as fluorine sulfonyl or nitro groups<sup>181</sup>. As the  $\sigma$ -hole increases, the possibility for a covalent reaction with a proteinogenic nucleophile (mainly cysteines) increases as well. Halogens are generally considered good leaving groups in the context of different organic reactions. Fluorine is in most cases an influential leaving group in an  $S_NAr$  type reaction, whereas iodine is considered to be the most reactive in an  $S_N2$  or  $S_N1$  type reaction. Chlorine and bromine usually are sorted in between. If the carbon halogen bond is strongly polarized, a covalent reaction might occur with less activated cysteines or other nucleophiles. The HEFLib was designed with the potential activity in mind, and fragments or building blocks with known reactivity were omitted<sup>7</sup>. As the study progressed, specific STD measurements raised the question of whether the reactivity of some fragments was underestimated. Other potential covalent moieties might be present besides the fine line between exceptionally tuned  $\sigma$ -hole and increased reactivity.

Six fragments were selected for a GSH stability assay (Figure 27). Fragments 1223 and 1255 have been reported as covalently binding compounds<sup>182, 183</sup>. Fragment 1234 showed strange behavior in the screening and validation phase. These three fragments featured a time-dependent degradation with GSH. The fastest reaction showed fragment 1223 being undetectable after one hour, making it impossible to calculate a half-life from only two points. Fragments 1234 and 1255 reacted with GSH and showed half-lives of 5 hours and 19.2 hours, respectively. The calculated half-lives of fragments 1224, 1213, and 1253 were considerably longer than 20 hours and therefore considered stable for this work. In the meantime, more fragments with peculiar behavior have been tested with this assay (e.g., 0474, 1216, 1217, 7409, 9595, and 9809), and none showed a half-life shorter than 70 hours. Although most of the fragments in the HEFLib do not feature stability issues, every hit in the future should be checked for stability.

The results were obtained after the initial STD screening, and the question arises if covalently bound fragments can be detected. There have been FBDD studies with covalent compounds utilizing the STD-NMR technique<sup>184</sup>. Nonetheless, the literature on STD-NMR with covalent binding compounds remains scarce. The covalent binding mode adds complexity to the data interpretation, as the STD technique relies on ligands' on and off kinetics. If a reactive compound can bind to a protein in a noncovalent fashion with suitable kinetic parameters, an STD signal could be observed. In addition, depending on the reactivity, a binding site might not have any suitable nucleophile. The potentially reactive compounds could bind in a noncovalent manner in addition to an unspecific covalent reaction. On the other hand, false-positive signals could arise from aggregated compounds and protein, or the proton signals could be shifted as the covalent moiety could exert different chemical shift anisotropy. It remains elusive, which species is responsible for such signals.

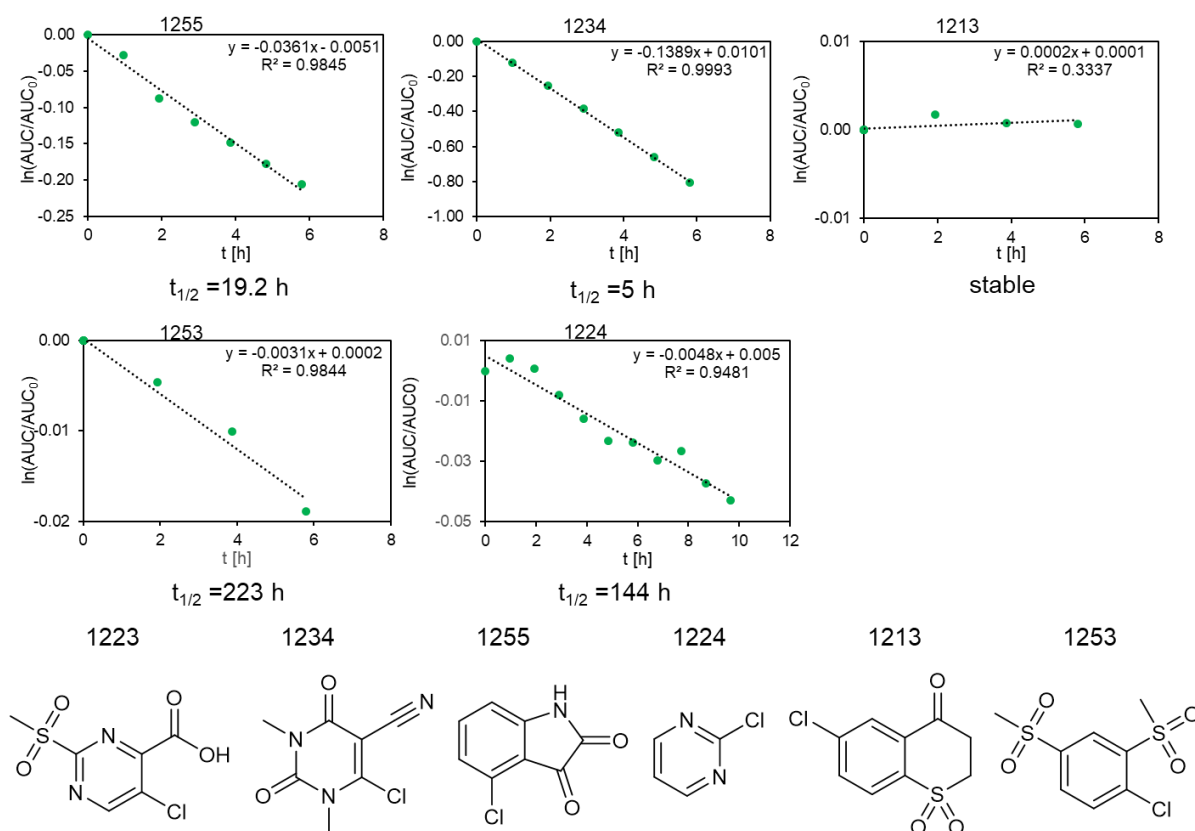


Figure 27. Stability assay for six fragments with GSH. The  $\ln \frac{AUC}{AUC_0}$  of the HPLC peak is plotted against the time. Fragments 1255 and 1234 are classified as unstable with an extrapolated half-life of 19.2 hours and 5 hours, respectively. Fragments 1213, 1253, and 1224 were considered stable as the half-life extends any possible type of testing or assay. No half-life could be estimated for fragment 1223, as it had reacted entirely with GSH after one hour.

### 2.3 Chemical stability of fragments

After these considerations, the decision was made to include the three unstable fragments in the screening and evaluation process. Furthermore, in an STD-NMR, fragment concentrations are typically 20 to 100 times in excess of the protein. Therefore, the fragment can engage in a noncovalent binding motif, albeit various nucleophiles might be covalently saturated. It must be mentioned that all results were interpreted with the possibility of covalent reactions in mind. In later ITC measurements, fragment 1234 showed prolonged and constant heat release with every tested protein, signifying an unspecific covalent reaction in the binding event. Consequently, the STD-NMR results of 1234, showing a signal with every protein, were deemed artifacts and excluded from further work (see section 2.5.8).

### 2.4 STD-NMR

The initial screening of targets was performed by STD-NMR<sup>30, 31, 33</sup> as a time and protein efficient technique, combined with the additional benefit of gaining rudimentary information about the relative position of fragment protons relative to the protein. The high sensitivity and broad spectrum of applicable proteins are further advantages, which led to the widespread adaptation of STD-NMR as a means of fragment screening<sup>5, 32</sup>. Part of the study was the initial testing, adaptation, and optimization of protocols as necessary.

#### 2.4.1 STD-NMR Optimization

One of the goals of this work was to establish a reasonable fast screening protocol for the nearly 200 fragments with the potential to increase the library at a later point in time. To accomplish this task, a streamlined protocol had to be developed. A compromise between signal intensity (or signal-to-noise ratio) and measuring time for all parameters had to be achieved. As a starting point, JNK3 was used as the target for optimization due to the vast amounts of available protein, and some fragments had been established as hits by DSF<sup>102</sup>. The main goal of the first optimization was the reduction of the protein concentration to a minimum. This would enable screening with the least amount of protein. 10  $\mu$ M of JNK3 resulted in poor spectra with the previously established fragments 9595 and 9601<sup>102</sup>. A final concentration of 20  $\mu$ M of protein proved to be sufficient. Secondly, the saturation time could be reduced from 4 seconds to 3 seconds while reducing the scans from 64 to 16 on-resonance scans. A water suppression pulse sequence (WATERGATE) was examined, but the resulting benefit in signal-to-noise ratio was deemed not worth the extra measuring time. A fragment concentration of 2 mM was used with the first 60 fragments. We found a hit rate exceeding 30% and reduced the screening concentration to 1 mM. In hindsight, we might reduce the screening concentration further by applying the WATERGATE sequence due to the increased signals.

#### 2.4.1 STD-NMR Optimization

NMR spectra measured in water are highly susceptible to changes in pH value. With several fragments bearing either acidic or basic groups, the buffer in the STD measurements had to stabilize the pH at millimolar concentrations of strong acids or bases. After searching through the library, ortho hydroxybenzoic acid was one reoccurring strong acid with a known  $pK_a$ , which was used as a benchmark for the buffer calculation. PBS was chosen as the NMR buffer with increased buffer capacity to ensure a change in pH to be  $<0.02$  at a 2 mM acid/fragment concentration. A standard  $^1\text{H-NMR}$  was recorded for every fragment to guarantee the identification of compound mixtures and detect eventual compound degradation over the years. In the final screening protocol, mixtures of two fragments were used (see Table 2) to decrease the measuring time further and increase throughput per session.

The fragment mixtures could be increased to 3 or 4 different fragments without signal overlap. At this rate, the complete library could be measured in one session (autosampler was limited to 60 samples), and the extra measuring time from water suppression could be neglected. A water suppression protocol would further increase the sensitivity, which could be helpful with more challenging targets. The fragment mixtures were derived from the  $^1\text{H-NMR}$  to ensure no overlap in proton signals. Fragments 0091, 1101, 1151, 1215, 1220, 1231, and 4484 lack an observable proton in an aqueous solution and could not be screened. This was confirmed by reference  $^1\text{H-NMR}$  in an aqueous solution. Every fragment has a barcode assigned to it; the last four digits are used as identifiers in the study. Table 2 summarizes the mixtures arranged for the screening of all proteins except JNK3.

Table 2. Mixtures of fragments in the STD-NMR screening. The last four digits of the barcode are sufficient to identify the fragment.

Mixture No.	Reference NMR	Barcode	Stock No.	Reference NMR	Barcode	Stock No.	Well
1	J3A 1	B02289595	1	J3A 52	B02294485	65	A1
2	J3A 2	B02289597	2	J3D 8	B02289613	46	A2
3	J3D 1	B02290522	3	J3A 16	B02290469	21	A3
4	J3D 2	B02290525	4	J3A 25	B02290464	32	A4
5	J3A 3	B02290523	5	J3B 45	B02297424	131	A5
6	J3A 4	B02297655	6	J3C 27	B02291272	198	A6
7	J3A 5	B02290658	7	J3A 32	B02289616	39	H12
8	J3A 6	B02290113	8	J3D 15	B02290660	70	A8
9	J3D 3	B02290114	9	J3D 40	B02291152	160	A9
10	J3A 7	B02290120	10	J3A 40	B02289617	48	A10
11	J3A 8	B02290118	11	J3A 34	B02289610	41	A11



## 2.4.1 STD-NMR Optimization

Mixture No.	Reference NMR	Barcode	Stock No.	Reference NMR	Barcode	Stock No.	Well
12	J3A 9	B02290117	12	J3C 7	B02294482	168	A12
13	J3A 10	B02290121	13	J3D 43	B02297765	167	B1
14	J3A 11	B02290115	14	J3A 17	B02290473	23	B2
15	J3A 12	B02290119	15	J3A 42	B02289603	50	B3
16	J3A 13	B02290122	16	J3D 44	B02290661	174	B4
17	J3D 4	B02290116	17	J3A 36	B02289612	43	B5
18	J3D 5	B02290123	18	J3C 6	B02291156	166	B6
19	J3A 14	B02290472	19	J3B 28	B02297417	108	B7
20	J3A 15	B02290476	20	J3A 55	B02290481	69	B8
21	J3D 6	B02290465	22	J3A 43	B02289604	51	B9
22	J3A 18	B02290467	24	J3C 4	B02291153	162	B10
23	J3A 19	B02290470	25	J3D 7	B02290459	31	B11
24	J3A 20	B02290461	26	J3A 41	B02289618	49	B12
25	J3A 21	B02290477	27	J3C 3	B02291154	161	C1
26	J3A 22	B02290463	28	J3D 38	B02291102	155	C2
27	J3A 23	B02290460	29	J3D 10	B02289596	59	C3
28	J3A 24	B02290475	30	J3D 9	B02290088	55	C4
29	J3A 26	B02290466	33	J3D 36	B02289608	153	C5
30	J3A 27	B02289607	34	J3D 39	B02290521	158	C6
31	J3A 28	B02289621	35	J3B 34	B02297415	117	C7
32	J3A 29	B02289619	36	J3B 58	B02289605	152	C8
33	J3A 30	B02289620	37	J3A 33	B02289614	40	C9
34	J3A 31	B02289600	38	J3A 50	B02297766	61	C10
35	J3A 35	B02289601	42	J3A 44	B02289615	52	C11
36	J3A 37	B02289609	44	J3A 45	B02289606	53	C12
37	J3A 38	B02289602	45	J3D 41	B02291150	163	D1
38	J3A 39	B02289611	47	J3D 11	B02290524	60	D2
39	J3A 46	B02290090	54	J3D 13	B02290483	68	D3
40	J3A 47	B02290408	56	J3A 48	B02290087	57	D4
41	J3D 14	B02294483	63	J3C 17	B02291226	185	D5
42	J3A 53	B02294486	66	J3D 37	B02290089	154	D6
43	J3A 54	B02290482	67	J3B 4	B02291247	79	D7
44	J3A 56	B02289809	71	J3C 2	B02290407	159	D8
45	J3A 57	B02291251	72	J3B 42	B02297419	128	D9
46	J3A 58	B02291255	73	J3D 18	B02291218	89	D10
47	J3A 59	B02291258	74	J3D 51	B02291237	192	D11
48	J3B 1	B02291256	75	J3C 20	B02291222	188	D12
49	J3B 2	B02291257	76	J3B 40	B02297400	125	E1
50	J3B 3	B02291245	77	J3D 33	B02291223	144	E2
51	J3D 16	B02291246	78	J3B 25	B02291267	105	E3
52	J3B 5	B02291244	80	J3B 39	B02297422	124	E4
53	J3B 6	B02291248	81	J3B 17	B02291260	96	E5
54	J3B 7	B02291249	82	J3B 22	B02291264	101	E6
55	J3B 8	B02291228	83	J3B 33	B02297401	116	E7

## 2.4.1 STD-NMR Optimization

Mixture No.	Reference NMR	Barcode	Stock No.	Reference NMR	Barcode	Stock No.	Well
56	J3B 9	B02291232	84	J3D 29	B02297396	133	E8
57	J3B 10	B02291235	85	J3B 48	B02297398	137	E9
58	J3B 11	B02291239	86	J3B 19	B02291262	98	E10
59	J3B 12	B02291240	87	J3D 24	B02297428	114	E11
60	J3D 17	B02291227	88	J3B 26	B02291224	106	E12
61	J3B 13	B02291219	90	J3B 37	B02297414	122	F1
62	J3D 19	B02291213	91	J3B 50	B02291241	141	F2
63	J3B 14	B02291212	92	J3C 1	B02291100	157	F3
64	J3D 20	B02291216	93	J3C 25	B02291230	194	F4
65	J3B 15	B02291254	94	J3B 23	B02291269	103	F5
66	J3B 16	B02291259	95	J3D 27	B02297404	127	F6
67	J3B 18	B02291263	97	J3B 24	B02291271	104	F7
68	J3B 20	B02291261	99	J3D 30	B02297394	136	F8
69	J3B 21	B02291265	100	J3B 54	B02290462	148	F9
70	J3D 21	B02291266	102	J3B 31	B02291253	113	F10
71	J3B 27	B02297412	107	J3C 16	B02297408	181	F11
72	J3D 22	B02297429	109	J3B 35	B02297427	119	F12
73	J3D 23	B02297423	110	J3B 38	B02297425	123	G1
74	J3B 29	B02297411	111	J3B 32	B02297399	115	G2
75	J3B 30	B02297426	112	J3D 31	B02297402	138	G3
76	J3D 25	B02297416	118	J3B 51	B02291238	142	G4
77	J3D 26	B02297420	120	J3B 49	B02291229	140	G5
78	J3B 36	B02297418	121	J3D 47	B02297407	179	G6
79	J3B 41	B02297410	126	J3C 21	B02291221	189	G7
80	J3B 43	B02297421	129	J3B 44	B02297413	130	G8
81	J3D 28	B02297405	132	J3B 47	B02297395	135	G9
82	J3B 46	B02297403	134	J3D 53	B02291225	197	G10
83	J3B 52	B02291236	143	J3D 35	B02291270	146	G11
84	J3B 53	B02289599	147	J3D 12	B02294481	62	G12
85	J3B 55	B02290474	149	J3C 9	B02290404	170	H1
86	J3B 56	B02290471	150	J3C 12	B02290659	173	H2
87	J3B 57	B02290468	151	J3D 34	B02291234	145	H3
88	J3D 42	B02291155	164	J3C 11	B02290401	172	H4
89	J3C 8	B02290403	169	J3C 13	B02289594	176	H5
90	J3C 10	B02290402	171	J3D 45	B02290112	175	H6
91	J3C 14	B02297397	177	J3D 49	B02291250	183	H7
92	J3D 46	B02297406	178	J3C 19	B02291217	187	H8
93	J3C 15	B02297409	180	J3D 48	B02291268	182	H9
94	J3C 18	B02291242	186	J3D 50	B02291252	184	H10
95	J3C 24	B02291233	193	J3D 52	B02291214	196	H11
96	J3D 32	B02291243	139		DMSO		

### 2.4.2 Fragment Screening

After the initial optimization phase, the JNK3 and the seven remaining proteins were screened consecutively. Within the library, 111 fragments were found to interact somehow with one of the eight proteins, and 80 fragments showed no hit event in any case. A hit event was defined as a signal intensity greater than three times the local noise in the difference spectrum. This criterion was loosely based on the %STD concept of Begley et al.<sup>32</sup>. Figure 28 A shows the general screening procedure, the mixtures were measured in a <sup>1</sup>H-NMR (red), and an STD spectrum was generated (blue). If a signal was present, it was compared to the reference of the mixture and the pure fragment (green). The signal intensities of STD-spectra vary by the efficiency of the saturation transfer, which in turn depends on a multitude of variables. Amongst others, it is affected by the binding mode, the proximity of a fragment's hydrogens to the protein, the association and dissociation constants, the relaxation times of the protein, and the fragment's influence on the saturation transfer. To identify the influence of each variable, vast resources would have been necessary, which exceed the use of the STD method for fragment screening on multiple target proteins. The example in Figure 28 A. illustrates the possibility of multiple fragments binding to the same protein in one sample. The two duplets of the *p*-chlorophenol (4485) fragment are barely recognizable.

The singlet of the 4-chloro-5*H*-pyrrolo[3,2-*d*]pyrimidine (9595), on the other hand, exerts an intense peak, and the other two protons show peaks close to the background noise. The *p*-chlorophenol (4485) was not counted as a hit, whereas fragment 9595 was further investigated. The possible competitive nature of these binding events can either benefit the screening due to the filtering of weaker binding fragments or hinder the process by diminishing signals below the detection threshold. In most cases, the benefit of time savings and intrinsic filtering is seen as an advantage of the STD technique compared to other screening techniques. Figure 28 B. demonstrates the typical differences in signal intensity. The shown fragment, 3-iodo-2,4-dimethylpyrazole (0459) with AAK1 (blue) and CAMK1G (red), was highly saturated, and a strong signal was received.

## 2.4.2 Fragment Screening

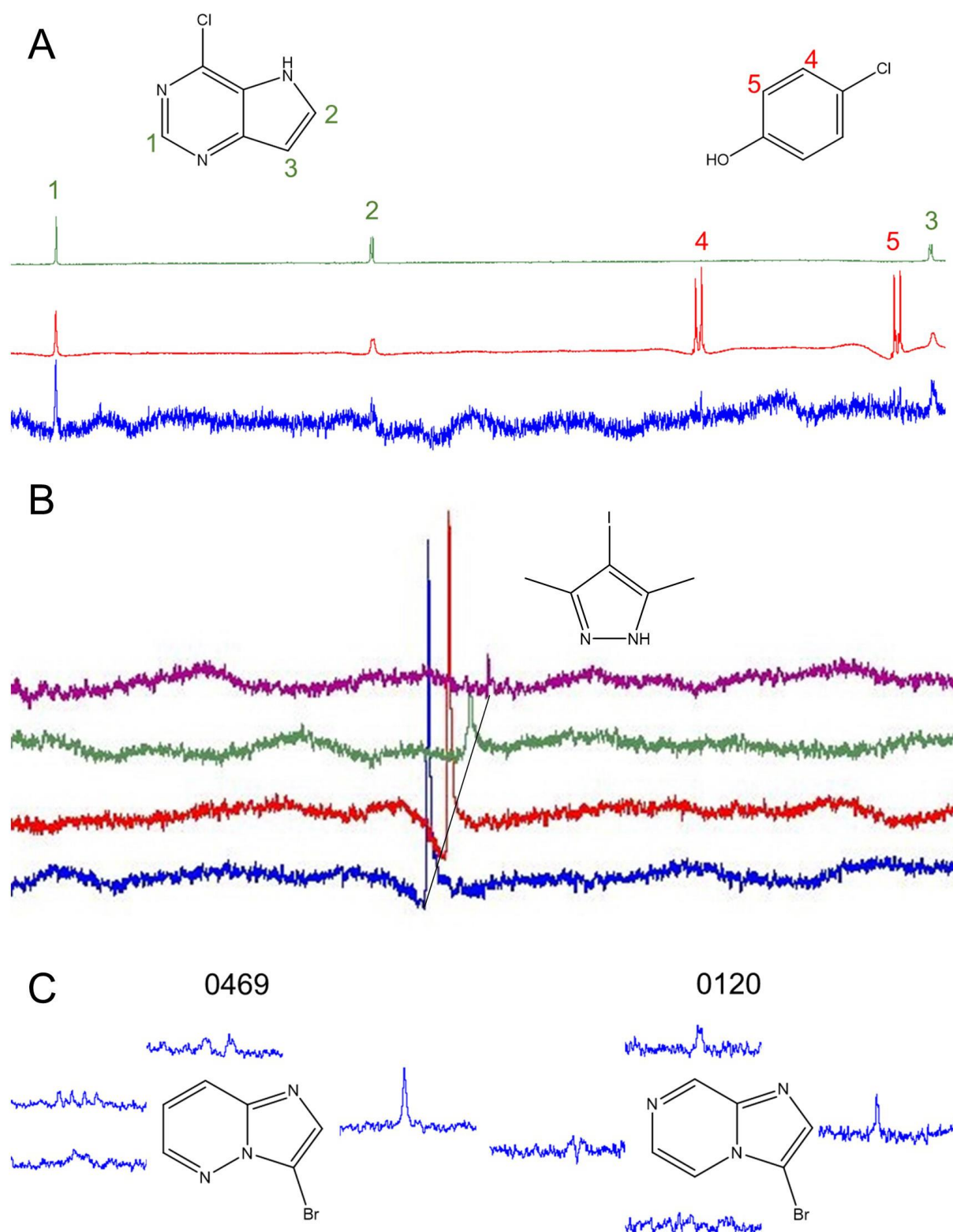


Figure 28. **A:** Comparison of the reference  $^1\text{H-NMR}$  spectrum from both fragments (red 9595+4485) with the spectrum of 9595 (green) and the STD spectrum with AAK1 (blue). While two of three protons of 9595 are visible in the difference spectrum, the four protons of 4485 are not significantly exceeding the background noise. **B:** Comparison of difference spectra of the same fragment mixture with AAK1 (blue), CAMK1G (red), IDO1 (green), and DOT1L (purple). All spectra show the exact peak with a 0.02 ppm offset for visibility. **C:** Two examples of hit fragments with the CAMK1G, NMR peaks were clipped and arranged to the corresponding proton. The fragments share a structural similarity as well as in the spectrum.

## 2.4.2 Fragment Screening

In contrast, the signals in the screening of IDO1 and even more DOT1L were considerably less intensive (Figure 28 B). The signal of DOT1L with 0459 was close to the detection limit. These differences are caused by the already mentioned possible fluctuations intrinsic to the change of the target. In panel C of Figure 28, two very similar fragments can be seen. Both have a good singlet signal next to the bromine, which will become important in fragment prioritization (see section 2.4.3). The signals of the remaining hydrogens would not have been sufficient to be counted as hits, although the signals can be differentiated from the background noise.

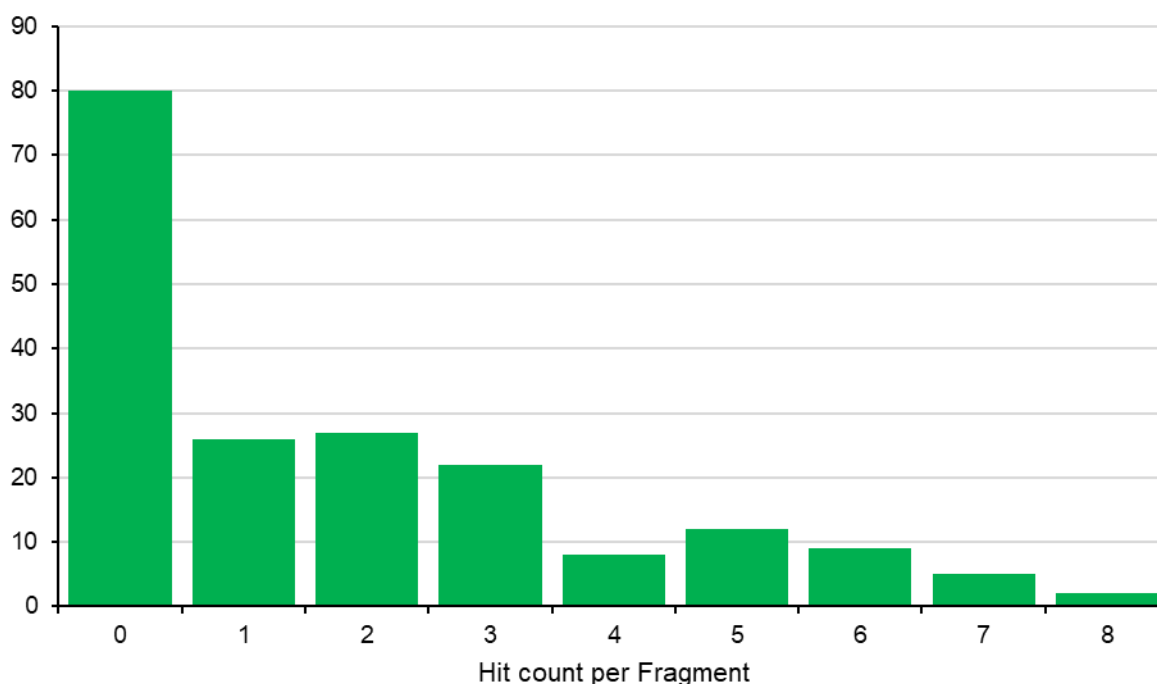


Figure 29. Bar plot of the fragments to interact with target proteins. 80 different fragments were found to interact with none of the tested proteins. Most of the fragments interacted with 1-3 proteins, a lesser part interacted with 4-6 proteins, and only a few fragments interacted with seven to eight proteins. Although five of the tested proteins were kinases and thus shared some similarities, it is remarkable to see the many fragments interacting somewhat uniquely.

Most fragments interact with only one to three different proteins, two fragments show a hit event with every protein, and five further fragments bind to every but one protein (Figure 29). A pleasant situation was found, with great freedom to choose from over 60 fragments, bearing a more or less unique hit pattern with a maximum of three proteins. An elaborate asset of criteria had to be found to rationalize and streamline the validation process to prioritize specific fragments further. A few of the universal binding fragments were of greater interest as well, to have the possibility to see if the protocols worked independently from the protein. One of the universal hits was

## 2.4.2 Fragment Screening

fragment 1234, which had already proven to react covalently with GSH. Therefore the STD data were carefully reviewed. It was impossible to verify or falsify any assay-disrupting behavior, and therefore it was subjected to the hit validation process (see section 2.5.8). In addition, some fragments binding to only one or all tested kinases were presumed worth further investigation. All of the screening results are summarized in Table 3. Every fragment, with or without hit events, is depicted by color code. The sum for every fragment is added on the right.

Table 3. Overview of all STD experiments with every screened protein. A one in a field indicates a hit, whereas a zero indicates no event.

Barcode	JNK3	JNK2	DYRK1a	CAMK1G	AAK1	DOT1L	BIRC5	IDO1	Sum
B02289594	0	0	0	0	0	0	0	0	0
B02289595	1	1	1	1	1	0	0	0	5
B02289596	0	0	1	1	0	0	0	0	2
B02289597	0	0	0	0	0	0	0	0	0
B02289599	0	0	0	0	0	0	0	0	0
B02289600	0	1	1	1	0	0	0	0	3
B02289601	1	1	1	0	1	0	0	1	5
B02289602	0	0	0	0	0	0	0	0	0
B02289603	0	0	0	0	0	0	0	0	0
B02289604	1	0	1	0	1	0	0	0	3
B02289605	1	1	0	1	0	0	0	0	3
B02289606	0	0	0	0	0	0	0	0	0
B02289607	1	1	1	0	1	0	0	0	4
B02289608	0	0	1	1	1	0	1	0	4
B02289609	0	0	0	0	0	0	0	0	0
B02289610	1	1	1	1	1	0	0	0	5
B02289611	0	0	0	0	0	0	0	0	0
B02289612	1	1	0	0	0	0	0	1	3
B02289613	1	1	0	0	0	0	0	0	2
B02289614	0	0	0	0	0	0	0	0	0
B02289615	1	0	1	0	0	0	0	0	2
B02289616	0	0	1	0	1	0	0	0	2
B02289617	1	1	0	1	0	0	0	0	3
B02289618	0	0	0	0	0	0	0	0	0
B02289619	1	0	0	1	0	0	0	0	2
B02289620	0	0	1	0	0	0	0	1	2
B02289621	1	1	1	1	1	0	0	0	5
B02289809	1	0	0	0	0	0	0	0	1
B02290087	0	1	0	1	0	0	0	0	2
B02290088	0	0	0	0	0	0	0	0	0
B02290089	0	0	1	0	0	0	0	0	1
B02290090	1	0	1	0	0	0	0	0	2
B02290112	0	0	0	0	0	0	0	0	0

## 2.4.2 Fragment Screening

Barcode	JNK3	JNK2	DYRK1a	CAMK1G	AAK1	DOT1L	BIRC5	IDO1	Sum
B02290113	0	1	1	0	1	0	0	1	4
B02290114	0	0	0	0	0	0	1	0	1
B02290115	0	1	1	1	0	0	0	0	3
B02290116	0	0	1	0	0	0	0	1	2
B02290117	0	0	0	0	0	0	0	0	0
B02290118	0	0	0	0	0	0	0	0	0
B02290119	0	0	0	0	0	0	0	0	0
B02290120	0	0	1	0	0	1	1	0	3
B02290121	0	0	0	0	0	0	0	0	0
B02290122	0	0	0	0	0	0	0	0	0
B02290123	0	0	1	0	0	0	0	1	2
B02290401	0	1	0	0	0	0	0	0	1
B02290402	0	0	0	0	0	0	0	0	0
B02290403	0	1	0	0	1	0	0	0	2
B02290404	0	0	0	0	0	0	0	0	0
B02290407	1	1	0	0	1	0	0	0	3
B02290408	0	0	0	0	0	0	0	0	0
B02290459	1	1	1	1	1	1	1	1	8
B02290460	1	1	0	1	0	0	0	0	3
B02290461	0	0	0	0	0	0	0	0	0
B02290462	0	0	0	0	0	0	0	0	0
B02290463	1	0	0	0	0	0	0	0	1
B02290464	1	1	1	1	1	0	0	1	6
B02290465	0	0	0	0	1	0	0	0	1
B02290466	0	0	0	0	0	0	0	0	0
B02290467	0	0	0	0	0	0	0	0	0
B02290468	0	0	1	0	0	0	0	0	1
B02290469	1	1	1	1	0	0	0	1	5
B02290470	0	0	0	0	0	0	0	0	0
B02290471	0	0	0	0	1	0	0	0	1
B02290472	0	0	0	0	0	0	0	0	0
B02290473	1	0	1	1	1	0	0	0	4
B02290474	0	1	0	0	1	0	0	0	2
B02290475	1	1	1	1	1	0	0	0	5
B02290476	0	1	0	0	0	0	0	0	1
B02290477	0	0	0	0	0	0	0	0	0
B02290481	1	0	0	0	0	0	0	0	1
B02290482	0	1	1	0	0	0	0	0	2
B02290483	0	0	0	0	0	0	0	0	0
B02290521	1	1	1	1	1	0	1	1	7
B02290522	0	0	1	0	1	0	0	1	3
B02290523	0	0	0	0	0	0	0	0	0
B02290524	0	0	1	0	0	0	0	0	1
B02290525	0	0	0	0	0	0	0	0	0
B02290658	0	0	0	0	0	0	0	0	0

## 2.4.2 Fragment Screening

Barcode	JNK3	JNK2	DYRK1a	CAMK1G	AAK1	DOT1L	BIRC5	IDO1	Sum
B02290659	0	0	0	0	0	0	0	0	0
B02290660	1	1	1	0	1	1	1	0	6
B02290661	1	1	1	1	1	0	0	1	6
B02291100	0	0	0	0	0	0	0	0	0
B02291102	0	0	0	1	1	0	0	1	3
B02291150	1	1	1	1	1	0	0	0	5
B02291152	0	0	1	0	0	0	1	0	2
B02291153	0	0	0	0	0	0	0	0	0
B02291154	0	0	0	0	0	0	0	0	0
B02291155	0	0	0	0	0	0	0	0	0
B02291156	0	0	0	0	0	0	0	0	0
B02291212	0	0	0	0	0	0	0	0	0
B02291213	0	0	1	0	0	0	1	0	2
B02291214	0	0	0	0	0	0	0	0	0
B02291216	1	1	1	1	1	0	1	1	7
B02291217	0	1	0	0	1	0	0	0	2
B02291218	0	0	0	0	0	0	0	0	0
B02291219	1	1	1	1	1	0	0	0	5
B02291221	0	0	0	0	0	0	0	0	0
B02291222	0	0	0	0	0	0	0	0	0
B02291223	0	0	1	0	0	0	0	0	1
B02291224	0	0	0	0	0	0	0	0	0
B02291225	0	0	1	0	0	0	0	1	2
B02291226	0	0	0	0	0	0	0	0	0
B02291227	0	0	1	0	1	0	0	0	2
B02291228	0	0	0	0	0	0	0	0	0
B02291229	0	0	0	0	0	0	0	0	0
B02291230	1	1	1	1	1	0	0	0	5
B02291232	0	1	1	0	1	0	0	0	3
B02291233	0	0	1	0	1	0	0	0	2
B02291234	1	1	1	1	1	1	1	1	8
B02291235	1	1	0	1	0	0	0	0	3
B02291236	0	1	0	0	0	0	0	0	1
B02291237	0	1	0	0	0	0	0	0	1
B02291238	0	0	1	0	0	0	0	0	1
B02291239	0	0	0	0	0	0	0	0	0
B02291240	0	0	0	0	0	0	0	0	0
B02291241	0	0	0	0	0	0	0	0	0
B02291242	0	0	0	0	0	0	1	0	1
B02291243	0	0	0	1	0	0	0	1	2
B02291244	0	1	1	0	0	0	0	0	2
B02291245	0	0	0	0	0	0	0	0	0
B02291246	1	1	1	0	1	0	1	1	6
B02291247	0	0	0	0	0	0	0	0	0
B02291248	0	1	1	1	1	0	0	0	4



## 2.4.2 Fragment Screening

Barcode	JNK3	JNK2	DYRK1a	CAMK1G	AAK1	DOT1L	BIRC5	IDO1	Sum
B02291249	0	0	0	0	0	0	0	0	0
B02291250	0	0	0	1	0	0	0	0	1
B02291251	0	0	0	0	0	0	0	0	0
B02291252	1	0	0	0	1	0	1	0	3
B02291253	0	0	1	1	0	0	0	0	2
B02291254	1	1	1	1	1	0	0	1	6
B02291255	0	0	0	0	0	0	0	0	0
B02291256	1	1	1	1	1	0	0	0	5
B02291257	0	1	0	0	1	0	0	0	2
B02291258	0	0	0	0	0	0	0	0	0
B02291259	0	0	0	0	0	0	0	0	0
B02291260	0	0	0	0	0	0	0	0	0
B02291261	0	0	1	0	1	1	0	0	3
B02291262	0	0	0	0	0	0	0	0	0
B02291263	0	0	0	0	0	0	0	0	0
B02291264	0	0	0	0	0	0	0	0	0
B02291265	0	1	0	1	1	0	0	0	3
B02291266	1	1	1	1	1	0	1	1	7
B02291267	1	1	1	0	0	0	0	0	3
B02291268	0	0	0	0	0	0	0	1	1
B02291269	0	0	0	0	0	0	0	0	0
B02291270	0	1	1	1	1	1	0	1	6
B02291271	0	0	1	0	0	0	0	0	1
B02291272	0	0	0	0	0	0	0	0	0
B02294481	0	0	1	0	0	0	0	0	1
B02294482	0	0	0	0	0	0	0	0	0
B02294483	0	1	1	0	1	0	0	0	3
B02294485	1	1	1	0	0	0	0	0	3
B02294486	1	1	1	1	1	0	0	1	6
B02297394	1	1	1	1	1	0	1	1	7
B02297395	0	0	0	0	0	0	0	0	0
B02297396	0	0	1	0	0	0	0	0	1
B02297397	0	0	0	0	0	0	0	0	0
B02297398	0	0	0	0	0	0	0	0	0
B02297399	0	0	0	0	0	0	0	0	0
B02297400	0	1	0	0	0	0	0	0	1
B02297401	1	1	1	0	1	0	0	0	4
B02297402	0	0	1	0	1	0	0	0	2
B02297403	0	0	0	0	0	0	0	0	0
B02297404	0	0	0	0	0	0	0	0	0
B02297405	0	1	1	1	1	1	1	0	6
B02297406	0	1	1	0	1	0	0	0	3
B02297407	0	1	1	0	1	1	0	0	4
B02297408	0	0	0	0	0	0	0	0	0
B02297409	1	0	0	0	1	0	0	0	2

## 2.4.2 Fragment Screening

Barcode	JNK3	JNK2	DYRK1a	CAMK1G	AAK1	DOT1L	BIRC5	IDO1	Sum
B02297410	0	1	1	0	1	0	0	0	3
B02297411	0	0	0	0	0	0	0	1	1
B02297412	0	0	0	0	0	0	0	0	0
B02297413	1	1	1	0	1	0	0	1	5
B02297414	0	0	0	0	0	0	0	0	0
B02297415	1	0	1	0	0	0	0	0	2
B02297416	0	0	0	0	0	0	0	0	0
B02297417	0	0	0	0	0	0	0	0	0
B02297418	0	0	0	0	0	0	0	0	0
B02297419	0	0	1	0	0	0	0	0	1
B02297420	1	1	1	0	1	1	1	1	7
B02297421	0	0	0	0	0	0	0	0	0
B02297422	1	0	1	1	1	0	0	0	4
B02297423	1	0	1	1	1	0	1	1	6
B02297424	1	0	1	1	1	0	1	0	5
B02297425	0	0	0	0	0	0	0	1	1
B02297426	0	0	0	0	0	0	0	0	0
B02297427	0	0	0	0	0	0	0	1	1
B02297428	0	0	0	0	0	0	0	0	0
B02297429	0	0	1	0	0	0	0	1	2
B02297655	0	0	0	0	0	0	0	0	0
B02297765	0	0	0	0	0	0	0	0	0
B02297766	1	0	1	0	0	0	1	0	3

#### 2.4.3 Library Analysis and Fragment Prioritization

Ranking fragments based on their STD signals are highly susceptible to fluctuations in saturation transfer, especially when comparing different proteins<sup>185</sup>. First, fragments have to bind in a comparable binding mode in the same pocket, generally being established in a competitive screening. In this case, every possible binding site was included to detect potential allosteric binding. Even if only one binding site would be present, as Aretz et al. showed, slight variations in the binding mode of a compound could result in extreme changes in the STD effect<sup>185</sup>. The fragment library is designed to employ diversity, and in conclusion, high variability in the STD spectrum has to be expected. The efficacy in the saturation transfer should, in theory, be different for every fragment. This might be the reason for the high variability in the 0459 hit (Figure 28 B). Aretz et al. could prove that a quantifiable correlation between STD effect and binding affinity is highly susceptible to experiment variables and not a reliable prioritization criterion<sup>185</sup>. Therefore, other criteria had to be utilized to filter the fragments for their potential binding affinity.

A first effort to overcome the missing STD-affinity correlation was to look at various patterns in the hit events of specific fragments (Table 3). A few patterns were of particular interest, especially fragments binding to only one protein (e.g., 0471, 1268, 7411), fragments binding to only one kinase, or every except one kinase (e.g., 0468, 0473, 9607). Some fragments bind to various proteins, either closely related (JNK2 and JNK3) or have nothing in common (AAK1 and IDO1). There are fragments with exciting patterns associated with the screening. For example, fragment 0660 binding to all proteins except CAMK1G and IDO1, or fragment 1248 does not bind to JNK3 but to the four remaining kinases.

Fragment 9612 acts as another example; it only binds to JNK2 and JNK3, which are closely related. On the other hand, it was unremarkable with DYRK1a, another close relative in terms of kinase similarity (Table 3). The structurally similar fragment 0523 could not be observed as a hit with any protein. These differences in binding pattern combined with structural similarities can give first hints and ideas for further validation and evaluation of the fragments.

Another selected fragment for further examination was 9595 (see Appendix B 13). It interacts with all five kinases and could lead to novel and interesting insights. Fragments 7422, 7423, and 7424 show hits with the same proteins (AAK1, CAMK1g, DYRK1a, and JNK3) without structural similarities (Appendix C: HEFLib Properties). These findings could point to multiple binding sites on a protein or a chance to use fragment linking if the same or adjacent pockets are addressed. This cherry-picking approach selected some fragments for validation, but too many remained for streamlined analysis. Therefore, further criteria had to be applied to prioritize certain fragment-protein pairs for ITC validation.

Table 4. The number of fragment hits found with each protein, the diagonal shows the number of uniquely binding fragments. Apart from the diagonal, the number of fragments is shown interacting with at least the proteins of the column and row.

	JNK3	JNK2	DYRK1a	CAMK1G	AAK1	DOT1L	BIRC5	IDO1
JNK3	3	35	37	28	33	4	13	17
JNK2	35	5	41	30	40	7	10	18
DYRK1a	37	41	9	31	46	9	17	24
CAMK1G	28	30	31	1	28	4	10	15
AAK1	33	40	46	28	2	8	14	19
DOT1L	4	7	9	4	8	0	6	4
BIRC5	13	10	17	10	14	6	2	9
IDO1	17	18	24	15	19	4	9	4
Hit-rate [%]	26.2	30.9	39.8	21.5	29.8	4.7	10.5	16.2

As shown in Table 4, uniquely binding fragments were found for every protein except DOT1L (diagonal line). The overlap of binding fragments to the kinases was, as expected, the most significant and could be found in the screening data (Table 3) as well. The overall hit rates of the kinases ranged from 21.5% for the CAMK1G up to 39.8% for the DYRK1a, which would generally be considered too high for a standard fragment screening<sup>6</sup>. The hit rate was one of the reasons for the high number of fragments in need of further investigation. In stark contrast, the hit rates of the IDO1 and BIRC5 are 16.2% and 10.5%. Both rates are considerably lower and are generally much more practical for hit validation.

The hit rate of DOT1L is typically considered too low for a successful screening and validation campaign, with only nine fragments found<sup>6</sup>.

The broad range of hit rates found between proteins further demonstrates the tremendous impact of the protein itself on the screening. It suggests that minor optimization of the screening protocol should be incorporated into the standard procedure for every new protein. This optimization could include modifying the fragment concentration, especially to reduce the hit rate. Increasing the hit rate by increasing the fragment concentration (above 2 mM) is not feasible due to the limited solubility of some fragments. However, a higher sensitivity could be achieved by trading efficiency in measuring time for more scans per experiment or incorporating a water suppression pulse sequence to diminish the receiver coil saturation by the vast water peak. A more complex mixture of fragments could be used to counteract the rise in measuring time. As mentioned before, the increase in hit events might benefit a drug discovery campaign, but generated hits are likely to have millimolar affinity and are more challenging in the validation with the tools on hand in this study.

The enormous hit rate of the kinases might be attributed to the large and particular deep ATP-binding pocket compared to the active sites of DOT1L and IDO1. The BIRC5 has no enzymatic activity and has only small crevices on the surface, without a canonical binding site. The active site of DOT1L is accustomed to SAM, with the methionine residue pointing into a tunnel and the adenosine laying atop a relatively shallow surface recess<sup>134</sup>. Other allosteric pockets have been described and are either a sub-pocket or are situated in the vicinity of the active site<sup>137-139</sup>. Overall, the pockets of the DOT1L are shallower than the ATP pocket of the kinases, and fragments might experience a higher cost of desolvation energy upon binding. The same but to a lesser extent applies to the IDO1 binding pocket. The numerous states and flexibility of the binding pocket with its heme cofactor enable many possible partially bound states of fragments. A manifold of protein states coexisting with each other would decrease the STD effect of only a small portion is possible to bind a fragment. The kinases, in general, have stable, deep, and aromatic compounds accepting active sites. These features could be reasons the STD was so effective in these kinase cases.

As shown in Figures 30, 31, and 32, no clear picture of a binding mode can be determined; fragments seem random and have minor to no similarities. The lack of similarity is a favorable observation, suggesting a diverse set of binding modes. The diversity optimization in the library design could be one reason for the diversity in the uniquely binding fragments, even though there are matched pairs and satellite groups of similar fragments incorporated in the library<sup>7</sup>. Deducing similarities and estimating binding modes is highly speculative and, therefore, should not and was not attempted to prioritize fragments. In the case of the universally binding fragments, though, a large portion of fragments shared the Xanthine or pyrimidinedione scaffold (Figure 33), which could be a sign of unspecific binding events, and some similar compounds have been reported to bear PAIN characteristics<sup>157, 161</sup>.

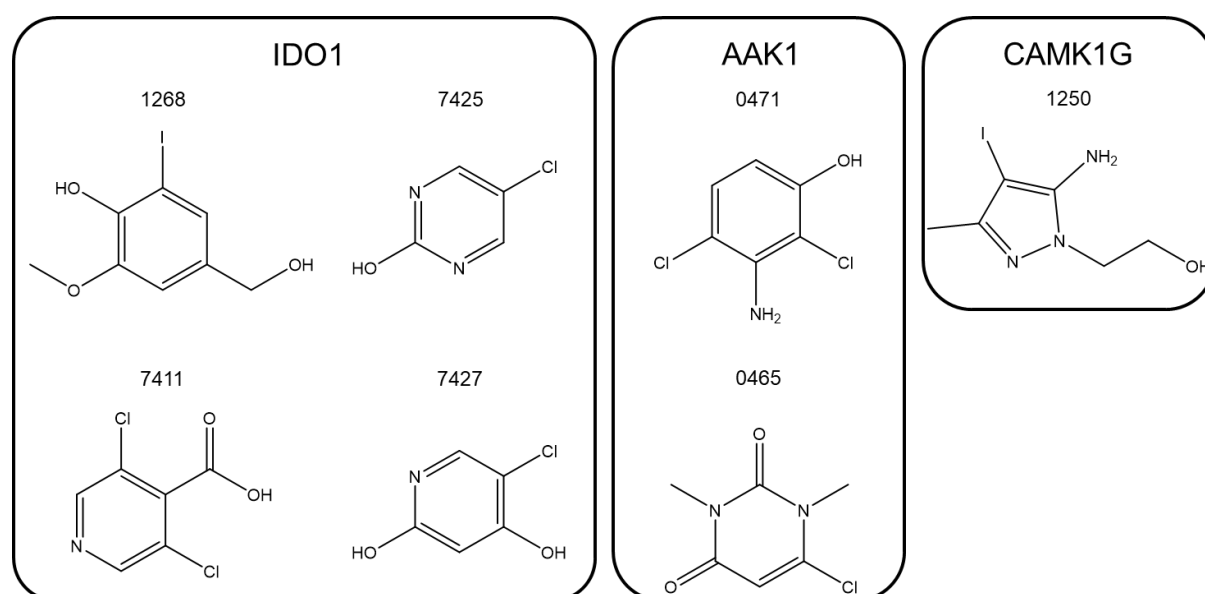


Figure 30. Uniquely interacting hit fragments for the proteins IDO1, AAK1, and CAMK1G.

### 2.4.3 Library Analysis and Fragment Prioritization

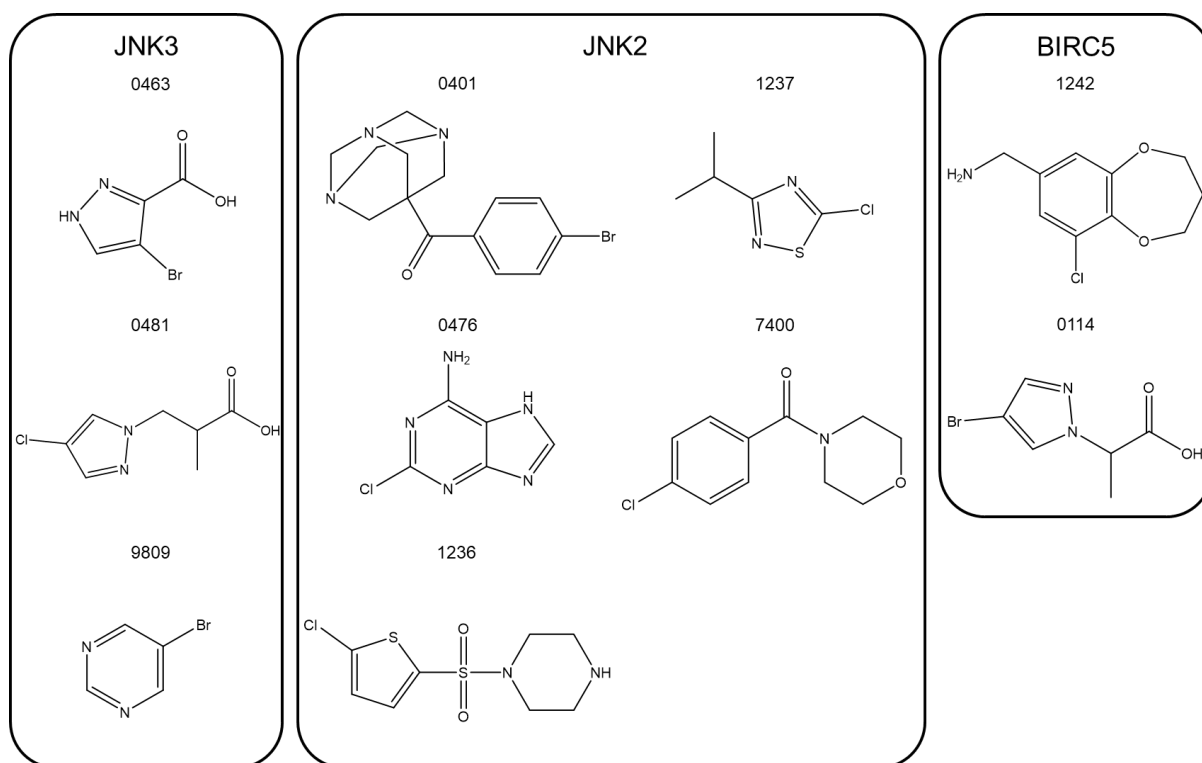


Figure 31. Uniquely interacting hit fragments for the proteins JNK3, JNK2, and BIRC5.

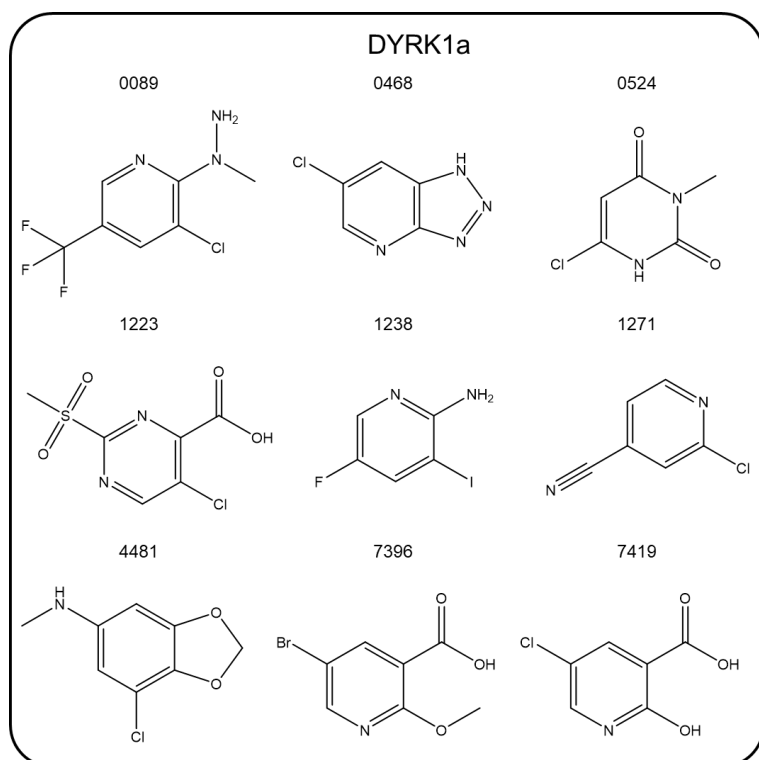


Figure 32. Uniquely interacting hit fragments for DYRK1a.

### 2.4.3 Library Analysis and Fragment Prioritization

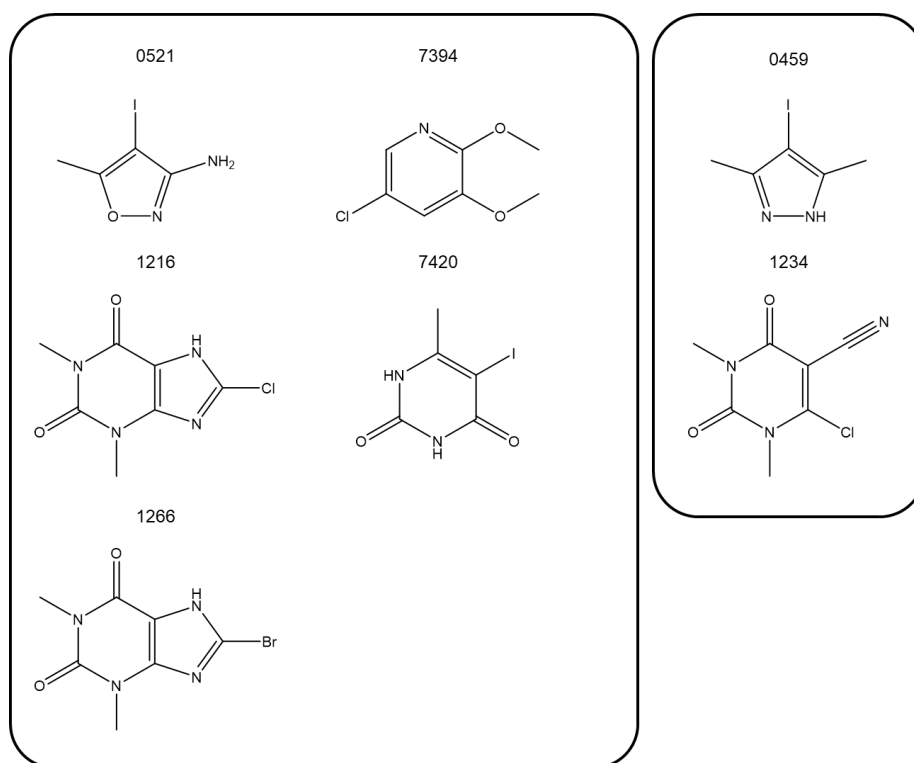


Figure 33. On the left are Fragments binding to seven out of eight proteins; fragments 0459 and 1234 scored a hit with all screened proteins. A clustering of Xanthine or pyrimidinedione-like structures can be seen with this group of fragments.



### 2.4.3 Library Analysis and Fragment Prioritization

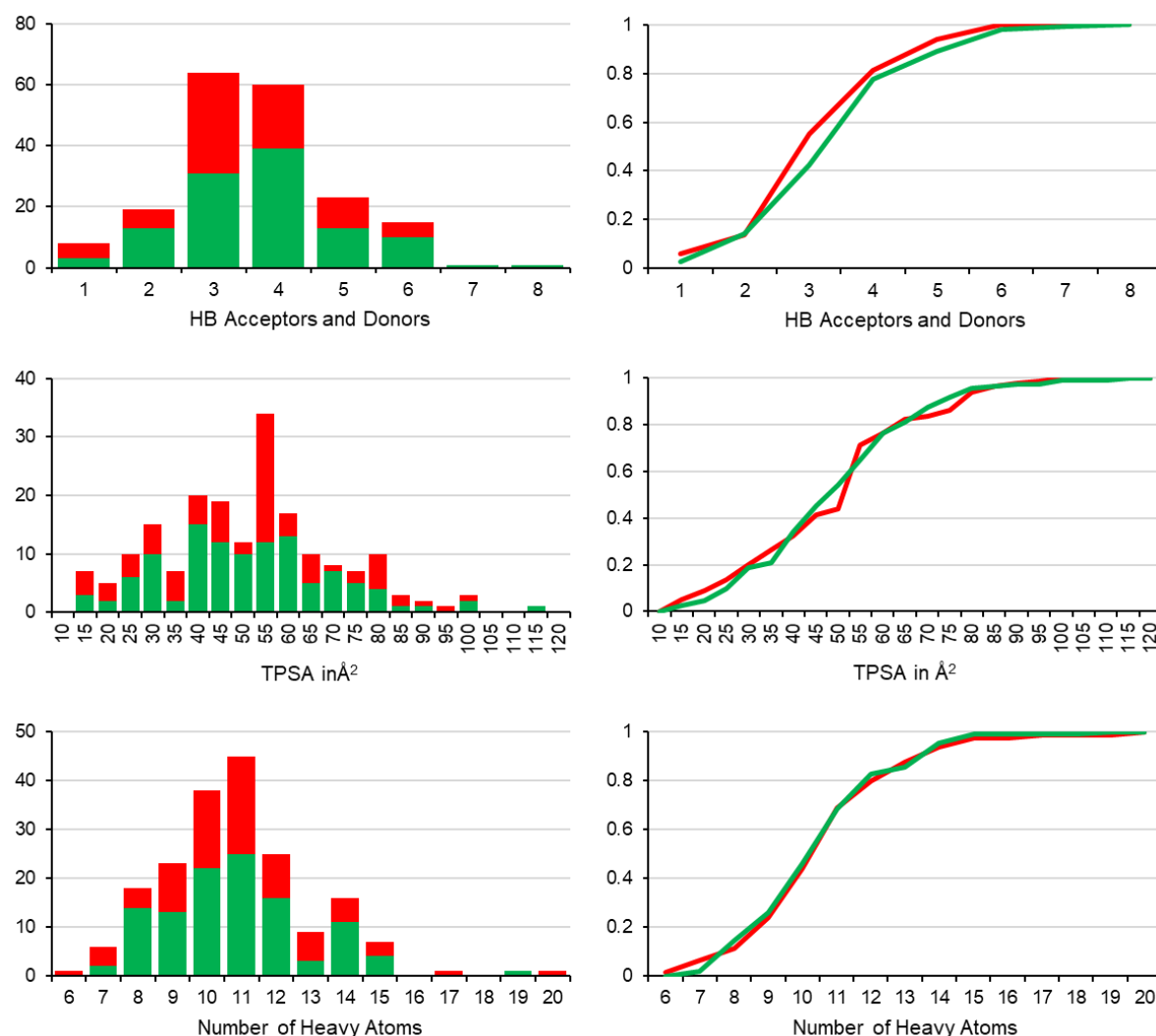


Figure 34. Bar plots on the left show the parameter distribution of the non-binders (red) on top of the hit fragments (green). The cumulative distributions of the same parameters depicted in the bar plots are on the right. No significant differences in distribution between hits and non-hits or the whole library can be determined. Partitions were calculated from all hit fragments or all uneventful fragments, respectively.

As mentioned above, the number of hit events was too large to validate every hit event on a reasonable scale. The cherry-picking approach with the addition of uniquely binding fragments can be reasonable. However, for prioritization, other metrics were introduced with the idea of selecting fragments with suitable drug potential. Standard metrics such as the number of hydrogen bond acceptors, donors, TPSA, and the number of heavy atoms gave no further insight (Figure 34). These metrics had been applied in the design of the library, and it was interesting to see if any significant shift between hits and uneventful fragments could be observed. No significant difference between inconspicuous fragments and hits could be observed. A specific aberration was found with the calculated logP values of the fragments (Figure 35).

Hit fragments tended to have lower solubility than non-binders. This observation is often made in fragment screenings<sup>6, 8, 92</sup>. Due to the lesser solubility, the desolvation cost of fragments is lower, and the energetic threshold of a binding event can be overcome easier<sup>92</sup>. Perhaps this signifies that specific fragments in the library are too soluble to be engaged in interactions. This study is not comprehensive enough to estimate the thermodynamic consequences of binding, but this trend should not be overlooked in further studies. Possible reevaluation of hit fragments and their respective solubility after additional screening could lead to the exclusion of fragments with low chances of hit generation.

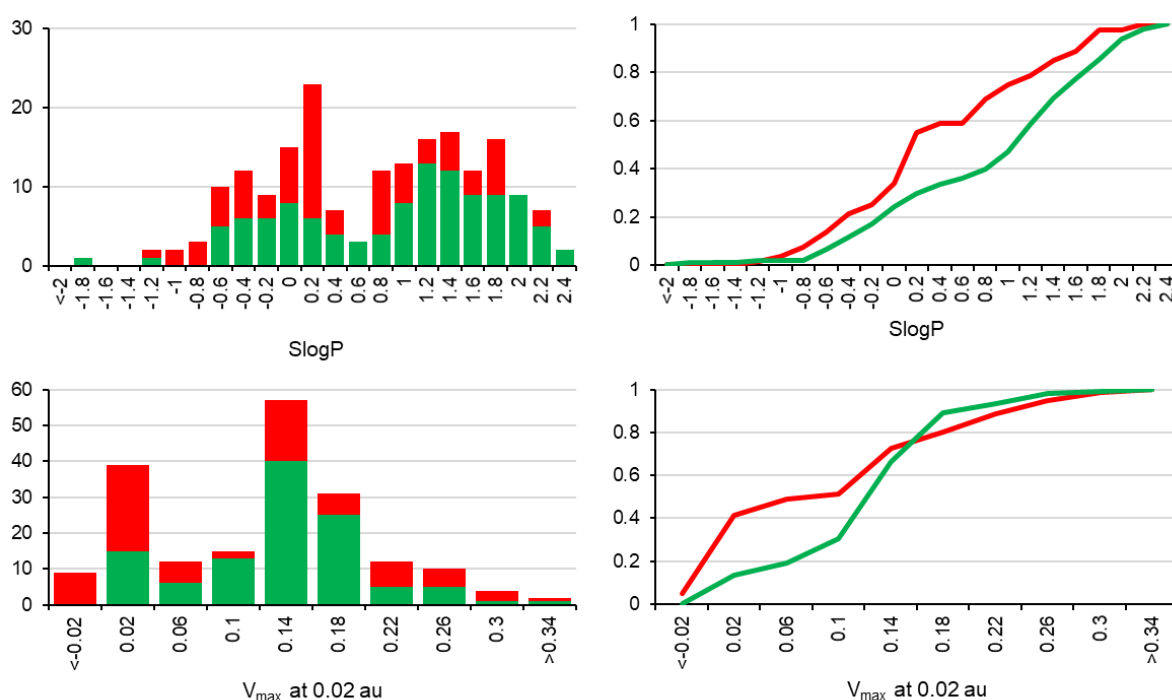


Figure 35. Bar plot on the left and cumulative distribution of calculated SlogP values and  $V_{\max}$ -values of hits (green) and non-binders (red). Small SlogP values indicate high solubility in water. Bar plots depict the absolute numbers, whereas the cumulative distribution shows portions of all hits or non-binders. Both plots showcase the substantial right shift of hits compared to non-binders, associating hit fragments with lower solubility than non-binders. The same correlation can be seen in the  $V_{\max}$  plot. Many of the hit fragments have a  $V_{\max} > 0.12$ . The first indicator for halogen bonding is a valuable part of various binding modes.

The second value of interest was the  $V_{\max}$ ; with a shift towards higher values, the  $\sigma$ -hole is increased and could lead to stronger halogen bonds in theory<sup>86</sup>. Figure 35 shows that more than 60% of the hit fragments have a  $V_{\max} > 0.12$ . In contrast, more than 50% of the non-binders have a  $V_{\max} < 0.12$ . In other words, a high  $V_{\max}$ -value is found in most hit events, whereas a low  $V_{\max}$ -value is found with the most inconspicuous fragments.

A first sign of the halogen bond could be a valuable part in many binding modes yet to be revealed. As the library was designed with the electrostatic potential of halogens in mind, the fragments were initially selected, among other things, for maximum diversity in the halogen bonding interface<sup>7</sup>. As a result, fragments with highly tuned halogens were incorporated as well as fragments with anti-tuned halogens, mainly by the introduction of negatively charged substituents (e.g., carboxyl group). In the clustering of  $V_{\max}$ -values, we can see that most hit fragments bear halogens with a  $V_{\max} > 0.12$  at 0.02 au (Figure 35). This is a reassuring sign of our original intention to create a fragment library with halogen bonding as a significant contribution to binding motifs.

The shift observed in Figure 35 led to the statistical analysis of the fragment library regarding the halogens present in hit fragments. An increase in hit-events from chlorine through bromine to iodine can be observed by clustering the hits by their respective halogen. The HEFLib contains 114 fragments with chlorine, and 61 of these show an event with at least one protein (53.5%), iodine-containing fragments are only found 14 times in the library, but 11 of these display at least one hit event (78.6%), the bromine fraction lies in between these two halogens. If all hit events (a fragment can have a hit with more than one protein) are added, all chlorine-containing fragments showed 174 hits, on average 1.5 hits events per fragment. The average hits per-fragment increase to two for bromine, and iodine is capping at three hit events per fragment, illustrating the impact of halogen type on the HEFLib. The number of hits per protein and halogen is normalized against the sum per protein and the halogen type (Table 5 bold numbers), illustrating the trend toward larger halogens. Seven out of the eight proteins show a tendency to bind more iodine-containing fragments than bromine- or chlorine-containing (except for JNK2). An excellent example is the CAMK1g, which features a normalized hit rate of 0.42 for chlorine-bearing fragments, which increases to 1.02 for iodine-bearing fragments (Table 5). The trend is even more pronounced within the DOT1I And BIRC5 screening, but the low overall number of hits can easily skew the normalization. The initial acquisition of the library was hindered, among other factors, by the extraordinary price of iodine compounds, thus explaining the low number of iodine-containing fragments<sup>102</sup>.

Table 5. The hits are dissected by the halogen of the fragment. Bold proportions are normalized to the number of hits per protein and the number of fragments with the halogen type.

	Cl		Br		I		Sum
Halogen Type	114	59.7%	63	33.0%	14	7.3%	191
Hit Fragments	61	55.0%	39	35.1%	11	9.9%	111
	<b>0.48</b>		<b>0.56</b>		<b>0.70</b>		
JNK3	22	44.0%	22	44.0%	6	12.0%	50
	<b>0.39</b>		<b>0.70</b>		<b>0.86</b>		
JNK2	30	50.8%	24	40.7%	5	8.5%	59
	<b>0.45</b>		<b>0.65</b>		<b>0.61</b>		
DYRK1a	41	53.9%	26	34.2%	9	11.8%	76
	<b>0.47</b>		<b>0.54</b>		<b>0.85</b>		
CAMK1g	20	47.6%	16	38.1%	6	14.3%	42
	<b>0.42</b>		<b>0.60</b>		<b>1.02</b>		
AAK1	30	52.6%	21	36.8%	6	10.5%	57
	<b>0.46</b>		<b>0.58</b>		<b>0.75</b>		
DOT1L	4	44.4%	3	33.3%	2	22.2%	9
	<b>0.39</b>		<b>0.53</b>		<b>1.59</b>		
BIRC5	9	45.0%	6	30.0%	5	25.0%	20
	<b>0.39</b>		<b>0.48</b>		<b>1.79</b>		
IDO1	18	58.1%	9	29.0%	4	12.9%	31
	<b>0.51</b>		<b>0.46</b>		<b>0.92</b>		
Sum	174	50.6%	127	36.9%	43	12.5%	344

The number of hits of the three halogens does not show the whole picture as the scaffold or EWGs can tune the  $V_{\max}$ -value. To showcase the increasing  $V_{\max}$  from uneventful hits to hit fragments independent of halogen type, Table 6 compares the average  $V_{\max}$  of non-binders with hits and the complete library. The average increases from chlorine to iodine in all three categories from left to right. The same can be observed if the average  $V_{\max}$ -values are compared from the non-binders to the hit-fragments. A two-sided t-test with unequal variances (Welsh's t-test) was calculated to assess whether the effect was pure coincidence. A p-value of 0.01 for being similar was calculated for  $V_{\max}$  averages of the non-binders ( $V_{\max}=0.104$ ) and hits ( $V_{\max}=0.137$ ), without regarding the halogen type. As shown in Table 6, the magnitude of the  $V_{\max}$ -values differs highly for every halogen. Thus, the significance is flawed in the overall comparison. The same statistical effect could be calculated for the chlorine-containing fragments; however, the significance was lower with a p-value was 0.026.

As mentioned above, the HEFLib has a small amount of bromine and even scarcer iodine-bearing fragments. Thus, no significance level could be calculated as the sample was too small for the effect to be observed with enough precision. The higher  $V_{\max}$ -values of hits are partially explained by the higher overall values of iodine compounds combined with the higher hit rate of these compounds. In stark contrast, iodine-bearing fragments account for only 10% of all hits; chlorine-bearing fragments, on the other hand, for over five times as many hits (Table 5), whereas the average of iodine fragment's  $V_{\max}$  is only two times greater than that of chlorine-bearing fragments. This should diminish any potential skewing of the distribution towards larger halogens. Encouraged by these results, we attempt to improve the library by expanding our search for optimized halogen bonding interfaces in commercially available fragments. In addition, we undertake efforts to incorporate iodine into existing fragments and synthetically increase the  $\sigma$ -hole of existing compounds by variation of other substituents.

Table 6. Descriptive statistics of the complete library's  $V_{\max}$ -values were divided between hit fragments and not interacting fragments.

$V_{\max}$		Cl	Br	I
Complete Library	$\bar{x}$	0.107	0.136	0.202
	n	114	63	14
	$\sigma$	0.079	0.078	0.059
No Hits	$\bar{x}$	0.089	0.13	0.184
	n	53	24	3
	$\sigma$	0.094	0.099	0.108
Hits	$\bar{x}$	0.123	0.14	0.207
	n	61	39	11
	$\sigma$	0.059	0.064	0.045

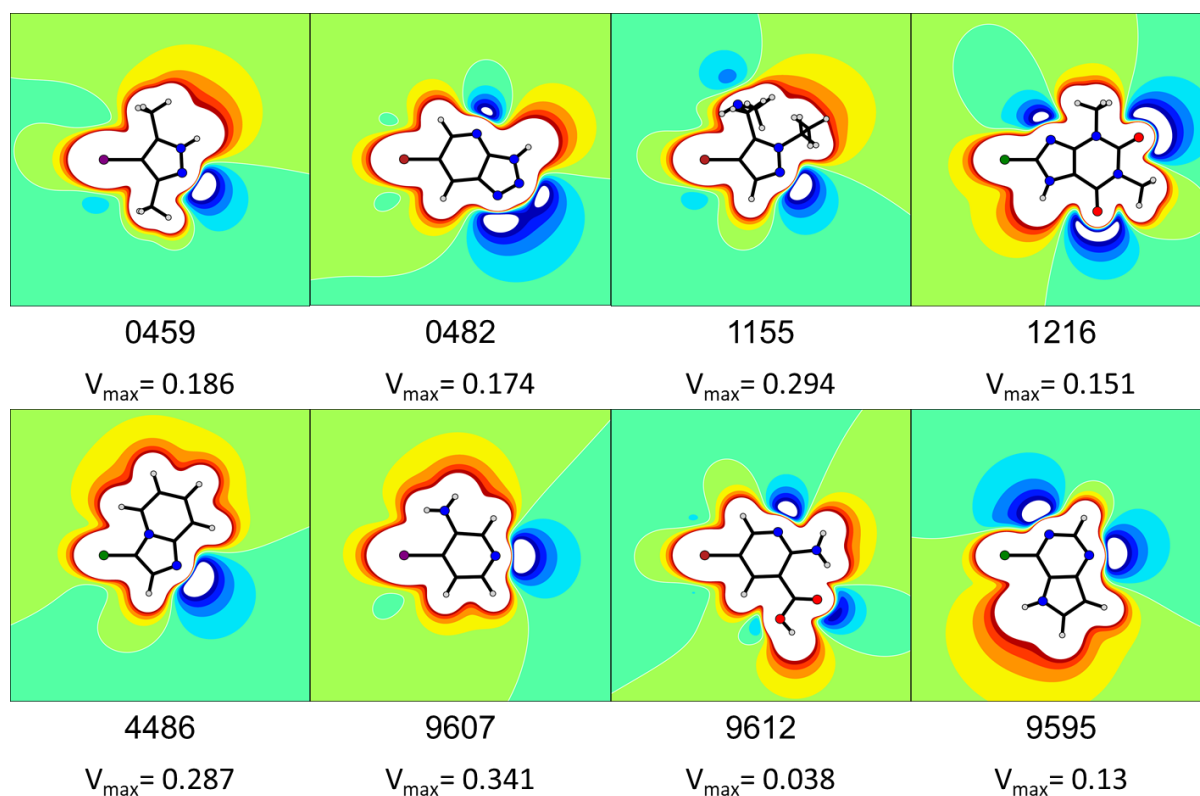


Figure 36. Electrostatic potential of an example set of fragments. Blue (-0.05 au) was chosen for a negative potential and red (0.05 au) for positive potential. Fragments with the numbers 0459, 0482, 9612, and 9595 are fragments to be discussed in more detail in further chapters. Fragment 1155 has the highest potential for a bromine fragment, 4486 has the highest for chlorine, and 9607 for iodine. All three potentials are close together, showing the immense effects of the scaffold and substituents on the ESP. ESPs range from -0.21 to 0.341 at a pH of 7 and at 0.02 au.

The conclusions derived from Table 5 and Table 6 lead to the incorporation of  $V_{\max}$ -values into selecting fragments and the generation of ESP-plots, which would support selection and extend the  $V_{\max}$  concept from a simple number to a more elaborate viewing of the electrostatic potential of a fragment (Figure 36). The depicted ESP-plots are examples of the variability of halogen bonding interfaces (XB-motif) of the HEFLib. Besides the different sizes, susceptibility to being polarized, and  $\sigma$ -hole potency of the three halogens, the environment of said halogen plays a vital role in the binding interface of these fragments. Neighboring atoms with a positive potential can shift the position of the  $V_{\max}$  towards said substituent, or if a negative potential is present, push the position away from the inducing substituent. An excellent example of this effect can be seen in the ESP-plot of 9595. The pyrimidine-nitrogen next to the chlorine exerts a negative potential and shifts the potential of the chlorine. This suggests that the most positive electron density on the halogen does not always have to be precisely in front of the halogen. In conclusion, XB geometries with deviation from the 180° angle could be

more attractive than previously thought. The protonation state is essential to these plots, and much care should be taken if protomers are not clearly distinguishable. In this study, the protomers were generated by MOE<sup>186</sup>, and in doubt, ESP-plots of multiple protomers were generated and reviewed. The halogen itself not only shapes the halogen bonding motif but also combines with these neighboring atoms, allowing a variety of secondary interactions next to the halogen. ESP-plots are a valuable tool to illustrate this complex situation at a glance, which was possible due to the high planarity of the aromatic scaffold in all fragments. In Figure 36, the fragments with the highest  $V_{\max}$ -values of each halogen are depicted (9607 for iodine, 1155 for bromine, and 4486 for chlorine) with values ranging from 0.287 for chlorine to 0.341 for iodine; in all three cases, a positive (partial) charge in the scaffold is responsible for the high  $\sigma$ -hole tuning. Fragments 4486 and 9607 show STD signals with a multitude of proteins, whereas fragment 1155 has not interacted with any of the targets. Fragments 0459, 0482, 1216, 9595, and 9612 are depicted due to their interesting behavior in the validation and crystallization experiments (see sections 2.5.9 and 2.6). Except for 9612, all of the fragments, which would later prove most valuable, show a  $V_{\max} > 0.13$ . In retrospect, the use of  $V_{\max}$  as a criterion for fragment selection proved to be helpful but not omnipotent.

The main drawback of the ESP-plots is, as the name implies, only electrostatics are viewed, and other effects of binding events are neglected. The most significant influence on electrostatics is the induction effects of the protein on the electrostatics of fragments upon binding. These are unpredictable without knowledge of the binding mode, and even with information of the binding motif cannot be not completely understood<sup>37</sup>. One example mentioned earlier is the importance of protomers and tautomers in the calculations. Binding pockets can easily bind a tautomer of a compound, which is unfavorable in solution and consequently changes the complete ESP. These deficiencies in the  $V_{\max}$ -value are why we introduced a third criterion to prioritize fragments. It was of great importance to utilize at least one experimental criterion to aid the ESP-plots of a calculated one, therefore balancing calculations and experiments in the prioritization. Although STD spectra should not be ranked, as discussed previously<sup>185</sup>, we could use the spectra to select fragments with STD signals of protons in a vicinal position to the halogen. The vicinity would increase the chance of the halogens engaging in direct protein-ligand contacts instead of pointing to the

#### 2.4.5 Kinases in the Screening

solvent. As the STD effect increases with more intense contact of the fragment to the protein, a more robust signal should indirectly point to higher proximity of the halogen and protein. These multifactorial priorities were used to select a total of 132 fragment-protein combinations (out of 344 in total) for further investigation by ITC. These included the uniquely binding fragments as well as fragments showing a hit-event with a multitude of targets.

#### 2.4.5 Kinases in the Screening

As stated in section 2.1, the proteins were selected for the screening due to various reasons. To test the library, starting with the JNK3, AAK1, CAMK1G, DOT1L, and IDO1, as a diverse set of proteins. As the kinases turned out to be promising targets with higher chances of crystal structures and a higher hit rate (see section 2.4.2), the DYRK1a and JNK2 were screened as well. The proteins were interpreted together as we could see the general message and behavior of the HEFLib screening did not change after the first four proteins were screened<sup>187</sup>. The fragment shift towards lesser solubility and higher  $V_{max}$  can also be seen with the initial four proteins<sup>187</sup>. The preference for iodine-containing fragments over bromine and chlorine could be seen as well.

The relatively high conservation in the binding site of kinases led to the assumption that fragment hits should share certain similarities. Figure 15 showcases the evolutionary distance between the five kinases; AAK1 and CAMK1G are evolutionary as far apart on the kinome as possible. The difference in the binding pocket is further elaborated in Figure 16 (see section 2.1). Figure 19 and Table 4 show that most of the fragments bind to less than three target proteins. Although the kinases share a certain body of promiscuous binding from certain fragments, the hit rate could have been much higher with five kinases tested if fragments would unspecifically bind kinases. Many fragments found in the screening only bind to two kinases at most. In conclusion, the high proportion of kinases seems not to be problematic.



The common element of all kinases is the hinge region between the N-terminal and C-terminal domain, responsible for the adenine's binding. The central hinge region consists of five varying amino acids; the backbone nitrogen and oxygen can engage in up to four strong hydrogen bonds<sup>188</sup>. Most optimized inhibitors utilize this conformation and exhibit at least two hydrogen bonds to the hinge region<sup>188</sup>. The so-called hinge binding motif usually includes a hydrogen bond donor and an acceptor spaced by either one or two bond lengths, depending on which backbone atoms are targeted<sup>188</sup>. By checking the screening results (Table 3) for “pan-kinase” binding fragments, fourteen interact with all five kinases. Nine of these fourteen include a hinge binding motif; only fragments 1256, 1266, 4486, 7396, and 9595 did not include this motif (Figure 37). Arguably 1256 and 1266 could interact with the hinge region, although the HB donor and acceptor are three bond lengths apart. The remaining four fragments might bind in multiple motifs distinct at different surface locations.

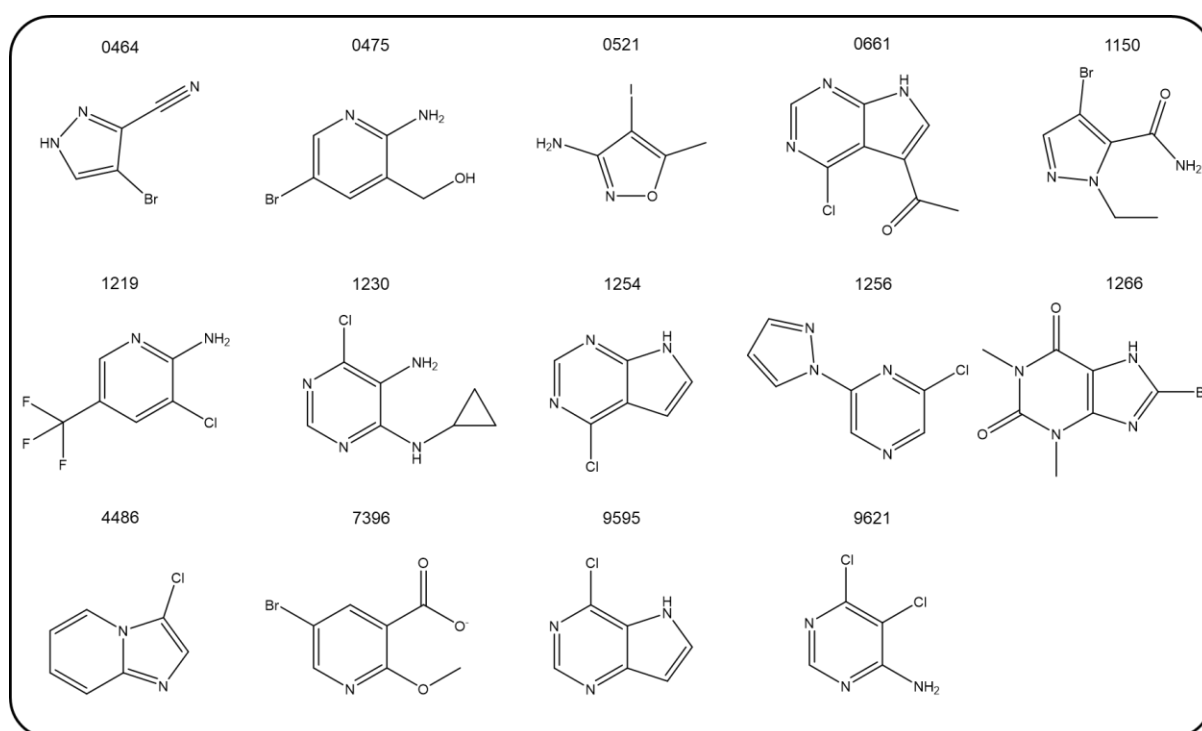


Figure 37. Fragments binding to all five kinases of the 14 fragments nine show a standard hinge binding motif.

### 2.4.6 IDO1 Screening

The IDO1 occupies a special place in the screening as it is the only protein screened with a cofactor present. The heme in the tryptophan pocket (so-called A-pocket) has been used as an anchor point for many inhibitors<sup>149, 150, 165</sup>. In the past decade, inhibitors for IDO1 targeting every state of the heme cofactor have been reported<sup>170, 171, 173</sup>. Problematic is the lack of studies regarding the oxidation state of the iron in the inhibited state. Inhibitors have been found to bind to either both states or preference one oxidation state. The literature lacks a comparison of binding affinities with both states. By design of the STD experiment, no reducing agent is present in the measuring buffer. As a result, the iron in the IDO should oxidize from the ferrous to the ferric state. The SEC buffer includes DTT and prevents the iron from oxidization (see section 2.2.6). Unfortunately, no kinetics for the reaction of oxygen with the iron of IDO1 is known, and these could greatly vary in different buffers.

As the reaction with tryptophan and oxygen is reasonably fast with a turnover number of  $1-1.7\text{s}^{-1}$ <sup>169, 189</sup>, it is reasonable to assume the iron is oxidized until NMR measurements. The actual state of the iron stays unclear; the last point of measurement shows the IDO1 in the ferrous state, an hour before NMR measurements. A more recent study could prove that roughly 85% of IDO1 are not bound to heme at all in the cell<sup>173</sup>. Nelp et al. could prove to inhibit heme binding to the apo IDO1 and effectively shut off the enzymatic activity. This literature leads to the conclusion that a further screen of the HEFLib with apo-IDO1 could be an interesting way to find further hit fragments. In addition, a larger binding site would be available for interactions. The independence of heme would alleviate the challenges discussed with heme saturation in purification and subsequent measurements.

### 2.4.7 BIRC5 Screening

BIRC5 is another special case within the proteins studied, as it has neither enzymatic activity nor a sophisticated binding pocket. Furthermore, the protein exists as a dimer in solution, as can be seen in the SEC chromatogram (Figure 19) (see section 2.2.2). To compensate for the dimeric form, BIRC5 was used at a concentration of 40  $\mu\text{M}$ , which should result in a concentration of 20  $\mu\text{M}$  dimer. To date, only a few direct inhibitors of BIRC5 have been described<sup>148</sup>. Some target the dimerization site and interfere with the formation of the dimer<sup>148</sup>. This dimerization leads to an inherent problem with the general measurement principle of STD-NMR. In general, proteins with molecular weights greater than 30 kDa are considered suitable for STD because of their low  $T_2$  time. There are known cases of proteins weighing as little as 20 kDa being measured with STD<sup>190</sup>. Therefore, the 32 kDa of the BIRC5 dimer should be suitable for STD-NMR. The problem arises when the proteins dissociate to a monomeric form, which is always present in small amounts and could be further increased by binding fragments.

Consequently, some fragments could bind to BIRC5 and inhibit dimer formation, biasing the STD measurement. The screening of BIRC5 was an attempt to test the method's limitations while increasing the diversity of proteins studied. Since the dimerization and low molecular mass complicate the interpretation of the spectra, for these reasons, the data was interpreted carefully and with a fair amount of skepticism.

As no fragment could be validated, the hits for BIRC5 generated by the STD should be reevaluated, and orthogonal screening methods like differential scanning fluorometry could be used to increase the credibility of the STD data.

### 2.5 Hit Characterization by ITC

Isothermal Titration Calorimetry was chosen to incorporate a completely orthogonal principle of measurement. In addition to  $K_D$ -values being determined, thermodynamic properties can be directly measured and used to increase knowledge of the binding mode. Another advantage is the measurement in solution without any labeling or immobilization, allowing allosteric binding events to be observed—a feature proving to be difficult with other techniques (e. g., SPR and FP). The drawbacks of ITC are the high amounts of protein necessary and the long measuring times, which are even longer with weak binding fragments, as the injection spacing has to be elongated to reach a stable baseline after injection of large compound quantities<sup>36</sup>. As described in section 2.4.3, hits had to be prioritized to facilitate the hit validation. Rühmann et al. and Turnbull et al. suggested that a  $K_D$ -value above 1 mM is not measurable with a direct ITC approach, and competitive inhibitors have to be utilized<sup>36, 106</sup>. The limit is imparted by the limited solubility of the fragment or the protein. There are cases in material sciences where millimolar affinities of ions to chelating compounds have been observed<sup>34</sup>. Some fragments in this study had an affinity above this threshold, and a value has been given in some cases to allow comparison. However, these values are highly inaccurate and should be interpreted as  $K_D > 1$  mM in case of doubt. It should be kept in mind that the solubility and amount of protein per measurement pushed the ITC technique to its limits in these cases.

#### 2.5.1 Fragment Validation Overview

From the 764 STD experiments (1528 screenings of fragments), 344 hit events could be observed, of which 132 fragment-protein combinations were selected. These included 19 fragments measured with JNK3, 23 with JNK2, 30 with DYRK1a, 12 with CAMK1G, 14 with AAK1, 7 fragments with DOT1L, 5 fragments with BIRC5, and 22 fragments with IDO1. 25 fragment protein combinations measured exhibited a  $K_D > 800 \mu\text{M}$ . Additional five fragments could be validated in an ITC experiment as binders. However, the affinities were too weak to fit a reliable  $K_D$ -value, or the protein saturation was too low (<80 %) for reliable curve fitting. Corresponding to roughly 20 % of all tested STD hits could be validated by ITC. If this percentage is projected to all 344 hits found overall, another 47 fragment protein combinations could be validated in

ITC measurements. A few reasons probably skew an overall projection. Firstly, none of the BIRC5 (25 % of hits tested) and DOT1L (78 % of hits tested) hits could be validated, and secondly, the rate of validated hits for each protein and the percentage of tested hits varies by a large margin. For IDO1, a validation rate of 9 % (71 % of hits tested) was the lowest, followed by the JNK2 with 17 % (39 % of hits tested), DYRK1a with 20 % (39 % of hits tested), JNK3 with 21 % (38 % of hits tested), CAMK1G with 33 % (29 % of hits tested), and AAK1 with 36 % (25 % of hits tested). On average, 43 % of hits per protein were tested. AAK1 and CAMK1G are the most yielding targets with hit validation rates of over 30 %, although only 25-29 % of the hits were tested.

Suppose the validation rate for AAK1 and CAMK1G is reliable. In that case, although only a quarter to a third of hits was tested, the remaining unvalidated hits (72 in total) hold tremendous potential. Due to the main body of non-validated fragments binding to multiple targets in the NMR screening, there is reason to expect many fragments to be validated with multiple kinases. The STD hits of the remaining three kinases (JNK2, JNK3, and DYRK1a) were tested more thoroughly and still show reasonable validation rates of about 20 %. In summary, of 1528 screened fragment-protein combinations, 344 were initial hits, from which 132 were tested, and 25 could be confirmed. This results in 1.6 % of initial experiments completing the validation stage and could be subjected to further characterization in this study. There are still many fragments needing confirmation by ITC (212 fragment protein combinations remaining). Projecting the percentages to the whole set of hits is rather difficult due to the high variance in the validation rates. A conservative estimation of hits with the lowest validation rate of 9 % (IDO1) is compared to the highest rate of 36 % (AAK1). From these numbers, potentially 28 to 113 hits could be validated with affinities in the high micromolar to low millimolar range (DOT1L and BIRC5 data excluded). An investigation of the remaining AAK1 and CAMK1G hits seems to be the most promising. With the lowest confirmation rate and the highest portion of IDO1 hits being tested, resources should be guided into characterization instead of adding further fragments. Table 7 depicts the fragments, which are either weak binders or false positives in the screening.

Table 7. Overview of all unsuccessful fragment validation with their respective target.

<b>JNK3</b>	<b>JNK2</b>	<b>DYRK1a</b>	<b>CAMK1G</b>	<b>AAK1</b>	<b>BIRC5</b>	<b>DOT1L</b>	<b>IDO1</b>
0090	0087	0089	0087	0403	0120	7655	0123
0114	0401	0090	1102	0465	0459	9613	0464
0459	0403	0116	1234	1102	1234	9615	0521
0463	0474	0123	1243	1227	1270	9619	0661
0481	0476	0522	1250	1234	7405	9809	1225
0523	0482	0524	1253	7402	7407		1234
1152	0523	1152	9596	9616	7420		1243
1234	1217	1213					1246
1242	1234	1223					1254
7395	1236	1227					1266
7396	1237	1234					1268
7403	1244	1238					4486
7415	1257	1244					7394
	7395	1253					7411
	7396	1271					7420
	7400	4481					7423
	7403	7396					7425
	7655	7402					7427
	9613	7415					7429
		7429					9620
		9596					
		9615					
		9616					
		9620					

### 2.5.2 JNK3 Fragment Validation

The affinity of four fragments could be determined with JNK3. By the ITC thermograms depicted in Figure 38, four fragments could be fitted with a  $K_D$ -value. Binding affinities vary from 127  $\mu$ M (9595) up to 312  $\mu$ M (0459) (Table 8). The ITC measurements of 7409 show distinctive heat emissions with baseline shifts after every injection. The baseline is corrected in the peak integration process, and the baseline shifts are skewed to look like very broad peaks (Figure 38 B). Even when the compound was titrated into the buffer solution, this behavior was observed. No precipitation of the fragment was observed after the measurements, which might have explained this behavior. Additional stability measurements with water and GSH show no instability at all. The reason for the baseline shift remains elusive. The problem is independent of proteins, and the buffer contains only HEPES, NaCl, MgCl<sub>2</sub>, and TCEP.

## 2.5.1 Fragment Validation Overview

As a conclusion, we assume the unconventional behavior to be a compound property. It is also possible, although unrealistic, that the compound reacts with the Hastelloy of the titration cell. Although less pronounced, the same problems were observed with AAK1 (see section 2.5.6). All four validated fragments bind to at least another protein. It was impossible to measure a  $K_D$ -value of fragments 0463 and 0481, the fragments binding uniquely to JNK3 (Figure 31). Fragment 9809 has not been surveyed, but it might be worthwhile to follow up.

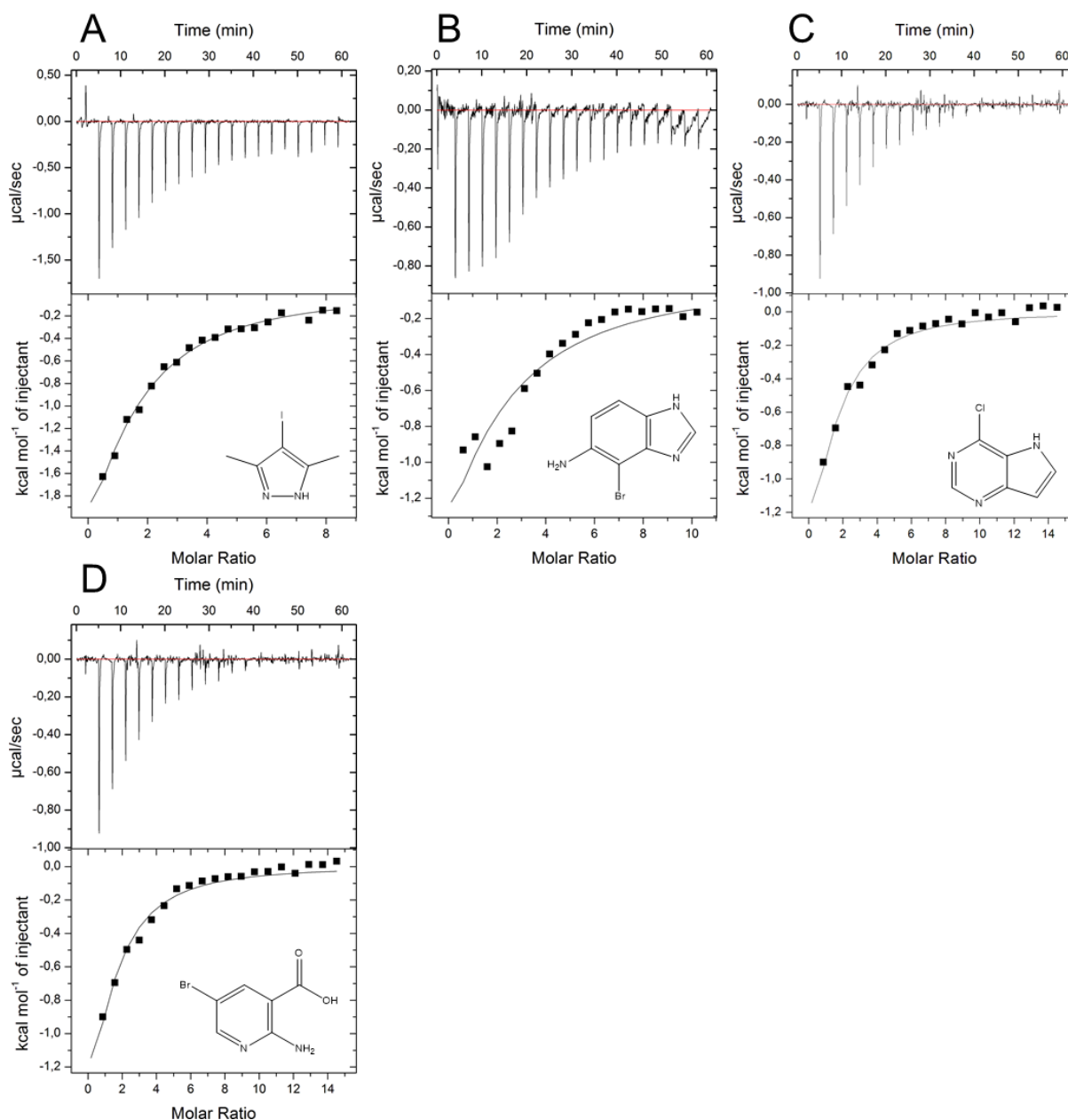


Figure 38. Exemplary ITC thermograms of JNK3 validated fragments **A**: 0459, **B**: 7409, **C**: 9595, **D**: 9612. The peaks in panel B seem large due to the baseline correction applied. The fragment exerts a downward stepping baseline, even if no protein is present. The resulting affinity is flawed due to this compound peculiarity.

### 2.5.3 JNK2 Fragment Validation

With the JNK2, four fragments could be verified. Three out of these four fragments bind to JNK3 as well (Figure 38 and Figure 39). The dissociation constants measured with JNK2 varied widely from the JNK3 dissociation constants. Every fragment binds much stronger to JNK2 than to JNK3. The best binding fragment had a  $K_D=18\mu\text{M}$  (9605), and the weakest of the verified fragments binds with a dissociation constant of  $81\mu\text{M}$  (0459). Due to the high similarity of JNK2 and JNK3 (see section 2.1), fragments binding to both kinases are of great interest and are worth further evaluation, as preferences could be optimized into selectivity. Especially fragments binding only to the two JNKs and not to the remaining three kinases of this work. When the binding constants of these fragments are compared, 9595 and 0459 bind three to four times better to JNK2 than JNK3 (Table 8). Fragment 9612 has the advantage of binding roughly six times stronger to JNK2 than JNK3, with the added benefit of scoring no hit with any of the other three kinases. Thus, although the fragment features a classic hinge binding motif, a certain selectivity might be found. The high affinity combined with the preference for JNK2 over JNK3 were the main reasons fragment 9612 was further evaluated, and a set of SAR compounds were acquired. The crystallization of a structure with JNK3 gave great insights into the binding mode (see sections 2.5.13 and 2.6.4). Of the five uniquely binding fragments, none showed measurable binding in the ITC (Figure 31).



### 2.5.3 JNK2 Fragment Validation

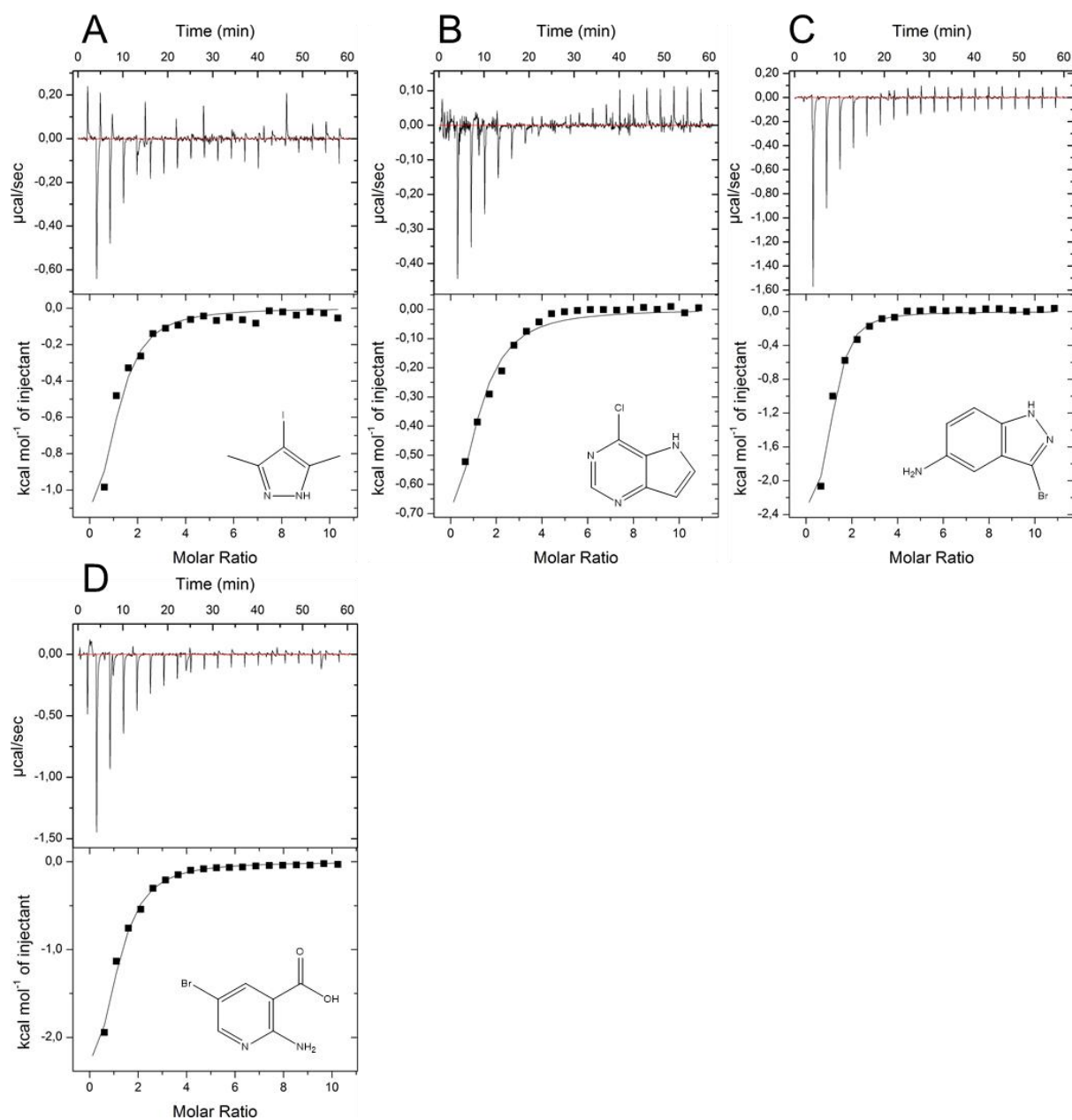


Figure 39. Exemplary ITC thermograms of JNK2 validated fragments **A**: 0459, **B**: 9595, **C**: 9605, **D**:9612.

## 2.5.4 DYRK1a Fragment Validation

Five fragments were verified with the DYRK1a, with affinities ranging from 163  $\mu\text{M}$  (9595) up to 680  $\mu\text{M}$  (7419) (Figure 40, Table 8). In addition to the fragments 0459 and 9595, which bind to all kinases, three more fragments were identified as binders to the DYRK1a. All nine uniquely binding fragments were tested (Figure 32), and 0468 and 7419 are the ones with detectable heat emissions. The signal-to-noise ratio of fragment 1233 measurements is the weakest, and integration and curve-fitting pose high uncertainties in this magnitude of affinity. As a result, of even less heat emission observed with fragment 0468, a binding affinity could not reliably be attained. The strong resemblance of the fragments 7419 and 7396, binding uniquely to DYRK1a, is notable. A methyl ether (7396) instead of a hydroxy group (7419) is the single difference between both fragments (Figure 40) and results in the complete abolishment of the affinity towards DYRK1a. Both fragments are part of a satellite group with high structural similarity, of whom fragment 9612 is a part as well. This fragment binds to JNK2 and JNK3. Although all these fragments are similar, the 5-chloro-2-hydroxynicotinic acid (7419) binds to DYRK1a and CAMK1G, but only the 5-bromo-2-aminonicotinic acid (9612) binds to the JNKs (see section 2.5.12). The minuscule structural differences lead to significant changes in affinity. A better understanding of fragment binding modes in kinases could be achieved with more elucidated binding modes for these fragments. A more in-depth study of the structure-affinity relationship of these compounds to the different kinases might reveal completely different changes in the binding mode from minimal changes in the binding pocket or fragment.

Another noteworthy fragment is 0482 (6-bromo-1*H*-[1,2,3]triazolo[4,5-*b*]pyridine, Figure 40 B), as the screening showed a hit with the DYRK1a and the CAMK1G, the chloro-variant of the same fragment (0468) was seen to only generate a unique hit with the DYRK1a. The 6-chloro-1*H*-[1,2,3]triazolo[4,5-*b*]pyridine (0468) showed signs of binding with distinct heat emission in the ITC, the effect was too weak for  $K_D$  determination, and insufficient protein saturation was achieved (Figure 43 B). This affinity difference might indicate a halogen bond-dependent binding, as the larger halogen is capable of a stronger bond (compare fragment 0468  $V_{\text{max}}=0.141$  to fragment 0482  $V_{\text{max}}=0.174$ ). An SAR study of the fragment was conducted (see section 2.5.11) to assess this interesting behavior.

## 2.5.4 DYRK1a Fragment Validation

As the dissociation constants must be fitted with a fixed stoichiometry due to the limitations of the ITC approach, the  $K_D$ -value of 0482 has to be reevaluated as subsequent crystallization showed (see section 2.6.3).

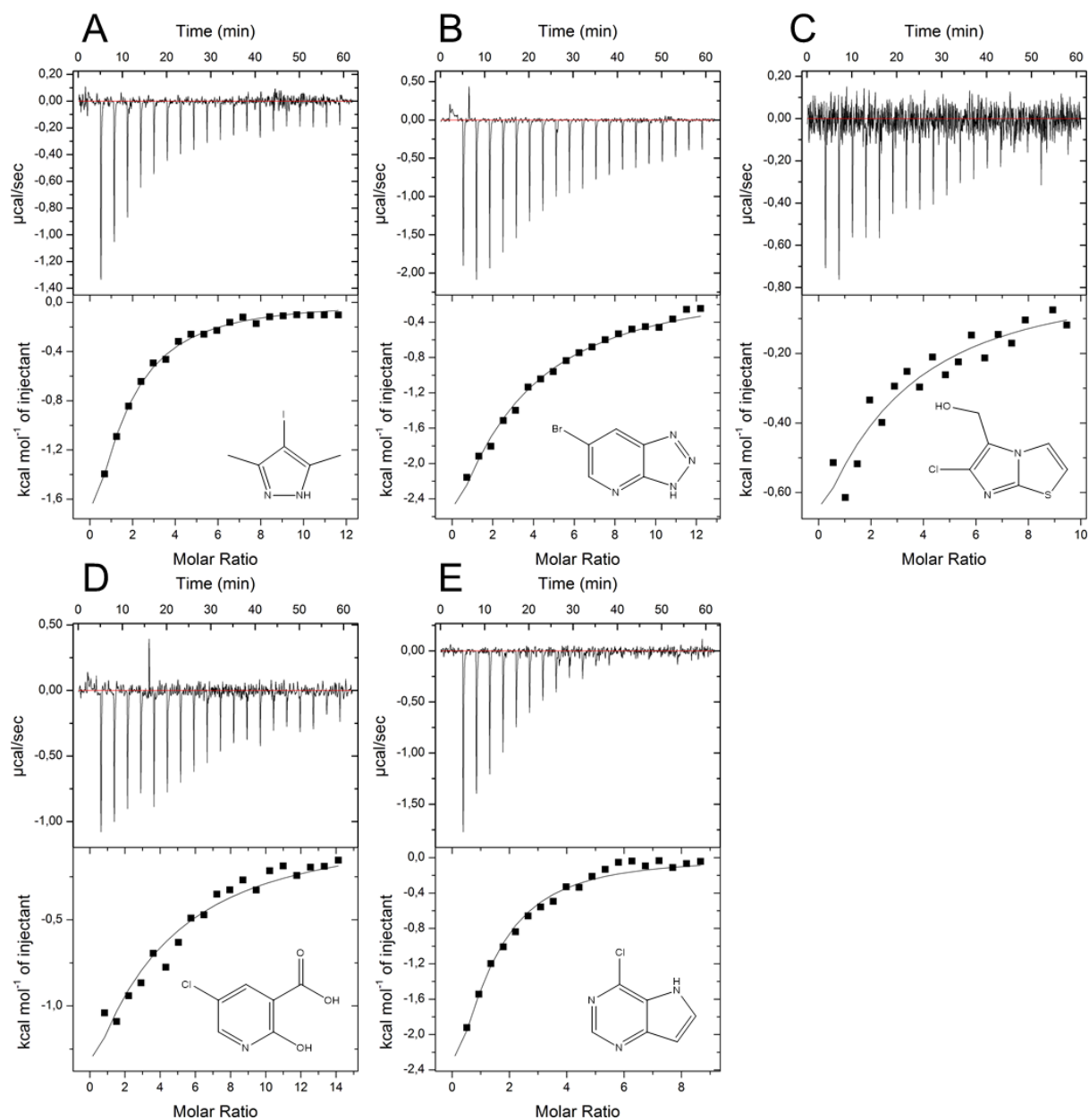


Figure 40. Exemplary ITC thermograms of DYRK1a validated fragments **A:** 0459, **B:** 0482, **C:** 1233, **D:** 7419, **E:** 9595. The affinity fitted to 1233 in panel C has a significant standard deviation as the signal-to-noise ratio is becoming so small.

## 2.5.5 CAMK1G Fragment Validation

Five fragments were verified as binders to CAMK1G, with the tightest binding occurring with  $K_D=133\ \mu\text{M}$  (7405) and fragment 9605 with a  $K_D=700\ \mu\text{M}$  (Figure 41, Table 8). Fragment 1250 was the only fragment with a unique binding event in the screening but could not be verified with the ITC. All of the fragments verified with the CAMK1G bind to another target. Due to the limited knowledge of crystallization conditions of CAMK1G (PDB entry 2JAM), the hits were postponed in favor of the DYRK1a and JNK3.

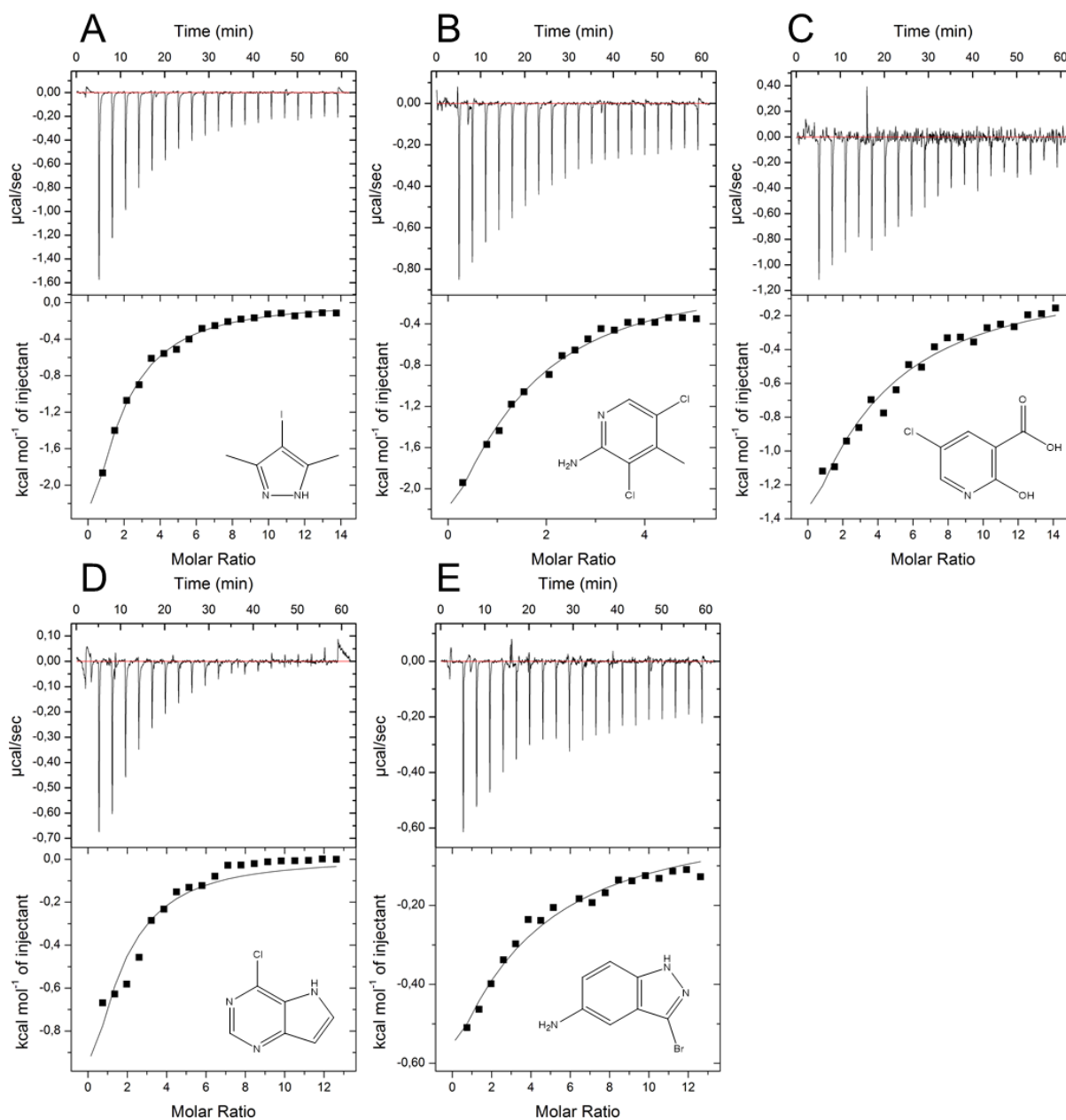


Figure 41. Exemplary ITC data of CAMK1G validated fragments **A:** 0459, **B:** 7405, **C:** 7419, **D:** 9595, **E:** 9605.

### 2.5.6 AAK1 Fragment Validation

With AAK1, five fragments were validated as binders (Figure 42). All fragments bind considerably more potent to AAK1 than most of the verified fragments on the other targets. The weakest binding fragment has an affinity of 184  $\mu\text{M}$  (0474). As seen with the JNK3, fragment 7409 shows the same downward stepping baseline with AAK1. Although the shift is not as well observed as with the JNK3, the associated curve determination has to be viewed with caution. Fragments 0459 and 9595 were verified with every kinase, including AAK1. Especially the latter fragment stands out; with a  $K_D = 6 \mu\text{M}$ , the fragment was the best binding hit in the entire work. In Addition, 9595 binds AAK1 20-30 times stronger than JNK3 or CAMK1G (Table 8). With this affinity, the fragment was at least binding seven times stronger to AAK1 than the next best hit (JNK2  $K_D = 40 \mu\text{M}$ ) and 34 times stronger to the least affine target (CAMK1G  $K_D = 200 \mu\text{M}$ ). A set of SAR compounds were acquired to understand this behavior, and crystallization efforts were undertaken (see section 2.6.10). The same but less intense pattern can be observed with 0459, which binds the strongest to AAK1 with a  $K_D = 18 \mu\text{M}$ . The next best hit is bound 4.5 times weaker (JNK2), and 17 times less potent to its weakest hit (JNK3). The affinity towards CAMK1G and DYRK1a lies in between (Table 8).

Of the uniquely binding fragments (Figure 30), 0471 has been tested and verified as a binder, and fragment 0465 could not be verified as a hit. The results with the AAK1 were the most promising due to the number of verified fragments and the high affinities exerted by all of the fragments.

## 2.5.6 AAK1 Fragment Validation

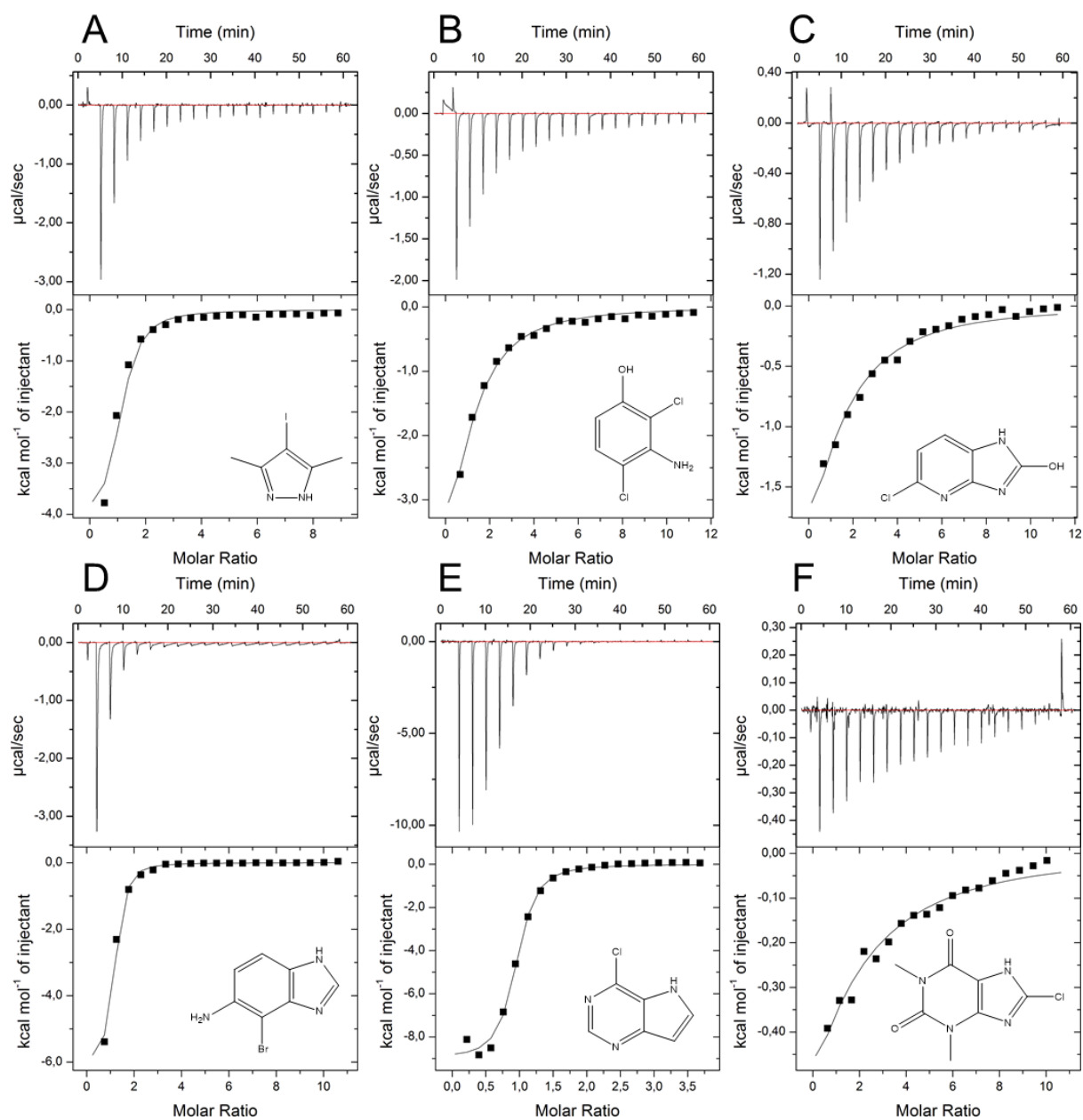


Figure 42. Example ITC data of AAK1 validated fragments **A**: 0459, **B**: 0471, **C**: 0474, **D**: 7409, **E**: 9595. **F**: Fragment 1216 validated with IDO1.

### 2.5.7 IDO1 Fragment Validation

For the IDO1, only one fragment could be reliably verified. Fragment 1216 has a  $K_D$ -value of  $394\ \mu\text{M}$  (Figure 42 F, Table 8). The thermogram of fragment 0459 showed significant heat emission in the ITC, but protein saturation was incomplete, and consequently, no reliable affinity could be determined (Figure 43 E). As previously shown, there were four uniquely binding fragments (Figure 30). From these, none could be verified as a binder. In addition to the verified fragment 1216, the bromo-variant of the fragment (1266) was tested as well. Unfortunately, no binding could be verified. Attenuation of the binding by the introduction of bromine instead of chlorine might be attributed to the increased space necessary for the bromine. As the hit could be verified, but indications for a halogen being ineffective in the binding motif, the hit was of lesser interest in the study and received no follow-up.

### 2.5.8 Problematic Fragment Validation

In most cases, the fragment response to ITC was successful, but some problems arose as sporadically addressed previously. The most frequent problem in the validation process was the unsuccessful protein saturation with the fragment<sup>106</sup>. Due to limited solubility in the buffer, the measurements were confined to a 5 mM concentration. With higher fragment concentrations, a higher saturation could have been achieved. On the other hand, it was unresourceful to increase the protein concentration above  $100\ \mu\text{M}$ , and high protein concentration would have made excessive protein purification necessary. The solubility and stability of proteins at high concentrations at  $25\ ^\circ\text{C}$  were unknown and could have given false-positive or false-negative results. Partial unfolding could emit heat and lead to a skewing of the thermogram or unspecific binding events in the measurements. With a  $K_D=800\ \mu\text{M}$ , a receptor concentration of  $100\ \mu\text{M}$  (resulting in a  $c=0.125$ ) has been used to calculate affinities<sup>34, 106</sup>.

## 2.5.8 Problematic Fragment Validation

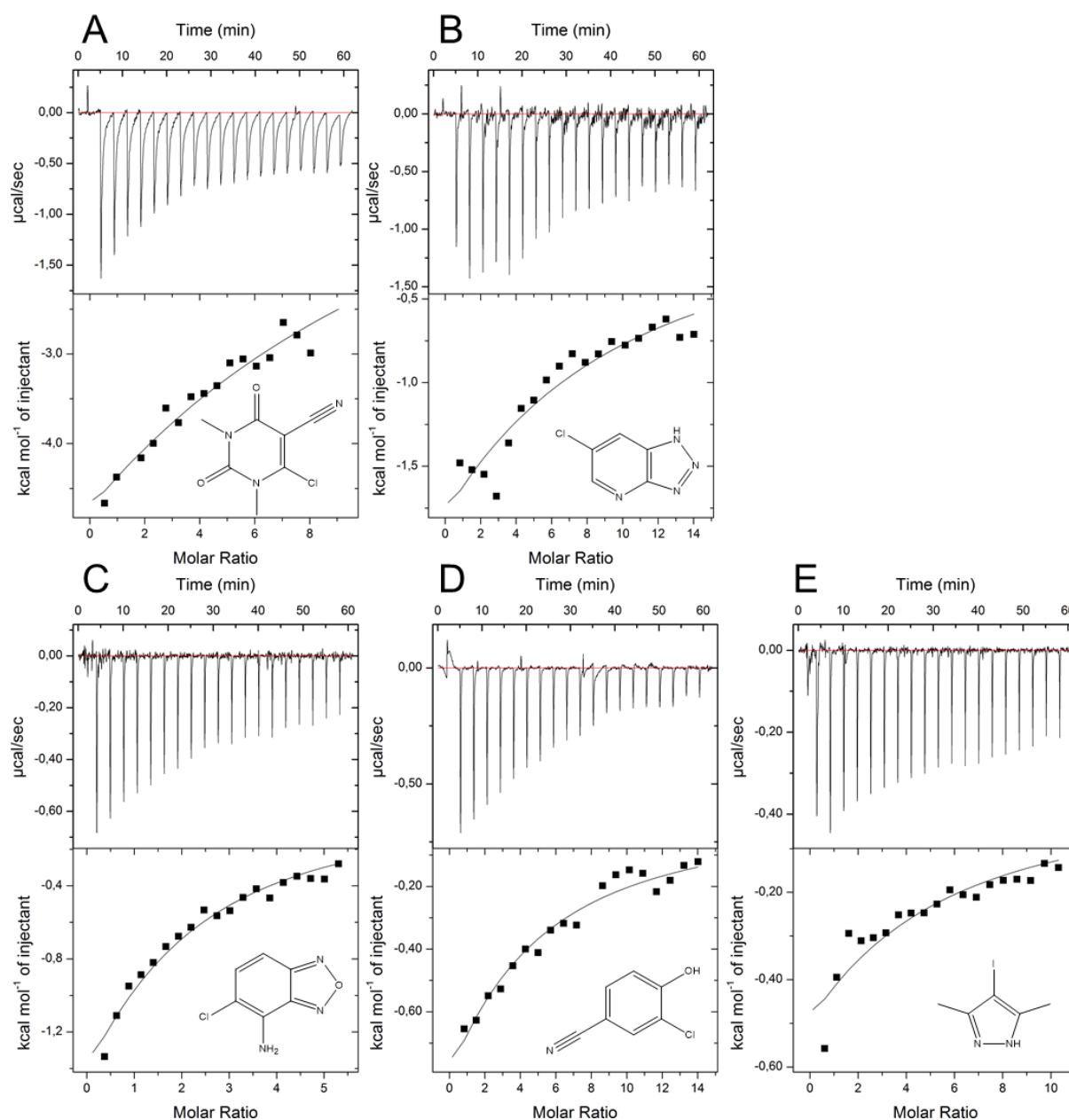


Figure 43. Example ITC data of fragments with various issues in the validation process. **A**: 1234 with JNK3, the exemplified prominent peaks without baseline return could be observed with other proteins as well. A buffer mismatch could be ruled out, as the behavior only occurred with protein present. The most common problem was the low saturation of the protein, as binding affinities were too low. This is exemplified in pane **B**: 0468 with DYRK1a, **C**: 1217 with AAK1, **D**: 1257 with AAK1, and **E**: 0459 with IDO1. A binding event can be observed, but the saturation is too low for  $K_D$  estimation.

However, affinities have been estimated with the analyte saturation as low as 50%<sup>34</sup>. For this study, a more conservative lower limit of 80% protein saturation was used. The accuracy of affinity determination is also highly dependent on the signal-to-noise ratio in the thermogram. Curve fitting becomes impossible if the dilution heat is as high as the binding heat.



### 2.5.9 Ranking of Fragments

In Figure 43 (C-E), the heat emitted through a binding event is significantly smaller than the heat emissions in Figure 38 to Figure 42. The low protein saturation illustrated by the missing plateau in the thermogram makes it unreliable to determine an accurate affinity.

Nonetheless, heat emissions are observed, indicating some binding event even though it is not quantifiable. These fragments were included in Table 8 to complete the validation overview, although the  $K_D$ -values assigned to the fragments are prone to accuracy and precision errors. Figure 43 A was mentioned earlier as further evidence of the reactive nature of 1234. The prominent peaks, which do not reach baseline even after 180 seconds of delay, and the broad peaks remaining even after an excess of the fragment was added, show some reaction occurs, although initial saturation is observed. Perhaps some specific binding is occurring, but the overlapping of different unspecific or covalent reactions makes it impossible to deconvolute both events from each other. As the large peaks could not be observed with buffer alone, there was no reason to believe dilution was the cause of the peaks remaining this large. The typical dilution heat emission is in the range of 0.2 kcal/mol, which is exceeded by a factor of 15. This behavior was observed with all tested proteins, suggesting an unspecific covalent reaction.

### 2.5.9 Ranking of Fragments

Sophisticated SAR series are generally ranked based on their  $IC_{50}$  or  $K_D$ -values. An easy and intuitive method, optimizing simply for the best binding inhibitor, disregarding gaining affinity, is associated with adding functional groups or scaffolds. As fragments are substances with high variability in affinity, size, and scaffold, the simple affinity ranking is not expedient. The most common way to overcome the problem of ranking completely unrelated structures is the normalization of affinities of differently sized fragments by the number of (non-hydrogen) heavy atoms (HA)<sup>191, 192</sup>. As the binding affinity is based on the Gibbs free energy, which is defined as:

$$\text{Eq. 2.} \quad \Delta G_0 = -RT \cdot \ln \left( \frac{K_D}{C_0} \right)$$

With R being the ideal gas constant and T being the temperature in K,  $C_0$  is the standard concentration.

Assuming 300K, neutral pH, and standard concentration of 1 M<sup>193</sup>, equation two approximates:

$$\text{Eq. 3.} \quad \Delta G_0 \approx -1.37 \cdot \lg K_D$$

The ligand efficiency (LE) is defined as the quotient of binding energy divided by the HA number. Thus LE is defined as<sup>192</sup>:

$$\text{Eq. 4.} \quad \text{LE} = \frac{\Delta G_0}{\text{HA}} = -\frac{1.37}{\text{HA}} \cdot \lg K_D$$

It has to be stressed that the LE is based on conventions, mainly to divide the affinity by 1 M<sup>193</sup>. Due to the convention, LE should not be used as a predictor of potency or a complex decision metric<sup>194</sup>. Instead, it should be used as a guideline for comparison in the context of additional data. Great caution has to be used to calculate and interpret LE values. There are many possible ways to skew ligand efficiencies unintentionally<sup>194</sup>.

As the incorporation of size enables us to compare fragments with variable sizes, weak binding fragments can be ranked higher than larger and stronger binding fragments. Having subsequent rounds of optimization in mind, possible vectors for fragment growing or linking are desirable attributes. Figure 44 summarizes the LE of all validated hits, and fragment 1234 is the only one with a ligand efficiency lower than 0.3. The fragment has been identified as reactive with GSH (see Section 2.3). Consequently, side reactions likely distorted the ITC measurement. Therefore, it should not be counted as a hit and only serves an illustrative purpose in Figure 44 and Table 8. Direct ITC can only reliably detect affinities  $K_D < 800 \mu\text{M}$  with reasonable accuracy (within the given limits). A fragment with less than 15 HA automatically has a  $\text{LE} > 0.3$ . Thus, the ITC serves the purpose of selecting favorable fragments, in addition to verifying screening hits.

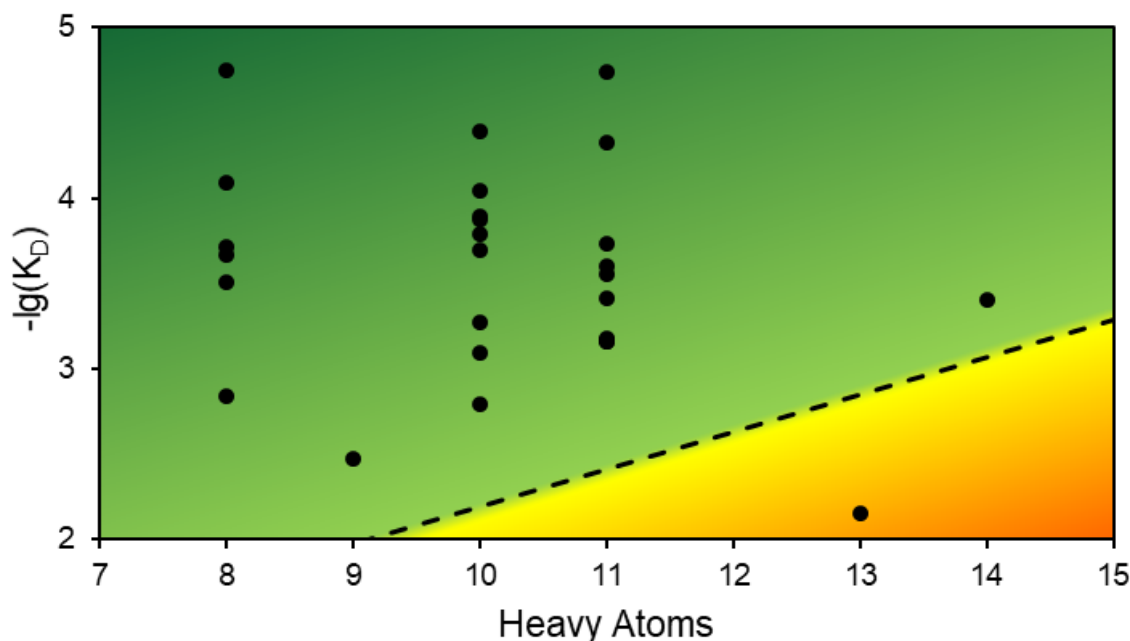


Figure 44. The number of heavy atoms plotted against the negative decadic logarithm of dissociation constants. The dotted line corresponds to a ligand efficiency of 0.3, and every fragment except 1234 has a LE > 0.3.

With LE as the main criterion to compare the potential of a fragment, Table 8 ranks the fragments by ligand efficiency over pure affinity. The fragments 0459 and 9595 are the highest-ranked fragments in the validation. Both exhibit varying affinities with different proteins, and 0459 has the highest LE due to its small size and the strong interaction with AAK1. Fragment 9595 has a 3-fold better affinity but is ranked second due to the larger size. Except for 1216, all fragments have 11 or fewer heavy atoms. As a result, the ranking of the narrow size margin, the affinity, is less important in the LE calculation, which leads to the high ranking of 0459 with many targets. However, the K<sub>D</sub>-values are considerably larger than the affinities of 9612 or 0482.

In summary, there are nine hits with an LE > 0.6 and 15 hits with a LE > 0.5, which is considered more than promising for further evaluation. The size-independent ligand efficiency (SILE)<sup>192</sup> can be used to normalize the ligand efficiency and counteract the nonlinear relationship between affinity and heavy atom count. Compared to LE, the SILE is calculated with<sup>192</sup>:

$$\text{Eq. 5.} \quad \text{SILE} = \frac{\Delta G_0}{\text{HA}^{0.3}} = -1.37 \cdot \lg(K_D) / \text{HA}^{0.3}$$

Equation 5 establishes a linear relation of the affinity to the HA number. Due to the logarithm in the LE equation, the HA number has a much more significant impact on the LE than the affinity. If the SILE is used for ranking, the extremely weak binding fragments 0459 with IDO1 and 1257 with AAK1 are rated lower, and the effect of affinity is more prominent than the size. The measured affinities are exceptionally strong for the size of the fragment. In some FBDD campaigns, even millimolar affinities are used for hit-to-lead optimization, whereas millimolar affinities were not considered for characterization in this study<sup>11</sup>. The generally small size of the fragments further reinforces the unusual high affinities observed.

Size alone is often not enough to describe the potential of a fragment, and additional information is considered. Solubility is the critical parameter for ligand optimization next to affinity, and thus, logP values are factored into the ligand efficiency. There are different ways of including solubility into fragment parameters with varying advantages and disadvantages. The most prominent are lipophilic ligand efficiency (LLE) and lipophilicity corrected ligand efficiency (LELP). These are defined as Eq. 6 and Eq. 7<sup>192, 195</sup>.

$$\text{Eq. 6.} \quad \text{LLE} = -\lg K_D - \lg P$$

$$\text{Eq. 7.} \quad \text{LELP} = \frac{\lg P}{\text{LE}}$$

An LLE below five and a LELP value between -10 and 10 is considered optimal<sup>195</sup>. As no experimental logP values are available for the library, calculated SlogP values are used as surrogates, as is standard practice<sup>192, 195</sup>. The SlogP for all fragments was calculated on the same basis (*RDKit*)<sup>102, 103</sup> and thus can be used to compare fragments relative to each other. All verified fragments comply with the stricter ruleset of  $-5 < \text{LELP} < 5$ . In most cases, the LELP is preferred over the LLE as the values are more comparable between proteins, and the size of the molecules is factored in. By comparing and balancing the  $K_D$ , LE, LELP, and the target similarities, fragments 9595, 0482, and 9612 were selected for further evaluation and characterization.

## 2.5.9 Ranking of Fragments

Table 8. Validated Fragment hits sorted by LE, with various optimization metrics.  $K_D$ -values are in  $\mu\text{M}$ , LE, SILE, LLE, and LELP. Standard deviations for affinities below  $800\ \mu\text{M}$  were below 15%.  $K_D > 800\ \mu\text{M}$  were derived from single measurements.

Protein	Fragment	$K_D$	HA	LE	SILE	SlogP	LLE	LELP
AAK1	0459	18	8	0.81	2.54	1.63	3.12	2.01
AAK1	9595	6.3	10	0.71	2.61	1.61	3.59	2.26
JNK2	0459	81	8	0.70	2.19	1.63	2.46	2.33
AAK1 <sup>a</sup>	7409	5.9	11	0.65	2.55	1.91	3.32	2.93
CAMK1G	0459	190	8	0.64	1.99	1.63	2.08	2.57
DYRK1a	0459	215	8	0.63	1.96	1.63	2.04	2.60
JNK2	9595	41	10	0.60	2.20	1.61	2.78	2.68
JNK3	0459	310	8	0.60	1.88	1.63	1.87	2.72
JNK2	9605	18	11	0.59	2.31	1.91	2.83	3.23
AAK1	0471	90	10	0.55	2.03	2.28	1.76	4.12
JNK2	9612	47	11	0.54	2.11	-0.21	4.54	-0.39
JNK3	9595	127	10	0.53	1.95	1.61	2.28	3.02
CAMK1G <sup>b</sup>	7405	133	10	0.53	1.94	2.28	1.60	4.29
DYRK1a	9595	163	10	0.52	1.90	1.61	2.18	3.10
CAMK1G	9595	203	10	0.51	1.85	1.61	2.08	3.19
IDO1	0459	1440	8	0.49	1.52	1.63	1.21	3.35
AAK1	0474	184	11	0.47	1.82	0.90	2.83	1.94
JNK3 <sup>a</sup>	7409	251	11	0.45	1.75	1.91	1.69	4.25
DYRK1a	0482	530	10	0.45	1.64	0.74	2.53	1.66
JNK3	9612	277	11	0.44	1.73	-0.21	3.77	-0.47
AAK1 <sup>c</sup>	1217	385	11	0.43	1.66	1.46	1.96	3.43
AAK1	1257	810	10	0.42	1.55	1.29	1.81	3.03
DYRK1a	1233	660	11	0.40	1.55	1.54	1.64	3.89
DYRK1a	7419	680	11	0.39	1.54	-0.61	3.78	-1.54
CAMK1G	7419	690	11	0.39	1.54	-0.61	3.77	-1.54
CAMK1G	9605	700	11	0.39	1.54	1.91	1.25	4.85
DYRK1a	0468	1600	10	0.38	1.40	0.64	2.16	1.66
JNK3	9601	3400	9	0.38	1.28	1.08	1.39	2.87
IDO1	1216	394	14	0.33	1.54	-0.65	4.06	-1.96
JNK3	1234	7000	13	0.23	1.00	-0.39	2.55	-1.72

<sup>a</sup> downward stepping baseline throughout measurements, the fitting becomes unreliable.

<sup>b</sup> fragment concentration reduced to 2 mM. <sup>c</sup> Fragment concentration reduced to 2.5 mM.

Fragment 9595 was evaluated with the AAK1 as the fragment shows the best affinity. At the same time, other kinases bind significantly weaker, and the fragment has not been reported as a kinase inhibitor by itself. 5*H*-pyrolo[3,2-*d*]-pyrimidines have been used as synthetic building blocks, but no affinity data is known so far<sup>196-198</sup>. In addition, the proton vicinal to the chlorine exhibits the strongest saturation transfer signal in the NMR spectrum (Figure 28 A). Fragment 0482 was evaluated with DYRK1a as the chlorine version (0468), and the bromine version were both conspicuous in the screening as well as in the ITC.

#### 2.5.10 Structure-Affinity Relationship of 9595 with AAK1

Again, the STD spectrum shows the most potent signal next to the halogen atom (Appendix B 12). Fragment 9612 was evaluated with JNK2 and JNK3. The main reason is the remarkable preference for JNK2 over JNK3, with a nearly six-fold decrease in affinity from JNK2 to JNK3. Additionally, the fragment showed no hits with the remaining three kinases in the NMR. Although the LE of fragment 9612 is mediocre compared to other fragments ( $0.44 < LE > 0.54$ ), the high solubility of the fragment ( $SlogP = -0.21$ ) is a powerful argument to include it in a more in-depth examination.

#### 2.5.10 Structure-Affinity Relationship of 9595 with AAK1

The fragment was measured additional times to a quadruplicate as a first step. With increasing protein concentration, higher c-values were achieved, and consequently, interpretation of the thermodynamic parameters of the binding was possible. A c-value of 45 could be achieved with a protein concentration of 300  $\mu$ M. Thus, stoichiometry could be included in the fit of the curve, and thus, reliable thermodynamic characteristics can be derived. At low c-values, in the validation process, the stoichiometry had to be fixed to one; the additional experiments showed a 1:1 complex formation with a fitted  $n = 0.91$  and  $n = 0.89$ , respectively. The deviation from a perfect one-to-one ratio is caused by the error in protein concentration estimation. At these high concentrations pipetting errors are potentiated by the 40-fold dilution to achieve lambert-beer conditions for protein estimation. In addition, small impurities in the compound stock can alter the stoichiometry further. The purity of 9595 was given as 95% by the supplier. The thermodynamic signature of the fragment is consistent over all four measurements, with a minimum standard deviation of the Gibbs free energy, although protein concentration varied from 111  $\mu$ M to 300  $\mu$ M. An enthalpy-driven binding mode can be deduced as the enthalpy generated exceeds the entropic part five-fold (Figure 45). The affinity was determined to be  $K_D = 6.3 \mu$ M with a standard deviation of 0.7  $\mu$ M. Due to the lack of alternative conformations in the fragment, fewer degrees of freedom are lost at binding. In addition, there are four heteroatoms with the potential for polar interactions. With this information, a set of analogs were acquired to test the influence of each polar interaction of the fragment. The commercial availability of pyrrolopyrimidine analogs was the main reason for the small sample size in the set.

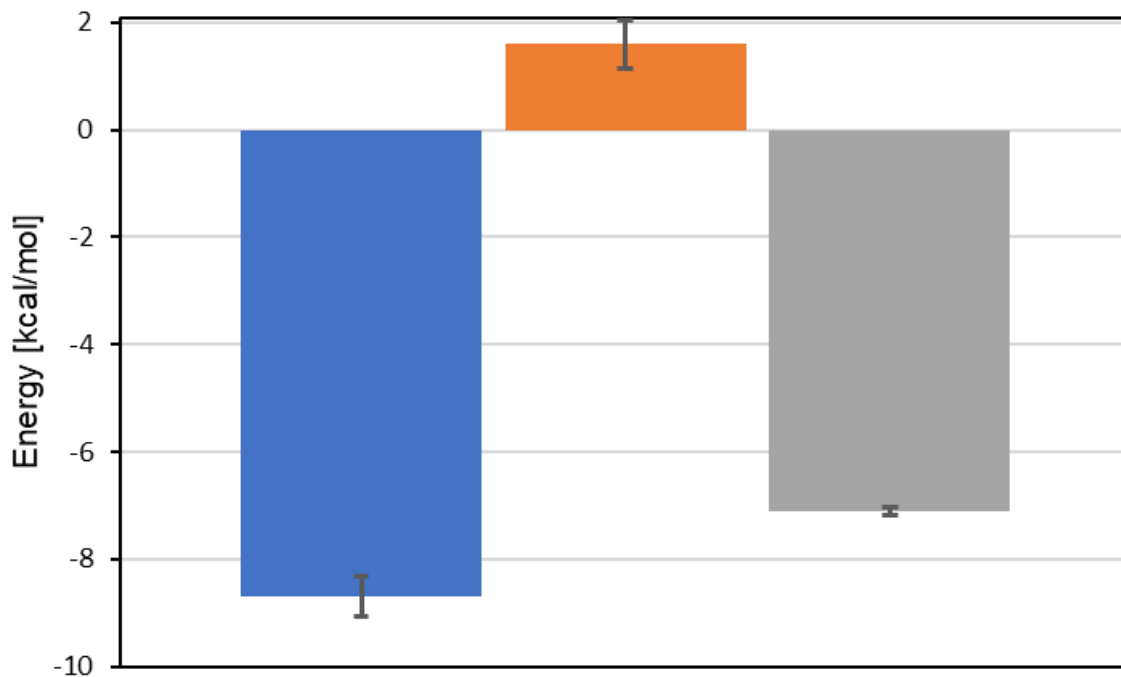


Figure 45. Averaged thermodynamic parameters for the binding of 9595 to AAK1 over four measurements.  $\Delta H^0$  in blue,  $-T\Delta S^0$  in orange, and  $\Delta G^0$  in grey in kcal/mol. The standard deviations of each parameter are indicated as error bars. The binding motif is notably enthalpy-driven.

As Figure 46 and Table 9 illustrate, replacing either pyrimidine nitrogens with carbons reduced the affinity to 50-60  $\mu\text{M}$  (A1, A2). In contrast, completely removing the chlorine atom abolished the affinity with a  $K_D > 1000 \mu\text{M}$ . A roughly 200-fold decrease of affinity from A2 to A3, proving the halogen is the major contributor to the binding affinity. However, this is no evidence for a classical halogen bond per se, as the chlorine might engage in multipolar interactions with multiple residues. The replacement of the pyrrole nitrogen to oxygen (A4) resulted in a 50-fold affinity decrease, although the compound bound considerably stronger than A3, with a  $K_D = 300 \mu\text{M}$ . The interaction pattern of the pyrrole nitrogen could not be studied in detail due to the small set of compounds. A simple hydrogen bond could be possible, but the tuning effect of the chlorine to the nitrogen and vice versa could play an essential role in the binding motif. The effects of the pyrimidine nitrogens were shown to be significant. However, to a lesser extent, as A5 showed, thiophene instead of a furan (A5 versus A4) could restore most of the binding affinity, being only four times weaker than fragment 9595 with a  $K_D = 26 \mu\text{M}$  (Figure 46).

## 2.5.10 Structure-Affinity Relationship of 9595 with AAK1

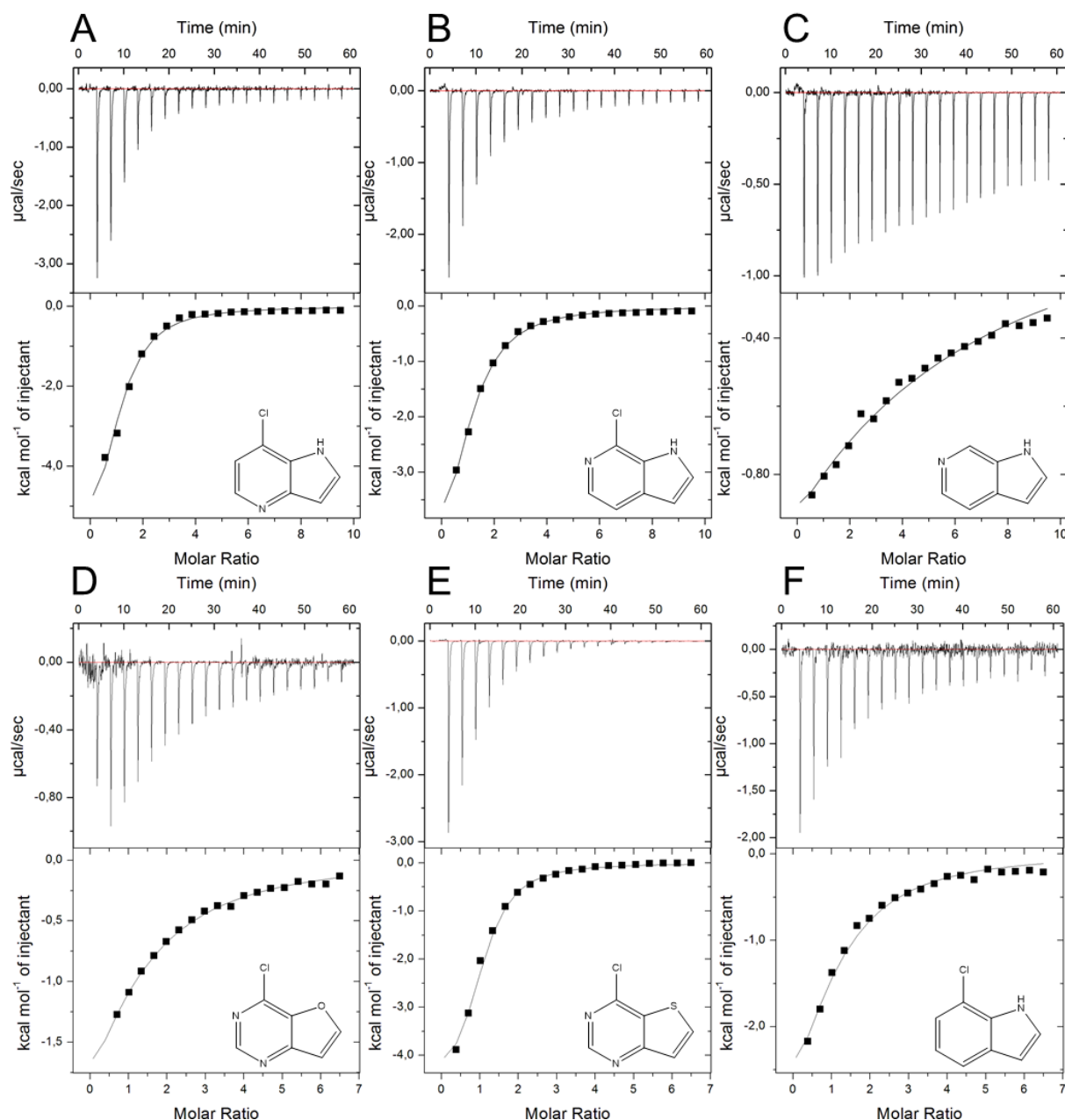


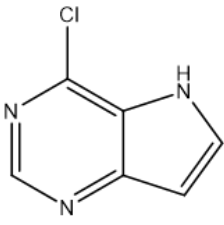
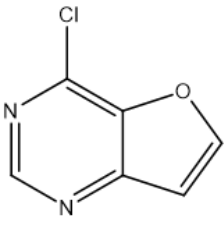
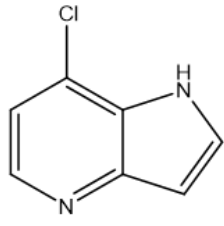
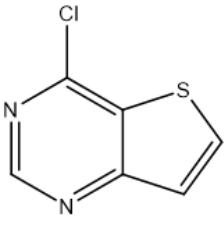
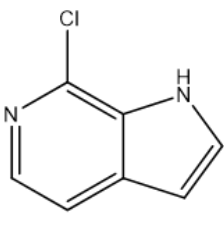
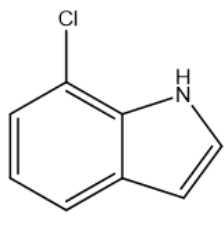
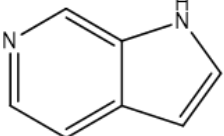
Figure 46. ITC data for SAR-compounds of 9595 with AAK1, from **A-F** the compounds A1-A6 are depicted.

A standard hinge binding motif, as seen with other pyrrolopyrimidines, especially the [2,3-*d*] version with the pyrrole nitrogen adjacent to the pyrimidine nitrogen, is not present in fragment 9595. Most ATP competitive inhibitors utilize two nitrogens facing the hinge region of kinases. In theory, a hinge binding motif with the chlorine and the pyrrole nitrogen could be envisioned, where the chlorine would engage in an XB to the hinge region. However, the fragment was measured with all five kinases, and we could observe that the binding mode is likely specific within the five tested kinases (Table 8).



The fragment had a six-fold decreased affinity to JNK2 and a 20 to 34-fold decreased affinity to the JNK3, DYRK1a, and CAMK1G. The measured preferences might indicate different binding modes with the five proteins or a similar binding mode with less optimal interaction geometry.

Table 9. Name, structure, affinity, and standard deviation of 9595 analogs with AAK1.

Name	Structure	$K_D$ [ $\mu\text{M}$ ]	$\sigma$ [ $\mu\text{M}$ ]	Name	Structure	$K_D$ [ $\mu\text{M}$ ]	$\sigma$ [ $\mu\text{M}$ ]
9595		6.3	0.7	A4		300	N/A
A1		49	2.9	A5		26	7.3
A2		61	11	A6		170	N/A
A3		1900	N/A				

In an ideal setting, a crystal structure would be generated. As the group of Professor S. Knapp in Frankfurt nor we could obtain a structure with AAK1 and fragment 9595. I assume the protein crystallizes with reasonably large inhibitors by strengthening a specific protein conformation. Although the fragment binds with strength comparable to substances with a long optimization history, the compound might be too small to interact with specific residues to stabilize the protein in a conformation that can crystallize. It might be possible to crystallize the AAK1 with an already elucidated compound and exchange the compound within this structure. On the other hand, the fragment is readily available, and further synthetic expansion could lead to compounds amenable to protein crystallography.

## 2.5.11 Structure-Affinity Relationship of 0482 with DYRK1a

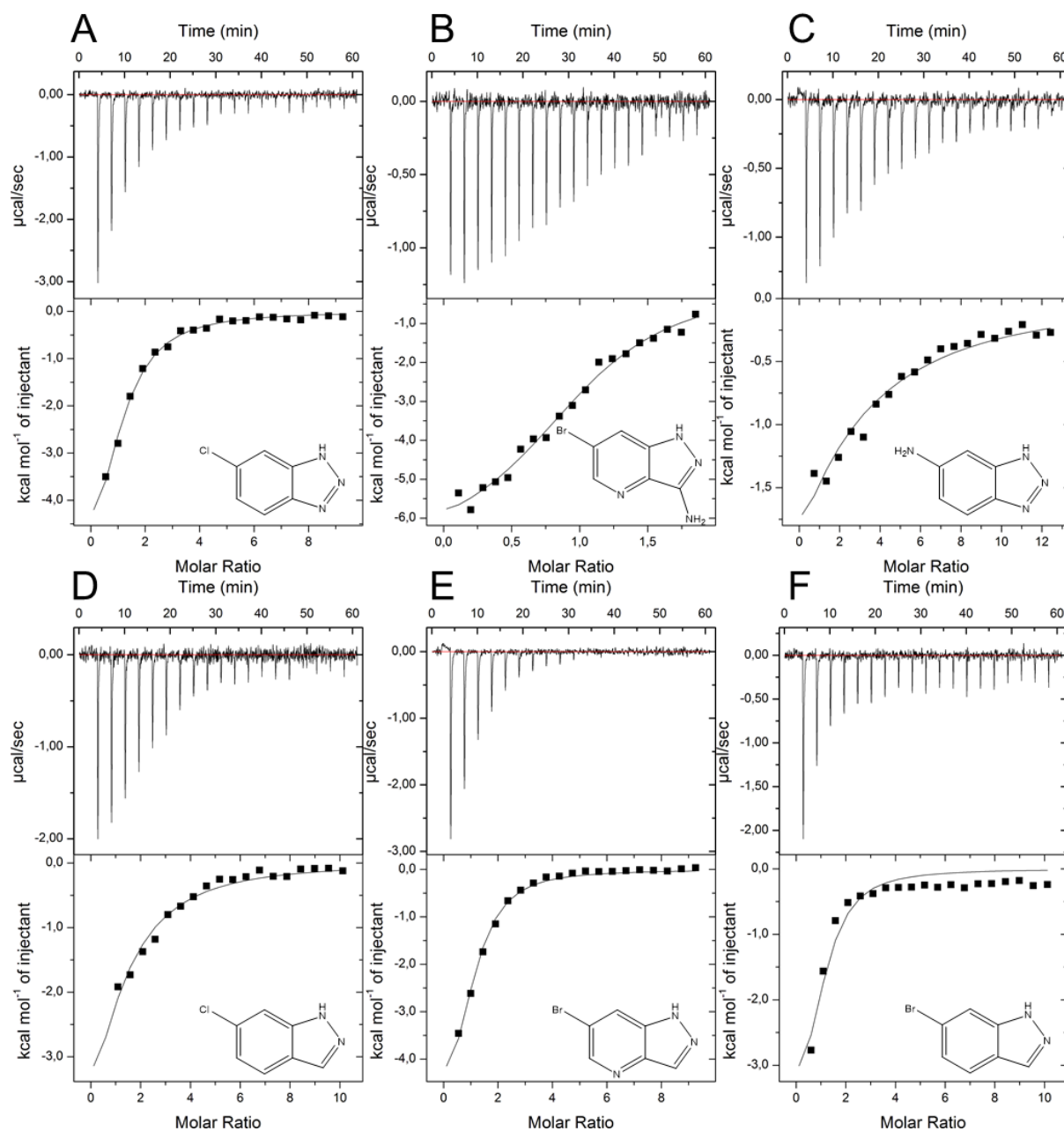
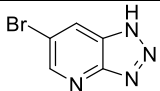
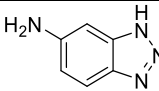
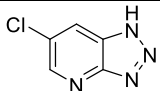
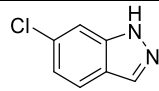
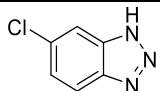
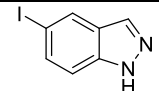
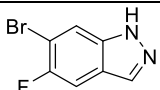
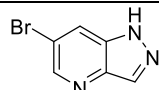
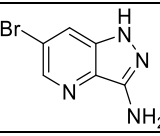
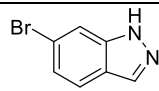
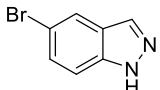
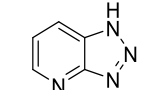


Figure 47. ITC data for SAR-compounds of 0482 with DYRK1a, **A**: D1, **B**: D3, **C**: D5, **D**: D6, **E**: D8, **F**: D9.

Compared to the AAK1, a crystallization protocol for the DYRK1a 119, 125 (see section 2.6.2) was established. Therefore, a reasonable chance was provided to elucidate the binding mode of fragments with a strong affinity. The measured  $K_D$ -values for the five verified DYRK1a binders were in the three-digit micromolar range. Not the best binding fragment was prioritized for further investigation. Fragments 0482 and 0468 are nearly identical with bromine instead of chlorine and an increased affinity of the bromine over

the chlorine. This finding leads to the hypothesis that halogen bonding was present in the binding mode. Therefore, we acquired a set of analogs of 0482 and tested its affinity to DYRK1a. Unfortunately, the most exciting compounds were either insoluble (iodine analogs) or were not commercially available. The derivable SARs was hard to comprehend, and building a hypothesis of the binding mode from the affinity data proved to be impractical. The seeming contradiction could later be elicited by the crystal structure of 0482 (see section 2.6.3). Comparing the numbers and the thermograms in Figure 47 and Table 10, we can derive that the halogen is vital to the binding mode. Fragment 0468 binds so weakly that no reliable affinity was determined, whereas 0482 binds with approximately 530  $\mu\text{M}$  (Table 10). In addition, comparing the SAR pairs D1 to D5 and D6 to D9, we see a substantial increase in affinity from chlorine compared to an amine and an even better affinity for bromine over chlorine. Compound D10 was measured, but no heat emissions at all were visible, which could be used for any fit; therefore, the affinity towards DYRK1a has to be even worse than 0468. Although the compound features a known hinge binding motif, no binding at all could be observed. This is another indicator of a halogen bond in the initial fragment, and it is even more interesting that the hinge binding motif might not be the key interaction of the binding mode. The SAR is less intuitive than the relationships of the 9595 analogs. Suppose D1 is compared to the fragments 0468 and D5. The replacement of the pyridine nitrogen with a carbon seems to increase the affinity from undeterminable to 80  $\mu\text{M}$ , and the replacement of the chlorine with an amine (D5  $K_D=430\ \mu\text{M}$ ) results in roughly the same affinity as 0482. Apparently, the triazolopyridine scaffold hinders the binding. This assumption is contradicted as compounds D8 and D9 bind even better than any of the previously discussed compounds. Both compounds feature bromine, which binds better than D1 and D6 ( $K_D=180\ \mu\text{M}$ ). In general, removing heteroatoms seemed to increase the binding affinity. The sole exception for this behavior was D3. For this compound, an affinity was determined at  $K_D=47\ \mu\text{M}$  resulting in a roughly 10 times higher affinity than the initial hit of 0482. Furthermore, the amine at the 3-position could be a possible fragment growing vector for this compound. However, aminopyrazoles have been described as possible hinge binding motifs, and a change in the binding mode should be evaluated for this compound and the potential following amide compounds<sup>120</sup>.

Table 10. Name, structure, affinity, and standard deviation of 0482 analogs with DYRK1a. D2, D4, and D7 were insoluble.

Compound	Structure	$K_D$ [ $\mu\text{M}$ ]	$\sigma$ [ $\mu\text{M}$ ]	Compound	Structure	$K_D$ [ $\mu\text{M}$ ]	$\sigma$ [ $\mu\text{M}$ ]
482		530	43	D5		430	200
468		>1000	-	D6		180	22
D1		80	9.9	D7		N/A	-
D2		N/A	-	D8		43	2.7
D3		47	28	D9		38	9.6
D4		N/A	-	D10		>>1000	-

The set of analogs could paint no clear picture, most likely, due to the possibility of multiple binding modes of these fragments. In summary, a halogen seems to be a valuable part of these binding modes. However, more compounds with varying substituents as replacements for the bromine would be needed to analyze the potential of the halogen. The changing effect of nitrogen in the pyrimidine and triazole scaffold is not entirely comprehensible without a crystal structure.

## 2.5.12 Structure-Affinity Relationship of 9612 with JNK3 and JNK2

An advantage of the HEFLib is the integrated sets of compounds for validation and comparison. Fragment 9612 belongs to one of these small sets of integrated SAR compounds in the library. The group of satellite compounds is extensive for nicotinic acids, partly due to the excellent availability of nicotinic acid analogs. As previously mentioned, the fragment had the best preference between two almost identical kinases regarding the binding site. The affinity of 9612 with JNK2 was in the two-digit micromolar range and could be accurately defined with a  $c$ -value of two and a protein saturation of 94%. Eleven analogs were either bought or repurposed from the HEFLib. Nearly every fragment showed no activity with either JNK2 or JNK3 (Table 11). We determined from these negative results that neither the amino group, the carboxyl group, or the halogen could change the position, as every change resulted in a complete loss of binding. Even small changes, such as replacing the amino group with a hydroxyl group, resulted in no detectable signal in the ITC. This lack of distinguishable information made it difficult to gain further insight into the binding mode. The only compounds showing measurable affinities were S1 and S3 (Figure 48). The alcohol showed significantly decreased affinity towards both targets, but only with JNK2 was the extent quantifiable ( $K_D=310\ \mu\text{M}$ ).

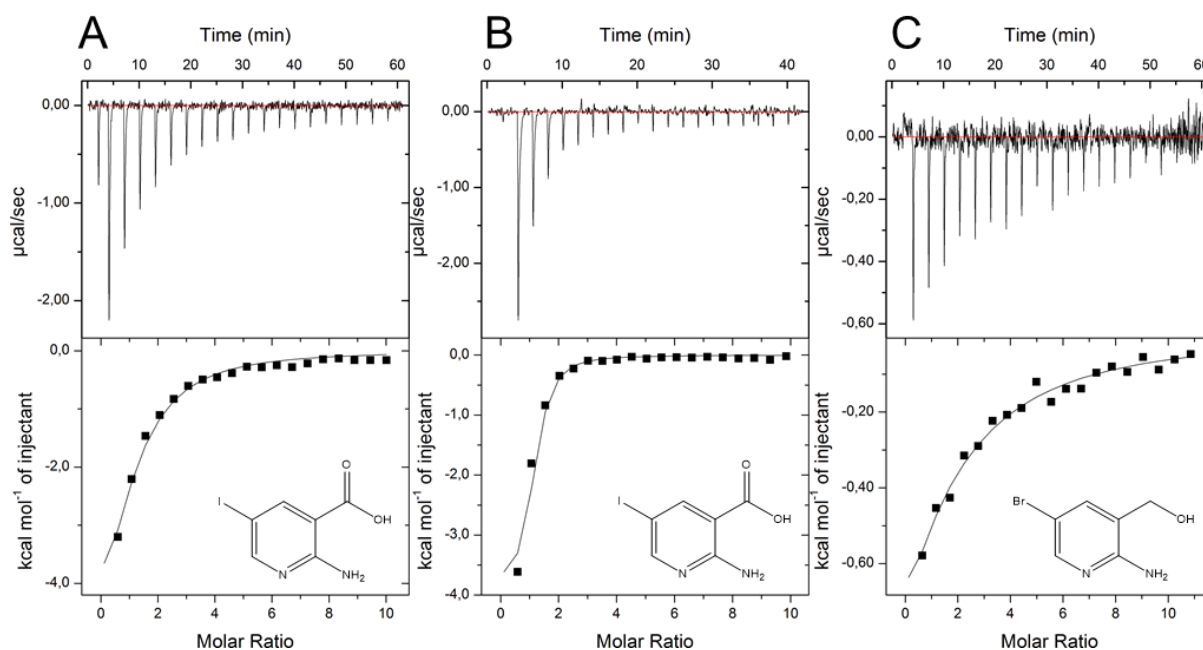
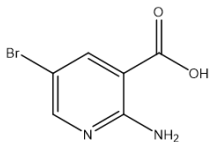
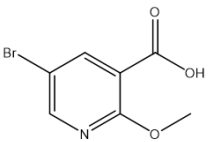
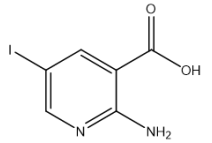
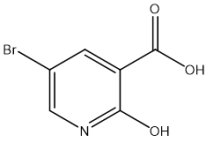
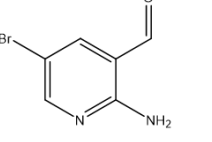
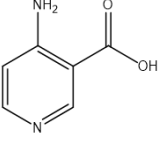
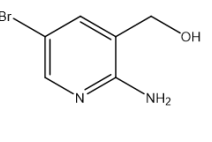
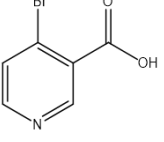
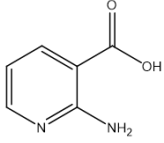
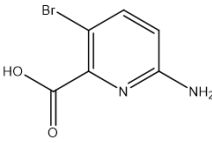
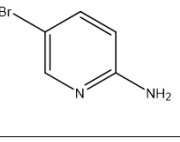
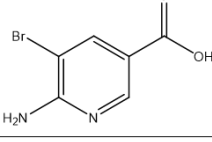
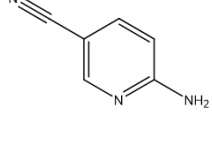


Figure 48. ITC data for SAR compound of 9612 with **A**: JNK3 with S1, **B**: JNK2 with S1, **C**: JNK2 with S3.

It would have been interesting to know whether the aldehyde would have had an affinity between the acid and the alcohol, which was the hypothesis. The most promising compound was S1, with iodine instead of bromine. The affinities increased to 11  $\mu\text{M}$  towards JNK2 and 90  $\mu\text{M}$  towards JNK3. As can be seen, the preference for JNK2 over JNK3 increased from six-fold to eight-fold. The increase in affinity by a larger halogen could point to a halogen bond being present but again is no definitive evidence. The larger electron shell and the greater surface might interact by dispersion effects or other weak interactions with additional residues. However, this was a promising result, and the crystal structure for 9612 and S1 could be obtained with JNK3 (see section 2.6.4), giving insight into the binding mode and the halogen bond involved.

Table 11. Name, structure, affinity, and standard deviation of 9612 analogs with JNK2 and JNK3.

Name	Structure	$K_D$ [ $\mu\text{M}$ ] JNK2	$\sigma$ [ $\mu\text{M}$ ]	$K_D$ [ $\mu\text{M}$ ] JNK3	$\sigma$ [ $\mu\text{M}$ ]	Name	Structure	$K_D$ [ $\mu\text{M}$ ] JNK2 and JNK3
9612		46	6.7	275	33	7396		>1000
S1		11	1.3	90	5.5	0523		>1000
S2		Insoluble		Insoluble		4ANA		>1000
0475		310	55	>1000		7403		>1000
2ANA		>1000		>1000		7395		>1000
1232		>1000		>1000		7655		>1000
S5		>1000		>1000				

## 2.5.13 Comparison of Gatekeeper mutants of JNK3 with 9612

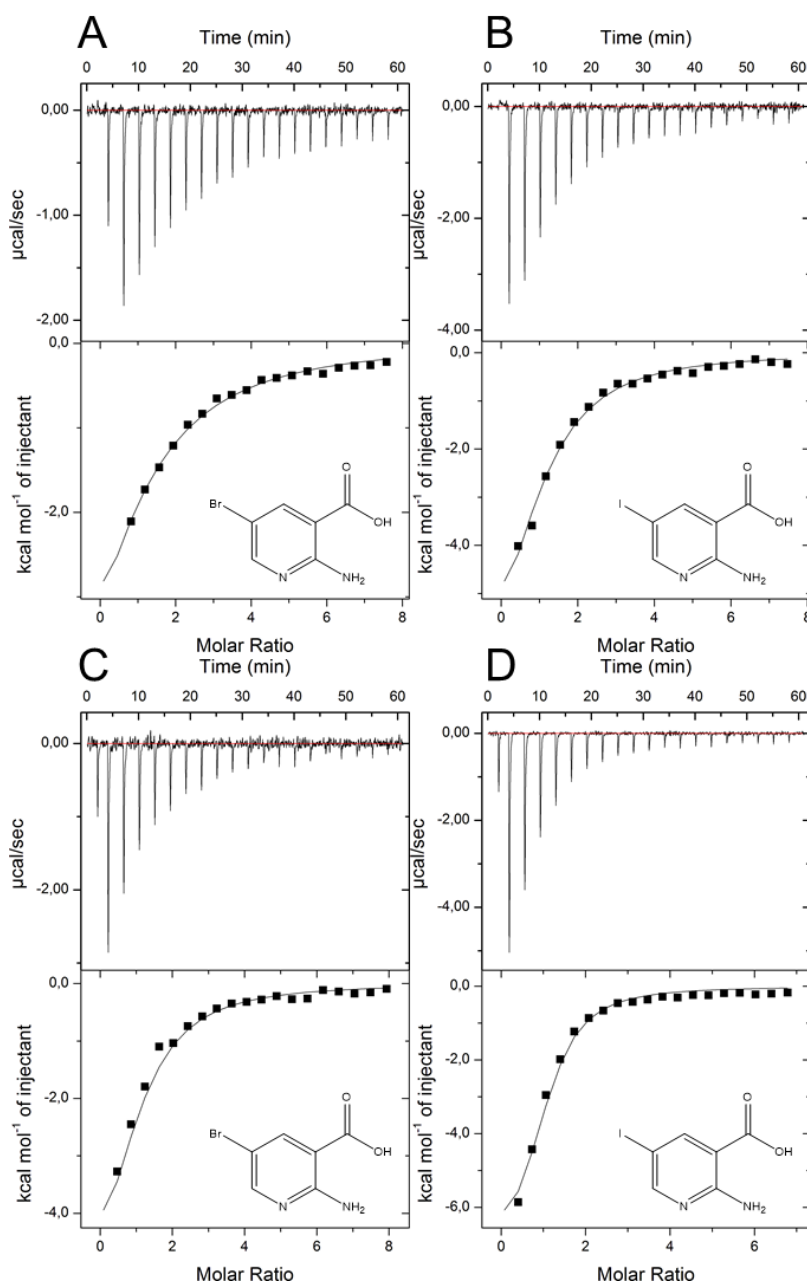


Figure 49. Exemplified ITC data of JNK3M115A with **A**: 9612 and **B**: S1, JNK3M115L with **C**: 9612 and **D**: S1.

Due to the high similarity of JNK2 and JNK3, we constructed two mutants of JNK3. The first mutant M115L corresponds to a pseudo-JNK2 as the only difference in the binding pocket and the surrounding sequence is this one particular amino acid. In the JNK3, methionine 115 is positioned right behind the gatekeeper residue (M146). As previously shown<sup>63</sup>, both can form a three-center chalcogen bond with a ligand and each other.



We assumed that the polar interaction of these two amino acids might interfere with the binding of the fragment, as JNK2 features a leucine (L77) in the corresponding position behind the gatekeeper (M108). The second mutant was a relatively conservative approach to test whether the space in the hydrophobic pocket behind the gatekeeper impacted binding. Figure 49 shows ITC measurements of JNK3 M115A (A and B) and JNK3 M115L (C and D) with the fragments 9612 (A and C) and S1 (B and D).

The affinity of fragment 9612 (2-amino-5-bromonicotinic acid) has a high affinity towards JNK2 at 46  $\mu\text{M}$  and a much lower affinity towards JNK3 at 275  $\mu\text{M}$ . The pseudo-JNK2 variant M115L increases the affinity close to the level of JNK2 with 85  $\mu\text{M}$ . The M115A variant increases the affinity, but only with a minimal effect (249  $\mu\text{M}$ ). The SAR compound S1 (2-amino-5-iodonicotinic acid) shows a similar pattern, although the initial dissociation constants were generally lower with  $K_D(\text{JNK3}) = 90 \mu\text{M}$  and  $K_D(\text{JNK2}) = 11 \mu\text{M}$  (Table 12). The M115L mutant showed a  $K_D = 34 \mu\text{M}$ , and M115A has a binding constant of 108  $\mu\text{M}$ . Proving the strong influence of the methionine 115 (JNK3) and leucine 77 (JNK2), respectively, on the binding mode without direct contact with the fragment.

Table 12. Affinities of the fragments 9612 (Bromo) compared to S1 (Iodo) with the JNK3 mutants and wild-type proteins. N denotes the number of measurements.

	Compound	$K_D$ [ $\mu\text{M}$ ]	% St.dev.	N
JNK3	Bromo	275	11.8	2
	Iodo	90	6.1	3
JNK2	Bromo	46	14.5	2
	Iodo	11	11.9	2
M115L	Bromo	85	7.7	3
	Iodo	34	4.9	3
M115A	Bromo	249	0.0	2
	Iodo	108	8.0	2

On the other hand, the data suggest that the space behind the gatekeeper (M146) is not the affinity determining factor, as the alanine mutant is less strongly bonded. Therefore, we assume the M115 in JNK3 might engage in the previously mentioned chalcogen bond<sup>63</sup> with M146 and restrict the gatekeeper methionine's interaction geometries with fragment 9612. The leucine sidechain in the JNK2 cannot form polar interactions with the gatekeeper M108, thus granting more conformational freedom to the gatekeeper methionine. As the binding mode of 9612 must be restricted by other interactions, the methionine might form an interaction with the fragment in an induced-fit-like manner. The energy balance for this interaction might be more favorable with the JNK2 as the methionine can interact less with the surrounding amino acids. The pseudo-JNK2 (JNK3 M115L) does not bind to both fragments with the same affinity as the natural JNK2; minor differences in the binding site geometry due to scaffolding effects could be responsible for the remaining difference in dissociation constants.

### 2.6. Crystallization

To accomplish the necessary binding mode elucidation of validated fragments, the crystallization of AAK1, DYRK1a, JNK2, and JNK3 was undertaken. For JNK3 and DYRK1a, extensive data for various crystallization conditions have been published. Only a few crystal structures of JNK2 and AAK1 have been published so far <sup>111, 164</sup>. The crystal structures of AAK1 have exclusively been produced by the group of S. Knapp in Frankfurt. After extensive trials, it was impossible to produce a structure with one of the validated fragments. We tried an AAK1 screening for co-crystallization conditions ourselves, but the results were the same.

There are three crystal structures of JNK2; two are linked to a publication with detailed information about the crystallization conditions<sup>128, 133</sup>. Both publications stress the importance of mutated constructs to facilitate the crystallization of JNK2<sup>128, 133</sup>. Although the correct protein construct was used, no crystals could be grown with the reported conditions or commercially available screens.

#### 2.6.1 DYRK1a crystallization

The crystallization of DYRK1a was established based on a protocol of Knapp et al. <sup>120</sup>, with the inhibitor KuRoM118 (2-cyclopentyl-7-iodo-1*H*-indole-3-carbonitrile), which the group of Professor Kunick in Braunschweig provided <sup>15</sup>. The compound was used to test whether the soaking conditions were sufficient to reach occupancy without damaging the crystals. The fragments have an insufficient solubility without any DMSO, but the addition of 5% allows for soaking concentrations of up to 10 mM for most fragments. The crystals co-crystallized with AMP-PCP are typically in a rhombohedral shape with 0.2 mm in the longest dimension (Figure 50), with occasional larger crystals up to 0.5 mm. The soaking can be performed for 24 or up to 36 hours; longer durations in 5% DMSO tended to dissolve the crystals. Soaked crystals were sent to the Swiss light source (SLS) of the Paul Scherrer Institute in Switzerland and measured with the help of Georg Zocher<sup>199</sup>.

### 2.6.1 DYRK1a crystallization



Figure 50. Image of a typical drop of DYRK1a AMP-PCP cocrystals on the left. A crystal was sent to the SLS for diffraction on the right. Crystals are typically between 0.2 and 0.3 mm in the longest dimension, with occasional larger ones up to 0.5 mm.

A DYRK1a crystal with the compound KuRoM118 was solved to a resolution of 2.22 Å and a  $R_{\text{work}}/R_{\text{free}}$  of 0.196/0.242. Another DYRK1a crystal with the fragment 0482 could be solved to a similar resolution (2.3 Å) and an  $R_{\text{work}}/R_{\text{free}}$  of 0.209/0.242. Further statistics of the structures are shown in Table 13. Additional crystals with DYRK1a binding fragments were soaked (i.e., 9595, 0459, and 7419). The crystals showed resolutions of around 3 Å, and as a result, the electron density was sufficient to see the exchange of AMP-PCP, but no binding mode of the fragments could be deduced.

Table 13. Overview of data collection and refinement statistics of the DYRK1a crystals with KuRoM118 and 0482.

Data Collection and Refinement	KuRom118	B02290482
Beamline	SLS-PXIII Pilatus 2M-F	SLS-PXIII Pilatus 2M-F
Space group	C 1 2 1	C 1 2 1
Cell dimensions a/b/c (Å)	244.24 / 64.47 / 147.33	245.09 / 64.69 / 147.38
Cell angles $\alpha/\beta/\gamma$ (°)	90 / 115.58 / 90	90 / 115.57 / 90
Wavelength (Å)	0.9998	0.9999
Resolution (Å)	50-2.22 (2.35-2.22)	50-2.3 (2.44-2.3)
CC 1/2 (%)	99.8 (42.5)	99.8 (53.1)
Completeness (%)	99.6 (98.6)	99.8 (99.5)
Redundancy	4.54 (4.42)	6.81 (6.6)
I/ $\sigma$ (I)	9.66 (1.02)	12.48 (1.25)
R <sub>meas</sub> (%)	11.9 (150.5)	13.3 (157.8)
Wilson B-factor (Å <sup>2</sup> )	52.68	53.92
Refinement	Phenix v1.19.2	Phenix v1.19.2
Resolution (Å)	49.09 (2.22)	48.6- (2.3)
R <sub>work</sub> /R <sub>free</sub>	0.196 / 0.242	0.209 / 0.242
R <sub>free</sub> test set	1536 (1.5%)	1487 (1.6%)
Number of atoms	12010	12099
Protein	1387	1386
Water	613	641
Other molecules	39	33
B-factors (Å <sup>2</sup> )	59.74	59.5
Protein	59.68	59.34
Water	57.65	58.39
Other	67.6	68.96
r.m.s.d		
Bond length (Å)	0.011	0.01
Bond angles (°)	1.197	1.093
Ramachandran (%)		
favorable/allowed/outlier	95.98 / 4.02 / 0	95.46 / 4.47 / 0.07
Rotamer outlier (%)	2.14	0.68
All-atom clash score	3.7	2.69

## 2.6.2 KuRoM118 crystal

The first successful crystallization of this study was performed with KuRoM118. The asymmetric unit shows four copies of DYRK1a in Figure 51 A, an overview of the structure with the protein in green cartoon representation, the compound in yellow sticks, and all residues around 5 Å of the compound are represented in green sticks. The structure was solved from residue V135 to K481 without unresolved loops. Two loops were found in different conformations. The amino acids from H213 to I221 in chains A and C are ordered in a short loop followed by a small  $\alpha$ -helix with one whole turn. The chains B and D form one larger loop without the  $\alpha$ -helix. Only chains A and D form crystal contacts in this region with different parts of different protein chains. The second loop with two observed conformations starts at K406 and ends with Y415. Only chain C occupies an entirely different geometry (Figure 51 B). The loop engages in crystal contacts in all four chains, but the contact partners differ in every chain, hence the changing loop conformation. The electron density of the compound KuRoM118 was found in all four chains with high certainty in the ATP binding pocket. The omit-maps generated for all four chains were evident at  $3\sigma$  shown in Figure 51 C.

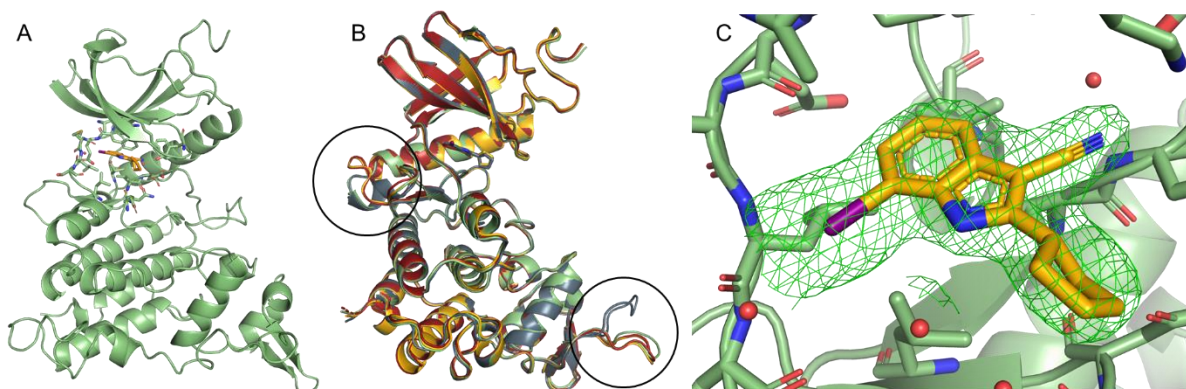


Figure 51. **A:** A green cartoon view of the complete DYRK1a with the bound compound (KuRoM118) in yellow, residues within 5 Å are shown as sticks. **B:** Alignment of all four chains of the DYRK1a, there are two loops with variable conformations mainly due to crystal contacts (black circles). **C:** Detailed view of the ATP-binding site (chain A) with an omit map contoured at  $3\sigma$  of KuRoM118.

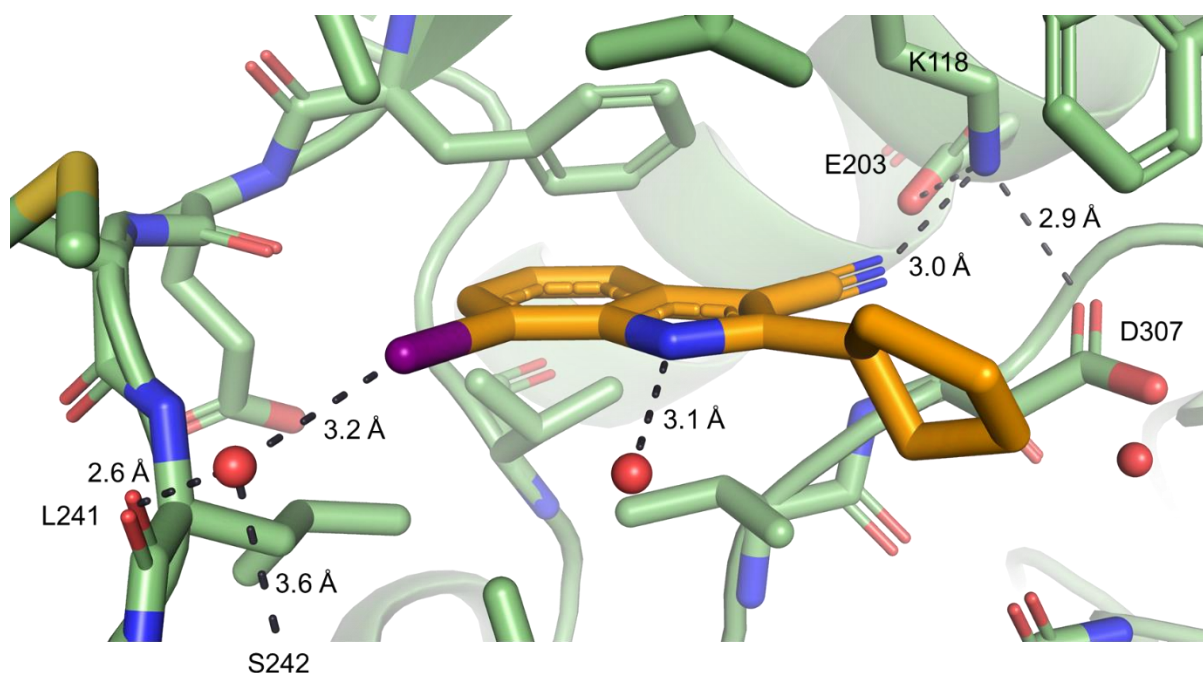


Figure 52. Detailed view of the DYRK1a (green) binding mode of KuRoM118 (yellow), the main interactions of the compound are the numerous CH- $\pi$  and the strong hydrogen bond of the nitrile to K188. The indole nitrogen forms a hydrogen bond with a water molecule, and the iodine forms a halogen bond with water at a distance of 3.2 Å and an angle of 167.8°.

The indole scaffold of the compound forms various CH- $\pi$ -interactions, namely with the residues V173, V306, and L294. The gatekeeper F238 interacts in a T-shape  $\pi$ - $\pi$ -interaction (Figure 52), but this conformation makes the hydrophobic pocket inaccessible. The cyclopentane ring is surrounded by water, and no close contacts (<4.0 Å) are observable. The primary polar interaction is the hydrogen bond between the nitrile nitrogen and the sidechain of K118 (Figure 52). This lysine is highly coordinated to two residues (E203 and D307). No polar contacts with the hinge region are present. A water molecule is bound to the indole nitrogen at a distance of 3.1 Å. In the best-resolved chain, a second water molecule lies in front of the iodine atom and is coordinated by two additional residues (L241 and S242). The water molecule is in a near-perfect geometry for an XB to the iodine with a distance of 3.2 Å and a C-I-O angle of 167.8°. The halogen bond in this instance would explain the gain of affinity reported for similar compound series reported by Meine et al.<sup>200</sup>. The affinity towards DYRK1a of these 7-halo-2-phenyl-1*H*-indole-3-carbonitriles increases with the size of the halogen, from chlorine (IC<sub>50</sub>=40 nM) over bromine (IC<sub>50</sub>=25 nM) to iodine (IC<sub>50</sub>=10 nM), compounds with hydrogen (IC<sub>50</sub>=400 nM), or a methyl group (35% inhibition at 10  $\mu$ M) instead of the halogen feature even weaker affinities<sup>15, 65, 200</sup>.

### 2.6.3 Fragment 0482 crystal

The present compound was measured to an inhibitory constant of  $IC_{50}=70$  nM, owing to the cyclopentane ring instead of a phenyl ring<sup>15</sup>. The same water was observed in another structure (PDB code 4YLJ) with a similar iodine-bearing compound at nearly the same position (Figure 53)<sup>65</sup>. As the water could only be observed in the best-resolved chain, the presence of this water molecule in the remaining three chains is uncertain. Without the water, the iodine forms no close contact, and as a result, the affinity increase through heavier halides might be explained by desolvation or dispersion effects.

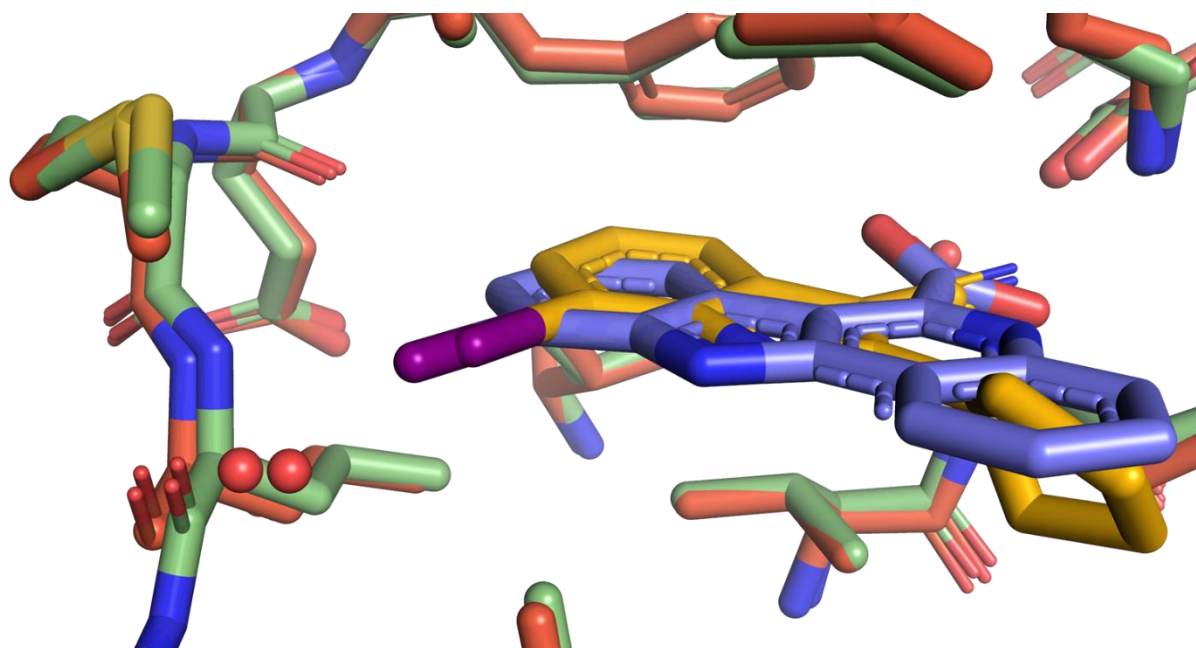


Figure 53. The binding mode of KuRoM118 (yellow) compared to the iodo-indolo-quinolinic acid (4YLJ) (blue) in their respective binding mode. The iodine in both inhibitors points to a water molecule directly in front of the  $\sigma$ -hole. Both geometries lie within the van der Waals radii of the atoms (distances of 3.2 Å versus 3.0 Å) and share a similar angle of 167.8° and 168.5°.

### 2.6.3 Fragment 0482 crystal

The crystal solved with fragment 0482 showed the same space group and overall similar statistics as the KuRoM118 crystal (Table 13). With 2.3 Å, the resolution was only marginally worse. Again, no unresolved loops were present, but two fragment instances were found in each chain (Figure 54 A)<sup>199</sup>. The loop near the hinge region (H213 to I221) was found in only one conformation in all four chains. In this crystal, the conformation without the  $\alpha$ -helix was present. In both crystals, chain C performs no crystal contact at this part, and the loop conformation is present in the crystal soaked



with fragment 0482, whereas the same chain exerts the helix-loop conformation in the crystal soaked with KuRoM118. This finding leads to the assumption that both conformations are of similar thermodynamic stability, as crystals grew at 4 °C over weeks. The loop from K406 to Y415 conforms to the other crystal, again with a second conformation only in chain C. The two binding sites of the fragment were validated by the generation of omit-maps contoured at  $3\sigma$ , with a well-defined electron density in all four chains (Figure 54 C and D)<sup>199</sup>. Of the two binding sites found in the crystal, one fragment instance was found in the ATP-binding pocket next to the hinge region. The second fragment was found in a small crevice between a loop and an  $\alpha$ -helix near the c-terminus of the kinase domain (Figure 54 A).

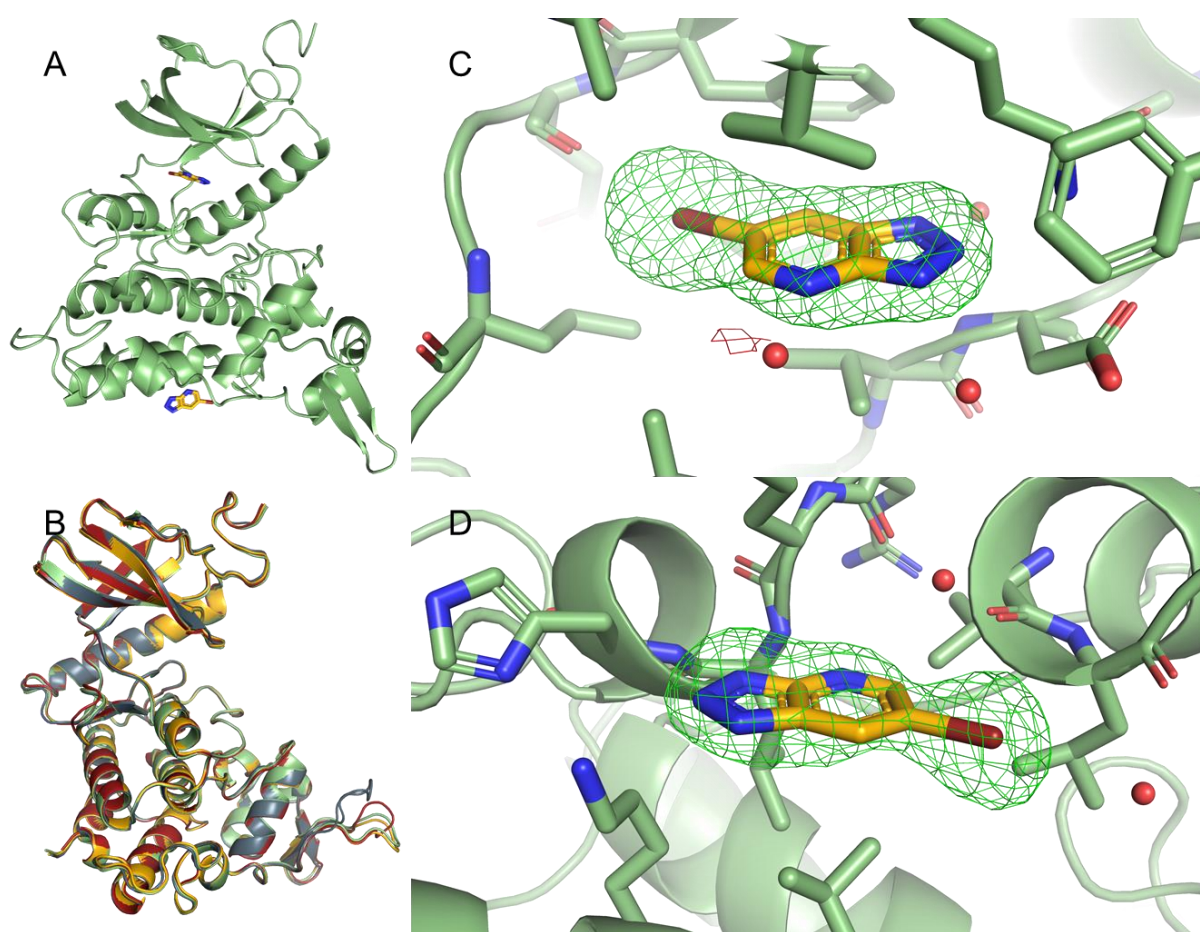


Figure 54. **A:** Overview of DYRK1a in green cartoon representation with the fragment 0482 in yellow sticks. **B:** Alignment of all four chains of the DYRK1a ASU, with the already mentioned loop in two conformations **C:** Detailed view of the ATP-binding site (chain A) with an omit-map contoured at  $3\sigma$  of fragment 0482. **D:** Detailed view of the second binding site (chain A) with an omit-map contoured at  $3\sigma$  of fragment 0482.

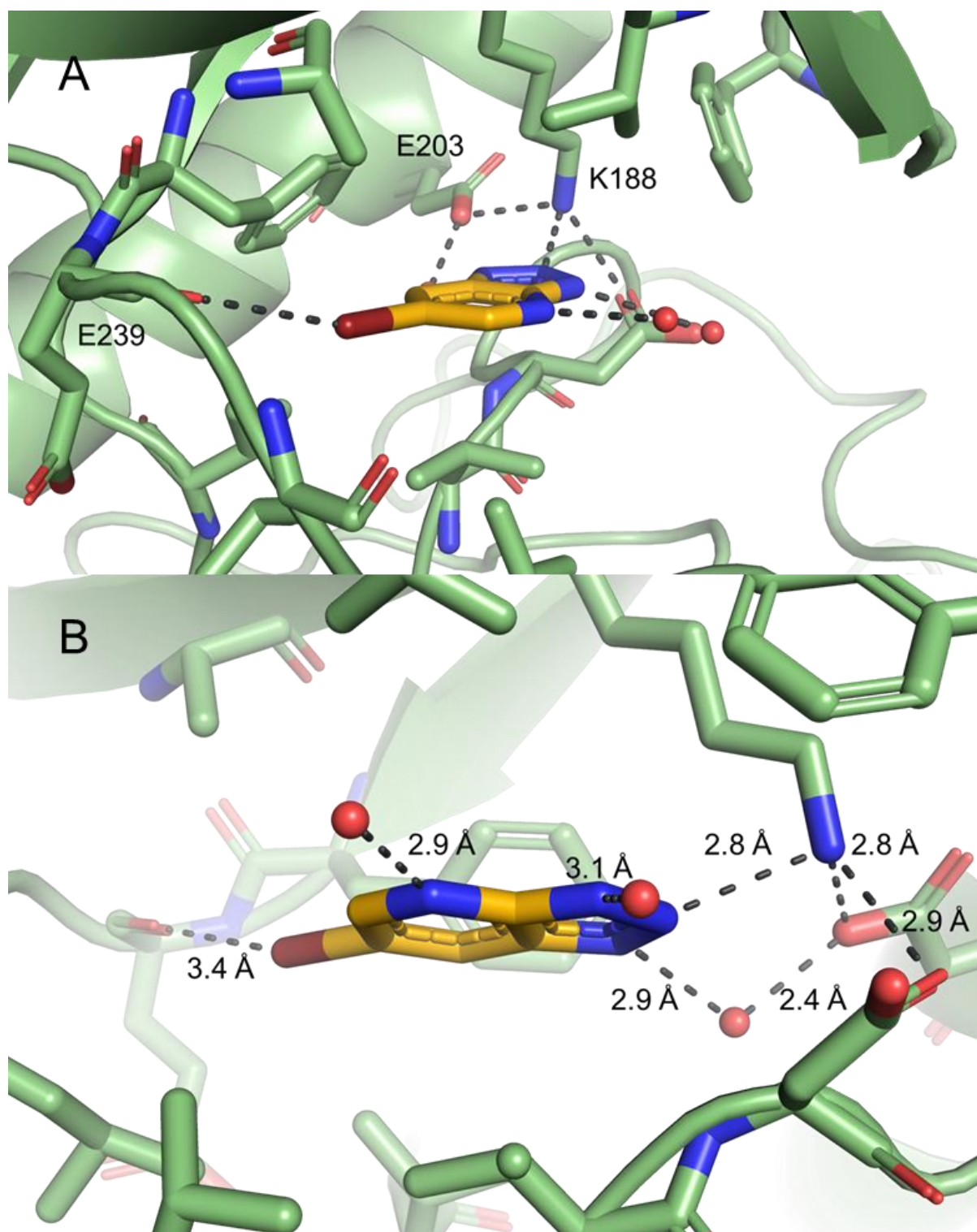


Figure 55. **A:** Binding mode with key interactions of fragment 0482 in the ATP-binding site, the key polar interaction partners, are marked. **B:** The same binding mode from another perspective, with the distances of the key interactions marked; a third coordinated water molecule is visible in this view. The halogen bond angle from the bromine to E239 is  $162.0^\circ$ .

Many lipophilic side chains dominate the ATP-binding pocket. Fragment 0482 is bound by the residues V173, V306, and F238 through either CH- $\pi$  or  $\pi$ - $\pi$ - interactions (Figure 55). The key hydrogen bond is formed by the lysine 188, glutamate 203, aspartate 307, and a water molecule as a sophisticated hydrogen bond network. The triazole ring can interact with this network through two hydrogen bonds. Two additional water molecules are coordinated by the pyridine nitrogen and the third triazole nitrogen—a halogen bond to the backbone oxygen of glutamate 239 anchors the fragment to the hinge region. The distance from the bromine to the oxygen is 3.4 Å, and the C-Br-O angle is 162°, a geometry if not perfect but definitely in the favorable energetic region<sup>199</sup>. With this evidence, one of the critical goals of the study could be achieved, to showcase the feasibility of finding halogen bonds containing binding modes.

The second binding mode of fragment 0482 revealed another sophisticated hydrogen bond network for such a small fragment (Figure 56). The pyridine nitrogen forms a hydrogen bond to the backbone nitrogen of leucine 243. Two of the three triazole nitrogens form perfect hydrogen bonds with the backbone nitrogen of histidine 424 and the lysine 453 sidechain. The lysine 453 is bound to the sidechain of said histidine 424. The third nitrogen must be protonated for the other two nitrogens to act as hydrogen bond acceptors, and an HB donor function would not be tolerated in the binding mode.

Lysine 422 sidechain might interact with the  $\pi$ -electrons of the aromatic scaffold. However, the interaction energy is neglectable with a distance of over 6 Å from the amine to the triazole ring. Leucine 457, on the other hand, forms CH- $\pi$  interactions with the aromatic scaffold at a distance of 4.0 Å. A weak hydrogen bond is formed between the backbone oxygen of valine 376 and the hydrogen at the C5 position in the Triazolopyridine (not shown in Figure 56), with a distance of 3.2 Å from the carbon to the oxygen. This interaction can add a significant amount to the free Gibbs energy of binding. The bromine forms a halogen bond to a water molecule; the distance (3.5 Å) and angle (162.5°) are comparable to the halogen bond in the ATP-binding site. Unlike the earlier case of halogen bonding to a water molecule with KuRoM118 (see section 2.6.2), the water molecule in the second binding site of fragment 0482 is found in all but the least defined chain at the same position<sup>199</sup>.

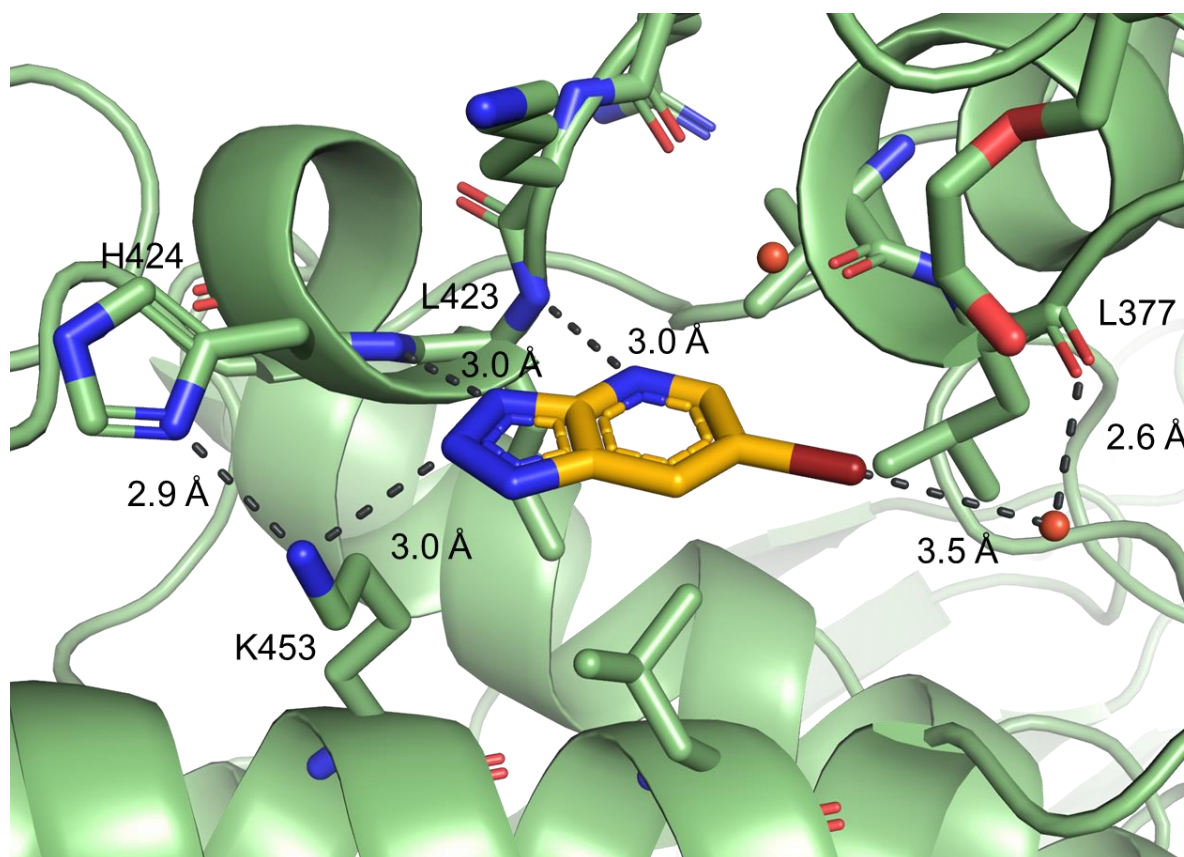


Figure 56. View of the second binding site at the C-lobe of the kinase domain. Key interaction partners are L423, H424, and K453, forming hydrogen bonds to the scaffold nitrogens. L377 engages a water molecule in an HB, which is bound by an XB with a distance of 3.5 Å and an angle of 162.5°.

If the binding sites of both crystals are compared, no significant change within the ATP pocket can be observed in either crystal (Figure 57 A). The water network changes due to the difference in molecule size and the additional hydrogen bond acceptors of 0482. The second binding site of fragment 0482 shows no significant change in the protein structure (Figure 57 B). The only shifted sidechain belongs to K422, but the long-distance and the variable position of the sidechain in every chain, indicate the irrelevance of the binding mode. The sidechain is reasonably well defined in the chains to be confident of its position. In the crystal with the indole-carbonitrile, the second binding site of 0482 is occupied by a water molecule and a sulfate ion. The sulfate forms ionic interactions with both lysine sidechains and the backbone nitrogen of the H424. As both crystals were grown under the same conditions, the sulfate ion has to be replaced by the fragment.

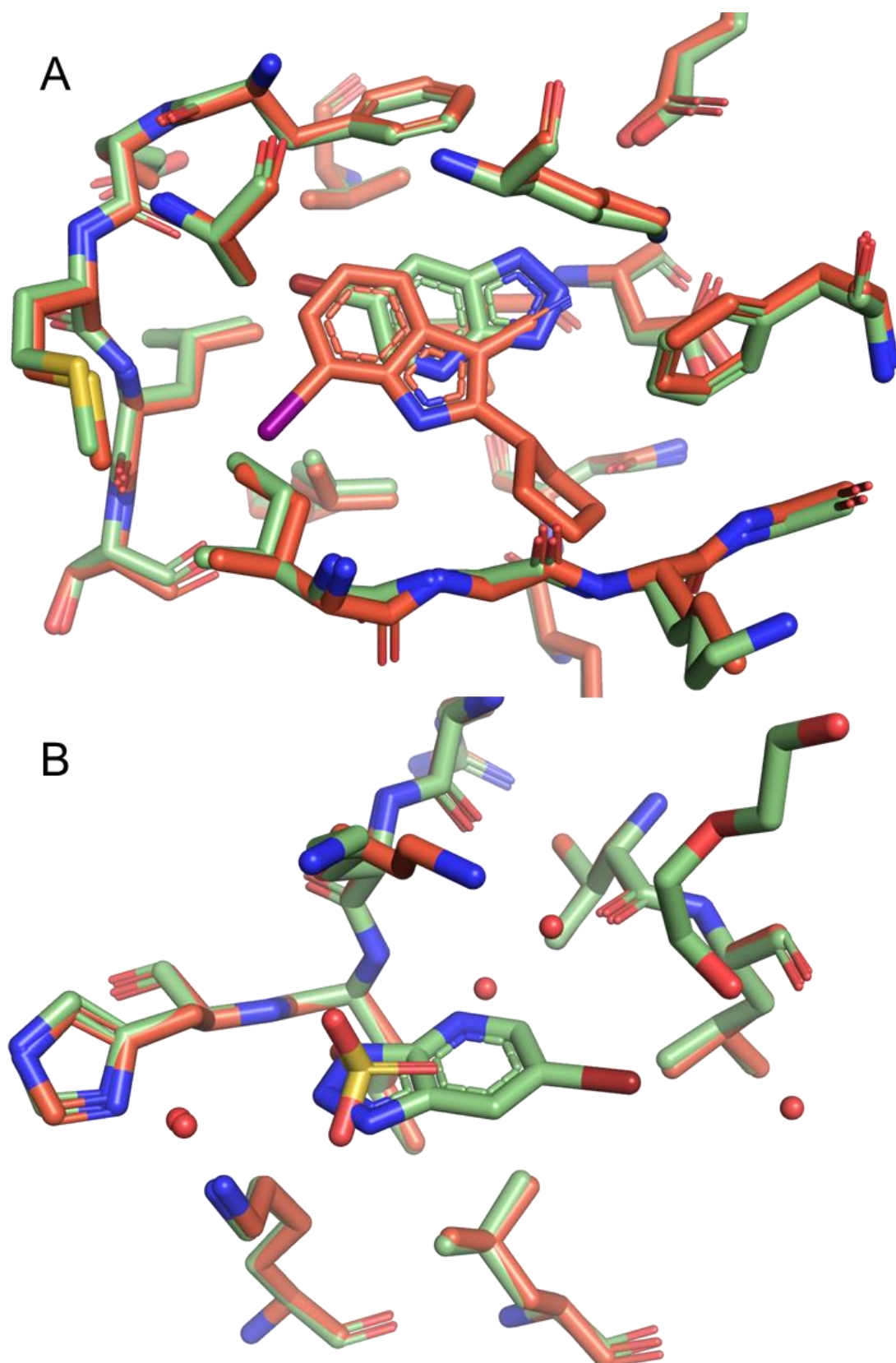


Figure 57. **A**: The aligned ATP-binding site of both crystals, with 0482 in green and KuRoM118 in orange. **B**: The second binding site of the fragment in green compared to the structure solved with the KuRoM118 compound. Only K422 assumes a slightly different conformation upon binding fragment 0482.

The sulfate ion's salt bridges and hydrogen bonds are responsible for a considerable amount of  $\Delta G$ . In addition, 0482 has to replace a water molecule bound to the sulfate ion, L423, and V376 via hydrogen bonds. The  $\Delta G$  upon binding of 0482 has to overcompensate the loss of these interactions. The crystals grew and soaked at a pH of 8.5, and it is imaginable that either the protonated or deprotonated molecule is bound to the secondary binding site. It is impossible to assess the pH level in a self-contained pocket without freely moving water molecules (such as in the ATP pocket). However, the second binding site is merely a crevice, and a fast exchange of water and protons should be possible. Therefore, it can be assumed that the triazole ring is deprotonated upon binding. Although delocalized through the scaffold, the negative charge decreases the  $\sigma$ -hole in the bromine; on the other hand, it strengthens the hydrogen bonds of all three accepting nitrogens. This effect might be why the fragment can displace the water and sulfate ions in the crystal<sup>199</sup>.

In section 2.5.11, a set of analogs of 0482 were tested. The results were unequivocal and not entirely explained by the assumed ATP competitive binding mode. With the knowledge about a second binding mode, some questions arise. The dissociation constant was fitted with an assumed 1:1 ligand-receptor stoichiometry, which no longer holds. Fitting the titration curve to a sequential or a 2:1 binding model was unsuccessful. The impossible deconvolution could arise from dissociation constants of similar magnitude for both binding modes, or the second binding mode has only partial occupancy at a pH of 7.5, at which the ITC measurements took place. As a result, the heat emissions of both binding events overlap and cannot be deconvoluted. It is unclear if the sulfate and water are bound tightly to the protein in solution or if these are crystallization artifacts. The ITC measurements were performed without the  $\text{Li}_2\text{SO}_4$  concentration used in crystallization. In conclusion, it might be possible for the protonated fragment to bind in the second binding site.

Although questions arose, the SAR previously generated (see section 2.5.11) were now elucidated. All four nitrogens are needed in the scaffold; if a carbon atom replaces one, the binding mode is impossible. Looking back at the ITC data, we saw that fragments 0482 ( $K_D=530\ \mu\text{M}$ ) and 0468 ( $K_D>1000\ \mu\text{M}$ ) had the lowest dissociation constants but were the only fragments capable of binding at this site. In every other analog, either the pyrimidine or a triazole nitrogen was replaced with a carbon (see Table 10: D1, D6, D8, and D9).

Exclusion of the second binding mode led to an enormous increase in affinity, disregarding which nitrogen was replaced. There are several reasons for this behavior.

Firstly, it is possible that the jump in affinity is no actual gain in affinity. However, the second binding mode is made impossible, and the titration data can be correctly interpreted in a 1:1 binding model. If we look at the closely related casein kinase 2 (CK2), a 5-bromo-1*H*-benzo[*d*][1,2,3]triazole has been reported as a low micro to high nanomolar binding fragment ( $IC_{50}=26\ \mu\text{M}$ ,  $K_D=0.93\ \mu\text{M}$ )<sup>18, 23, 51, 55, 201, 202</sup>. The affinity is in the same range as the tested 5-chloro-1*H*-benzo[*d*][1,2,3]triazole (D1  $K_D=80\ \mu\text{M}$ ) or the 6-bromo-1*H*-pyrazolo[4,3-*b*]pyridine (D8  $K_D=43\ \mu\text{M}$ , Table 10). The electron density found in the crystal for 0482 suggests a high occupancy and, thus, a micromolar affinity towards one or both binding sites is reasonable. These findings point to a potentially lower affinity of 0482 (6-bromo-1*H*-[1,2,3]triazolo[4,5-*b*]pyridine) to either the ATP binding site or to both sites. The actual dissociation constant of the fragment remains elusive in this study, as the measurement of one binding event without the other was impossible. In the future, we hope to address this issue by generating multiple mutants which incapacitate either one binding site to measure the other site without interference; another way would be incorporating <sup>13</sup>C-labeled amino acids and measuring HSQC-NMR. To accomplish this, we are currently investigating the mutants K118M and K118A to reduce the affinity towards the ATP-site. The mutants H424A, H424P, and K453A shall reduce the affinity towards the second binding site by eliminating the critical interactions of the fragment. Parallel to the mutation study, new analogs will be synthesized to only be capable of binding to one or the other binding site.

Secondly, it is possible that the water network observable in the structure with 0482 is destroyed by the SAR compounds, and the affinity is increased through gain in entropy by the release of water molecules into bulk water. These effects have been studied with halogen-containing compounds by Poznanski et al.<sup>26</sup> Further crystal structures of DYRK1a in complex with D1, D3, or D8 might aid this hypothesis.

### 2.6.3 Fragment 0482 crystal

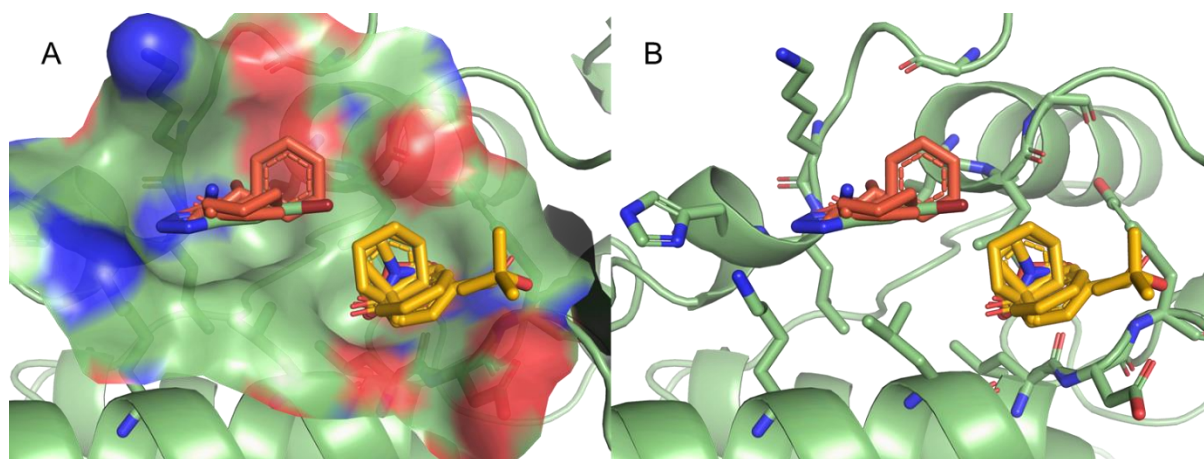


Figure 58. DYRK1a in green with the cluster of solvent mapping performed by Yueh et al. Data pictures were generated with data from the Kinase atlas. Fragment 0482 overlaps with one cluster, and another is roughly 3-4 Å away.

The second binding mode of fragment 0482 represents the first time a possible allosteric binding site has been found for the DYRK1a. All 75 crystal structures published (November 2021) contain compounds binding in the ATP-binding site, with some adjacent residues being utilized. In 2019, Yueh et al. analyzed and reviewed the available structural data on kinases and sorted known kinase inhibitors into groups dependent on the binding pocket<sup>203</sup>. Allosteric or secondary binding sites are well established in some classes and families of kinases. They published a druggability analysis of potential allosteric sites of all kinases with structural data present<sup>203</sup>. In a solvent mapping approach, they found two clusters of solvents for the DYRK1a in direct proximity to fragment 0482 (Figure 58). Both clusters contain 23 different molecules, categorized as potentially druggable in their web application Kinase Atlas<sup>203</sup>. These clusters were categorized in the EDI (EGFR dimerization interface), which had the second-highest median site mapping population in the DYRK1a after the PDK1 interacting fragment<sup>203</sup>.

The protein-protein interactions (PPI) and functionalities of many parts of the kinases surface are poorly understood. DYRK1a is no exception to this lack of knowledge. There are roughly twenty proteins known to interact somehow with the DYRK1a, but no data on a molecular level has been generated detailing the interactions<sup>118, 120, 126</sup>. The DYRK1a consists of a kinase domain, additional 160 disordered amino acids at the N-terminus with a small nuclear localization domain (amino acids 117-134), and another 280 disordered amino acids at the C-terminus.



It remains unknown what kind of functionality many of these significant parts of the sequence are responsible for to this date. Like other allosteric kinase inhibitors, the secondary binding site could directly impact the kinase activity, or the region could be responsible for an upstream or downstream PPI<sup>126</sup>. However, a direct allosteric effect is unlikely as the binding site is neither deeply buried, which could alter the conformation of the domain, nor is it close to the activation loop or other essential parts of the kinase domain. An indirect effect by changing the interactions with other proteins could be envisioned and might change the activity of interacting proteins. Therefore, the secondary binding mode could be developed into a highly affine binder and could change the activity pattern of the DYRK1a or attached proteins.

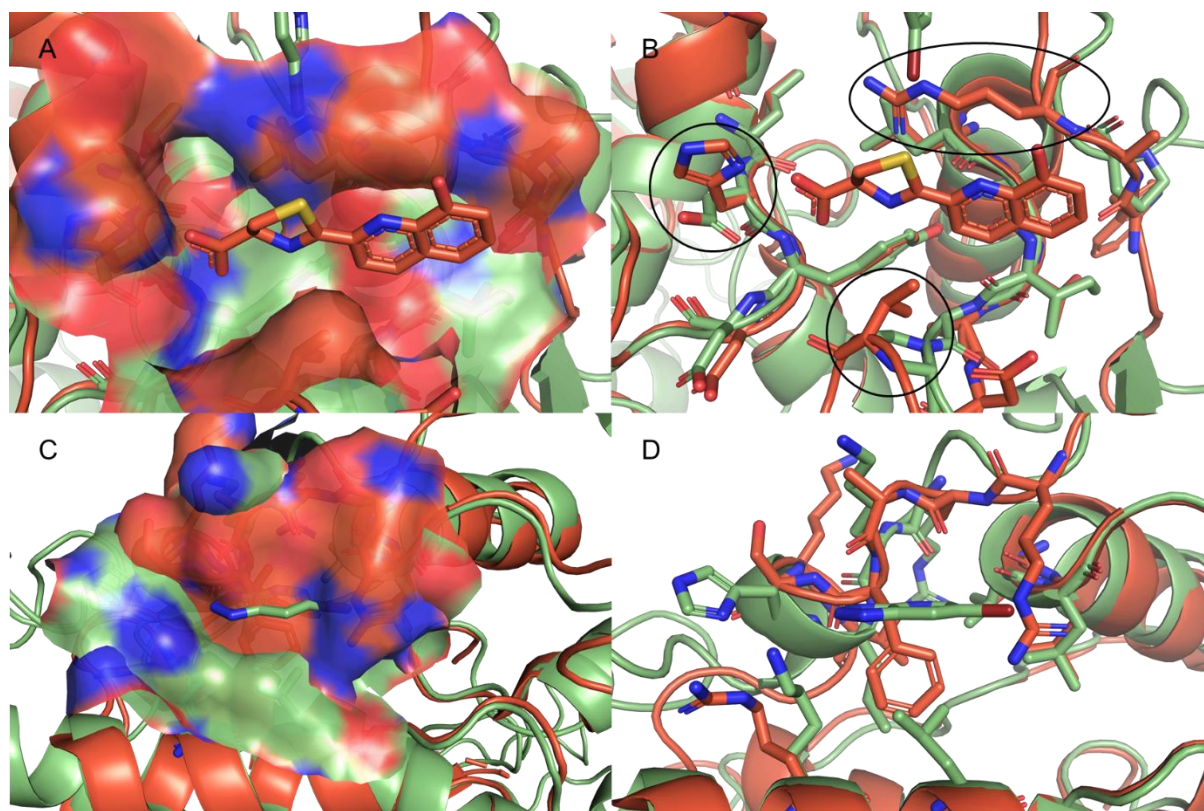


Figure 59. **A:** Alignment of CDK2 (4D1Z)<sup>124</sup> in orange and DYRK1a 0482 structure in green with surfaces present. The EDI pocket for compound 15 is much more encased than the same part in the DYRK1a. **B:** The exact alignment without the surface; the encasement of the crevice is primarily due to three key residues (black circles). The residues R245, H268, and V226 are the main difference in structure compared to DYRK1a. **C:** Both structures' surface overlay illustrates the pocket's absence in the CDK2. **D:** The same view without surface illustrates the high conformational difference of both kinases in that particular region.

Rothweiler et al. reported a luciferin-derived compound (compound 15) with a secondary binding site in the EDI region of the cyclin-dependent kinases 2 (CDK2)<sup>124, 125</sup>. The EDI region exists in the EGFR-family and forms an asymmetric dimer with the N-terminal domain of a second EGFR-kinase<sup>204</sup>. More importantly, an analog region is present in the CDK2, a CMGC kinase, like DYRK1a. They are related to each other much closer than to the EGFR. The EDI region in the CDK and EGFR are considered functional analogs<sup>203</sup>. The compound 15<sup>124</sup> binds directly to a secondary binding site adjacent to fragment 0482 (Figure 59 A). The binding site of compound 15 is highly conserved in the DYRK1a (Figure 59, panels A and B). Although the crevice in which compound 15 binds is more encased than the respective DYRK1a surface, this difference is caused mainly by R245 (Figure 59 B). The residues H268 and V226 are the second and third differing residues, whereas the rest of the structure is conserved. On the other side, the indentation 0482 binds to is entirely blocked by a loop in the CDK2 (Figure 59 C and D).

The bromine of 0482 points directly into the crevice, to which compound 15 binds and could be used as a vector for further synthetic development of the fragment. The indentation next to the secondary binding site features a possible interaction partner but no highly coordinated water. One of the water molecules is held in place by the halogen bond and a hydrogen bond to L377 (Figure 60 B), and D461 holds another two water molecules. These three molecules and I379 and Y462 could be potential guidelines for fragment growing. A method must be validated to distinguish between ATP-site binding compounds and secondary site binding events to tackle this approach. Again, the utilization of structural mutants in combination with the existing crystallization protocol will be needed to develop this fragment into either an ATP-competitive inhibitor or establish the secondary binding site.

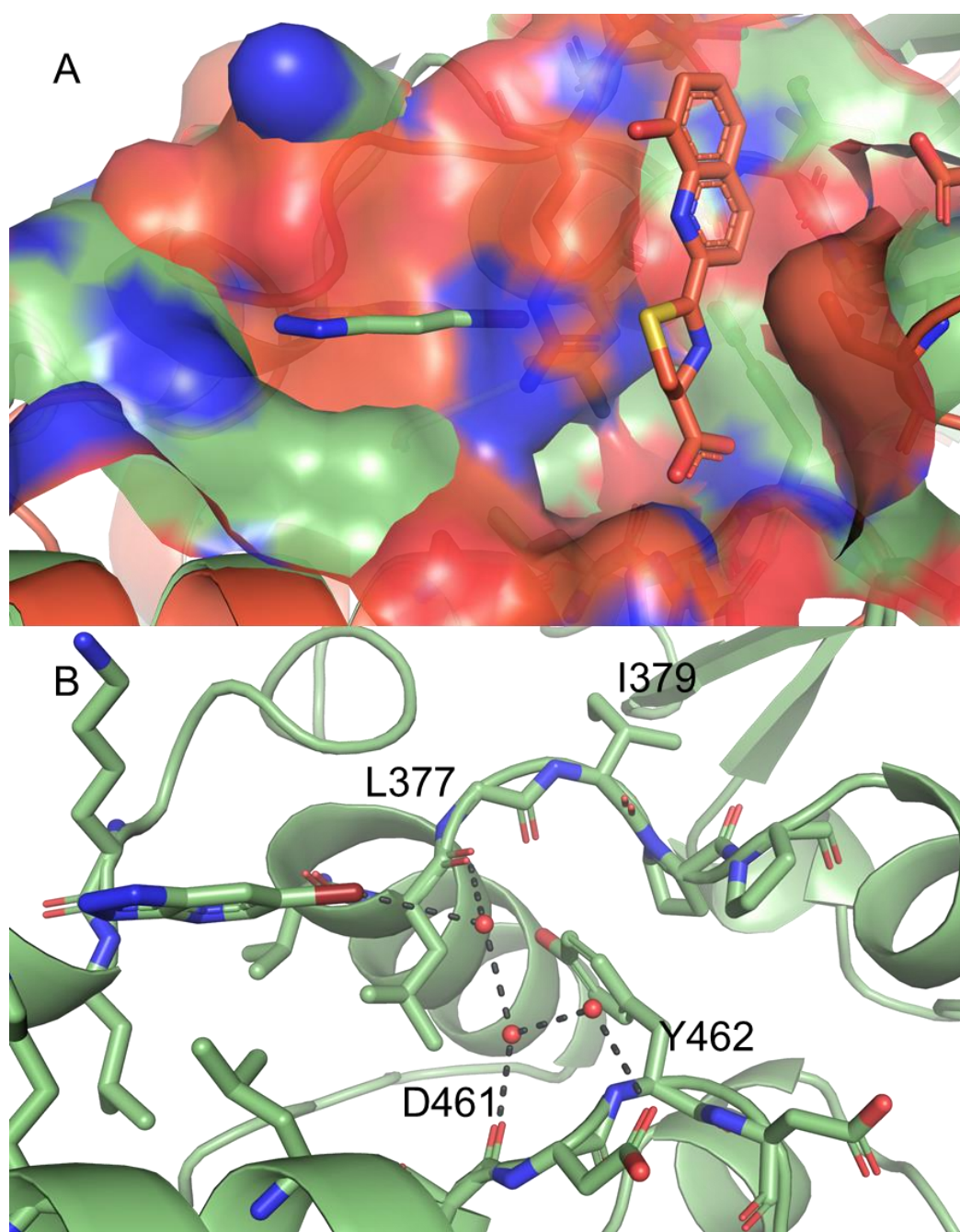


Figure 60. **A:** Combined view of the CDK2 (4D1Z<sup>124</sup>) surface and the surface of the DYRK1a around the secondary binding sites of fragment 0482 and compound 15. **B:** Overview of residues with prominent features pointing towards the surface. The residues L377, I379, D461, and Y462 might be valuable interaction partners for developing fragment 0482 into an inhibitor.

## 2.6.4 Fragment crystallization of JNK

The crystallization of JNK3 is well established and has been done under various conditions in the last decade<sup>63, 127</sup>. The JNK2, on the other hand, proves to be a completely different problem. With only two publications regarding structural data of the JNK2, it is the least studied JNK of the three. The two publications stress the importance of surface mutations to facilitate crystallization but used different constructs in the published structures<sup>128, 133</sup>. Neither published conditions nor three commercially available crystallization screens (Classic II, JCSG++, and PACT++) from Jena Bioscience with 288 conditions in total resulted in any form of crystallization. The published structure (3E7O)<sup>128</sup> was co-crystallized with an inhibitor; therefore, AMP-PCP might not be a viable crystallization agent. In addition, octyl- $\beta$ -D-glucopyranoside was used as an additive in the published conditions, which could be incorporated into screening experiments. Different online tools predict a wide array of possible mutations, which could reduce the structural variability or increase the area of crystal contact, which could alleviate crystallization.

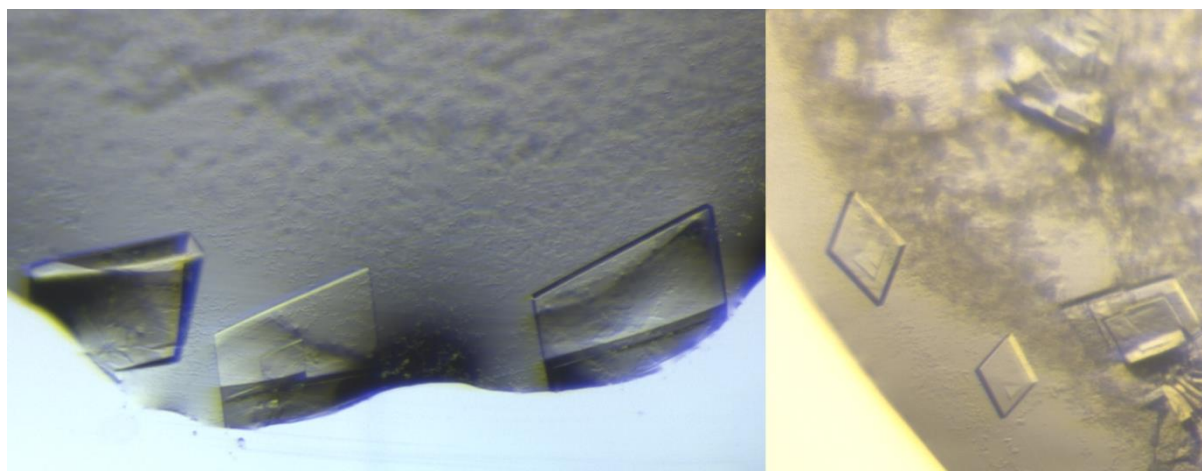


Figure 61. Images of typical JNK3 crystals co-crystallized with AMP-PCP. Crystals typically form rhomboid plates, and the crystals often form macroscopic twins or cracks. Typical crystals are 0.3 mm in the longest direction, with occasional larger crystals.

JNK3 crystals grew with various degrees of macroscopic twinning in rhomboid or cuboid manifestations (Figure 61). The drops always contained aggregated protein to various degrees, which made the fishing process tedious, and the area around crystals had to be freed of aggregates beforehand. Although macroscopic cracks were present in some crystals, diffraction data of these were not inferior to those without cracks.

#### 2.6.4 Fragment crystallization of JNK

Typical crystals grew to a diagonal of 0.2 mm, with the occasional larger crystal up to 0.4 mm (Figure 61 left). The soaking protocol used with the DYRK1a crystals was performed with JNK3 crystals in the corresponding cryoprotection buffer (see section 4.10.2). Again, the soaking in DMSO-containing buffer reduced the diffraction quality of the crystals, which could be observed with the DYRK1a crystal but to a lesser extent. Three crystal structures could be solved. A structure with the ATP analog AMP-PCP could be solved to a resolution of 1.57 Å and an  $R_{\text{work}}/R_{\text{free}}$  of 0.197/0.219, which is the best-resolved structure to date (Table 14). The structure with the initial hit 9612 was solved to a 2.38 Å resolution with a  $R_{\text{work}}/R_{\text{free}}$  of 0.238/0.246. To gain enough completeness, two data sets of the same crystal were collected with a  $\chi$ -angle of 20° in between them. The data sets were scaled and merged (by XSCALE of the XDS package), which is the reason for the high redundancy of 25.5 (Table 14). The structure of the JNK3 in complex with the 9612 analog S1 was solved to 1.8 Å with a  $R_{\text{work}}/R_{\text{free}}$  of 0.226/0.278. More sophisticated collection and refinement parameters are summarized in Table 14<sup>199</sup>.

Table 14. Overview of data collection and refinement statistics of the JNK3 crystals with fragment 9612, its analog S1, and the reference crystal with AMP-PCP.

Data Collection and Refinement	9612	9612S1	AMP-PCP
Beamline	SLS-PXIII Pilatus 2M-F	SLS-PXIII Pilatus 2M-F	SLS-PXI Eiger 16MX
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2	P 2 <sub>1</sub> 2 <sub>1</sub> 2	P 2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions a/b/c (Å)	113.18 / 156.88 / 43.99	112.26 / 157.11 / 44.00	110.65 / 156.81 / 43.96
Cell angles $\alpha/\beta/\gamma$ (°)	90 / 90 / 90	90 / 90 / 90	90 / 90 / 90
Wavelength (Å)	0.919	0.9198	0.9999
Resolution (Å)	50-2.38 (2.44-2.38)	50-1.8 (1.91-1.8)	50-1.57 (1.67-1.57)
CC 1/2 (%)	99.9 (48.9)	100 (49.5)	99.9 (54.4)
Completeness (%)	100 (99.9)	84.8 (99.7)	99.3 (98.2)
Redundancy	25.5 (23.8)	12.63 (12.86)	13.14 (11.68)
I/ $\sigma$ (I)	14.6 (1.24)	17.07 (0.98)	17.34 (1.10)
R <sub>meas</sub> (%)	20.9 (327.5)	7.5 (222.1)	7.4 (202.1)
Wilson B-factor (Å <sup>2</sup> )	62.8	46.76	37.09
Refinement	Phenix v1.19.2	Phenix v1.19.2	Phenix v1.19.2
Resolution (Å)	47.47-2.38	45.61-1.8	47.26-1.57
R <sub>work</sub> /R <sub>free</sub>	0.246 / 0.238	0.226 / 0.278	0.197 / 0.219
R <sub>free</sub> test set	1505 (4.65%)	1515 (2.44%)	1509 (1.41%)
Number of atoms	5243	5523	5803
Protein	680	675	675
Water	50	209	347
Other molecules	5	8	12
B-factors (Å <sup>2</sup> )	72.59	58.51	34.63
Protein	72.6	58.54	34.15
Water	65.14	55.64	38.61
Other	79.14	64.93	42.82
r.m.s.d			
Bond length (Å)	0.009	0.008	0.008
Bond angles (°)	1.153	0.948	1.015
Ramachandran (%)			
favorable/ allowed/ outlier	96.11 / 3.89 / 0	96.23 / 3.77 / 0	98.35 / 1.65 / 0
Rotamer outlier (%)	9.80	7.78	3.41
All-atom clash score	10.95	6.84	3.53

All three structures were solved in space group 18 ( $P 2_1 2_1 2$ ) with two chains in one asymmetric unit. The A chain is better resolved than the B chain in all three structures. The AMP-PCP structure coincides with already published data; Figure 62 A shows an alignment of chain A with the JNK3 structure of 6EQ9<sup>205</sup>. The only part not resolved is the A-loop<sup>188</sup>. In this loop, the residues A211 to T226 are poorly resolved in all chains of the three structures. In the AMP-PCP structure, residues F218 to V224 cannot be fitted to the electron density. In structure 9612, the residues F218 to V225 and I376 to D381, and in the 9612S1 structure residues, M219 to V225 could not be fitted into the electron density. Few amino residues lack electron density in the loop in chain A, partly due to the A-loop's crystal contacts. These contacts cause the conformation allotted to the A-loop in the structures to be unlikely to represent meaningful confirmation in a physiological manner. The loop is not or only partially resolved in most structures and has been assigned to many conformations. Crystal contacts influence the majority of these conformations. The loop is essential for the protein recognition of the kinase and assumes different conformations based on the substrate.

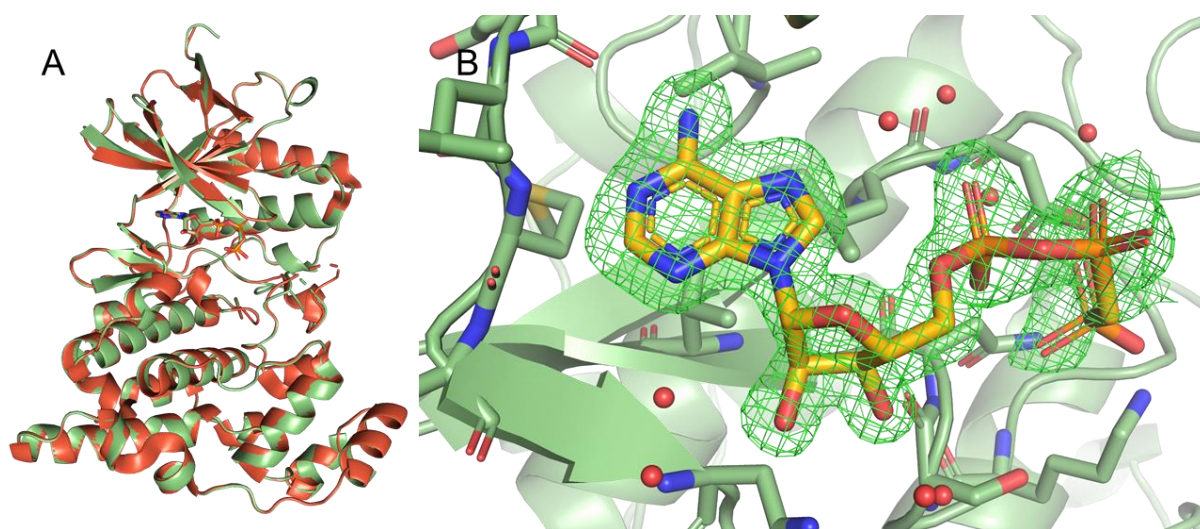


Figure 62. **A:** Alignment of The JNK3 AMP-PCP (green) complex with the PDB-ID 6EQ9<sup>205</sup> (orange); besides the A-loop, the whole structure is well defined and coincides with already published structures. **B:** Detailed view of the ATP-binding pocket with an omit map in green contoured at  $3\sigma$  around the AMP-PCP compound.

The AMP-PCP is well defined by the electron density, which can be seen in the omit map in Figure 62 B contoured at  $3\sigma^{199}$ . Only the  $\gamma$ -phosphate group was found in two conformations if chains A and B were compared. The omit map was used as a reference for the soaking of fragments to differentiate between the electron density of a soaked fragment or the remaining AMP-PCP electron density if fragments would not occupy the ATP binding site.

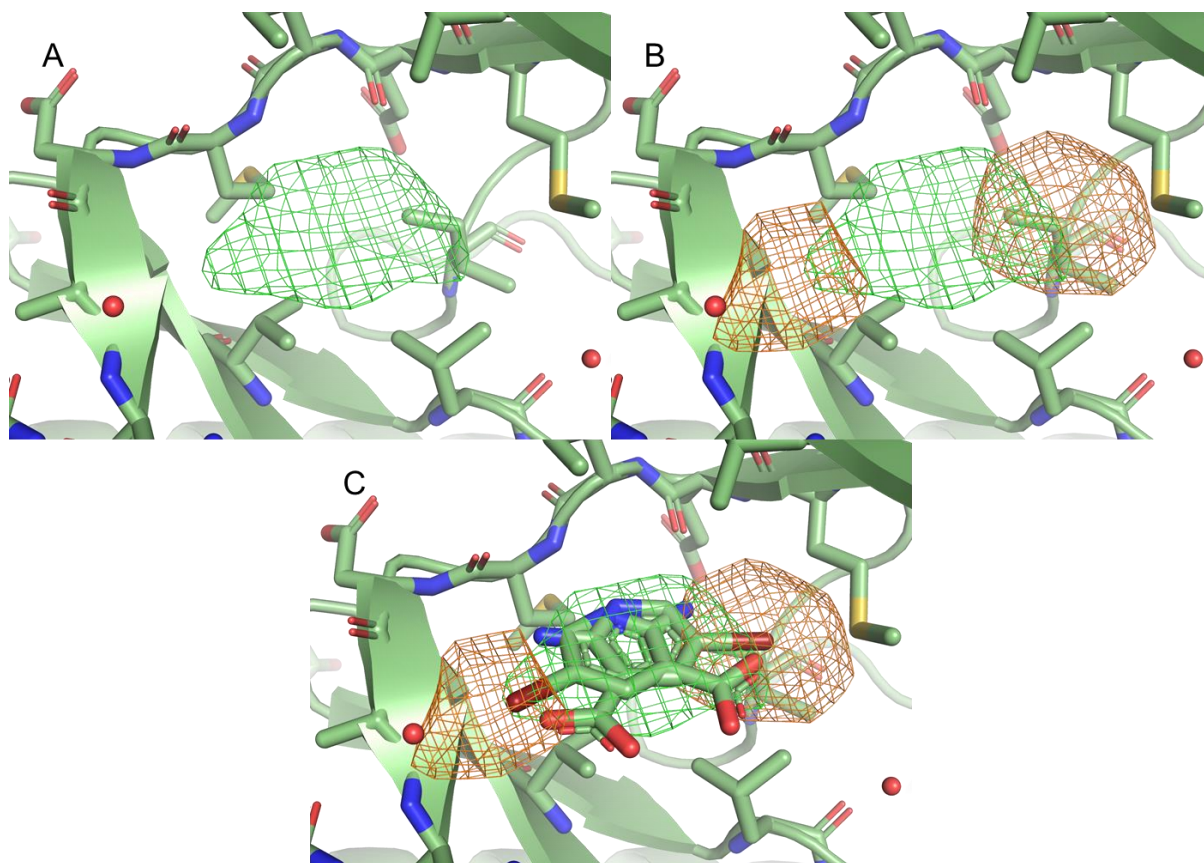


Figure 63. **A:** Omit map contoured at  $3\sigma$  (green), with ambiguous shape, AMP-PCP is displaced, but the binding mode is unclear. **B:** The same view with an additional map generated from anomalous scattering (orange) contoured at  $5\sigma$ , showing two peaks. **C:** Two conformations of fragments 9612 with around 50 % occupancy each.

In the 9612 structure an ambiguous electron density could be found in the ATP-binding pocket. In roughly the same region, the adenosine could be found in the AMP-PCP structure, next to the hinge region, composed of the residues E147 to D150. The standard omit-map generated (Figure 63 A) could not determine the exact binding mode. Due to the bromine in fragment 9612, a wavelength with higher energy than the anomalous X-ray absorption edge ( $0.92\text{ \AA}$  for the K-edge of bromine) was used. The electron map with a resolution of  $5\text{ \AA}$  generated from the anomalous scattering (shown at  $5\sigma$ ) proved the presence of fragment 9612 (Figure 63 B), but instead of one signal



for the bromine atom, two signals at each end of the electron density in the F+F-map were discovered. Two conformations could be found in chain A, whereas only one anomalous signal and one conformation were found in chain B. With the help of the anomalous scattering, the position of the bromine could be pinpointed, and two conformations with around 50% occupancy each were fitted to the electron density in chain A (Figure 63 C).

The two binding modes of fragment 9612 engage different amino acids in the ATP-binding site. In binding-mode A (Figure 64 A), the aminopyridine scaffold forms two hydrogen bonds with the backbone of M149 and E147 at a near-perfect distance. The second binding mode, or conformation B (Figure 64 B), engages with M149 through the pyridine nitrogen. However, the ring is flipped nearly 180° along a thought axis through the pyridine nitrogen and the para-positioned carbon. As a result, the amino function performs a hydrogen bond to the backbone oxygen of M149. The carboxyl group in conformation A lies in the vicinity of methionine 146 (4.1 Å), also known as the gatekeeper residue. The carboxyl group has to be assumed as deprotonated, and the interaction with the methionine sulfur could be discarded as unfavorable with two electron clouds in a certain proximity.

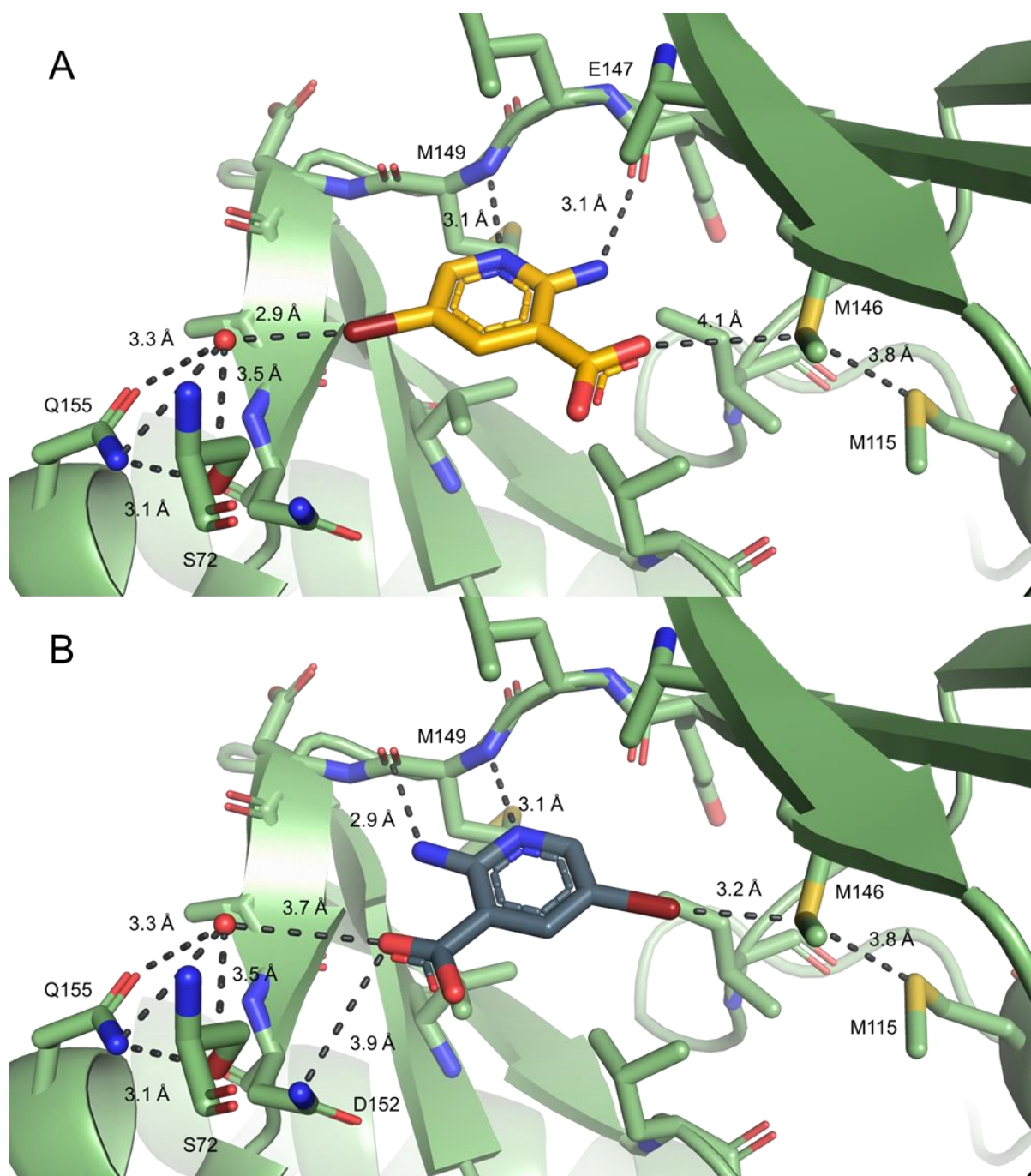


Figure 64. **A:** Binding mode of 9612 with key interactions of conformation A. The halogen bond to the water has an angle of  $170.6^\circ$ . **B:** Binding mode of 9612 with key interactions of conformation B. The halogen bond to M146 has an angle of  $160^\circ$ .

The literature shows that sulfur can engage in an electrostatic interaction somewhat analog to a halogen bond, called a chalcogen bond<sup>63</sup>. The methionine 115 forms such a bond with the M146 as both sulfurs are  $3.8 \text{ \AA}$  apart. The position of the methyl group in the sidechain of M146 is in a position, enabling the  $\sigma$ -hole of the sulfur to interact with the negative charge of the carboxyl group in 9612. Methionine 146 is the nearest residue to the carboxy group in this conformation.

Every other potential residue is more than 5 Å away, rendering them impossible as key interaction partners. The carboxy group engages in multiple hydrogen bonds in the binding mode B. A water molecule and the residues Q155, D152, and S72 form a network of hydrogen bonds with the carboxyl group in this conformation. The proportion in which the serine 72 participates in this network is debatable, as the electron density of the sidechain is not resolved in the structure. The carboxyl group forms a direct hydrogen bond with the amide nitrogen of the D152 sidechain and links with the amide of Q155 through a hydrogen bond with a water molecule. Compared to the water molecule in front of the iodine in the KuRoM118 structure (see section 2.6.3), this water has a lower B-factor and a well-defined electron density. The SAR data in section 2.5.12 proved the carboxyl group vital in the binding mode. The reduction of the carboxyl group to hydroxymethyl abolished the affinity toward JNK3 completely, whereas the affinity towards the JNK2 was significantly reduced (7 fold). The crystal structure proved the halogen's importance, forming a halogen bond to the M146 in conformation B and the mentioned water molecule in conformation A. The bond distance and angle are near perfect for a halogen bond in both conformations. The halogen bond has an angle of 170.6° at a distance of 2.9 Å in conformation A and has an angle of 160° at a distance of 3.2 Å in conformation B. With these geometric parameters, the halogen bond is a highly attractive interaction, and as the SAR data proved, the halogen is essential for binding to JNK3 and JNK2.

With two binding modes competing, it is up to debate which is more critical for the affinity observed and how much the crystallized structure is comparable to the protein in solution. As the binding mode B was found in both chains, while mode A was only found in chain A, we assumed mode B was the more prominent mode. The less resolved chain would only show electron density for the more stable binding mode. On the other hand, the halogen bond to the M146 and the resulting chalcogen bond of the methionine to its neighboring M115 could be shown to be detrimental to the affinity (see section 2.5.13). The assumed interaction of the carboxyl group with the gatekeeper M146 might be a repelling force after all. If the sidechain of the methionine has a slightly different angle, the  $\sigma$ -hole of the sulfur will not point towards the electrons of the carboxyl oxygen, but two pairs of electrons would clash. With more room to accommodate the electron pairs of the M146, the interaction might not be repelling, thus explaining the increase of affinity from wild-type JNK3 to the M115L mutant.

As previously mentioned, the leucine allows the gatekeeper greater conformational freedom without locking the sulfur in a chalcogen bond. Thus, the binding mode A of 9612 could explain the observed behavior of the mutant JNK3.

The third structure with a JNK3 complex with the analog 9612S1<sup>199</sup> showed an unambiguous electron density in the ATP-binding pocket (Figure 65 A). The shape of the compound was well defined, and the iodine could be pinned by the omit map alone. In addition, the generated map from anomalous scattering showed only one signal at  $5\sigma$  (Figure 65 B). A small signal appeared at  $3\sigma$  in the anomalous density but could not be allotted to a second binding mode as the Fo-Fc electron density map did not support another binding mode. The anomalous signal might stem from the nearby methionine sulfur, or only a fraction of the compound binds in an alternative mode, which is insufficient to provide a decent signal in the electron density.

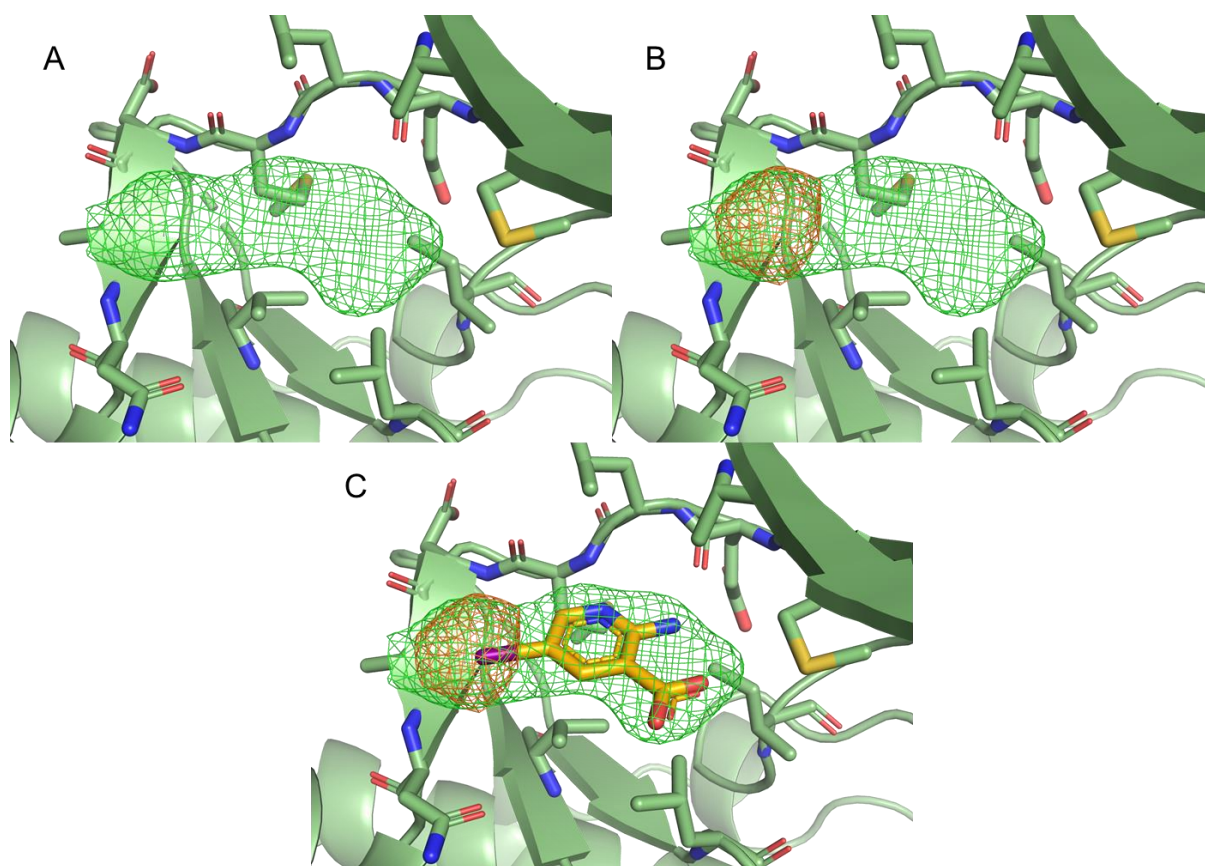


Figure 65. **A:** Omit map contoured at  $3\sigma$  (green), with an unambiguous shape. **B:** The same view with an additional map generated from anomalous scattering (orange) contoured at  $5\sigma$  with one peak. **C:** Only one binding mode of the 9612 analog S1.

Therefore, one binding mode was fitted to the electron density (Figure 65 C). The binding mode is analog to the binding mode A of the bromo-variant of the compound. However, the distances and angles of the interactions have changed slightly, accompanied by an affinity gain of three-fold (see section 2.5.12). The iodo-variant performs the same hydrogen bonds as the bromo-variant to the hinge region. However, the distance of each bond is slightly decreased from 3.1 Å to 3 Å and 2.9 Å, respectively (Figure 66). The iodine performs a halogen bond to the same water observed in all three structures, held by the glutamine 155. The iodine halogen bond has a length of 3.3 Å and an angle of 169.8°. The angle is comparable to the bromo-variant, whereas the greater volume of the iodine could explain the difference in the distance compared to bromine<sup>199</sup>.

The carboxyl group is situated near the gatekeeper M146 sidechain. In the iodo-variant, the distance is even closer than with the bromo-variant (3.4 Å versus 4.1 Å). At the same time, the distance of the two sulfurs of M146 and M115 increased from 3.8 Å in the bromo-variant to 4.4 Å in the iodo-variant. The nature of this contact is complicated to grasp. Suppose the SAR data with the JNK3 mutants is considered again. The JNK3 M115L saw a larger increase in affinity from fragment 9612 to the iodo-variant S1 than the wild-type JNK3 (Table 12). Therefore, a greater distance would be expected if the contact is of a repulsive nature. The decrease in the distance can, in part, result from the error associated with the position of the terminal methyl group in a methionine sidechain. This is particularly true for the structure in complex with 9612, as the resolution is not definitive enough to accurately determine this methyl group's position. The 1.8 Å resolution and the electron density of the iodine structure, on the other hand, allowed for the positioning of the methyl group. As mentioned earlier, a slight change in the methyl-sulfur-carboxyl oxygen angle can make the difference between attractive and repulsive interaction.

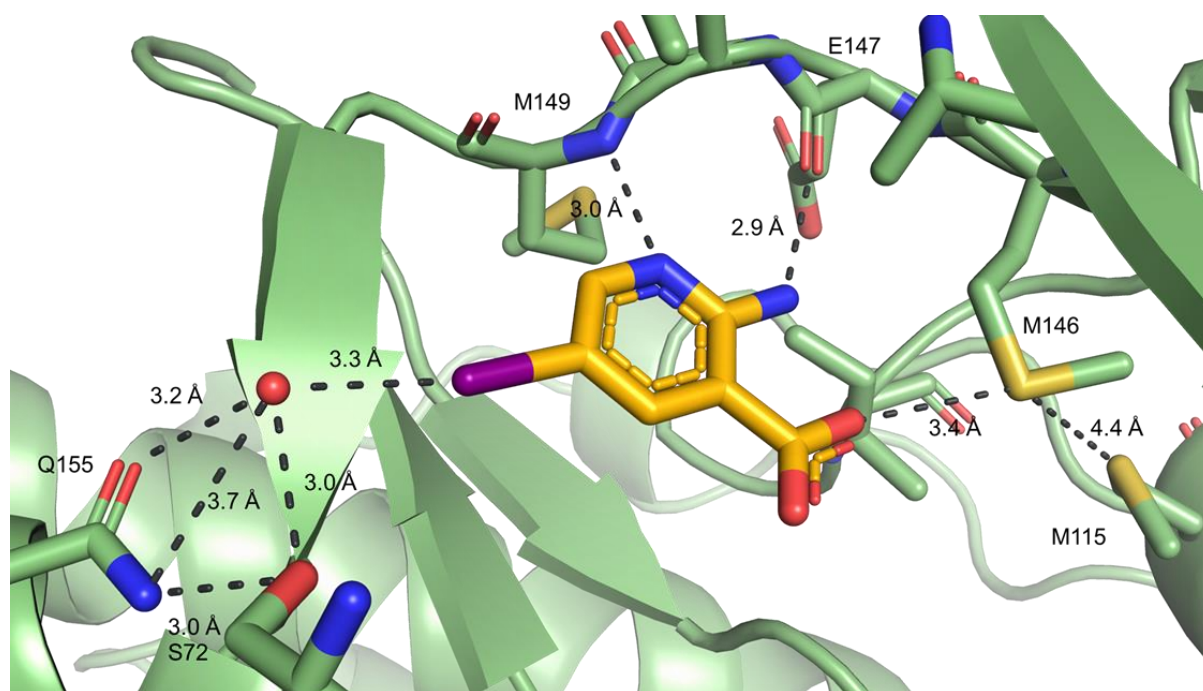


Figure 66. Binding mode of the 9612 S1 with key interactions. The halogen bond of the iodine to the water has an angle of  $169.8^\circ$ .

In Figure 67 A, the binding modes of 9612 (yellow) and 9612S1 (green) are overlaid. Both compounds share an excellent overlap, and the halogens are positioned in nearly the same position. The water molecule bound by the halogen bond, on the other hand, is pushed a small amount, owing to the greater volume of the iodine (Figure 67 B). The glutamine 155, on the other hand, remains in its position resulting in the slightly decreased distance of the hydrogen bond. The position of the methionine is variable compared to the other residues in the binding pocket. The exact position of the methionine's methyl group is vital for the attractiveness of the interaction with the carboxyl group or the halogen in binding mode A. In the case of the carboxyl group, the  $\sigma$ -hole of the sulfur should overlap with the free electron pairs of the carboxyl oxygen. In the case of the halogen, a different geometry is necessary for an attractive interaction, as the sulfur's free electrons have to align with the  $\sigma$ -hole of the halogen. The methionine in the 9612S1 structure is well defined, and only one binding mode can be observed, forming an electrostatic interaction with the M146 sulfur. In the 9612 structure, two binding modes are present, which require different geometries in the M146 sidechain. This could be the reason for the less defined electron density for the M146.

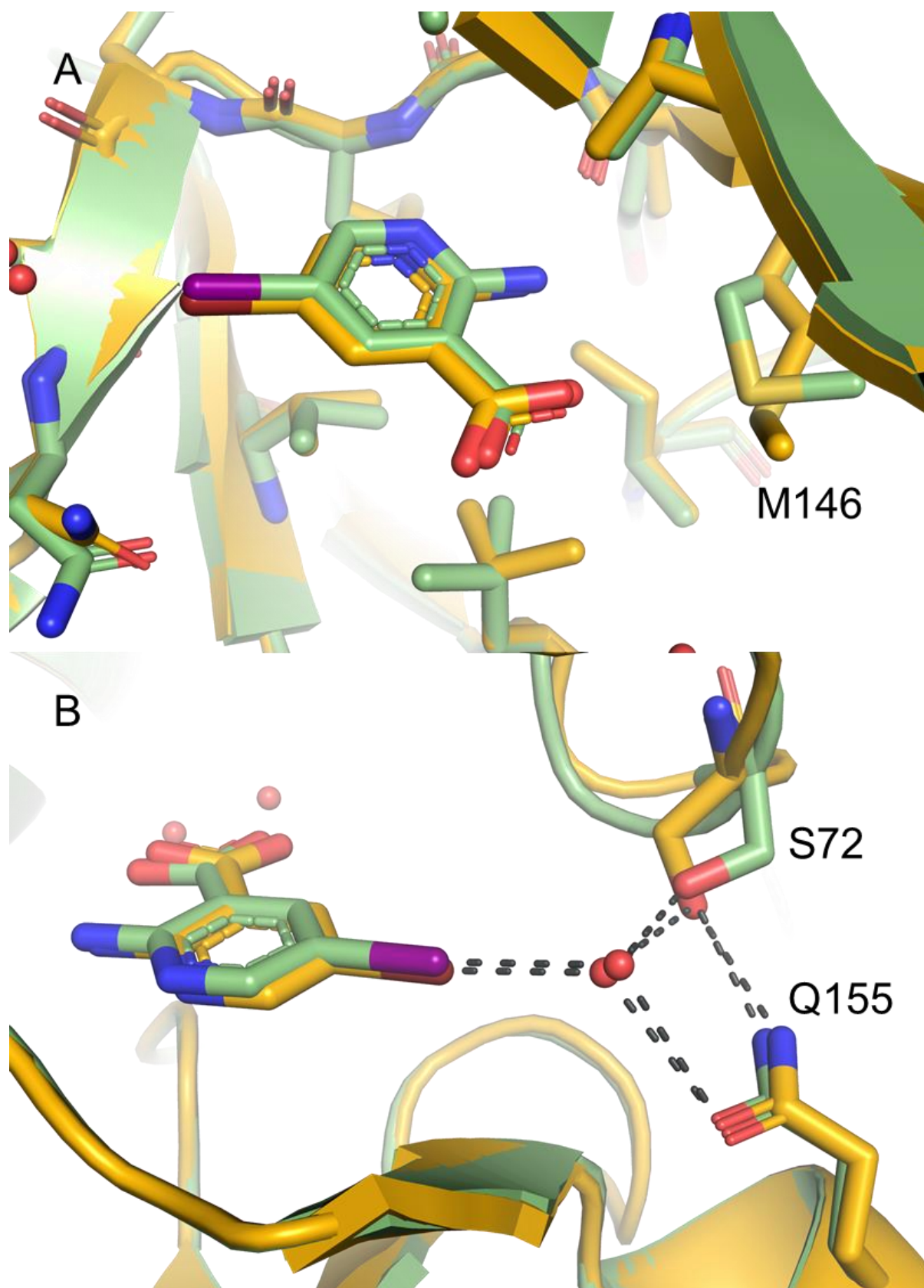


Figure 67. Two views of the binding modes of 9612 conformation A (yellow) and the analog S1 (green). The compounds' binding modes align near perfectly. The P-loop the serine 72 is part of has high degrees of freedom and is, therefore, lesser resolved in the structure. The position of the M146 sidechain is only reasonably defined in the 9612S1 structure.

The serine 72 depicted in Figure 65, 66, and 67 shows the most significant deviation in both structures. This fact should not be overinterpreted as the residue is poorly resolved in the 9612S1 structure, and the sidechain is not traceable in both complexes. The position assigned to the residue is most likely due to the hydrogen bond formed with the water. As the residue is poorly resolved in the electron density map, the sidechain will likely be highly flexible and partly capable of forming this bond.

In fact, the whole region around the serine, from G71 to I77, is wholly changed upon binding of the fragment compared to the AMP-PCP structure (Figure 68). The  $\beta$ 1-sheet and the P-loop of the JNK3 are entirely disordered and form a continuous loop. The P-loop has been known as a more flexible part of the kinase domain due to the three glycine residues in the short sequence<sup>188</sup>, but only one structure of JNK3 (PDB-ID: 2R9S<sup>132</sup>) shows the destruction of the  $\beta$ 1-sheet in a similar manner<sup>132</sup>. As there is no apo-structure of JNK3, the binding event of AMP-PCP could stabilize the  $\beta$ 1-sheet and the P-loop in its canonical conformation. On the other hand, structures of different kinases in its apo-form exist, showing the canonical  $\beta$ 1 and P-loop configuration. The possibility of a naturally occurring conformation without this secondary structure seems unlikely.

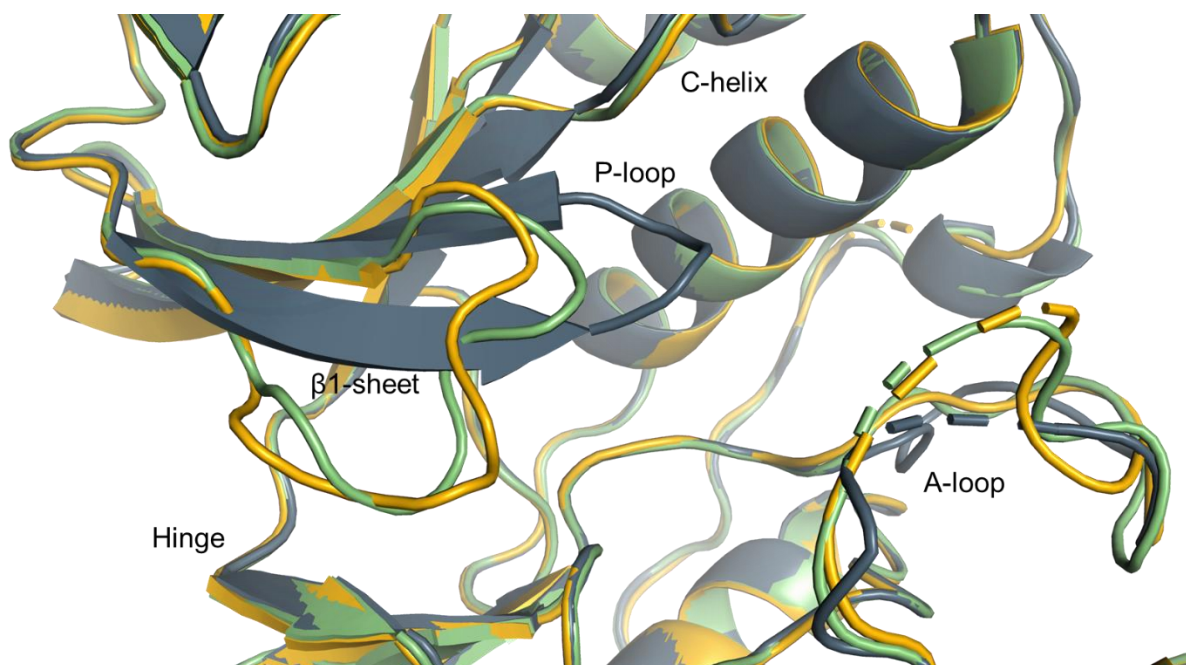


Figure 68. Alignment of the JNK3 structures with AMP-PCP (grey), 9612 (yellow), and S1 (green). The hinge region and the C-helix coincide well, the A-loop is partially unresolved, and the resolved part could only be captured due to crystal contacts in the region. Therefore, the conformation is unlikely to be physiological. The P-loop and the  $\beta$ 1-sheet in the AMP-PCP structure have entirely changed in the other two structures.



The binding of fragments 9612 and S1 could be why this part of the ATP-binding pocket is changed. The same protocol generated all three structures, and the compounds were soaked into JNK3-AMP-PCP crystals. Therefore the conformation change has to occur in the already formed crystal, induced by the binding of the fragments. This region's electron density is less defined than the rest of the structure. So far, the loop cannot be assigned any conformation in chain B. Furthermore, the conformation of the amino acids is different in every crystal structure. The loop is slightly better defined in the 9612 structure (Figure 69 A) than in the 9612S1 structure (Figure 69 B), but both structures lack some density for the side chains. Namely, I77 and S72 in the 9612 structure and Q75 and S72 in the S1 structure. In reality, the loop might retain a certain degree of freedom as no close contact of the compounds is formed, and in turn, a specific loop conformation cannot be fixed.

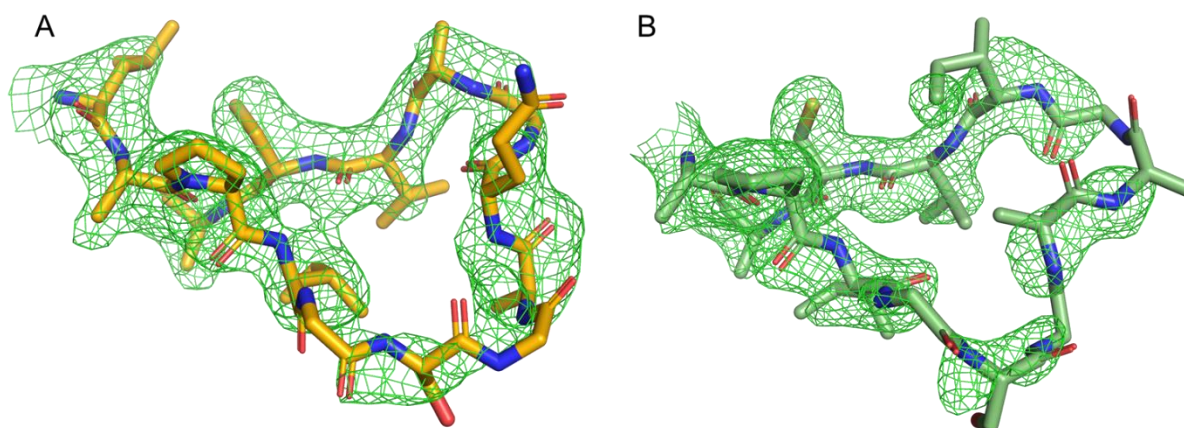


Figure 69. **A:** Stick representation of the P-loop and adjacent residues in the structure of 9612 with an electron density of the 2F<sub>o</sub>-F<sub>c</sub> map contoured at 1σ (green). **B:** Image of the identical residues in the structure of 9612 analog S1 with an electron density of the 2F<sub>o</sub>-F<sub>c</sub> map contoured at 1σ (green).

Although both loop conformations are different from residue G73 to I77, the difference in the canonical conformation remains. Even more astonishing is that fewer and less optimal hydrogen bonds are formed in the loop region compared to the AMP-PCP structure (Figure 70). It remains elusive how the loss of interaction energy is compensated in this secondary structure. An opportunity arises from the conformational shift, as new amino acid sidechains can be addressed by ligands binding in the new conformation. The S72 is a prime example; the backbone is typically fixed in the β1-sheet hydrogen bonds, while the sidechain points away from the pocket.

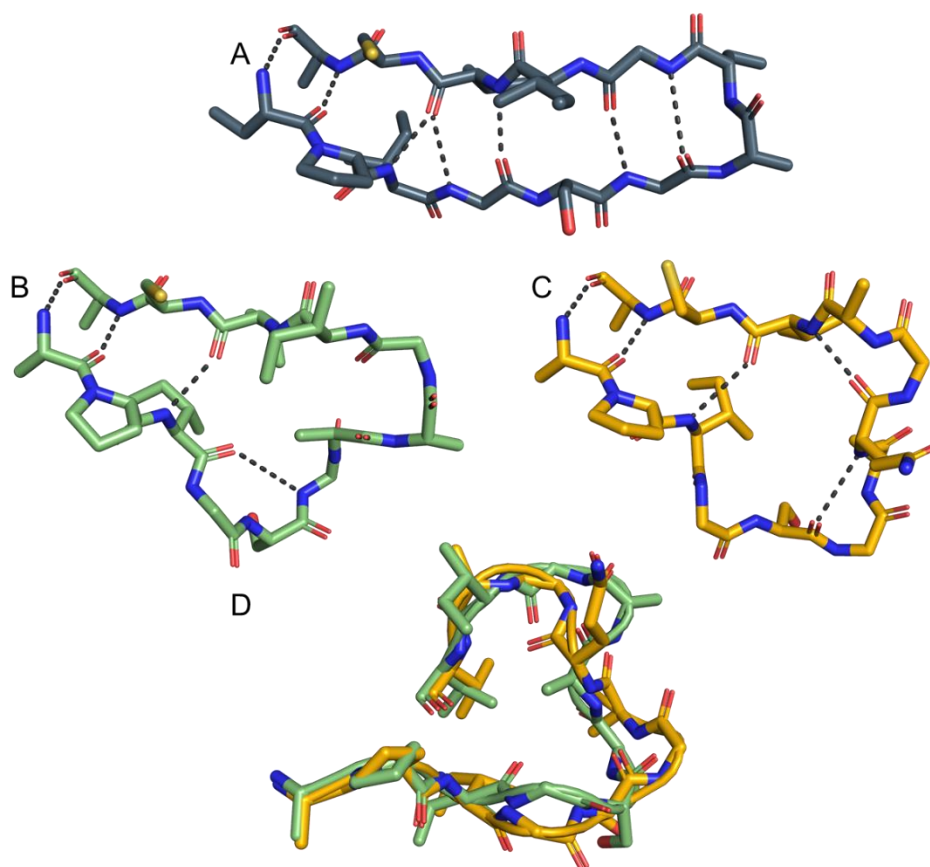


Figure 70. **A:** Stick representation of the P-loop and the  $\beta$ 1-sheet in the AMP-PCP structure, with hydrogen bonds represented in grey. **B** and **C:** The same amino acids are viewed from the same angle in the 9612, and S1 structures with hydrogen bonds marked by the grey dashed lines. **D:** Alignment of B and C viewed from a different angle with an additional cartoon representation.

The residue is poorly resolved in the current structure, but it might be possible to determine more exact sidechain positions with different compounds. With the destruction of the  $\beta$ 1-sheet and the P-loop, ATP can no longer bind to the pocket, leading to interesting dynamic effects or some induced fit in the ATP-pocket for the compound. In addition, the compound's key interactions are not influenced by the conformational shift, and the fragment could bind to the canonical conformation. It remains unclear why the  $\beta$ -sheet dissolves and forms a less interactive conformation.

Further investigations into how the change in conformation could aid the search for new inhibitors are necessary and if the shift is relevant to the development of new inhibitors. Additionally, it must be examined if the conformational shift can and would occur in other kinases. If it has not occurred so far, the JNK3 might be prone to the conformation shift, and it could be used to generate selectivity outside of the JNK-family. All three JNKs show the same sequence in this region (PIGSGAQQIV).

#### 2.6.4 Fragment crystallization of JNK

Therefore, it is likely, that the change in conformation can occur in all three JNKs. We hypothesize that the same shift is present in the JNK2, with slightly different geometries, which could explain the stronger affinity of all the tested compounds to JNK2 compared to JNK3.

## 2.6.5 Development of the fragment 9612

The iodo-variant of 9612 or 9612S1 could be a valuable starting point for a new series of JNK inhibitors. We used a two-pronged strategy to select compounds with an enhanced affinity towards JNK3 or JNK2.

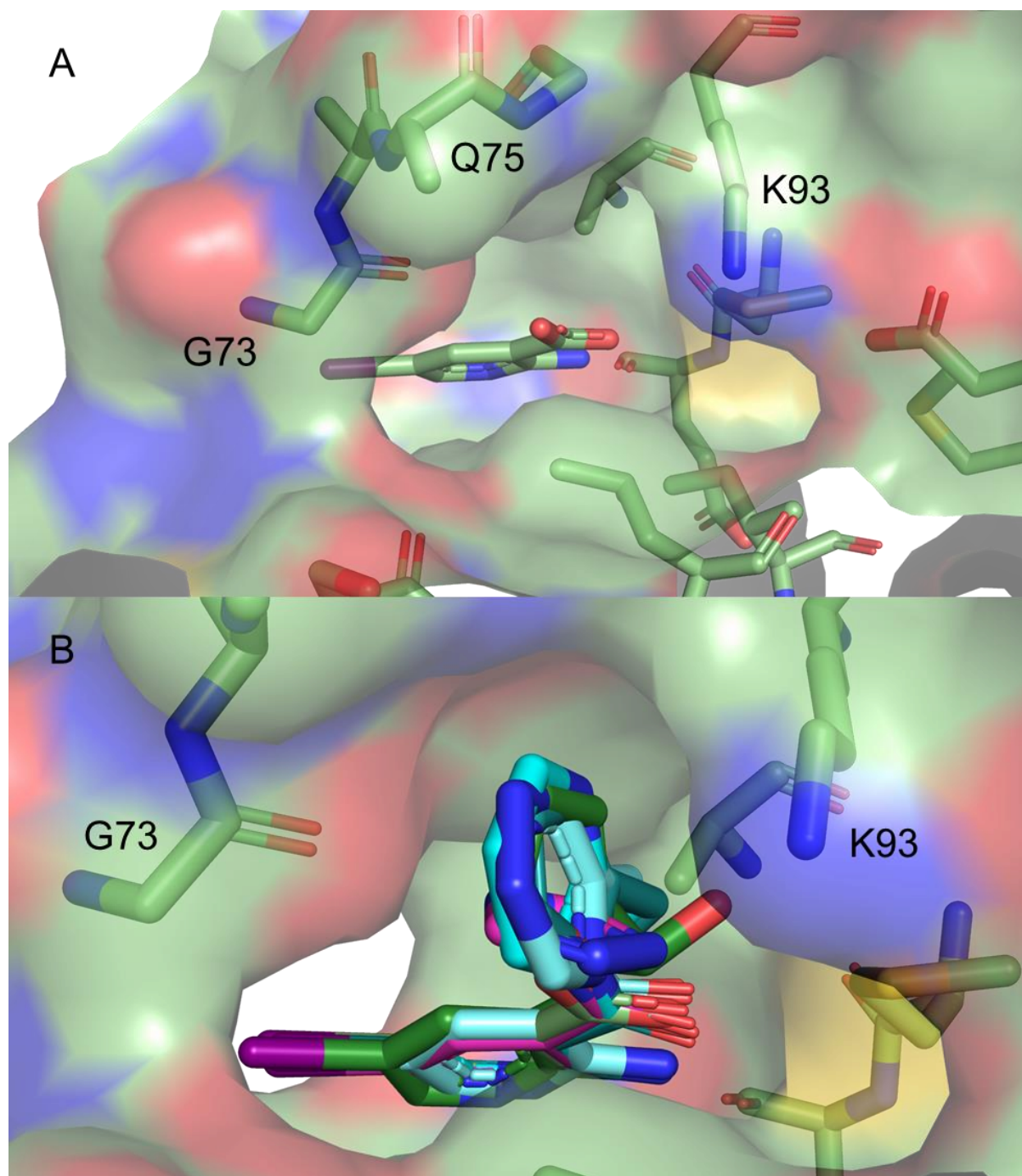


Figure 71. **A:** Closeup of the binding site of S1 with a surface, showing the more closed state of the ATP-binding pocket and the gap between K93 and the loop. **B:** A similar view of the pocket with some docking poses of S1 analogs overlaid. The position of the iodo-aminopyridine scaffold remains static. The substituents primarily target the K93 and various parts of the loop structure (G73, Q75, G76).

We generated 102 analogs of S1 with different groups and substituents at the carboxy group. This group was deemed the easiest way to gain affinity without losing established interactions, and an unoccupied gap in the binding pocket lies in front of one carboxyl oxygen. Esters are often metabolically unstable. Thus, we focused on amides and sulfonamides as replacements for the carboxyl group. The analogs carried differently polarized and sized residues. We aimed at the lysine 93 and the newly formed loop as possible interaction partners for these residues (Figure 71 A).

The docking results were first filtered by the RMSD of the iodine position of the S1 compound to the docked compound. If both positions overlapped significantly, the docking pose was deemed valid. This fix needed to be used to maintain the halogen bond to the water, as the docking program (PLANTS) would not recognize or calculate the halogen bond correctly<sup>1, 83</sup>. For the 102 compounds, a total of 1200 binding poses were found to have a score below -30. These positions were filtered for the positional overlap of the aminopyridine scaffold and, in particular, of the iodine. Sixty positions were selected as the most promising after reviewing (Figure 71 B). The docking results were sent to the group of Pierre Koch, who is synthesizing some compounds to be tested and see whether the docking predictions hold. Suppose compounds show higher affinity than the two crystallized fragments. In that case, they could be subjected to soaking of JNK3 crystals, which would elucidate the binding mode and give further insight into the importance of the loop conformation<sup>199</sup>.

In a second approach, M. Zimmermann conducted a PDB search of small molecules already crystallized with kinases. These fragments were filtered for an amino-pyridine scaffold. In total, 25 fragments with an amino-pyridine scaffold and 18 or fewer heavy atoms were found binding to kinases. Nine of these fragments had a somewhat similar binding mode. The most similar binding mode was occupied with the most similar fragment (5-bromo-2,3-diaminopyridine) (Figure 72 A). The fragment was crystallized with the CK2 in a fragment screening<sup>94</sup>. The fragment obtains a similar binding mode as 9612 mode A, with a potential halogen bond to the gatekeeper, phenylalanine, in this instance. The bond is relatively weak as the distance to the phenylalanine is about 4.4 Å. Although binding modes of 9612 and the CK2 fragment are similar, the  $\beta$ 1-sheet of the CK2 remains unchanged. The sequence in the CK2 differs widely from the JNKs in this region, which eliminates potential assumptions.

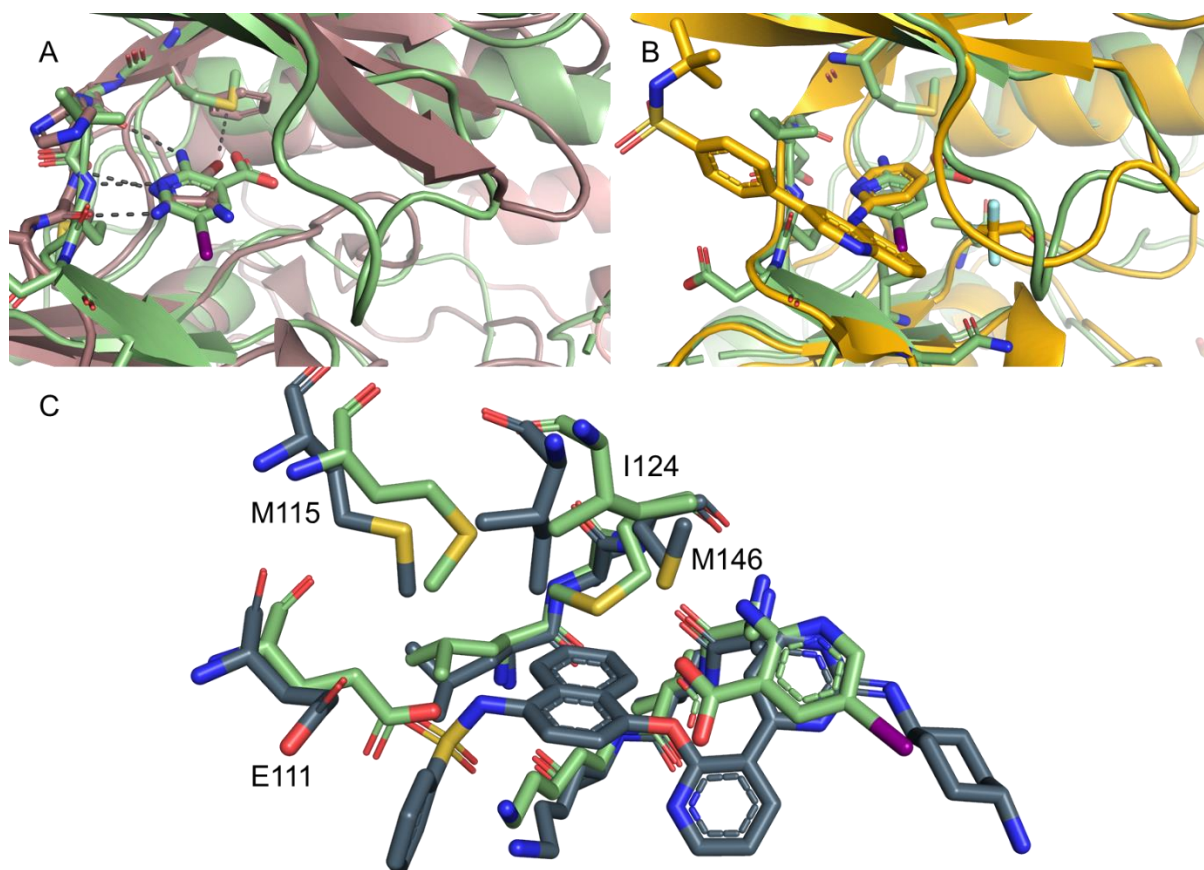


Figure 72. Comparison of different kinase structures with the JNK3 S1 structure. **A:** Hinge region of JNK3 S1 (green) and 5MOW<sup>94</sup> (brown). **B:** ATP-binding pocket of S1 and 2R9S<sup>132</sup> (yellow) with similar binding mode and conformational shift in the  $\beta$ 1-sheet. **C:** Binding mode of S1 and compound 19 from 4U79<sup>206</sup>, the naphthyl moiety could be used as a guideline for development.

Additionally, every structure of the three JNKs were searched for inhibitors with an amino-pyridine moiety. Twenty-three structures were identified with inhibitors bearing aminopyridine-like scaffolds; 2-amino-pyrimidines were included in the search. In 18 of these structures, the aminopyridine moieties engage in hydrogen bonds with the same amino acids as 9612 in binding mode A does. None of the inhibitors have the amino group flipped to the S1 or 9612 mode B position. As mentioned above, the  $\beta$ 1-sheet of 2R9S shows a similar conformational shift as seen in our structures (Figure 72 B). In addition, the complexed inhibitor does not occupy the sugar or phosphate pocket in the ATP-binding site but wraps around the hinge region and engages in hydrogen bonds on the backside of the hinge<sup>206</sup>.

One structure in the literature search stood out, as the complexed inhibitor aligns with the fragment growing vector proposed in the docking. The pyridine-naphthyl ether of compound 19 has substantial overlap with the proposed structures in the docking (Figure 72 C)<sup>206</sup>. Although four residues need to change in their conformation (E111, M115, I124, and M146) to clear the way for the sulfonamide moiety, this substituent could be an inspiration for the newly synthesized compounds.

### 3 Conclusion

In this study, the evaluation and characterization of the HEFLib was the a priori goal. We wanted to showcase the ability of the previously generated library as a valuable tool in finding new fragment binding modes on the one hand. As an additional benefit, the binding modes would be highly likely to utilize a halogen bond<sup>7, 28, 102</sup>. The importance of the halogen bond in binding modes has been discussed controversially in the past. Some skeptics see the halogen bond as a one-off interaction, only contributing significantly to the binding energy in exceptional cases. We see this opinion as misguided, as most halogen bonds are found by serendipity or in the late stages of lead optimization, which leaves less room for optimal halogen bond geometries. The HEFLib shall contribute to disseminating halogen bonds as a valuable tool in medicinal chemistry and drug discovery<sup>199</sup>.

Eight different proteins were evaluated to assess the broad application of the library in a drug discovery project. It was necessary to work with highly variable protein properties, showcasing the general usability instead of specially selected cases. With our now-established screening protocol, we were able to identify a diverse set of fragments binding to various classes of proteins. Due to the high amount of initial hits in some cases, we have utilized electrostatic potential calculations to prioritize initial hit fragments with higher halogen bonding capabilities. It should be noted that even in the primary hit results, an accumulation of fragments with higher  $V_{\max}$ -values was observed. Other factors, such as the number of hydrogen bonds or heavy atoms, were not enriched in the primary hits.

The overabundance of hits was further prioritized and subjected to hit validation by ITC. Due to the reduced sensitivity of ITC compared to STD-NMR, only fragments with desirable LE were validated. The validation of kinase hits was remarkably more successful than IDO1, BIRC5, and DOT1L. Twenty-four protein-fragment combinations with suitable ligand optimization parameters could be validated, as their lipophilicity was relatively favorable, leading to a remarkably good LE of up to 0.81 and low or even negative LELP values. In part, the high solubility and low HA count are responsible for these good metrics.



A few selected fragments were further characterized. The best binding fragment 9595 with the AAK1, the fragment 0482 binding to DYRK1a, and fragment 9612 with JNK2 and JNK3 were further characterized.

With the AAK1 and 9595, the SAR obtained by several close analogs highlighted the importance of chlorine in the binding mode. Although 4-Chloro-5*H*-pyrrolo[3,2-*d*]pyrimidine has been described as a synthetic building block for various kinase inhibitors, the binding capacities of the fragment have not been evaluated before<sup>196-198</sup>. Due to the importance of the chlorine atom for maintaining the affinity, we assume that it shows a novel binding motif, which has not been described before, featuring either a halogen bond or multiple polar interactions. This assumption is supported by the preference of the fragment towards AAK1 with a factor of 10-30 compared to the other five kinases we have tested. Unfortunately, neither our group nor the group of S. Knapp successfully crystallized AAK1 with 9595. The elucidation of the binding mode is essential for further development. Therefore, the now available dephosphorylated protein will be used for testing various soaking strategies and crystallization conditions.

Although the SAR obtained for fragment 0482 (6-Bromo-3*H*-[1,2,3]triazolo[4,5-*b*]pyridine) was elusive at first, the successful crystallization of DYRK1a with 0482 completed the picture. Although the fragment possesses a typical hinge-binding motif, it does not form classical contacts with the hinge region. The bromine forms a halogen bond to the glutamate 239, but no additional hinge amino acid is included in the binding mode. This result raised the question of how strong the hinge binding motif is, compared to the binding mode seen in the crystal structure. With a secondary binding mode in a previously postulated potential allosteric site at the c-lobe<sup>203</sup>, the affinities measured in the ITC could be translated into reasonable SAR. The influence and affinity of fragment 0482 to either binding mode are now subject to investigation. Mutants will be created, and ITC measurements should elicit affinities of either binding mode in the intermediate future. Simultaneously, fragment growing strategies should increase the affinity towards either binding mode. A new possible allosteric binding site in the DYRK1a kinase must be evaluated if and what kind of impact a binding ligand can have. Protein-protein interactions are numerous in the kinase signaling pathways and are not entirely understood. An optimized ligand could give new insights into these pathways and help the understanding of the complex signaling cascades and networks. Again, a halogen bond is one of the few key interactions in both binding

sites. The bond is formed either to a water molecule in the second binding site or backbone oxygen in the ATP-binding site. Without the bromine, the affinities dropped significantly, and it can be assumed that the binding mode would change as well.

The third fragment investigated in detail was 9612 (2-amino-5-bromonicotinic acid) with the 5-iodo-variant S1. Both compounds showed good micromolar affinity towards JNK2 and JNK3 with stringent SAR; of all the tested analogs, only the iodine did not altogether abolish the affinity towards the targets. The preference for JNK2 over JNK3 was subject to a mutation study and could prove the strong influence of one indirectly participating amino acid (L77 in JNK2 equivalent to M115 in JNK3). The crystallization of both fragments and an AMP-PCP structure showed a significantly changed conformation of the  $\beta$ 1-sheet and the P-loop upon binding these fragments. One continuous loop was formed, significantly changing the ATP-binding pocket's shape. The new conformation had not been reported for the JNKs before, and it is unclear which dynamic effects this shift brings.

Further studies are necessary to understand and establish the physiological relevance of the conformational shift. In addition, the resolution of the loop was relatively poor, which speaks for a loose conformation with many degrees of freedom. Developing the fragment into a real inhibitor could alleviate some of these questions while embracing the new interaction possibilities granted by the new conformation. Currently, the group of Pierre Koch is synthesizing new compounds based on a small set of docked compounds. These new compounds will be tested and crystallized with JNK3 and, if possible, with either JNK2 or the pseudo JNK2 variant JNK3 M115L.

All three hits evaluated in detail show halogen bonding as a main contributor to the binding mode. Although the prioritization process was optimized to select for such interactions, it is remarkable that all selected fragments showcased halogen bonds with near-perfect geometrical parameters. With these results, the HEFLib proved to be a fragment library capable of generating highly affine binding modes of tiny fragments with a high probability of halogen bonds with nearly perfect geometries.

#### 3.1. Outlook

Besides the three fragments evaluated in detail, many hits validated by ITC could be further explored either in a SAR campaign by in-house generated compounds or by X-ray crystallography. Especially, fragments binding to the CAMK1G and IDO1 could not be addressed adequately so far. The crystallization of JNK3 and DYRK1a is established with a soaking protocol capable of identifying micromolar binding fragments. The binding mode for the validated hits for these proteins could be easily elucidated. Elucidation of multiple fragments' binding modes would open up the possibility of fragment merging or linking, as more fragment binding modes are known.

Going a step back in the screening process, many unvalidated hits from the STD-NMR screening still need validation. Especially the number of validated hits with the AAK1 and CAMK1G showcases the potential that might still be hidden in the remaining hits. The SAR compounds acquired in the 9595 evaluation with AAK1 could be tested against the other five kinases. Fragment 9595 binds to all five kinases, and thus, the SAR compounds could give further insight into those binding modes as well. Similar binding modes could be possible if comparable affinities were generated with all SAR compounds in the four remaining kinases. Substantial deviations in the affinity pattern could lead to assumptions about each binding mode and the importance of the chlorine atom in all of these.

The HEFLib proved to be a successful library in the drug discovery campaign conducted in this study. Based on this library characterization, we will continue to develop the library in several ways. Firstly, we will aim to extend our library with additional fragments featuring a firmly tuned  $\sigma$ -hole of the halogen. This has already happened in part by acquiring SAR compounds featuring a larger halogen. SAR satellites were a part of the library from the design phase on, and we want to engage in the possible benefits of hit validation by analogs. Although the incorporation of similar fragments decreases the overall diversity of the library, these will only be part of a sub-library used if the diverse subset of fragments generates a hit. We hope to find a compromise between library size, diversity, and validation by analogs with this strategy.

### 3.1. Outlook

Correlation between  $V_{\max}$  and the number of hit fragments, combined with the overabundance of iodine within these hits, concluded that an increase in iodine-containing fragments would improve the HEFLib characteristics. The overall iodine content in the HEFLib is limited so far because of reduced commercial availability and higher costs of suitable iodinated fragments. As an alternative, the focus of the development of the library will be on fragments with more tuned  $\sigma$ -holes. One way to increase the  $V_{\max}$  of fragments is the addition of fluorine. Adding a large amount of fluorine to a compound drastically decreases its solubility, which is detrimental to our screening protocol. In recent years the idea of using halogenated difluoro-methyl groups could expand the capabilities of halogen bonds. Utilizing this substructure would relieve the XB from an aromatic scaffold, as those groups are chemically and metabolically stable. In a halogen-difluoro-methyl group, the halogen can rotate and, in principle, could adapt more easily to form good XB contacts.

The STD protocol was based on an initial optimization for JNK3. Here, optimization for other targets might prove reasonable, especially with different sizes. Hits generated with the kinases proved to have a much higher affinity, mainly caused by the well-defined and enclosed ATP-binding site. Targets with less structured binding sites could be overlooked, and the screening could be unsuccessful if the protocol were not adapted.

Every crystal structure solved in the study showed the involvement of water molecules not only as a conveyor of hydrogen bonds but, more important, as halogen bond acceptors. The accumulation of halogen-water interactions is remarkable, but the information on this interaction is limited and primarily based on very confined systems, such as clathrates or rotaxanes<sup>207</sup>. As water is unlikely to be held just by a halogen bond, at least one other amino acid is involved in the interaction<sup>208</sup>. Evaluating these three centered halogen bonds in a biochemical environment could aid the estimation and classification of such interactions compared to the already established halogen bond donors (e.g., protein backbone,  $\pi$ -systems, or electron-rich sidechains). Furthermore, the often-discussed desolvation penalty for a hydrogen or halogen bond would be eliminated in most cases. Therefore, a greater potential for higher gains in net adduct formation energies could be possible.

## 4. Materials and Methods

### 4.1 Media Preparation

LB-Medium contained (1 L):

15 g tryptone  
5 g yeast extract  
10 g NaCl

The contents were mixed with ddH<sub>2</sub>O, pH was adjusted to 7, and the media was autoclaved at 121 °C for 20 minutes. For plates, 1.5 g of agar per 100 mL was added before autoclaving.

2xYT-Medium contained (1 L):

16 g tryptone  
10 g yeast extract  
5 g NaCl

The contents were mixed with ddH<sub>2</sub>O, pH was adjusted to 7, and the media was autoclaved at 121 °C for 20 minutes. For plates, 1.5 g of agar per 100 mL was added before autoclaving.

TB-Medium contained (900 mL):

12 g tryptone  
24 g yeast extract  
5 g glycerol

10x TB-salt (1 L):

2.31 g KH<sub>2</sub>PO<sub>4</sub> (0.17 M)  
12.54 g K<sub>2</sub>HPO<sub>4</sub> (0.72 M)

The contents were mixed with ddH<sub>2</sub>O, pH was adjusted to 7, and the media was autoclaved at 121 °C for 20 minutes. After the medium was autoclaved, 100 mL of sterile 10x TB-salts was added.

## 4.2 Buffer Preparation

SOC-Medium contained (1 L):

20 g	tryptone
5 g	yeast extract
5 g	NaCl
0.186 g	KCl

The contents were mixed with ddH<sub>2</sub>O; pH was adjusted to 7 and autoclaved at 121 °C for 20 minutes. Before use, 5 mL of a 2 M MgCl<sub>2</sub>·6 H<sub>2</sub>O (40.7 g in 100 mL water) and 20 mL of a sterile filtered 1 M glucose solution (18 g in 100 mL) were added.

## 4.2 Buffer Preparation

Buffers were calculated with the temperature correction and Debye-Hueckel correction for ionic strength in mind by using the buffer calculator tool from Rob J. Beynon (<http://phbuffers.org/BufferCalc/Buffer.html>). Buffers used on the ÄKTA (Cytiva) system were sterile filtered before use. The pH values given correspond to the temperature at which the buffers were used.

Kanamycin and Ampicillin stocks were prepared at 100 mg/mL in ddH<sub>2</sub>O and sterile filtered. Ampicillin solutions were always used within two days. Chloramphenicol stocks were prepared as 34 mg/mL in ethanol.

## 4.3 Competent Cell Preparation

*E. coli* cells were plated on LB-agar plates, and suitable colonies were transferred to 5 mL of LB medium the next day. DH5 $\alpha$  cells were cultivated without antibiotics, whereas Rosetta 2 (DE3) pLysS and BL21 (DE3) pLysS cells were cultivated with chloramphenicol (34  $\mu$ g/mL) to retain the LysS plasmid. Cultures grew overnight at 37 °C while shaking at 220 rpm with a 25 mm shaking throw on an HT Multitron (INFORS). 100 mL LB medium was inoculated with 1 mL preculture and grown to an OD<sub>600</sub>=0.6. The cells were harvested by centrifugation for 10 min at 4000 rpm on a Heraeus Multifuge X3-FR (Thermo Fisher) at 4 °C. The pellet was resuspended in

#### 4.4 DNA Preparation

10 mL of ice-cold 0.1 M CaCl<sub>2</sub> with 10% glycerol. The suspension was centrifuged again, and the pellet was washed with the CaCl<sub>2</sub> solution and resuspended again. 100 µL aliquots of the suspension were flash-frozen in liquid nitrogen and stored at -80 °C.

#### 4.4 DNA Preparation

##### 4.4.1 Heat Shock Transformation

An aliquot of the CaCl<sub>2</sub> competent *E. coli* cells was thawed on ice for 5 min. 2 µL (50-100 ng) of the plasmid were added and incubated on ice for 30 min. The Heat Shock was performed at 42 °C for 2 minutes. Afterward, cells were cooled on ice for another 5 min. 900 µL of SOC medium were added, and the culture was incubated at 37 °C and 400 rpm on an MB-102 mixing block (Bioer) for an hour. Cells were plated on LB agar with suitable antibiotics and incubated overnight at 37 °C.

##### 4.4.2 Mutagenesis

Primers for mutagenesis were designed with the NEBaseChanger™ tool. These were then synthesized by Merck (EMD). Lyophilized primers were dissolved in water to a 100 µM concentration. These stock solutions were then further diluted to 10 µM for use in the NEB Q5® site directed mutagenesis kit. A mutagenesis PCR mixture consisted of:

Q5 Hot Start High Fidelity 2x Master Mix	12.5 µL
10 µM forward primer	1.25 µL
10 µM reverse primer	1.25 µL
Template DNA (10-50 ng/µL)	0.5-1 µL
Water	9-9.5 µL

#### 4.4.2 Mutagenesis

The PCR program consisted of 25 Cycles:

Initial Denaturation	98 °C	30 s
Denaturation 98 °C	10 s	
Annealing	50-72 °C	30 s
Polymerization	72 °C	180 s
Final Extension	72 °C	120 s
Hold on at	4 °C	

Annealing temperatures were chosen according to the temperatures calculated by the NEBaseChanger™ tool. PCRs were conducted in a Mastercycler gradient (Eppendorf).

PCR primers for mutagenesis, in a 5'-3' order.

JNK3 M115A      Forward: 5'GCTGGTCCTCGCGAAGTGTGTGAACC3'

Reverse: 5'TCCCGGTACGCTCTCTTG3'

JNK3 M115L      Forward: 5'GCTGGTCCTCCTGAAGTGTGTGAAC3'

Reverse: 5'TCCCGGTACGCTCTCTTG3'

IDO-TEV          Forward: 5'CGAAAACCTGTACTTCCAGAGCC3'

Reverse: 5'CCGCTGCTGTGATGATGA3'

Subsequently, the PCR product was cleaned, and the following protocol digested the template DNA.

PCR Product	1 µL
2x Kinase, Ligase, DpnI buffer	5 µL
10x Kinase, Ligase, DpnI Enzyme mix	1 µL
Water	3 µL



#### 4.4.3 Cloning

This mixture was incubated for 30-60 minutes at room temperature. 5  $\mu$ L of the mixture was added to the *E. coli* DH5 $\alpha$  cells, which were provided with the kit, and a heat shock transformation was performed. Only the duration of the heat shock was changed to 30 s to comply with the kit instructions.

#### 4.4.3 Cloning

The DNA sequences of the AAK1 (amino acid 121-471), CAMK1g (amino acid 1-320), DOT1L (amino acid 1-416), GSK3 (amino acid 1-420), and JNK2 (amino acid 2-362) were optimized for *E. coli* expression and synthesized by Thermo Fisher's GeneArt service. The synthesized genes came in pMAT vectors with BamHI and EcoRI restriction sites attached to the genes. First, the lyophilized genes were dissolved in nuclease-free water and then transformed into *E. coli* DH5 $\alpha$  cells with Ampicillin as an antibiotic. The next day suitable colonies were picked, and 100 mL LB medium was inoculated and cultivated overnight. The following day a QIAprep Spin Maxiprep Kit (Qiagen) was used with the manufacturer's protocol for extraction and purification of the plasmids. Eurofins sequenced all synthesized genes to verify the sequence after this step. DNA of JNK3 (amino acid 39-402) was cloned into the pET24a\_HLT vector in a previous work by A. Lange from the Kinase ORF-Kit (Addgene)<sup>209</sup>. The pET15B\_IDO1 (amino acid 1-399) construct was a generous gift of H. Sugimoto<sup>169</sup>. Susanne Hennig previously generated the pET24\_HLT\_BIRC5 (amino acid 1-142) construct from an insert synthesized by Eurofins.

The plasmids were then simultaneously digested by BamHI HF (NEB) and EcoRI HF (NEB) enzymes for 15 min at 37 °C, followed by an inactivation step for 15 min at 65 °C. The digestion contained the following ingredients:

Cutsmart buffer (NEB) 10x	5 $\mu$ L
Appropriate plasmid	10 $\mu$ L
BamHI HF	2 $\mu$ L
EcoRI HF	2 $\mu$ L
Water	31 $\mu$ L

#### 4.5 Protein Expression and Purification

The reaction was cleaned using a preparative 0.8% Agarose gel with TAE buffer (40 mM TRIS, 20 mM acetic acid, 1 mM EDTA). The gel was stained with SYBR gold (Thermo Fisher) for 20 minutes and viewed on a Gel Doc XR+ (Biorad); the gene band was cut out. Clean-up was performed with a QIAquick PCR & Gel Cleanup Kit (Qiagen) according to the manufacturer's protocol. Instead of eluting with 50  $\mu$ L elution buffer, 30  $\mu$ L was used.

Genes were ligated into the according vector with a T4 DNA Ligase (NEB). The ligation was carried out at 4 °C overnight with the following composition:

T4 buffer 10x	2 $\mu$ L
linear plasmid	1 $\mu$ L
appropriate Insert	2-3 $\mu$ L
T4 DNA Ligase	1 $\mu$ L
Water	13-14 $\mu$ L

The reaction was transformed into *E. coli* DH5 $\alpha$  cells, cultivated and sequenced. The correct clones were retransformed, and a maxiprep was done to obtain enough plasmid for storage. The stored constructs were sequenced a second time.

#### 4.5 Protein Expression and Purification

##### 4.5.1 SDS-PAGE

Protein purification was monitored with a 10% polyacrylamide BisTris (pH 6.6, 357 mM) gel. The loading dye for the gel was prepared as five times stock solution (5% DTT, 0.02% Bromophenol blue, 30% glycerol, 10% SDS, 250 mM pH 6.8). The running buffer was also prepared as five times stock solution (250 mM TRIS, 250 mM MOPS, 4 mM EDTA, 5 mM Na<sub>2</sub>S<sub>2</sub>O<sub>5</sub>, 0.25% SDS). The gel was run at 200 V and 100 mA. Gels were stained with InstantBlue® solution (Expedeon) for at least 30 minutes, with two subsequent water wash steps.

#### 4.5.2 TEV

#### 4.5.2 TEV

The TEV protease (S219P mutant) plasmid (pRK792) was acquired from David Waugh through Addgene (ID: #8830)<sup>210</sup>. The plasmid was transformed into *E. coli* BL21 (DE3) pLysS cells with Ampicillin and Chloramphenicol as antibiotics. The colonies were washed into 100 mL of LB medium with antibiotics as a pre-culture. The next day 6 L of 2xYT medium were inoculated and grown at 37 °C and 200 rpm with a 25 mm shaking throw to an OD<sub>600</sub>=0.5. The temperature was lowered to 20 °C, and IPTG was added to a final concentration of 0.8 mM. The expression lasted for 16-18 hours. Then the cultures were centrifuged at 4000 rpm for 30 minutes in a J6-MI (Beckman-Coulter) centrifuge. The pellet was resuspended with lysis buffer (50 mM TRIS, 500 mM NaCl, 10 mM Imidazole, 5 mM β-ME, pH of 7.5). After adding DNase and RNase, the suspension was sonified five times for 4 minutes with a Sonopuls HD3200 (Bandelin) and a KE76 probe. The program consisted of 15 s pulses with an amplitude of 20 % alternating with a 30 s pause (3.6 kJ per cycle). The suspension was centrifuged in an Avanti J-30-I (Beckman-Coulter) for one hour at 18,500 rpm. The supernatant was then sterile filtered and loaded onto a Nickel-NTA Column (Cytiva) equilibrated in TEV lysis buffer. The column was washed with 5 CV with the lysis buffer. TEV protease was eluted with the TEV elution buffer (50 mM TRIS, 500 mM NaCl, 300 mM Imidazole, 5 mM β-ME, pH of 7.5). The protein purification was monitored via SDS-PAGE. 10 % glycerol was added, and the protein was concentrated to circa 1 mg/mL, flash-frozen in liquid nitrogen and stored at -80 °C for further use.

#### 4.5.3 AAK1 / CAMK1g

The pET24a\_HLT\_AAK1 or pET24a\_HLT\_CAMK1g construct was transformed in *E. coli* BL21 (DE3) pLysS cells with Kanamycin (50 µg/mL) and Chloramphenicol (34 µg/mL) as antibiotics. The co-expression with the pET13S-A\_λ-phosphatase (Addgene #79748) spectinomycin (50 µg/mL) was used<sup>211</sup>. The expression was performed as described with the TEV protease. The lysis buffer consisted of 50 mM HEPES, 500 mM NaCl, 10 mM imidazole, 5 % (v/v) glycerol, and 0.5 mM TCEP with a pH of 7.5. The nickel column was washed with 5 CVs of lysis buffer and eluted with 30 % of the buffer mentioned above, in which the imidazole concentration was 300 mM. The eluted fractions were pooled and TEV protease was added to compose 1/10<sup>th</sup> of

#### 4.5.4 JNK2

the volume. The eluate was dialyzed overnight against the lysis buffer lacking imidazole and glycerol. A reverse nickel run was performed the next day, and the flow-through and washing steps were collected. The solution was concentrated and loaded onto a HiLoad 26/60 column filled with Superdex 75 (Cytiva) equilibrated with SEC buffer (50 mM HEPES, 300 mM NaCl, 5 % (v/v) glycerol, 0.5 mM TCEP, pH of 7.5). The peak containing the protein of interest were collected, concentrated to 50  $\mu$ M, flash-frozen in liquid nitrogen and stored at -80 °C for further use. The purification was monitored with SDS-PAGE.

#### 4.5.4 JNK2

The pET24a\_HLT\_JNK2 construct was transformed in *E. coli* BL21 (DE3) pLysS cells with Kanamycin and Chloramphenicol as antibiotics. The expression was performed as described with the TEV protease. The purification was done as described with the AAK1. The lysis buffer for JNK2 consisted of 20 mM TRIS, 500 mM NaCl, 15 mM imidazole, 5 mM  $\beta$ -ME pH of 7.5, and the elution buffer was substituted with 300 mM imidazole. The dialysis buffer again lacked the imidazole. The SEC buffer consisted of 20 mM TRIS, 250 mM NaCl, 5 % (v/v) glycerol, and 5 mM  $\beta$ -ME with a pH of 7.5. The peak containing the protein of interest were collected, concentrated to 50  $\mu$ M, flash-frozen in liquid nitrogen and stored at -80 °C for further use. The purification was monitored with SDS-PAGE.

#### 4.5.5 JNK3

The pET24a\_HLT\_JNK3 construct was transformed in *E. coli* BL21 (DE3) pLysS cells with Kanamycin and Chloramphenicol as antibiotics. The expression was performed as described with the TEV protease. The purification was done as described with the AAK1. The lysis buffer for JNK3 contained 50 mM TRIS, 500 mM NaCl, 10 mM imidazole and 5 mM  $\beta$ -ME with a pH of 7.4. In the elution buffer, imidazole concentration was 300 mM. The SEC buffer contained 50 mM HEPES, 100 mM NaCl, 2 mM MgCl<sub>2</sub>, 5 mM  $\beta$ -ME, and 5 % (v/v) glycerol and a pH of 7. The peak containing the protein of interest were collected, concentrated to 50  $\mu$ M, flash-frozen in liquid

#### 4.5.6 DOT1L

nitrogen and stored at -80 °C for further use. The purification was monitored with SDS-PAGE.

#### 4.5.6 DOT1L

The pET24a\_HLT\_DOT1L construct was transformed in *E. coli* BL21 (DE3) pLysS cells with Kanamycin and Chloramphenicol as antibiotics. The expression was performed as described with the TEV protease. The purification was done as described with the AAK1. The lysis buffer for DOT1L contained 20 mM TRIS, 500 mM NaCl, 10 mM imidazole, 5 mM  $\beta$ -ME, and 10% (v/v) glycerol with a pH of 7.8. The imidazole concentration in the elution buffer was 300 mM. The SEC buffer contained 20 mM HEPES, 200 mM NaCl, 1 mM EDTA, and 1 mM DTT with a pH of 7.8. The peak containing the protein of interest were collected, concentrated to 50  $\mu$ M, flash-frozen in liquid nitrogen and stored at -80 °C for further use. The purification was monitored with SDS-PAGE. A cation exchange, as described in the literature<sup>136</sup>, was not necessary for purification.

#### 4.5.7 DYRK1a

The pNIC28-Bsa4\_DYRK1a construct was cloned into *E. coli* BL21 (DE3) pLysS cells with Kanamycin and Chloramphenicol as antibiotics. The expression was performed as described with the TEV protease. The lysis buffer contained 50 mM HEPES, 500 mM NaCl, 10 mM imidazole, and 5 mM  $\beta$ -ME with a pH of 7.5. The nickel column was washed with 5 CVs of lysis buffer and eluted with 30% of the buffer mentioned above except for an imidazole concentration of 300 mM. The eluted fractions were pooled and TEV protease was added to compose 1/10<sup>th</sup> of the volume. The eluate was dialyzed overnight against the lysis buffer lacking imidazole and glycerol. A reverse nickel run was performed the next day, and the flow-through and washing steps were collected. DYRK1a was not digested with TEV for crystallization experiments, and direct SEC chromatography was performed. The SEC buffer contained 50 mM HEPES, 500 mM NaCl, and 5 mM DTT with a pH of 7.5. The fractions containing DYRK1a were pooled, concentrated to 50  $\mu$ M, flash-frozen in liquid nitrogen, and stored at -80 °C. The purification was monitored with SDS-PAGE.

## 4.5.8 IDO1

The pET15bT\_IDO1 construct was transformed into *E. coli* Rosetta 2 (DE3) pLysS cells (Novagen) with Ampicillin and Chloramphenicol as antibiotics. The colonies were washed into 100 mL of LB medium with antibiotics as a pre-culture. The next day 6 L of TB medium were inoculated and grown at 37 °C and 200 rpm with a 25 mm shaking throw to an OD<sub>600</sub>=0.5. The temperature was lowered to 18 °C, and IPTG was added to a final concentration of 0.5 mM. For heme incorporation, 5-ALA was added to a final concentration of 0.5 mM. The expression lasted 15 hours. The cultures were centrifuged at 4000 rpm for 30 minutes in a J6-MI (Beckman-Coulter) centrifuge. The pellet was resuspended with lysis buffer (25 mM TRIS, 150 mM NaCl, 10 mM Imidazole, 5 mM β-ME, pH of 7.4). After adding DNase and RNase, the suspension was sonified five times for 4 minutes with a Sonopuls HD3200 (Bandelin) and a KE76 probe. The program consisted of 15 s pulses with an amplitude of 20 % alternating with a 30 s pause (3.6 kJ per cycle). The suspension was centrifuged in an Avanti J-30-I (Beckman-Coulter) for one hour at 18,500 rpm. Afterward, the supernatant was sterile filtered and loaded onto a Nickel-NTA Column (Cytiva) equilibrated in lysis buffer. The column was washed with 5 CV with the lysis buffer. Protein was eluted with 40% of elution buffer, which was the lysis buffer supplemented with 300 mM imidazole. The fractions were pooled, and TEV was added and left overnight to digest. No dialysis was performed to prevent heme from diffusing out of the pocket. To subject the protein to a reverse nickel run, the protein was buffer exchanged into lysis buffer on a HiPrep Desalting 26/10 column (Cytiva). After the reverse nickel run, four times excess of Hemin was added and left overnight. Before loading onto the ÄKTA system, the solution was filtered through a 0.22 μm filter. As the next step, an SEC chromatography with 25 mM TRIS, 150 mM NaCl, and 5 mM β-ME was performed. The fractions containing IDO1 were pooled, concentrated to 50 μM, flash-frozen in liquid nitrogen, and stored at -80 °C. The purification was monitored with SDS-PAGE. The absorption at 280 nm and 405 nm was measured to estimate the heme incorporation, and a quotient Q ( $Q = \frac{280}{405}$ ) was calculated with  $Q = 2.2 = 100\%$  <sup>166</sup>.

#### 4.5.9 BIRC5

#### 4.5.9 BIRC5

The pET24a\_HLT\_BIRC5 construct was transformed into *E. coli* BL21 (DE3) pLysS cells with Kanamycin and Chloramphenicol as antibiotics. The expression was performed as described with the TEV protease. The purification was done as described with AAK1. The lysis buffer for BIRC5 contained 5 mM TRIS, 500 mM NaCl, 10 mM imidazole, 5 mM  $\beta$ -ME, and 5% (v/v) glycerol with a pH of 7.9. The imidazole concentration in the elution buffer was 250 mM. 80  $\mu$ M ZnSO<sub>4</sub> was added to the dialysis buffer. The SEC buffer contained 25 mM TRIS, 100 mM NaCl, 5 mM  $\beta$ -ME, and 5% (v/v) glycerol with a pH of 7.5. The protein was separated on a GE HiLoad 26/60 with Superdex 75 pg. The peak containing the protein of interest were collected, concentrated to 50  $\mu$ M, flash-frozen in liquid nitrogen and stored at -80 °C for further use. The purification was monitored with SDS-PAGE.

#### 4.5.10 GSK3

The various GSK constructs were transformed into either different *E. coli* strains or *V. natriegens* Vmax (BioCat). Different expression conditions were tested (see section 2.2.8). The purification of soluble fraction or inclusion bodies was done on a HisTrap column (Cytiva) with standard buffers<sup>174</sup>. Samples were subjected to Western Blot and ELISA either crude or after Ni-affinity chromatography.

#### 4.5.11 Inclusion body preparation

After centrifugation at 4000 rpm for 30 minutes, the pellet was resuspended in lysis buffer (50 mM TRIS, 500 mM NaCl, 5 mM  $\beta$ -ME, 1 mM PMSF, pH of 8), and 300 mg/L Lysozyme, DNase and RNase were added and incubated for 20 min at RT. The suspension was subjected to the ultrasound as described above. The suspension was then centrifuged for one hour at 22,000 g. The supernatant was removed and used for further analysis. The pellet was resuspended with a handheld homogenizer in buffer Wash1 (50 mM TRIS, 500 mM NaCl, 2 M urea, 2% Triton X100, 5 mM  $\beta$ -ME, pH of 8). After centrifugation at 22,000 g for 30 min, the supernatant was discarded. The washing step was repeated. The third washing was done with buffer Wash2 (Wash1 without

#### 4.5.12 Western Blot and ELISA

Triton X100 and urea), following another 30 min centrifugation step at 22,000 g. The remaining pellet is resuspended in Extraction buffer (50 mM TRIS, 500 mM NaCl, 6 M guanidine, 5 mM imidazole, 5 mM  $\beta$ -ME, pH of 8) and stirred for one hour at RT. The solution was then centrifuged at 100,000 g for one hour, and the supernatant was filtered afterward. The solution was loaded onto a HisTrap Ni-NTA column (Cytiva) equilibrated in the Extraction buffer. The column is subsequently washed with 5 CVs of buffer Wash3 (50 mM TRIS, 500 mM NaCl, 6 M urea, 20 mM imidazole, 5 mM  $\beta$ -ME, pH of 8). A linear gradient lasting for 30 min with buffer Wash4 (50 mM TRIS, 500 mM NaCl, 20 mM imidazole, 5 mM  $\beta$ -ME, pH of 8) is applied. The protein was eluted with a linear gradient of 50 mM TRIS, 500 mM NaCl, 300 mM Imidazole, and 5 mM  $\beta$ -ME with a pH of 8<sup>212, 213</sup>.

#### 4.5.12 Western Blot and ELISA

After an SDS-PAGE, the gel was blotted onto an Amersham Protran 0.45  $\mu$ m (Cytiva) nitrocellulose membrane. The protein was transferred via wet blotting in 25 mM TRIS, 190 mM glycine, and 20% methanol at a pH of 8.3 for 90 minutes at 30 V. The ELISA was performed with the following protocol:

5 % BSA in TBS	60 min
TBS	3x 30 s
TBS	2x 5 min
Antibody 1	60 min
TBS	3x 5 min
Antibody 2	60 min
TBS	2x 5 min
TBS+ 0.1 % Tween 20	2x 5 min
Detection buffer	2x 5 min
Substrate buffer	2-10 min



#### 4.6 Fragment Library Preparation

The antibodies were diluted at 1:5000 into TBS with 5% BSA. Antibody one was a murine, monoclonal anti-6xHis antibody (Sigma). A caprine, monoclonal anti-mouse antibody coupled with alkaline phosphatase (Sigma), was used as antibody two. The detection buffer contained 50 mM TRIS, 50 mM NaCl, and 25 mM MgCl<sub>2</sub> with a pH of 9.5. The substrate buffer contained the detection buffer with BCIP, and NBT was added to a final concentration of 520 μM and 400 μM, respectively. The reaction was stopped by adding TBS with 2 mM EDTA.

#### 4.6 Fragment Library Preparation

All fragments were stored at -20 °C as solids in argon flushed glass bottles. For NMR screening, fragments were weighed and dissolved in DMSO-d<sub>6</sub> to a stock concentration of 300 mM. Mother plates were prepared with a 100 mM concentration. From these plates, daughter plates were prepared with the fragments mixed in the plates to a concentration of 50 mM for each fragment. Daughter plates were opened a maximum of eight times. The plates were argon flushed and sealed with an HT121TS (HTA) and Foil Heat Seal (Biozym).

#### 4.7 STD-NMR

For NMR experiments, proteins were buffer exchanged to 100 mM sodium phosphate, 250 mM NaCl, and 2 mM MgCl<sub>2</sub> with a pH of 7. For long-time measurements, like the protein <sup>1</sup>H spectrum, the buffer exchange was performed twice to reduce residual buffer peaks. All experiments were performed on a Bruker Avance III HDX 700 instrument, equipped with a 5 mm Prodigy TCI cryo-probehead. A standard <sup>1</sup>H-NMR experiment with 1 k scans and water suppression through presaturation was performed for every protein. The STD experiments adapted the pulse sequence published by Mayer et al.<sup>30, 31, 33</sup>. The on-resonance frequency, which was determined from the <sup>1</sup>H-NMR spectra of the protein, was between 0.5-0.6 ppm. 40 ppm was used as the off-resonance frequency. For an interleaved acquisition of the on- and off-resonance, a pseudo-2D scheme was applied. The saturation was done by Gaussian pulses with a length of 50 ms and 60 dB of attenuation, done with an interpulse delay of 1 ms leading to an excitation bandwidth of about 42 Hz. The screening was done

#### 4.8 ITC

with 16 scans of on- and off-resonance scans each, with a 3 second saturation time. The samples contained final concentrations of 20  $\mu\text{M}$  protein, 10% ( $\text{V}/\text{V}$ )  $\text{DMSO-d}_6$ , and 1 mM of each of the two fragments. A  $^1\text{H}$  NMR experiment was performed for each compound to act as a reference spectrum in the STD experiments. NMR experiments were carried out at 25  $^\circ\text{C}$ . Spectra were processed and analyzed with TopSpin v 4.1.1 (Bruker).

#### 4.8 ITC

The protein was buffer exchanged into 50 mM HEPES, 100 mM NaCl, 2 mM  $\text{MgCl}_2$ , 1 mM TCEP, and a pH of 7.4. The protein was concentrated to 50-100  $\mu\text{M}$ , and 5% ( $\text{V}/\text{V}$ ) DMSO was added. The fragments were dissolved as a 100 mM stock solution in DMSO and diluted with buffer to 5 mM. A MicroCal iTC200 (Malvern) instrument was used with the measuring cell set to 25  $^\circ\text{C}$  while the jacket was set to 15  $^\circ\text{C}$ . After a 120 s initial delay, the first 0.5  $\mu\text{L}$  over 2 s was injected. After 180 s of equilibration, 19 injections with 2  $\mu\text{L}$  over 4 s were performed. 1000 rpm were used for stirring, and the reference heat rate was set to 10  $\mu\text{cal}/\text{s}$ . During equilibration, the measurements were aborted if the measuring cell did not reach a heat rate above 9  $\mu\text{cal}/\text{s}$ . After each experiment, the cell and syringe were washed at least three times thoroughly with water. The syringe was dried with methanol. After each day, the cell was cleaned with a 14% solution of Decon 90 (Decon) for 15 minutes. Before each run, the cell was primed with buffer twice. Thermograms were processed and analyzed with the Origin 7 plug-in Autocal.

#### 4.9 GSH Stability Assay

The performed GSH-stability assay for the fragments was based on the established assay by G. M. Keserű et al.<sup>181, 184</sup> for heterocyclic electrophilic fragments. A 500  $\mu\text{M}$  solution of the fragment in PBS buffer pH of 7.4, 10% acetonitrile, and 200  $\mu\text{M}$  Indoprofen as an internal standard was added to 10 mM GSH solution in a 1:1 ratio. The reaction temperature was 40  $^\circ\text{C}$ . The mixture was analyzed by HPLC (Column: Phenomenex Kinetex 2.6u C8 100A 150 x 4,6 mm) with an injection volume of 5  $\mu\text{L}$ , a flow rate of 0.5 mL/min at 23  $^\circ\text{C}$  after 0, 1, 2, 4, 8, 12, and 24 hours.

#### 4.10 Crystallization

The reaction of fragments with GSH was determined by measuring the compound's decreasing area under the curve (AUC) relative to the internal standard Indoprofen. The gradient was chosen depending on the polarity of fragments to be either one of the following:

HPLC-run normal (28 min): 0 min: 40 % MeOH 60 % phosphate buffer pH of 2.3; 15 min: 85 % MeOH 15 % phosphate buffer pH of 2.3; 20 min: 85 % MeOH 15 % phosphate buffer pH of 2.3; 22 min: 40 % MeOH 60 % phosphate buffer pH of 2.3; 28 min: 40 % MeOH 60 % phosphate buffer pH of 2.3

HPLC-run for polar fragments (28 min): 0 min: 10 % MeOH 90 % phosphate buffer pH of 2.3; 15 min: 85 % MeOH 15 % phosphate buffer pH of 2.3; 20 min: 85 % MeOH 15 % phosphate buffer pH of 2.3; 22 min: 10 % MeOH 90 % phosphate buffer pH of 2.3; 28 min: 10 % MeOH 90 % phosphate buffer pH of 2.3

#### 4.10 Crystallization

##### 4.10.1 DYRK1a

Crystals for DYRK1 a were obtained with the sitting drop vapor diffusion method in Cryschem 24 well plates (Hampton). The protein was buffer exchanged into 25 mM HEPES, 500 mM NaCl, and 5 mM DTT with a pH of 7.5 and was concentrated to 12 mg/mL. The reservoir solution contained 100 mM TRIS, 100 mM Li<sub>2</sub>SO<sub>4</sub>, and 34 % (v/v) PEG 300 with a pH of 8.5. 4 μL of the protein solution were mixed with 2 μL of the reservoir solution, and the complete drop contained 1-2 mM of AMP-PCP. The reservoir was filled with 500 μL. Crystals grew after one to four weeks at 4 °C. The first-generation crystals were crushed and used for streak seeding for sufficiently uniform crystals. Octahedral crystals were fished and soaked into reservoir solution with 1 mM compound or 5 mM fragment for 24 to 36 h. The final DMSO concentration in the soaking drop was 5 % (v/v). Crystals were fished and flash-frozen in liquid nitrogen.

#### 4.10.2 JNK3

Crystals of JNK3 were obtained with the sitting drop vapor diffusion method in MRC Maxi 48 well plates as described (Swissci) before. The protein was concentrated to 7-9 mg/mL in the SEC buffer. The solution was supplemented with 1 mM AMP-PCP and 0.4 mM Zwittergent 3-14. The reservoir solution contained 0.1 M BisTris pH of 6-6.5 and 27 % PEG 3350. The plates were set up with a 200  $\mu$ L reservoir, and a 2  $\mu$ L protein drop was mixed with 2  $\mu$ L of reservoir solution. Rectangular cuboid crystals grew after one day and into sufficient size within a week. For soaking, a saturated solution of the fragment in cryoprotectant was used. The cryo-protectant consisted of a reservoir solution supplemented with 10% ethylene glycol and 15% glycerol. The compounds were soaked for 24 h and flash-frozen in liquid nitrogen.

#### 4.10.3 Data Collection and Processing

Protein crystals were measured at the SLS beamlines X06DA (PXIII) and X06SA (PXI). Integration and indexing were done with XDS (version Feb 5, 2021)<sup>214</sup>. Phenix (v19.2)<sup>210</sup> and Phaser in the CCP4 suite (v7.1)<sup>211</sup> were used for refinement, and the PDB entry 2VX3<sup>120</sup> was used as the search model for DYRK1a. PDB entry 4X21<sup>63</sup> was used as the search model for the JNK3 structures. Modelbuilding was done with COOT (v0.9.5)<sup>215</sup>.

#### 4.11 ESP-Plots and $V_{\max}$ -values

The ESP-plots and  $V_{\max}$  assessment calculations were performed using TURBOMOLE Version 7.4.1<sup>216</sup>. A triple- $\zeta$  basis set (def2-TZVPP) was used throughout the study. MP2 calculations were performed in combination with the resolution of identity (RI) technique and the frozen core approximation. The frozen core orbitals were defined using default settings by which all orbitals possessing energies below  $-3.0$  au were considered core orbitals. The SCF convergence criterion was increased to 10–8 Hartree for all calculations. Custom python scripts were used for plotting.

#### 4.12 Figures and Tables

Figures of proteins were created with PyMOL 2.3.3 (Schrödinger). Tables were created with Excel (Microsoft). Chemical structures were created with ChemDraw 18 (Perkin Elmer). Figures were assembled with Photoshop CS6 (Adobe) and PowerPoint (Microsoft).

#### 4.13 Author Contributions

F.M. Boeckler envisioned the overall idea of the Study. The experiments were planned and executed by me. Datasets were processed and analyzed by me. M. Zimmermann calculated and plotted the ESP-plots and  $V_{\max}$ -values. He performed the docking studies in this work. W. Jason Stahlecker helped with the SAR-ITC measurements, and with the crystallization of DYRK1a and JNK3, he performed most of the electron density mapping, model building, and refinement.

## 5 Zusammenfassung / Summary (German)

Fragment basierte Wirkstoffforschung (FBDD) ist in den letzten Jahren zu einer Standardmethodik geworden. In der industriellen und akademischen Forschung wurden vielfältige Fragment-Bibliotheken entwickelt und erfolgreich eingesetzt, um neuartige Medikamente zu entwickeln<sup>9, 217</sup>. Das erste zugelassene Medikament, dass auf diese Weise gefunden wurde war Vemurafenib in 2011<sup>10, 93</sup>. Im industriellen Kontext dient dieser Ansatz der Generierung neuer Startpunkte in der Wirkstoffentwicklung. Die akademische Forschung richtet ihr Augenmerk auf die zusätzlich gewonnenen Erkenntnisse über die molekularen Prozesse bei der Bindung von Inhibitoren und Fragmenten an Proteine. Durch die Aufklärung von Bindungsmodi kleiner Moleküle kann die Abschätzung von Interaktionsstärken leichter fallen, da weniger konkurrierende Interaktionen das Bild verfälschen können. Ein weiterer Vorteil der Fragmente sind die passgenauere Bindung an das Zielprotein. Durch die zuvor erwähnte geringe Anzahl an möglichen Wechselwirkungen sind die Fragment Hits in ihrer Affinität schwächer, normiert auf die Anzahl der Atome jedoch oft besser, als Hits generiert durch HTS-Verfahren<sup>217</sup>. Durch die geringere Größe der Fragmente im Vergleich zu traditionellen Lead-Verbindungen, können die Fragment-Bibliotheken einen größeren Teil des chemischen Raumes mit einer überschaubaren Zahl an Verbindungen abdecken. Um eine effiziente Abdeckung eines bestimmten Bereichs zu gewährleisten, muss beim Design dieser Bibliotheken große Sorgfalt auf die Auswahl der Fragmente gelegt werden. Wenn ein Fragment Screening erfolgreich ist, kann es durchaus vorkommen, dass mehrere Fragmente mit komplett verschiedener Struktur als Hits identifiziert werden. Diese können diverse Bindungsmodi besitzen, was wiederum ausgenutzt werden kann, um Fragmente entweder, zu verbinden (Fragment-linking), zu verschmelzen (Fragment-merging), oder durch klassische medizinische Chemie das Pharmakophor zu vergrößern (Fragment-growing). So können Fragmente zu kompletten Inhibitoren oder Aktivatoren erweitert werden<sup>92</sup>.

Der zweite Fokus dieser Arbeit liegt auf der Halogen-Bindung. Analog zu der Wasserstoffbrückenbindung, beruht auch diese Interaktion zum größten Teil auf elektrostatischer Anziehung.

Durch eine Vielzahl an Kristallstrukturen von organischen Molekülen und theoretischen Überlegungen lässt sich das Verhalten von Halogenen am besten durch eine anisotrope Elektronendichte erklären. Diese Anisotropie kann durch die Analyse der natürlichen Bindungorbitale (NBO) erklärt werden<sup>53</sup>. Die resultierenden Orbitale befinden sich in einer  $s^2p_x^2p_y^2p_z^1$ -Konfiguration (mit der Kohlenstoff-Halogen-Bindung in der z-Achse). Diese Orbital-konfiguration resultiert in einer erhöhten Elektronendichte in einem äquatorialen Gürtel in Kombination mit einem Fleck oder Loch mit relativem Elektronenmangel in der Verlängerung der Halogen-Kohlenstoff-Bindung. Dieser Unterschied in der Dichte wird im Allgemeinen und im Folgenden als  $\sigma$ -Hole bezeichnet. Durch die Partialladungen am Halogen sind elektrostatische Wechselwirkungen entweder im 180°- oder im 90°-Winkel möglich, wobei das Halogen den Scheitelpunkt bildet. Potenzielle Wechselwirkungen in einem Winkel von 90° sind aufgrund des höheren sterischen Anspruchs durch andere Substituenten in ortho-Position zum Halogen seltener anzutreffen als Wechselwirkungen mit annähernd 180°<sup>1</sup>.

Die in dieser Arbeit verwendete Halogen-Enriched-Fragment-Library (HEFLib) wurde von J. Heidrich designend<sup>7, 11, 28, 102</sup>. Die experimentelle Charakterisierung, Evaluation und Anwendung wurde basierend auf einen GSH-Stabilitäts-Assay, gefolgt von einem dreistufigen Screening Prozess durchgeführt. Die chemische Stabilität wurde exemplarisch geprüft, für Fragmente, die entweder als instabil beschrieben wurde oder Ähnlichkeiten zu instabilen Verbindungen besitzen. Dabei wurde drei Fragmente identifiziert, die eine Halbwertszeit von unter 20 Stunden haben und damit signifikante Mengen an Fragment während des STD-NMR Screenings reagieren. Die Fragmente 1223 und 1255 waren zuvor schon als mögliche kovalent bindende Strukturelemente postuliert worden<sup>182, 183</sup>. Fragment 1234 war in der Literatur noch nicht als instabil beschrieben worden.

Für das Screening wurden acht Proteine (AAK1, BIRC5, CAMK1G, DOT1L, DYRK1a, IDO1, JNK2 und JNK3) aus unterschiedlichsten Gesichtspunkten ausgewählt, dabei wurde vor allem auf eine Durchmischung von strukturell unterschiedlichen und sehr ähnlichen Proteinen geachtet. Damit wird einerseits die Anwendbarkeit der Bibliothek auf eine breite Basis an Proteinen gezeigt werden. Andererseits sollte das Screening von sehr ähnlichen Proteinen die Möglichkeit für Kreuzvalidierung der verwendeten Assays beitragen.

Durch die große Variabilität im Verwandtschaftsgrad der Kinasen, insbesondere von JNK2, JNK3 und AAK1, können Rückschlüsse über die Selektivität von Fragmenten gezogen werden. Gleichzeitig können Vermutungen zu den Bindungsmodi aufgestellt werden, die in einem späteren Schritt überprüft werden können.

Im ersten Schritt des Screenings wurde STD-NMR als Methode zur Hit Identifikation verwendet. Dabei wurden insgesamt mehrer hundert Hit-Events ermittelt. Für alle Proteine bis auf DOT1L konnten Fragmente identifiziert werden, die an ein einziges der getesteten Proteine binden. Im Gegensatz dazu binden fünf Fragmente an nahezu alle (sieben von acht) Proteine und zwei Fragmente (0459 und 1234) binden an alle getesteten Proteine. Die Hit-Events im NMR von 1234 könnten auch teilweise auf die Reaktivität des Fragments mit Cysteinen zurückzuführen sein. Bei der Analyse der Hit Fragmente und ihrer Eigenschaften viel besonders auf, dass Fragmente mit einer geringeren Löslichkeit (höherer SlogP-Wert) häufiger binden. Dazu kommt, dass Fragmente mit einem größeren  $\sigma$ -Hole, gemessen an ihrem  $V_{\max}$ -Wert, häufiger als Hit identifiziert wurden. Andere Eigenschaften wie molekulare Masse, polare Oberfläche, Anzahl an H-Brücken Donoren oder Akzeptoren hatte keinen Einfluss. Diese Entdeckung ist ein erstes Indiz, dass Halogen-Bindungen in den Bindungsmodi vorkommen. Dieser Verdacht wurde erhärtet, da ein Welch-t-Test einen signifikanten Unterschied im  $V_{\max}$ -Wert von bindenden und nicht bindenden Fragmenten nachweisen konnte.

Um aus den 344 Events die vielversprechendsten Kandidaten auszuwählen, wurden Hits nach mehreren Kriterien priorisiert. Da der  $V_{\max}$ -Wert einen Einfluss auf die Hit-Events zu haben schien, wurden ESP-Plots erstellt, mit denen ein besserer Überblick über die elektrostatischen Eigenschaften der Fragmente zu bekommen. Da der Großteil der Verbindungen hohe Anteile an planaren Strukturen haben, sind zweidimensionale Plots ausreichend die elektrostatischen Eigenschaften der Moleküle abzuschätzen. Diese Plots dienen einer ersten Klassifizierung. Weiterhin wurden die STD-Spektren zu Rate gezogen, da ein starkes Protonensignal in direkter Nachbarschaft zu einem Halogen auf eine etwaige Halogen-Bindung hindeutet. Gleichzeitig wurden Fragmente ausgewählt, die nur an bestimmte Proteine gebunden hatten. Zum Beispiel ist hier Fragment 9595 erwähnt, das an alle fünf Kinasen gebunden hat, oder Fragment 9612, das ausschließlich an JNK2 und JNK3 gebunden hat.



Die Hit-Events wurden mittels ITC einer Validierung und Charakterisierung unterzogen. 132 Fragment-Protein Kombinationen wurden mit Hilfe der ITC vermessen. Dabei zeigten 25 Fragmente eine deutlich messbare Affinität ( $K_D < 800 \mu\text{M}$ ). Weitere fünf Fragmente konnten verifiziert werden, die Affinitäten sind jedoch zu gering, um sie quantifizieren zu können. Aufgrund der limitierten Löslichkeit der Proteine und der Fragmente, konnten in der ITC nur mikromolare Affinitäten bestimmt werden. Dies führt in der Validierung automatisch dazu, dass nur Fragmente mit einer Ligand Efficiency (LE) über 0,3 sicher erkannt werden. Mit einem  $K_D = 6 \mu\text{M}$  weist Fragment 9595 zur AAK1 die größte Affinität auf. Weitere sieben Fragment-Protein Kombinationen weisen Affinitäten im zweistellig mikromolaren Bereich auf. Ein großer Teil der validierten Fragmente zeigt  $LE > 0,5$  und können somit als Hits mit gutem Potenzial zur Weiterentwicklung angesehen werden. Von den validierten Fragmenten wurden drei ausgewählt zur weiteren Charakterisierung.

Die Bindung von Fragment 9595 (4-Chloro-5*H*-pyrrolo[3,2-*d*]pyrimidin) an AAK1 wurde intensiver untersucht, da dieses die stärkste Affinität zu allen Kinasen hat, und das Fragment kein klassisches Hinge-Binding Motiv zeigt. Die thermodynamischen Parameter der Bindung zeigen einen Bindungsmodus hauptsächlich bestimmt durch enthalpische Effekte. Um mehr über den Bindungsmodus zu erfahren, wurden mehrere analoge Fragmente getestet. Aus dieser kleinen SAR-Serie konnte geschlossen werden, dass das Chloratom den größten Einfluss auf die Affinität hat, gefolgt vom Pyrrol-Stickstoff. Damit ist das Vorhandensein einer Halogen Bindung naheliegend.

Als zweites Fragment wurde der Bindungsmodus von Fragment 0482 (6-Bromo-3*H*-[1,2,3]triazolo[4,5-*b*]pyridin) mit der DYRK1a untersucht. Die bestimmte Affinität lag mit  $530 \mu\text{M}$  hoch und die SAR-Serie ließ keine eindeutigen Schlussfolgerungen zu. Nachdem der Protein-Fragment Komplex kristallisiert und die Struktur gelöst werden konnte, konnte klar gezeigt werden, dass das Fragment zwei Bindungsstellen an der DYRK1a besetzt. Mit diesem Wissen konnten auch die Daten aus der SAR-Reihe erklärt werden. Das Fragment bindet einerseits in der ATP-Bindungstasche, dabei bildet das Brom eine Halogen-Bindung zu einem Backbone-Sauerstoff, während die Triazol-Stickstoffe H-Brücken zu einem Lysin und einem gebundenen Wasser bilden. Der zweite Bindungsmodus befindet sich an dem sogenannten EGFR Dimerization Interface (EDI) auf der C-terminalen Domäne.

In diesem Bindungsmodus sind alle vier Stickstoffe im Molekül für den Bindungsmodus essenziell, sodass der Austausch zu Kohlenstoff, den zweiten Bindungsmodus unmöglich macht. Der Bindungsmodus in der ATP-Bindungsstelle wurde zuvor mit ähnlichen Molekülen an anderen Kinasen kristallisiert. Die Bindung am EDI wurde zuvor für andere Kinasen postuliert, aber noch nicht bewiesen, zumal die DYRK1a keine EGFR ähnliche Kinase ist und es nicht bekannt ist, welche Funktion dieser Bereich in der Interaktion mit anderen Proteinen einnimmt.

Als drittes Fragment wurde die Bindung von 9612 (2-Amino-5-bromonicotinsäure) an JNK2 und JNK3 untersucht. Dieses Fragment besticht durch die deutlich stärkere Affinität zur JNK2 im Vergleich zur JNK3. Die Affinität und Selektivität konnten weiter gesteigert werden durch den Austausch des Brom-Atoms zu einem Iod. Gleichzeitig konnte der Bindungsmodus beider Verbindungen mit JNK3 aufgeklärt werden. Die Fragmente binden in der ATP-Tasche und zeigen klassische Hinge-Binding Kontakte über H-Brücken. Während in der Iod-Variante nur ein Bindungsmodus gebunden ist, liegen in der Brom-Variante zwei alternative Bindungsmodi des Fragments vor. In Bindungsmodus A bildet das Halogen (sowohl Brom als auch Iod) eine Halogen-Bindung zu einem Wasser aus, welches wiederum durch Q155 koordiniert wird.

In Bindungsmodus B bildet das Brom eine Halogen-Bindung zum Gatekeeper (M146) aus. Der Bindungsmodus mit der JNK2 bleibt ungeklärt, da es nicht möglich war diese Kinase zu kristallisieren. Um dennoch einen Einblick in die Ursache für die Selektivität der Fragmente zu bekommen, wurde die einzige Aminosäure, die die JNK2 von der JNK3 in der Bindungstasche unterscheidet, mutiert. Besagte Aminosäure ist das M115 in der JNK3, das einem Leucin (L77) in der JNK2 entspricht. Dieser Unterschied reicht aus um eine sechs bis neunfache Selektivität der Iod-Variante zu erzeugen. Daher ist anzunehmen, dass die Carboxylgruppe eine deutliche Wechselwirkung mit Seitenkette des Gatekeepers eingeht und die Umgebung dieser Aminosäure und damit seine Konformation maßgeblichen Einfluss auf die Affinität ausüben. Zudem war die Sekundärstruktur der JNK3 in den Fragmentkomplexen deutlich verändert, im Vergleich zur AMP-PCP Struktur. Mit einem der beiden Fragmente verändert sich die Konformation des  $\beta$ 1-Faltblatts und P-loops zu einem durchgehenden Loop. Diese deutliche Änderung in der Proteinstruktur wurde zuvor nicht in diesem Maße beobachtet und könnte zu Inhibitoren mit bisher nicht für möglich gehaltenen Bindungsmodi führen. In der gefundenen Konformation zeigen Aminosäuren in

Richtung der Bindungstasche, die zuvor außerhalb der Reichweite für Protein-Inhibitor Interaktionen lagen. Noch ist nicht geklärt wie weit diese Konformationsänderung in der aktiven JNK3 vorkommt, da nur Strukturen mit einem Liganden bekannt sind und keine Apostruktur.

Diese Arbeit liefert die Grundlage für viele neue Anknüpfungspunkte, besonders die Evaluierung der noch nicht validierten Hits könnte noch viele weitere Fragmente mit guten Eigenschaften für die Weiterentwicklung zu einer Lead-Verbindung aufweisen. Die drei intensiver untersuchten Fragmente können die Grundlage für neue Inhibitoren werden. Besonders die zweite Bindungsstelle and der DYRK1a bietet die Möglichkeit die Funktion dieses Bereichs aufzuklären, wenn der Compound weiterentwickelt wird. Obwohl es bisher nicht gelungen ist, die AAK1 mit dem Fragment 9595 zu kristallisieren und den Bindungsmodus aufzuklären, bestehen hier eventuell die Möglichkeiten mit neuen Proteinkonstrukten Einblicke zu erhalten. Ein Punkt, der in dieser Arbeit nicht mehr durchgeführt werden konnte, sind die restlichen validierten Fragmente mit der JNK3 und der DYRK1a zu kristallisieren, dadurch bestünde die Möglichkeit Fragmente mit aufgeklärten Bindungsmodi zu linken oder zu mergen.

## 6. Literature

1. Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M., Principles and applications of halogen bonding in medicinal chemistry and chemical biology. *J. Med. Chem.* **2013**, *56* (4), 1363-88.
2. Carter, M.; Ho, P. S., Assaying the Energies of Biological Halogen Bonds. *Cryst Growth Des* **2011**, *11* (11), 5087-5095.
3. Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimagi, A.; Resnati, G.; Terraneo, G., The Halogen Bond. *Chem. Rev.* **2016**, *116* (4), 2478-601.
4. Politzer, P.; Murray, J. S.; Clark, T., Halogen bonding and other sigma-hole interactions: a perspective. *Phys. Chem. Chem. Phys.* **2013**, *15* (27), 11178-89.
5. Jacquemard, C.; Kellenberger, E., A bright future for fragment-based drug discovery: what does it hold? *Expert Opin. Drug Discov.* **2019**, *14* (5), 413-416.
6. Keseru, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D., Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.* **2016**, *59* (18), 8189-206.
7. Heidrich, J.; Sperl, L. E.; Boeckler, F. M., Embracing the Diversity of Halogen Bonding Motifs in Fragment-Based Drug Discovery-Construction of a Diversity-Optimized Halogen-Enriched Fragment Library. *Front. Chem.* **2019**, *7* (9), 9.
8. Giordanetto, F.; Jin, C.; Willmore, L.; Feher, M.; Shaw, D. E., Fragment Hits: What do They Look Like and How do They Bind? *J. Med. Chem.* **2019**, *62* (7), 3381-3394.
9. Jahnke, W.; Erlanson, D. A.; de Esch, I. J. P.; Johnson, C. N.; Mortenson, P. N.; Ochi, Y.; Urushima, T., Fragment-to-Lead Medicinal Chemistry Publications in 2019. *J. Med. Chem.* **2020**, *63* (24), 15494-15507.
10. Sancineto, L.; Massari, S.; Iraci, N.; Tabarrini, O., From small to powerful: the fragments universe and its "chem-appeal". *Curr. Med. Chem.* **2013**, *20* (11), 1355-81.
11. Wilcken, R.; Liu, X.; Zimmermann, M. O.; Rutherford, T. J.; Fersht, A. R.; Joerger, A. C.; Boeckler, F. M., Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J. Am. Chem. Soc.* **2012**, *134* (15), 6810-8.

## 6. Literature

12. Shi, Y.; von Itzstein, M., How Size Matters: Diversity for Fragment Library Design. *Molecules* **2019**, *24* (15), 2838.
13. Murray, C. W.; Rees, D. C., The rise of fragment-based drug discovery. *Nat. Chem.* **2009**, *1* (3), 187-92.
14. Wilcken, R.; Zimmermann, M. O.; Lange, A.; Zahn, S.; Kirchner, B.; Boeckler, F. M., Addressing Methionine in Molecular Design through Directed Sulfur-Halogen Bonds. *J. Chem. Theory. Comput.* **2011**, *7* (7), 2307-15.
15. Lechner, C.; Flasshoff, M.; Falke, H.; Preu, L.; Loac, N.; Meijer, L.; Knapp, S.; Chaikuad, A.; Kunick, C., [b]-Annulated Halogen-Substituted Indoles as Potential DYRK1A Inhibitors. *Molecules* **2019**, *24* (22).
16. Voth, A. R.; Hays, F. A.; Ho, P. S., Directing macromolecular conformation through halogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (15), 6188-93.
17. Wood, D. J.; Lopez-Fernandez, J. D.; Knight, L. E.; Al-Khawaldeh, I.; Gai, C.; Lin, S.; Martin, M. P.; Miller, D. C.; Cano, C.; Endicott, J. A.; Hardcastle, I. R.; Noble, M. E. M.; Waring, M. J., FragLites-Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *J. Med. Chem.* **2019**, *62* (7), 3741-3752.
18. Czapinska, H.; Winiewska-Szajewska, M.; Szymaniec-Rutkowska, A.; Piasecka, A.; Bochtler, M.; Poznanski, J., Halogen Atoms in the Protein-Ligand System. Structural and Thermodynamic Studies of the Binding of Bromobenzotriazoles by the Catalytic Subunit of Human Protein Kinase CK2. *J. Phys. Chem. B* **2021**, *125* (10), 2491-2503.
19. Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J. M.; Hartmann, G.; Isshiki, Y.; Morikami, K.; Shimma, N.; Haap, W.; Banner, D. W.; Diederich, F., Halogen bonding at the active sites of human cathepsin L and MEK1 kinase: efficient interactions in different environments. *ChemMedChem* **2011**, *6* (11), 2048-54.
20. Mondal, S.; Manna, D.; Raja, K.; Mugesh, G., Halogen Bonding in Biomimetic Deiodination of Thyroid Hormones and their Metabolites and Dehalogenation of Halogenated Nucleosides. *Chembiochem* **2020**, *21* (7), 911-923.

## 6. Literature

21. Kortagere, S.; Ekins, S.; Welsh, W. J., Halogenated ligands and their interactions with amino acids: implications for structure-activity and structure-toxicity relationships. *J. Mol. Graph. Model.* **2008**, *27* (2), 170-7.
22. Carlsson, A. C.; Scholfield, M. R.; Rowe, R. K.; Ford, M. C.; Alexander, A. T.; Mehl, R. A.; Ho, P. S., Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds. *Biochemistry* **2018**, *57* (28), 4135-4147.
23. Battistutta, R.; Mazzorana, M.; Sarno, S.; Kazimierczuk, Z.; Zanotti, G.; Pinna, L. A., Inspecting the structure-activity relationship of protein kinase CK2 inhibitors derived from tetrabromo-benzimidazole. *Chem. Biol.* **2005**, *12* (11), 1211-9.
24. Bayse, C. A.; Rafferty, E. R., Is halogen bonding the basis for iodothyronine deiodinase activity? *Inorg. Chem.* **2010**, *49* (12), 5365-7.
25. Peintner, S.; Erdelyi, M., Pushing the limits of characterising a weak halogen bond in solution. *Chem. Eur. J.* **2021**, *28* (5).
26. Wasik, R.; Lebska, M.; Felczak, K.; Poznanski, J.; Shugar, D., Relative role of halogen bonds and hydrophobic interactions in inhibition of human protein kinase CK2alpha by tetrabromobenzotriazole and some C5-substituted analogues. *J. Phys. Chem. B* **2010**, *114* (32), 10601-11.
27. Ho, P. S., Structure of the Holliday junction: applications beyond recombination. *Biochem. Soc. Trans.* **2017**, *45* (5), 1149-1158.
28. Zimmermann, M. O.; Lange, A.; Wilcken, R.; Cieslik, M. B.; Exner, T. E.; Joerger, A. C.; Koch, P.; Boeckler, F. M., Halogen-enriched fragment libraries as chemical probes for harnessing halogen bonding in fragment-based lead discovery. *Future Med. Chem.* **2014**, *6* (6), 617-639.
29. Mashalidis, E. H.; Sledz, P.; Lang, S.; Abell, C., A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat. Protoc.* **2013**, *8* (11), 2309-24.
30. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **1999**, *38* (12), 1784-1788.
31. Mayer, M.; Meyer, B., Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **2001**, *123* (25), 6108-17.

32. Begley, D. W.; Moen, S. O.; Pierce, P. G.; Zartler, E. R., Saturation transfer difference NMR for fragment screening. *Curr. Protoc. Chem. Biol.* **2013**, 5 (4), 251-68.
33. Meyer, B.; Klein, J.; Mayer, M.; Meinecke, R.; Möller, H.; Neffe, A.; Schuster, O.; Wülfken, J.; Ding, Y.; Knaie, O.; Labbe, J.; Palcic, M. M.; Hindsgaul, O.; Wagner, B.; Ernst, B. In *Saturation Transfer Difference NMR Spectroscopy for Identifying Ligand Epitopes and Binding Specificities*, Leucocyte Trafficking, Berlin, Heidelberg, Hamann, A.; Asadullah, K.; Schottelius, A., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; pp 149-167.
34. Herrera, I.; Winnik, M. A., Differential binding models for isothermal titration calorimetry: moving beyond the Wiseman isotherm. *J. Phys. Chem. B* **2013**, 117 (29), 8659-72.
35. Velazquez-Campoy, A.; Freire, E., Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protoc.* **2006**, 1 (1), 186-91.
36. Ruhmann, E.; Betz, M.; Fricke, M.; Heine, A.; Schafer, M.; Klebe, G., Thermodynamic signatures of fragment binding: Validation of direct versus displacement ITC titrations. *Biochim. Biophys. Acta* **2015**, 1850 (4), 647-56.
37. Bissantz, C.; Kuhn, B.; Stahl, M., A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010**, 53 (14), 5061-84.
38. Newberry, R. W.; Raines, R. T., Secondary Forces in Protein Folding. *ACS Chem. Biol.* **2019**, 14 (8), 1677-1686.
39. Steiner, T.; R. Desiraju, G., Distinction between the weak hydrogen bond and the van der Waals interaction. *ChemComm* **1998**, (8), 891-892.
40. Li, Z.; Lazaridis, T., The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. *J. Phys. Chem. B* **2005**, 109 (1), 662-670.
41. Mardirossian, M.; Rubini, M.; Adamo, M. F. A.; Scocchi, M.; Saviano, M.; Tossi, A.; Gennaro, R.; Caporale, A., Natural and Synthetic Halogenated Amino Acids--Structural and Bioactive Features in Antimicrobial Peptides and Peptidomimetics. *Molecules* **2021**, 26 (23).
42. Wang, C.; Du, W.; Lu, H.; Lan, J.; Liang, K.; Cao, S., A Review: Halogenated Compounds from Marine Actinomycetes. *Molecules* **2021**, 26 (9).
43. Wang, C.; Lu, H.; Lan, J.; Zaman, K. A.; Cao, S., A Review: Halogenated Compounds from Marine Fungi. *Molecules* **2021**, 26 (2).

## 6. Literature

44. Marsan, E. S.; Bayse, C. A., A Halogen Bonding Perspective on Iodothyronine Deiodinase Activity. *Molecules* **2020**, *25* (6).
45. Valadares, N. F.; Salum, L. B.; Polikarpov, I.; Andricopulo, A. D.; Garratt, R. C., Role of halogen bonds in thyroid hormone receptor selectivity: pharmacophore-based 3D-QSSR studies. *J. Chem. Inf. Model.* **2009**, *49* (11), 2606-16.
46. Christoffel, F.; Ward, T. R., Palladium-Catalyzed Heck Cross-Coupling Reactions in Water: A Comprehensive Review. *Catal. Lett.* **2018**, *148* (2), 489-511.
47. González-Sebastián, L.; Morales-Morales, D., Cross-coupling reactions catalysed by palladium pincer complexes. A review of recent advances. *J. Organomet. Chem.* **2019**, *893*, 39-51.
48. Cerreia Vioglio, P.; Szell, P. M. J.; Chierotti, M. R.; Gobetto, R.; Bryce, D. L., (79/81)Br nuclear quadrupole resonance spectroscopic characterization of halogen bonds in supramolecular assemblies. *Chem. Sci.* **2018**, *9* (20), 4555-4561.
49. Borissov, A.; Marques, I.; Lim, J. Y. C.; Felix, V.; Smith, M. D.; Beer, P. D., Anion Recognition in Water by Charge-Neutral Halogen and Chalcogen Bonding Foldamer Receptors. *J. Am. Chem. Soc.* **2019**, *141* (9), 4119-4129.
50. Govindaraj, V.; Ungati, H.; Jakka, S. R.; Bose, S.; Mugesh, G., Directing Traffic: Halogen-Bond-Mediated Membrane Transport. *Chem. Eur. J.* **2019**, *25* (48), 11180-11192.
51. Dobes, P.; Rezac, J.; Fanfrlik, J.; Otyepka, M.; Hobza, P., Semiempirical quantum mechanical method PM6-DH2X describes the geometry and energetics of CK2-inhibitor complexes involving halogen bonds well, while the empirical potential fails. *J. Phys. Chem. B* **2011**, *115* (26), 8581-9.
52. Scholfield, M. R.; Ford, M. C.; Carlsson, A. C.; Butta, H.; Mehl, R. A.; Ho, P. S., Structure-Energy Relationships of Halogen Bonds in Proteins. *Biochemistry* **2017**, *56* (22), 2794-2802.
53. Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P., Halogen bonding: the sigma-hole. Proceedings of "Modeling interactions in biomolecules II", Prague, September 5th-9th, 2005. *J. Mol. Model.* **2007**, *13* (2), 291-6.



## 6. Literature

54. Zimmermann, M. O.; Lange, A.; Boeckler, F. M., Evaluating the potential of halogen bonding in molecular design: automated scaffold decoration using the new scoring function XBScore. *J. Chem. Inf. Model.* **2015**, *55* (3), 687-99.
55. Battistutta, R.; De Moliner, E.; Sarno, S.; Zanotti, G.; Pinna, L. A., Structural features underlying selective inhibition of protein kinase CK2 by ATP site-directed tetrabromo-2-benzotriazole. *Protein Sci.* **2001**, *10* (11), 2200-6.
56. Lange, A.; Heidrich, J.; Zimmermann, M. O.; Exner, T. E.; Boeckler, F. M., Scaffold Effects on Halogen Bonding Strength. *J. Chem. Inf. Model.* **2019**, *59* (2), 885-894.
57. Lange, A.; Zimmermann, M. O.; Wilcken, R.; Zahn, S.; Boeckler, F. M., Targeting histidine side chains in molecular design through nitrogen-halogen bonds. *J. Chem. Inf. Model.* **2013**, *53* (12), 3178-89.
58. Zimmermann, M. O.; Boeckler, F. M., Targeting the protein backbone with aryl halides: systematic comparison of halogen bonding and  $\pi\cdots\pi$  interactions using N-methylacetamide. *MedChemComm* **2016**, *7* (3), 500-505.
59. Wilcken, R.; Zimmermann, M. O.; Lange, A.; Zahn, S.; Boeckler, F. M., Using halogen bonds to address the protein backbone: a systematic evaluation. *J. Comput. Aided Mol. Des.* **2012**, *26* (8), 935-45.
60. Zimmermann, M. O.; Lange, A.; Zahn, S.; Exner, T. E.; Boeckler, F. M., Using Surface Scans for the Evaluation of Halogen Bonds toward the Side Chains of Aspartate, Asparagine, Glutamate, and Glutamine. *J. Chem. Inf. Model.* **2016**, *56* (7), 1373-83.
61. Steiner, T., The Hydrogen Bond in the Solid State. *Angew. Chem. Int. Ed.* **2002**, *41* (1), 48-76.
62. Kuhn, B.; Gilberg, E.; Taylor, R.; Cole, J.; Korb, O., How Significant Are Unusual Protein-Ligand Interactions? Insights from Database Mining. *J. Med. Chem.* **2019**, *62* (22), 10441-10455.
63. Lange, A.; Gunther, M.; Buttner, F. M.; Zimmermann, M. O.; Heidrich, J.; Hennig, S.; Zahn, S.; Schall, C.; Sievers-Engler, A.; Ansideri, F.; Koch, P.; Laemmerhofer, M.; Stehle, T.; Laufer, S. A.; Boeckler, F. M., Targeting the Gatekeeper MET146 of C-Jun N-Terminal Kinase 3 Induces a Bivalent Halogen/Chalcogen Bond. *J. Am. Chem. Soc.* **2015**, *137* (46), 14640-52.

## 6. Literature

64. Heroven, C.; Georgi, V.; Ganotra, G. K.; Brennan, P.; Wolfreys, F.; Wade, R. C.; Fernandez-Montalvan, A. E.; Chaikuad, A.; Knapp, S., Halogen-Aromatic pi Interactions Modulate Inhibitor Residence Times. *Angew. Chem. Int. Ed.* **2018**, *57* (24), 7220-7224.
65. Falke, H.; Chaikuad, A.; Becker, A.; Loaec, N.; Lozach, O.; Abu Jhaisha, S.; Becker, W.; Jones, P. G.; Preu, L.; Baumann, K.; Knapp, S.; Meijer, L.; Kunick, C., 10-iodo-11H-indolo[3,2-c]quinoline-6-carboxylic acids are selective inhibitors of DYRK1A. *J. Med. Chem.* **2015**, *58* (7), 3131-43.
66. Engelhardt, M. U., unpublished data.
67. Tatko, C. D.; Waters, M. L., Effect of halogenation on edge-face aromatic interactions in a beta-hairpin peptide: enhanced affinity with iodo-substituents. *Org. Lett.* **2004**, *6* (22), 3969-72.
68. Perera, M. D.; Aakeröy, C. B., Organocatalysis by a multidentate halogen-bond donor: an alternative to hydrogen-bond based catalysis. *New J. Chem.* **2019**, *43* (21), 8311-8314.
69. Sutar, R. L.; Huber, S. M., Catalysis of Organic Reactions through Halogen Bonding. *ACS Catal.* **2019**, *9* (10), 9622-9639.
70. Priimagi, A.; Cavallo, G.; Metrangolo, P.; Resnati, G., The Halogen Bond in the Design of Functional Supramolecular Materials: Recent Advances. *Acc. Chem. Res.* **2013**, *46* (11), 2686-2695.
71. Meazza, L.; Foster, J. A.; Fucke, K.; Metrangolo, P.; Resnati, G.; Steed, J. W., Halogen-bonding-triggered supramolecular gel formation. *Nat. Chem.* **2013**, *5* (1), 42-47.
72. Manna, D.; Mugesh, G., Regioselective deiodination of thyroxine by iodothyronine deiodinase mimics: an unusual mechanistic pathway involving cooperative chalcogen and halogen bonding. *J. Am. Chem. Soc.* **2012**, *134* (9), 4269-79.
73. Wojtczak, A.; Cody, V.; Luft, J. R.; Pangborn, W., Structure of rat transthyretin (rTTR) complex with thyroxine at 2.5 Å resolution: first non-biased insight into thyroxine binding reveals different hormone orientation in two binding sites. *Acta Crystallogr. D* **2001**, *57* (Pt 8), 1061-70.
74. Fierro, A.; Matthies, D. J.; Cassels, B. K.; Jaque, P.; Zapata-Torres, G., 5-HT2 Receptor Subfamily and the Halogen Bond Promise. *J. Chem. Inf. Model.* **2021**.

## 6. Literature

75. Schoepfer, J.; Jahnke, W.; Berellini, G.; Buonamici, S.; Cotesta, S.; Cowan-Jacob, S. W.; Dodd, S.; Drucekes, P.; Fabbro, D.; Gabriel, T.; Groell, J. M.; Grotzfeld, R. M.; Hassan, A. Q.; Henry, C.; Iyer, V.; Jones, D.; Lombardo, F.; Loo, A.; Manley, P. W.; Pelle, X.; Rummel, G.; Salem, B.; Warmuth, M.; Wylie, A. A.; Zoller, T.; Marzinzik, A. L.; Furet, P., Discovery of Asciminib (ABL001), an Allosteric Inhibitor of the Tyrosine Kinase Activity of BCR-ABL1. *J. Med. Chem.* **2018**, *61* (18), 8120-8135.
76. Kunfermann, A.; Witschel, M.; Illarionov, B.; Martin, R.; Rottmann, M.; Hoffken, H. W.; Seet, M.; Eisenreich, W.; Knolker, H. J.; Fischer, M.; Bacher, A.; Groll, M.; Diederich, F., Pseudilins: halogenated, allosteric inhibitors of the non-mevalonate pathway enzyme IspD. *Angew. Chem. Int. Ed.* **2014**, *53* (8), 2235-9.
77. Matter, H.; Nazare, M.; Gussregen, S.; Will, D. W.; Schreuder, H.; Bauer, A.; Urmann, M.; Ritter, K.; Wagner, M.; Wehner, V., Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity. *Angew. Chem. Int. Ed.* **2009**, *48* (16), 2911-6.
78. Roehrig, S.; Straub, A.; Pohlmann, J.; Lampe, T.; Pernerstorfer, J.; Schlemmer, K.-H.; Reinemer, P.; Perzborn, E., Discovery of the Novel Antithrombotic Agent 5-Chloro-N-((5S)-2-oxo-3-[4-(3-oxomorpholin-4-yl)phenyl]-1,3-oxazolidin-5-yl)methylthiophene-2-carboxamide (BAY 59-7939): An Oral, Direct Factor Xa Inhibitor. *J. Med. Chem.* **2005**, *48* (19), 5900-5908.
79. Wylie, A. A.; Schoepfer, J.; Jahnke, W.; Cowan-Jacob, S. W.; Loo, A.; Furet, P.; Marzinzik, A. L.; Pelle, X.; Donovan, J.; Zhu, W.; Buonamici, S.; Hassan, A. Q.; Lombardo, F.; Iyer, V.; Palmer, M.; Berellini, G.; Dodd, S.; Thohan, S.; Bitter, H.; Branford, S.; Ross, D. M.; Hughes, T. P.; Petruzzelli, L.; Vanasse, K. G.; Warmuth, M.; Hofmann, F.; Keen, N. J.; Sellers, W. R., The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. *Nature* **2017**, *543* (7647), 733-737.
80. Dessi, A.; Peluso, P.; Dallochio, R.; Weiss, R.; Andreotti, G.; Allocca, M.; Aubert, E.; Pale, P.; Mamane, V.; Cossu, S., Rational Design, Synthesis, Characterization and Evaluation of Iodinated 4,4'-Bipyridines as New Transthyretin Fibrillogenesis Inhibitors. *Molecules* **2020**, *25* (9).
81. Neumann, C. S.; Fujimori, D. G.; Walsh, C. T., Halogenation Strategies In Natural Product Biosynthesis. *Chem. Biol.* **2008**, *15* (2), 99-109.

## 6. Literature

82. Tosstorff, A.; Cole, J. C.; Taylor, R.; Harris, S. F.; Kuhn, B., Identification of Noncompetitive Protein-Ligand Interactions for Structural Optimization. *J. Chem. Inf. Model.* **2020**, *60* (12), 6595-6611.
83. Korb, O.; Stützle, T.; Exner, T. E., PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *Lecture Notes in Computer Science*, , Dorigo M., G. L. M., Birattari M., Martinoli A., Poli R., Stützle T., Ed. Springer, Berlin Heidelberg: Heidelberg, 2006; Vol. 4150.
84. Xie, J.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G., The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **2004**, *22* (10), 1297-301.
85. FDA approves asciminib for Philadelphia chromosome-positive chronic myeloid leukemia. <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-asciminib-philadelphia-chromosome-positive-chronic-myeloid-leukemia> (accessed 29.10.2021).
86. Heidrich, J.; Exner, T. E.; Boeckler, F. M., Predicting the Magnitude of  $\sigma$ -Holes Using VmaxPred, a Fast and Efficient Tool Supporting the Application of Halogen Bonds in Drug Discovery. *J. Chem. Inf. Model.* **2019**, *59* (2), 636-643.
87. Zimmermann, M. O., unpublished data.
88. Nordqvist, A.; Nilsson, M. T.; Lagerlund, O.; Muthas, D.; Gising, J.; Yahiaoui, S.; Odell, L. R.; Srinivasa, B. R.; Larhed, M.; Mowbray, S. L.; Karlén, A., Synthesis, biological evaluation and X-ray crystallographic studies of imidazo[1,2-a]pyridine-based Mycobacterium tuberculosis glutamine synthetase inhibitors. *Med. Chem. Commun.* **2012**, *3* (5), 620-626.
89. van Vlijmen, H.; Ortholand, J.-Y.; Li, V. M. J.; de Vlieger, J. S. B., The European Lead Factory: An updated HTS compound library for innovative drug discovery. *Drug Discov.* **2021**, *26* (10), 2406-2413.
90. von Ahsen, O.; Bömer, U., High-Throughput Screening for Kinase Inhibitors. *Chembiochem* **2005**, *6* (3), 481-490.
91. Davis, B. J.; Erlanson, D. A., Learning from our mistakes: the 'unknown knowns' in fragment screening. *Bioorg. Med. Chem. Lett.* **2013**, *23* (10), 2844-52.

## 6. Literature

92. Erlanson, D.; Jahnke, W., *Fragment-based Drug Discovery Lessons and Outlook*. Wiley-VCH: 2016.
93. Bollag, G.; Tsai, J.; Zhang, J.; Zhang, C.; Ibrahim, P.; Nolop, K.; Hirth, P., Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discov.* **2012**, *11* (11), 873-886.
94. De Fusco, C.; Brear, P.; Iegre, J.; Georgiou, K. H.; Sore, H. F.; Hyvonen, M.; Spring, D. R., A fragment-based approach leading to the discovery of a novel binding site and the selective CK2 inhibitor CAM4066. *Bioorg. Med. Chem.* **2017**, *25* (13), 3471-3482.
95. Bergström, C. A. S.; Larsson, P., Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting. *Int. J. Pharm.* **2018**, *540* (1), 185-193.
96. Işık, M.; Bergazin, T. D.; Fox, T.; Rizzi, A.; Chodera, J. D.; Mobley, D. L., Assessing the accuracy of octanol–water partition coefficient predictions in the SAMPL6 Part II log P Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 335-370.
97. Dalvit, C.; Vulpetti, A., Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, *62* (5), 2218-2244.
98. Boeckler, F. M.; Joerger, A. C.; Jaggi, G.; Rutherford, T. J.; Veprintsev, D. B.; Fersht, A. R., Targeted rescue of a destabilized mutant of p53 by an in silico screened drug. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (30), 10360-5.
99. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H., A 'Rule of Three' for fragment-based lead discovery? *Drug Discov.* **2003**, *8* (19), 876-877.
100. Shultz, M. D., Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *J. Med. Chem.* **2019**, *62* (4), 1701-1714.
101. Lipinski, C. A., Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. *Adv. Drug Deliv. Rev.* **2016**, *101*, 34-41.
102. Heidrich, J. Facilitating the Application of Halogen Bonding in Fragment-based Drug Discovery: From QM Calculations, Machine Learning and Tailormade Tools for Library Design to the Experimental Evaluation of Diversity-optimized HEFLibs. Monography, University of Tübingen, 2018.

## 6. Literature

103. Wildman, S. A.; Crippen, G. M., Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (5), 868-873.
104. Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J., Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88* (7), 990-994.
105. Zhou, M.; Li, Q.; Wang, R., Current Experimental Methods for Characterizing Protein–Protein Interactions. *ChemMedChem* **2016**, *11* (8), 738-756.
106. Turnbull, W. B.; Daranas, A. H., On the value of  $c$ : can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **2003**, *125* (48), 14859-66.
107. Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L.-N., Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **1989**, *179* (1), 131-137.
108. Song, C.; Zhang, S.; Huang, H., Choosing a suitable method for the identification of replication origins in microbial genomes. *Front. Microbiol.* **2015**, *6* (1049).
109. Russo Krauss, I.; Merlino, A.; Vergara, A.; Sica, F., An Overview of Biological Macromolecule Crystallization. *Int. J. Mol. Sci.* **2013**, *14* (6).
110. Neveu, G.; Ziv-Av, A.; Barouch-Bentov, R.; Berkerman, E.; Mulholland, J.; Einav, S.; Ou, J. H. J., AP-2-Associated Protein Kinase 1 and Cyclin G-Associated Kinase Regulate Hepatitis C Virus Entry and Are Potential Drug Targets. *J. Virol.* **2015**, *89* (8), 4387-4404.
111. Verdonck, S.; Pu, S.-Y.; Sorrell, F. J.; Elkins, J. M.; Froeyen, M.; Gao, L.-J.; Prugar, L. I.; Dorosky, D. E.; Brannan, J. M.; Barouch-Bentov, R.; Knapp, S.; Dye, J. M.; Herdewijn, P.; Einav, S.; De Jonghe, S., Synthesis and Structure–Activity Relationships of 3,5-Disubstituted-pyrrolo[2,3-b]pyridines as Inhibitors of Adaptor-Associated Kinase 1 with Antiviral Activity. *J. Med. Chem.* **2019**, *62* (12), 5810-5831.
112. Mercer, J.; Schelhaas, M.; Helenius, A., Virus Entry by Endocytosis. *Annu. Rev. Biochem.* **2010**, *79* (1), 803-833.
113. Stebbing, J.; Phelan, A.; Griffin, I.; Tucker, C.; Oechsle, O.; Smith, D.; Richardson, P., COVID-19: combining antiviral and anti-inflammatory treatments. *Lancet Infect. Dis.* **2020**, *20* (4), 400-402.

## 6. Literature

114. Martínez-Gualda, B.; Schols, D.; De Jonghe, S., A patent review of adaptor associated kinase 1 (AAK1) inhibitors (2013-present). *Expert Opin. Ther. Pat.* **2021**, *31* (10), 911-936.
115. Harding, S. D.; Armstrong, Jane F.; Faccenda, E.; Southan, C.; Alexander, S. P. H.; Davenport, A. P.; Pawson, Adam J.; Spedding, M.; Davies, Jamie A.; Nc, I., The IUPHAR/BPS guide to PHARMACOLOGY in 2022: curating pharmacology for COVID-19, malaria and antibacterials. *Nucleic Acids Res.* **2021**, *50* (D1), D1282-D1294.
116. Berginski, M. E.; Moret, N.; Liu, C.; Goldfarb, D.; Sorger, Peter K.; Gomez, S. M., The Dark Kinase Knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Res.* **2021**, *49* (D1), D529-D535.
117. Fernández, A.; Crespo, A.; Tiwari, A., Is there a case for selectively promiscuous anticancer drugs? *Drug Discov. Today* **2009**, *14* (1), 1-5.
118. Becker, W.; Sippl, W., Activation, regulation, and inhibition of DYRK1A. *FEBS J.* **2011**, *278* (2), 246-56.
119. Alexeeva, M.; Aberg, E.; Engh, R. A.; Rothweiler, U., The structure of a dual-specificity tyrosine phosphorylation-regulated kinase 1A-PKC412 complex reveals disulfide-bridge formation with the anomalous catalytic loop HRD(HCD) cysteine. *Acta Crystallogr. D* **2015**, *71*, 1207-15.
120. Soundararajan, M.; Roos, A. K.; Savitsky, P.; Filippakopoulos, P.; Kettenbach, A. N.; Olsen, J. V.; Gerber, S. A.; Eswaran, J.; Knapp, S.; Elkins, J. M., Structures of Down syndrome kinases, DYRKs, reveal mechanisms of kinase activation and substrate recognition. *Structure* **2013**, *21* (6), 986-96.
121. Ogawa, Y.; Nonaka, Y.; Goto, T.; Ohnishi, E.; Hiramatsu, T.; Kii, I.; Yoshida, M.; Ikura, T.; Onogi, H.; Shibuya, H.; Hosoya, T.; Ito, N.; Hagiwara, M., Development of a novel selective inhibitor of the Down syndrome-related kinase Dyrk1A. *Nat. Commun.* **2010**, *1* (1), 86.
122. Lee Walmsley, D.; Murray, J. B.; Dokurno, P.; Massey, A. J.; Benwell, K.; Fiumana, A.; Foloppe, N.; Ray, S.; Smith, J.; Surgenor, A. E.; Edmonds, T.; Demarles, D.; Burbridge, M.; Cruzalegui, F.; Kotschy, A.; Hubbard, R. E., Fragment-Derived Selective Inhibitors of Dual-Specificity Kinases DYRK1A and DYRK1B. *J. Med. Chem.* **2021**, *64* (13), 8971-8991.

## 6. Literature

123. Meine, R.; Becker, W.; Falke, H.; Preu, L.; Loaëc, N.; Meijer, L.; Kunick, C., Indole-3-Carbonitriles as DYRK1A Inhibitors by Fragment-Based Drug Design. *Molecules* **2018**, *23* (2).
124. Rothweiler, U.; Eriksson, J.; Stensen, W.; Leeson, F.; Engh, R. A.; Svendsen, J. S., Luciferin and derivatives as a DYRK selective scaffold for the design of protein kinase inhibitors. *Eur. J. Med. Chem.* **2015**, *94*, 140-8.
125. Rothweiler, U.; Stensen, W.; Brandsdal, B. O.; Isaksson, J.; Leeson, F. A.; Engh, R. A.; Svendsen, J. S., Probing the ATP-Binding Pocket of Protein Kinase DYRK1A with Benzothiazole Fragment Molecules. *J. Med. Chem.* **2016**, *59* (21), 9814-9824.
126. Guard, S. E.; Poss, Z. C.; Ebmeier, C. C.; Pagratis, M.; Simpson, H.; Taatjes, D. J.; Old, W. M., The nuclear interactome of DYRK1A reveals a functional role in DNA damage repair. *Sci. Rep.* **2019**, *9* (1), 6539.
127. Mishra, P.; Gunther, S., New insights into the structural dynamics of the kinase JNK3. *Sci. Rep.* **2018**, *8* (1), 9435.
128. Shaw, D.; Wang, S. M.; Villasenor, A. G.; Tsing, S.; Walter, D.; Browner, M. F.; Barnett, J.; Kuglstatter, A., The crystal structure of JNK2 reveals conformational flexibility in the MAP kinase insert and indicates its involvement in the regulation of catalytic activity. *J. Mol. Biol.* **2008**, *383* (4), 885-93.
129. Wada, T.; Penninger, J. M., Mitogen-activated protein kinases in apoptosis regulation. *Oncogene* **2004**, *23* (16), 2838-49.
130. Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G. L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C.; Edkins, S.; O'Meara, S.; Vastrik, I.; Schmidt, E. E.; Avis, T.; Barthorpe, S.; Bhamra, G.; Buck, G.; Choudhury, B.; Clements, J.; Cole, J.; Dicks, E.; Forbes, S.; Gray, K.; Halliday, K.; Harrison, R.; Hills, K.; Hinton, J.; Jenkinson, A.; Jones, D.; Menzies, A.; Mironenko, T.; Perry, J.; Raine, K.; Richardson, D.; Shepherd, R.; Small, A.; Tofts, C.; Varian, J.; Webb, T.; West, S.; Widaa, S.; Yates, A.; Cahill, D. P.; Louis, D. N.; Goldstraw, P.; Nicholson, A. G.; Brasseur, F.; Looijenga, L.; Weber, B. L.; Chiew, Y.-E.; deFazio, A.; Greaves, M. F.; Green, A. R.; Campbell, P.; Birney, E.; Easton, D. F.; Chenevix-Trench, G.; Tan, M.-H.; Khoo, S. K.; Teh, B. T.; Yuen, S. T.; Leung, S. Y.; Wooster, R.; Futreal, P. A.; Stratton, M. R., Patterns of somatic mutation in human cancer genomes. *Nature* **2007**, *446* (7132), 153-158.



## 6. Literature

131. Wu, Q.; Wu, W.; Jacevic, V.; Franca, T. C. C.; Wang, X.; Kuca, K., Selective inhibitors for JNK signalling: a potential targeted therapy in cancer. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2020**, *35* (1), 574-583.
132. Jiang, R.; Duckett, D.; Chen, W.; Habel, J.; Ling, Y. Y.; LoGrasso, P.; Kamenecka, T. M., 3,5-Disubstituted quinolines as novel c-Jun N-terminal kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17* (22), 6378-6382.
133. Kuglstatler, A.; Ghate, M.; Tsing, S.; Villasenor, A. G.; Shaw, D.; Barnett, J. W.; Browner, M. F., X-ray crystal structure of JNK2 complexed with the p38alpha inhibitor BIRB796: insights into the rational design of DFG-out binding MAP kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20* (17), 5217-20.
134. Feng, Q.; Wang, H. B.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Struhl, K.; Zhang, Y., Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* **2002**, *12* (12), 1052-1058.
135. Min, J.; Feng, Q.; Li, Z.; Zhang, Y.; Xu, R. M., Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. *Cell* **2003**, *112* (5), 711-23.
136. Chen, C.; Zhu, H.; Stauffer, F.; Caravatti, G.; Vollmer, S.; Machauer, R.; Holzer, P.; Mobitz, H.; Scheufler, C.; Klumpp, M.; Tiedt, R.; Beyer, K. S.; Calkins, K.; Guthy, D.; Kiffe, M.; Zhang, J.; Gaul, C., Discovery of Novel Dot1L Inhibitors through a Structure-Based Fragmentation Approach. *ACS Med Chem Lett* **2016**, *7* (8), 735-40.
137. Gibbons, G. S.; Chakraborty, A.; Grigsby, S. M.; Umeano, A. C.; Liao, C.; Moukha-Chafiq, O.; Pathak, V.; Mathew, B.; Lee, Y. T.; Dou, Y.; Schurer, S. C.; Reynolds, R. C.; Snowden, T. S.; Nikolovska-Coleska, Z., Identification of DOT1L inhibitors by structure-based virtual screening adapted from a nucleoside-focused library. *Eur. J. Med. Chem.* **2020**, *189*, 112023.
138. Mobitz, H.; Machauer, R.; Holzer, P.; Vaupel, A.; Stauffer, F.; Ragot, C.; Caravatti, G.; Scheufler, C.; Fernandez, C.; Hommel, U.; Tiedt, R.; Beyer, K. S.; Chen, C.; Zhu, H.; Gaul, C., Discovery of Potent, Selective, and Structurally Novel Dot1L Inhibitors by a Fragment Linking Approach. *ACS Med Chem Lett* **2017**, *8* (3), 338-343.
139. Scheufler, C.; Mobitz, H.; Gaul, C.; Ragot, C.; Be, C.; Fernandez, C.; Beyer, K. S.; Tiedt, R.; Stauffer, F., Optimization of a Fragment-Based Screening Hit toward Potent DOT1L Inhibitors Interacting in an Induced Binding Pocket. *ACS Med Chem Lett* **2016**, *7* (8), 730-4.

## 6. Literature

140. Vaidergorn, M. M.; da Silva Emery, F.; Ganesan, A., From Hit Seeking to Magic Bullets: The Successful Union of Epigenetic and Fragment Based Drug Discovery (EPIDD + FBDD). *J. Med. Chem.* **2021**.
141. Daigle, S. R.; Olhava, E. J.; Therkelsen, C. A.; Basavapathruni, A.; Jin, L.; Boriack-Sjodin, P. A.; Allain, C. J.; Klaus, C. R.; Raimondi, A.; Scott, M. P.; Waters, N. J.; Chesworth, R.; Moyer, M. P.; Copeland, R. A.; Richon, V. M.; Pollock, R. M., Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood* **2013**, *122* (6), 1017-1025.
142. ClinicalTrials.gov Pinometostat With Standard Chemotherapy in Treating Patients With Newly Diagnosed Acute Myeloid Leukemia and MLL Gene Rearrangement. <https://www.clinicaltrials.gov/ct2/show/NCT03724084?term=EPZ-5676&draw=2&rank=4>.
143. Li, F.; Aljahdali, I.; Ling, X., Cancer therapeutics using survivin BIRC5 as a target: what can we do after over two decades of study? *J. Exp. Clin. Cancer Res.* **2019**, *38* (1), 368.
144. Li, F.; Ambrosini, G.; Chu, E. Y.; Plescia, J.; Tognin, S.; Marchisio, P. C.; Altieri, D. C., Control of apoptosis and mitotic spindle checkpoint by survivin. *Nature* **1998**, *396* (6711), 580-584.
145. Wang, H.; Holloway, M. P.; Ma, L.; Cooper, Z. A.; Riolo, M.; Samkari, A.; Elenitoba-Johnson, K. S. J.; Chin, Y. E.; Altura, R. A., Acetylation Directs Survivin Nuclear Localization to Repress STAT3 Oncogenic Activity\*. *J. Biol. Chem.* **2010**, *285* (46), 36129-36137.
146. Pavlyukov, M. S.; Antipova, N. V.; Balashova, M. V.; Vinogradova, T. V.; Kopantzev, E. P.; Shakhparonov, M. I., Survivin Monomer Plays an Essential Role in Apoptosis Regulation. *J. Biol. Chem.* **2011**, *286* (26), 23296-23307.
147. Liu, Y.; Lear, T.; Iannone, O.; Shiva, S.; Corey, C.; Rajbhandari, S.; Jerome, J.; Chen, B. B.; Mallampalli, R. K., The Proapoptotic F-box Protein Fbx17 Regulates Mitochondrial Function by Mediating the Ubiquitylation and Proteasomal Degradation of Survivin\*. *J. Biol. Chem.* **2015**, *290* (19), 11843-11852.
148. Martínez-García, D.; Manero-Rupérez, N.; Quesada, R.; Korrodi-Gregório, L.; Soto-Cerrato, V., Therapeutic strategies involving survivin inhibition in cancer. *Med. Res. Rev.* **2019**, *39* (3), 887-909.
149. Lewis-Ballester, A.; Pham, K. N.; Batabyal, D.; Karkashon, S.; Bonanno, J. B.; Poulos, T. L.; Yeh, S. R., Structural insights into substrate and inhibitor binding sites in human indoleamine 2,3-dioxygenase 1. *Nat. Commun.* **2017**, *8* (1), 1693.

## 6. Literature

150. Wu, Y.; Xu, T.; Liu, J.; Ding, K.; Xu, J., Structural insights into the binding mechanism of IDO1 with hydroxylamide based inhibitor INCB14943. *Biochem. Biophys. Res. Commun.* **2017**, *487* (2), 339-343.
151. van Baren, N.; Van den Eynde, B. J., Tryptophan-Degrading Enzymes in Tumoral Immune Resistance. *Front. Immunol.* **2016**, *6* (34).
152. Uyttenhove, C.; Pilotte, L.; Théate, I.; Stroobant, V.; Colau, D.; Parmentier, N.; Boon, T.; Van den Eynde, B. J., Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2,3-dioxygenase. *Nat. Med.* **2003**, *9* (10), 1269-1274.
153. Metz, R.; DuHadaway, J. B.; Kamasani, U.; Laury-Kleintop, L.; Muller, A. J.; Prendergast, G. C., Novel Tryptophan Catabolic Enzyme IDO2 Is the Preferred Biochemical Target of the Antitumor Indoleamine 2,3-Dioxygenase Inhibitory Compound D-1-Methyl-Tryptophan. *Cancer Res.* **2007**, *67* (15), 7082.
154. A Phase 3 Study of Pembrolizumab + Epcadostat or Placebo in Subjects With Unresectable or Metastatic Melanoma (Keynote-252 / ECHO-301). <https://clinicaltrials.gov/ct2/show/NCT02752074>.
155. Fang, K.; Dong, G.; Li, Y.; He, S.; Wu, Y.; Wu, S.; Wang, W.; Sheng, C., Discovery of Novel Indoleamine 2,3-Dioxygenase 1 (IDO1) and Histone Deacetylase (HDAC) Dual Inhibitors. *ACS Med Chem Lett* **2018**, *9* (4), 312-317.
156. Sequence Identity And Similarity. <http://imed.med.ucm.es/Tools/sias.html>.
157. Baell, J. B.; Nissink, J. W. M., Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017-Utility and Limitations. *ACS Chem. Biol.* **2018**, *13* (1), 36-44.
158. Luo, S.; Xu, K.; Xiang, S.; Chen, J.; Chen, C.; Guo, C.; Tong, Y.; Tong, L., High-resolution structures of inhibitor complexes of human indoleamine 2,3-dioxygenase 1 in a new crystal form. *Acta Crystallogr. F* **2018**, (F74), 717-724.
159. Richon, V. M.; Johnston, D.; Sneeringer, C. J.; Jin, L.; Majer, C. R.; Elliston, K.; Jerva, L. F.; Scott, M. P.; Copeland, R. A., Chemogenetic analysis of human protein methyltransferases. *Chem. Biol. Drug. Des.* **2011**, *78* (2), 199-210.
160. Serena, M.; Bastos, R. N.; Elliott, P. R.; Barr, F. A., Molecular basis of MKLP2-dependent Aurora B transport from chromatin to the anaphase central spindle. *J. Cell Biol.* **2020**, *219* (7), e201910059.

## 6. Literature

161. Capuzzi, S. J.; Muratov, E. N.; Tropsha, A., Phantom PAINS: Problems with the Utility of Alerts for Pan-Assay INterference CompoundS. *J. Chem. Inf. Model.* **2017**, *57* (3), 417-427.
162. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science* **2002**, *298* (5600), 1912-34.
163. Hartz, R. A.; Ahuja, V. T.; Nara, S. J.; Kumar, C. M. V.; Brown, J. M.; Bristow, L. J.; Rajamani, R.; Muckelbauer, J. K.; Camac, D.; Kiefer, S. E.; Hunihan, L.; Gulianello, M.; Lewis, M.; Easton, A.; Lippy, J. S.; Surti, N.; Pattipati, S. N.; Dokania, M.; Elavazhagan, S.; Dandapani, K.; Hamman, B. D.; Allen, J.; Kostich, W.; Bronson, J. J.; Macor, J. E.; Dzierba, C. D., Discovery, Structure–Activity Relationships, and In Vivo Evaluation of Novel Aryl Amides as Brain Penetrant Adaptor Protein 2-Associated Kinase 1 (AAK1) Inhibitors for the Treatment of Neuropathic Pain. *J. Med. Chem.* **2021**, *64* (15), 11090-11128.
164. Sorrell, F. J.; Szklarz, M.; Abdul Azeez, K. R.; Elkins, J. M.; Knapp, S., Family-wide Structural Analysis of Human Numb-Associated Protein Kinases. *Structure* **2016**, *24* (3), 401-11.
165. Meininger, D.; Zalameda, L.; Liu, Y.; Stepan, L. P.; Borges, L.; McCarter, J. D.; Sutherland, C. L., Purification and kinetic characterization of human indoleamine 2,3-dioxygenases 1 and 2 (IDO1 and IDO2) and discovery of selective IDO1 inhibitors. *Biochim. Biophys. Acta* **2011**, *1814* (12), 1947-54.
166. Takikawa, O.; Kuroiwa, T.; Yamazaki, F.; Kido, R., Mechanism of interferon-gamma action. Characterization of indoleamine 2,3-dioxygenase in cultured human cells induced by interferon-gamma and evaluation of the enzyme-mediated tryptophan degradation in its anticellular activity. *J. Biol. Chem.* **1988**, *263* (4), 2041-8.
167. Austin, C. J.; Mizdrak, J.; Matin, A.; Sirijovski, N.; Kosim-Satyaputra, P.; Willows, R. D.; Roberts, T. H.; Truscott, R. J.; Polekhina, G.; Parker, M. W.; Jamie, J. F., Optimised expression and purification of recombinant human indoleamine 2,3-dioxygenase. *Protein Expr. Purif.* **2004**, *37* (2), 392-8.
168. Qian, S.; He, T.; Wang, W.; He, Y.; Zhang, M.; Yang, L.; Li, G.; Wang, Z., Discovery and preliminary structure-activity relationship of 1H-indazoles with promising indoleamine-2,3-dioxygenase 1 (IDO1) inhibition properties. *Bioorg. Med. Chem.* **2016**, *24* (23), 6194-6205.
169. Sugimoto, H.; Oda, S.; Otsuki, T.; Hino, T.; Yoshida, T.; Shiro, Y., Crystal structure of human indoleamine 2,3-dioxygenase: catalytic mechanism of O<sub>2</sub> incorporation by a heme-containing dioxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (8), 2611-6.

## 6. Literature

170. Ning, X. L.; Li, Y. Z.; Huo, C.; Deng, J.; Gao, C.; Zhu, K. R.; Wang, M.; Wu, Y. X.; Yu, J. L.; Ren, Y. L.; Luo, Z. Y.; Li, G.; Chen, Y.; Wang, S. Y.; Peng, C.; Yang, L. L.; Wang, Z. Y.; Wu, Y.; Qian, S.; Li, G. B., X-ray Structure-Guided Discovery of a Potent, Orally Bioavailable, Dual Human Indoleamine/Tryptophan 2,3-Dioxygenase (hIDO/hTDO) Inhibitor That Shows Activity in a Mouse Model of Parkinson's Disease. *J. Med. Chem.* **2021**, *64* (12), 8303-8332.
171. Pham, K. N.; Yeh, S. R., Mapping the Binding Trajectory of a Suicide Inhibitor in Human Indoleamine 2,3-Dioxygenase 1. *J. Am. Chem. Soc.* **2018**, *140* (44), 14538-14541.
172. Peng, Y. H.; Ueng, S. H.; Tseng, C. T.; Hung, M. S.; Song, J. S.; Wu, J. S.; Liao, F. Y.; Fan, Y. S.; Wu, M. H.; Hsiao, W. C.; Hsueh, C. C.; Lin, S. Y.; Cheng, C. Y.; Tu, C. H.; Lee, L. C.; Cheng, M. F.; Shia, K. S.; Shih, C.; Wu, S. Y., Important Hydrogen Bond Networks in Indoleamine 2,3-Dioxygenase 1 (IDO1) Inhibitor Design Revealed by Crystal Structures of Imidazoleisoindole Derivatives with IDO1. *J. Med. Chem.* **2016**, *59* (1), 282-93.
173. Nelp, M. T.; Kates, P. A.; Hunt, J. T.; Newitt, J. A.; Balog, A.; Maley, D.; Zhu, X.; Abell, L.; Allentoff, A.; Borzilleri, R.; Lewis, H. A.; Lin, Z.; Seitz, S. P.; Yan, C.; Groves, J. T., Immune-modulating enzyme indoleamine 2,3-dioxygenase is effectively inhibited by targeting its apo-form. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (13), 3249-3254.
174. Stamos, J. L.; Chu, M. L.; Enos, M. D.; Shah, N.; Weis, W. I., Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. *Elife* **2014**, *3*, e01998.
175. Bax, B.; Carter, P. S.; Lewis, C.; Guy, A. R.; Bridges, A.; Tanner, R.; Pettman, G.; Mannix, C.; Culbert, A. A.; Brown, M. J. B.; Smith, D. G.; Reith, A. D., The Structure of Phosphorylated GSK-3 $\beta$  Complexed with a Peptide, FRATtide, that Inhibits  $\beta$ -Catenin Phosphorylation. *Structure* **2001**, *9* (12), 1143-1152.
176. Weinstock, M. T.; Heseck, E. D.; Wilson, C. M.; Gibson, D. G., *Vibrio natriegens* as a fast-growing host for molecular biology. *Nat. Methods* **2016**, *13* (10), 849-51.
177. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **2014**, *5* (172), 172.
178. Lorenz, R.; Bernhart, S. H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; Hofacker, I. L., ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6* (1), 26.

## 6. Literature

179. Gruber, A. R.; Lorenz, R.; Bernhart, S. H.; Neuböck, R.; Hofacker, I. L., The Vienna RNA Websuite. *Nucleic Acids Res.* **2008**, *36* (suppl\_2), W70-W74.
180. Kim, E.; Ilic, N.; Shrestha, Y.; Zou, L.; Kamburov, A.; Zhu, C.; Yang, X.; Lubonja, R.; Tran, N.; Nguyen, C.; Lawrence, M. S.; Piccioni, F.; Bagul, M.; Doench, J. G.; Chouinard, C. R.; Wu, X.; Hogstrom, L.; Natoli, T.; Tamayo, P.; Horn, H.; Corsello, S. M.; Lage, K.; Root, D. E.; Subramanian, A.; Golub, T. R.; Getz, G.; Boehm, J. S.; Hahn, W. C., Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles. *Cancer Discov.* **2016**, *6* (7), 714.
181. Keeley, A.; Abranyi-Balogh, P.; Keseru, G. M., Design and characterization of a heterocyclic electrophilic fragment library for the discovery of cysteine-targeted covalent inhibitors. *MedChemComm* **2019**, *10* (2), 263-267.
182. Bauer, M. R.; Joerger, A. C.; Fersht, A. R., 2-Sulfonylpyrimidines: Mild alkylating agents with anticancer activity toward p53-compromised cells. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (36), E5271-80.
183. Hyatt, J. L.; Moak, T.; Hatfield, M. J.; Tsurkan, L.; Edwards, C. C.; Wierdl, M.; Danks, M. K.; Wadkins, R. M.; Potter, P. M., Selective inhibition of carboxylesterases by isatins, indole-2,3-diones. *J. Med. Chem.* **2007**, *50* (8), 1876-85.
184. Petri, L.; Ábrányi-Balogh, P.; Tímea, I.; Pálffy, G.; Perczel, A.; Knez, D.; Hrast, M.; Gobec, M.; Sosič, I.; Nyíri, K.; Vértessy, B. G.; Jänsch, N.; Desczyk, C.; Meyer-Almes, F.-J.; Ogris, I.; Golič Grdadolnik, S.; Iacovino, L. G.; Binda, C.; Gobec, S.; Keserű, G. M., Assessment of Tractable Cysteines for Covalent Targeting by Screening Covalent Fragments. *Chembiochem* **2021**, *22* (4), 743-753.
185. Aretz, J.; Rademacher, C., Ranking Hits From Saturation Transfer Difference Nuclear Magnetic Resonance-Based Fragment Screening. *Front. Chem.* **2019**, *7* (215), 215.
186. ULC, C. C. G. *Molecular Operating Environment (MOE)*, 2019.01; Chemical Computing Group ULC,; 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.
187. Dammann, M.; Kramer, M.; Zimmermann, M. O.; Boeckler, F. M., Quadruple Target Evaluation of Diversity-Optimized Halogen-Enriched Fragments (HEFLibs) Reveals Substantial Ligand Efficiency for AP2-Associated Protein Kinase 1 (AAK1). *Front. Chem.* **2022**, *9*.

## 6. Literature

188. Fabbro, D.; Cowan-Jacob, S. W.; Moebitz, H., Ten things you should know about protein kinases: IUPHAR Review 14. *Br. J. Pharmacol.* **2015**, *172* (11), 2675-2700.
189. Yuasa, H. J.; Takubo, M.; Takahashi, A.; Hasegawa, T.; Noma, H.; Suzuki, T., Evolution of Vertebrate Indoleamine 2,3-Dioxygenases. *J. Mol. Evol.* **2007**, *65* (6), 705.
190. Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M., Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chem. Eur. J.* **2010**, *16* (26), 7803-12.
191. Schultes, S.; de Graaf, C.; Haaksma, E.; de Esch, I. J.; Leurs, R.; Krämer, O., Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discov. Today Technol.* **2010**, *7* (3), e147-202.
192. Hopkins, A. L.; Keseru, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H., The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.* **2014**, *13* (2), 105-21.
193. Matta, C. F.; Massa, L.; Gubskaya, A. V.; Knoll, E., Can One Take the Logarithm or the Sine of a Dimensioned Quantity or a Unit? Dimensional Analysis Involving Transcendental Functions. *J. Chem. Educ.* **2011**, *88* (1), 67-70.
194. Kenny, P. W., The nature of ligand efficiency. *J. Cheminform.* **2019**, *11* (1), 8.
195. Tarcsay, A.; Nyiri, K.; Keseru, G. M., Impact of lipophilic efficiency on compound quality. *J. Med. Chem.* **2012**, *55* (3), 1252-60.
196. Oguro, Y.; Miyamoto, N.; Okada, K.; Takagi, T.; Iwata, H.; Awazu, Y.; Miki, H.; Hori, A.; Kamiyama, K.; Imamura, S., Design, synthesis, and evaluation of 5-methyl-4-phenoxy-5H-pyrrolo[3,2-d]pyrimidine derivatives: novel VEGFR2 kinase inhibitors binding to inactive kinase conformation. *Bioorg. Med. Chem.* **2010**, *18* (20), 7260-73.
197. Ishikawa, T.; Seto, M.; Banno, H.; Kawakita, Y.; Oorui, M.; Taniguchi, T.; Ohta, Y.; Tamura, T.; Nakayama, A.; Miki, H.; Kamiguchi, H.; Tanaka, T.; Habuka, N.; Sogabe, S.; Yano, J.; Aertgeerts, K.; Kamiyama, K., Design and synthesis of novel human epidermal growth factor receptor 2 (HER2)/epidermal growth factor receptor (EGFR) dual inhibitors bearing a pyrrolo[3,2-d]pyrimidine scaffold. *J. Med. Chem.* **2011**, *54* (23), 8030-50.
198. Gangjee, A.; Pavana, R. K.; Li, W.; Hamel, E.; Westbrook, C.; Mooberry, S. L., Novel water-soluble substituted pyrrolo[3,2-d]pyrimidines: design, synthesis, and biological evaluation as antitubulin antitumor agents. *Pharm. Res.* **2012**, *29* (11), 3033-9.

## 6. Literature

199. Dammann, M. unpublished data.
200. Meine, R.; Becker, W.; Falke, H.; Preu, L.; Loaec, N.; Meijer, L.; Kunick, C., Indole-3-Carbonitriles as DYRK1A Inhibitors by Fragment-Based Drug Design. *Molecules* **2018**, *23* (2).
201. Battistutta, R.; Mazzorana, M.; Cendron, L.; Bortolato, A.; Sarno, S.; Kazimierczuk, Z.; Zanotti, G.; Moro, S.; Pinna, L. A., The ATP-binding site of protein kinase CK2 holds a positive electrostatic area and conserved water molecules. *Chembiochem* **2007**, *8* (15), 1804-9.
202. Chojnacki, K.; Lindenblatt, D.; Winska, P.; Wielechowska, M.; Toelzer, C.; Niefind, K.; Bretner, M., Synthesis, biological properties and structural study of new halogenated azolo[4,5-b]pyridines as inhibitors of CK2 kinase. *Bioorg. Chem.* **2021**, *106*, 104502.
203. Yueh, C.; Rettenmaier, J.; Xia, B.; Hall, D. R.; Alekseenko, A.; Porter, K. A.; Barkovich, K.; Keseru, G.; Whitty, A.; Wells, J. A.; Vajda, S.; Kozakov, D., Kinase Atlas: Druggability Analysis of Potential Allosteric Sites in Kinases. *J. Med. Chem.* **2019**, *62* (14), 6512-6524.
204. Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J., An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **2006**, *125* (6), 1137-1149.
205. Ansideri, F.; Macedo, J. T.; Eitel, M.; El-Gokha, A.; Zinad, D. S.; Scarpellini, C.; Kudolo, M.; Schollmeyer, D.; Boeckler, F. M.; Blaum, B. S.; Laufer, S. A.; Koch, P., Structural Optimization of a Pyridinylimidazole Scaffold: Shifting the Selectivity from p38 $\alpha$  Mitogen-Activated Protein Kinase to c-Jun N-Terminal Kinase 3. *ACS Omega* **2018**, *3* (7), 7809-7831.
206. Harrington, P. E.; Biswas, K.; Malwitz, D.; Tasker, A. S.; Mohr, C.; Andrews, K. L.; Dellamaggiore, K.; Kendall, R.; Beckmann, H.; Jaeckel, P.; Materna-Reichelt, S.; Allen, J. R.; Lipford, J. R., Unfolded Protein Response in Cancer: IRE1 $\alpha$  Inhibition by Selective Kinase Ligands Does Not Impair Tumor Cell Viability. *ACS Med Chem Lett* **2015**, *6* (1), 68-72.
207. Zhou, P.; Lv, J.; Zou, J.; Tian, F.; Shang, Z., Halogen-water-hydrogen bridges in biomolecules. *J. Struct. Biol.* **2010**, *169* (2), 172-82.
208. Riel, A. M. S.; Rowe, R. K.; Ho, E. N.; Carlsson, A. C.; Rappe, A. K.; Berryman, O. B.; Ho, P. S., Hydrogen Bond Enhanced Halogen Bonds: A Synergistic Interaction in Chemistry and Biochemistry. *Acc. Chem. Res.* **2019**, *52* (10), 2870-2880.



## 6. Literature

209. Yang, X.; Boehm, J. S.; Yang, X.; Salehi-Ashtiani, K.; Hao, T.; Shen, Y.; Lubonja, R.; Thomas, S. R.; Alkan, O.; Bhimdi, T.; Green, T. M.; Johannessen, C. M.; Silver, S. J.; Nguyen, C.; Murray, R. R.; Hieronymus, H.; Balcha, D.; Fan, C.; Lin, C.; Ghamsari, L.; Vidal, M.; Hahn, W. C.; Hill, D. E.; Root, D. E., A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* **2011**, *8* (8), 659-61.
210. Kapust, R. B.; Tozser, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S., Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **2001**, *14* (12), 993-1000.
211. Albanese, S. K.; Parton, D. L.; Işık, M.; Rodríguez-Laureano, L.; Hanson, S. M.; Behr, J. M.; Gradia, S.; Jeans, C.; Levinson, N. M.; Seeliger, M. A.; Chodera, J. D., An Open Library of Human Kinase Domain Constructs for Automated Bacterial Expression. *Biochemistry* **2018**, *57* (31), 4675-4689.
212. Singh, A.; Upadhyay, V.; Upadhyay, A. K.; Singh, S. M.; Panda, A. K., Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microb. Cell Factories* **2015**, *14* (1), 41.
213. Palmer, I.; Wingfield, P. T., Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. *Curr. Protoc. Protein Sci.* **2004**, Chapter 6, Unit 6.3.
214. Kabsch, W., XDS. *Acta Crystallogr. D* **2010**, *66* (Pt 2), 125-32.
215. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D* **2010**, *66* (4), 486-501.
216. TURBOMOLE V7.4.1 2017, a. d. o. U. o. K. a.; Forschungszentrum Karlsruhe GmbH, -; TURBOMOLE GmbH, s. a. f.; <http://www.turbomole.com>.
217. Revillo Imbernon, J.; Jacquemard, C.; Bret, G.; Marcou, G.; Kellenberger, E., Comprehensive analysis of commercial fragment libraries. *Rsc Med Chem* **2022**.

## 7 Appendix

### Appendix A: DNA Constructs

#### AAK1 in a pET24a vector:

```
ATGCACCACCACCACCACCACAGCGGCGCTTTTGAATTTAAGCTGCCGGACATTGGCG
AAGGCATCCACGAAGGTGAAATTGTCAAATGGTTTGTGAAACCGGGCGATGAAGTGAAC
GAAGACGATGTATTGTGCGAAGTGCAAATGACAAGGCGGTTGTGCAAATCCCTCCCC
GGTCAAAGGGAAAGTGCTTCAAATCCTCGTCCCGGAGGGAACAGTGGCAACGGTCCG
GCAAACGCTCATCACGCTCGATGCGCCGGGTTATGAAAACATGACGACCGGCAGCGAC
ACCGGCGAAAACCTGTACTTCCAGGGTGGATCCATGAGCGGTTTAGGTAGCGGTTATAT
TGGTCGTGTTTTTGGTATTGGTCGCCAGCAGTTACCGTTGATGAAGTTCTGGCAGAAG
GTGGTTTTTGAATTGTTTTTCTGGTTCGTACCAGCAATGGTATGAAATGTGCACTGAAAC
GCATGTTTGTGAACAACGAACATGATCTGCAGGTTTGTAAACGCGAAATTCAGATTATGC
GTGATCTGAGCGGTCATAAAAAACATTGTGGGCTATATTGATAGCAGCATCAATAATGTTA
GCAGCGGTGATGTTTGGGAAGTTCTGATTCTGATGGATTTTTGTCGTGGTGGTCAGGTT
GTTAATCTGATGAATCAGCGCCTGCAGACCGTTTTACCGAAAATGAAGTGCTGCAGAT
TTTTTGCGATACCTGTGAAGCAGTTGCACGTCTGCATCAGTGTAACCCCGATTATTCA
TCGTGACCTGAAAGTGAAAACATCCTGCTGCATGATCGTGGTCATTATGTTCTGTGTG
ATTTTGGTAGCGCCACCAACAAATTTTCAAGATCCCGCAGACCGAAGGTGTTAATGCAGTT
GAAGATGAGATCAAAAAATACACCACACTGAGCTATCGTGCACCGGAAATGGTTAATCT
GTATAGCGGTAAAATCATCACCACCAAGCAGATATTTGGGCACTGGGTTGTCTGCTGT
ATAAACTGTGTTATTTTACCCTGCCGTTTGGTGAAAGCCAGGTTGCAATTTGTGATGGCA
ATTTACCATTCCGGATAATAGCCGTTATAGCCAGGATATGCATTGCCTGATTGTTATA
TGCTGGAACCTGATCCGGATAAACGTCCGGATATTTATCAGGTGAGCTACTTTTTCATTCA
AACTGCTGAAAAAAGAATGCCCGATTCCGAATGTTTCAAGATAGTCCGATTCCGGCAAAA
CTGCCGGAACCGGTTAAAGCAAGCGAAGCAGCAGCAAAAAAACCCAGCCGAAAGCAC
GTCTGACCGATCCGATTCTACCACCGAAACCAGCATTGCACCGCGTCAGCGTCCGAA
AGCAGGTCAGTAA
```

#### BIRC5 in a pET24a vector:

```
ATGCACCACCACCACCACCACAGCGGCGCTTTTGAATTTAAGCTGCCGGACATTGGCG
AAGGCATCCACGAAGGTGAAATTGTCAAATGGTTTGTGAAACCGGGCGATGAAGTGAAC
GAAGACGATGTATTGTGCGAAGTGCAAATGACAAGGCGGTTGTGCAAATCCCTCCCC
GGTCAAAGGGAAAGTGCTTCAAATCCTCGTCCCGGAGGGAACAGTGGCAACGGTCCG
GCAAACGCTCATCACGCTCGATGCGCCGGGTTATGAAAACATGACGACCGGCAGCGAC
ACCGGCGAAAACCTGTACTTCCAGGGTGGATCCATGGGAGCACCAACCTTGCCACCTG
CTTGGCAACCGTTTCTGAAAGACCATCGCATTAGCACCTTCAAGAATTGGCCGTTTCTG
GAGGGATGTGCGTGTACTCCGGAACGCATGGCAGAAGCAGGGTTTATCCATTGCCCAA
CGGAGAATGAACCGGATATGGCGCAATGCTTCTTTTGTCTCAAGGAGCTTGAAGGCTGG
GAGCCTGATGACGATCCTATTGAGGAACACAAGAAGCACAGTTCTGGGTGTGCGTTTCT
GTCTGTGAAGAAACAGTTTGAAGAGCTGACGTTGGGCGAATTTCTCAAACCTGGATCGGG
AACGGGCAAGAACAATAATCGCCAAGAGACAAACAACAAGAAGAAAGAGTTTGAGGAA
ACGGCCAAGAAAGTTTCGCCGTGCCATTGAGCAGTTAGCAGCCATGGATTGA
```

#### CAMK1G in a pET24a vector:

```
ATGCACCACCACCACCACCACAGCGGCGCTTTTGAATTTAAGCTGCCGGACATTGGCG
AAGGCATCCACGAAGGTGAAATTGTCAAATGGTTTGTGAAACCGGGCGATGAAGTGAAC
GAAGACGATGTATTGTGCGAAGTGCAAATGACAAGGCGGTTGTGCAAATCCCTCCCC
GGTCAAAGGGAAAGTGCTTCAAATCCTCGTCCCGGAGGGAACAGTGGCAACGGTCCG
```

GCAAACGCTCATCACGCTCGATGCGCCGGGTTATGAAAACATGACGACCGGCAGCGAC  
ACCGGCGAAAACCTGTACTTCCAGGGTGGATCCATGGGTTCGTAAAGAAGAGGACGATT  
GTAGCAGCTGGAAAAACAGACCACCAATATTCGAAAACCTTTATCTTCATGGAAGTGT  
TAGGTAGCGGTGCATTTAGCGAAGTTTTTCTGGTTAAACAGCGTCTGACCGGTAAACTG  
TTTGCACGTGAAATGCATCAAAAAAAGTCCGGCATTTCGTGATAGCTCCCTGGAAAATGAA  
ATTGCCGTTCTGAAAAAATCAAACACGAGAACATTGTGACCCTGGAAGATATTTATGAA  
AGCACCACGCATTATTACCTGGTTATGCAGCTGGTTAGCGGTGGTGAAGTGTGATCG  
TATTCTGGAACGTGGTGTGTATACCGAAAAAGATGCAAGCCTGGTGATTCAGCAGGTTT  
TGAGCGCAGTTAAATATCTGCATGAAAATGGTATCGTGCACCGTGATCTGAAACCGGAA  
AATCTGCTGTATCTGACTCCGGAAGAAAACAGCAAAATTATGATTACCGATTTTCGGCCTG  
AGCAAGATGGAACAGAATGGTATTATGAGCACCGCATGTGGTACACCGGGTTATGTTGC  
ACCGGAAGTTCTGGCACAGAAACCGTATAGCAAAGCAGTTGATTGTTGGAGCATTGGCG  
TGATTACCTATATTCTGCTGTGTGGTTATCCGCCTTTTTATGAAGAAACCGAAAGCAAAC  
TGTTTCGAGAAAATCAAAGAAGGCTATTACGAATTTGAAAGCCCCTTCTGGGATGATATTA  
GCGAAAGCGCAAAGATTTTATCTGCCACCTTCTGGAAAAAGATCCGAATGAACGTTATA  
CCTGTGAAAAAGCACTGAGCCATCCGTGGATTGATGGTAATACCGCACTGCATCGTGAT  
ATTTATCCGAGCGTTAGCCTGCAGATTCAGAAGAATTTGCCAAAAGCAAATGGCGTCA  
GGCATTAAATGCAGCAGCAGTTGTTTCATCATATGCGCAAACCTGTAA

DOT1L in a pET24a vector:

ATGCACCACCACCACCACCACAGCGGCGCTTTTTGAATTTAAGCTGCCGGACATTGGCG  
AAGGCATCCACGAAGGTGAAATTGTCAAATGTTTTGTGAAACCGGGCGATGAAGTGAAC  
GAAGACGATGTATTGTGCGAAGTGCAAATGACAAGGCGGTTGTGCAAATTCCTCCCC  
GGTCAAAGGGAAAGTGCTTGAATCCTCGTCCCGGAGGGAACAGTGGCAACGGTCGG  
GCAAACGCTCATCACGCTCGATGCGCCGGGTTATGAAAACATGACGACCGGCAGCGAC  
ACCGGCGAAAACCTGTACTTCCAGGGTGGATCCATGGGTGAAAACTGGAAGTGCCTC  
TGAAAAGTCCGGTTGGTGCAGAACCAGGTTTATCCGTGGCCTCTGCCGGTTTTATGAT  
AAACATCACGATGCAGCCCATGAAATCATTGAAACCATTTCGTTGGGTTTGCGAAGAAAT  
CCGGATCTGAAACTGGCCATGGAAAATTATGTGCTGATCGATTATGACACCAAAAAGCTTT  
GAAAGCATGCAGCGTCTGTGCGATAAATACAATCGTGCAATTGATAGCATTACCCAGCT  
GTGGAAAGGTACAACCCAGCCGATGAAACTGAATACCCGTCCGAGCACCGGTCTGCTG  
CGTCATATTCTGCAGCAGGTTTATAATCATAGCGTTACCGATCCGGAAAAGCTGAATAAC  
TATGAACCGTTTAGTCCGGAAGTTTATGGCGAAACCTCATTGATCTGGTTGCACAGATG  
ATCGATGAGATCAAATGACCGATGATGACCTGTTTGTGATTTAGGTAGCGGTGTTGG  
TCAGGTTGTTCTGCAGGTTGCAGCAGCAACCAATTGTAAACATCATTATGGTGTGAAA  
AAGCCGACATTCCGGCAAAGTATGCAGAAACCATGGATCGTGAATTTGCAAATGGATG  
AAATGGTACGGTAAAAACATGCCGAATATACCCTGGAACGTGGTGAATTTCTGAGCGA  
AGAATGGCGTGAACGTATTGCAAATACCAGCGTTATTTTCGTGAACAACCTTTGCATTTGG  
TCCCGAAGTTGATCATCAGCTGAAAGAACGTTTTGCCAATATGAAAGAAGGTGGTCGTA  
TTGTTAGCAGCAAACCGTTTTGCACCGCTGAATTTTCGTATTAATAGCCGTAATCTGAGCG  
ATATTGGCACCATTTATGCGTGTGTTGAACTGAGTCCGCTGAAAGGTAGCGTTAGCTGG  
ACCGGTAACCGGTTAGCTATTATCTGCATACCATTGATCGTACCATCCTGGAAACTAT  
TTTAGCAGCCTGAAAAATCCGAAACTGCGCGAAGAACAAGAAGCAGCACGTCGTCGTCA  
GCAGCGTGAAAGCAAAGCAATGCAGCAACCCCGACCAAAGGTCCGGAAGGTAAAGTT  
GCAGGTCCGGCAGATGCACCGATGGATAGCGGTGCCGAAGAAGAAAAAGCAGGCGCA  
GCAACCGTTAAAAAACCAGTCCGAGCAAAGCACGTAAAAAGAACTGAACAAAAAGGG  
TCGTAATGTCAGGTCGTAAACGTGGTCGTCCGAAAAATGA

DYRK1a in a pNIC28 vector:

ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGAGAACCTGTACTTC  
CAATCCATGAGCTCCCATAAGAAGGAACGTAAGGTGTACAACGATGGTTATGACGACGA  
TAATTACGACTACATCGTTAAGAACGGCGAGAAGTGGATGGATCGCTATGAAATTGACT

CGCTGATCGGTAAAGGCTCTTTCCGGTCAGGTCGTAAAAGCCTATGATCGTGTGGAGCAA  
 GAATGGGTCGCAATAAAGATCATTAAAGAACAAGAAGGCGTTCTTGAACCAGGCACAGAT  
 CGAGGTCCGCCTGTTGGAAGTATGAATAAGCACGATACCGAGATGAAATACTATATCG  
 TGCATCTGAAACGTCACCTTCATGTTTCGCAATCACCTGTGTTTGGTATTCGAGATGCTGT  
 CATAAACCTGTATGACCTGTTGCGTAATACTAACTTCCGCGGCGTTAGTTTGAATCTGA  
 CACGCAAATTCGCCAACAATGTGCACCGCTTTGTTGTTTCTGGCGACTCCAGAATTG  
 AGCATTATTCATTGTGATCTGAAACCTGAGAACATCCTGTTGTGCAACCCGAAACGCTCC  
 GCAATCAAATTGTCGACTTTGTTTCTCATGTCAGCTGGGCCAGCGTATCTACCAATAC  
 ATTCAGAGTCGCTTCTATCGTTCCGCCGAAGTGTGCTGGGTATGCCATACGATTTGGC  
 CATCGACATGTGGAGCCTGGGCTGCATTCTGGTAGAAATGCATACGGGAGAACCTTTGT  
 TTTCCGGAGCAAACGAAGTGGATCAAATGAACAAGATCGTTGAGGTCCTGGGTATTCCG  
 CCCGCGCACATCTTGGACCAGGCCCAAAAGCACGCAAATCTTTGAAAAGCTGCCTGA  
 TGGCACCTGGAACCTCAAAAAACGAAGGACGGTAAACGTGAATATAAGCCCCCGGC  
 ACACGCAAATTGCATAACATCCTGGGTGTAGAAACCGGCGGTCCAGGCGGTCTGTCGCG  
 CTGGCGAATCTGGTCATACGGTAGCCGATTACTTGAAGTTCAAAGACCTGATTTTGCCT  
 ATGCTGGATTATGACCCTAAAACCCGCATCCAACCGTACTATGCGCTGCAGCATTCAAT  
 CTTTAAAGAAAACAGCAGATGAGTGA

JNK2 in a pET24a vector:

ATGCACCACCACCACCACCACAGCGGCGCTTTTTGAATTTAAGCTGCCGGACATTGGCG  
 AAGGCATCCACGAAGGTGAAATTGTCAAATGTTTTGTGAAACCGGGCGATGAAGTGAAC  
 GAAGACGATGTATTGTGCGAAGTGCAAATGACAAGGCGGTTGTGCAAATTCCTCCCC  
 GGTCAAAGGGAAAGTGCTTGAATCCTCGTCCCGGAGGGAACAGTGGCAACGGTCCG  
 GCAAACGCTCATCACGCTCGATGCGCCGGGTTATGAAAACATGACGACCGGCAGCGAC  
 ACCGGCGAAAACCTGTACTTCCAGGGTGGATCCATGGACAGCCAGTTTTATAGCGTTCA  
 GGTTGCAGATAGCACCTTTACCGTTCTGAAACGTTATCAGCAGCTGAAACCGATTGGTA  
 GCGGTGCACAGGGTATTGTTTGTGCAGCATTGATAACCGTGGTGGTATTAATGTTGCC  
 GTTAAAAACTGAGCCGTCCGTTCCAGAATCAGACCCATGCAAACGTGCATATCGTGA  
 ACTGGTTCTGCTGAAATGCGTGAACCATAAAAACATTATTAGCCTGCTGAACGTGTTTAC  
 ACCGCAGAAAACACTGGAAGAATTCAGGATGTTTACCTGGTTATGGAAGTATGGATG  
 CAAATCTGTGTCAGGTGATTCACATGGAAGTGGATCATGAACGTATGAGCTATCTGCTG  
 TATCAGATGCTGTGTGGTATTAACATCTGCATAGCGCAGGTATTATTCACCGTGATCTG  
 AAACCGTCAAACATTGTTGTTAAAAGCGATTGCACCCTGAAAATTCTGGATTTTGGTCTG  
 GCACGTACCGCAAGCACCAATTTATGATGACCCCGTATGTTGTTACCCGTTATTACCG  
 TGCACCGGAAGTTATTCTGGGTATGGGTTATAAAGAAAACGTGGATATTTGGAGCGTGG  
 GTTGTATTATGGGCGAACTGGTTAAAGGTAGCGTTATTTTTAGGGCACCAGTACATT  
 GATCAGTGGAAATAAAGTTATTGAACAGCTGGGCACCCCGAGCGCAGAATTTATGGCAGC  
 ACTGCAGCCGACCGTTTCGTAATTATGTTGAAAATCGTCCGGCATATCCGGGTATTGCAT  
 TTGAAGAAGTGTTCGCGATTGGATTTTTCCGAGCGAAAGCGAACGTGATAAAATCAAAA  
 CCAGCCAGGCACGTGATCTGCTGAGCAAATGCTGGTTATTGATCCGGATAAACGTATC  
 AGCGTTGATGAAGCACTGCGTCATCCGTATATTACAGTTTGGTATGATCCGGCAGAAGC  
 AGAAGCACCGCCTCCGCAGATTTATGATGCACAGCTGGAAGAACGTGAACATGCAATTG  
 AAGAATGGAAGAAGTACTGATCTATAAAGAAGTATGGATTGA

JNK3 in a pET24a vector:

ATGCACCACCACCACCACCACAGCGGCGCTTTTTGAATTTAAGCTGCCGGACATTGGCG  
 AAGGCATCCACGAAGGTGAAATTGTCAAATGTTTTGTGAAACCGGGCGATGAAGTGAAC  
 GAAGACGATGTATTGTGCGAAGTGCAAATGACAAGGCGGTTGTGCAAATTCCTCCCC  
 GGTCAAAGGGAAAGTGCTTGAATCCTCGTCCCGGAGGGAACAGTGGCAACGGTCCG  
 GCAAACGCTCATCACGCTCGATGCGCCGGGTTATGAAAACATGACGACCGGCAGCGAC  
 ACCGGCGAAAACCTGTACTTCCAGGGTGGATCCATGAGCAAAGCAAAGTTGACAACCA  
 GTTCTACAGTGTGGAAGTGGGAGACTCAACCTTCACAGTTCTCAAGCGCTACCAGAATC

TAAAGCCTATTGGCTCTGGGGCTCAGGGCATAGTTTGTGCCGCGTATGATGCTGTCCCTT  
GACAGAAATGTGGCCATTAAGAAGCTCAGCAGACCCTTTCAGAACCAAACACATGCCAA  
GAGAGCGTACCGGGAGCTGGTCCTCATGAAGTGTGTGAACCATAAAAACATTATTAGTT  
TATTAATGTCTTCACACCCCAGAAAACGCTGGAGGAGTTCCAAGATGTTTACTTAGTAA  
TGGAAGTATGGATGCCAACTTATGTCAAGTGATTCAGATGGAATTAGACCATGAGCGA  
ATGTCTTACCTGCTGTACCAAATGTTGTGTGGCATTAAAGCACCTCCATTCTGCTGGAATT  
ATTCACAGGGATTTAAAACCAAGTAACATTGTAGTCAAGTCTGATTGCACATTGAAAATC  
CTGGACTTTGGACTGGCCAGGACAGCAGGCACAAGCTTCATGATGACTCCATATGTGGT  
GACACGTTATTACAGAGCCCCTGAGGTCATCCTGGGGATGGGCTACAAGGAGAACGTG  
GATATATGGTCTGTGGGATGCATTATGGGAGAAATGGTTCGCCACAAAATCCTCTTTCC  
AGGAAGGGACTATATTGACCAGTGGAATAAGGTAATTGAACAACCTAGGAACACCATGTC  
CAGAATTCATGAAGAAATTGCAACCCACAGTAAGAACTATGTGGAGAATCGGCCCAAG  
TATGCGGGACTCACCTTCCCCAACTCTTCCCAGATTCCCTCTTCCCAGCGGACTCCGA  
GCACAATAAACTCAAAGCCAGCCAAGCCAGGGACTTGTGTCAAAGATGCTAGTGATTG  
ACCCAGCAAAAAGAATATCAGTGGACGACGCCTTACAGCATCCCTACATCAACGTCTGG  
TATGACCCAGCCGAAGTGGAGGCGCCTCCACCTCAGATATATGACAAGCAGTTGGATG  
AAAGAGAACACACAATTGAAGAATGGAAAGAACTTATCTACAAGGAAGTAATGAATTCAG  
AATGA

TEV-protease in a pMal-C2 vector:

GAAAATCTTTATTTTCAAGGTCATCATCATCATCATCATGGAGAAAGCTTGTTTAAGG  
GGCCGCGTGATTACAACCCGATATCGAGCACCATTTGTCATTTGACGAATGAATCTGAT  
GGGCACACAACATCGTTGTATGGTATTGGATTTGGTCCCTTCATCATTACAAACAAGCAC  
TTGTTTAGAAGAAATAATGGAACACTGTTGGTCCAATCACTACATGGTGTATTCAAGGTC  
AAGAACACCACGACTTTGCAACAACACCTCATTGATGGGAGGGACATGATAATTATTCG  
CATGCCTAAGGATTTCCACCATTTCCCTCAAAGCTGAAATTTAGAGAGCCACAAAGGG  
AAGAGCGCATATGTCTTGTGACAACCAACTTCCAACTAAGAGCATGTCTAGCATGGTG  
TCAGACACTAGTTGCACATTCCCTTCATCTGATGGCATATTCTGGAAGCATTGGATTCAA  
ACCAAGGATGGGCAGTGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTTATTGTTGG  
TATACTCAGCATCGAATTTACCAACACAAACAATTATTTACAAGCGTGCCGAAAAA  
CTTCATGGAATTGTTGACAAATCAGGAGGCGCAGCAGTGGGTTAGTGTTGGCGATTAA  
ATGCTGACTCAGTATTGTGGGGGGGCCATAAAGTTTTTCATGCCGAAACCTGAAGAGCCT  
TTTCAGCCAGTTAAGGAAGCGACTCAACTCATGAATCGTCGTCGCCGTCGCTAA

GSK3 $\beta$ :

Optimized for *E. coli* in a pET24a vector:

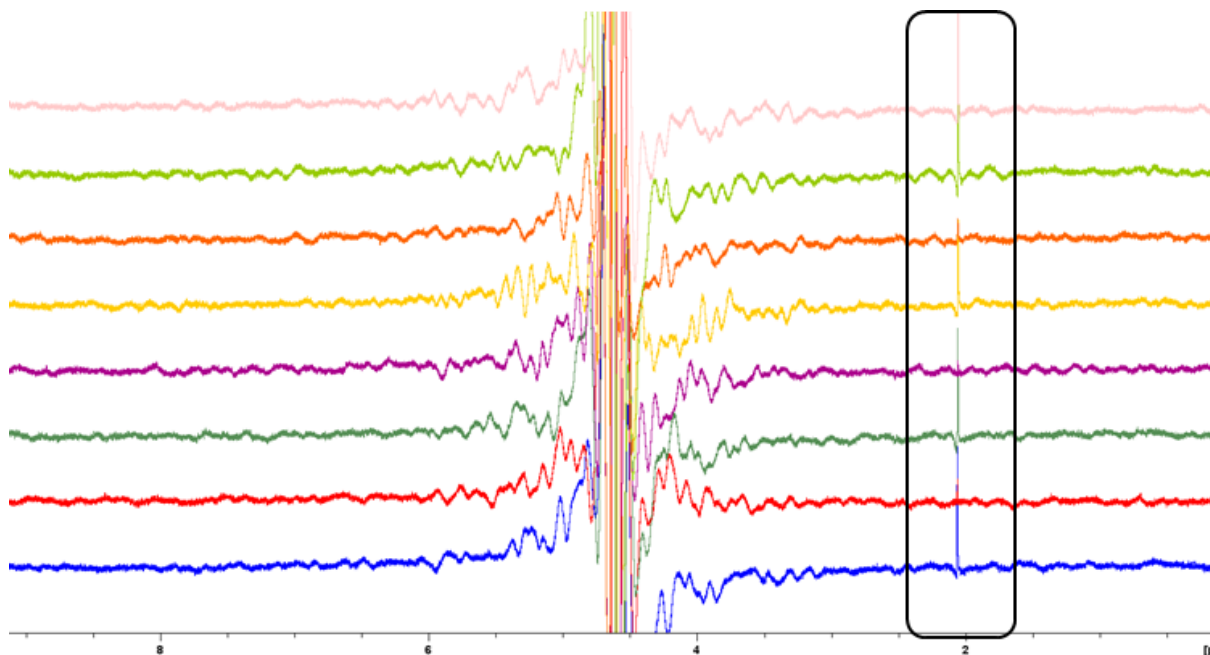
ATGCACCACCACCACCACAGCGGCGCTTTTGAATTTAAGCTGCCGGACATTGGCG  
AAGGCATCCACGAAGGTGAAATTGTCAAATGTTTTGTGAAACCGGGCGATGAAGTGAAC  
GAAGACGATGTATTGTGCGAAGTGCAAAATGACAAGGCGGTTGTGCGAAATTCCTCCCC  
GGTCAAAGGGAAAGTGCTTGAAATCCTCGTCCCGGAGGGAAACAGTGGCAACGGTCCG  
GCAAACGCTCATCACGCTCGATGCGCCGGGTTATGAAAACATGACGACCGGCAGCGAC  
ACCGGCGAAAACCTGTACTTCCAGGGTGGATCCATGAGCGGTGTCGCGTACCACCA  
GTTTTGCAGAAAGCTGTAAACCGTTTCCAGCAGCCGAGCGCATTGTTGGTAGCATGAAAGTT  
AGCCGTGATAAAGATGGTAGCAAAGTTACCACCGTTGTTGCAACCCCTGGTCAGGGTCC  
TGATCGTCCGCAAGAAGTTAGCTATACCGATAACAAAGTTATTGGCAATGGTAGCTTTG  
GTGTTGTGTATCAGGCAAACTGTGTGATAGCGGTGAACTGGTTGCAATCAAAAAAGTT  
CTGCAGGACAAACGCTTTAAAACCGTGAAGTGCAGATTATGCGTAAACTGGATCATTG  
CAATATTGTGCGTCTGCGCTATTTTTCTATAGCAGCGGTGAAAAAAGATGAAGTGTA  
TCTGAATCTGGTGTGATTATGTTCCGGAAACCGTTTATCGTGTTGCACGTCAATTATAG  
CCGTGCAAAACAGACCCTGCCGGTTATTTATGTTAAGCTGTATATGTATCAGCTGTTTCG

TAGCCTGGCATATATTCATAGCTTTGGCATTGTCACCGTGATATTAACCGCAGAATCT  
GCTGCTTGATCCGGATACCGCAGTTCTGAAACTGTGCGATTTTGGTAGCGCAAACAGC  
TGGTTCGTGGTGAACCGAATGTGAGCTATATTTGTAGCCGTTATTATCGTGCACCGGAA  
CTGATTTTTGGTGCAACCGATTATACCAGCAGCATTGATGTTTGGAGCGCAGGTTGTGT  
TCTGGCCGAAGTCTGCTGGGTCAGCCGATTTTTCCGGGTGATAGTGGTGTGATCAG  
CTGGTTGAAATTATCAAAGTTCTGGGCACCCCGACACGTGAGCAGATTCGTGAAATGAA  
TCCGAATTATACCGAGTTCAAATTCCCGCAGATTAAGCACATCCGTGGACCAAAGTTTT  
TCGTCCTCGTACACCGCCTGAAGCAATTGCACTGTGTAGCCGTCTGCTGGAATATACCC  
CGACCGCACGTCTGACACCGCTGGAAGCATGTGCCATAGTTTTTTTTGATGAACTGCGT  
GATCCGAATGTTAAACTGCCGAATGGTCGTGATACACCGGCACTGTTTAACTTTACCAC  
ACAAGAAGTGTGAGCAGCAATCCGCCTCTGGCAACCATTCTGATTCCGCCTCATGCACGTA  
TTCAGGCAGCAGCATCAACCCCGACCAATGCAACCGCAGCAAGTGATGCAAATACCGG  
TGATCGTGGTTCAGACCAATAATGCAGCAAGCGCAAGCGCCAGCAATAGCACCTAA

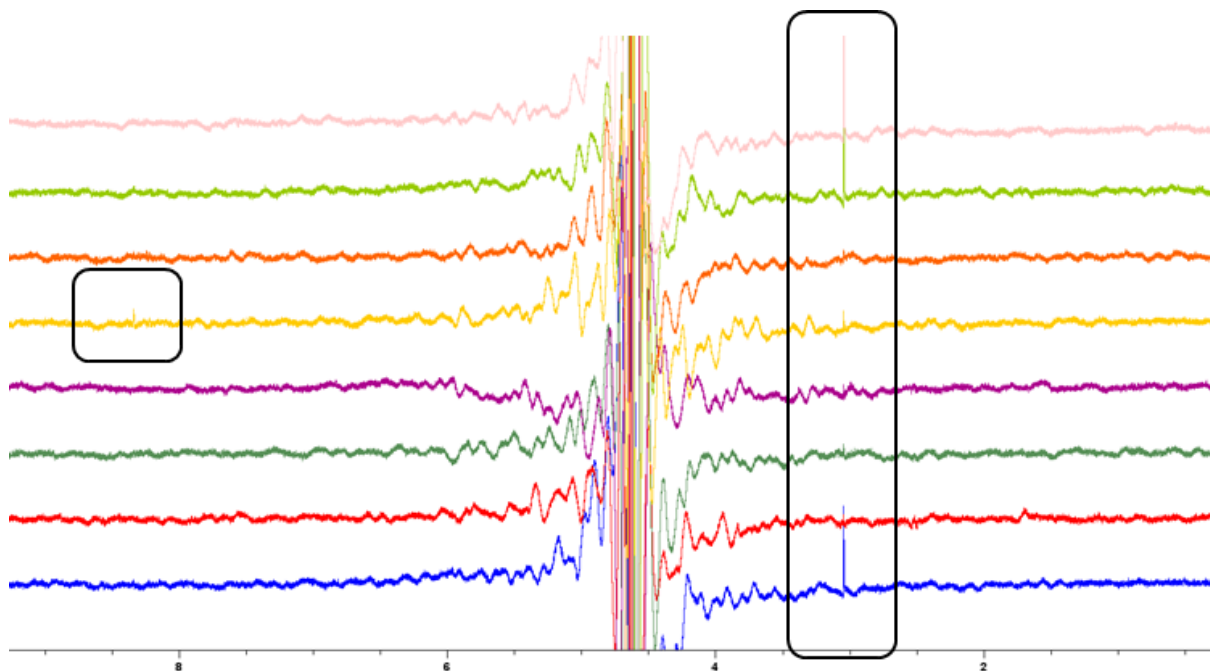
Wildtype in a pET15b vector:

ATGTCAGGGCGGCCAGAACCCACCTCCTTTGCGGAGAGCTGCAAGCCGGTGCAGCAG  
CCTTCAGCTTTTGGCAGCATGAAAGTTAGCAGAGACAAGGACGGCAGCAAGGTGACAA  
CAGTGGTGGCAACTCCTGGGCAGGGTCCAGACAGGCCACAAGAAGTCAGCTATACAGA  
CACTAAAGTGATTGAAATGGATCATTGTTGGTGTGGTATATCAAGCCAACTTTGTGATTC  
AGGAGAACTGGTCGCCATCAAGAAAGTATTGCAGGACAAGAGATTTAAGAATCGAGAGC  
TCCAGATCATGAGAAAGCTAGATCACTGTAACATAGTCCGATTGCGTTATTTCTTCTACT  
CCAGTGGTGAGAAGAAAGATGAGGTCTATCTTAATCTGGTGTGACTATGTTCCGGAA  
ACAGTATACAGAGTTGCCAGACACTATAGTCGAGCCAAACAGACGCTCCCTGTGATTTA  
TGTCAAGTTGTATATGTATCAGCTGTTCCGAAGTTTAGCCTATATCCATTCCTTTGGAATC  
TGCCATCGGGATATTAACCGCAGAACCTCTTGTGGATCCTGATACTGCTGTATTA  
CTCTGTGACTTTGGAAGTGCAAAGCAGCTGGTCCGAGGAGAACCCAATGTTTCGTATAT  
CTGTTCTCGGTACTATAGGGCACCAGAGTTGATCTTTGGAGCCACTGATTATACCTCTA  
GTATAGATGTATGGTCTGCTGGCTGTGTGTTGGCTGAGCTGTTACTAGGACAACCAATA  
TTTCCAGGGGATAGTGGTGTGGATCAGTTGGTAGAAATAATCAAGGTCTGGGAACTCC  
AACAAGGGAGCAAATCAGAGAAATGAACCCAACTACACAGAATTTAAATTCCTCAAT  
TAAGGCACATCCTTGGACTAAGGTCTTCCGACCCCGAACTCCACCGGAGGCAATTGCA  
CTGTGTAGCCGTCTGCTGGAGTATACACCAACTGCCCGACTAACACCACTGGAAGCTTG  
TGCACATTCATTTTTTGTGATGAATTACGGGACCCAAATGTCAAACCTACCAAATGGGCGAGA  
CACACCTGCACTCTTCAACTTCACCACTCAAGAAGTGTCAAGTAATCCACCTCTGGCTAC  
CATCCTTATTCCTCCTCATGCTCGGATTCAAGCAGCTGCTTCAACCCCAACAATGCCA  
CAGCAGCGTCAGATGCTAATACTGGAGACCGTGGACAGACCAATAATGCTGCTTCTGCA  
TCAGCTTCCAACCTCCACC

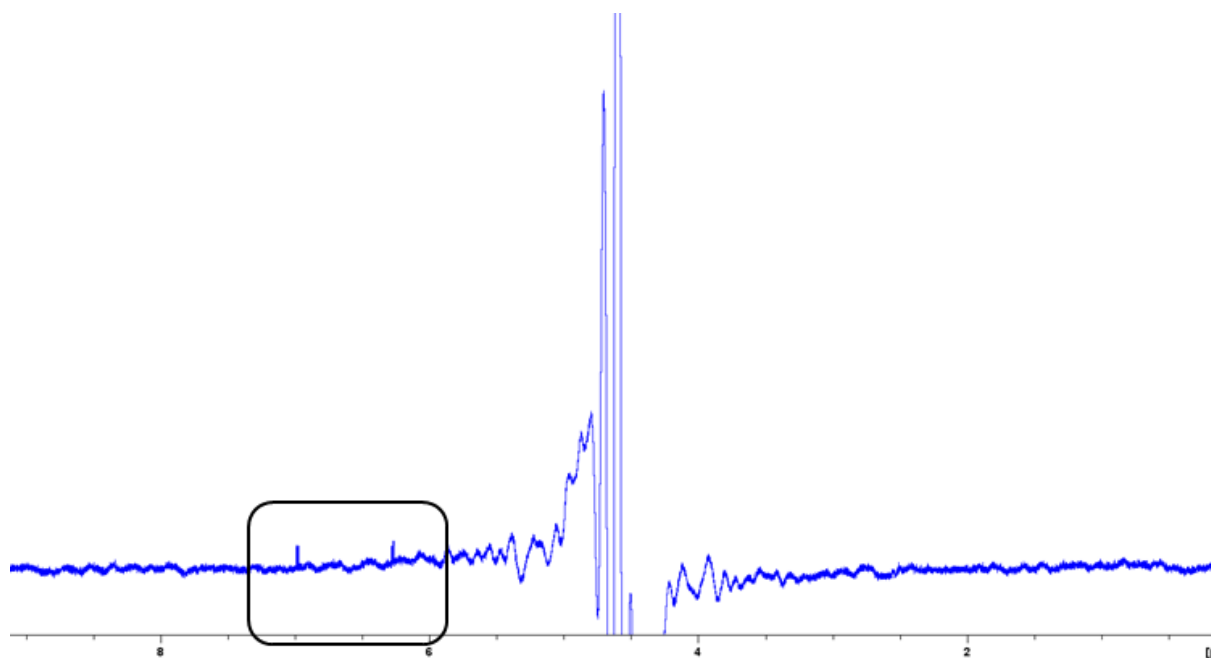
## Appendix B: STD-NMR spectra



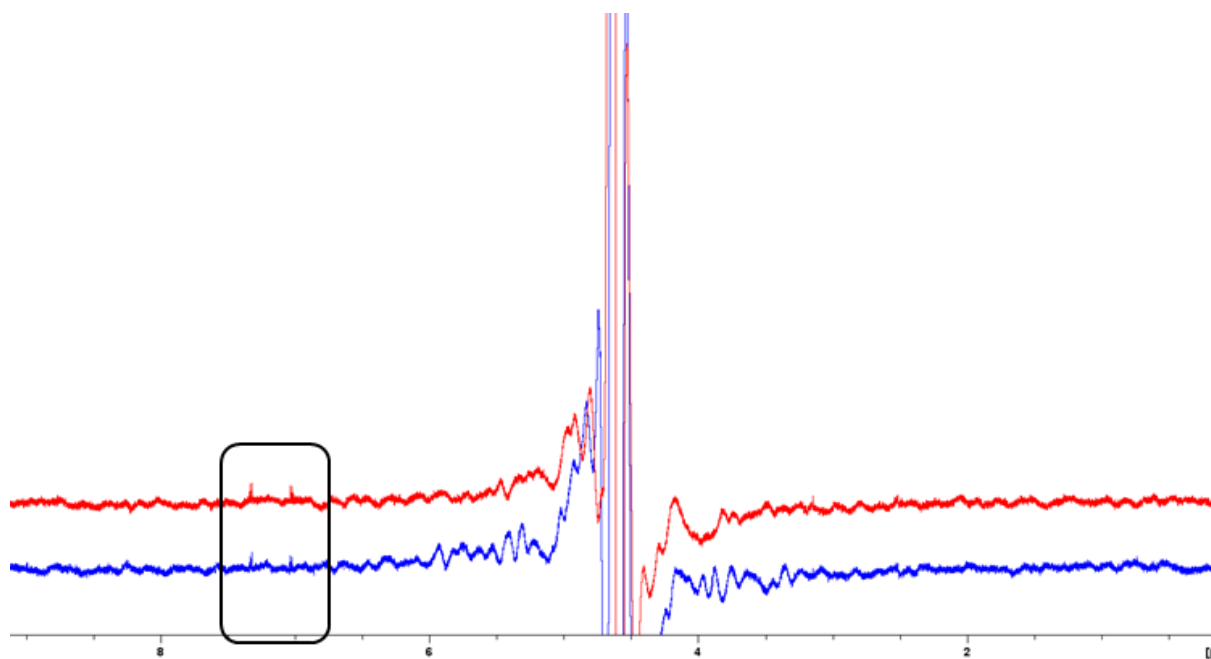
Appendix B 1. STD spectra overlay of 0459 with AAK1 (blue), BIRC5 (red), CAMK1G (green), DOT1L (violet), DYRK1a (yellow), IDO1 (orange), JNK2 (light green), and JNK3 (pink). The black box marks the positive signal.



Appendix B 2. STD spectra overlay of 0468 and 1234 with AAK1 (blue), BIRC5 (red), CAMK1G (green), DOT1L (violet), DYRK1a (yellow), IDO1 (orange), JNK2 (light green), and JNK3 (pink). The black box on the left marks the positive signal for 0468 and the box on the right for 1234.

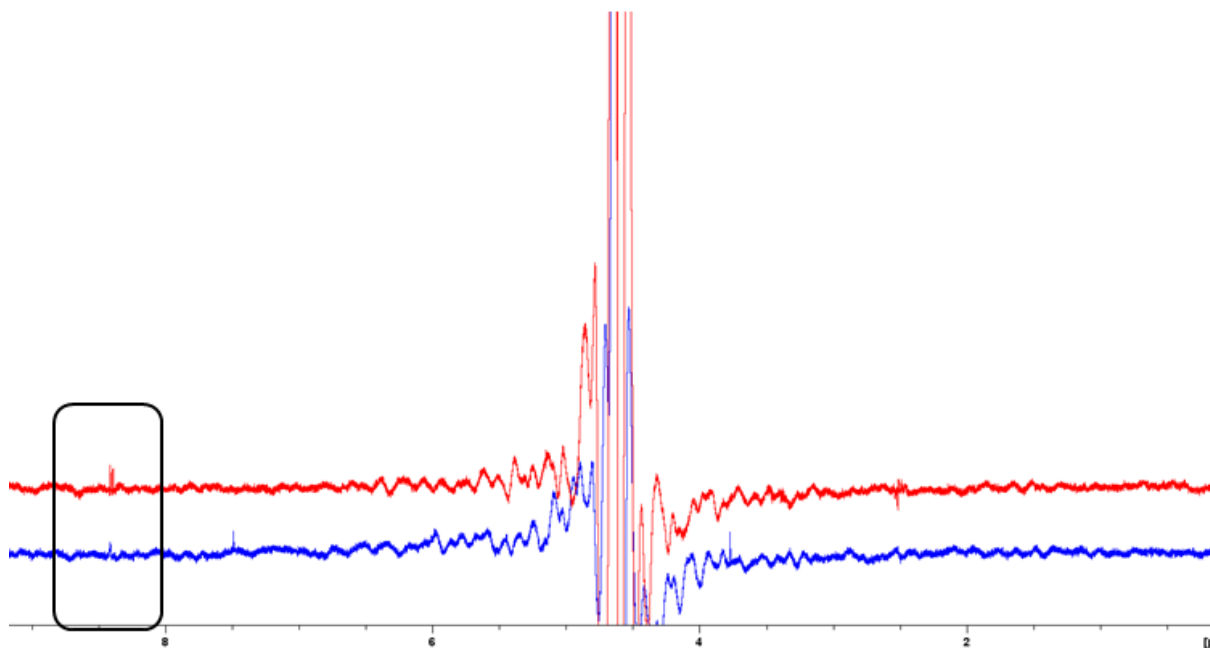


Appendix B 3. STD spectrum of 0471 with AAK1 (blue). The black box marks the positive signals.

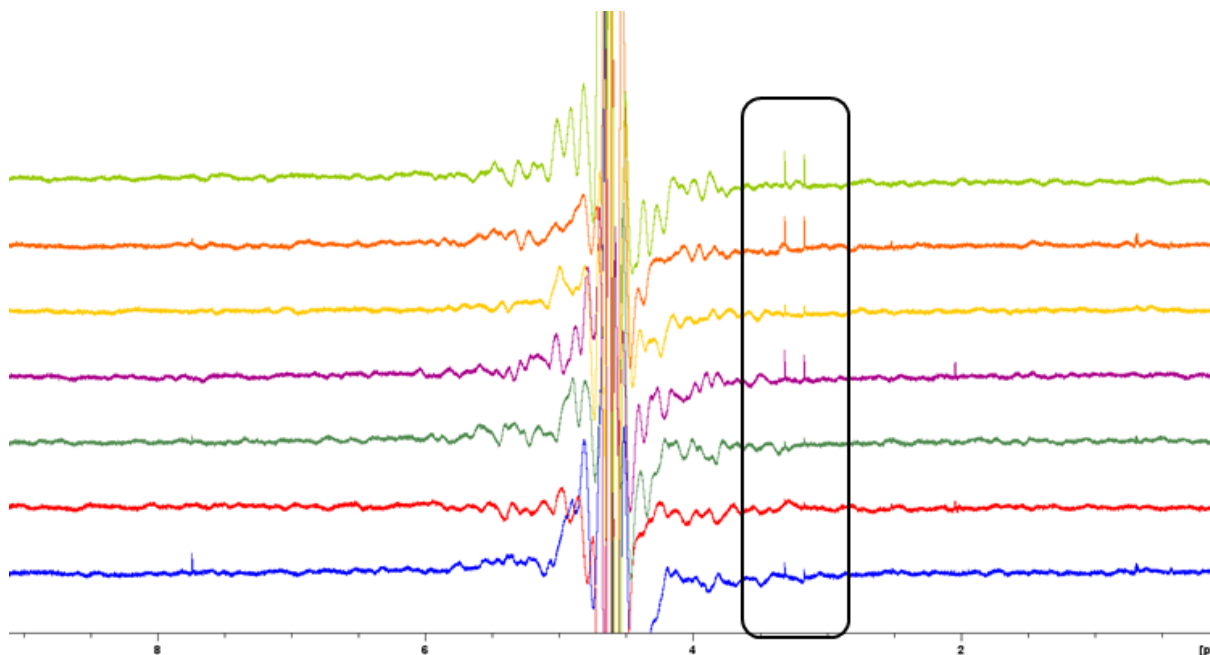


Appendix B 4. STD spectra overlay of 0474 with AAK1 (blue) and JNK2 (red). The black box marks the positive signals.

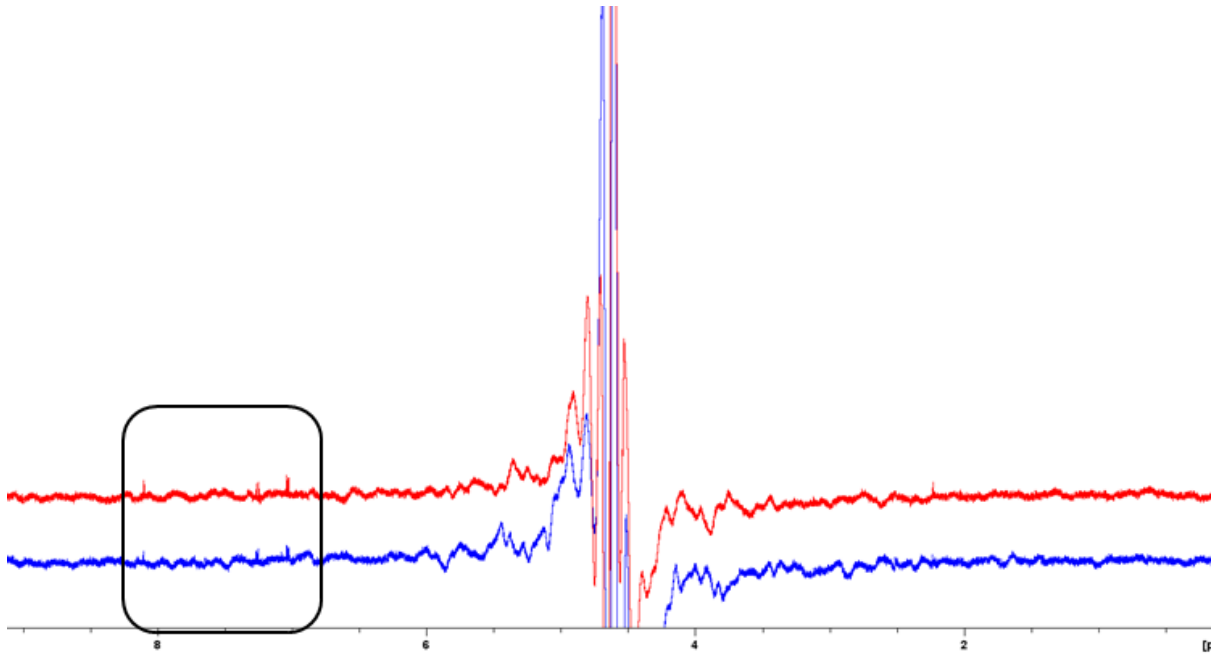




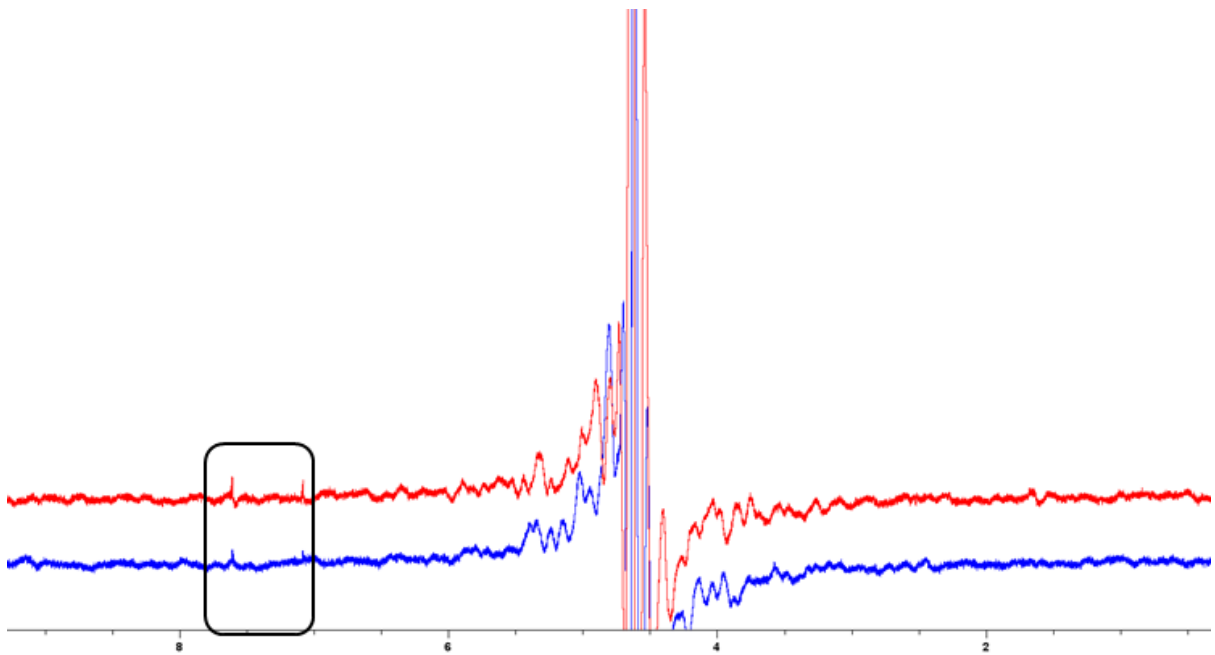
Appendix B 5. STD spectra overlay of 0482 with DYRK1a (blue) and JNK2 (red). The black box marks the positive signal.



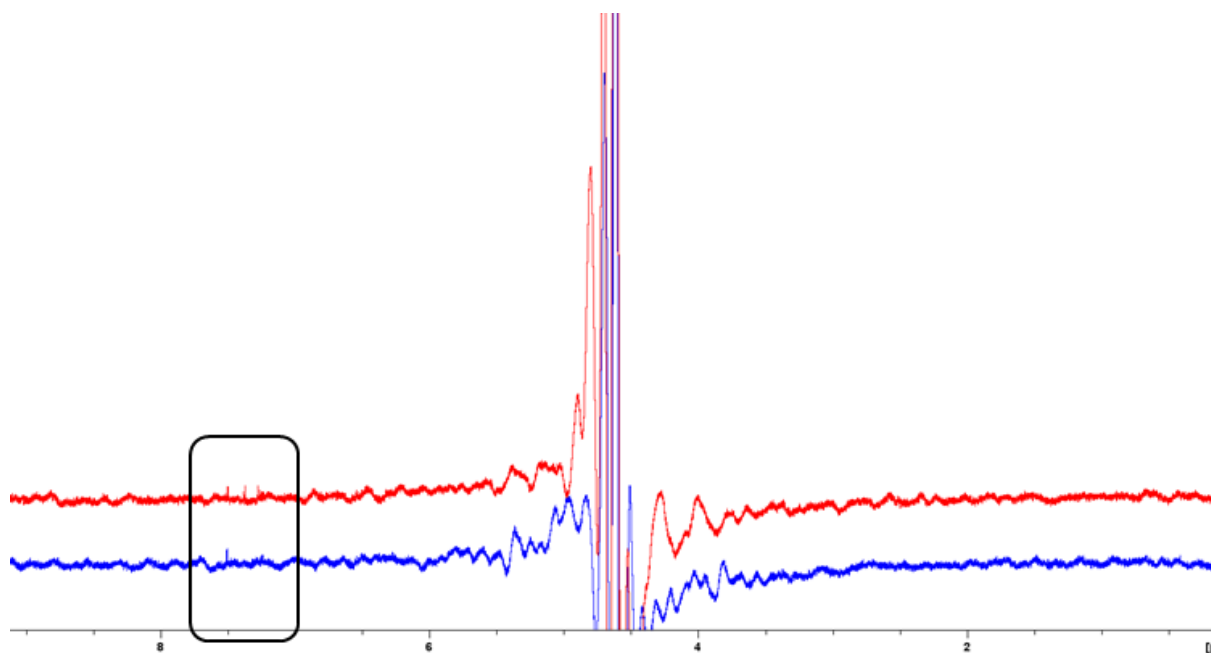
Appendix B 6. STD spectra overlay of 1216 with AAK1 (blue), BIRC5 (red), CAMK1G (green), DYRK1a (violet), IDO1 (yellow), JNK2 (orange), and JNK3 (light green). The black box marks the positive signals.



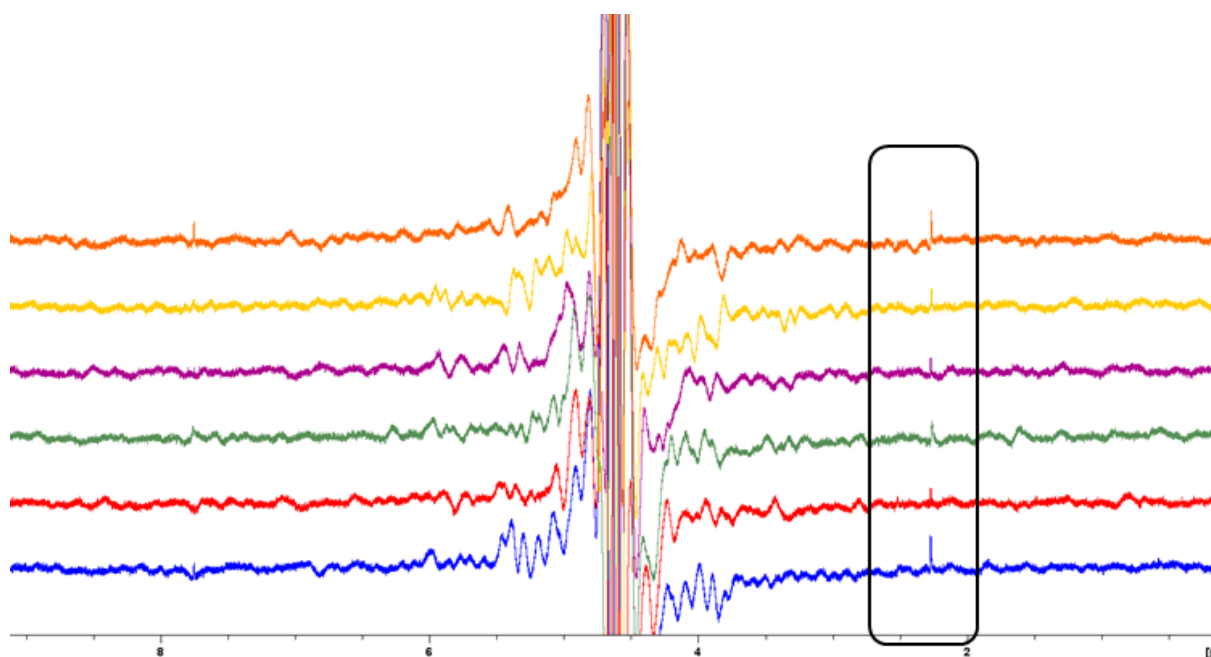
Appendix B 7. STD spectra overlay of 1217 with AAK1 (blue) and JNK2 (red). The black box marks the positive signals.



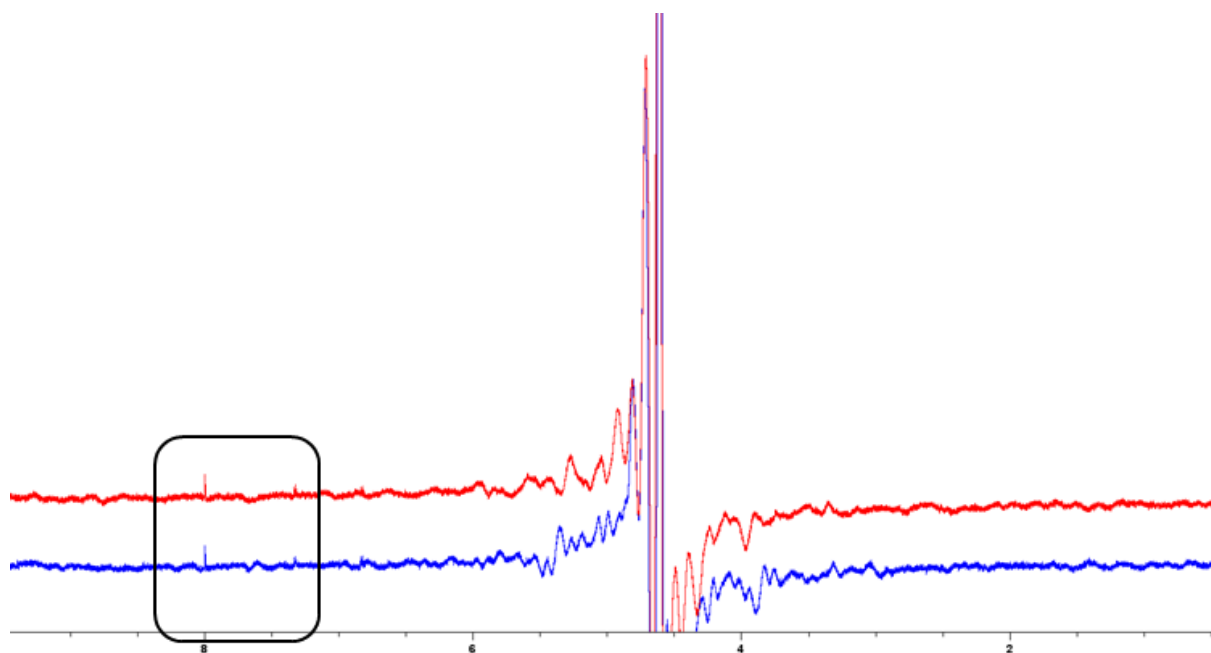
Appendix B 8. STD spectra overlay of 1233 with AAK1 (blue) and DYRK1a (red). The black box marks the positive signals.



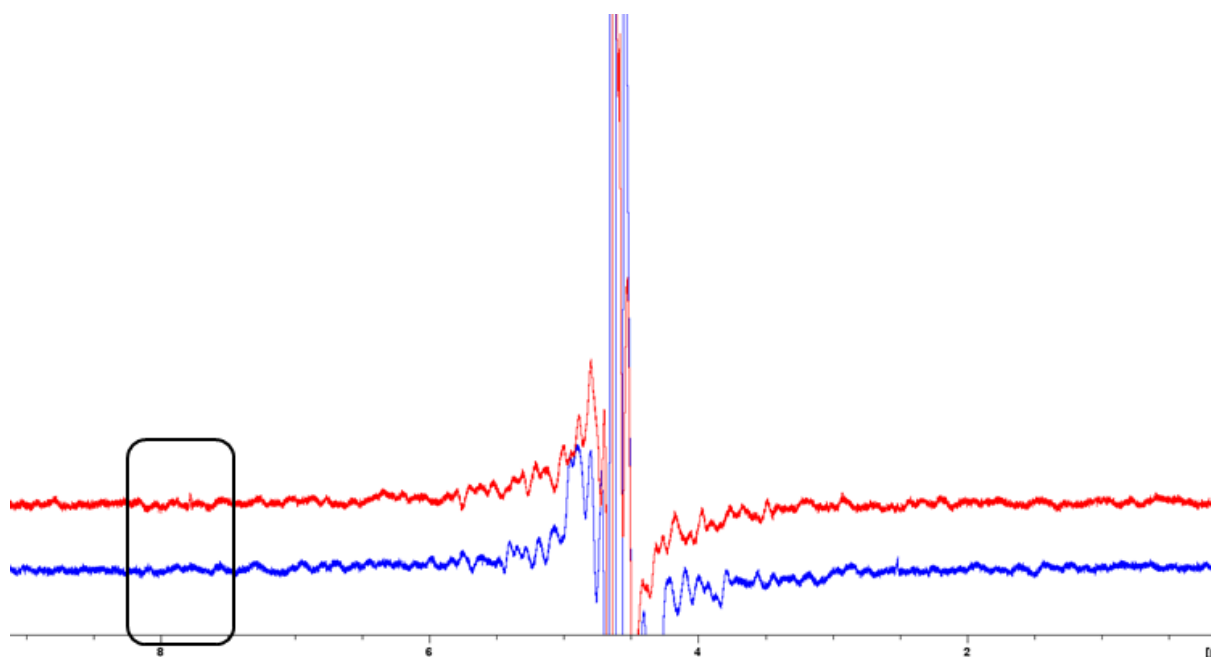
Appendix B 9. STD spectra overlay of 1257 with AAK1 (blue) and JNK2 (red). The black box marks the positive signal.



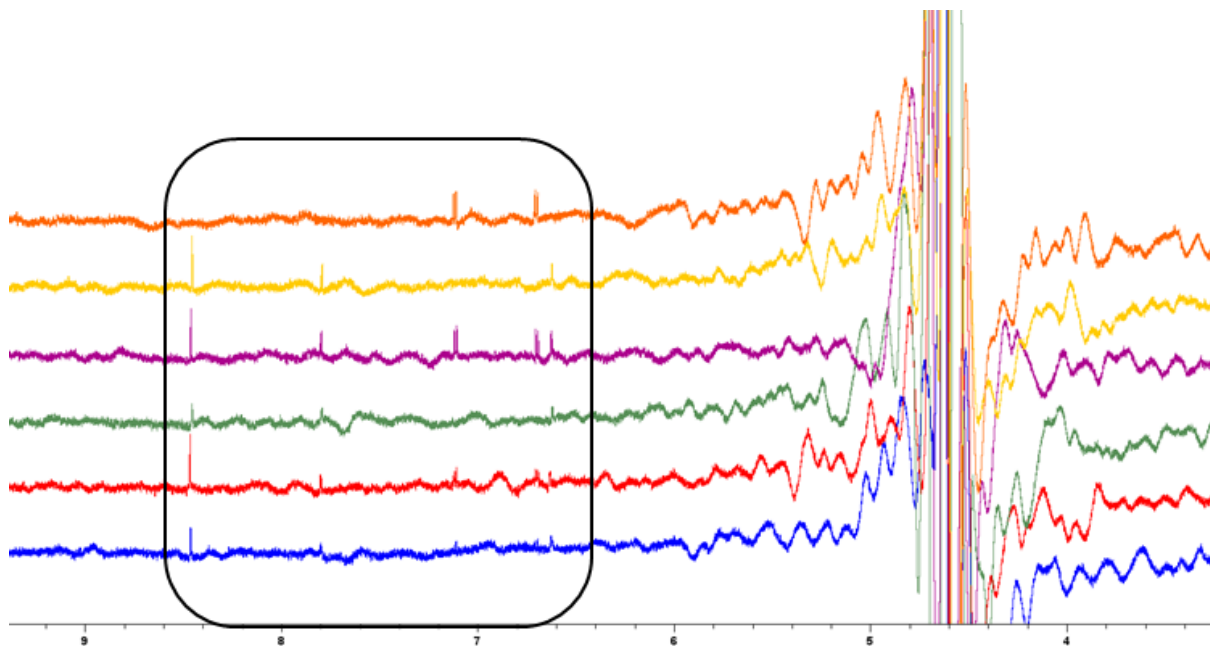
Appendix B 10. STD spectra overlay of 7405 with AAK1 (blue), BIRC5 (red), CAMK1G (green), DOT1L (violet), DYRK1a (yellow), and JNK2 (orange). The black box marks the positive signal.



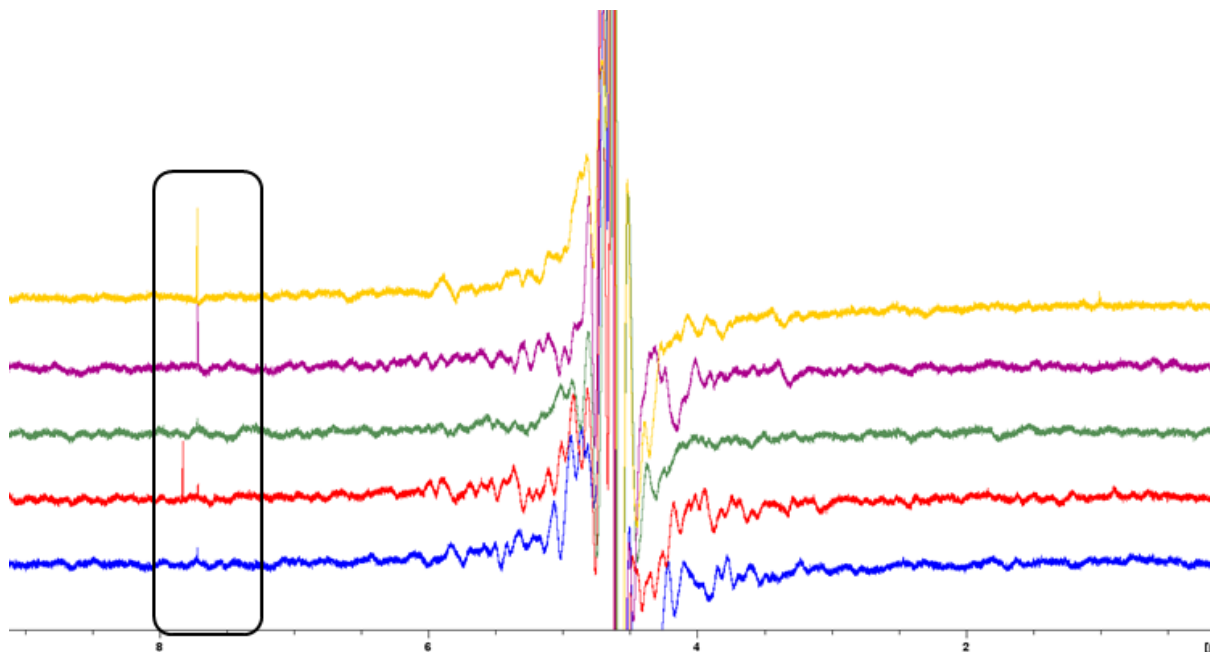
Appendix B 11. STD spectra overlay of 7409 with AAK1 (blue) and JNK3 (red). The black box marks the positive signals.



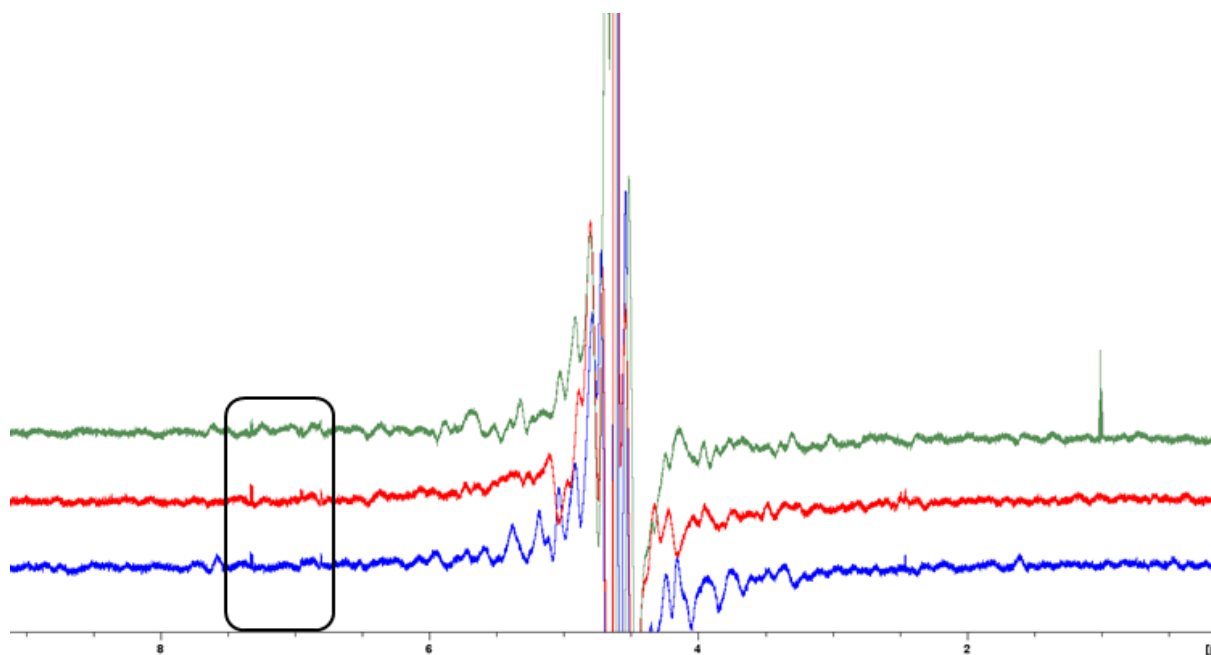
Appendix B 12. STD spectra overlay of 7419 with CAMK1G (blue) and DYRK1a (red). The black box marks the positive signal.



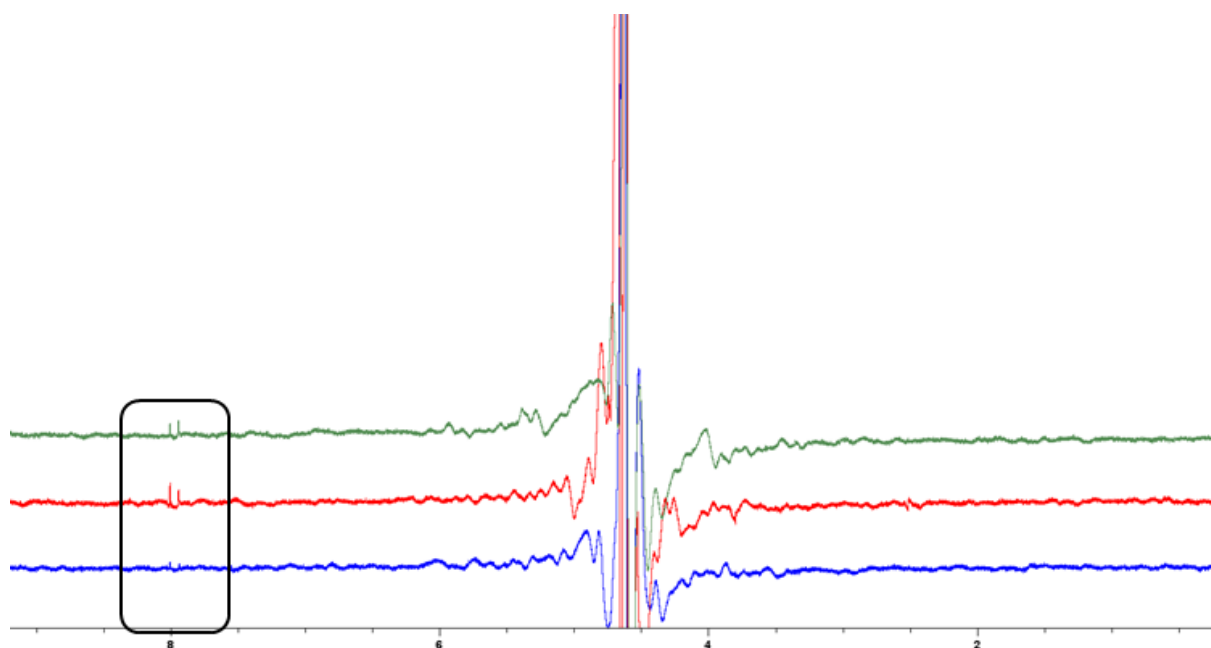
Appendix B 13. STD spectra overlay of 9595 and 4485 with AAK1 (blue), CAMK1G (red), DYRK1a (green), JNK2 (violet), and JNK3 (9595 in yellow and 4485 in orange). The black box marks the positive signals.



Appendix B 14. STD spectra overlay of 9601 with AAK1 (blue), DYRK1a (red), IDO1 (green), JNK2 (violet), and JNK3 (yellow). The black box marks the positive signal.



Appendix B 15. STD spectra overlay of 9605 with CAMK1g (blue), JNK2 (red), and JNK3 (green). The black box marks the positive signals.



Appendix B 16. STD spectra overlay of 9612 with IDO1 (blue), JNK2 (red), and JNK3 (green). The black box marks the positive signals.

## Appendix C: HEFLib Properties

Barcode	SMILES	SlogP	TPSA	Num HBD	Num HBA	HA	MW	V <sub>max</sub>
9594	<chem>[NH3+]C(CCO)c1ccc(Cl)cc1Cl</chem>	1.659	47.87	2	1	13	221.107	0.266/0.229
9595	<chem>Clc1ncnc2cc[nH]c12</chem>	1.611	41.57	1	2	10	153.572	0.13
9596	<chem>Cc1noc(C)c1Br</chem>	2.054	26.03	0	2	8	176.013	0.17
9597	<chem>O=C([O-])c1cccc(Cl)n1</chem>	0.099	53.02	0	3	10	156.548	0.003
9599	<chem>OB(O)c1ccc(Cl)nc1F</chem>	-0.446	53.35	2	3	11	175.355	0.121
9600	<chem>Nc1ncccc1I</chem>	1.268	38.91	1	2	8	220.013	0.178
9601	<chem>O=[N+](O)c1cnc(Br)[nH]1</chem>	1.080	71.82	1	3	9	191.972	0.191
9602	<chem>O=C([O-])c1ccc(Cl)cn1</chem>	0.099	53.02	0	3	10	156.548	0.025
9603	<chem>Nc1cccc(Br)c1O</chem>	1.737	46.25	2	2	9	188.024	0.157
9604	<chem>O=C([O-])c1cncc(Cl)c1</chem>	0.099	53.02	0	3	10	156.548	0.018
9605	<chem>Nc1ccc2[nH]nc(Br)c2c1</chem>	1.908	54.7	2	2	11	212.05	0.149
9606	<chem>OCc1ccc(Cl)nc1</chem>	1.227	33.12	1	2	9	143.573	0.114
9607	<chem>Nc1cnccc1I</chem>	1.268	38.91	1	2	8	220.013	0.341
9608	<chem>COc1nc(N)ncc1I</chem>	0.672	61.03	1	4	10	251.027	0.187
9609	<chem>[NH3+]Cc1cc(Br)no1</chem>	0.179	53.67	1	2	8	178.009	0.286
9610	<chem>BrC1cnn2cccnc12</chem>	1.492	30.19	0	3	10	198.023	0.161
9611	<chem>O=C([O-])c1cncc(Cl)n1</chem>	-0.507	65.91	0	4	10	157.536	0.015
9612	<chem>Nc1ncc(Br)cc1C(=O)[O-]</chem>	-0.210	79.04	1	4	11	216.014	0.038
9613	<chem>CCOC(=O)c1coc(Cl)n1</chem>	1.505	52.33	0	4	11	175.571	0.154
9614	<chem>O=C([O-])c1ncccc1Br</chem>	0.208	53.02	0	3	10	200.999	0.029
9615	<chem>O=C([O-])c1ccc(Cl)c(O)c1</chem>	0.409	60.36	1	3	11	171.559	0.028
9616	<chem>Cc1ccc(N)c(Cl)n1</chem>	1.626	38.91	1	2	9	142.589	0.113
9617	<chem>BrC1cnc2cnccn12</chem>	1.492	30.19	0	3	10	198.023	0.186
9618	<chem>O=c1[nH]cnc([O-])c1Br</chem>	-0.394	68.81	1	3	9	189.976	0.141
9619	<chem>Cc1ncccc1Br</chem>	2.153	12.89	0	1	8	172.025	0.158
9620	<chem>Cn1ccc(Br)n1</chem>	1.183	17.82	0	2	7	161.002	0.149

Barcode	SMILES	SlogP	TPSA	Num HBD	Num HBA	HA	MW	V <sub>max</sub>
9621	<chem>Nc1ncnc(Cl)c1Cl</chem>	1.366	51.8	1	3	9	163.995	0.144/0.127
9809	<chem>Brc1cncnc1</chem>	1.239	25.78	0	2	7	158.986	0.174
0087	<chem>Brc1cnc(N2CCC[NH2+]CC2)nc1</chem>	0.013	45.63	1	3	14	258.143	0.251
0088	<chem>COc1ccc(C[NH3+])c(Cl)c1</chem>	1.091	36.87	1	1	11	172.635	0.258
0089	<chem>CN(N)c1ncc(C(F)(F)F)cc1Cl</chem>	2.064	42.15	1	3	14	225.601	0.143
0090	<chem>Nc1c(Cl)cc(Cl)cc1C[NH3+]</chem>	1.318	53.66	2	1	11	192.069	0.247/0.232
0112	<chem>COc1ccc(I)cc1C[NH2+]CCO</chem>	0.356	46.07	2	2	14	308.139	0.281
0113	<chem>Brc1cnn(-c2nnn[n-]2)c1</chem>	-0.223	70.59	0	5	11	214.006	0.063
0114	<chem>CC(C(=O)[O-])n1cc(Br)cn1</chem>	-0.044	57.95	0	4	11	218.03	0.051
0115	<chem>O=C(c1cc(Br)c[nH]1)N1CCOCC1</chem>	1.250	45.33	1	2	14	259.103	0.154
0116	<chem>Cn1ncc(Br)c1C(=O)[O-]</chem>	-0.454	57.95	0	4	10	204.003	0.023
0117	<chem>[NH3+]CCc1cc(Cl)cn1</chem>	0.169	45.46	1	2	10	160.628	0.201
0118	<chem>NC(=O)c1nc(Cl)n[n-]1</chem>	-0.814	82.97	1	3	9	145.529	-0.021
0119	<chem>O=C([O-])CCc1nc(Br)n[nH]1</chem>	-0.750	81.7	1	4	11	219.018	0.025
0120	<chem>Cn1nc(Br)nc1NN</chem>	-0.137	68.76	2	5	9	192.02	0.158
0121	<chem>Clc1ccc(CN2CC[NH2+]CC2)cn1</chem>	0.114	32.74	1	2	14	212.704	0.193
0122	<chem>Brc1csc(CN2CC[NH2+]CC2)c1</chem>	0.890	19.85	1	2	13	262.196	0.233
0123	<chem>Cc1c(Cl)c([N+](=O)[O-])nn1CC(=O)[O-]</chem>	-0.497	101.09	0	6	14	218.576	0.036
0401	<chem>O=C(c1ccc(Br)cc1)C12CN3CN(CN(C3)C1)C2</chem>	1.438	26.79	0	4	19	322.206	0.243
0402	<chem>O=S(=O)(CCO)c1n[n-]c(Cl)n1</chem>	-1.147	94.25	1	5	12	210.622	-0.004
0403	<chem>COCn1nc(Br)nc1C(N)=O</chem>	-0.257	83.03	1	5	12	235.041	0.163
0404	<chem>COCn1nc(Br)[n-]c1=O</chem>	-0.433	58.22	0	4	10	207.007	0.188
0407	<chem>Brc1cnc(C2OCCO2)s1</chem>	1.951	31.35	0	4	11	236.09	0.176
0408	<chem>O=C([O-])C(=O)Nc1ccc(Cl)cn1</chem>	-0.577	82.12	1	4	13	199.573	0.042
0459	<chem>Cc1n[nH]c(C)c1I</chem>	1.631	28.68	1	1	8	222.029	0.186
0460	<chem>O=c1[nH]cc(Cl)cc1[N+](=O)[O-]</chem>	0.937	76	1	3	11	174.543	0.006
0461	<chem>Nc1ncc(Br)s1</chem>	1.488	38.91	1	3	7	179.042	0.168



Barcode	SMILES	SlogP	TPSA	Num HBD	Num HBA	HA	MW	V <sub>max</sub>
0462	<chem>Brc1nnc2n1CC[NH2+]C2</chem>	-0.882	47.32	1	3	10	204.051	0.177
0463	<chem>O=C([O-])c1n[nH]cc1Br</chem>	-0.464	68.81	1	3	9	189.976	0.023
0464	<chem>N#Cc1n[nH]cc1Br</chem>	1.044	52.47	1	2	8	171.985	0.18
0465	<chem>Cn1c(Cl)cc(=O)n(C)c1=O</chem>	-0.263	44	0	4	11	174.587	0.154
0466	<chem>O=C([O-])c1ccc(Cl)o1</chem>	0.297	53.27	0	3	9	145.521	0.025
0467	<chem>O=C1Nc2c(Cl)cccc2C1=O</chem>	1.475	46.17	1	2	12	181.578	0.146
0468	<chem>Clc1cnc2nn[n-]c2c1</chem>	0.635	52.77	0	3	10	153.552	0.141
0469	<chem>Brc1cnc2cccnn12</chem>	1.492	30.19	0	3	10	198.023	0.173
0470	<chem>O=C([O-])c1cc(Cl)nnc1Cl</chem>	0.147	65.91	0	4	11	191.981	0.025/0.002
0471	<chem>Nc1c(Cl)ccc(O)c1Cl</chem>	2.281	46.25	2	2	10	178.018	0.137/0.124
0472	<chem>Nc1ncnc(C(=O)[O-])c1Cl</chem>	-0.924	91.93	1	5	11	172.551	0.001
0473	<chem>Fc1cnccc1I</chem>	1.825	12.89	0	1	8	222.988	0.206
0474	<chem>O=c1[nH]c2ccc(Cl)nc2[nH]1</chem>	0.905	61.54	2	2	11	169.571	0.12
0475	<chem>Nc1ncc(Br)cc1CO</chem>	0.919	59.14	2	3	10	203.039	0.152
0476	<chem>Nc1nc(Cl)nc2[nH]cnc12</chem>	0.589	80.48	2	4	11	169.575	0.109
0477	<chem>Clc1cscn1</chem>	1.797	12.89	0	2	6	119.576	0.126
0481	<chem>CC(Cn1cc(Cl)cn1)C(=O)[O-]</chem>	-0.078	57.95	0	4	12	187.606	0.039
0482	<chem>Brc1cnc2[n-]nnc2c1</chem>	0.745	52.77	0	3	10	198.003	0.174
0483	<chem>C[NH+]1CCN(C(C[NH3+])c2cccc2Cl)CC1</chem>	-0.547	35.32	2	1	17	255.793	0.327
0521	<chem>Cc1onc(N)c1I</chem>	1.170	52.05	1	3	8	224.001	0.205
0522	<chem>Cc1nn(C)c(Cl)c1S(N)(=O)=O</chem>	0.029	77.98	1	4	12	209.658	0.147
0523	<chem>O=C([O-])c1cc(Br)c[nH]c1=O</chem>	-0.499	72.99	1	3	11	216.998	0.047
0524	<chem>Cn1c(=O)cc(Cl)[n-]c1=O</chem>	-0.644	53.17	0	3	10	159.552	0.159
0525	<chem>Cc1ccc(Cl)c(C[NH3+])c1F</chem>	1.529	27.64	1	0	11	174.626	0.265
0658	<chem>O=C([O-])c1c(I)ccnc1F</chem>	0.189	53.02	0	3	11	265.989	0.067
0659	<chem>O=C([O-])c1nccc(Cl)c1Cl</chem>	0.752	53.02	0	3	11	190.993	0.027/0.007
0660	<chem>CN(C)S(=O)(=O)n1nccc1Br</chem>	0.300	55.2	0	4	12	254.109	0.171

Barcode	SMILES	SlogP	TPSA	Num HBD	Num HBA	HA	MW	V <sub>max</sub>
0661	<chem>CC(=O)c1c[nH]c2ncnc(Cl)c12</chem>	1.814	58.64	1	3	13	195.609	0.13
1100	<chem>[O-]c1occ(O)c1Br</chem>	0.821	56.43	1	3	8	177.961	0.011
1102	<chem>Cc1nc(N)nc(Cl)c1Cl</chem>	1.674	51.8	1	3	10	178.022	0.124/0.131
1150	<chem>CCn1ncc(Br)c1C(N)=O</chem>	0.764	60.91	1	3	11	218.054	0.174
1152	<chem>Cc1nn(CC[NH3+])cc1Cl</chem>	0.087	45.46	1	2	10	160.628	0.212
1153	<chem>Oc1ccc(Cl)cc1/C=C/C12CN3CN(CN(C3)C1)C2</chem>	1.865	29.95	1	4	20	291.782	0.188
1154	<chem>O=C([O-])c1nn(Cn2cccn2)cc1Cl</chem>	-0.398	75.77	0	6	15	225.615	0.006
1155	<chem>CCn1ncc(Br)c1C[NH2+]C</chem>	0.359	34.43	1	2	11	219.106	0.294
1156	<chem>Nc1nn(CC(=O)[O-])cc1Br</chem>	-1.022	83.97	1	5	11	219.018	0.051
1212	<chem>BrC1CCCC2C1C[NH2+]C2</chem>	1.026	16.61	1	0	10	199.071	0.277
1213	<chem>CS(=O)(=O)c1ccc(Cl)c(S(C)(=O)=O)c1</chem>	1.147	68.28	0	4	15	268.743	0.147
1214	<chem>Cn1nc(C(F)(F)F)c(CO)c1Cl</chem>	1.585	38.05	1	3	13	214.574	0.15
1216	<chem>Cn1c2nc(Cl)nc-2c([O-])n(C)c1=O</chem>	-0.654	75.77	0	6	14	213.604	0.151
1217	<chem>Nc1c(Cl)ccc2nonc12</chem>	1.458	64.94	1	4	11	169.571	0.133
1218	<chem>Cc1c(Cl)c(C(=O)[O-])nn1C</chem>	-0.255	57.95	0	4	11	173.579	-0.011
1219	<chem>Nc1ncc(C(F)(F)F)cc1Cl</chem>	2.336	38.91	1	2	12	196.559	0.141
1221	<chem>O=C([O-])c1cc(Cl)c[nH]1</chem>	0.032	55.92	1	2	9	144.537	0.003
1222	<chem>NC(=O)c1cc2c(nc1Cl)CCC2</chem>	1.323	55.98	1	2	13	196.637	0.127
1223	<chem>CS(=O)(=O)c1ncc(Cl)c(C(=O)[O-])n1</chem>	-1.103	100.05	0	6	14	235.628	0.028
1224	<chem>Clc1nccn1</chem>	1.130	25.78	0	2	7	114.535	0.12
1225	<chem>CCOC(Cn1cc(Cl)cn1)OCC</chem>	1.936	36.28	0	4	14	218.684	0.121
1226	<chem>[NH3+]C1CCSc2ccc(Cl)cc21</chem>	2.119	27.64	1	1	12	200.714	0.228
1227	<chem>Cc1ccc(C(N)=O)c(Cl)n1</chem>	1.142	55.98	1	2	11	170.599	0.129
1228	<chem>O=C([O-])c1csc(Cl)n1</chem>	0.160	53.02	0	4	9	162.577	0.027
1229	<chem>Clc1cccc2nnnn12</chem>	0.778	43.08	0	4	10	154.56	0.157
1230	<chem>Nc1c(Cl)ncnc1NC1CC1</chem>	1.287	63.83	2	4	12	184.63	0.12
1232	<chem>Nc1ccc(Br)cn1</chem>	1.426	38.91	1	2	8	173.013	0.151

Barcode	SMILES	SlogP	TPSA	Num HBD	Num HBA	HA	MW	V <sub>max</sub>
1233	<chem>OCc1c(Cl)nc2sccn12</chem>	1.542	37.53	1	4	11	188.639	0.118
1234	<chem>Cn1c(Cl)c(C#N)c(=O)n(C)c1=O</chem>	-0.391	67.79	0	5	13	199.597	0.169
1235	<chem>OCc1cc2c(cc1Cl)OCO2</chem>	1.561	38.69	1	3	12	186.594	0.119
1236	<chem>O=S(=O)(c1ccc(Cl)s1)N1CCNCC1</chem>	0.995	49.41	1	4	15	266.775	0.242
1237	<chem>CC(C)c1nsc(Cl)n1</chem>	2.315	25.78	0	3	9	162.645	0.146
1238	<chem>Nc1ncc(F)cc1I</chem>	1.408	38.91	1	2	9	238.003	0.198
1239	<chem>O=C([O-])c1ccnc(Cl)c1</chem>	0.099	53.02	0	3	10	156.548	0.005
1240	<chem>Clc1cnccn1</chem>	1.130	25.78	0	2	7	114.535	0.13
1241	<chem>Clc1ccc(CN2CC[NH2+][CC2])s1</chem>	0.781	19.85	1	2	13	217.745	0.21
1242	<chem>[NH3+]Cc1cc(Cl)c2c(c1)OCCCO2</chem>	1.243	46.1	1	2	14	214.672	0.219
1243	<chem>Cc1nn(CC(=O)[O-])c(C)c1Cl</chem>	-0.097	57.95	0	4	12	187.606	0.017
1244	<chem>Nc1cc(Cl)cc([N+](=O)[O-])c1[O-]</chem>	0.904	92.22	1	4	12	187.562	0.128
1245	<chem>O=C([O-])c1ccc(Br)cn1</chem>	0.208	53.02	0	3	10	200.999	0.056
1246	<chem>COc1cc(OC)nc(Cl)n1</chem>	1.147	44.24	0	4	11	174.587	0.115
1247	<chem>O=C([O-])C(O)c1ccccc1Cl</chem>	0.123	60.36	1	3	12	185.586	0.007
1248	<chem>O=c1[n-]cc(Br)c(=O)[nH]1</chem>	-0.545	64.03	1	2	9	189.976	0.172
1249	<chem>O=C([O-])Cc1ccc(Cl)nc1</chem>	0.027	53.02	0	3	11	170.575	0.016
1250	<chem>Cc1nn(CCO)c(N)c1I</chem>	0.371	64.07	2	4	11	267.07	0.187
1251	<chem>O=C([O-])c1cnc(Cl)cn1</chem>	-0.507	65.91	0	4	10	157.536	0.022
1252	<chem>Cn1nc(Br)c(C(=O)NCC(=O)[O-])c1N</chem>	-1.756	113.07	2	6	15	276.07	0.05
1253	<chem>O=C1CCS(=O)(=O)c2ccc(Cl)cc21</chem>	1.700	51.21	0	3	14	230.672	0.138
1254	<chem>Clc1ncnc2[nH]ccc12</chem>	1.611	41.57	1	2	10	153.572	0.12
1255	<chem>O=C1Nc2cccc(Cl)c2C1=O</chem>	1.475	46.17	1	2	12	181.578	0.13
1256	<chem>Clc1cncc(-n2cccn2)n1</chem>	1.316	43.6	0	4	12	180.598	0.133
1257	<chem>N#Cc1ccc([O-])c(Cl)c1</chem>	1.285	46.85	0	2	10	152.56	0.149
1258	<chem>O=C([O-])c1cccnc1Cl</chem>	0.099	53.02	0	3	10	156.548	-0.016
1259	<chem>Nc1ccc(C(=O)[O-])c(Cl)c1</chem>	0.286	66.15	1	3	11	170.575	-0.017
1260	<chem>Nc1ccc(Cl)c(C(=O)[O-])c1</chem>	0.286	66.15	1	3	11	170.575	-0.018

Barcode	SMILES	SlogP	TPSA	Num HBD	Num HBA	HA	MW	V <sub>max</sub>
1261	<chem>O=C([O-])c1cncc(Br)c1</chem>	0.208	53.02	0	3	10	200.999	0.049
1262	<chem>Fc1ncccc1I</chem>	1.825	12.89	0	1	8	222.988	0.203
1263	<chem>NS(=O)(=O)c1cc(C(=O)[O-])ccc1Cl</chem>	-0.649	100.29	1	4	14	234.64	0.034
1264	<chem>O=C([O-])c1cc(Br)c[nH]1</chem>	0.141	55.92	1	2	9	188.988	0.035
1265	<chem>Clc1n[nH]c2cccc12</chem>	2.216	28.68	1	1	10	152.584	0.123
1266	<chem>Cn1c2nc(Br)nc-2c([O-])n(C)c1=O</chem>	-0.545	75.77	0	6	14	258.055	0.184
1267	<chem>Clc1cnc(Cl)nc1</chem>	1.783	25.78	0	2	8	148.98	0.149/0.128
1268	<chem>COc1cc(CO)cc(I)c1O</chem>	1.498	49.69	2	3	12	280.061	0.193
1269	<chem>Clc1ccncc1</chem>	1.735	12.89	0	1	7	113.547	0.131
1270	<chem>CCC([NH3+])c1ccc(Cl)cc1</chem>	2.033	27.64	1	0	11	170.663	0.222
1271	<chem>N#Cc1ccnc(Cl)c1</chem>	1.607	36.68	0	2	9	138.557	0.135
1272	<chem>BrC1noc(C2CC2)n1</chem>	1.710	38.92	0	3	9	189.012	0.174
4481	<chem>CNc1cc(Cl)c2c(c1)OCO2</chem>	2.110	30.49	1	3	12	185.61	0.118
4482	<chem>FC(F)(F)c1cnc(Cl)nc1</chem>	2.149	25.78	0	2	11	182.532	0.134
4483	<chem>Cc1n[nH]c(C(=O)[O-])c1Br</chem>	-0.156	68.81	1	3	10	204.003	0.023
4485	<chem>Oc1ccc(Cl)cc1</chem>	2.046	20.23	1	1	8	128.558	0.117
4486	<chem>Clc1cnc2cccn12</chem>	1.988	17.3	0	2	10	152.584	0.287
7394	<chem>COc1cc(Cl)cnc1OC</chem>	1.752	31.35	0	3	11	173.599	0.12
7395	<chem>Nc1ccc(Br)c(C(=O)[O-])n1</chem>	-0.210	79.04	1	4	11	216.014	0.025
7396	<chem>COc1ncc(Br)cc1C(=O)[O-]</chem>	0.216	62.25	0	4	12	231.025	0.046
7397	<chem>Fc1cc(C2CNCC[NH2+])2c(Cl)cc1Cl</chem>	1.340	28.64	2	1	15	250.124	0.236/0.219
7398	<chem>O=C([O-])c1sccc1Cl</chem>	0.765	40.13	0	3	9	161.589	-0.011
7399	<chem>OB(O)c1ccncc1Br</chem>	-0.476	53.35	2	3	10	201.816	0.148
7400	<chem>O=C(c1ccc(Cl)cc1)N1CCOCC1</chem>	1.812	29.54	0	2	15	225.675	0.125
7401	<chem>Oc1cccc(O)c1Br</chem>	1.860	40.46	2	2	9	189.008	0.174
7402	<chem>Cc1cnc(C)c(Cl)n1</chem>	1.747	25.78	0	2	9	142.589	0.121
7403	<chem>O=C([O-])c1cnccc1Br</chem>	0.208	53.02	0	3	10	200.999	0.027
7404	<chem>C[NH+]1CCN(c2ccc(Br)cn2)CC1</chem>	0.179	20.57	1	2	14	257.155	0.238


Barcode	SMILES	SlogP	TPSA	Num HBD	Num HBA	HA	MW	V <sub>max</sub>
7405	<chem>Cc1c(Cl)cnc(N)c1Cl</chem>	2.279	38.91	1	2	10	177.034	0.133/0.125
7406	<chem>Cc1cc(N)c(Br)c[nH+]1</chem>	1.154	40.16	1	1	9	188.048	0.157
7407	<chem>COC(=O)c1cc(N)c(Cl)s1</chem>	1.770	52.32	1	4	11	191.639	0.134
7408	<chem>Brc1ccc(C2C[NH2+]CCN2)s1</chem>	0.718	28.64	2	2	12	248.169	0.258
7409	<chem>Nc1ccc2[nH]cnc2c1Br</chem>	1.908	54.7	2	2	11	212.05	0.141
7410	<chem>OC1CCN(c2cc(Cl)ncn2)CC1</chem>	1.091	49.25	1	4	14	213.668	0.118
7411	<chem>O=C([O-])c1c(Cl)cncc1Cl</chem>	0.752	53.02	0	3	11	190.993	0.005/0.005
7412	<chem>Nc1nc(Cl)c(NC=O)c(Cl)n1</chem>	0.934	80.9	2	4	12	207.02	0.141/0.138
7413	<chem>O=C([O-])c1cc(Br)c(O)cc1O</chem>	0.224	80.59	2	4	12	232.009	0.051
7414	<chem>Clc1nsnc1N1CCOCC1</chem>	1.028	38.25	0	5	12	205.67	0.138
7415	<chem>Cn1cc(Br)c(N)n1</chem>	0.765	43.84	1	3	8	176.017	0.152
7416	<chem>CC(=O)Nc1ccc(Cl)c(C(=O)[O-])n1</chem>	0.057	82.12	1	4	14	213.6	0.003
7417	<chem>O=C([O-])C(O)c1cccc(Cl)c1</chem>	0.123	60.36	1	3	12	185.586	0.018
7418	<chem>O=C([O-])c1ncc(Cl)cc1Cl</chem>	0.752	53.02	0	3	11	190.993	0.032/0.008
7419	<chem>O=C([O-])c1cc(Cl)c[nH]c1=O</chem>	-0.608	72.99	1	3	11	172.547	0.024
7420	<chem>Cc1[nH]c(=O)[nH]c(=O)c1I</chem>	-0.024	65.72	2	2	10	252.011	0.2
7421	<chem>[O-][n+]1ccc(Cl)cc1</chem>	0.973	26.94	0	1	8	129.546	0.135
7422	<chem>Brc1cnc(N2CC[NH2+]CC2)nc1</chem>	-0.378	45.63	1	3	13	244.116	0.251
7423	<chem>COc1ncc(C(=O)[O-])cc1Cl</chem>	0.107	62.25	0	4	12	186.574	0.016
7424	<chem>Nc1cnccc1Br</chem>	1.426	38.91	1	2	8	173.013	0.161
7425	<chem>O=c1ncc(Cl)c[nH]1</chem>	0.423	45.75	1	2	8	130.534	0.145
7426	<chem>O=C([O-])c1ccncc1Br</chem>	0.208	53.02	0	3	10	200.999	0.026
7427	<chem>O=c1cc([O-])c(Cl)c[nH]1</chem>	0.102	55.92	1	2	9	144.537	0.117
7428	<chem>Cc1ccc(C(=O)[O-])c(Cl)n1</chem>	0.407	53.02	0	3	11	170.575	-0.017
7429	<chem>COc1ncc(Cl)cc1C(=O)[O-]</chem>	0.107	62.25	0	4	12	186.574	0.015
7655	<chem>Nc1ncc(C(=O)[O-])cc1Br</chem>	-0.210	79.04	1	4	11	216.014	0.049
7765	<chem>CC(=O)CS(=O)(=O)c1cc(Cl)ccc1F</chem>	1.842	51.21	0	3	15	250.678	0.135
7766	<chem>Cn1ncc(I)c1C(=O)NN</chem>	-0.372	72.94	2	4	11	266.042	0.192

## Appendix D: Permissions

The ACS AuthorChoice agreement can be found under:

[https://pubs.acs.org/page/policy/authorchoice\\_termsfuse.html](https://pubs.acs.org/page/policy/authorchoice_termsfuse.html)

ACS Publications granted permission for Figure 7:

 **Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding**  
Author: Aldino Viegas, João Manso, Franklin L. Nobrega, et al  
Publication: Journal of Chemical Education  
Publisher: American Chemical Society  
Date: Jul 1, 2011  
Copyright © 2011, American Chemical Society

**PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE**

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#) [CLOSE WINDOW](#)

© 2021 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions  
Comments? We would like to hear from you. E-mail us at [customer@copyright.com](mailto:customer@copyright.com)

The CC BY 4.0 license can be found under:

<https://creativecommons.org/licenses/by/4.0/legalcode>

The CC BY 3.0 license can be found under:

<https://creativecommons.org/licenses/by/3.0/legalcode>

