The design, synthesis and evaluation of Nrf2-Keap1 PPI inhibitors: a modular, virtual screening-led approach

Thesis submitted for the degree of Maser of Philosophy (MPhil)
in Medicinal Chemistry

Clément Martin Leopold Bechtler

Primary Supervisor: Dr Geoffrey Wells

Secondary Supervisor: Prof Frank Kozielski

School of Pharmacy

UCL - University College London

Plagiarism statement

This report describes research conducted in the UCL School of Pharmacy under the supervision of Dr Geoffrey Wells. I certify that the research described is original and that any parts of the work that have been conducted by collaboration are clearly indicated. I also certify that I have written all the text herein and have clearly indicated by suitable citation any part of this report that hast already appeared in publication.

Clément Bechtler

08.10.2017

Abstract

Nrf2 (nuclear factor erythroid 2-related factor 2) is a cap'n'collar bZIP transcription factor and is the main activator of the transcription of over 100 genes that play roles in responses to oxidative stress and detoxifying xenobiotics. The main control of Nrf2 levels is exercised by Keap1 (Kelch-like ECH-associated protein 1) which facilitates the ubiquitination of Nrf2 and therefore its degradation. Keap1 is oxidation-sensitive and upon exposure to oxidants, it changes its conformation and binds Nrf2 tightly. Consequently, *de novo*-synthesised Nrf2 can accumulate.

Following its discovery, Nrf2 received most attention in relation to cancer. Over the time, however, its implication in other pathologies have been more and more acknowledged, namely in inflammation and most importantly in neurodegenerative diseases. Especially Parkinson's disease (PD), which is the second most common neurodegenerative disease, caused by the progressive loss of dopaminergic neurons in the *substantia nigra*, has been linked to oxidative stress. The role Nrf2 plays has been demonstrated in animal models of α-synuclein aggregation or chemically induced parkinsonism, where an increase in Nrf2 expression provided neuroprotection and a slowing of disease progression. Therefore, the inhibition of Keap1-mediated Nrf2 degradation presents a promising strategy for the mechanistic study and the therapy of PD.

Several structures showing high potency towards Keap1 inhibition have been described, with activities in the nanomolar range. However, these compounds are large, or hydrophilic and charged. In order to develop new scaffolds, extensive virtual screening assays have been conducted which resulted in hits with promising molecular scaffolds. At the same time, chemical modifications on a known triazole structure have been performed in order to elucidate structure-activity relationships. In this thesis, the molecular modelling lead, as well the synthetic approach to both project components are described. Finally, the results of a competitive fluorescence polarisation (FP) assay for the second set of compounds are presented.

Acknowledgements

First and foremost, I want to thank Dr Geoff Wells for his guidance and support through all the times. His door was always open whenever there was a need for advice or help on every possible aspect. I want to thank the German Academic Exchange Service (DAAD) and Keregen Therapeutics Ltd for financial support which only allowed me to have this very valuable experience, which let me grow scientifically and personally. A thank you to all members of G25 for creating a stimulating work environment, and for interesting discussions. Finally, I want to thank my family and everyone who was by my side and supported me, who made difficult moments easier to deal with, good times even better, and coffee breaks in the park special.

Table of contents

List of abbreviations

ABC:	ATP-binding cassette
ACF:	colonic aberrant crypt foci
AD:	Alzheimer’s disease
AFAR:	aflatoxin B1 aldehyde reductase
AP-1:	activator protein-1
ARE:	antioxidant responsive element
ATF3:	activating transcription factor 3
ATF4:	activating transcription factor 4
Bach1:	BTB and CNC homology 1
BBB:	blood-brain barrier
BRCP:	breast cancer resistance protein
BRG1:	Brahma-related gene 1
βTrCP:	β-transducin repeat containing E3 ubiquitin protein ligase
BTB:	broad complex, tramtrack, bric-a-brac
bZIP:	basic leucine zipper
CAND1:	cullin-associated and neddylation-dissociated 1
CNC:	cap’n’collar
CNS:	central nervous system
CBP:	CREB binding protein
CDDO:	1-[2-cyano-3-,12-dioxooleana-1,9(11)-dien-28-oyl]
CK:	casein kinase
CREB:	cAMP responsive element binding protein
CRM1:	chromosomal maintenance 1
CTR:	C-terminal region
DCC:	N,N'-dicyclohexylcarbodiimide
DGR:	double-glycine repeat
DIPEA:	N,N’-diisopropylethylamine
4-DMAP:	4-dimethylamino pyridine
DMF:	dimethylformamide
ECH:	erythroid cell-derived protein with CNC homology/enoyl-CoA hydratase
EDCI:	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
ERα:	estrogen receptor α
ERK1/2:	extracellular signal-regulated kinases 1/2
FCC:	flash column chromatography

FP:	fluorescence polarisation
γ-GCS:	γ-glutamylcysteine synthetase
GDNF:	glia cell line-derived neurotrophic factor
GI-GPx:	glutathione peroxidase – gastrointestinal
GS:	glutathione synthetase
GSH:	glutathione
GSK3β:	Glycogen synthase kinase 3β
GST:	glutathione S-transferase
HB:	hydrogen bond
HBA:	hydrogen bond acceptor
HBD:	hydrogen bond donator
HO-1:	heme oxygenase 1
IL:	interleukin
ITC:	isothermal titration calorimetry
IVR:	intervening region
LC/MS:	liquid chromatography/mass spectrometry
maf:	musculoaponeurotic fibrosarcoma
MAPK:	mitogen-activated protein kinases
MDR:	multidrug resistance proteins
MRP:	multidrug resistance-associated proteins
MST:	microscale thermophoresis
mEH:	microsomal epoxide hydrolase
ME1:	malic enzyme 1
MPP^+:	1-methyl-4-phenylpyridinium
M_W:	molecular weight
ND:	neurodegenerative diseases
Neh:	Nrf2-ECH-homology
NES:	nuclear export sequence
NF-E2:	nuclear factor erythroid 2
NF- κB:	nuclear factor ‘kappa-light-chain-enhancer’ of activated B-cells
NLS:	nuclear localisation signal
NMR:	nuclear magnetic resonance
NQO1:	NAD(P)H:quinone oxidoreductase 1
Nrf1:	nuclear factor erythroid 2-related factor 1
Nrf2:	nuclear factor erythroid 2-related factor 2
Nrf3:	nuclear factor erythroid 2-related factor 3

NTR:	N-terminal region
PD:	Parkinson’s disease
PDB:	protein data bank
PEKR:	PKR-like endoplasmic reticulum-resident kinase
Pgp:	permeability glycoprotein
PI3K:	phosphoinositide 3-kinase
PMF1:	polyamine-modulated factor 1
PPARγ:	peroxisome proliferator-activated receptor γ
PPI:	protein-protein interaction
qPCR:	quantitative polymerase chain reaction
PyBroP:	Bromotripyrrolidinophosphonium hexafluorophosphate
PyBOP:	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
RARα:	retinoic acid receptor α
RMSD:	root mean square deviation
ROS:	reactive oxygen species
rt:	room temperature
SCF complex:	Skp, Cullin, F-box containing complex
SFERRβ:	short form estrogen-related receptor β
SOD:	superoxide dismutase
SMD:	steered molecular dynamics
SNpc:	substantia nigra pars compacta
SPR:	surface plasmon resonance
TAT:	trans-activating transcriptional activator
TBAB:	tetrabutylammonium bromide
TGF:	transforming growth factor
TLC:	thin layer chromatography
t_R:	retention time
βTrCP:	beta-transducin repeat containing protein
UDP:	uridine diphosphate
UGT:	UDP-glucuronosyltransferase
VS:	virtual screening
vdW:	van der Waals

1 Introduction

1.1 Nrf2

1.1.1 Structure

Human Nrf2 (nuclear factor erythroid 2-related factor 2) is a 597-amino acid protein and was described initially as a transcription factor that binds to the hypersensitive site 2 located in the β-globin locus control region on chromosome 11. Subsequently it has been shown that Nrf2 is found ubiquitously both in cells and across species.[1,2,11,17] The murine form of the Nrf2 protein is highly homologous, but 8 amino acids longer in sequence.[17] The name Nrf2 is derived from its similarity with NF-E2 (nuclear factor erythroid 2) and it belongs, like NF-E2, to the structural class of cap'n'collar (CNC) basic leucine zipper protein (bZIP) transcription factors. The C-terminus, containing the bZIP domain, is highly homologous to NF-E2 and Nrf1 (nuclear factor erythroid 2-related factor 1) whilst the N-terminus is distinct.[1] This CNC domain is highly conserved between the family members.[11] In total, 6 CNC family members have been described: NF-E2, Nrf1, Nrf2, Nrf3 (nuclear factor erythroid 2-related factor 3), Bach1 (BTB and CNC homology 1) and Bach2 (BTB and CNC homology 2).[15,16]

Nrf2 has been divided into 6 domains, labelled Neh1 – Neh6 (Nrf2-ECH homology).[5] Nrf2's CNC domain mainly constitutes Neh1 which is essential for DNA binding and heterodimerization with either one of three small Maf (musculoaponeurotic fibrosarcoma) molecules (MafF, MafG, MafK), Jun molecules (c-Jun, Jun-D, Jun-B), PMF1 (polyamine-modulated factor 1) or ATF4 (activating transcription factor 4).[5,10,12,14] Additionally, the NLS (nuclear localisation signal) RKRK can be found in the basic part of the bZIP (residues 515 – 518) which is a short nuclear import sequence, and has been labelled as NLS2. Furthermore, the domain contains another NLS, namely NLS3, reaching from residue 587 to 593.[15,16,17,19]

The *N*-terminal Neh2 region is the major negative regulation site of Nrf2 under homeostatic conditions and forms the binding region for Keap1 (Kelch-like ECH-associated protein 1), which is the most prominent negative regulator, but also contains Keap1-independent degradation motifs. Additionally, the subdomain contains seven lysine residues that can be ubiquitinated and function as degradation markers.[5,12,15] The important motifs for Keap1-mediated degradation are the ^{79}ETGE^{82} and ^{29}DLG^{31} sequences of Nrf2 (numbering for murine Nrf2) and these are found in the hydrophilic subdomain.[5,17] Together with its adjacent amino acids, the DLG motif is also called a DIDLID element, comprising amino acids 17-32.[19] Furthermore, the DIDLID element is redox-sensitive on its own and can promote proteasomal degradation independently of the interaction with Keap1.[12] Another important functional part of Neh2 are amino acids 42-53 which comprise another NLS, namely NLS1.[19]

Although the exact function of Neh3 is not as well documented as for Neh1 and Neh2, it appears to be important as a transactivation domain since deleting the domain results in a protein that has no transcriptional activity, despite being localised in the nucleus.[9]

Neh4 and Neh5 are transcriptional activation domains which recruit and bind cooperatively to the coactivators CBP (CREB binding protein) and BRG1 (Brahma-related gene 1). Both domains are indispensable for full activity, but Neh5 has a higher affinity.[9,11] CBP is known to activate gene expression in two ways:

- It recruits HAT (histone acetyltransferase), but can also acetylate histone and non-histone nuclear proteins itself. Through this, the structure of the chromatin is altered and consequently, gene expression as well.
- It bridges transcription factors to the transcriptional machinery and enables RNA polymerase II complex assembly.

Other CNC family members lack the Neh4 domain (*e.g.* Nrf1) or even Neh4 and Neh5 (*e.g.* Nrf3).[11] This might partially explain the stronger transcriptional activity of Nrf2 compared to these factors.

Neh6 comprises amino acids 329 – 379 of the protein. It destabilises Nrf2 under basal conditions as well as under oxidative stress and therefore acts as an oxidative stress-insensitive degron. It is a relatively serine-rich region, and these serine residues are essential for regulation through the GSK3-SCF/βTrCP pathway.[9,12,19,20] It is thought that this process takes place in the nucleus, in contrast to the Neh2-mediated degradation that takes place in the cytosol. The suggested reason behind the nuclear location is that the proteins required for degradation mediated by Neh6 are located exclusively in the nucleus.[12] Figure 1 gives an overview of the localisation of the different domains on Nrf2.

Interestingly, Nrf2 contains not only several NLS, which have been labelled NLS1 – NLS3, but also two NES (nuclear export sequence).[15,16,17,19] One of the NES is located in the leucine zipper region of the bZIP domain, being formed by residues 545-554, and is redox-insensitive, the other is located in the Neh5 domain, from amino acid 175-186.[16,17,19]

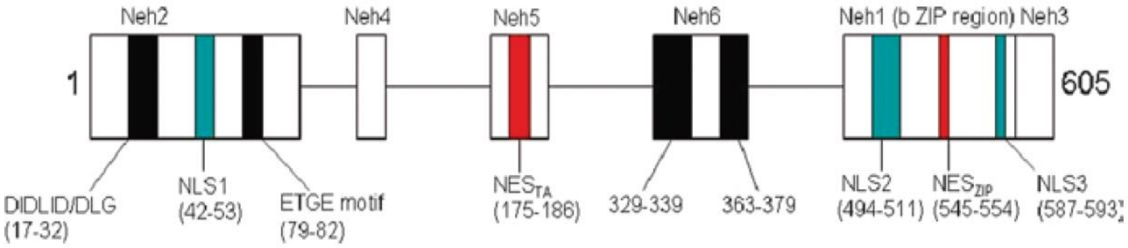


Figure 1: Domain structure of the Nrf2 protein, from reference 19 (A. Giudice, C. Arra, M. C. Turco; Review of Molecular Mechanisms Involved in the Activation of the Nrf2-ARE Signalling Pathway by Chemopreventive Agents; *Transcription Factors: Methods and Protocols, Methods in Molecular Biology* **647** 37-74 (2010)).

1.1.2 Function

Nrf2 is the major transcription factor responsible for rapidly inducing the expression of more than 100 cytoprotective genes under oxidative, chemical (xenobiotics, drugs, smoke, metals) or radiation-mediated stress. These stressors lead to the formation of reactive oxygen species (ROS) such as hydrogen peroxide, hydroxyl radicals and superoxide anions.[14,17,18] However, it should be noted that although high levels of ROS are harmful and cause damage to macromolecules and lipids, lower, transient levels of ROS are needed for cellular signalling and for defence mechanisms against microbes. The essential ROS are formed endogenously, either by NADP oxidases or as a by-product of oxidative phosphorylation.[18] Recently, Nrf2 has been shown to influence mitochondrial activity under homeostatic as well as stressed situations affecting fatty acid oxidation, mitochondrial membrane potential, ATP synthase subunit α expression and consequently ATP synthesis. Under oxidative stress, it upregulates uncoupling protein 3 and influences nuclear respiratory factor 1 as well as peroxisome proliferator-activated receptor γ (PPARγ) coactivator α. Additionally, it might be involved in an increased efficiency in the electron transport chain, leading to a higher fraction of oxygen being completely reduced instead of being partially reduced.[23]

Nrf2 is expressed ubiquitously and activated very rapidly; the time period between the insult and its nuclear import is less than 15 min.[14] Its transcriptional activity is at least 10 times more effective at the same concentrations than its family members Nrf1 and Nrf3 (nuclear factor erythroid 2-related factor 3), measured in a HO-1 (heme oxygenase 1) enhancer-reporter fusion gene assay.[11,73] The plethora of regulated proteins include: phase I- and phase II-enzymes, transporters, antioxidant proteins, proteins taking part in the inflammatory response, chaperones, proteasomes, and proteins with antiapoptotic activity and regulating cell growth such as growth factors, growth factor receptors, and transcription factors.[10,11,14-17,19] For a short description of the function of some of the transcription products, see Table 1.

The general importance of Nrf2 is demonstrated by the facts that Nrf2-knockout mice are, despite being viable, more susceptible to oxidative stress and carcinogenesis, caused for example by benzo[*a*]pyrene exposure.[10,11] It has been speculated that up to 10% of human genes may be controlled by the Nrf2/Keap1/ARE (antioxidant responsive element) system.[29]

The effects of Nrf2 are mediated through binding to DNA at ARE (antioxidant responsive element) sequences. These are *cis*-elements in the promoter regions of various genes, with the core sequence 5'-GTGACNNNGC-3'.[5,9,14,16] Nrf2 binds to the sites as a heterodimer with another bZIP protein, either a small Maf protein (MafF, MafG, MafK), a Jun molecule (c-Jun, Jun-D, Jun-B), PMF1 or ATF4.[5,7,10,16,19]

Globally, the activity of Nrf2 enables cells to reduce thiols, and detoxify ROS and a variety of harmful chemicals, including α,β-unsaturated carbonyl compounds, peroxides, quinones and epoxides.[17]

Abbreviation	Full name	Function
NQO1	NAD(P)H:quinone oxidoreductase 1	2e$^-$-reduction of endogenous and exogenous molecules, including antioxidants
γ-GCS	γ-Glutamylcysteine synthetase	Synthesis of γ-glutamylcysteine, first intermediate/building block for glutathione
GS	Glutathione synthetase	Synthesis of glutathione, a tripeptide crucial for protection of damage caused by ROS, and detoxification
GST	Glutathione S-transferase	Detoxification of chemicals by transferring GSH
GI-GPx	Gastrointestinal isoform of glutathione peroxidase	Protection against inflammation by removing hydroperoxides
UGT	UDP-glucuronosyltransferase	Transfer of glucoronate to chemicals for hydrophilisation and enhanced excretion
HO-1	Heme oxygenase 1	Catalysis of first step of heme catabolism
SOD	Superoxide dismutase	Catalysis of superoxide disproportionation
	Ferritin	Binds free iron to prevent it from mediating radical formation
mEH	Microsomal epoxide hydrolase	Intracellular deactivation of epoxides
MRP and MDR	Multidrug resistance proteins and multidrug resistance-associated proteins	Facilitation of chemicals’ export out of cells
	Peroxiredoxin	Reduction of hydrogen peroxide
	Thioredoxin	Reduction of oxidised protein thiols, including peroxiredoxin
	Thioredoxin reductase	Reduction of oxidised thioredoxin
	Sulfiredoxin	Reduction of sulfinic acids
ME1	Malic enzyme 1	Decarboxylation of maleic acid for NADP generation
AFAR	Aflatoxin B1 aldehyde reductase	Reduction of ketones and aldehydes
	leukotriene B4 dehydrogenase	Reductive deactivation of proinflammatory cytokine leukotriene B4
Nrf2	nuclear factor erythroid 2-related factor 2	Transcription factor upregulating expression of antioxidative proteins
MafF, MafG	musculoaponeurotic fibrosarcoma F and G, respectively	Heterodimerization factors for Nrf2
TGFα, TGFβ	transforming growth factor α and β, respectively	Induction of cell proliferation, development, immune modulation and healing
	TGFβ receptor II	Receptor for TGFβ

Table 1: Selection of transcription products positively regulated by Nrf2

1.1.3 Regulation

Under basal conditions, Nrf2 is expressed, but degraded rapidly through its negative regulators, leading to a half-life of 10 to 30 min which can increase to 40 min under oxidative stress.[9,12,16,19] In addition, since it has been identified in mice that the Nrf2 gene has an ARE in its promotor region, it is thought that Nrf2 activates its own expression.[16] When a cell is exposed to oxidative stress, the structure and consequently the activity of Nrf2 and its regulators are altered which leads to cellular accumulation of Nrf2. The most important regulators will be presented in the following paragraphs.

1.1.3.1 Keap1

The importance of the Nrf2-Keap1-complex is illustrated by the fact that it is conserved in different chordates, namely humans, rodents and zebrafish. Furthermore, the importance of Keap1 in the regulation of Nrf2 was demonstrated by measuring the half-life of Nrf2 in Keap1 knockout cells which increased from 0.6 h to 2.5 h. The human Keap1 (Kelch-like ECH-associated protein1) protein is a 627 (murine: 624) amino acid cytoplasmic, homodimeric metalloprotein that is the main regulator of Nrf2 and acts as a ubiquitin ligase adaptor protein. Its name is derived from its structural similarity to a *Drosophila* protein, Kelch, and its first description, which found that it binds to ECH, the chicken homologue of Nrf2. It has five domains:

- the NTR (*N*-terminal region) comprising amino acids 1-60,
- the BTB (broad complex, tramtrack, bric-a-brac) domain which recruits Cul3/Rbx1, a ubiquitin ligase,
- the cysteine-rich IVR (intervening region) from residue 180-314,
- the Kelch motif, which is also called DGR (double-glycine repeat) domain from residue 315-598, which interacts with Nrf2's Neh2 domain,
- the CTR (C-terminal region) from amino acid 599-624.[5,10,12,14,15,17,19,29]

Figure 2 illustrates the localisation of the different domains on Keap1.

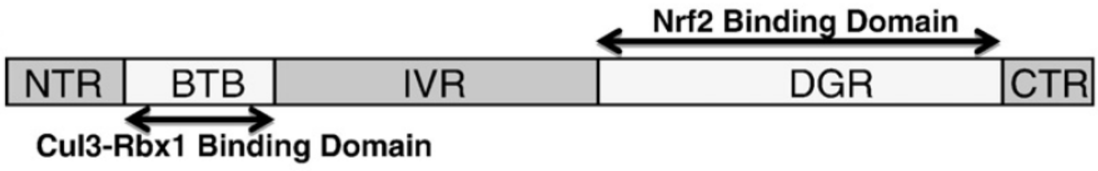


Figure 2: Domain structure of the Keap1 protein, from reference 14 (J. W. Kaspar, S. K. Niture, A. K. Jaiswal; Nrf2:INrf2 (Keap1) signalling in oxidative stress; *Free Radical Biology & Medicine* **47** 1304-1309 (2009)).

The dimer has been proposed to have a "cherry bob"-like structure: the two Kelch and CTR-domains correspond to the 'cherries' while the other domains form the 'stem' through which the dimerization is accomplished.[18]

Ubiquitin ligase complexes are generally formed out of three proteins:

- E1 which activates ubiquitin,
- E2, the ubiquitin-conjugating enzyme which transfers the small protein ubiquitin to a substrate protein,
- E3, the ubiquitin-protein ligase which recognizes the substrate and assists in the transfer of ubiquitin.[16,19]

Within the Cul3/Rbx1 E3 ligase, the cullin Cul3 acts as a scaffold by recruiting the RING finger protein Rbx1, an E2 enzyme and a substrate adaptor such as Keap1 to form a multi-subunit E3 ligase complex.[14,19] Binding of Keap1 to Nrf2 facilitates the ubiquitination and consequently the degradation of Nrf2 through the 26S proteasome.[10,14]

The Kelch domain of Keap1 consists of a six bladed β-propeller (so called Kelch repeats) with each blade formed of four antiparallel β-sheets. The Kelch domain binds to two Nrf2 motifs with different affinities.[15,17,29] A high affinity ($K_D \sim 5.6$ nM) interaction is formed with the Nrf2 ETGE sequence which can be found in a β-hairpin formed by residues ^{75}QLDEETGEFL^{84}. Keap1 interacts with this through a number of its charged and polar amino acids in the Kelch domain, namely Arg380, Arg415, Arg483, His436, Tyr334, Ser363, Asn382, Ser508, Tyr525, Ser555, and Tyr572.[17,30] They are not equally distributed over the domain, but can be found either in the loop between β-strand 2 and β-strand 3 (sometimes labelled as B and C) in Kelch-repeats 1, 2, 3 and 4 or in the loop that precedes β-strand 1 in Kelch repeats 2, 3, 5 and 6.[17] The second binding site of Nrf2, the DLG motif, binds to Keap1 with a lower affinity ($K_D \sim 1$ µM) although the interaction pattern is similar.[17,30] The complete interacting motif is ^{24}WRQDIDLG^{31}, illustrating the similarity to the ETGE sequence, as Gln26, Asp27, Asp29 correspond Glu79, Thr80, Glu82 in ETGE (Figure 3).[17]

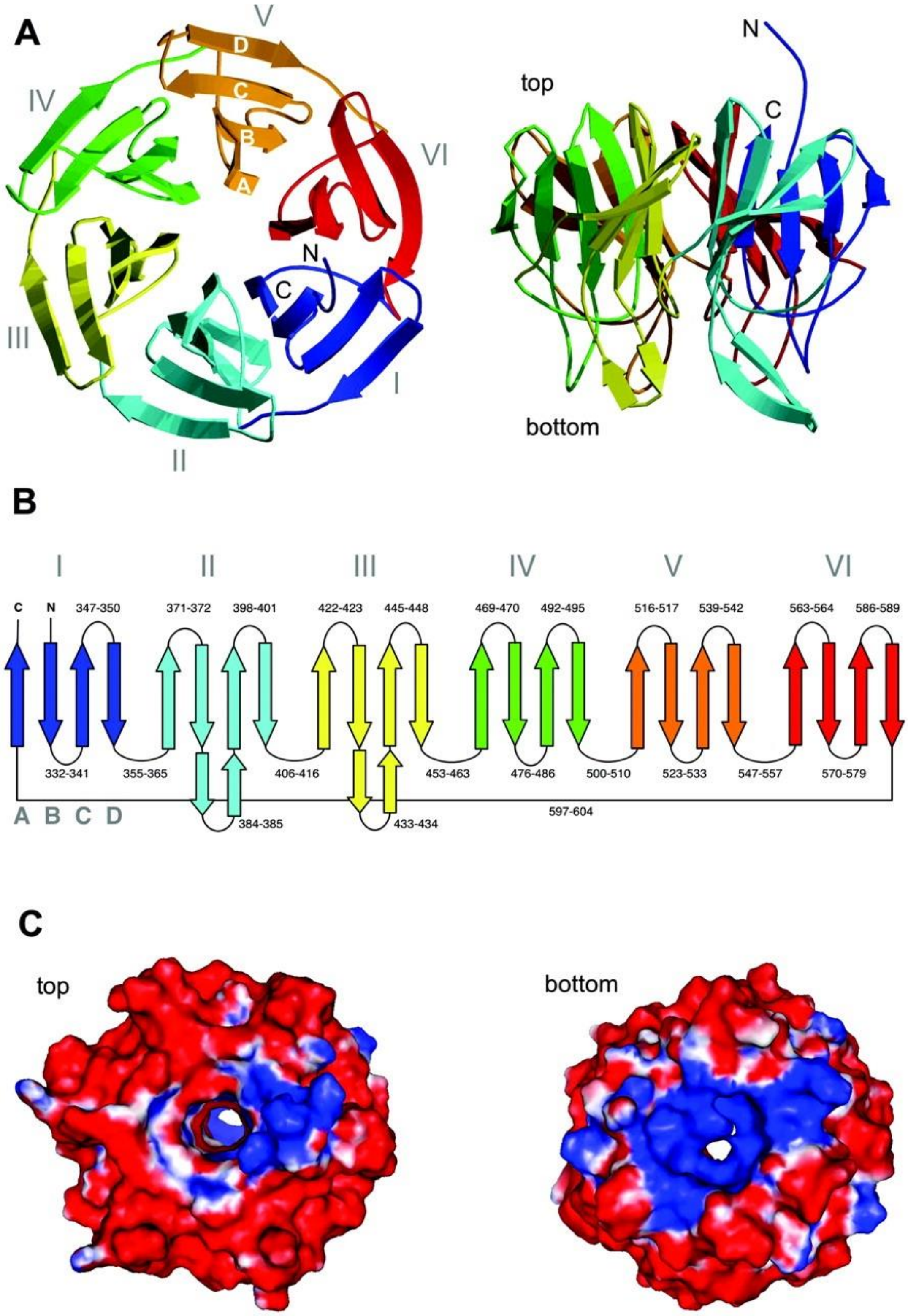


Figure 3: Structure of Keap1 Kelch domain. The different β-propellers are labelled as I – VI and the respective blades are labelled as A – D. From reference 28. **A**. Ribbon Diagram **B**. Topology Map **C**. Polar surface area. Red denotes a negative, blue a positive charge. Modified from ref. 28 (X. Li, D. Zhangs, M. Hannink, L. J. Beamer; Crystal Structure of the Kelch Domain of Human Keap1; *The Journal of Biological Chemistry* **279(50)** 54750-54758 (2004)).

In total, murine Keap1 contains 25 cysteines, which are conserved in rats and humans, with the human sequence containing two additional cysteines.[16,17] Around half of these 25 – 27 cysteines are expected to be reactive, the reactivity being based on vicinity to basic (positively charged) amino acids that favour deprotonation of the thiol moiety which is then more susceptible towards oxidation or alkylation.[17,29]

Under basal conditions, the Keap1/Cul3/Rbx1 complex is active and rapidly ubiquitinates Nrf2.[14] During oxidative stress, Cys151 of Keap1, which is located in the BTB domain, is oxidised, thereupon changing its conformation and consequently, reducing the rate of ubiquitination. Thus, Nrf2 accumulates in the cell.[10] Cys273 and Cys288 which are located in the IVR and Cys23 were shown to play an important role in the functioning of Keap1, the latter two being required for ubiquitination. Other reports have identified Cys257 and Cys297 residues being reactive in some cases.[14,16,17,19]

Several models have been published trying to explain the Keap1-mediated repression of Nrf2.[16] These will be described briefly here.

Firstly, a cytoplasmic retention model for Keap1 and Nrf2 has been proposed. As Keap1 can bind cytoskeletal actin or mitochondrial PGAM5, and Nrf2 binds Keap1, Nrf2 is also retained in the cytoplasm. Under oxidative stress, Keap1 changes its conformation and releases Nrf2 which can be imported into the nucleus.[17]

Secondly, it has been proposed that Keap1 shuttles between the nucleus and the cytoplasm and, under basal conditions, prevents Nrf2 accessing the ARE by retention in a sub-nuclear compartment. Under oxidative stress, prothymosin-α binds to the Kelch repeats and thus liberates Nrf2.[17]

Thirdly, a model based on ‘protein stabilisation’ suggests that under basal conditions Nrf2 is rapidly degraded by the 26S proteasome as a consequence of the Keap1-based E3 ubiquitin ligase activity. As the ETGE motif has a much higher affinity than the DLG motif for Keap1, it has been proposed that it binds first and thus facilitates the binding of the DLG motif to the adjacent Keap1 unit, which then fixes the Nrf2 across the Keap1 dimer. Therefore, this model has also been labelled the ‘hinge and latch’-model and the process has been called ‘two-site tethering’ (Figure 4). Through this fixation, Nrf2 is positioned correctly for efficient ubiquitination. It has been speculated that the oxidation of Cys151 leads to a conformational change that lowers the affinity of the DLG motif to Keap1 and therefore ubiquitination is hampered. In any case, it was shown that Cys151, Cys273 and Cys288 are essential for Keap1’s redox-sensing property. As no increased release of Nrf2 from the complex could be observed after treatment with electrophiles, it has been concluded that the complex stays tightly bound, which decreases the availability of free Keap1 to bind newly synthesised Nrf2.

Other propositions for protein stabilisation include the redox-dependent dissociation of Keap1 and Cul3 with results from immunoprecipitation supporting this hypothesis, as well as the redox-dependent proteasomal degradation of Keap1.[17,19,20] Mutation studies have shown that both the ETGE and DLG motifs are essential for Keap1-mediated Nrf2 degradation and have established the two-site model as the most promising.[17] Although the exact oxidation products of Keap1's reactive cysteines are unknown, it has been reported for other proteins that the sulfide anion can be serially oxidised from sulfenic to sulfinic then sulfonic acids. The oxidation to the sulfonic acid is considered to be irreversible.[18]

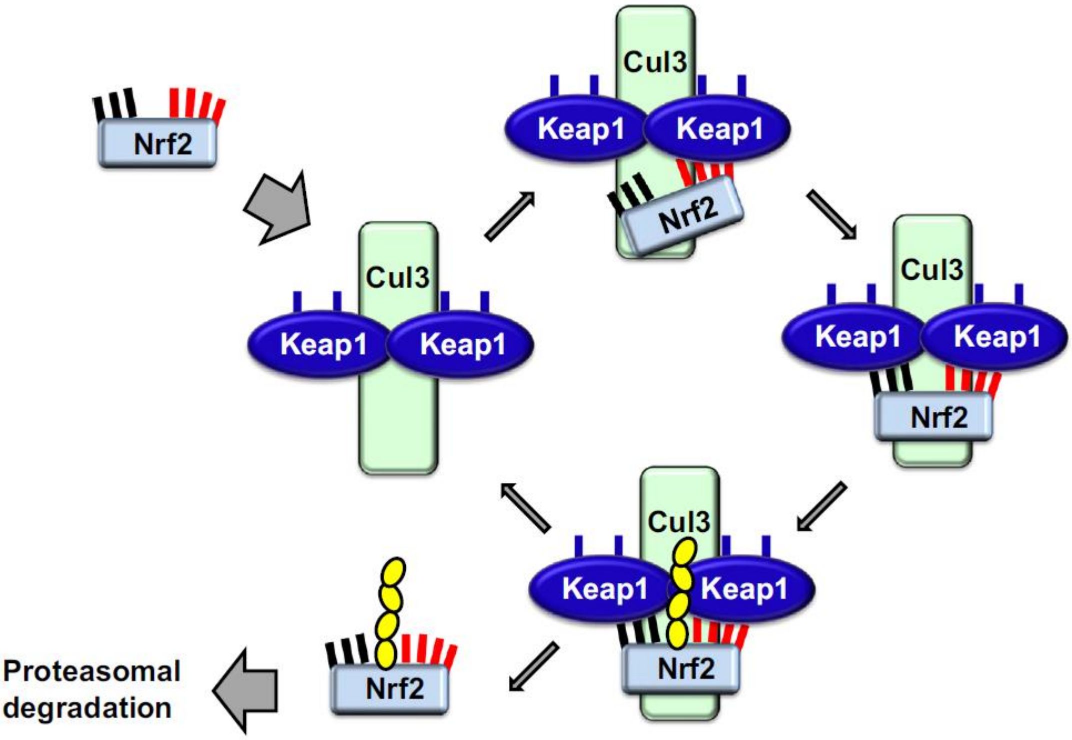


Figure 4: Illustration of the 'hinge and latch' model of Nrf2-Keap1 interaction. First, the high affinity ETGE motif of Nrf2 binds to Keap1, which allows consequently the binding of the low affinity DLG motif. Thus, the position of Nrf2 is fixed which facilitates its ubiquitination. From reference 23 (A. T. Dinkova-Kostova, A. Y. Abramov; The emerging role of Nrf2 in mitochondrial function; *Free Radical Biology and Medicine* **88** 179-188 (2015)).

Fourthly, it has been suggested that the NLS and NES motifs are redox-sensitive, hence the nuclear-cytoplasmic shuttling, the nuclear concentration and the activation of ARE-dependent genes are tuned by the level of oxidisers present. As the NES in Neh5 has been identified as ^{175}LLSIPELQCLNI^{186}, it has been proposed that the cysteine in the sequence is redox-sensitive.[17,19] However, it should be noted that the other NES, being ^{545}LKRRLSTLYL^{554} (murine) or ^{537}LKKQLSTLYL^{546} (human) and located in the Neh2 domain, is close to Tyr568 which is phosphorylated by Fyn under basal conditions. This phosphorylation probably leads to a conformational change exposing the leucine-rich NES, allowing it to interact with the exportin

CRM1 and preventing the nuclear accumulation of Nrf2. Fyn itself is a downstream target for inactivating phosphorylation of the central signalling pathways PI3K-Akt and PKC. PI3K activates Akt which then inhibits GSK3β, and GSK3β activates Fyn. Furthermore, GSK3β inhibits Keap1 directly. PI3K is activated through oxidative stress. However, GSK3β activity is also reduced directly by oxidative stress. [17,19] PKC interacts with this signalling pathway in two ways: first, it phosphorylates Nrf2 which increases the expression of Nrf2-dependent genes, probably through an increased release from Keap1; second, PKC can, like Akt, phosphorylate GSK3β and consequently inactivate it (Figure 5).[19]

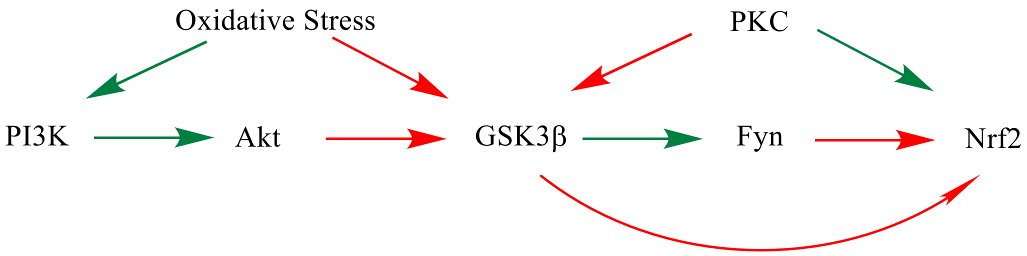


Figure 5: Influence on Nrf2 downstream of central PI3K-Akt and PKC pathways. Red arrow: inhibition, green arrow: activation

In addition to Nrf2, Keap1 has a number of binding partners. These include:

- PGAM5, a phosphoglycerate mutase,
- prothymosin α
- foetal Alz-50 clone 1[17]
- the autophagy-associated protein p62 which leads to the formation of Keap1-aggregates which cannot interact with Nrf2[10]
- PALB2 which has an ETGE motif and can use it to bind to Keap1, therefore preventing it from interacting with Nrf2.

In contrast, CK2 (Casein kinase 2) phosphorylates Thr55 of Keap1 which consequently recruits Hsp90 and stabilises Keap1 through this interaction.[10,17]

Nrf2 activity influences Keap1 in two ways: as the Keap1 gene has an ARE in its promotor region, the transcription is increased upon Nrf2 induction.[14] On the other hand, Keap1 is a ubiquitination target of Cul3 and therefore, its degradation (proteasome-independent) increases with higher Nrf2 concentrations.[14,19]

1.1.3.2 βTrCP

It has been shown that the F-box protein βTrCP degrades Nrf2 that has been phosphorylated by GSK3β.[10,15] It is, like Keap1, a substrate adaptor protein for the E3 ubiquitin ligase SCF. The SCF complex is made up of four parts: an F-box protein (such as βTrCP) that recognises the substrate and binds to the linker protein Skp1. Additionally, it interacts with Cul1 and Rbx1. Both isoforms of βTrCP, 1 and 2, recognise Nrf2. The known general βTrCP recognition motif $K(X)_nDSG(X)_{1\text{-}4}S$ is located between amino acids 322 and 338 and has the sequence DSGIS. In total, Nrf2 has a cluster of 6 serines close to this sequence: Ser335, Ser338, Ser342, Ser347, Ser351, Ser355 (numbering for murine Nrf2), where at least one is indispensable and phosphorylated by GSK3β for βTrCP-mediated degradation. GSK3β is, as previously mentioned, negatively regulated by the PI3K/Akt-pathway, which is upregulated by oxidative stress and growth factors.[20]

1.1.3.3 Other negative regulators

There are a number of other regulatory processes that have been described recently and may impact on Nrf2 activity and ARE gene expression:

- Recently, a third E3 ubiquitin ligase that is involved in Nrf2 regulation has been described: Hrd1, which is activated during endoplasmic reticulum stress.[23]
- Furthermore, some, so far unidentified, tyrosine kinases sense oxidative stress whereupon they phosphorylate Keap1, Fyn, and Bach1 (BTB and CNC homology 1) which are all negative regulators of Nrf2.[10,14]
- The small Maf proteins can form homodimers that are able to bind to the ARE without activating it.[19]
- c-Fos is a member of the AP-1 (AP-1: activator protein-1) which can form heterodimers with Nrf2 and negatively regulate the expression of ARE-controlled genes.[16,17]
- Similar properties have been reported for ATF3 (activating transcription factor 3), ERα (estrogen receptor α), PPARγ (peroxisome proliferator-activated receptor γ) and RARα (retinoic acid receptor α).[17,19]

1.1.3.4 Positive regulators

PKCδ is known to phosphorylate Ser40 of Nrf2 which enables its release from Keap1, conditionally on oxidative modification of Keap1 Cys151.[10] Intriguingly, it has also been reported that this release did not lead to an increased nuclear import of Nrf2 or gene expression, but mutation studies showed that Ser40 is important for nuclear import.[16,19] Also, it was demonstrated that induction of oxidative stress stimulates PKC.[19]

p21 is one of the few examples of a stabiliser of Nrf2. It appears to exert its function by directly interacting through its KKR motif with the ETGE and DLG motifs of Nrf2.[10,18]

Factors that increase basal and induced expression of Nrf2 are K-Ras, B-Raf and Myc.[10] Other activating kinases include PI3K (indirectly), ERK2, ERK5, and PERK.[14,16,17] p38 MAPK phosphorylates Nrf2, but the consequences of it are unclear: some report an increased interaction with Keap1, others a release from it.[14,19]

CK2 phosphorylates the Neh4 and Neh5 domains which seems to be relevant to transactivation.[19]

1.1.4 Pathological Importance

Oxidative stress and Nrf2 play an important role in a variety of conditions which include neurodegenerative diseases (Alzheimer's, Parkinson's, multiple sclerosis), cancer, inflammation, diabetes, pulmonary fibrosis, asthma, emphysema, ischemia, lupus-like autoimmune nephritis, cardiovascular diseases and macular degeneration.[14,15,19] Hence, Nrf2 is pivotal to these indications and its contribution to some of these pathologies will be elucidated here.

1.1.4.1 Neurodegenerative Diseases

The general and increasing importance of neurodegenerative diseases (ND) is widely recognised as they are expected to outstrip cancer as the major cause of death around the year 2050. Under the term ND, several pathologies are summarised, the most frequent ones being Alzheimer's disease (AD) and Parkinson's disease (PD). All NDs are characterised by a progressive neuronal cell loss in the central nervous system (CNS). This cell loss is more severe than in other tissues as neurones cannot replicate themselves. Although the exact mechanisms are not completely understood and differ depending on the disease, it was established that excitotoxicity, oxidative stress, inflammation, impaired neurotrophic support and apoptosis generally play a pivotal role in which oxidative stress appears to be the underlying mechanism of insult.[22,24]

On a histological level, PD is characterised by the loss of dopaminergic neurons in the *pars compacta* of the *substantia nigra* (SNpc), a part of the basal ganglia and mesencephalon playing a crucial role in movement. Consequently, the motor function of the basal ganglia is compromised and leads to the cardinal symptoms of hypokinesia, bradykinesia, rigidity and resting tremor. Additionally, non-motor functions are altered too, resulting in reduced cognitive ability as well as abnormal mood and sleep.[22]

Although the pathological role of α-synuclein is not well understood to date, its aggregates are found in so-called Lewy bodies in the brains of PD patients and it is widely suggested that the three known point mutations in α-synuclein genes or the overexpression of wild type α-synuclein leads to a toxic gain of function which is linked to neuroinflammation and oxidative stress. One possible connection between Nrf2 and α-synuclein aggregation is that reduced levels of Nrf2 impair α-synuclein proteasomal degradation and can therefore lead to an accumulation of proteins that would have been degraded under normal conditions.[27] In PD mouse models, where the animals have been transfected with adeno-associated viral vectors containing the human α-synuclein gene, Nrf2 knock out-mice experienced a 23% increased loss of dopaminergic neurons in the *substantia nigra*, determined by immunohistochemical staining of tyrosine hydroxylase, a marker for dopaminergic neurons. Astrocytes and microglia have been reported to be activated to a greater extent under Nrf2 knock-out conditions and proinflammatory NF-κB-dependent genes *e.g.* IL-6, are transcribed with a faster kinetics than Nrf2-dependent genes, but both pathways are activated by α-synuclein.[27]

The SNpc is predestined to be vulnerable to oxidative stress: as part of the brain, the levels of oxygen, and especially in the SNpc the levels of iron, are high, while the levels of GSH are low which tends to favour higher levels of ROS. Additionally, dopamine metabolism generates ROS.[22,24] It was established that inflammatory markers are increased in the SNpc and the striatum, another part of the basal ganglia, in PD patients. The importance that Nrf2 might play in this context was demonstrated by knocking out the Nrf2 gene in animals which lead to a considerable loss of dopaminergic neurons as well as by the fact that a reduction in GSH levels is one of the first biochemical changes observed in PD.[22,24] On the other hand, the genetic or pharmacological activation of Nrf2 was protective in animals when exposed to the mitochondrial respiratory chain inhibitor and neurotoxin 1-methyl-4-phenylpyridinium (MPP^+) or to 6-hydroxydopamine, which are used to induce parkinsonism in animal models. The neuroprotection by pharmacological activation of Nrf2 was demonstrated to be related to an inhibition of ERK1/2, whose activation has been linked to neuronal death.[23,24,25]

Furthermore, oxidative stress has been linked to multiple sclerosis which is supported by the fact that dimethyl fumarate, a licensed treatment for this disease, targets Nrf2.[24] Similarly, ethyl pyruvate, another activator of Nrf2, was found to have protective effects on astrocytes, which positively affected the survival of neuronal cell cultures exposed to excitotoxic stress. This effect was mediated by GDNF (glia cell line-derived neurotrophic factor) and GSH.[26]

1.1.4.2 Cancer

In 1951, long before the description of Nrf2 at the end of the 1990s, Richardson *et al.* found out that rats that were fed small amounts of aryl hydrocarbons were less susceptible to developing cancer upon administration of carcinogens, a process called pre-conditioning. In the following period, the principle of chemoprevention was developed: the administration of substances to reduce the probability of cancer development.[15] After the discovery of the Nrf2 pathway, the major therapeutic focus of the drug development targeting Nrf2 and Keap1 was chemoprevention. Additionally, a multitude of studies were able to demonstrate that there is an increased risk of a range of different cancers when ARE-controlled genes mutate.[16]

On one hand, a range of studies demonstrated the crucial role that Nrf2 plays in chemoprevention since Nrf2-knock out mice are prone to develop a whole range of different tumours as well as favouring metastasis in existing tumours if exposed to oxidants. Similarly, it could be shown in rats having ACF (colonic aberrant crypt foci), a precancerous lesion in the colon, that treatment with sulforaphane or oltipraz, both Nrf2 inducers, results in a reduced number of foci.[16] Additionally, it is known that some tumour suppressors like p21 and PALB2 increase the levels of Nrf2.[10]

On the other hand, some known tumour promoters, namely K-Ras, B-Raf and Myc, amplify Nrf2-expression in cancer cells leading to a cytoprotective effect. Furthermore, an increase in anabolic metabolism through activation of glycolytic enzymes can be observed as well as a downregulation of the antiapoptotic genes Bcl-2 and Bcl-xL. Additionally, increased Nrf2-levels have been linked to chemoresistant cells which have been shown to have increased levels of GSH and metabolising enzymes.[10,16,17] Nrf2-driven chemoresistance could be demonstrated for cancers being treated *in vitro* with etoposide, 5-fluorouracil, cisplatin, and carboplatin.[17]

A loss-of-function mutation of Keap1 can be found in some cancer cells, including breast, ovary, lung, bladder, stomach, liver, biliary and prostate cancer cells, which lead to an accumulation of Nrf2 and consequently to overexpression of cytoprotective proteins with increased longevity of these neoplastic cells.[10,14,17,18] The type of mutations include missense changes, deletions and insertions as well as epigenetic modifications, namely hypermethylation of CpG sites in the Keap1 promoter.[17] A range of mutations result in alterations in the conserved β-strand 2 of each Kelch repeat which contain a double-glycine motif.[17]

Furthermore, there are known mutations in Nrf2, especially in or around the ETGE and DLG motifs, that lead to loss of Keap1-interaction, giving rise to constitutively active Nrf2.[10,17] For instance, it has been found that 15% of lung cancer patients have Keap1 mutations preventing effective Nrf2 interaction. Similarly, 10% of the same patient group had mutations in Nrf2, enabling it to circumvent Keap1-mediated regulation, therefore around 25% of these patients are affected by alteration of Nrf2-signalling, both types being linked to a worse prognosis.[17] Moreover, activating Nrf2 mutations have been described in oesophagus, larynx, skin, head and neck cancers.[17,18] Intriguingly, cancer patients have either an Nrf2 or Keap1 mutation, but never both.[18]

Altogether, this provides a heterogeneous picture of the role of Nrf2 in terms of chemoprevention and tumour progression.[10]

1.1.4.3 Others

The general role that Nrf2 plays in inflammation is that oxidative stress increases the expression of NF-κB (nuclear factor 'kappa-light-chain-enhancer' of activated B-cells), a central activator of pro-inflammatory cytokines.[22]

Nrf2 has been investigated as a target to mitigate this effect as it has been shown that during ageing the levels of glutathione and its activity decrease. Lipoic acid was used as an Nrf2 inducer and successfully increased GSH levels in rats.[21]

Nrf2 has been proposed as one of the key defence factors for oxidative or electrophile-induced hepatotoxicity as many Nrf2-regulated enzymes such as HO-1, NQO1, mEH or the heavy chain of γ-GCS have been found to be upregulated after acetaminophen/paracetamol-induced hepatotoxicity.[16] Accordingly, hepatospecific Keap1-knock out mice are resistant to acetaminophen-induced hepatotoxicity.[16]

Genetic analyses of differently gallstone-susceptible mice have shown that more resistant strains have a higher basal expression of some Nrf2-dependent genes like GST and mEH, indicating that a higher activity of Nrf2 might have protective effects for gallstones.[16]

1.2 Nrf2 modulators

1.2.1 Irreversible

Most irreversible Nrf2 inhibitors are indirect inhibitors of Keap1 and act by oxidising or alkylating cysteine residues.[29] The main classes of these are (i) isothiocyanates, (ii) oxidisable phenols and (iii) α,β-unsaturated carbonyl compounds.

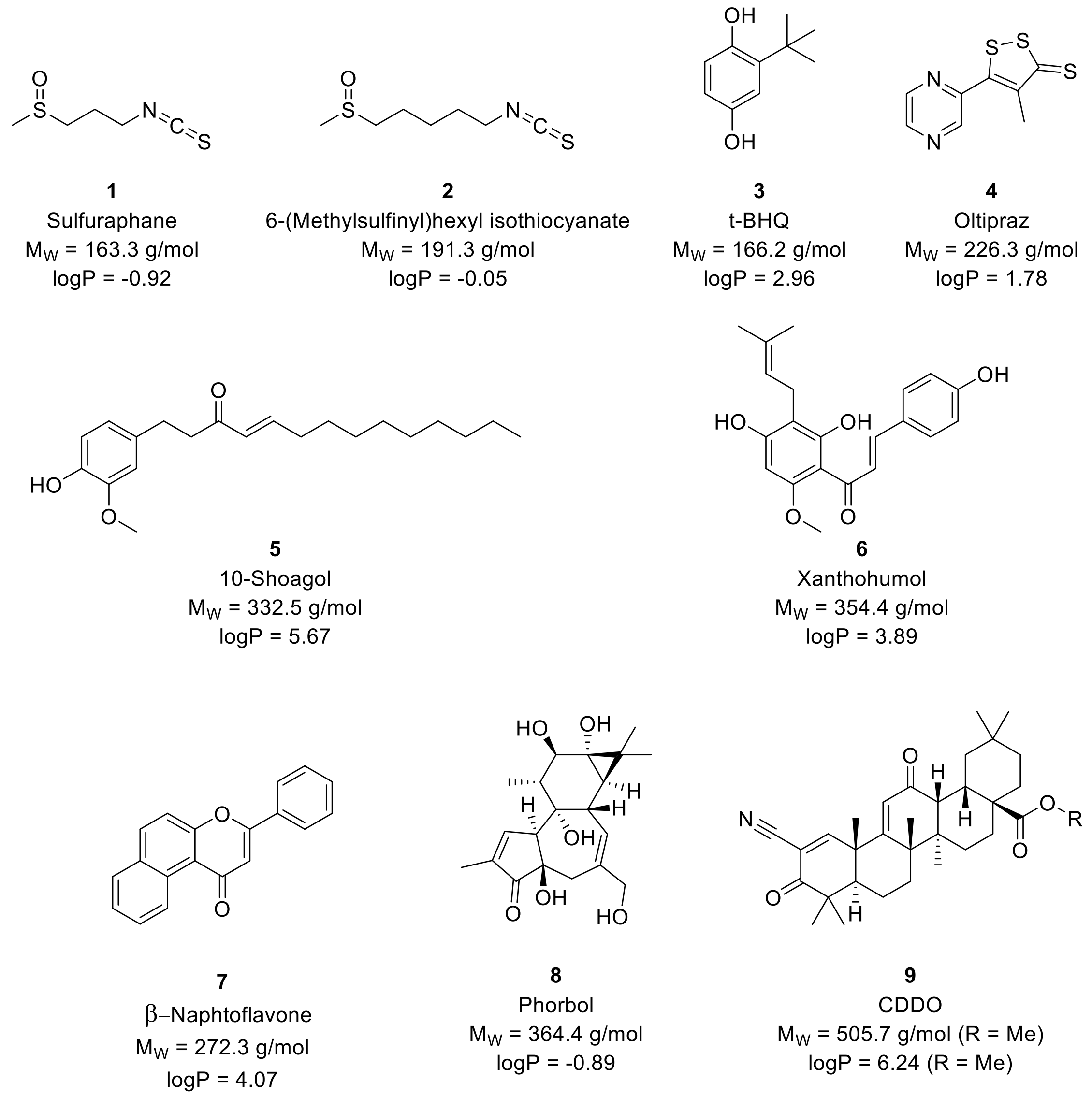
One of the best-known modulators is sulforaphane (6-(methylsulfinyl)butyl isothiocyanate) **1**, an isothiocyanate from *Crucifera* vegetables such as broccoli or Brussel sprouts. Another compound from this class and very closely related to **1** is 6-(methylsulfinyl)hexyl isothiocyanate **2**. These compounds oxidise cysteine residues in Keap1, probably Cys273 and Cys288. Thereupon, these form an intermolecular disulfide bond between both Keap1 monomers of the complex. On the other hand, it was demonstrated that Cys151 is crucial for the activity of sulforaphane in terms of Nrf2 activation. Overall, the compounds induce a change in the Keap1 conformation leading to an improved binding of Keap1 to the DLG-motif, but stalled ubiquitination, resulting in inefficient ubiquitination of Nrf2, allowing *de novo* synthesised Nrf2 to accumulate.[10,16,19,22,24] Besides these direct effects on Keap1, it is not completely clear if the isothiocyanates also mediate their effects either directly or indirectly via kinases. For example, sulforaphane activates ERK1 and 2 and suppresses p38 MAPK, due to the inhibition of the phosphorylation of the upstream kinases MKK3 and 6.[19] When investigated for neuroprotective effects, isothiocyanates showed a decrease in microglia activation and inflammatory markers after endotoxin injection. Moreover, the known effects of Nrf2 activation such as upregulation of phase II enzymes and an increase in GSH levels were observable in the basal ganglia. Furthermore, apoptosis was reduced, measured through caspase-3 activation and blood brain barrier penetration is possible.[22,24]

The dithiolethione oltipraz (3H-1,2-dithiole-3-thione) **4** is related to the isothiocyanates and contains a cyclic disulphide moiety which seems to be central for its activity as it can undergo thiol-disulfide exchange with cysteines.[29]

Examples of the other classes of proven or suspected irreversible Nrf2 activators, i.e. (ii) oxidisable phenols and (iii) α,β-unsaturated carbonyl compounds, are for (ii) tBHQ (*tert*-butylhydroxyquinone) **3** and for (iii) 10-shoagol **5**, xanthohumol **6,** β-naphtoflavone **7**, esters of phorbol **8** (a tricyclic diterpene, esters mostly of the cyclohexyl alcohols), and triterpenes, especially CDDO-esters **9**.[16,19] As for sulforaphane, it was shown that the activity of tBHQ in terms of reduced Nrf2 ubiquitination depends on the presence of Cys151. Additionally, t-BHQ **3** (but not sulforaphane **1**) was able to redirect the ubiquitination target from Nrf2 to Keap1, but this is Cys151-independent.[19] Both classes are connected as compounds from group (ii) are oxidised to quinones that then react as Michael acceptors, similarly to compounds from group (iii) and some of their structures are shown in figure 6.[29]

Additionally, the Nrf2 inducer dimethyl fumarate has been approved for treatment of psoriasis and multiple sclerosis, demonstrating the effectiveness of targeting Nrf2 for inflammatory diseases.[18]

It should be noted that the exact mechanism of activation of these molecules is not always clear, some are thought to act on Nrf2-regulating kinases as some kinase inhibitors have shown comparable activity.[18]



1
Sulfuraphane
M_W = 163.3 g/mol
logP = -0.92

2
6-(Methylsulfinyl)hexyl isothiocyanate
M_W = 191.3 g/mol
logP = -0.05

3
t-BHQ
M_W = 166.2 g/mol
logP = 2.96

4
Oltipraz
M_W = 226.3 g/mol
logP = 1.78

5
10-Shoagol
M_W = 332.5 g/mol
logP = 5.67

6
Xanthohumol
M_W = 354.4 g/mol
logP = 3.89

7
β–Naphtoflavone
M_W = 272.3 g/mol
logP = 4.07

8
Phorbol
M_W = 364.4 g/mol
logP = -0.89

9
CDDO
M_W = 505.7 g/mol (R = Me)
logP = 6.24 (R = Me)

Figure 6: Structures of irreversible Nrf2 inducers.

1.2.2 Reversible

As the irreversible Keap1 inhibitors show some off-target and side effects, *e. g.* weight loss, probably due to cross-reactivity of cysteine residues from other proteins, attempts have been made to develop reversible, competitive Nrf2-Keap1 protein-protein interaction (PPI) inhibitors.[74] This has been facilitated by the availability of X-ray crystal structures of the Nrf2-Keap1 complex which enabled rational inhibitor design. Two approaches have been pursued: peptidic and non-peptide drug-like inhibitors.

1.2.2.1 Peptides

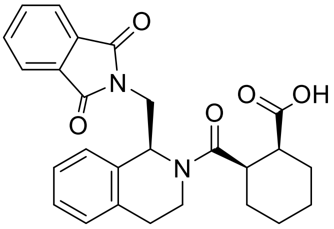
Peptidic inhibitors of the Nrf2-Keap1-interactions are based on Nrf2 ETGE-sequence. Generally, the shorter the peptide the weaker the affinity, with the hexadecamer ^{69}AFFAQLQLDEETGEFL^{84} and the tetradecamer ^{74}LQLDEETGEFLPIQ^{87} showing similar activities with K_D's of around 20 nM, but the decamer ^{76}LDEETGEFLP^{85} was less effective. A series of peptides with different lengths have been synthesised, and the minimal length required for maintaining activity was found to be 9 amino acids using the native sequence, being the peptide ^{76}LDEETGEFL^{84} with a K_D of 352 nM. Upon *N*-terminal acetylation, the potency was increased to a K_D of 21 nM.[29] However, shorter peptides with even higher affinity could be made by including residues from another Keap1 binding protein, p62, and further lipophilisation by coupling the *N*-terminus to stearic acid. This resulted in the heptamer stearyl-DPETGEL-OH with an IC$_{50}$ of 22 nM (in a fluorescence polarisation competition assay).[30] From these studies, it is clear that the inclusion of the acidic moieties is critical for potency as substituting the *C*-terminal glutamate, which shows more interactions than its *N*-terminal equivalent, with glutamine, reduces the IC$_{50}$ by one order of magnitude, compared to the glutamate analogue. However, the higher net charge of the glutamate analogue counteracts cellular uptake.[30]

1.2.2.2 Small molecules

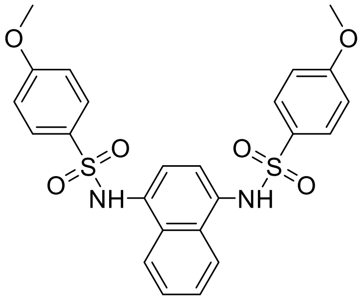
Several inhibitors have been described which all bind to Keap1. The first small molecule Nrf2-Keap1 PPI inhibitor was discovered by screening the NIH MLPCN library. This tetrahydroisoquinoline derivative **10** had a K_D of 1.0 µM (surface plasmon resonance, SPR). Interestingly, only the S,R,S stereoisomer showed activity, the other isomers that were investigated (R,S,R; S,S,R; R,R,S) were inactive.[31]

From another screen, the symmetrically substituted 1,4-diaminonaphthalene **11** was identified as a lead structure, having an IC$_{50}$ of 2.7 µM.[31] By alkylating the nitrogen of both sulphonamides with methylene carboxylic acid (-CH_2-COOH), the affinity of compound **12** was increased by three orders of magnitude, resulting in a K_D of 3.59 nM and an IC$_{50}$ of 28.6 nM in an FP assay.[32] This demonstrates, as for the peptide inhibitors, the importance of acidic groups for potent inhibition. There have been attempts to diversify the substitution pattern while maintaining the 1,4-diaminonaphthalene core, especially with regards to reducing the overall net charge. While losing some activity, the structures remained active. Compound **13** for example was found to have an IC$_{50}$ of 0.14 µM (FP).[33] However, it should be noted that these naphthalenes have rather high molecular masses, compound **12** for instance has a M_W of 614.6 g/mol, which makes them not very favourable as lead structures, especially for central nervous system applications.

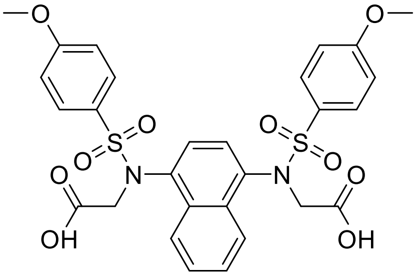
To date, the most potent Nrf2-Keap1 PPI inhibitor **14** has been developed by GSK with a K_D of 1.3 nM (isothermal titration calorimetry, ITC) and 95% inhibition of the interaction at 15 nM (FP). The structure has been identified through a fragment based approach, however, the structure is not dissimilar to the naphthalenes. Although the structure is rather large (M_W of 550.6 g/mol) and the pharmacokinetic properties were not optimal, the compound showed *in vivo* activity in rats after intravenous administration.[34] See figure 7 for their structures.



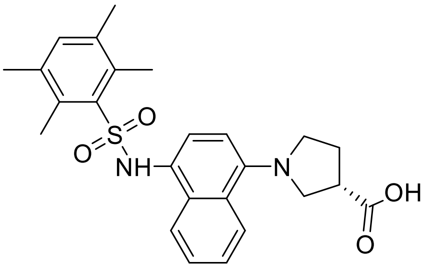
10
M_W = 446.5 g/mol
logP = 3.09
K_D = 1.0 μM



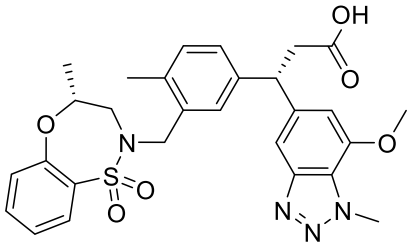
11
M_W = 498.1 g/mol
logP = 3.59
IC_{50} = 2.7 μM



12
M_W = 616.4 g/mol
logP = 2.60
K_D = 3.59 nM



13
M_W = 452.6 g/mol
logP = 5.29
IC_{50} = 0.14 μM



14
M_W = 550.2 g/mol
logP = 4.21
K_D = 1.3 nM

Figure 7: Structures, molecular weights, calculated logPs and reported binding affinity of known reversible Nrf2 inducers. For details, see text.

1.3 The challenge of targeting the CNS with drugs

1.3.1 Blood-brain barrier

It is essential for the correct functioning of the CNS that the neuronal microenvironment, especially ion concentrations, are maintained between narrow boundaries. The main point of exchange and control between the CNS and the rest of the body is the blood-brain barrier (BBB) which is mainly composed of the endothelial cells that line the capillaries in the brain and the spinal cord. Additionally, pericytes, astrocyte feet processes contribute to this barrier. Figure 8 gives an overview over their arrangement.

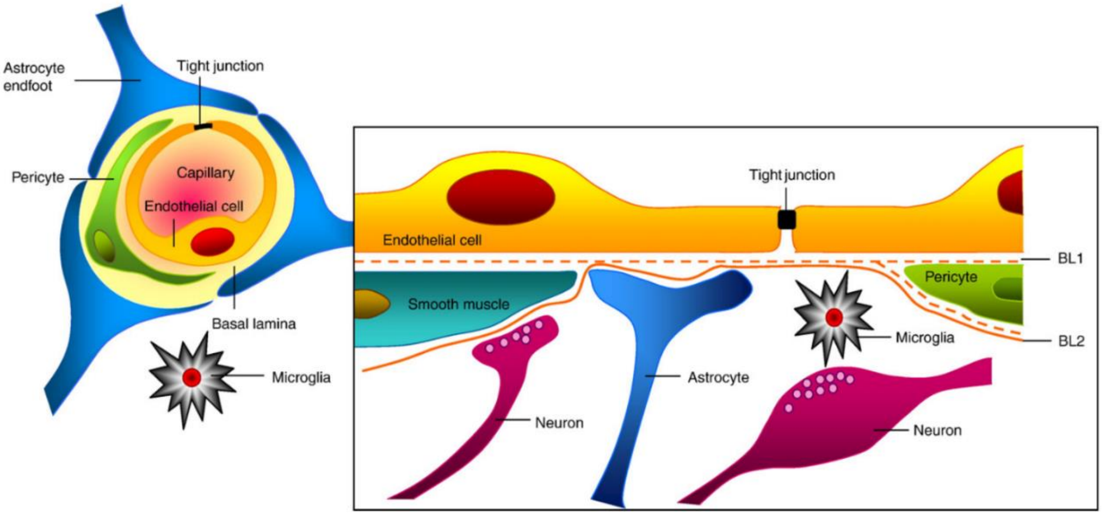


Figure 8: Sketch of a vertical (left) and horizontal (right) cross-section of a capillary in the central nervous system. The first and tightest part of the BBB is formed by the endothelial cells. On top of these, the astrocyte feet processes and pericytes are found which further seal the surface. Illustration taken from reference 35 (N. J. Abbott, A. A.K. Patabendige, D. E.M. Dolman, S. R. Yusof, D. J. Begley; Structure and Function of the blood-brain barrier; *Neurobiology of Disease* **37** 13-25 (2010))

The BBB guarantees that the uptake and concentration of ions, neurotransmitters, amino acids (especially the neurotransmitters glycine and glutamate), nutrients, xenobiotics and macromolecules (especially plasma proteins) are controlled strictly.

In the first place, the endothelium lacks the small pores, so called *fenestra*, which normally ensures a fast and quantitative exchange of molecules between the tissue and the blood.[36] Moreover, tight junctions form the major obstacle to passive diffusion of macromolecules and polar small molecules over the BBB. These junctions are particularly compact in the CNS compared to other tissues. Additionally, the cerebral endothelial cells are linked to each other by *adherens* junctions which ensure the cohesion of the cells. The tight junctions consist of three classes of proteins: claudins, occludin and junctional adhesion molecules, and are bound via the intracellular scaffold proteins ZO-1, ZO-2 and ZO-3 to the cytoskeleton. The claudins appear to be the protein class that mediates the barrier function, while the ZO proteins regulate, along with astrocytes, the induction, formation and maintenance of the tight junctions. The transport of polar molecules essential for the survival and functioning of the CNS (*e.g.* glucose, amino acids) is secured by the expression of specific transporters in the endothelial membranes whereas the uptake of macromolecules is achieved through endocytosis which is followed by exocytosis on the basolateral side, giving rise to a process called transcytosis. Transcytosis can be achieved by two different mechanisms: either receptor-mediated, which is therefore specific (receptor-mediated transcytosis) or with an excess of positive charge leading to an unspecific adsorption of the macromolecule to the endothelial membrane (adsorptive-mediated transcytosis).

Another way to control compound exposure to the CNS depends on ABC (ATP-binding cassette) transporters which are highly expressed in the central nervous endothelium. It is a protein family of efflux pumps, consisting of 48 members, which have a wide range of mostly lipophilic substrates, especially xenobiotics. The most important family members for the BBB are the permeability glycoprotein (Pgp, also known as multidrug resistance protein, MDR), multidrug resistance-associated proteins (MRPs) and the breast cancer resistance protein (BCRP).[35] Pgp is the most important member of this family as it has the widest range of substrates.[36] It is, like BRCP, only expressed on the luminal side of the endothelium and therefore has a clearing function, whereas MRPs can also be found on the basolateral side. As MRPs recognise more polar molecules, it was suggested that their clearing function is not as pivotal as that of Pgp and BRCP. [35] Other relevant classes of transporters are the organic ion transporters and the monocarboxylic acid transporters.[36]

1.3.2 Properties of CNS drugs

There is still an unmet need for drugs treating CNS disorders which is linked to the challenges of delivering them to the brain. Even if molecules are able to enter the endothelial cells that comprise the blood-brain barrier (BBB), ABC transporter may export them again. This fact has been identified as critical in drug development, as higher lipophilicity, which might be pivotal for crossing the BBB, increases the probability of a molecule being recognised as a substrate by ABC transporters.[35] However, it has been described for several CNS-related diseases, including PD, that the activity of Pgp is decreased in these conditions which might be advantageous from a drug development perspective.[35] Furthermore, an increased lipophilicity might increase metabolic lability due to cytochrome P450 metabolism. In order to avoid issues around lipophilicity, it has been suggested to begin a CNS drug development programme with a candidate having a logP of around 2. Although it is widely recognised that lipophilicity is a central parameter for CNS activity, the logPs of marketed drugs vary widely. Buchwald *et al.* found that if the logP range is broken down into intervals of 0.25, at least 5 out of the 405 studied drugs can be found in each of these intervals for a logP range from -0.25 to 5.75, with some drugs even having more extreme values. It should be noted that these values only measure the distribution of the neutral form (in contrast to the logD which takes into account all ionisation states) and most of these structures are ionisable, with basic nitrogens being commonly found. This reduces the apparent logP by 1 - 2 units for these molecules.[36] The ability of these basic, polar compounds to penetrate more easily through the BBB has been linked to their positive charge which enables them to interact with the negatively charged phospholipids and glycocalyx.[35] The vast majority of the 405 reviewed structures have 2 or fewer hydrogen bond donors (HBD), 2 - 4 hydrogen bond acceptors (HBA) and a M_W between 150 and 450, which fits the general belief that CNS drugs should not have a M_W larger than 450 Da and should not form more than 6 hydrogen bonds.[35,36] It has been proposed, as a general, approximate rule that the BBB permeability of drugs decreases 10-fold for the addition of each pair of hydrogen bonds.[72] Furthermore, the polar surface area should not exceed 80 Å².[35] As it has been suggested for an acceptable oral bioavailability that the total count of HBA and HBD should not exceed 12 and the polar surface 140 Å², this clearly shows the bigger difficulty to develop CNS-permeable molecules over other drugs.[37]

Another important factor is the number of rotatable bonds, although there is not a clear-cut threshold. A finding that supports this importance is that above a M_W of 400 Da, the BBB permeability does not increase in proportion with an increase in lipophilicity.

A possibility to circumvent these difficulties in uptake is that some synthetic structures, especially peptides that are close to the structure of transporter substrates, might be able to hijack these transporters and be actively transported. Alternatively, there have been attempts to develop prodrug or receptor-mediated targeting approaches by linking the drug to a molecule that is the substrate of a transporter.

Other approaches that intend to deliver the drug to the CNS but do not fall necessarily into the field of active substance-focused medicinal chemistry, include BBB disruption, nanoparticle delivery, receptor-mediated transport, cell-penetrating peptides, intracerebroventricular delivery, intracerebral delivery, intranasal delivery and prodrugs. Intranasal delivery is an intriguing approach as it is non-invasive and even peptides like insulin or vasopressin can be delivered to the brain. Cell-penetrating peptides, *e.g.* TAT-derived peptides, that have already been used successfully for enhancing the cellular uptake of Nrf2-derived peptides, are also interesting tools. As most of these peptides carry a substantial positive net charge, it is thought that their mechanism of action involves non-specific adsorption to the negatively charged phospholipids in the cell membrane.

Prodrugs are molecules that are pharmacologically inactive, but become active drugs through *in vivo* metabolism. The most important technique in creating a prodrug is esterification and it is particularly appealing for CNS drug applications as it increases lipophilicity. Depending on the structure that is used for the esterification, a selectivity in terms of organ and release time of the parent compound can be achieved. The concept of prodrugs has become more and more elaborate over the years, resulting in the creation of 'chemical delivery systems'. These systems comprise several chemical modifications to the active drug, resulting in a site-specific uptake, metabolism and activation. The idea of tissue-specific metabolism is that the molecule becomes more hydrophilic after being metabolised and gets locked-in at the target site. The most used system so far is 1,4-dihydrotrigonelline. The *N*-methyl-1,4-dihydropyridine core is oxidised *in vivo*, giving rise to the hydrophilic, quaternary amine trigonelline, a structure very closely related to NADH/NAD^+. The polar cation cannot cross the cell membrane and thus, is trapped within the site of oxidation. A useful aspect of this system is that it is not only able to deliver drugs to the brain, but to deliver it preferentially to the CNS.[36]

1.4 Molecular Modelling

Molecular modelling is a discipline which uses theoretical models, usually based on classical or quantum mechanics, to predict the structure and behaviour of molecules.[3] Here, the section of molecular modelling of interest is molecular docking where the interaction of two molecules is investigated, and usually one molecule is docked on or into another one using a docking programme. This is computer software, designed to find the most energetically favourable position of two molecules relative to each other and to estimate the energy of the resulting interaction. For the case of finding small molecules interacting with a macromolecule, a binding pocket is first defined for the macromolecule. Consequently, a potential ligand is placed into this pocket and moved (translationally and rotationally) until the energetically most favourable conformation is found.

In order to be able to fulfil its tasks, every molecular docking programme uses two algorithms: a search algorithm (sometimes called the sampling algorithm) which generates the multiple, possible conformations of the protein-ligand-complexes, and the scoring algorithm (which is usually called a scoring function) that determines the score of the respective complexes, constructed previously by the search algorithm. This score should ideally correlate with the thermodynamics of the interaction.[39]

1.4.1 Search Algorithms

There are three different types of search algorithms.

1.4.1.1 Rigid-body (or shape matching) search algorithms

These algorithms consider both protein and ligand as rigid. Hence, the conformations obtained are based on geometrical complementarity between both molecules. In MSDOCK and DOCK (original and newer versions) this kind of algorithm is implemented.[3,13]

1.4.1.2 Flexible-ligand search algorithms

Here, the target is kept rigid while the conformational space of the ligand is considered. This is the most widely used type of search algorithm. It is essential in order for these algorithms to perform well that the macromolecule conformation used is representative for that occurring in the real complex.

There are two subtypes of these algorithms[4,8]:

a) Systematic search algorithms explore either all degrees of freedom of the ligand (*e. g.* this is the case for Molecular Dynamics), or employ fragmentation-based methods (or alternatively retrieving such information from databases). Generally, this is a computationally expensive approach as the number of conformers increases exponentially with the number of rotatable bonds. The conformational systematic flexible-ligand search algorithm is implemented in DOCK (4.0 and more recent).[3,4,8]

b) Random (or stochastic) search algorithms apply stochastic changes to the conformation of the ligand. These changes can then be rejected or accepted by a predefined probability function. The generally implemented methods are Monte Carlo algorithms, Genetic algorithms or Tabu Search methods.[4,6]

   a. Monte Carlo algorithms take into account a Boltzmann probability function as the acceptance criterion of a ligand pose. The Monte Carlo algorithm docks the ligand in different poses by operating randomly created translations and rotations, decreasing the probability of the ligand being trapped in local minima. An energy-based selection criterion is applied in order to choose which poses are kept. Prodock, DockVision and AutoDock (Monte Carlo Simulated Annealing, specifically) have implemented this type of algorithm.[3,13]

   b. Genetic algorithms are a global search strategy that try to find the pose closest to the global energy minimum. The different degrees of freedom are encoded as genes (in the form of binary strings) which are then changed (‘mutations’) or exchanged (‘crossover’). They are heuristic algorithms and emerged from the concepts of genetics. If a specific combination surpasses an energetic threshold, the conformation is taken as seed for the next generation. GOLD and AutoDock (Lamarckian Genetic Algorithm) contain such an algorithm.[6,13]

c. Tabu search methods are meta-heuristic methods. ‘Heuristic’ methods are employed when classical approaches are impossible to use or would require too much computational power. It decides at every branching point of a searching algorithm which path to follow based on available information. However, it might not always give the best existing solution as it may for example approximate the exact solution. These algorithms follow an iterative procedure where a ligand is moved from one pose to another. While doing this, several restrictions are imposed that assure that a previously considered pose is not reconsidered, and these previously considered poses are saved in a Tabu list. From a newly created conformation, usually the RMSD (root mean square deviation) of the molecule’s atoms relative to the conformations in the Tabu list are calculated. The RMSD is then used as a criterion for acceptance or rejection of a newly created pose. PRO_LEADS uses this approach.[8.13]

1.4.1.3 Flexible-Ligand and Receptor Search Algorithms

Partial protein flexibility is taken into consideration, especially for side chains within the binding pocket. The approaches used are Molecular Dynamics methods, Monte Carlo methods, rotamer libraries, protein ensemble grids and soft-receptor modelling. The first two are very accurate by explicitly taking all the degrees of freedom into consideration, and may include solvent if necessary. The use of rotamer libraries is the most popular method. It represents the protein conformational change as a set of (experimentally observed) preferred rotameric states for each residue’s side chain. Soft-protein approaches compute a weighted average of different experimentally obtained and computed protein conformations to obtain one energy weighted average grid. It’s the computationally least demanding approach, but it cannot handle large-scale receptor motions.

1.4.2 Consideration of ligand flexibility in the chosen docking software

In addition to discussing general principles of ligand pose construction, it is also important to describe how the utilised software packages, listed below, specifically create ligand poses and how ligand flexibility is treated, enabling a more critical view of sampled poses.

1.4.2.1 UCSF DOCK

UCSF DOCK[43] uses an anchor-and-grow algorithm for creating different ligand conformations. First, one moiety of the ligand is defined as the anchor, which is defined as the largest set of atoms which are separated solely by non-rotatable bonds (any bond in a cyclic systems, multiple bond, terminal bond or NH-C=O-bond). This rigid anchor is docked into the binding pocket, its position scored and the conformation optimised. Then each flexible part of the molecule is added sequentially and the conformations are created and scored. This is thought to improve the sampling of ligand conformations as only those which are relevant within the binding pocket are considered.[43]

1.4.2.2 Autodock Vina

Autodock Vina uses a so-called iterated local search global optimiser to account for ligand flexibility. It consists of two main parts: one being the creation of 'mutations', using Genetic Algorithms (GA, s. 1.4.1.2.b). The second part is an optimisation algorithm, here the Boryden-Fletcher-Goldfarb-Shanno method. This approach considers the calculated score and the change in score gradient as the iterations progress. This is computationally more expensive in the first place as the derivative of the scoring function as a function of the mutation has to be calculated, however, this can improve the overall speed of ligand pose creation and optimisation.[39,45]

1.4.2.3 rDock

rDock uses a combination of Genetic Algorithms (GA, s. 1.4.1.2.b), Monte Carlo-based refinement and Simplex energy minimisation.[42] First, three rounds of a GA search are performed. In the different rounds, certain parameters of the scoring function are varied. The Lennard-Jones-potential, *e.g.*, is hardened from a 4-8 to a 6-12 potential, making it less tolerant towards close contacts.[13,42] Angles (dihedral as well as between bonds and the axes of the coordinate system) and the centre of mass of the ligand are encoded as genes. These parameters are randomly changed and the newly created poses assessed. Finally, a retained pose is refined using a low temperature Monte Carlo algorithm and energy minimised using a Simplex routine.[42]

1.4.2.4 LeDock

LeDock uses a combination of simulated annealing and a Genetic Algorithm (GA). Simulated annealing originates from materials science, used for the simulation of cooling of metals. Random changes to the ligand's conformation are performed, and depending on a pre-defined probability function, which is linked to a "temperature", are retained. The higher the "temperature", the more tolerant the probability function is towards accepting newly created poses, even if these are considered to be energetically less favourable. This can mitigate the risk of being trapped in a local energy minimum. Here, simulated annealing is used to create a first pose, the optimisation is done with a GA.[44]

1.4.3 Scoring functions

There are three types of scoring functions: force field-based, knowledge-based and empirical.[4,39]

1.4.3.1 Force field-based scoring functions

Classical electrostatic and van der Waals-forces (vdW) are estimated explicitly to compute the binding energy between receptor and ligand, sometimes the internal energy of the ligand is included (*e.g.* AutoDock and DOCK). Two molecular mechanical force fields are widely used: the Amber force field (used by AutoDock and DOCK) and Tripos (used by D-Score for example). Their accuracy is generally lower than of the two other methods as they have been usually established for enthalpic changes in the gas phase, and not processes where entropic effects play a major role and solvation and desolvation processes take place.[3,4,13]

1.4.3.2 Knowledge-based scoring functions

They use purely statistical energy potentials. These are derived from the structural information that is embedded in experimentally determined atomic structures: pairs of atoms that are frequently found in close proximity (*i.e.* below some pre-defined average distance) are judged to be energetically favourable. Other typical interactions are derived from structural data. Therefore, they are usually more helpful for reproducing experimental structures than binding affinities. DrugScore and DSX use such scoring functions.[4,13]

1.4.3.3 Empirical scoring functions

They have the form of $\Delta G = \Sigma W_i \cdot \Delta G_i$ with ΔG_i being different energetic terms, *e.g.* for electrostatic, hydrophobic, vdW interactions or the ligand's conformational entropy. W_i is an empirically derived weighting factor, obtained from fitting a training data set of protein-ligand complexes with known binding affinities. Therefore, they estimate the binding energy by calculating the weighted sum of all parts of the interaction. Though many of the individual contributing terms have counterparts in the force-field based terms, the functional form is simpler. The main purpose for developing these was to account better for hydrophobic interactions. Cyscore, SCORE and X-SCORE are empirical scoring functions. An open question is if such scoring functions are suitable for protein-ligand interactions other than those used in the training set.[3,4,13]

1.4.4 Assessment of accuracy of docking programmes

As already described, docking programmes fulfil two tasks: generating realistic binding poses and assessing the energy of the resulting interactions. Therefore, accuracy should be assessed in two different parts: by the sampling power which describes how well a programme is able to reproduce known binding poses (and therefore is expected to predict correct binding poses for new ligands) and the scoring power which determines how well a programme is able to estimate the energy of an interaction.[39]

The accuracy of the generation of binding poses is usually given as RMSD compared to the experimentally obtained binding pose. Generally, a constructed binding pose with a RMSD < 2 Å is considered to be successfully docked.[39]

The accuracy of the scoring function is usually expressed in two different ways, and frequently both are used for comparing the scoring power. Firstly, there is the Pearson correlation coefficient and secondly, the Spearman's rank correlation coefficient. Mathematically, both are defined in the same way:

$$Correlation\ coefficient = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where x_i, y_i are the output values of one input in two different datasets (*e.g.* the predicted binding energy by a programme and the experimentally determined value of the same ligand) while $\bar{x}, \bar{y}$ are the average values of the respective property (*e.g.* score and binding energy) of the whole dataset. The difference between the two coefficients is that the Pearson correlation coefficient relates to the binding affinity itself, whereas for the Spearman rank correlation coefficient, the data points (*i.e.* ligands) are sorted according to their value (*i.e.* score or binding affinity) and assigned ranks based on that order and only then are they correlated. In general, the Spearman rank correlation is thought to be more important in virtual screening as at this point, discriminating between good and bad structures is more useful than the quantitative evaluation of the binding interaction. The values for the coefficients vary between -1 and +1, where +1 means a perfect positive correlation, 0 no correlation, and -1 a perfect negative correlation.

Generally, it has been found that consensus predictions, relying on the output of several docking programmes, for example by averaging the results, give superior outcomes compared to using a single software.[4,39] Although a large number of reviews assessing the accuracy of different programmes are available, most of them are of limited use for the end user as many focus on optimising the source code for enhancing results or are simply outdated. However, a review from Wang *et al.* compared 10 different docking programmes, of which 5 are of academic origin and freely available for use by universities. These 5 academic programmes were LeDock, rDock, Autodock, Autodock Vina, and DOCK. For the sampling power, LeDock, rDock and AutoDock Vina performed very well, as 57.4%, 50.3% and 49.0%, respectively, predicting their top scored poses within a RMSD of 2 Å of the native binding pose of a dataset of 2002 protein-ligand complexes. It is noteworthy that LeDock outperformed most of the commercial programmes. The number of rotatable bonds appears to be a critical issue for predicting an accurate binding pose, as the scoring pose of all programmes drops significantly if a ligand contains more than 20 rotatable bonds. However, most approved drugs have fewer than 10 rotatable bonds, therefore the practical relevance of this problem remains unclear. Importantly, it has been reported that, depending on the programme, the starting conformation influences how well a programme predicts a binding pose, even though this should, theoretically, not be the case for a robust algorithm. Autodock Vina appears to be rather sensitive to the starting configuration, in contrast to DOCK, LeDock and rDock, which are less so.

It is important to note that docking and scoring power do not necessarily correlate with each other. For example, even though rDock and Autodock Vina have a similar docking power, their scoring power varies widely: on the same dataset as mentioned above, the Spearman's rank correlation coefficient for AutoDock Vina was 0.580 while it was only 0.017 for rDock, making them the best and the worst performing academic software, respectively. As docking and scoring algorithms can be combined relatively easily with each other, this gives the opportunity to dock ligands with a well-performing docking algorithm while scoring it with another, potentially superior scoring function.[39]

1.5 Biological Evaluation: Fluorescence Polarisation Assay

Polarisation refers to the direction of a wave’s oscillation. If the direction of the oscillation remains constant over time, the wave is said to be polarised. When a molecule absorbs light and reemits it spontaneously (a process called fluorescence), the polarisation, if present, of the irradiated lights is preserved in the radiated light if the fluorescent molecule does not rotate in the plane perpendicular to the direction of propagation of the light during the time span between light absorption and reemission. This is because the fluorophore has the highest chance of being excited if the absorption dipole is parallel to the electric dipole of the incoming light.[38] In general, the bigger the volume of an object (be it a particle, a molecule or a ligand-protein complex) the slower it rotates. Therefore, the smaller a fluorescent entity, the bigger the angle between the polarisation of the absorbed and the emitted light. This property is made use of in a fluorescence polarisation (FP) assay, where the binding affinity of a fluorescent probe can be correlated with the emitted light intensity at 90° from the irradiated light: the stronger the binding, the weaker this intensity.

Two values are used frequently to describe the magnitude of this phenomenon: polarisation ratio (p) and optical anisotropy (r). They are defined as follows: $p = \frac{I_{II} - I_{\perp}}{I_{II} + I_{\perp}}$ and $r = \frac{I_{II} - I_{\perp}}{I_{II} + 2I_{\perp}}$, where I_{II} and $I_{\perp}$ denote the measured light intensities parallel and perpendicular to the illumination light, respectively. Therefore, p varies between -1 and 1: -1 if all reemitted light has a perpendicular, +1 if it has a parallel polarisation.[38] The r value is the physically more meaningful measure, as the denominator corresponds to the total intensity.

In a competitive FP assay, the potential inhibitors (or ligands in general), the protein and a fluorescently labelled ligand are incubated. If the inhibitors show affinity to the target, they will displace the known binder. This will be observable as an increased intensity of the light perpendicular to the irradiated light, and the degree of increase will depend on the affinity of the inhibitors. Therefore, the lower the *p* value, the greater the displacement of the fluorescent probe by the inhibitor. By using different inhibitor concentrations, a concentration-intensity curve can be constructed, enabling the determination of the IC_{50}. In general, these assays are performed in 96- or 384-well plates and can be read out with a suitably equipped fluorescence spectrophotometer plate reader.

A possible setup for an experimental arrangement is displayed in Figure 9.

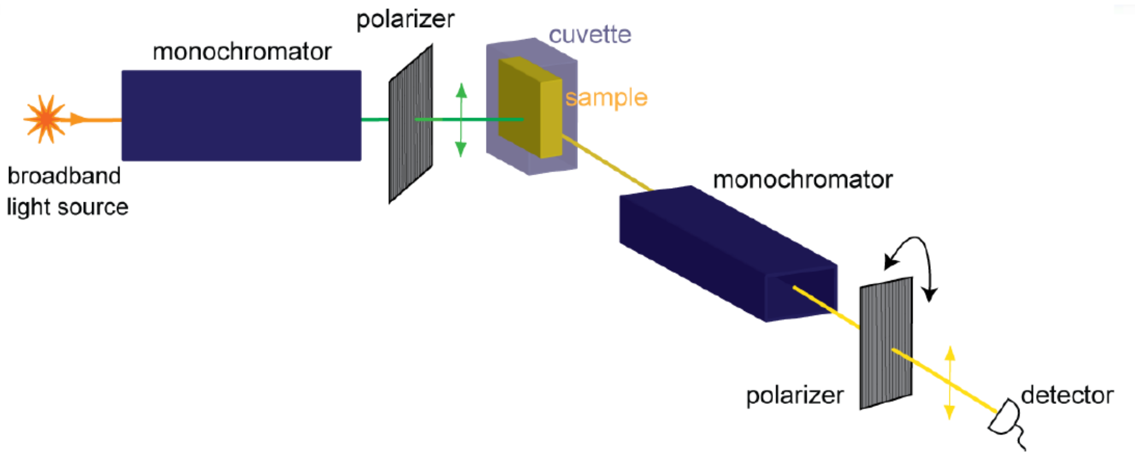


Figure 9: A possible setup for a fluorescence spectrophotometer used for an FP assay. Monochromators are grids or crystals that only let through light of a specific wave length. Alternatively, a filter can be used to select the wavelength; this is typically employed in plate readers. The polariser selects waves with a specific polarisation. This arrangement is called L-shaped, the output polariser is rotated in order to obtain the parallel and the perpendicular light intensities. There is as well a T- shaped arrangement possible, where the emitted is split into two and the parallel and perpendicular intensities are registered simultaneously. From reference 38 (T. Erdogan; Fluorescence Polarization in Life Sciences; *Semrock White Paper Series*).

1.6 Project Aims

To date, no drug treatment that slows down neuronal cell loss is available for patients suffering from PD. Due to the ageing of the population, the prevalence of this condition is expected to further increase. The Nrf2-Keap1 PPI represents a promising target for a new PD treatment as this interaction is central to the body's response to oxidative stress, a key process in neurodegeneration. Although much is already known in terms of the genes regulated by Nrf2 and its role in different cancers, its importance for neurodegenerative diseases is only beginning to emerge. The overall goal was to design and synthesise small molecule Nrf2-Keap1 PPI inhibitors that have physiochemical properties in accordance with central nervous availability in order to provide the basis of a new way to fight this frequent and severe disease.

As known irreversible inhibitors of Keap1 showed some safety and toxicity concerns, despite having promising activity, the main focus of the research community switched to reversible inhibitors. Some of these structures are quite active, especially if taking into consideration that targeting PPIs is rather challenging, but most lack promising physicochemical properties, either due to high polarity or a poor solubility, that would make them suitable for CNS applications.

In order to find new potential scaffolds, we decided to virtually screen large compound libraries with diverse sets of structures. Based upon the screening results, the structures have been iteratively refined until lead structures were picked. These structures were then prioritised for synthesis and testing using a competitive FP assay with the aim of subsequently confirming or rejecting the docking results.

In parallel, structural variations on a previously developed lead scaffold have been made to develop the respective structural-activity relationships. Consequently, the results obtained were used as a basis for further virtual screening.

Both approaches were based on molecules that are built from relatively easily accessible fragments. The intention was on one hand to test the fragments themselves and on the other hand to be able to link different building blocks from both series which would enable the exploration of synergistic effects and further optimisation. After having defined the crucial parts of the scaffold(s), an optimisation for blood brain-barrier crossing could be performed more rationally.

2 Methods and Experimental Procedures

2.1 Virtual screening

2.1.1 Receptor preparation

All Keap1 crystal structure files were obtained from the protein data bank (PDB)[40] in the pdb file format. The protein structures were converted into a mol2 file format using UCSF Chimera[41], version 1.10.2 using the Dock Prep tool with default settings. This functionality deletes ions and solvent molecules, adds hydrogens and partial charges and replaces incomplete side chains. These mol2 receptor files were used as input files for rDock[42]. For UCSF DOCK[43] version 6.7, all hydrogens were deleted and the receptor files saved in the pdb format. The receptor for LeDock[44] was prepared using the incorporated LePro command which uses the raw pdb file as input. For AutoDock Vina[45] version 1.1.2, the input receptor file in the pdbqt format was created using AutoDock Tools[46] version 1.5.6. The pdbqt file retains the polar hydrogens and includes partial charges assigned in AutoDock Tools.[46]

The chosen Keap1 structures were aligned to each other, taking the entry 4IQK as a reference, and using the Match→Align functionality of UCSF Chimera[41] with default settings.

2.1.2 Ligand preparation

The initial compounds libraries were obtained from the ZINC database[47] in mol2 format. Subsequent libraries were created with ChemDraw Professional 15.1, then energy minimised and saved as mol2 files using Chem3D. Partial charges were added with Chimera[41]. These files were then used as input for DOCK[43], and compiled into a single list for use with LeDock[44]. rDock[42] required the ligands in the sd file format which were created using Open Babel. The input ligand files for Autodock Vina[45] were made with AutoDock Tools[46] by removing all non-polar hydrogens and merging the charges.

2.1.3 Binding pocket definition

As all chosen proteins were aligned to the coordinates of the 4IQK protein structure, the ligand from this crystal structure complex was selected as a reference for defining the binding pocket. For AutoDock Vina[45], the grid was constructed visually with AutoDock Tools[46] and was set to be large enough (20 x 20 x 20 Å) to ensure the totality of the binding pocket would be captured. The coordinates of the centre of the grid box were read from the graphical interface. These coordinates were used for LeDock[44] where the binding cavity is defined by boundary coordinates in each of the x, y and z dimensions. Here as well, the grid box was a cube with an edge length of 20 Å.

rDock[42] defines the binding pocket with a sphere around a reference ligand. The programme's RbtLigandSiteMapper function was used to construct the pocket around the 4IQK ligand.
DOCK[43] also constructs the binding site using a reference ligand and spheres filling the cavities, but these are not constructed automatically. First, a surface file (DMS) of the respective receptor in mol2 format was created. Using the sphgen command, spheres were created that fill the empty spaces. Afterwards, the spheres were clustered, choosing 1.4 and 4.0 Å as the minimum and maximum sphere radii. Consequently, the spheres within a radius of 5 Å of the ligand were selected using the sphere_selector function of DOCK[43]. Using the showbox and grid commands with the default settings, the grid was finally created.

2.1.4 Docking parameters

For AutoDock Vina[45] the input parameters were set at 4 kcal/mol for the energy range (greatest energy difference between the best and worst binding mode exhibited), the exhaustiveness (a measure for the time spent to find the global energy minimum) to 9, the number of binding modes that are retained to 8, and the number of CPUs used for the calculations to 8.
DOCK[43], rDOCK[42] and LeDock[44] were used with default settings.

2.1.5 Analysis of the results

The results were analysed quantitatively with MS Office Excel 2016 and visually with UCSF Chimera[41] for assessing the ligand binding modes and for determination of the protein-ligand interactions. Furthermore, UCSF Chimera's[41] "Find Hydrogen Bond" tool was used with default settings (relax constraints by 0.4 Å, 20.0 degrees) for detecting ligand-protein hydrogen bonds.

2.2 Organic chemistry

2.2.1 Analytical Chemistry

Thin layer chromatography (TLC) was used to monitor reaction progress, determine end points, and analyse fractions isolated using flash chromatography (FC). This was done by using aluminium-backed silica gel plates (Merck 60F-254), purchased from Merck, and visualisation at 254 nm. Solvents and reagents were purchased from VWR, Sigma Aldrich, Fisher Scientific or Alfa Aesar and used without further purification.

^{1}H nuclear magnetic resonance (NMR) spectra were obtained from a Bruker Advance 400 Spectrophotometer at 400.13 MHz or Bruker Advance 500 Spectrophotometer at 500 MHz. The chemical shifts (δ) were measured in parts per million (ppm) relative to the internal standard tetramethylsilane (δ = 0), using the following solvent signals as references: $CDCl_3$ (δ = 7.26), CD_3OD (δ = 3.31), DMSO-*d6* (δ = 2.50) and acetone-*d6* (δ = 2.05). Multiplicities and peak shapes are labelled as follows: s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of a doublet, dt = doublet of a triplet, m = multiplet, br = broad.

^{13}C nuclear magnetic resonance (NMR) spectra were obtained from a Bruker Advance 400 Spectrophotometer at 100.13 MHz or Bruker Advance 500 Spectrophotometer at 125 MHz. The chemical shifts (δ) were measured in parts per million (ppm) relative to the internal standard tetramethylsilane (δ = 0), using the following solvent signals as references: $CDCl_3$ (δ = 77.16), DMSO-*d6* (δ = 39.52) and acetone-*d6* (δ = 206.26, 29.84).

High resolution mass spectra (HRMS) were obtained from the EPSRC UK National Mass Spectrometry Facility at the Institute of Mass Spectrometry, Swansea, UK. The mass spectrometer was a Thermo Scientific LTQ Orbitral XL using electrospray ionisation. The instrument was calibrated for the ionisation externally with caffeine, MRFA (MET-ARG-PHE-ALA), and Ultramark 1621, and additionally with sodium dodecyl sulfate and sodium taurocholate for the negative ionisation mode.[49]

Liquid chromatography/mass spectrometry (LC/MS) was performed on a Shimadzu LC/MS 2020. Column: xTerra® MS C_{18}; particle diameter: 2.5 µm; column diameter and length: 4.6 mm and 50 mm, respectively. Solvent A was H_2O + 0.1% formic acid, solvent B acetonitrile + 0.1% formic acid. The samples were dissolved in either MeOH or acetonitrile and their m/z ratio and purity determined by one of the following methods:

- Method I: running time: 7 min, LC UV detector: 220+254 nm, m/z range: 155-2000. Eluent system: Linear gradient from 0.00 to 4.00 min with a starting mixture of A 90% : B 10% reaching A 5% : B 95%. This proportion was maintained from 4.00 min to 5.00 min, then a linear gradient to A 90% : B 10% from 5.00 min to 5.20 min was run. This proportion was maintained from 5.20 min to 7.00 min.
- Method II: running time: 12 min, LC UV detector: 254+280 nm, m/z range: 155-2000. Eluent system: Initial proportion of A 90% : B 10% was maintained from 0.00 min to 1.00 min, followed by a linear gradient to A 5% : B 95% from 1.00 to 8.00 min. This proportion was maintained from 8.00 min to 10.00 min, then a linear gradient to A 90% B 10% from 10.00 min to 10.20 min was run. This proportion was maintained from 10.20 min to 12.00 min.
- Method III: running time: 12 min, LC UV detector: 254+280 nm, m/z range: 80-1000. Eluent system: Initial proportion of A 90% : B 10% was maintained from 0.00 min to 1.00 min, followed by a linear gradient to A 5% : B 95% from 1.00 to 8.00 min. This proportion was maintained from 8.00 min to 10.00 min, then a linear gradient to A 90% B 10% from 10.00 min to 10.20 min was run. This proportion was maintained from 10.20 min to 12.00 min.
- Method IV: running time: 7 min, LC UV detector: 220+254 nm, m/z range: 80-1000. Eluent system: Linear gradient from 0.00 to 4.00 min with a starting mixture of A 90% : B 10% reaching A 5% : B 95%. This proportion was maintained from 4.00 min to 5.00 min, then a linear gradient to A 90% : B 10% from 5.00 min to 5.20 min was run. This proportion was maintained from 5.20 min to 7.00 min.

High performance liquid chromatography (HPLC) was performed on an Agilent 1200 series. Column: XSELECT^{TM} CSHTM: C_{18}; particle diameter: 2.5 µm; column diameter and length: 6 mm and 50 mm, respectively. Solvent A was H_2O + 0.1% formic acid, solvent B acetonitrile + 0.1% formic acid. The samples were dissolved in acetonitrile, 10 µL injected and their purity determined by the following method:

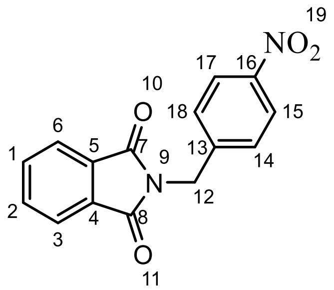
- Method I: running time: 25 min, UV detector: 254 nm. Eluent system: Linear gradient from 0.00 to 20.00 min with a starting mixture of A 95% : B 5% reaching A 50% : B 50%. This was followed by a linear gradient to A 5% : B 95 % from 20.00 min to 22.00 min. This proportion was maintained until 24.00 min. From 24.00 min to 25.00 min, a linear gradient was run, reaching A 95% : B 5%.

2.2.2 Synthetic Chemistry

General procedures:

A: Click-chemical synthesis of 1,2,3-triazoles

The alkyne (1 equiv.), (3-azidophenyl)(methyl)sulfane **65** (1 equiv.), sodium ascorbate (0.2 equiv.) and $CuSO_4 \cdot 5H_2O$ (0.05 equiv.) were combined in a mixture of H_2O (7.57 mL/mmol) and *t*-BuOH (5 mL/mmol). The mixture was sealed in a microwave flask and reacted under microwave irradiation at 130 °C for 45 min. The precipitated 1,2,3-triazole products were isolated by filtration and tested without further purification, if not stated otherwise.



2-(4-Nitrobenzyl)isoindoline-1,3-dione (47)

Exact Mass: 282.0641 Molecular Weight: 282.2550

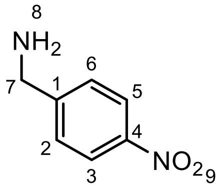
4-Nitrobenzylbromide **45** (2.000 g, 9.259 mmol), potassium phthalimide **46** (1.5 equivalents, 2.5719 g, 13.89 mmol) and tetrabutylammonium bromide (0.0090 g, 0.001 equivalents, 0.028 mmol) were dissolved in DMF (6 mL) and stirred at 100° C for 16 h. The cooled reaction mixture was quenched with H_2O (80 mL), upon which a white precipitate formed. The mixture was filtered and washed with EtOAc (3 x 40 mL). The filtrate was extracted and the aqueous phase washed once with EtOAc (20 mL). The organic extracts were combined and washed with sat. aqueous brine (30 mL). The solvent was evaporated under reduced pressure, giving 2.234 g of 2-(4-nitrobenzyl)isoindoline-1,3-dione **47** (8.659 mmol, 93.5% yield).

^{1}H NMR ($CDCl_3$, 500 MHz): δ 8.18 (2H, d, *J* = 8.5 Hz, H15 & H17), 7.87 (2H, m, H1 & H2), 7.75 (2H, m, H3 & H6), 7.59 (2H, d, *J* = 8.5 Hz, H14 & H18), 4.93 (2H, s, H12)

^{13}C NMR (DMSO-*d6*, 125 MHz): δ 167.66 (C7 & C8), 146.82 (C16), 144.31 (C13), 134.65 (C3 & C6), 131.60 (C4 & C5), 128.47 (C14 & C18), 123.72 (C15 & C17), 123.34 (C1 & C2), 40.38 (C12)

LC/MS (method II): t_R 3.93 min, purity ≥92%, no molecular ion detected

In accordance with previously reported values by Baumgartner et al.[78]



(4-Nitrophenyl)methanamine (48)

Exact Mass: 152.0586 Molecular Weight: 152.1530

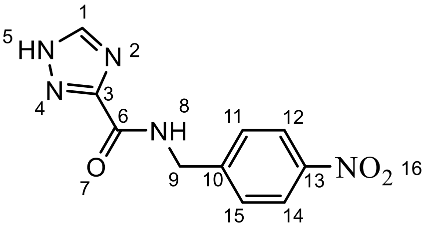
2-(4-Nitrobenzyl)isoindoline-1,3-dione **47** (2.1003 g, 8.141 mmol) was dissolved in MeOH (25 mL). Hydrazine monohydrate (1.9 equivalents, 15.47 mmol, 772.0 mg, 1.16 mL) was added dropwise with stirring. The reaction was set to stir for 5 hours under reflux. A white, voluminous cake formed. The solution was filtered and the precipitate washed with MeOH (2 x 10 mL). The filtrate was dried under reduced pressure and the residue dissolved in DCM (40 mL). The solution was washed with aqueous 1 M NaOH (3 x 20 mL). The organic layer was dried with $MgSO_4$ and the solvent removed under reduced pressure to yield (4-nitrophenylmethanamine) **48** as orange needles (1.1248 g, 7.489 mmol, 92.0% yield).

^{1}H NMR ($CDCl_3$, 500 MHz): δ 8.19 (2H, d, *J* = 10.9 Hz, H3 & H4), 7.50 (2H, d, *J* = 10.9 Hz, H2 & H6), 4.01 (2H, s, H7), 1.25 (2H, br, H8)

^{13}C NMR ($CDCl_3$, 125 MHz): δ 150.68 (C4), 147.03 (C1), 127.82 (C3 & C5), 123.87 (C2 & C6), 45.88 (C7)

In accordance with previously reported values by Baumgartner et al.[78] and Richy et al.[79]

LC/MS (method IV): t_R 0.72 min, purity ≥80%, m/z = 194.100 ($M+MeCN+H^+$)



N-(4-Nitrobenzyl)-1*H*-1,2,4-triazole-3-carboxamide (49)

Exact Mass: 247.0705 Molecular Weight: 247.2140

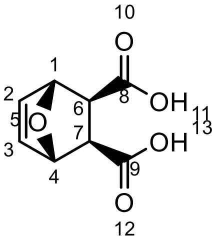
(4-Nitrophenyl)methanamine **48** (152.0 mg, 1.0 mmol) were dissolved in ice cooled, dry DMF (1 mL). 1*H*-1,2,4-triazole-3-carboxylic acid **40** (133.0 mg, 1.0 equivalent, 1.0 mmol) and PyBroP (bromotripyrrolidinophosphonium hexafluorophosphate) **50** (556.2 mg, 1.19 equivalents, 1.2 mmol) were added. The reaction flask was purged twice with argon. DIPEA (di-isopropyl ethyl amine) (0.52 mL, 3.0 equivalents, 3.0 mmol) was added dropwise and the ice bath was removed. The reaction was set to stir for 16 h. H_2O (25 mL) was added, upon which the clear red solution turned opaque. The reaction mixture was extracted with EtOAc (25 mL). Overnight, crystallisation occurred in the aqueous phase and the crystals were filtered off to yield 87.4 mg (35.4% yield) of *N*-(4-nitrobenzyl)-1*H*-1,2,4-triazole-3-carboxamide **49**.

^{1}H NMR (DMSO-*d6*, 500 MHz): δ 14.68 (1H, s, br, H5), 9.35 (1H, s, br, H8), 8.52 (1H, s, br, H1), 8.20 (2H, d, *J* = 8.8 Hz, H12 & H14), 7.57 (2H, d, *J* = 8.8, H11 & H15), 4.56 (2H, d, *J* = 6.3 Hz, H9)

^{13}C NMR (DMSO-*d6*, 125 MHz): δ 158.63 (br), 147.38, 146.47, 128.29, 123.53, 41.80

LC/MS (method II): t_R 3.62 min, purity 98%, m/z = 246.400 (M-H$^+$)

HRMS: expected: 248.0778 m/z, found: 248.0782 m/z



(1*R*,2*S*,3*R*,4*S*)-7-oxabicyclo[2.2.1]hept-5-ene-2,3-dicarboxylic acid (55)

Exact Mass: 184.0372 Molecular Weight: 184.1470

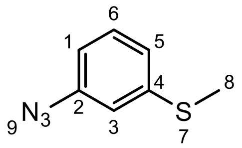
(3a*R*,4*S*,7*R*,7a*S*)-3a,4,7,7a-Tetrahydro-4,7-epoxyisobenzofuran-1,3-dione **42** (203.0 mg, 1.22 mmol) was dissolved in water (60 mL) and stirred at room temperature for 68 h. The solvent was removed under reduced pressure to yield (1*R*,2*S*,3*R*,4*S*)-7-oxabicyclo[2.2.1]hept-5-ene-2,3-dicarboxylic acid **55** as a white powder (216.5 mg, 1.18 mmol, 96.7% yield).

^{1}H NMR (acetone-*d6*, 500 MHz): δ 6.50 (2H, s, H2 & H3), 5.16 (2H, s, H1 & H4), 2.80 (2H, s, H6 & H7)

^{13}C NMR (acetone-*d6*, 125 MHz): δ 172.88 (C8 & C9), 137.52 (C2 & C3), 81.33 (C1 & C4), 47.29 (C6 & C7)

LC-MS (method IV): t_R = 2.48 min, m/z 183.1000 (M-H$^+$)

HRMS: (M-H$^+$): expected: 183.0299 m/z, found: 183.0302 m/z



(3-Azidophenyl)(methyl)sulfane (65)

Exact Mass: 165.0361 Molecular Weight: 165.2140

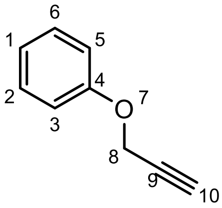
3-(Methylthio)aniline **63** (500.0 mg, 3.59 mmol, 0.44 mL) was suspended in H_2O (10 mL). Concentrated HCl (1.0 mL) was added and the solution stirred. The solution was cooled to 0 °C and consequently, a solution of $NaNO_2$ (297.4 mg, 1.2 equivalents, 4.31 mmol) in H_2O was added while maintaining the temperature at 0 °C. The solution was stirred for 20 min. After this period, a solution of NaN_3 (350.2 mg, 1.5 equivalents, 5.39 mmol) was added at 0 °C. The mixture was stirred for 3 h while ensuring that the temperature was maintained below 10 °C. The solution was extracted with Et_2O (1 x 20 mL, then 3 x 10 mL) and the combined organic extracts were washed with saturated brine. The organic layer was dried over $MgSO_4$, the solvent removed under reduced pressure and (3-azidophenyl)(methyl)sulfane **65** isolated as a brown oil (578.4 mg, 98.7% yield).

^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.25 (1H, t, *J* = 8.0 Hz, H6), 7.01 (1H, ddd, *J* = 1.0, 1.8, 7.9 Hz, H1), 6.88 (1H, t, *J* = 2.0 Hz, H3), 6.80 (1H, ddd, *J* = 1.0, 1.8, 7.9 Hz, H5), 2.48 (3H, s, H8)

^{13}C NMR ($CDCl_3$, 100 MHz): δ 140.91, 140.75, 130.01, 122.89, 116.71, 115.62, 15.62

LC/MS (method I): t_R 6.74 min, purity 91%, no molecular ion detected

In accordance with previously reported values by Bertrand et al.[51]



(Prop-2-yn-1-yloxy)benzene (70)

Exact Mass: 132.0575 Molecular Weight: 132.1620

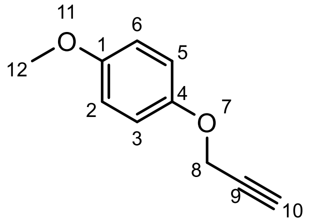
Phenol **67** (300 mg, 3.2 mmol) was dissolved in DMF (3.2 mL) and stirred for 5 min at room temperature. Then, K_2CO_3 (2.5 equivalents, 8.0 mmol, 1.101 g) was added and the mixture stirred for a further 5 min. After adding propargyl bromide **66** (1.5 equivalents, 4.8 mmol, 569 mg, 0.41 mL) dropwise, the reaction mixture was stirred for 19 h at room temperature. The reaction was quenched with H_2O (50 mL) and extracted with Et_2O (30 mL, 20 mL and 10 mL). The organic extracts were combined and washed with H_2O (10 mL) and saturated brine (10 mL) and dried with $MgSO_4$. The solvent was removed under reduced pressure and (prop-2-yn-1-yloxy)benzene **70** isolated as a yellow oil (226.1 mg, 53.5% yield).

^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.28-7.34 (2H, m, H2 & H6), 6.97-7.02 (3H, m, H1, H3 & H5), 4.70 (2H, d, *J* = 2.5 Hz, H8), 2.52 (1H, t, *J* = 2.5 Hz, H10)

^{13}C NMR ($CDCl_3$, 100 MHz): δ 157.73 (C4), 129.63 (C2 & C6), 121.74 (C1), 115.07 (C3 & C5), 78.79 (C9), 75.57 (C10), 55.91 (C8)

LC/MS (method I): t_R 3.93 min, purity 90%, no molecular ion detected.

In accordance with previously reported values by Orbisaglia et al.[80]



1-Methoxy-4-(prop-2-yn-1-yloxy)benzene (71)

Exact Mass: 162.0681 Molecular Weight: 162.1880

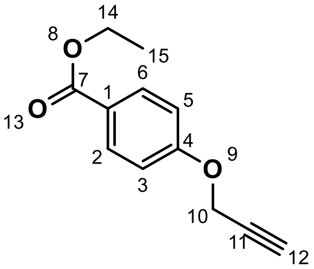
4-Methoxyphenol **68** (900.1 mg, 7.3 mmol) was dissolved in DMF (10 mL). K_2CO_3 (2.576 g ,2.6 equivalents, 18.6 mmol) was added and the solution was stirred for 5 min at room temperature. Afterwards, propargyl bromide **66** (0.71 mL (1.13 equivalents, 8.2 mmol, 975 mg) was added dropwise. The reaction mixture was stirred for 25 h at room temperature and then for 5 h at 50 °C. The reaction was quenched with H_2O and extracted with Et_2O (2 x 20 mL and 1 x 5 mL). The combined organic extracts were washed with H_2O (10 mL) and dried over $MgSO_4$. The solvent was removed under reduced pressure. 4-Methoxyphenyl propiolate **71** was isolated as an orange oil (1.1322 g, 95.6% yield).

^{1}H NMR ($CDCl_3$, 500 MHz): δ 6.93 (2H, d, *J* = 9.2 Hz, H2 & H6), 6.85 (2H, d, *J* = 9.2 Hz, H3 & H5), 4.64 (2H, d, *J* = 2.3 Hz, H8), 3.78 (3H, s, H12), 2.50 (1H, t, *J* = 2.3 Hz, H10)

^{13}C NMR ($CDCl_3$, 400 MHz): δ 154.66 (C1), 151.85 (C4), 116.31 (C2 & C6), 114.77 (C3 & C5), 79.06 (C9), 75.41 (C10), 56.79 (C8), 55.84 (C12)

LC/MS (method I): t_R 3.62 min, purity 100%, no molecular ion detected

In accordance with previously reported values by Chen et al.[81]



Ethyl 4-(prop-2-yn-1-yloxy)benzoate (72)

Exact Mass: 204.0786 Molecular Weight: 204.2250

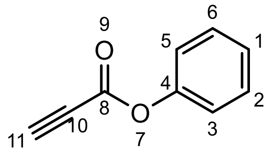
Ethyl-4-hydroxybenzoate **69** (1000 mg, 6.0 mmol) was dissolved in dry acetone (10 mL) and K_2CO_3 (1.4 equivalents, 1164 mg, 8.4 mmol) was added. The reaction was stirred for 5 min, then propargyl bromide **66** (2.0 equivalents, 12.0 mmol, 1432 mg, 1.04 mL) was added dropwise. The reaction was heated to 75 °C for 3 h. After cooling down to room temperature, DCM (100 mL) was added. The solution was washed with H_2O (50 mL), sat. $NaHCO_3$ solution (25 mL) and saturated brine (50 mL). The organic extract was dried with Na_2SO_4 and the solvent removed under reduced pressure. Ethyl 4-(prop-2-yn-1-yloxy)benzoate **72** was isolated as a yellow oil (1.1236 g, 91.7% yield).

^{1}H NMR ($CDCl_3$, 400 MHz): δ 8.02 (2H, m, H2 & H6), 7.00 (2H, m, H3 & H5), 4.75 (2H, d, *J* = 2.3 Hz, H10), 4.35 (2H, q, J = 7.1 Hz, H14), 2.54 (1H, t, *J* = 2.3 Hz, H12), 1.38 (3H, t, *J* = 7.1 Hz, H15)

^{13}C NMR ($CDCl_3$, 100 MHz): δ 166.22 (C7), 161.09 (C4), 131.50 (C2 & C6), 123.86 (C1), 114.45 (C3 & C5), 77.86 (C11), 76.02 (C12), 60.71 (C14), 55.73 (C10), 14.37 (C15)

LC/MS (method I): t_R 3.95 min, purity 100%, no molecular ion detected

In accordance with previously reported values by Stockmaier et al.[50]



Phenyl propiolate (73)

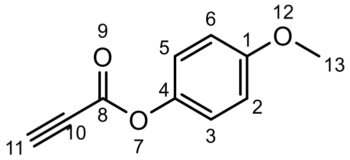
Exact Mass: 14.0368 Molecular Weight: 146.1450

Phenol **67** (600 mg, 6.4 mmol), DCC (*N*,*N*'-dicyclohexylcarbodiimide) (1315 mg, 1.0 equivalents, 6.4 mmol) and 4-DMAP (4-dimethylaminopyridine) (71.5 mg, 0.1 equivalents, 0.64 mmol) were added to a dry round bottom flask. The flask was evacuated then filled with argon. Dry DCM (10 mL) was added and the mixture stirred for 5 min. The reaction mixture was cooled in an ice bath then propiolic acid **43** (0.43 mL, 491 mg, 1.1 equivalents) was added dropwise over 20 min. The reaction mixture was stirred for 22 h at room temperature. The solution was filtered and the precipitate washed with DCM. The filtrate was washed twice each with H_2O and saturated brine, dried over Na_2SO_4 and the solvent removed under reduced pressure. The crude product was purified by FC (eluent: *n*-Hex:EtOAc = 9:1) to afford phenyl propiolate **73** (935.7 mg, 43.6% yield) as a yellow oil.

^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.41 (2H, t, *J* = 7.9 Hz, H2 & H6), 7.28 (1H, t, *J* = 7.6 Hz, H1), 7.15 (2H, d, *J* = 7.6 Hz, H3 & H5), 3.07 (1H, s, H11)

^{13}C NMR (DMSO-*d6*, 500 MHz): δ 150.78 (C4 or C8), 149.45 (C4 or C8), 129.80 (C2 & C6), 126.75 (C1), 121.56 (C3 & C5), 81.52 (C11), 74.27 C10)

In accordance with previously reported values by Yanada et al.[82] and Nagel et al.[83]



4-Methoxyphenyl propiolate (74)

Exact Mass: 176.0473 Molecular Weight: 176.1710

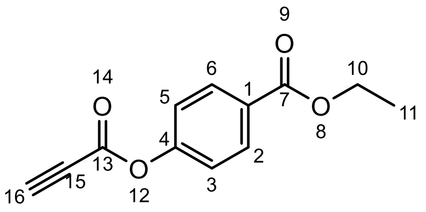
4-Methoxyphenol **68** (600 mg, 4.8 mmol), DCC (*N*,*N*'-dicyclohexylcarbodiimide) (998 mg, 1.0 equivalents, 4.8 mmol) and 4-DMAP (4-dimethylaminopyridine) (41.4 mg, 0.07 equivalents, 0.34 mmol) of were added to a dry round bottom flask. The flask was evacuated then filled with argon. Dry DCM (6 mL) were added and the mixture stirred for 5 min. The reaction mixture was cooled in an ice bath then propiolic acid **43** (0.30 mL, 339 mg, 1.0 equivalents) was added dropwise over 20 min. The reaction mixture was stirred for 48 h at room temperature. A solution of DCC (500 mg, 2.4 mmol, 1.0 equivalents) in dry DCM (1.8 mL) was added and the reaction was stirred at room temperature for a further 16 h. The reaction mixture was filtered and the precipitate washed with DCM. The filtrate was washed with H_2O (20 mL),1 M HCl (10 mL), saturated $NaHCO_3$ solution (10 mL) and saturated brine (10 mL), dried over Na_2SO_4 and the solvent removed under reduced pressure. The crude product was purified by FC (eluent: *n*-Hex:EtOAc = 9:1) to afford 4-methoxyphenyl propiolate **74** (525.7 mg, 61.7% yield)as a yellow solid.

^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.07 (2H, d, *J* = 9.2 Hz, H2 & H6), 6.90 (2H, d, *J* = 9.1 Hz, H3 & H5), 3.80 (3H, s, H13), 3.06 (1H, s, H11)

^{13}C NMR (DMSO-*d6*, 500 MHz): δ 157.93 (C1), 151.45 (C8), 143.41 (C4), 122.19 (C3 & C5), 114.72 (C2 & C6), 76.81 (C11), 74.46 (C10), 55.75 (C13)

LC-MS (method II): t_R = 5.61 min, purity ≥80%, no molecular ion detected

In accordance with previously reported values by Aparece et al.[84]



Ethyl 4-(propioloyloxy)benzoate (75)

Exact Mass: 218.0579 Molecular Weight: 218.2080

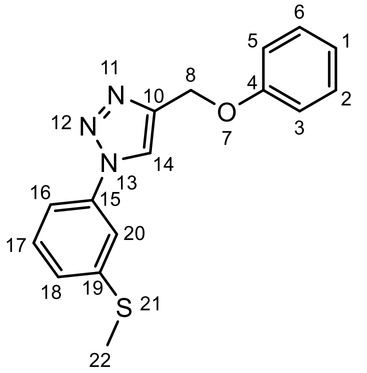
Ethyl-4-hydroxybenzoate **69** (600 mg, 3.6 mmol), DCC (*N*,*N*'-dicyclohexylcarbodiimide) (745 mg, 1.0 equivalents, 3.6 mmol) and 4-DMAP (4-dimethylaminopyridine) (3.2 mg, 0.01 equivalents, 0.036 mmol) of were added to a dry round bottom flask. The flask was evacuated and then filled with argon. Dry Et_2O (10 mL) was added and the mixture stirred for 5 min. The reaction mixture was cooled in an ice bath then propiolic acid **43** (0.22 mL, 250 mg, 0.99 equivalents) was added dropwise over 20 min. The reaction mixture was stirred for 18 h at room temperature. The solution was filtered and the precipitate washed with Et_2O. The filtrate was washed three times with 1 M HCl, dried over Na_2SO_4 and the solvent removed under reduced pressure. The crude product was purified by FC (eluent: *n*-Hex:EtOAc = 9:1) to afford ethyl 4-(propioloyloxy)benzoate **75** (788.2 mg, 43.6% yield) as a white, flocculent solid.

^{1}H NMR ($CDCl_3$, 400 MHz): δ 8.10 (2H, d, *J* = 8.7 Hz, H2 & H6), 7.23 (2H, d, *J* = 8.7 Hz, H3 & H5), 4.38 (2H, q, *J* = 7.1 Hz, H10), 3.11 (1H, s, H16), 1.39 (3H, t, *J* = 7.1 Hz, H11)

^{13}C NMR (DMSO-*d6*, 400 MHz): δ 164.83 (C7), 152.88 (C13 or C4), 150.12 (C13 or C4), 130.87 (C2 & C6), 128.29 (C1), 122.03 (C3 & C5), 81.98 (C16), 73.97 (C15), 60.91 (C10), 14.08 (C11)

LC-MS (method II): t_R = 6.29 min, purity 98%, no molecular ion detected

HRMS: ($M+H^+$): expected: 219.0657 m/z, found: 219.0657 m/z



1-(3-(methylthio)phenyl)-4-(phenoxymethyl)-1*H*-1,2,3-triazole (57)

Exact Mass: 297.0936 Molecular Weight: 297.3760

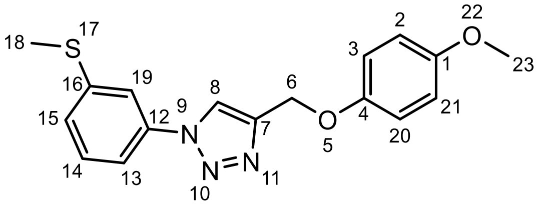
1-(3-(Methylthio)phenyl)-4-(phenoxymethyl)-1*H*-1,2,3-triazole **57** was synthesised from (prop-2-yn-1-yloxy)benzene **70** (39.6 mg, 0.42 mmol) according to the general procedure A. 1-(3-(Methylthio)phenyl)-4-(phenoxymethyl)-1*H*-1,2,3-triazole **57** was isolated as a grey solid (18.7 mg, 15.0% yield).

^{1}H NMR (DMSO-*d6*, 500 MHz): δ 8.99 (1H, s, H14), 7.75 (1H, t, *J* = 1.9 Hz, H20), 7.68 (1H, ddd, *J* = 0.9, 1.9, 8.0 Hz, H18), 7.52 (1H, t, *J* = 8.0 Hz, H17), 7.37 (1H, ddd, *J* = 0.9, 1.9, 8.0 Hz, H16), 7.32 (2H, m, H3 & H5), 7.07 (2H, m, H2 & H6) 6.97 (1H, tt, *J* = 7.3 Hz, *J* = 1.0 Hz, H1), 5.23 (2H, s, H8), 2.57 (3H, s, H22)

^{13}C NMR (DMSO-*d6*, 125 MHz): δ 158.01 (C4), 143.98 (C7), 140.85 (C19), 137.13 (C15), 130.33 (C17), 129.63 (C2 & C16), 125.82 (C16), 123.06 (C14), 121.04 (C1), 116.74 (C20), 116.42 (C18), 114.75 (C3 & C5), 60.94 (C8), 14.46 (C22)

LC-MS (method II): t_R = 6.81 min, purity ≥90%, m/z = 298.200 (M+H$^+$)

HRMS: (M+H$^+$): expected: 298.1009 m/z, found: 298.1008 m/z



4-((4-Methoxyphenoxy)methyl)-1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole (58)

Exact Mass: 327.1041 Molecular Weight: 327.4020

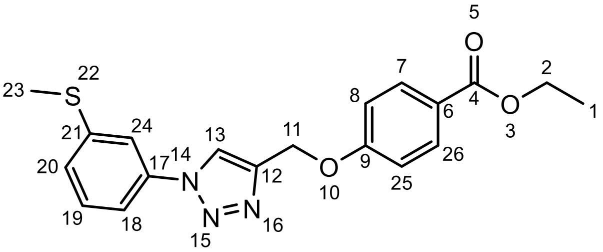
4-((4-Methoxyphenoxy)methyl)-1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole **58** was synthesised from 4-methoxyphenly propiolate **71** (68.7 mg, 0.42 mmol) according to the general procedure A. 4-((4-Methoxyphenoxy)methyl)-1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole **58** was isolated as a black solid (65.4 mg, 47.2% yield).

^{1}H NMR (DMSO-*d6*, 400 MHz): δ 8.97 (1H, s, H8), 7.74 (1H, t, *J* = 1.9 Hz, H19), 7.67 (1H, d, *J* = 8.1 Hz, H15), 7.52 (1H, t, *J* = 8.0 Hz, H14), 7.37 (1H, d, *J* = 8.1 Hz, H13), 7.01 (2H, m, H2 & H21), 6.88 (2H, m, H3 & H20), 5.16 (2H, s, H6), 3.70 (1H, s, H23), 2.57 (3H, s, H18)

^{13}C NMR (DMSO-*d6*, 125 MHz): δ 153.69 (C1 or C4), 152.01 (C1 or C4), 144.20 (C7), 140.84 (C16), 137.14 (C12), 130.33 (C14), 125.79 (C13), 122.93 (C8), 116.71 (C19), 116.39 (C15), 115.79 (C2 & C21), 114.67 (C3 & C20), 61.59 (C6), 55.39 (C23), 14.46 (C18)

LC-MS (method II): t_R = 6.71 min, purity 96%, m/z = 328.250 (M+H$^+$)

HRMS: (M+H$^+$): expected: 328.1114 m/z, found: 328.1114 m/z



Ethyl 4-((1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazol-4-yl)methoxy)benzoate (59)

Exact Mass: 369.1147 Molecular Weight: 369.4390

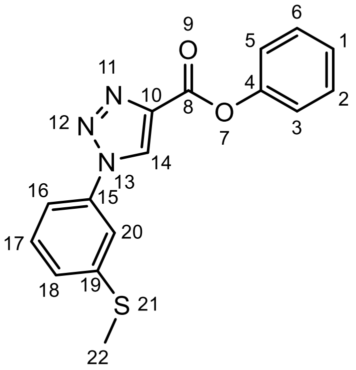
Ethyl 4-((1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazol-4-yl)methoxy)benzoate **59** was synthesised from ethyl 4-(prop-2-yn-1-yloxy)benzoate **72** (62.6 mg, 0.31 mmol) according to the general procedure A. Ethyl 4-((1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazol-4-yl)methoxy)benzoate **59** was isolated as a brown solid (32.4 mg, 28.8% yield).

^{1}H NMR (DMSO-*d6*, 500 MHz): δ 9.03 (1H, s, H13), 7.93 (2H, d, *J* = 8.8 Hz, H7 & H26), 7.75 (1H, s, H24), 7.68 (1H, d, *J* = 8.2 Hz, H20), 7.53 (1H, t, *J* = 7.9 Hz, H19), 7.38 (1H, d, *J* = 8.2 Hz, H18), 7.20 (2H, d, *J* = 8.8 Hz, H8 and H25), 5.34 (2H, s, H11), 4.28 (2H, q, *J* = 7.1 Hz, H2), 2.57 (3H, s, H23), 1.30 (3H, t, *J* = 7.1 Hz, H1)

^{13}C NMR (DMSO-*d6*, 125 MHz): δ 165.31 (C4), 161.71 (C9), 143.31 (C12), 140.78 (C21), 137.02 (C17), 131.18 (C7 & C26), 130.25 (C19), 125.79 (C18), 123.22 (C13), 122.52 (C6), 116.72 (C24), 116.39 (C20), 114.74 (C8 & C25), 61.23 (C11), 60.35 (C2), 14.38 (C1), 14.20 (C23)

LC-MS (method I): t_R = 7.11 min, purity 95%, m/z = 371.150 ($M+H^+$)

HRMS: ($M+H^+$): expected: 370.1220 m/z, found: 370.1221 m/z



Phenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate (60)

Exact Mass: 311.0728 Molecular Weight: 311.3590

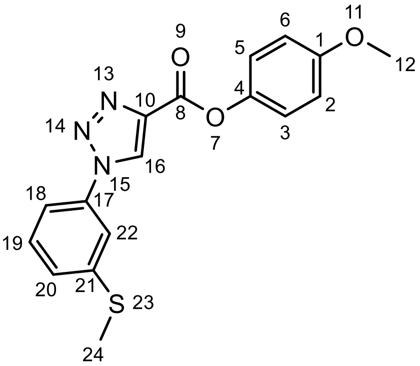
Phenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate **60** was synthesised from phenyl propiolate **73** (57.5 mg, 0.39 mmol) according to the general procedure A. The crude product was purified by FCC (*n*-Hexane:EtoAc = 9:1) and phenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate **60** isolated as a pale yellow powder (58.7 mg, 47.9% yield).

^{1}H NMR (DMSO-*d6*, 400 MHz): δ 9.80 (1H, s, H14), 7.86 (1H, t, *J* = 1.9 Hz, H20), 7.78 – 7.81 (1H, m, H18), 7.56 (1H, t, *J* = 8.0 Hz, H17), 7.48 – 7.53 (2H, m, C2 & C6), 7.42 – 7.45 (1H, m, H16), 7.31 – 7.37 (3H, m, H1 & H3 & H5), 2.57 (3H, s, H22)

^{13}C NMR (DMSO-*d6*, 100 MHz): δ 158.67 (C8), 149.95 (C4), 141.02 (C10), 138.97 (C19), 136.61 (C15), 130.29 (C17), 129.73 (C2 & C6), 128.42 (C1), 126.50 (C14), 126.30 (C15), 121.85 (C3 & C5), 117.17 (C20), 116.83 (C18), 14.45 (C22)

LC-MS (method II): t_R = 6.77 min, purity: 99%, m/z = 312.000 (M+H$^+$)

HRMS: (M+H$^+$): expected: 312.0801 m/z, found: 312.0802 m/z



4-Methoxyphenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate (61)

Exact Mass: 341.0834 Molecular Weight: 341.3850

4-Methoxyphenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate **61** was synthesised from 4-methoxyphenyl propiolate **74** (85.3 mg, 0.48 mmol) according to the general procedure A. 4-Methoxyphenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate **61** was isolated as a grey solid (28.7 mg, 29.0% yield).

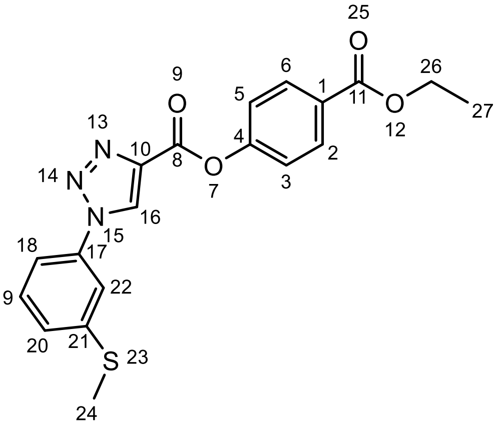
^{1}H NMR (DMSO-*d6*, 500 MHz): δ 9.80 (1H, s, 16H), 7.86 (1H, s, H22), 7.79 (1H, d, *J* = 7.9 Hz, H20), 7.56 (1H, t, *J* = 8.1 Hz, H19), 7.43 (1H, d, *J* = 7.9 Hz, H18), 7.24 (2H, d, *J* = 8.9 Hz, H3 & H5), 7.03 (2H, d, *J* = 9.2 Hz, H2 & H6), 2.59 (3H, s, H24)

^{13}C NMR (DMSO-*d6*, 125 MHz): δ 159.02 (C8), 157.18 (C1), 143.26 (C4), 141.03 (C10), 139.04 (C21), 136.62 (C17), 130.29 (C19), 128.32 (C18), 126.43 (C16), 122.71 (C3 & C5), 117.04 (C22), 116.74 (C20), 114.62 (C2 & C6), 55.46 (C12), 14.40 (C24)

LC-MS (method II): t_R = 6.80 min, m/z = 342.150 (M+H$^+$)

HPLC (method I): t_R = 21.828 min, purity 98%

HRMS: (M+H$^+$): expected: 342.0907 m/z, found: 342.0913 m/z



4-(Ethoxycarbonyl)phenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate (62)

Exact Mass: 383.0940 Molecular Weight: 383.4220

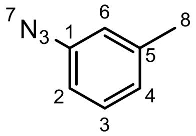
4-(Ethoxycarbonyl)phenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate **59** was synthesised from ethyl 4-(propioloyloxy)benzoate **75** (85.9 mg, 0.39 mmol) according to the general procedure A. The crude product was purified by FCC (eluent: *n*-Hexane:EtOAc= 8:2) and 4-(Ethoxycarbonyl)phenyl 1-(3-(methylthio)phenyl)-1*H*-1,2,3-triazole-4-carboxylate **62** isolated as a pale yellow crystals (40.0 mg, 26.5% yield).

^{1}H NMR (DMSO-*d6*, 500 MHz): δ 9.83 (1H, s, H16), 8.09 (2H, d, *J* = 8.6 Hz, H2 & H6), 7.86 (1H, t, *J* = 1.8 Hz, H22), 7.79 (1H, d, *J* = 7.8 Hz, H20), 7.57 (1H, t, *J* = 8.1 Hz, H19), 7.50 (2H, d, *J* = 8.8 Hz, H3 & H5), 7.44 (1H, d, *J* = 7.6 Hz, H18), 4.34 (2H, q, *J* = 7.1 Hz, H26), 2.57 (3H, s, H24), 1.34 (3H, t, *J* = 7.1 Hz, H27)

^{13}C NMR (DMSO-*d6*, 500 MHz): δ 165.06 (C11), 158.24 (C4 or C8), 153.61 (C4 or C8), 141.11 (C10), 138.72 (C21), 136.60 (C17), 130.99 (C2 & C6), 130.37 (C19), 128.72 (C1), 127.96 (C14), 126.54 (C16), 122.43 (C3 & C5), 117.11 (C22), 116.82 (C20), 61.00 (C26), 14.44 (C24), 14.20 (C27)

LC-MS (method II): t_R = 7.13 min, purity ≥88%, m/z = 384.000 (M+H$^+$)

HRMS: (M+H$^+$): expected: 384.1013 m/z, found: 384.1013 m/z



1-Azido-3-methylbenzene (78)

Exact Mass: 133.0640 Molecular Weight: 133.1540

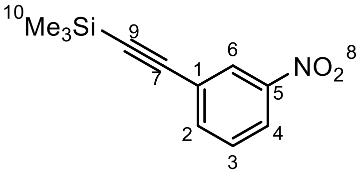
3-Methylanilline **76** (3.247 mL, 3214.8 mg, 30.0 mmol, 1 equiv.) was dissolved in H_2O (84 mL) and the solution stirred over an ice bath. Concentrated aqueous HCl (8.4 mL) was added and consequently, a solution of $NaNO_2$ (2484.0 mg, 36.0 mmol, 1.2 equiv.) in H_2O (18 mL) was added dropwise. After 15 min, a solution of NaN_3 (2925.5 mg, 45.0 mmol, 1.5 equiv.) in H_2O (25 mL) was added dropwise under ice cooling. The solution was stirred and the temperature allowed to increase to room temperature. The reaction was monitored by TLC (*n*-Hex:EtOAc = 3:7). After four hours, the reaction was extracted four times with Et_2O (total volume: 135 mL), the combined organic extracts washed with sat. brine (60 mL), dried over $MgSO_4$ and the solvent removed under reduced pressure to yield 1-azido-3-methylbenzene **78** (3684.0 mg, 27.7 mmol, 92.2% yield) as a brown oil.

^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.23 (1H, t, *J* = 8.2 Hz, H3), 6.95 (1H, d, *J* = 7.6 Hz, H4), 6.84 (2H, m, H6 and H2), 2.35 (3H, s, H8)

^{13}C NMR ($CDCl_3$, 125 MHz): δ 140.03 (C1 or C5), 139.97 (C1 or C5), 129.69 (C3), 125.89 (C4), 119.72 (C6), 116.25 (C2), 21.50 (C8)

LC-MS (method III): t_R = 6.77 min, purity 91%, no molecular ion detected

In accordance with previously reported values by Bertrand et al.[51]



Trimethyl((3-nitrophenyl)ethynyl)silane (81)

Exact Mass: 219.0716 Molecular Weight: 219.3150

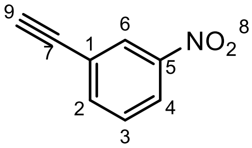
1-Iodo-3-nitrobenzene **79** (3984.2 mg, 16.0 mmol, 1.0 equiv.) was dissolved in anhydrous DMF (8 mL) and Et_2NH (16.7 mL, 11.7 g, 160.0 mmol, 10.0 equiv.) and argon was bubbled through this solution for 20 min. TMS-acetylene **80** (2.46 mL, 1744 mg, 17.8 mmol, 1.1 equiv.), CuI (121.8 mg, 0.64 mmol, 0.04 equiv.) and $Pd(PPh_3)_4$ (203.4 mg, 0.176 mmol, 0.011 equiv.) were added sequentially. The reaction mixture was monitored by TLC (*n*-Hex:EtOAc = 100:3). After 16 h, the reaction mixture was poured into 1 M HCl (48 mL) and extracted three times with DCM (total volume: 70 mL). The combined organic extracts were washed with 1 M HCl (40 mL) and sat. brine (40 mL), dried over $MgSO_4$ and the solvent removed under reduced pressure. The crude product was purified by FCC (*n*-Hex:EtOAc = 30:1) to yield trimethyl((3-nitrophenyl)ethynyl)silane **81** (1463 mg, 6.67 mmol, 41.7% yield).

^{1}H NMR ($CDCl_3$, 500 MHz): δ 8.31 (1H, t, *J* = 1.7 Hz, H6), 8.16 (1H, ddd, *J* = 0.8, 2.2, 8.4 Hz, H4), 7.75 (1H, dt, *J* = 1.1, 7.9 Hz, H2), 7.49 (1H, t, *J* = 7.9 Hz, H3), 0.27 (9H, s, H10)

^{13}C NMR ($CDCl_3$, 125 MHz): δ 148.15 (C5), 137.71 (C4), 129.39 (C3), 126.95 (C6), 125.11 (C1), 123.29 (C2), 102.29 (C7), 97.78 (C9), -0.11 (C10)

LC-MS (method III): t_R = 7.94 min, purity 85%, no molecular ion detected

In accordance with previously reported values by Bertrand et al.[51]



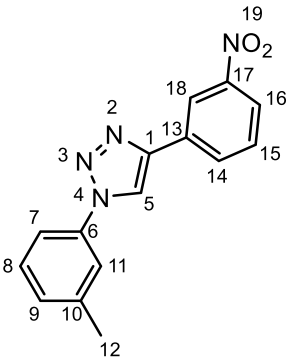
1-Ethynyl-3-nitrobenzene (82)

Exact Mass: 147.0320 Molecular Weight: 147.1330

Trimethyl((3-nitrophenyl)ethynyl)silane **81** (1400.0 mg, 6.38 mmol, 1 equiv.) and K_2CO_3 (1588.1 mg, 12.3 mmol, 1.8 equiv.) were suspended in MeOH (66 mL) and stirred under argon for 4 h. The solution was filtered and the solvent was removed under reduced pressure from the filtrate. The residue was taken up in H_2O (50 mL) and extracted twice with EtOAc (total volume: 70 mL). The combined organic extracts were washed with sat. brine (50 mL) and dried over $MgSO_4$. The solvent was removed under reduced pressure to yield 1-ethynl-3-nitrobenzene **82** as a brown oil (797.2 mg, 5.42 mmol, 85.0 % yield). The crude product was used in the next step without purification.

^{1}H NMR ($CDCl_3$, 500 MHz): δ 8.34 (1H, t, *J* = 1.8 Hz, H6), 8.20 (m, 1H, H4, overlapping with H4 of trimethyl((3-nitrophenyl)ethynyl)silane **81**), 7.79 (1H, dt, *J* = 1.2, 7.7 Hz, H2), 7.52 (3H, t, *J* = 8.0 Hz, H3), 3.22 (1H, s)

In accordance with previously reported values by Bertrand et al.[51]



1-(3-Methylphenyl)-4-(3-nitrophenyl)-1,2,3-triazole (83)

Exact Mass: 280.0960 Molecular Weight: 280.2870

1-Ethynyl-3-nitrobenzene **82** (176.6 mg, 1.20 mmol, 1 equiv.), 1-azido-3-methylbenzene **78** (159.8 mg, 2.30 mmol, 1 equiv.), sodium ascorbate (47.6 mg, 0.24 mmol, 0.2 equiv.) and $CuSO_4 \cdot 5H_2O$ (14.6 mg, 0.06 mmol, 0.05 equiv.) were combined in a mixture of H_2O (9.0 mL) and *t*-BuOH (6.0 mL). The mixture was sealed in a microwave flask and reacted under microwave irradiation at 100 °C for 30 s and at 130 °C for 30 min. The precipitated 1,2,3-triazole product **83** was isolated by filtration as a grey-brown solid (216.9 mg, 0.77 mmol, 64.5% yield) and not further purified.

^{1}H NMR (DMSO-*d6*, 500 MHz): δ 9.59 (1H, s, H5), 8.76 (1H, t, *J* = 1.9 Hz, H18), 8.41 (1H, d, *J* = 7.9 Hz, H16), 8.25 (1H, ddd, *J* = 0.8, 2.2, 8.3 Hz, H14), 7.83 (2H, m, H15 & H11), 7.77 (1H, d, *J* = 8.2 Hz, H7), 7.53 (1H, t, *J* = 7.9 Hz, H8), 7.36 (1H, d, *J* = 7.6 Hz, H9), 2.45 (3H, s, H12)

^{13}C NMR (DMSO-*d6*, 125 MHz): δ 148.44 (C17), 145.29 (C1), 139.79 (C10), 136.43 (C6), 131.98 (C13), 131.39 (C16), 130.79 (C15), 129.84 (C8), 129.57 (C9), 122.83 (C14), 121.02 (C11), 120.44 (C5), 119.64 (C18), 117.13 (C7), 20.96 (C12)

LC-MS (method II): t_R = 7.00 min, purity 98%, m/z = 281.000 ($M+H^+$)

HRMS: ($M+H^+$): expected: 281.1033 m/z, found: 281.1033 m/z

In accordance with previously reported values by Bertrand et al.[51]

2.3 Biological evaluation: competitive fluorescence polarisation assay

The competitive FP assay was performed according to the method reported by Hancock et al.[52] Stock solutions of the inhibitors in DMSO with a concentration of 10 mM (Stock 1) were prepared. Prior to plating, the Stock 1 solutions were diluted 1:100 (2 µL Stock + 200 µL DMSO) to obtain Stock 2 solutions with a concentration of 100 µM.

The Stock 2 solutions were plated onto untreated Corning® black 96 well plates (Sigma, CLS3991) containing a solution of the Keap1 Kelch domain (200 nM) and the fluorescent peptide FITC-β-DEETGEF-OH (1 nM) in Dulbecco's Phosphate Buffered Saline (DPBS) (Sigma, D8537) at pH 7.4 (11% final DMSO concentration, 100 µL final volume, 10 µM inhibitor concentration). After incubating for 150 min under slow agitation at room temperature in the dark, the plates were transferred to a PHERAstar microplate reader (BMG Labtech, Ortenberg, Germany), the fluorescence intensity recorded and the FP calculated according to equation 1: $p = \frac{I_{II} - I_{\perp}}{I_{II} + I_{\perp}}$. All measurements were recorded in triplicate. The baseline FP (fluorescent peptide only) was subtracted and the data was normalised to the control (fluorescent peptide plus Keap1 Kelch domain), corresponding to a full FP response.

2.4 Summary of methodological workflow

Figure 10 gives an overview of the different methods employed, their order and how they interact with each other.

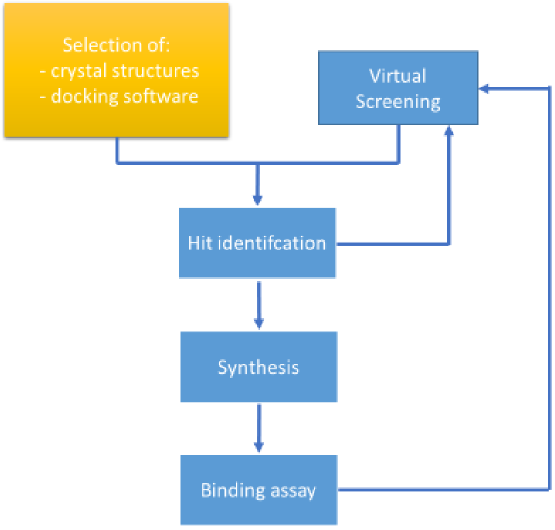


Figure 10: First, virtual screening along with the selection of appropriate Keap1 crystal structures and docking software validation were performed. After five iterations, a suitable hit candidate was found and its synthesis initiated (see: Virtual screening, de novo-scaffold development and fragment synthesis). For the second part of the project (Synthesis, evaluation and virtual structure-activity relationship of 1,2,3-triazoles), a hit had previously been identified. Analogues of this structure were synthesised and tested in a competitive FP binding assay. Subsequently, new rounds of virtual screen were performed and a potentially improved hit compound was identified.

3 Results: Virtual screening, *de novo*-scaffold development and fragment synthesis

In order to find new possible Keap1-Nrf2 PPI inhibitor scaffolds, a series of compound libraries were screened *in silico* to assess their potential interactions with the Keap1 Kelch domain. This approach has the potential to identify new hits comparatively quickly and easily, and has been employed successfully by other groups to target the Keap1-Nrf2 PPI.[86]

Out of 17 available Keap1 structures of human origin, five structures were chosen for the docking studies; this was later refined to three structures. Additionally, four different docking programmes and six associated scoring functions were evaluated for their fitness to predict the rank order of 17 known ligands of Keap1 for which experimental binding affinities and co-crystal structures were available. Two programmes showed promising predictive potential in this context.

In the compound screening stage of the study, first, two libraries from the ZINC database[47], one fragment-, one drug-like, with a total of around 10,000 structures were screened. Based on the results, a refined library was designed comprising 4,140 structures. This resulted in the detection of a preferred scaffold that was further refined by three rounds of screening doing systematic virtual SAR investigations. Compound **39** was found to be the most promising structure based on docking scores, physicochemical properties and synthetic accessibility. The chemical synthesis of its fragments was completed.

3.1 Molecular modelling

3.1.1 Choice of Keap1 crystal structures

To ensure that different Keap1 protein conformations were represented in the screen, the different available structures of the Kelch domain in the PBD database[40] were assessed. As of 11 March 2016, there were 28 structures containing human or murine Keap1, of which 25 contained the Kelch domain. Although human and murine Keap1 are highly homologous, we chose to focus on human Keap1 to avaoid any possible bias towards murine structures, and as there is abundant choice of Keap1 crystal structures, enabling to use rather restrictive filter criteria.

Different crystal structure criteria were assessed for inclusion in our study, in particular the following requirements were evaluated: $RMSD_{C_\alpha}(i) < \overline{RMSD_{C_\alpha}} + SD$ and $Resolution\ (i) < \overline{Resolution} + SD$, where $RMSD_{C_\alpha}$is the root mean square deviation of the C_α-atom positions, the top bar indicates the arithmetic mean of the respective quantity, (i) is a variable for the respective structure and SD is the standard deviation (see Tables 2 and 3 for detailed data).

Generally, it could be observed that the $RMSD_{C_\alpha}$ are lower when calculated for crystal structures that have been cocrystallised with ligands from the same class, namely small molecules, peptides or without ligand (Table 4). To account for these structural differences, at least one structure from each of these three classes was included. Further qualitative criteria were the R_{free} value, clashscore, Ramachandran outliers, side chain outliers, and real space R-value Z-score (RSRZ). The R_{free} value is a measure of the fit of the model to an experimental data subset that has not been used in the structure refinement and is therefore a measure for how close the experimental values are to those in the refined model (lower values are better).[53] The clashscore is based on the number of atom pairs that are unusually close.[53] The Ramachandran plots show the dihedral angles (around the N-C_α and C_α-C') for each amino acid of the protein backbone; only a subset of these angle combinations are favourable.[54] Ramachandran (plot) outliers are the percentage of dihedral angles that are outside of these energetically allowed combinations.[52] Similarly, the sidechain outliers are the percentage of unusual torsion angles of the sidechains.[52] RSRZ is a normalisation in terms of residue and resolution of the RSR, the quality of fit between residues and data in real space.[52] The available structures containing the Keap1 Kelch domain are listed in Table 2, for the $RMSD_{C\alpha}$ values see Tables 3 and 4.

After taking into consideration these criteria (in total five structures, at least one structure out of each ligand class, resolution, $RMSD_{C_\alpha}$, PDB's quality criteria) three structures with cocrystallised small molecules (4xmb[55], 4iqk[56], 3vng[57]), one with a peptide ligand (3zgc[58]) and one without a ligand (1zgk[59]) were chosen for docking studies.

Name	Publication Date	Description	Source organism	Resolution	Quality	Ligand
4xmb	2015/09	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.24	++	Disulfoneamidenaphthalene (2,2'-(naphthalene-1,4-diylbis(((4-methoxyphenyl)sulfonyl)azanediyl)) diacetamide)
4ifj	2014/08	Kelch domain of hKeap (residues 321 - 609)	Human	1.8	++	None (Apostructure)
4l7d	2014/04	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.25	-	(1S,2R)-2-{[(1S)-5-methyl-1-[(1-oxo-1,3-dihydro-2H-isoindol-2-yl)methyl]-3,4-dihydroisoquinolin-2(1H)-yl]carbonyl}cyclohexanecarboxylic acid
4l7c	2014/04	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.40	-	2-{[(1S)-2-{[(1R,2S)-2-(1H-tetrazol-5-yl)cyclohexyl]carbonyl}-1,2,3,4-tetrahydroisoquinolin-1-yl]methyl}-1H-isoindole-1,3(2H)-dione
4l7b	2014/04	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.41	++	(1S,2R)-2-{[(1S)-1-[(1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)methyl]-3,4-dihydroisoquinolin-2(1H)-yl]carbonyl}cyclohexanecarboxylic acid
4n1b	2014/04	Kelch domain of hKeap1 (residues 321 - 611)	Human	2.55	+	(1S,2R)-2-[(1S)-1-[(1-oxo-2,3-dihydro-1H-isoindol-2-Yl)methyl]-1,2,3,4-tetrahydroisoquinoline-2-Carbonyl]cyclohexane-1-carboxylic acid
4ifl	2013/12	Kelch domain of hKeap1 (residues 321 - 609)	Human	1.8	++	None (Apostructure)
4ifn	2013/12	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.4	--	(1R,2R)-2-{[(1S)-1-[(1,3-dioxo-1,3-dihydro- 2H-isoindol-2-yl)methyl]-3,4-dihydroisoquinolin- 2(1H)-yl]carbonyl}cyclohexanecarboxylic acid

3zgd	2013/06	Kelch domain of hKeap1 (mutant: E540A/E542A, residues 321 - 609)	Human	1.98	++	None (Apostructure)
3zgc	2013/06	Kelch domain of hKeap1 (mutant: E540A/E542A, residues 321 - 609)	Human	2.2	++	Nrf2-derived cyclic peptide
4iqk	2013/05	Kelch domain of hKeap1 (mutant: D349N, E540A, E542A, residues 321 - 609)	Human	1.97	++	Disulfoneamidenaphthalene (N,N'-naphthalene-1,4-diylbis(4-methoxybenzenesulfonamide))/cpd 16
4in4	2013/05	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.59	0	Disulfoneamidenaphthalene (2-({5-[(2,4-dimethylphenyl)sulfonyl]-6-oxo- 1,6-dihydropyrimidin-2-yl}sulfanyl)-N-[2-(trifluoromethyl)phenyl]acetamide)/cpd 15
3vng	2013/01	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.1	++	2-(3-((3-(5-(furan-2-yl)-1,3,4-oxadiazol-2-yl)ureido)methyl)phenoxy)acetic acid
3vnh	2013/01	Kelch domain of hKeap1 (residues 321 - 609)	Human	2.1	++	2-(3-((3-(5-(furan-2-yl)-1,3,4-oxadiazol-2-yl)ureido)methyl)phenoxy)acetic acid
2flu	2006/08	Kelch domain of hKeap1	Human	1.5	+	ETGE-containing peptide (Nrf2-derived, 16mer)
1zgk	2005/10	Kelch domain of hKeap1	Human	1.35	++	None (Apostructure)
1u6d	2004/10	Kelch domain of hKeap1	Human	1.85	++	None (Apostructure)
5cgj	2015/10	mKeap1 (residues 309 - 624)	Murine	3.36	+	Sulfoneamide-pyrrolidine-naphthalene ((3S)-1-(4-{[(2,3,5,6-tetramethylphenyl)sulfonyl]amino} naphthalen-1-yl)pyrrolidine-3-carboxylic acid)
3wn7	2013/12	mKeap1 (residues 321 - 609)	Murine	1.57	+	Long DLG-containing peptide: DLGex(Met17-Gln51)

3wdz	2013/09	mKeap1 (residues 321 - 609)	Murine	2.6	+	phosphorylated p62(=Nucleoporine)-peptide (aa 346 - 359)
3ade	2010/03	mKeap1 (residues 309 - 624)	Murine	2.8	++	p62-KIR peptide (residues 346 - 359)
2z32	2008/03	mKeap1 (residues 309 - 607)	Murine	1,9	-	Prothymosin α-peptide (residues 39 - 54)
2dyh	2007/09	mKeap1-DC	Murine	1.9	-	DLG-containing peptide from mNrf2 (residues Ile-22 to Val-36)
1x2j	2006/03	Kelch domain of mKeap1 (residues 309 - 624)	Murine	1.6	-	None (Apostructure)
1x2r	2006/03	Kelch domain of mKeap1 (residues 309 - 624)	Murine	1.7	+	ETGE-containing peptide (Leu-76 to Leu-84)

Table 2: available Keap1 crystal structures in the PDB as of 11 March 2016. The quality assignment refers to a qualitative evaluation of the quality parameters provided by PDB (s. text) where the assignments were made as follow: ++ = very superior, + = superior, 0 = average, - = inferior, -- = very inferior. It should be noted that this assignment has been made not as an absolute rating, but relative to this group of structures. All proteins have been expressed in *E. coli*.

		Small molecules										Peptides			Apostructures			
		4xmb	**4l7d**	**4l7c**	**4l7b**	**4n1b**	**4ifn**	**4in4**	**4iqk**	**3vng**	**3vnh**	**2flu**	**3zgc**	**4ifl**	**4ifj**	**3zgd**	**1zgk**	**1u6d**
Small molecules	**4xmb**	0	0.412	0.462	0.442	0.457	0.392	0.349	0.320	0.323	0.375	0.388	0.331	0.422	0.407	0.330	0.373	0.379
	4l7d	0.412	0	0.186	0.267	0.288	0.407	0.397	0.334	0.442	0.472	0.468	0.359	0.527	0.541	0.300	0.499	0.499
	4l7c	0.462	0.186	0	0.247	0.357	0.418	0.432	0.305	0.451	0.463	0.473	0.347	0.532	0.508	0.307	0.500	0.451
	4l7b	0.442	0.267	0.247	0	0.294	0.348	0.412	0.338	0.420	0.438	0.508	0.318	0.570	0.509	0.290	0.456	0.438
	4n1b	0.457	0.288	0.357	0.294	0	0.402	0.426	0.350	0.482	0.498	0.527	0.384	0.567	0.571	0.336	0.542	0.511
	4ifn	0.392	0.407	0.418	0.348	0.402	0	0.410	0.303	0.394	0.421	0.455	0.369	0.500	0.465	0.336	0.419	0.423
	4in4	0.349	0.397	0.432	0.412	0.426	0.410	0	0.321	0.383	0.388	0.392	0.393	0.431	0.431	0.314	0.406	0.398
	4iqk	0.320	0.334	0.305	0.338	0.350	0.303	0.321	0	0.347	0.359	0.386	0.262	0.432	0.402	0.254	0.346	0.332
	3vng	0.323	0.442	0.451	0.420	0.482	0.394	0.383	0.347	0	0.414	0.365	0.367	0.409	0.410	0.324	0.389	0.398
	3vnh	0.375	0.472	0.463	0.438	0.498	0.421	0.388	0.359	0.414	0	0.488	0.410	0.511	0.283	0.381	0.263	0.243
Peptides	**2flu**	0.388	0.468	0.473	0.508	0.527	0.455	0.392	0.386	0.365	0.488	0	0.417	0.180	0.495	0.382	0.454	0.458
	3zgc	0.331	0.359	0.347	0.318	0.384	0.369	0.393	0.262	0.367	0.410	0.417	0	0.440	0.438	0.204	0.392	0.371
	4ifl	0.422	0.527	0.532	0.570	0.567	0.500	0.431	0.432	0.409	0.511	0.180	0.440	0	0.495	0.416	0.474	0.475
Apostructures	**4ifj**	0.407	0.541	0.508	0.509	0.571	0.465	0.431	0.402	0.410	0.283	0.495	0.438	0.495	0	0.425	0.217	0.206
	3zgd	0.330	0.300	0.307	0.290	0.336	0.336	0.314	0.254	0.324	0.381	0.382	0.204	0.416	0.425	0	0.389	0.383
	1zgk	0.373	0.499	0.500	0.456	0.542	0.419	0.406	0.346	0.389	0.263	0.454	0.392	0.474	0.217	0.389	0	0.134
	1u6d	0.379	0.499	0.451	0.438	0.511	0.423	0.398	0.332	0.398	0.243	0.458	0.371	0.475	0.206	0.383	0.134	0
Mean		0.385	0.425	0.400	0.402	0.393	0.437	0.461	0.404	0.336	0.363	0.337	0.393	0.395	0.400	0.427	0.391	0.381
SD of Mean		0.045	0.104	0.100	0.097	0.093	0.092	0.089	0.047	0.057	0.060	0.045	0.035	0.042	0.079	0.080	0.105	0.103

Table 3: $RMSD_{C_\alpha}$ of the 17 human Keap1 crystal structures available from PDB. Colour code: red/inferior: $RMSD_{C_\alpha}(i) > \overline{RMSD_{C_\alpha}} + SD$, yellow/average: $\overline{RMSD_{C_\alpha}} - SD \leq RMSD_{C_\alpha}(i) \leq \overline{RMSD_{C_\alpha}} + SD$, green/superior: $RMSD_{C_\alpha}(i) < \overline{RMSD_{C_\alpha}} - SD$,values in Å.

	Total	Superior	Average	Inferior
Peptide/Peptide	6	33%	33%	33%
Peptide/small molecule	25	17%	37%	47%
Peptide/Apo	9	8%	33%	58%
Small molecule/Apo	34	35%	30%	35%
Small molecule/small molecule	90	36%	32%	22%
Apo/Apo	12	50%	50%	0%

Table 4: Comparison of crystal structure clusters. The protein structures were grouped according to the cocrystallised ligand type. The $RMSD_{C_\alpha}$ between a structure and the other structures within the same group were assessed (s. description of table 3). The number of $RMSD_{C_\alpha}$ values within each quality range was expressed as percentage of the total number of structure comparisons within each groups' comparison.

3.1.2 Virtual Screening

3.1.2.1 First screening: large library of fragments and drug-like molecules

As they have already been used in our group and are generally some of the most widely used docking programs, we docked the small molecules using DOCK[43] version 6.7, and Autodock Vina[45], version 1.1.2. In addition to the docking programme's own scoring function, CyScore[60] and DSX[61] scoring functions were applied to the poses docked by Autodock Vina[45].

In order to find new potential scaffolds, we decided to screen first two different, diverse libraries that were obtained from the ZINC database[47]: one of fragments with a molecular weight of 100 – 350 and a net charge of -1, and another one with a molecular weight of 350 - 450 and a net charge of -1. The -1 charge was fixed as most of the available data suggests that one negative charge is essential, mimicking one of the Glu residues in the ETGE sequence, being the high affinity sequence of Nrf2's Neh2 domain binding to Keap1. The two different molecular sizes were used to evaluate the contributions of ligand efficiency and complexity on the predicted binding affinity. Both libraries contained around 5000 molecules. Figure 11 summarises the workflow of this virtual screen.

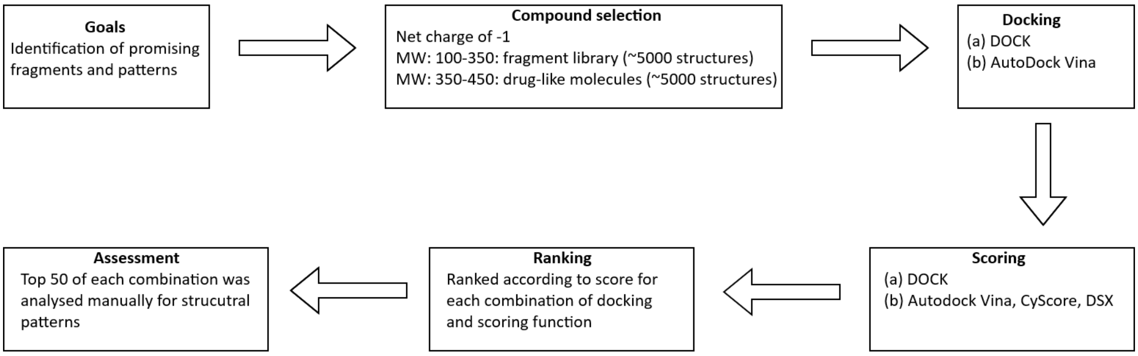


Figure 11: The workflow of this virtual screening consisting of goal definition, compound selection, docking with subsequent scoring. Consequently, the structures were ranked according to their scores and assessed visually for docking poses and structural redundancies or similarities.

From the fragment screen, the general scaffold of compounds with high scores appeared to follow a scheme in which a lipophilic group is linked to a hydrophilic moiety followed by a linker group of 2 – 7 atoms to an acidic group (typically a carboxylic acid) (Figure 12A). Scaffold **15**, based upon this core structure (Figure 12B), was considered as a potential lead due to the appealing three-dimensional structure, relatively easy synthetic accessibility and variability as different amines can be used to form the amide.

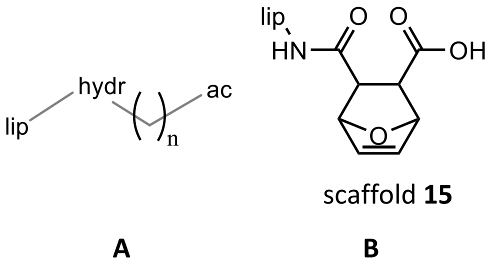
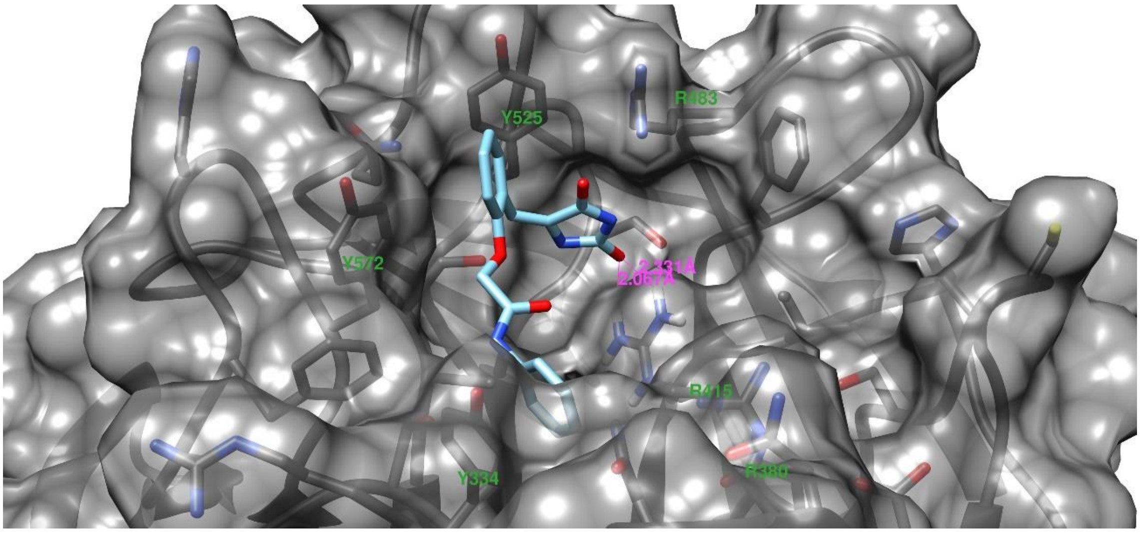
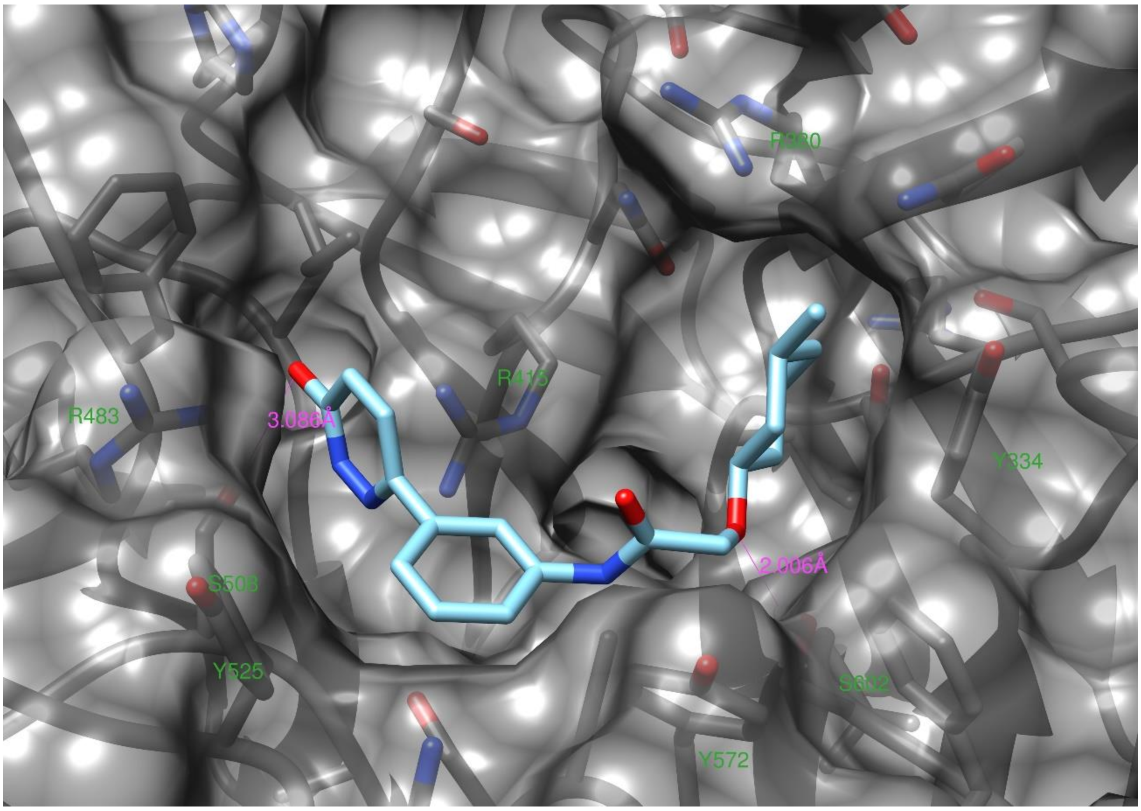


Figure 12: A: general connectivity of highly scored scaffolds. lip: lipophilic group (usually cyclic), hydr: hydrophilic group (mostly amide), n: number of atoms in linker (between 2 and 7, but mostly 4 to 6 & may contain cyclic moieties), ac: acidic group (most frequently a carboxylic acid) B**:** scaffold **15** as an example of the general structure.

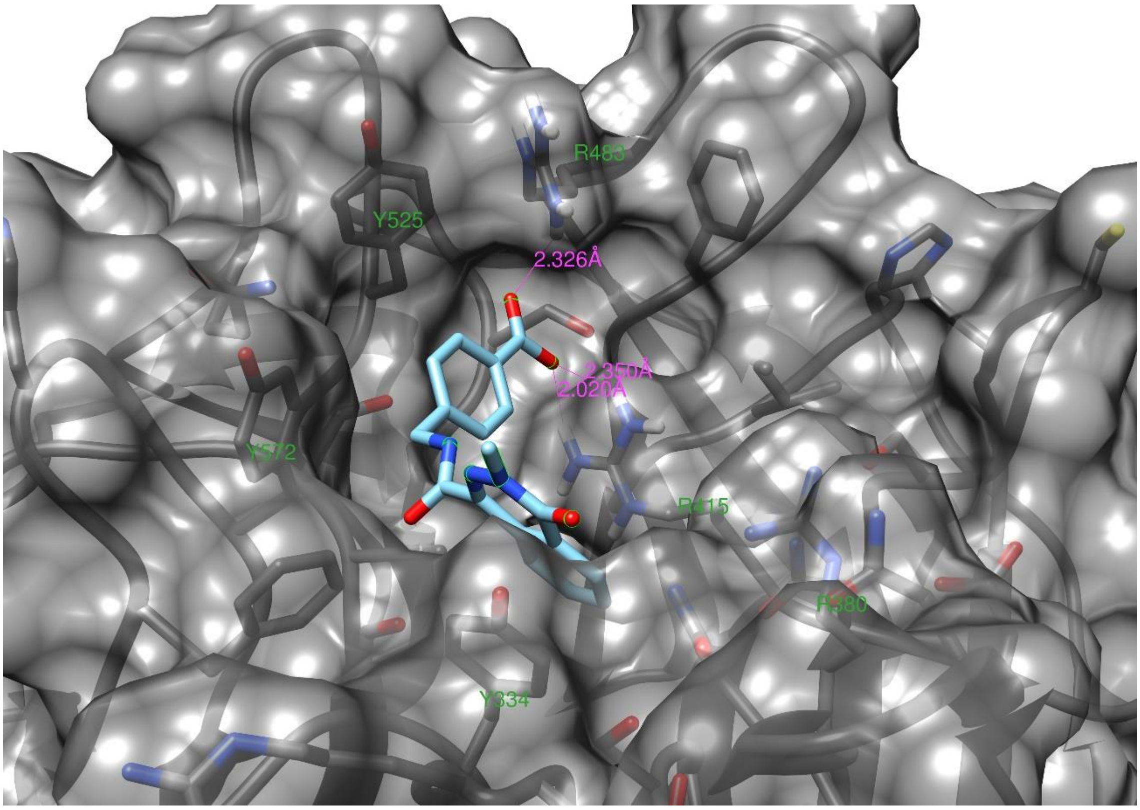
Representative examples of the interaction of highly ranked (9th best, on average on rank 339 ± 559 over the five protein structures and both programmes; and best, rank 227 ± 557) compounds from the fragment library are ZINC00316575 and ZINC71902874. Both fit nicely in the top part of the binding cavity and have hydrogen bonds to the Arg dyad. ZINC08313872 was the third best structure for Vina when docked to 3zgc and show a similar interaction pattern (Figure 13). Table 5 gives an overview of important physicochemical properties of representative structures and the respective interacting residues of Keap1.


Y525
R483
Y572
2.331Å
2.067Å
R415
R380
Y334

A


R380
R415
R483
3.086Å
Y334
2.006Å
S508
Y525
S602
Y572

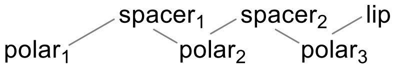
B



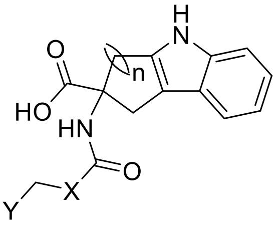
C

Figure 13: Interaction of ZINC00316575 (A), ZINC81902874 (B) and ZINC08313872 (C) (light blue) with the Keap1 Kelch domain from 3zgc (grey). Both ligands are slightly acidic ZINC00316575 (A): cpK$_a$ = 8.23 (hydantoin), and ZINC71902874 (B): cpK$_a$ = 10.34 (pyrazinone)[62], which might increase the strength of the hydrogen bonds (HB) of these groups in an *in vitro* or *in vivo* setup, as they might be deprotonated in the binding pocket, whereas ZINC08313872 (C) has an anionic carboxylate. **A**: ZINC00316575 forms two hydrogen bonds to R415 with its hydantoin moiety and the phenyl π-stacks with the Y334 side chain on the left-hand side. **B**: ZINC71902874 forms one hydrogen bond each with S508 and S602 on the opposite side of the pocket. The adjacent methylcyclohexyl moiety occupies the space between S602 and R380. **C**: Like ZINC00316575 (A), ZINC08313872 forms two hydrogen bonds to R415. Additionally, it forms a hydrogen bond to R483, and the phtalazinone moiety π-stacks with Y334.

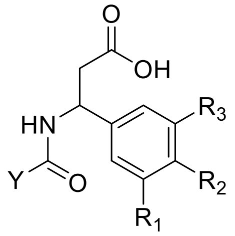
From the screen of the compounds with a molecular weight of 350 – 450, the structures had a similar composition to the smaller molecules, although were more complex (Figure 14). The interaction of a 3-amido-3-phenylpropionic acid-type compound (scaffold **17**) with Keap1 is illustrated in Figure 15. Table 5 gives an overview of important physicochemical properties of representative structures and the respective interacting residues of Keap1.



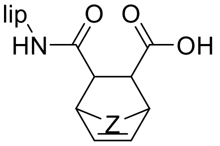
polar1: mostly acidic, thereof mostly carboxylate
spacer1: length 1 – 5, mostly 2 – 4
polar2: mostly amide
spacer2: length 1 – 4
polar3: only present sometimes
lip: bulky lipophilic



scaffold **16**
n = 0,1
X = NH, O
Y: mostly amide linked, bi– or tricycles



scaffold **17**
R_1, R_2, R_3: CN, OH, H, methylenedioxyl



scaffold **18**
Z = CH_2, O

Figure 14: general pattern and lead scaffolds **16** - **18** resulting from the screen of the library with a molecular weight of 350 – 450

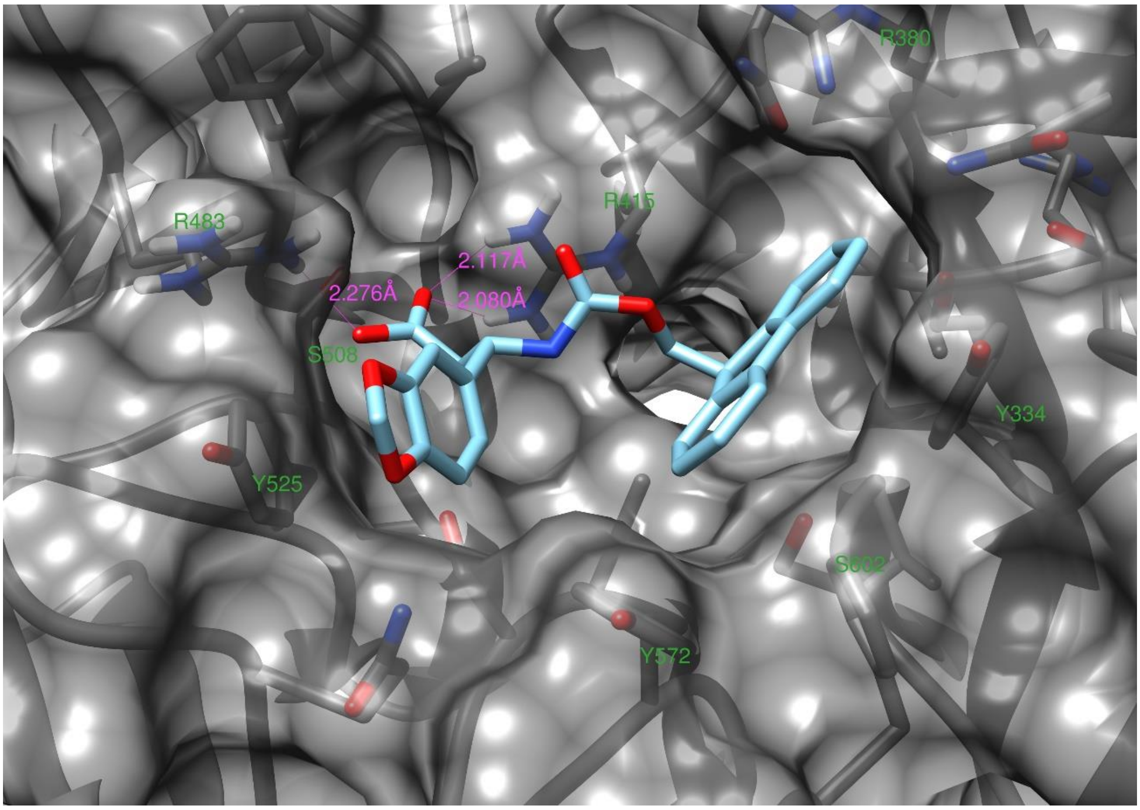


Figure 15: Interaction of ZINC02598929 (light blue) with the Keap1 Kelch domain from 3zgc (grey). The carboxylate forms three electrostatic interactions with the Arg dyad (two with R415, one with R483), while the benzodioxole has a sandwich π-stacking interaction with the Tyr525 sidechain and the fluorene ring has a T-shaped π-stacking interaction with a the Tyr334 sidechain. ZINC02598929 had the fifth best overall docking score for the Keap1 structures when docked with AutoDock Vina[45] (rank: 358 ± 501, the best compound having a rank of 235 ± 356).

Name	HBA	HBD	tPSA (in Å²)	Interacting residues of Keap1
ZINC00316575	7	1	98[§]	Y334, R415
ZINC71902874	6	1	68[*]	S508, S602
ZINC08313872	6	1	67[§]	Y334 R415, R483
ZINC02598929	7	1	96[§]	Y334, R415, R483, Y525

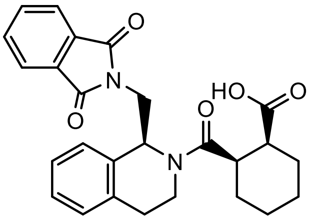
Table 5: Selected physicochemical properties of representative, docked molecules. All Os and Ns were considered HBA and all hydrogens bound to O or N counted as HBD, as suggested in the original publication by Lipinksi *et al.*[87]. tPSA was either obtained from the zinc database[47] (§) or calculated with ChemDraw 15.1 (*), and is given for the ionisation state of the compound that was docked.

The fact compounds that perform well overall have high ranks (and errors) show that individual docking and scoring functions have their own peculiarities and might give very different results for the same dataset. Additionally, it emphasises the challenge associated with finding suitable measures that discriminate between efficient and inefficient virtual ligands. At this point of our studies, the ligands were ranked according to their predicted binding energy, and these ranks were used for further assessment. This approach has two distinct advantages and disadvantages: on one hand, this measure is independent of the absolute numerical value attributed to the binding energy (which can differ 10-fold between different docking functions). On the other hand, it has the disadvantage that relatively small differences in binding energy might lead to a big change of rank, especially if it concerns molecules within middle-ranking group of structures which tend to have similar predicted efficiencies. In order to avoid this shortcoming, another type of measure was used in later in the project.

3.1.2.2 Second screening: Evaluation of Docking programmes

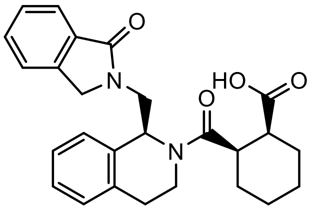
After conducting the first screening exercise, a study was published by Wang et al.[39] demonstrating that the programmes rDock[42] and LeDock[44] have a high docking power, and a high docking and scoring power, respectively. Consequently, these two programmes were implemented in the following screens, in addition to DOCK[43] and Vina[45]. However, to maintain the computing time at an acceptable level, we wanted to evaluate which programmes and which protein structures to use for the further studies as it is known that different programmes can perform very differently depending on the protein structure of interest.[39]

In this evaluation, the four different docking programmes and the five protein structures were tested against 13 known small molecule ligands (Figure 16) of Keap1 for which experimental affinities (K_D or IC_{50}) and crystal structures while bound to Keap1 were available, in order to determine if the programmes could reproduce the correct order of the molecules' binding affinities. The predicted binding poses were analysed visually to confirm that they are realistic. In addition to implementing the new programmes themselves, the scoring functions of Vina[45], CyScore[60] and DSX[61] were applied on the docked structures to explore suitable combinations of docking software and scoring algorithm. The quantified measure of the performance was the Spearman rank correlation coefficient between the experimental and computed (predicted) ranking of binding affinity. Figure 17 summarises the workflow. Table 6 gives an overview of important physicochemical properties of the structures and the interacting residues of Keap1.



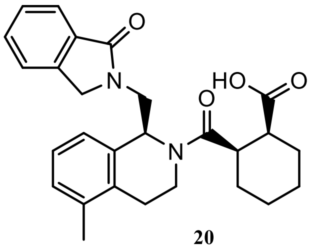
10

IC$_{50}$ = 2.3 µM[85]



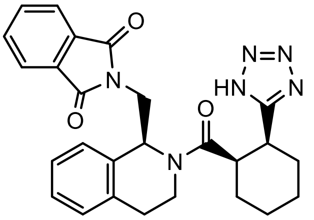
19

IC$_{50}$ = 1.1 µM[85]



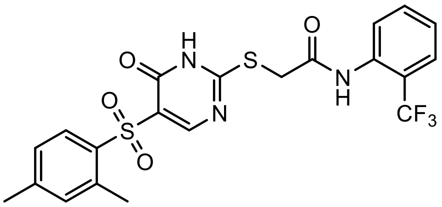
20

IC$_{50}$ = 0.75 µM[85]



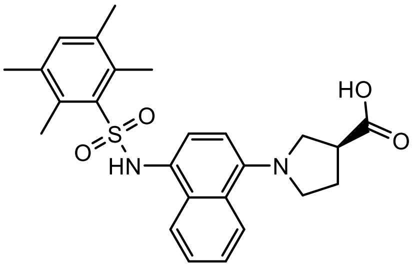
21

IC$_{50}$ = 7.4 µM[85]



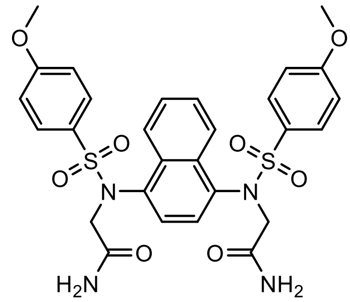
22

IC$_{50}$ = 118 µM[56]



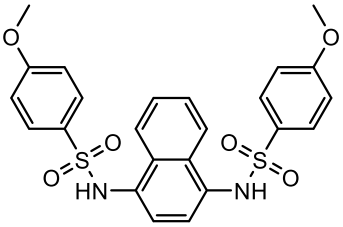
13

IC$_{50}$ = 0.14 µM[33]



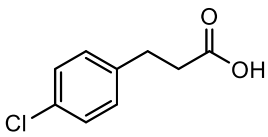
23

IC$_{50}$ = 0.063 µM[55]



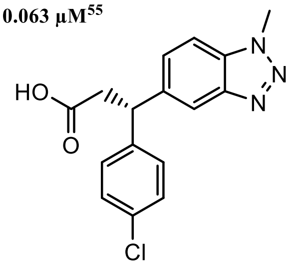
24

IC$_{50}$ = 2.7 µM[56]



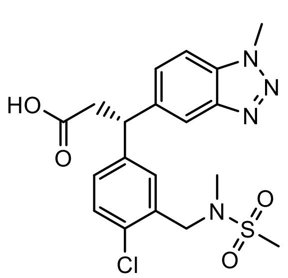
25

IC$_{50}$ = > 1000 µM[33]



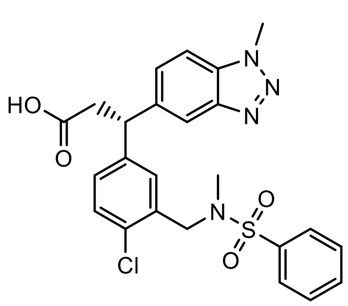
26

IC$_{50}$ = 59 µM[33]



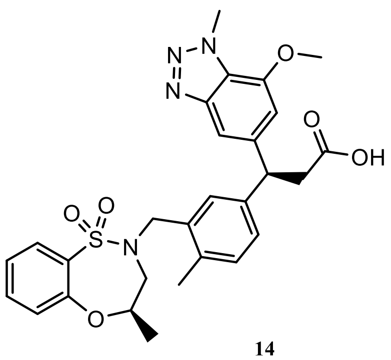
27

IC$_{50}$ = 4.0 µM[33]



28

IC$_{50}$ = 0.67 µM[33]



14

IC$_{50}$ = 0.0013 µM[33]

Figure 16: Structures of known ligands **10**, **13**, **14**, **19** – **28** for which experimentally determined binding affinities and crystal structures are available. These have been docked on five Keap1 crystal structures to assess the docking and scoring power of four docking programmes (DOCK, AutoDock Vina, LeDock, rDock).

Name	HBA	HBD	tPSA (in Å^2)	Interacting residues of Keap1
10	7	0	95	R415, R483
13	6	1	87	Y525
14	10	0	121	R415, R483, Y525
19	6	0	78	Y334
20	6	0	78	N414, R415, R483
21	9	0	107	Y334
22	7	2	105	R415
23	12	4	179	R415
24	8	2	111	R415, R483
25	0	0	37	N414
26	5	0	65	Y334
27	8	0	103	R483
28	8	0	103	R415

Table 6: Selected physicochemical properties of representative, docked molecules. All Os and Ns were considered HBA and all hydrogen bound to O or N counted as HBD, as suggested in the original publication by Lipinksi *et al.*[87]. tPSA was calculated with ChemDraw 15.1, and is given for the ionisation state of the compound that was docked.

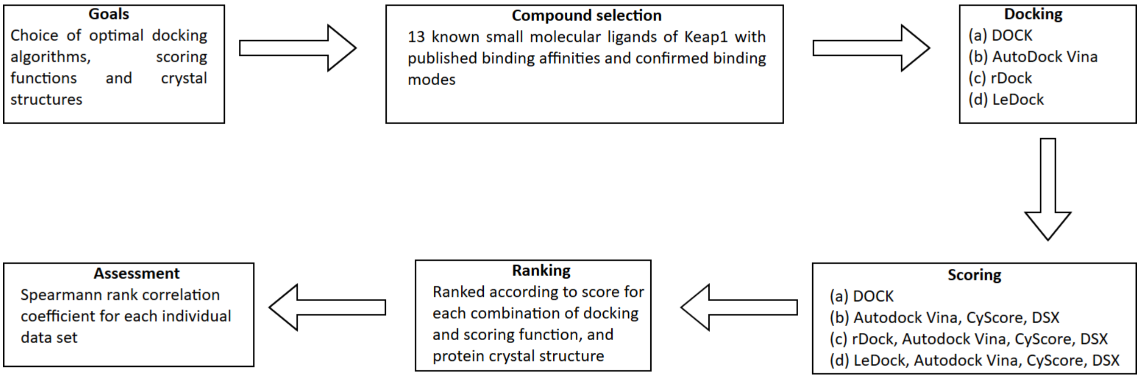


Figure 17: Workflow for the second screen. 13 previously Keap1 ligands with published binding mode and affinity have been used to evaluate the four docking software packages and five Keap1 crystal structures.

From the four examined docking programmes, DOCK[43] performed consistently badly (correlation coefficient on every protein structure < 0.5) which is in general accordance with the review by Wang et al.[39]. In contrast to the review we found that scoring with the Vina scoring function performed the least well after docking with Vina, rDock and LeDock. One further, important insight was that DSX score is a very promising scoring function as it scored best after Vina docking, second best after rDock docking and best after LeDock docking.

When the results were assessed for each protein crystal structure individually, the results were less clear. However, DOCK performed the least well of the four programmes (four times worst, one time worst but one).

Furthermore, when the correlation coefficients of the four programmes (ranks of consensus scoring vs. experimental affinity) were averaged for each protein structure, it could be observed that 4iqk and 3vng gave less good results (correlation coefficients of 0.352 and 0.429, respectively) compared to 1zgk, 3zgc and 4xmb (correlation coefficients between 0.569 and 0.694) (see Tables 7 and 8 for further details).

As a result of this evaluation and of the study by Wang *et al.*[39], we decided to use only LeDock and Autodock Vina as the docking software, and the 1zgk, 3zgc and 4xmb Keap1 Kelch domain crystal structures.

Docking programme	Scoring function	Experimental Rank Correlation
Total	Total	0.552
Vina	Total	0.618
	Vina	0.442
	CyScore	0.596
	DSX	0.785
DOCK	DOCK	0.116
rDock	Total	0.499
	rDock	0.513
	CyScore	0.363
	DSX	0.481
	Vina	0.301
LeDock	Total	0.525
	LeDock	0.538
	CyScore	0.446
	DSX	0.557
	Vina	0.363
All	Vina	0.323
	CyScore	0.512
	DSX	0.719

Table 7: Correlation coefficients between different docking and scoring functions as well as the average of a scoring function, and the experimental rank, over all protein structures.

Protein structure	Docking programme	Experimental Rank Correlation
Total	Total	0.605
1zgk	Total	0.569
	Vina	0.447
	DOCK	0.285
	rDock	0.631
	leDock	0.449
3vng	Total	0.429
	Vina	0.494
	DOCK	-0.749
	rDock	0.414
	leDock	0.696
3zgc	Total	0.613
	Vina	0.725
	DOCK	0.335
	rDock	0.245
	leDock	0.577
4iqk	Total	0.352
	Vina	0.577
	DOCK	0.006
	rDock	0.271
	leDock	0.336
4xmb	Total	0.694
	Vina	0.703
	DOCK	0.478
	rDock	0.574
	leDock	0.504

Table 8: Correlation coefficients between different docking programmes and their consensus, and the experimental rank for each of the five crystal structures as well as over all five structures. If one docked pose has been scored with more than one scoring function (the case for Vina, rDock and LeDock), the values have been averaged over these different scoring functions.

Colour code for both tables:
green: correlation $\geq 0{,}7$;
yellow: $0{,}6 \leq correlation < 0{,}7$;
orange: $0{,}5 \leq correlation < 0{,}6$;
red: $correlation < 0{,}5$

3.1.2.3 Third screening: refined library

Based on the results of the first screen, a new library, library1, comprising of 4,140 compounds from the ZINC database[47], was created. The designed scaffolds **29** and **30** were based on the general scaffolds **16** – **18** in terms of linker length and functional groups observed in the first two screens (fragments and lead-like) and that were present in many structures. Furthermore, similarity with the most promising scaffolds and some individual, high scoring structures were subjective selection criteria that were applied. Structural similarity was defined by a Tanimoto coefficient, a measure of the similarity between two bit sequences, of ≥ 0.6 (Figure 18).

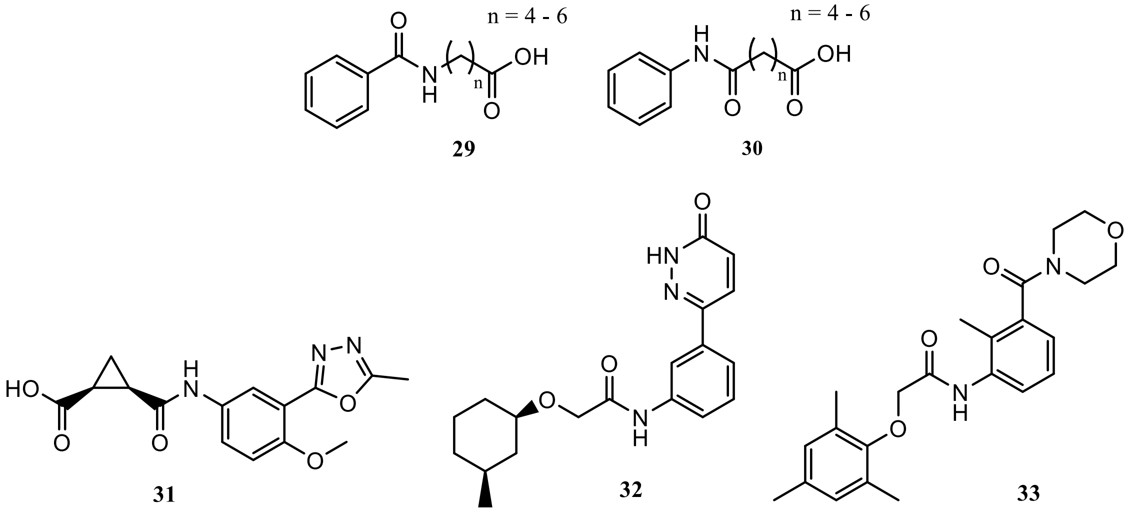


Figure 18: Structures used for the assembly of library1. From the ZINC database, all compounds with scaffold **29** or **30** and n = 4-6 were obtained. Additionally, similar structures to **29** and **30** for n = 5, and similar structures to **31**, **32** and **33** were selected.

Following the assembly of the library, the compounds were docked using AutoDock Vina and LeDock and scored using the AutoDock Vina, CyScore, DSX, and LeDock, Autodock Vina, CyScore, DSX scoring functions, respectively. The structures were ranked according to their computed affinity, the average of all ranks calculated and the best 100 compounds analysed visually for their binding poses, interactions and structural patterns (Figure 19).

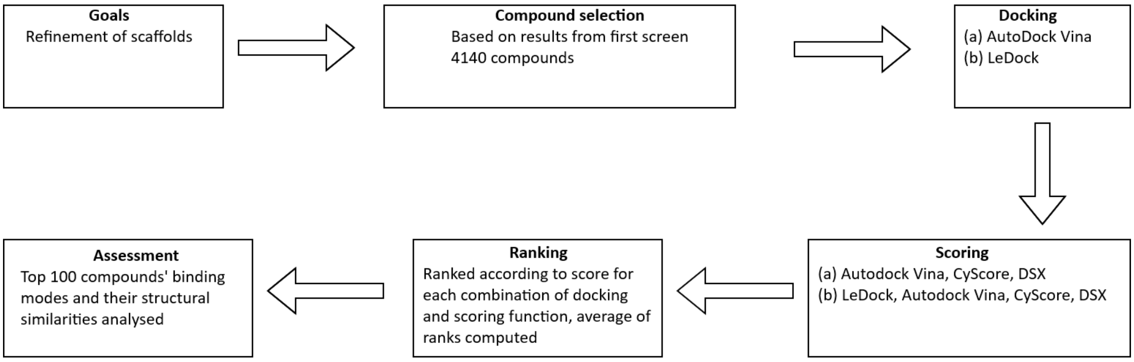
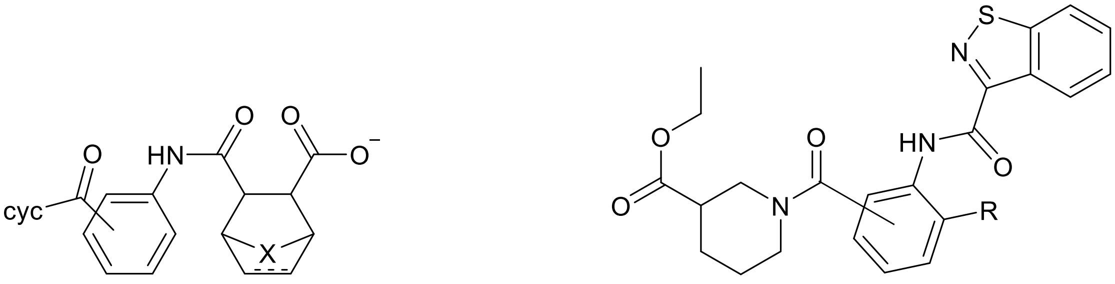


Figure 19: Workflow of the third screen. The compounds were docked using AutoDock Vina and LeDock, scored with three or four, respectively, different scoring functions and the best 100 compounds, according to their average rank, were analysed thoroughly.

The results from the screening of library1 suggested two main scaffolds, scaffolds **34** and **35**; see Figures 20 and 21 for structures and representative examples of compounds docked with Keap1, respectively. Important physicochemical properties of representative examples as well as the interacting residues of Keap1 are given in Table 9. In terms of potential target selectivity, the benzoisothiazole moiety appeared promising as it was predicted to form at least two hydrogen bonds, mostly with the Arg dyad of Keap1 and hydrogen bonding is generally considered as being important for selectivity in biomolecular interactions.



Scaffold **34**
X = CH_2CH_2, CH_2, O
cyc: cycle, meta- or para-linked

Scaffold **35**
R = H, Me
part on left hand: ortho or meta to amide

Figure 20: Best performing scaffolds of the screen of library1

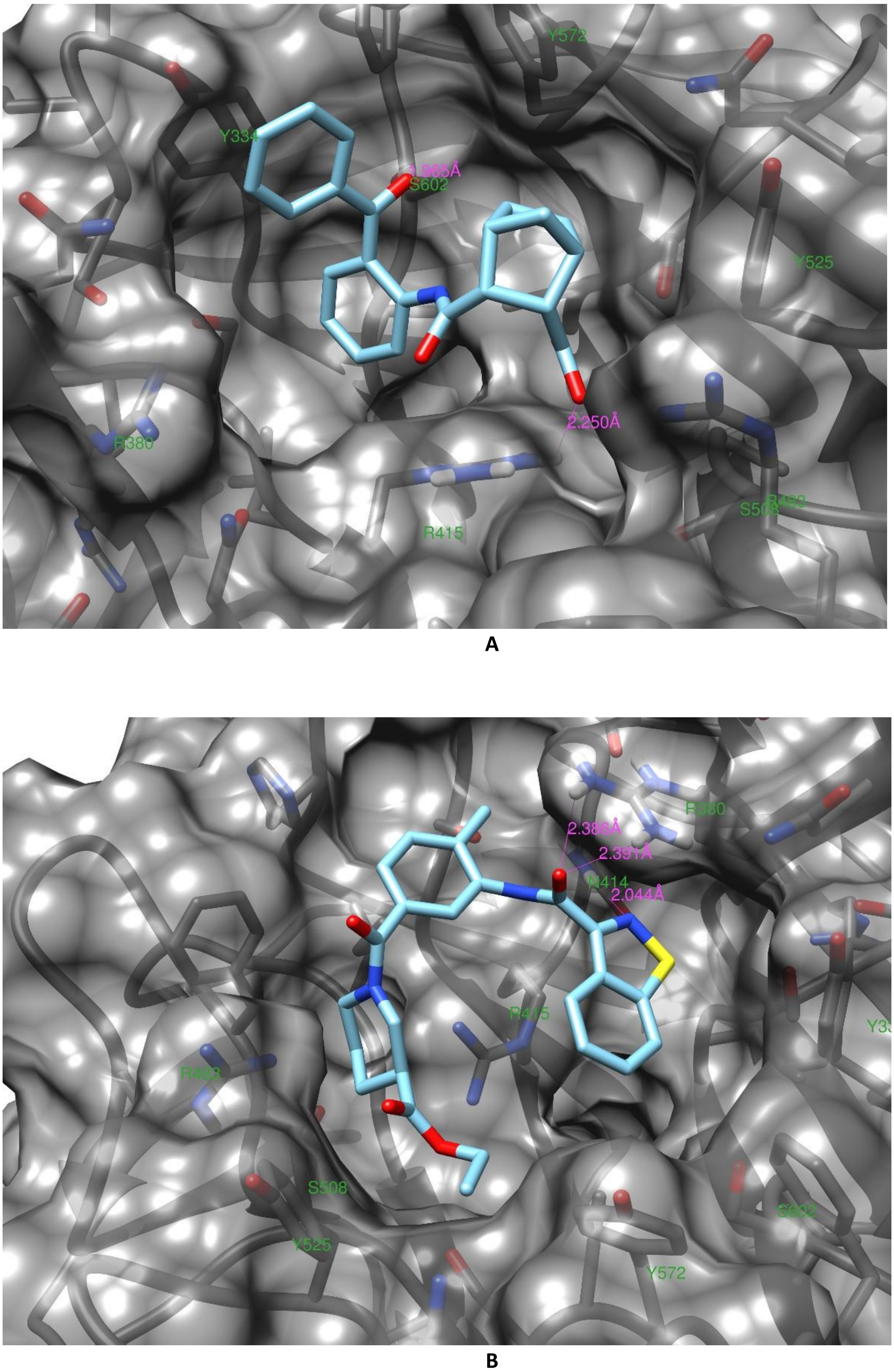


Figure 21: Interaction of ZINC035913019 (A) and ZINC008600061 (B) (light blue) with the Keap1 Kelch domain from 3zgc (grey). **A**: The carboxylate forms an electrostatic interaction with R415 of the Arg dyad on the right-hand side of the pocket, while the terminal phenyl ring forms a sandwich π-stacking interaction with Y334. Additionally, the ketone forms a hydrogen bond with S602. **B**: The benzoisothiazole forms one hydrogen bond with N414. Furthermore, the compound also forms two hydrogen bonds with R380 and occupies the space between R380 and Y572.

Name	HBA	HBD	tPSA (in Å²)	Interacting residues of Keap1
ZINC035913019	5	1	86	Y334, R415, S602
ZINC008600061	7	1	88	R380, N414

Table 9: Selected physicochemical properties of representative, docked molecules. All Os and Ns were considered HBA and all hydrogen bound to O or N counted as HBD, as suggested in the original publication by Lipinksi *et al.*[87]. tPSA was either obtained from the zinc database[47].

3.1.2.4 Fourth screening: systematic variation

In order to establish whether the benzoisothiazole moiety is essential for the potential binding activity, the activity of other synthetically accessible heterocycles was investigated. Benzofuran-2-yl, furan-2-yl, indol-2-yl, indol-3-yl, and 5-methyl-1,3-thiazol-4-yl containing structures were investigated and compared to the 1,2-benzoisothiazol-3-yl analogues. Whereas for scaffold **34**, the bicyclic structure was attractive in terms of spatial filling of the binding pocket as well as accessibility, because if X = O and the double bond is present, the bicycle can be synthesised from a Diels-Alder reaction of furan and maleic anhydride, followed by opening of the anhydride to form an amide. Depending on the amine used for this final step, a diverse set of amides might be created (Figure 22).

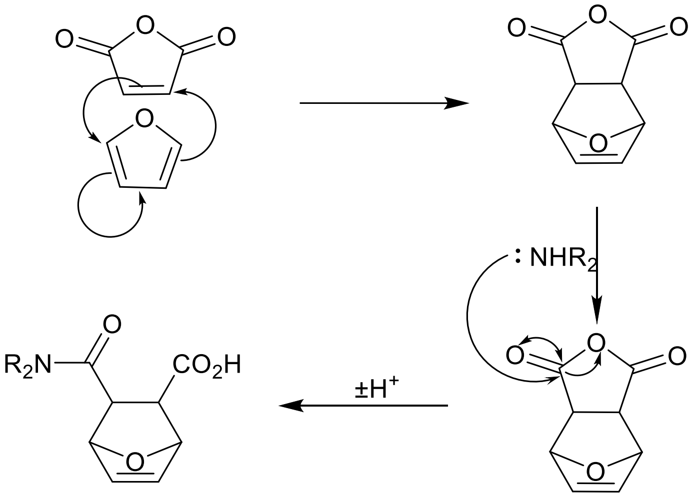


Figure 22: Scheme of the Diels-Alder reaction of maleic anhydride with furan and subsequent ring opening of the anhydride with an amine. Reaction with a diverse set of amines would allow to form a variety of amides.

As different connectivity (*ortho*-, *meta*- or *para*-disubstituted) on the central phenyl ring on scaffolds **34** and **35** appeared to be similarly high scoring within one structural group, another point of exploration was to systematically evaluate the *ortho*, *meta* and *para* versions of each molecule. For scaffold **34**, the preferred relative positions of the two groups are *meta* or *para*, whereas for **35**, they are *ortho* or *meta*. A reason behind this might be that the general binding conformation of both groups was different, as molecules with the structure of **35** tended to be closer to both Arg residues in the dyad and therefore prefer a more extended arrangement.

In order to determine the linker length between the central rings and the rings that are attached to them via the carbonyl group (on the left side in Figure 20), we explored a range of linkers. Several variations to the central ring were made, as so far only phenyl rings had been explored, but many known potent inhibitors contain a bicyclic naphthalene. Finally, as only amide linkages were considered in the previous studies, we investigated ether linkers between the central ring and the left ring (according to Figure 20) to examine less hydrophilic linkers and include some form of potential selectivity for the synthesis. The library was labelled library2 and contained 492 molecules. For an overview of the changes, see Figure 23. Figure 24 shows the workflow.

scaffold **36**

A

B

Figure 23: A: scaffold **36**: structural properties of library2 compounds. **A** = 1,2-, 1,3- or 1,4-linked (relative positions) phenyl, phenoxymethylphenyl, benzofuran, indole; n = 0 – 2; Y = (cyc)-CH_2O-(A), (cyc)-(C=O)-NH-(A); cyc: 1,2-benzoisothiazol-3-yl, benzofuran-2-yl, furan-2-yl, indol-2-yl, indol-3-yl, 5-methyl-1,3-thiazol-4-yl; **B**: representative example of structure with scaffold **36**: A = 1,2-linked benzofuran, n = 0, Y = CH_2O, cyc: furan-2-yl

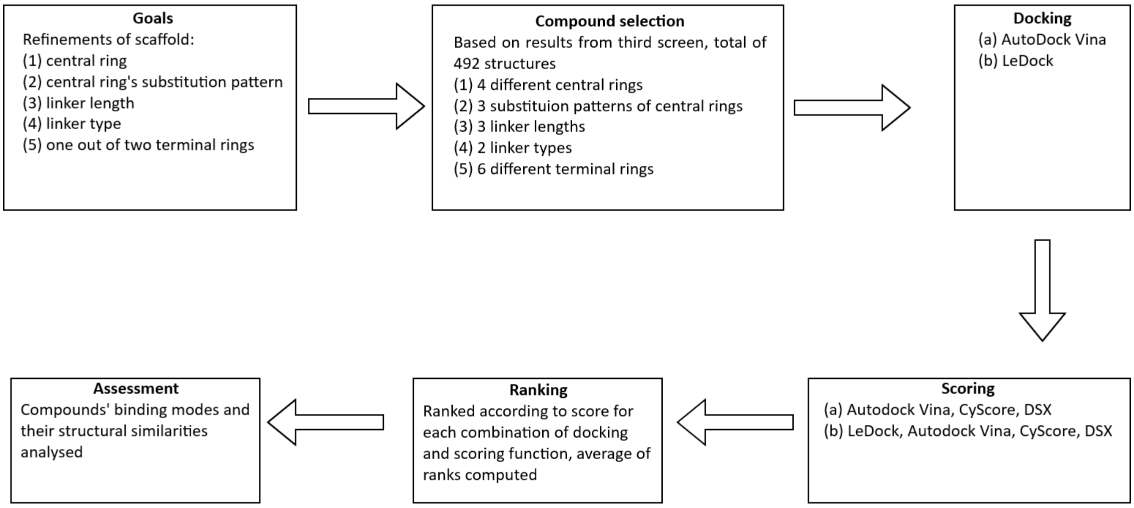


Figure 24: Workflow of the fourth screen. The compounds were docked using AutoDock Vina and LeDock, scored with three or four, respectively, different scoring functions and the compounds were ranked according to their average rank and analysed visually and for preferred structural motifs.

In terms of heterocycles, no clear trend was observable for the score or predicted ligand efficiencies with an increasing size of the "cyc" moiety. The amide linkage was always preferred over an ether, which was also the case for shorter linkers (n = 0 or 1) as well as *meta*- and *para*-substituted central A rings. There was no clear preference for which A ring was preferred.

Interestingly, in contrast to the previous screens for many of the molecules, it is not the carboxylic acid (if there was one present) that interacts with the Arg dyad, but the aromatic heterocycle. A possible explanation to this might be that in addition to the hydrogen bonds, which both groups can form, the heterocycle can form T-shaped π-cation interactions with one of two Arg (either Arg483 or Arg380). Usually this occurred with Arg483 as it is more exposed on the surface, in contrast to Arg380 which is located on the opposite side of the pocket and more withdrawn from the surface. Consequently, these properties tended to increase the predicted binding energy of the series.

Two examples of high ranking structures are represented in Figure 25.

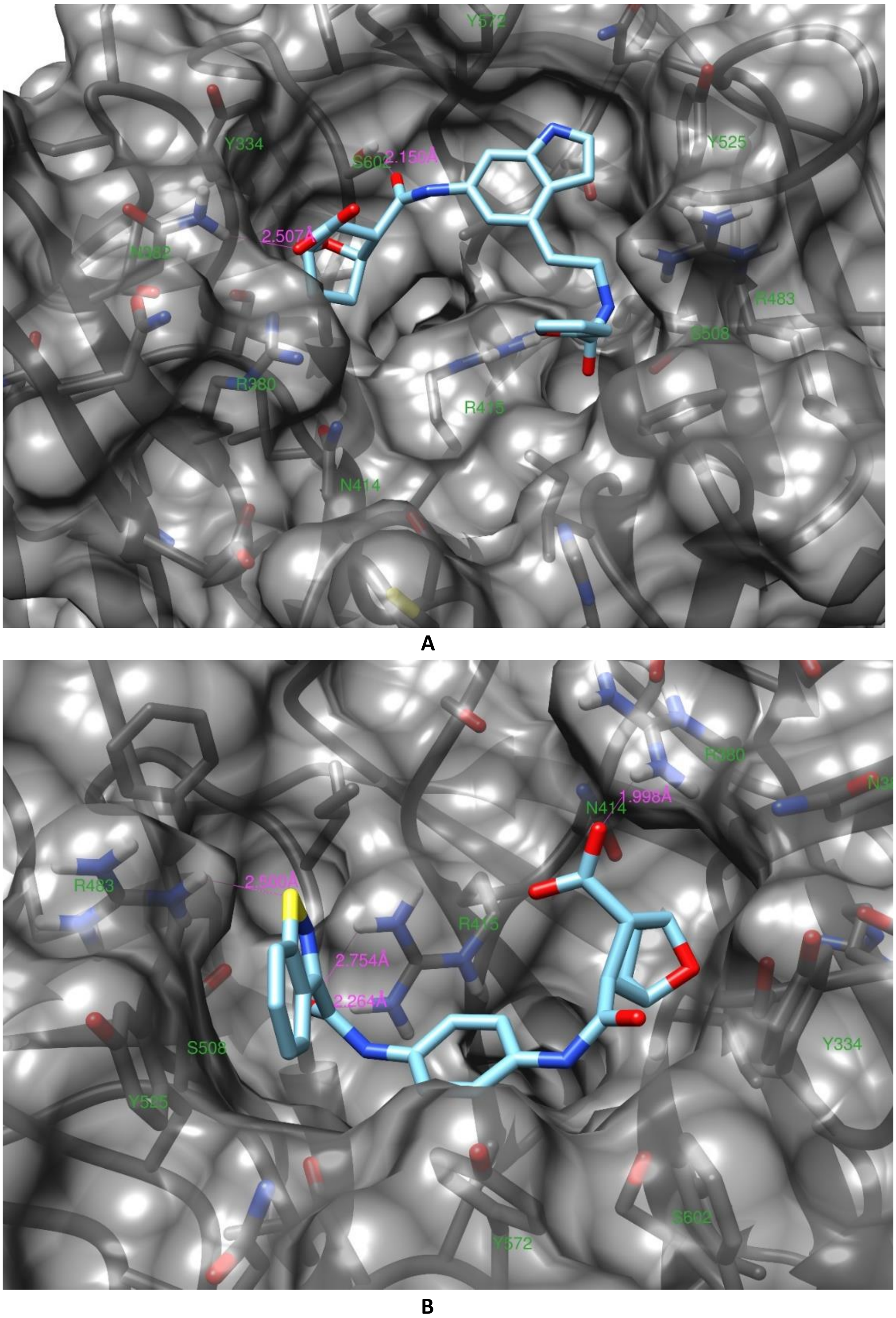


Figure 25: Interaction of 13-1-cin-sfu-a (A) and 14-0-cph-sbi-a (B) (light blue) with the Keap1 Kelch domain from 3zgc (grey). **A**: The carboxylate forms an electrostatic interaction with N382 which is located on the opposite to the Arg dyad. The furan moiety fills the pocket between R415 and R483, potentially forming π-cation interactions with the Arg dyad. Additionally, the carbonyl of the amide forms a hydrogen bond to S602. **B**: The carboxylate forms an electrostatic interaction with R380. The benzoisothiazole moiety forms an HB with an Arg of the dyad, R483, and the adjacent carbonyl groups forms two hydrogen bonds to R415.

3.1.2.5 Fifth screening: structure optimisation

A disadvantage of the compounds screened so far was that despite them showing promising binding poses and scores, they were all relatively large (molecular weight ~450) and lipophilic, with the highest scoring compounds having a clogP of 4 - 4.5. Although these properties might be suitable for a drug molecule[63], they were considered to be too high for an *in silico* hit candidate which would require further optimisation. Therefore, a further library of compounds, library3, was constructed with the aim of reducing the size and lipophilicity of the molecules while maintaining the calculated binding affinity.

Therefore, library3 had a similar scaffold to library2 with some elements remaining unchanged. The central ring was fixed to phenyl because larger aromatic systems did not generally have higher scores or facilitate specific interactions (*e.g.* HB, π-stacking) in the Keap1 binding pocket. All possible connectivities (*ortho*, *meta*, *para*) of the phenyl ring substituents were explored. Additionally, the investigated rings for the “cyc” part, on the left side, were expanded to include additional smaller heterocycles (namely triazoles) and the stereochemistry of the right-handed Diels-Alder product ring was fixed to *endo*, as this was predicted to be slightly more potent. The library contained 144 structures. For details, see Figure 26. Figure 27 summarises the workflow.

scaffold **37**

A

B

Figure 26: A: Scaffold **37** of library3; A = 1,2-, 1,3- or 1,4-linked phenyl; n = 0 – 2; Y = (cyc-)CH_2O-(A), (cyc-)CONH-(A); cyc: 1,2-benzoisothiazol-3-yl, benzofurane-2-yl, furane-2-yl, indole-2-yl, indole-3-yl, 5-methyl-1,3-thiazol-4-yl, 1H-1,2,3-triazo-4-yl, 1H-1,2,4-triazo-3-yl. B: representative example of structure with scaffold **37**: A = 1,2-linked phenyl, n = 0, Y = CONH, cyc: 1,2-benzoisthiazol-3-yl

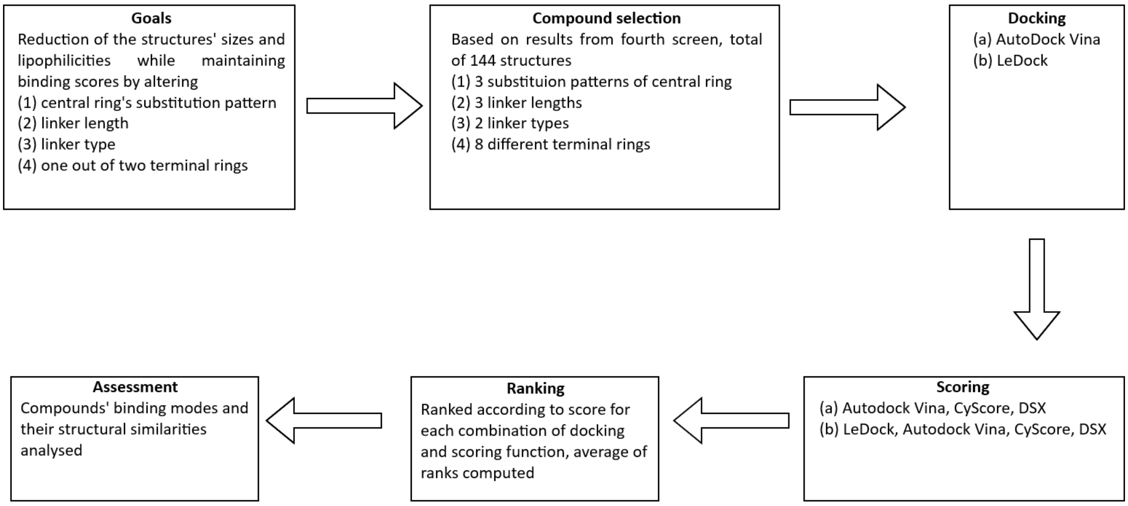


Figure 27: Workflow of the fifth screen. The compounds were docked using AutoDock Vina and LeDock, scored with three or four, respectively, different scoring functions and the compounds were ranked according to their average rank and analysed visually and for preferred structural motifs.

The docking results showed again that the amide linkage was preferred over an ether in terms of score, but longer linker lengths were preferred (*i.e.* n = 1 or 2 *vs* n = 0). The triazole containing compounds did not show the highest score; nevertheless, promising predicted affinities were achieved with scores less than -9 kcal/mol in Autodock Vina. This was potentially interesting considering the structures are relatively small (molecular weights between 350 and 400). Since the hydrophilic contacts are usually considered to be central for selectivity, the lipophilic efficiency, calculated as LipE = pIC_{50} – LogP, is a good metric for this and one of the best measures to identify early, promising hit series as well as later stage leads.[77] The group of triazole containing structures showed the highest ligand efficiencies with particularly high lipophilic efficiencies, typically higher than 8, qualifying them as interesting starting points for structure-based drug design (Tables 10 and 11).

In addition, the 1,2,4-triazoles substituted with an acyl group at the 3- or 5-position, have the interesting property that their pK_as are close to physiologic pH. Therefore, this heterocycle was incorporated into the first compound that were synthesised. This latter property is intriguing because generally, at least one acidic group appears to be required for effectiveness in most of peptidic or small molecular Nrf2 Keap1 PPI inhibitors (Figure 15). This might be linked to mimicking glutamates 79 and 82 of the ^{79}ETGE^{82} sequence of Nrf2, as it could be shown that replacing one of these residues results in a reduced binding affinity of peptides based on that sequence. Furthermore, *N*-alkylation of the sulphonamides of structure **24** with two acetic acid residues resulted in a 100-fold increase in activity, with both carboxylates interacting with Arg415.[21,48] However, an increase in the number of acidic groups entails a higher charge and therefore reduces the ability of the compound to diffuse through the cytoplasmic membrane, reducing the activity in cells. With a pK_a near the physiologic pH, roughly half of the molecules will have a protonated (uncharged) triazole, which facilitates cellular uptake, whereas the deprotonated form could have a higher binding affinity for Keap1.

	Score	**SD**	**LE**	**SD**	**Lip E**	**SD**
14-2-cph-sbf-a	-10.2	0.9	0.309	0.028	6.514	0.911
14-2-cph-s3i-a	-10.2	0.7	0.308	0.022	6.757	0.726
14-0-cph-s3i-a	-10.2	0.2	0.328	0.008	6.773	0.235
14-2-cph-s2i-a	-10.1	0.8	0.307	0.026	6.701	0.845
13-0-cph-s3i-a	-10.1	0.3	0.325	0.009	6.695	0.291
14-2-cph-sbi-a	-10.0	0.9	0.303	0.027	6.440	0.895
13-2-cph-s3i-a	-10.0	0.5	0.302	0.015	6.557	0.492
14-0-cph-s2i-e	-9.9	0.4	0.329	0.013	6.230	0.384
13-1-cph-s3i-a	-9.9	0.5	0.308	0.016	6.477	0.518
13-2-cph-sbf-a	-9.8	0.5	0.298	0.015	6.147	0.485
13-1-cph-sbf-a	-9.8	0.5	0.307	0.017	6.179	0.540
13-2-cph-s2i-a	-9.8	0.7	0.297	0.022	6.390	0.715
13-1-cph-sbi-a	-9.8	0.4	0.315	0.018	6.260	0.354
14-1-cph-sbf-e	-9.8	0.6	0.305	0.011	5.727	0.570

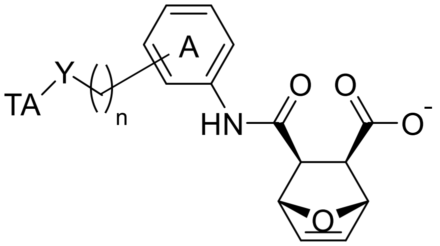
Table 10: Scores, ligand efficiencies (LE) and lipophilic efficiencies (LipE) of all compounds with a score of at least -9.8, using Autodock Vina. Out of these 14 compounds, 7 have a linker length of 2, 4 have a linker length of 1 and 3 have a linker length of 0. The higher predicted affinity produced by the amide instead of the ether is emphasised by the fact that 12 of these molecules contain an amide, and just 2 an ether. The names of the molecules reflect their structure in the following way: XX-n-A-cyc-Y (s. Figure 21) with XX: 12 = 1,2-, 13 = 1,3- or 14 = 1,4-linked phenyl; n: 0 – 2; A: cph = phenyl; Y: e = (cyc-)CH_2O-(A), a = (cyc-)CONH-(A); cyc: sbi =1,2-benzoisothiazol-3-yl, sbf = benzofuran-2-yl, sfu = furan-2-yl, s2i = indole-2-yl, s3i = indole-3-yl, sti = 5-methyl-1,3-thiazol-4-yl, s123ta = 1H-1,2,3-triazo-4-yl, s124ta = 1H-1,2,4-triazo-3-yl; SD: standard deviation

	Score	**SD**	**LE**	**SD**	**Lip E**	**SD**
14-2-cph-s123ta-a	-9.3	0.4	0.322	0.014	8.276	0.415
13-2-cph-s124ta-a	-9.3	0.2	0.322	0.006	8.209	0.166
14-1-cph-s124ta-a	-9.2	0.2	0.330	0.006	8.229	0.174
13-2-cph-s123ta-a	-9.2	0.3	0.317	0.014	8.142	0.409
13-1-cph-s124ta-a	-9.2	0.3	0.328	0.009	8.174	0.252

Table 11: Scores of all triazole containing compounds with a score of at least -9.2, using Autodock Vina. These five compounds contained an amide linkage and a linker length group of 1 or 2.

3.1.2.6 Sixth screening: final optimisation

When maleic anhydride and furan are reacted together the main product is the *exo*-4,10-dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione rather than the *endo*-form. Thus, it was investigated how changing the ring configuration and its saturation affected the predicted binding activity. If the *exo* and unsaturated ring system were predicted to be similarly high scoring, this would be advantageous as *exo*-4,10-dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione is commercially available. The scaffold of the assembled library, labelled library4, is show in Figure 28. The library contained 36 structures.



scaffold **38**

Figure 28: scaffold **38**: structural properties of library4 compounds; A = 1,2-, 1,3- or 1,4-linked phenyl; n = 0 – 2; Y = (cyc)-CH$_2$O-(A), (cyc)-(C=O)-NH-(A); TA: 1*H*-1,2,3-triazo-4-yl, 1*H*-1,2,4-triazo-3-yl

To evaluate this library, the two most promising combinations of docking and scoring functions were of particular interest. Firstly, it is the combination of LeDock[44] docking with the Vina[45] scoring function as these have been found by Wang *et al.*[39] to be generally the best docking and scoring algorithms, respectively, within the examined programmes and therefore, their combination can be expected to perform well. Secondly, the combination of Vina[45] docking and DSX[61] scoring was chosen as this was found to be the best performing combination in our own validation with known Keap1 ligands.

The evaluation was done by adding up all scores each of the *exo* and *endo* analogues and comparing these. No large differences could be observed, with the *exo* compounds seeming to perform moderately better. In order to avoid as far as possible, the individual weaknesses of a docking or scoring function, the consensus rank was used as a final selection criterion (Figure 29). Figure 30 summarises the workflow of this screen.

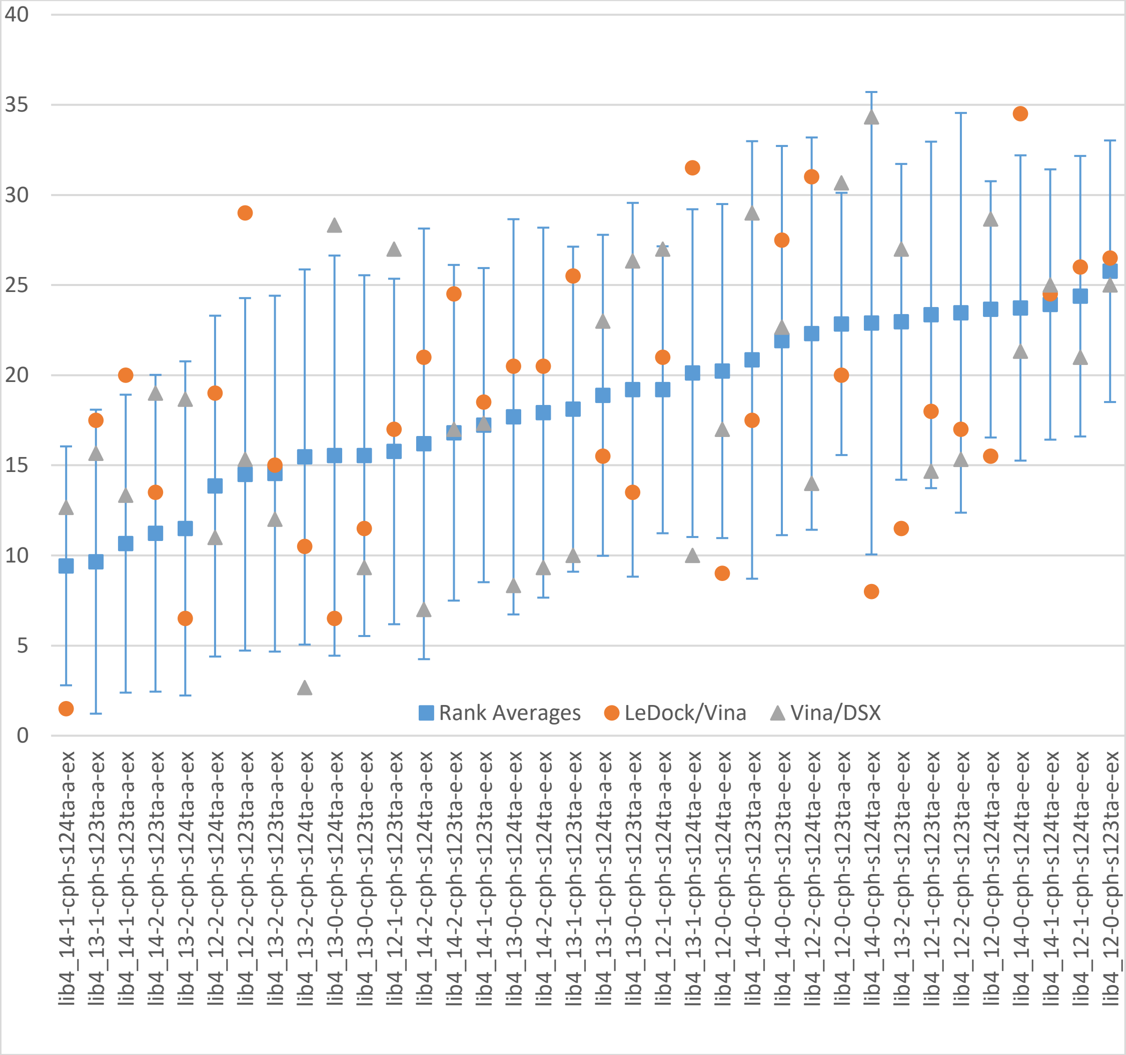


Figure 29: On the horizontal axis, the compounds of library4 are listed. On the vertical axis, the average ranks of the compounds are plotted. Blue squares: total average rank over all protein structures, docking and scoring functions (21 data points). The blue lines display the standard deviation (in both ways) of this average. Orange circles: average rank over all protein structures after LeDock docking and AutoDock Vina scoring (3 data points). Grey triangles: average rank over all protein structures after AutoDock Vina docking and DSXScore scoring (3 data points). The structure names are of the form lib4-A-n-cph-TA-linker type. A: 12, 13, 14 = 1,2-, 1,3- and 1,4-linked phenyl; n: 0, 1, 2 = linker length; TA: s123ta, s124ta = 1,2,3- and 1,2,4-triazole; linker type: e, a = ether, amide.

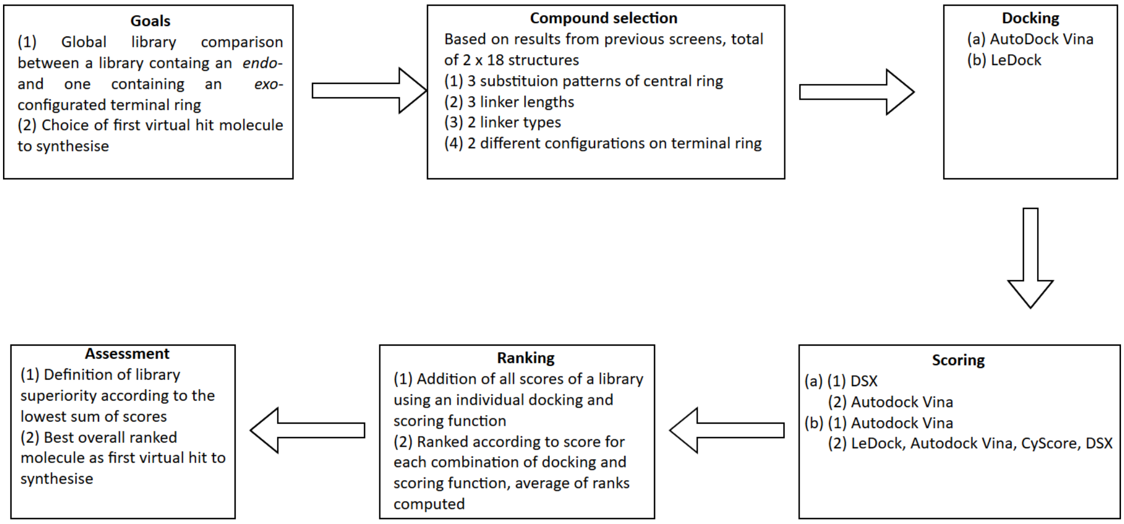


Figure 30: Workflow of the sixth screen. The compounds were docked using AutoDock Vina and LeDock. The global difference between the sub-libraries was assessed by using DSX scoring for AutoDock Vina docking and Autodock Vina scoring for LeDock docking, whereas the whole set of scoring functions was used for the choice of the first molecule to synthesise.

The similarity between the *exo*- and *endo*-compound series is emphasised by the fact that the binding poses are similar. The amide between the phenyl ring and bicycle is rotated by ~90° which enables both cyclic structures to occupy roughly the same space (Figure 31).
Therefore, compound lib4-14-1-cph-s124ta-a-ex **39** (Figure 32) was selected to be the first structure to be synthesised.

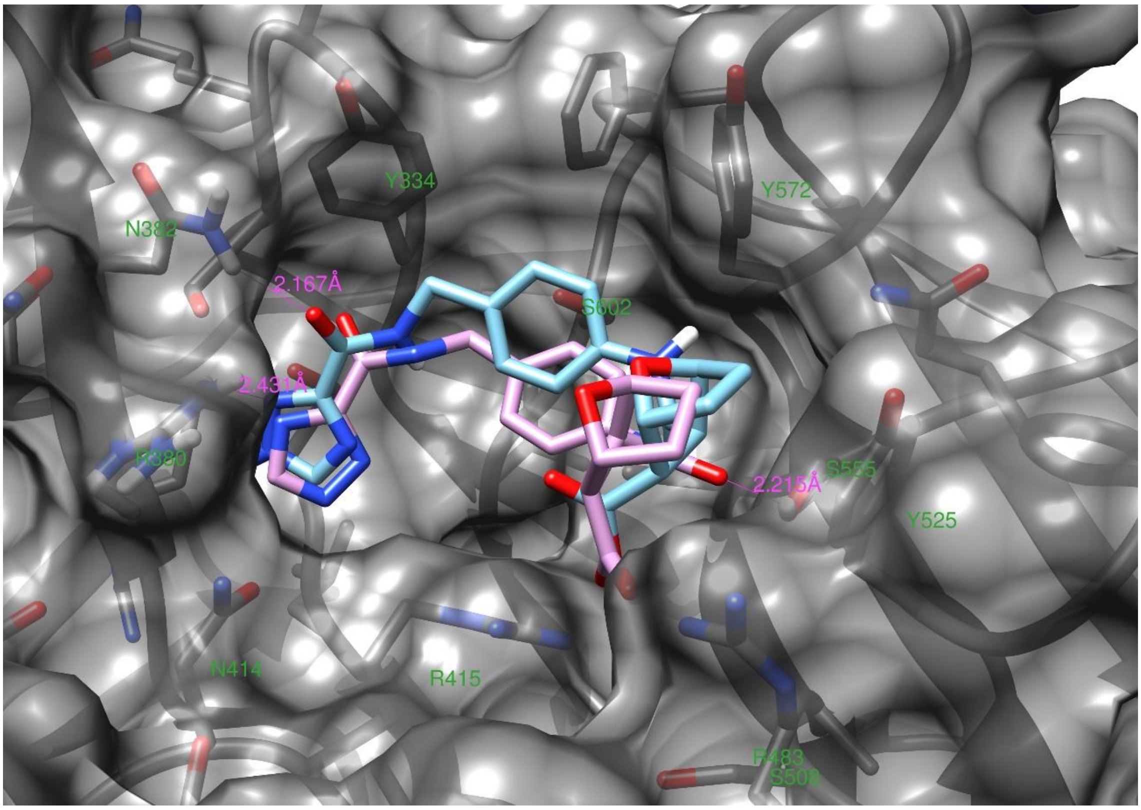


Figure 31: Interaction of the diamides lib4-14-1-cph-s124ta-a-ex **39** (light blue) and the *endo* analogue (purple) with the Keap1 Kelch domain from 3zgc (grey). Both molecules occupy similar space, however the carboxylate of the *endo* compound is orientated more towards the bottom of the binding pocket. Compound **39** forms two hydrogen bond: one via its triazole to R380 and one with the adjacent carbonyl oxygen to N382. On the other hand, the *endo* analogue forms only one hydrogen bond between a carbonyl oxygen and S555.

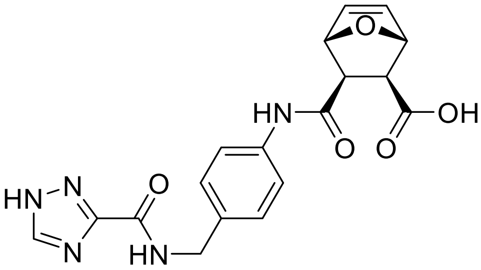


Figure 32: Structure of virtual hit molecule **39**

3.2 Fragment synthesis

The retrosynthetic analysis of compound **39** reveals three different fragments, labelled **40**, **41** and **42** (Figure 33).

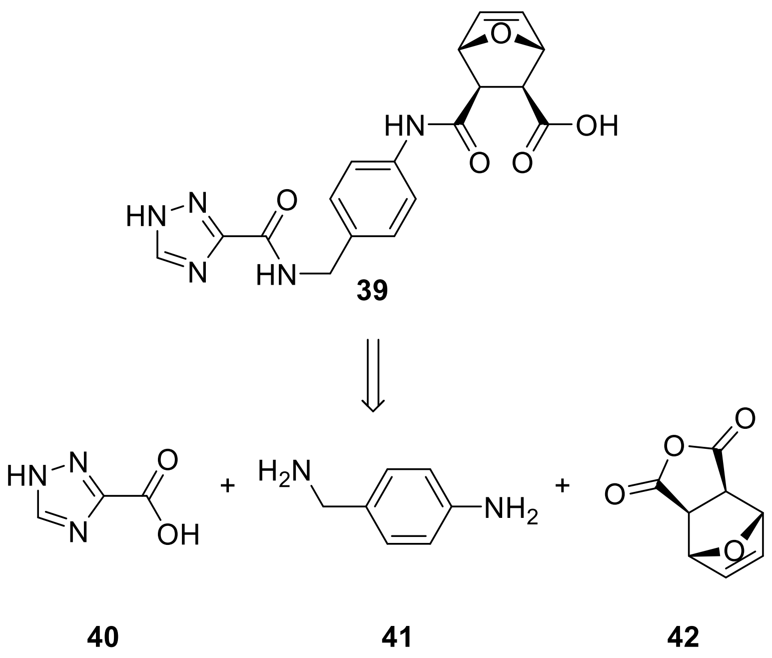
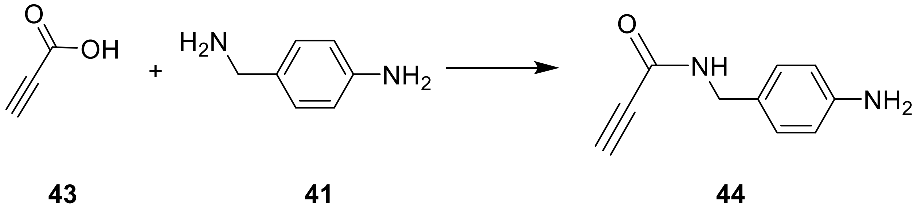


Figure 33: Retrosynthetic analysis of lead structure **39**

As all three fragments are commercially available, it was decided to use these three fragments for a relatively short synthesis using the following sequence.

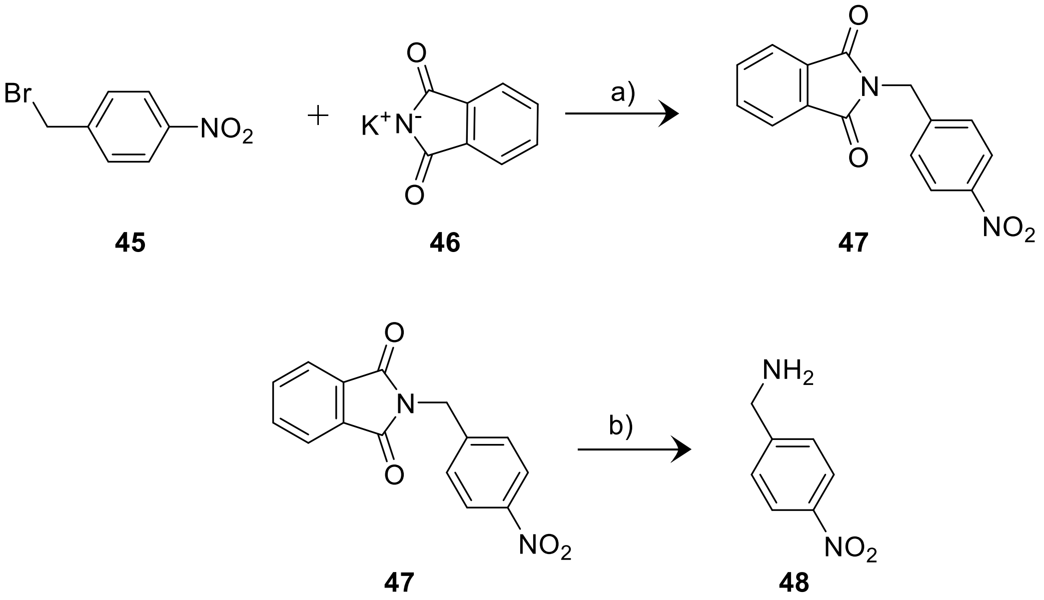
In order to take advantage of the anticipated higher reactivity of the aliphatic amine of **41**, the initial plan consisted of reacting **41** with **40** in an amide coupling reaction, followed by the reaction of **40-41** with **42**. As we encountered supply issues with **40**, we decided to first explore the related 1,2,3-triazole analogues of **39** as this was also a high-ranked structure (ranked 3rd from library4) and it would be interesting to examine how the position of the nitrogens within the ring systems affects activity.

Therefore, the synthetic route was adjusted to react **41** with propiolic acid **43** (Scheme 1). Despite the fact that a reaction occurred, the resulted mixture of products was inseparable. Side products may have resulted from regioisomeric acylation as well as from diacylation and polymerisation of the propiolic acid **43**.



Scheme 1: Synthesis of **44**. Conditions: **41** (1.2 eq.), EDCI-HCl (2.2 eq.), 4-DMAP (0.21 eq.), DIPEA (6.0 eq.), DMF, 0°C to rt, 12 h

In order to avoid these potential selectivity issues, we envisaged a route in which the triazole carboxylic acid **40** was coupled with 4-nitrobenzylamine **48**. Therefore, **48** was synthesised starting from 4-nitrobenzyl bromide **45** via the *N*-(4-nitrobenzyl)phthalimide **47**. The two step Gabriel synthesis sequence afforded **48** in an overall yield of 86% (Scheme 2). Since other phthalimides have been described to have Nrf2-Keap1 PPI inhibitory activity, structure **47** itself was an interesting fragment on its own.[21, 48]

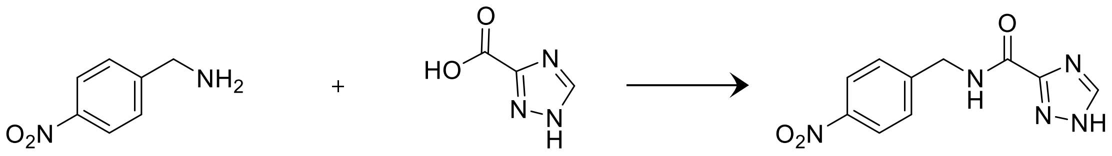


Scheme 2: Gabriel synthesis of 4-nitrobenzylamine **48** Conditions a): potassium phthalimide **46** (1.5 eq.), TBAB (0.001 eq.), DMF, 105°C, 5 h; 93.5% yield b): hydrazine monohydrate (1.9 eq.), MeOH, reflux, 5 h; 92.0% yield

Consequently, several routes were explored to synthesise amide **49**. After attempts with the coupling reagents EDCI (either with 4-DMAP or HOBt addition) and PyBOP which were unsuccessful, the reaction using PyBroP **50** (Figure 34) appeared to give better results. When 4-DMAP was used as an additive, a mixture of complex products was obtained as could be observed by the complex signals in the aromatic region of the ^{1}H NMR spectrum of the reaction product. The switch to HOBt as an additive improved the reaction by giving a cleaner result;

however, the desired product was still not obtained. Finally, when using PyBOP, no reaction occurred.

PyBroP **50** is a useful reagent for the coupling of steric hindered or less reactive amines.[64-66] Although steric hindrance might play a role, the slightly reduced nucleophilicity of amine **48**, demonstrated by a predicted pK_{aH} of 8.36[67], compared for example to Et_3N (pK_{aH} of 10.76[68]), $EtNH_2$ (pK_{aH} of 10.75[68]), or DIPEA (pK_{aH} of 10.57[68]), may be more important in this case. The reduced basicity is due to the *-I* effect of the 4-nitrobenzyl group, which reduces the electron density at nitrogen. As a sufficient electron density is important for the reactivity, this might explain the low reactivity. Couplings involving PyBroP **50** (Figure 33) are thought to involve acyl bromides as intermediates. These are highly reactive and less bulky than HOBt esters which might explain the increased reactivity.

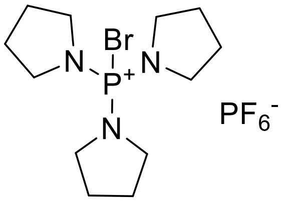


48 **40** **49**

Scheme 3: Amide coupling of amine **48** with carboxylic acid **40**. Conditions used were as listed in the following Table 12:

Eq. Amine	Coupling agent (eq.)	Additive (eq.)	Base (eq.)	Solvent	Temp.	Time	Product isolated?
1	EDCI (1.5)	4-DMAP (0.1)	DIPEA (6.0)	DMF	0 °C → rt	15 h	X
1	EDCI (1.2)	HOBt (1.05)	Et_3N (1.3)	DCM	rt	16.5 h	X
1	EDCI (1.5)	HOBt (1.1)	Et_3N (1.5)	DCM	rt	22 h	X
1.36	PyBOP (1.0)	-	Et_3N (3.0)	CH_3CN	rt	24 h	X
1	PyBroP (1.2)	-	DIPEA (3.0)	DMF	0 °C → rt	15 h	✓ 35.4% yield

Table 12: Reaction conditions explored to synthesise amide **49**.



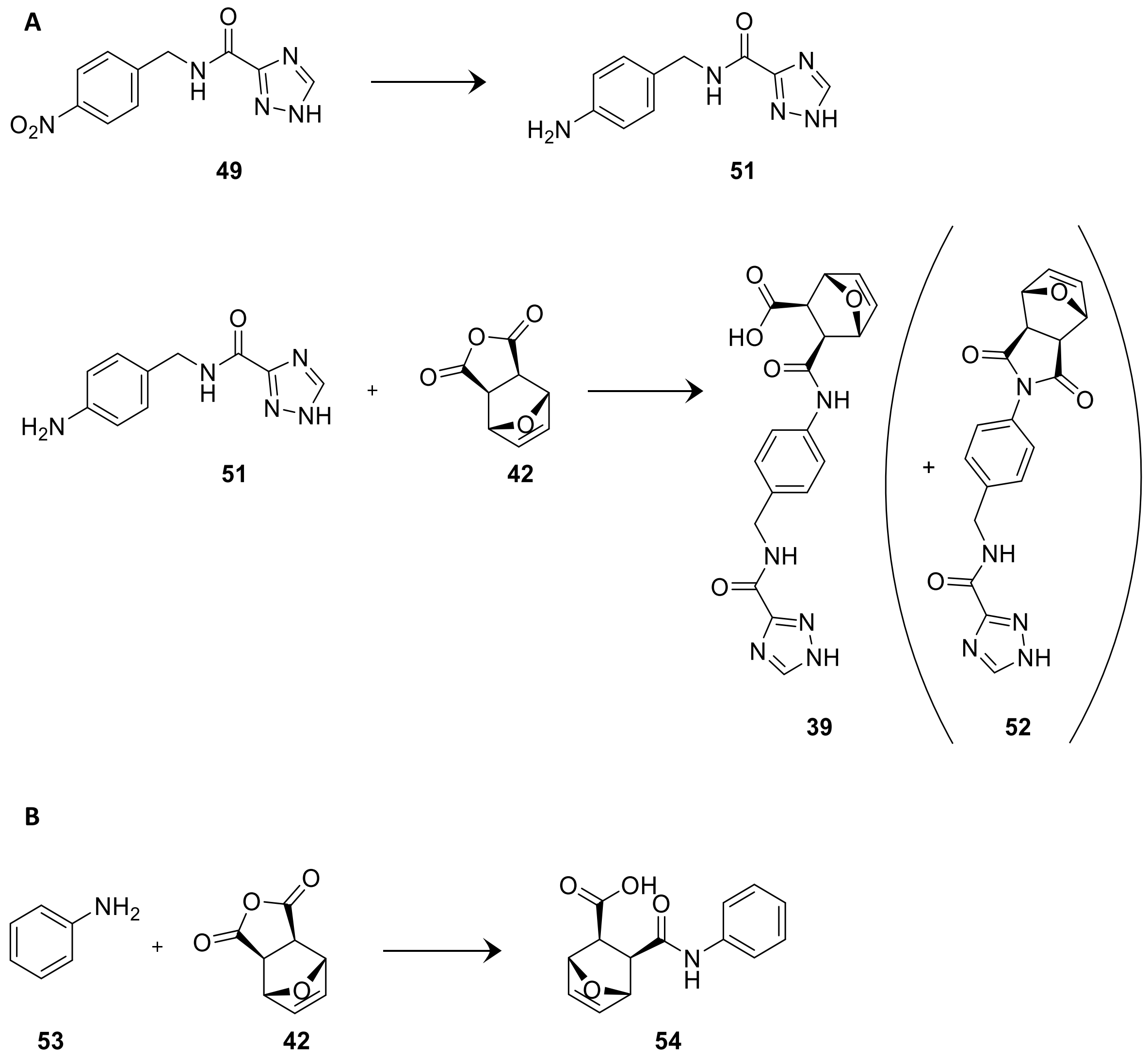
50

Figure 34: Structure of PyBroP **50**

As it was challenging to make **49**, a test reaction was set up to examine if the last step of the synthesis of the diamide **39** would be feasible**.** This last step would be the coupling of the amine **51** with the anhydride **42**, following the reduction of the nitro group of **49**. It was anticipated that it could be difficult to react an aniline with an anhydride in the first place. Additionally, it might then be difficult to stop the reaction cleanly at the amide formation as the primary amine might react twice to form the stable, cyclic imide **52**. The most rudimentary analogue of **51** is aniline **53** which was chosen for the test reaction in order to establish clearly if the coupling works. Both synthetic routes are drafted in Scheme 4 A and B, respectively.

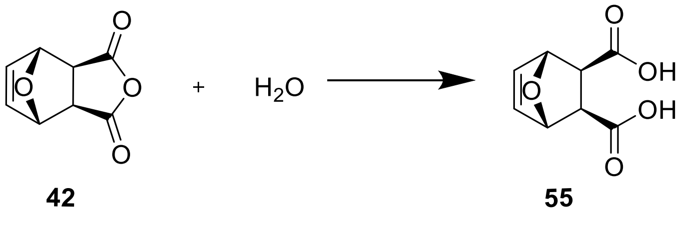
Omuaru *et al*. heated the anhydride **42** and aniline **53** (1.2 eq.) in $CHCl_3$, to synthesise **54**.[69] These conditions were applied to a mixture of **42** and **53** which were stirred in refluxing DCM for 24 h and monitored by TLC (eluent DCM:MeOH = 9:1). A new weak spot was observed by TLC after 2 h, however, this spot was not detectable after 4 (or more) hours, or after the work-up, so this route was not pursued further.

Subsequently, an attempt was made to ring open the anhydride **42** using MeOH in order to obtain the mono-acid/mono-ester which could then be coupled to aniline **53** or amine **51**. For this, **54** was reacted with 2 eq. of MeOH in THF in the presence of DMAP (0.1 eq.).[70] However, an inseparable mixture of products was obtained that prevented this route from being used in the synthesis of **54**.



Scheme 4: **A**: reaction scheme for the synthesis of diamide **39** from nitro compound **49**. After reduction of the nitro group, the aromatic amine **51** would be coupled with the cyclic anhydride **42** to afford amide **39**. **B**: Respective coupling of aniline **53** with the cyclic anhydride **42** in order to determine reaction conditions that might be suitable for the synthesis of **39**.

Another possibility which was explored was to ring open anhydride **42** with H_2O to form the diacid **55** (Scheme 5). The anhydride was dissolved in water and stirred at room temperature for 68 hours, and diacid **55** was obtained in 96.7% yield.



Scheme 5: Hydrolysis of anhydride **42** to diacid **55**.

As a next step, the diacid **55** could be coupled using standard amide coupling chemistry to the amines **51** or **53** to yield the respective amides. Additionally, the diacid **55** would be an interesting fragment to be tested on its own, as it contains the carboxylic acid of final compound **39** as well as a substantial part of its three-dimensional shape and either one or both carboxylic acids may form interactions with Keap1 when bound.

3.3 Conclusion

Through two rounds of virtual screening, it was possible to obtain a promising lead scaffold in terms of predicted binding affinity starting from more than 10,000 compounds. This scaffold was further refined through three iterations of systematic virtual structure activity relationship investigation. By these means, it was possible to maintain predicted affinity of the scaffold while reducing its size and increasing its ligand efficiency and lipophilic efficiency, which have been described as important parameters in medicinal chemistry.[77] Compound **39** was identified as the most promising candidate for synthesis based on its predicted binding affinity, lipophilic efficiency and synthetic accessibility. The synthesis of **39** was initiated and the synthesis of its fragments completed.

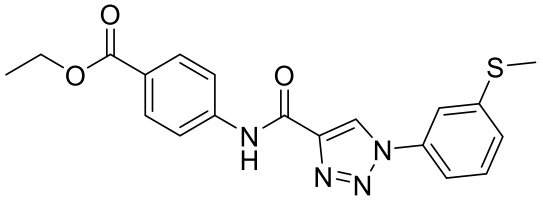
Furthermore, a comprehensive analysis of available Keap1 crystal structures was performed and a thorough selection algorithm for protein structures suitable for virtual screening was developed. Five Keap1 Kelch domain structures were selected and tested to establish whether docking a series of known ligands reproduced the correct rank order of experimental binding affinities. This procedure was also used to evaluate four docking programmes with a total of six scoring functions. Finally, this validation allowed the informed selection of two docking programmes (Autodock Vina[45], LeDock[44]) and three crystal structures (1zgk, 3zgc, 4xmb), performing better than the global consensus, for subsequent virtual screening exercises.

4 Results: Synthesis, evaluation and virtual structure-activity relationship of 1,2,3-triazoles

In parallel to the virtual screening approach (Chapter 3), another, independent approach was followed. The 1,2,3-triazole **56** was recently discovered within our group and showed promising inhibitory potential for the Nrf2-Keap1 PPI as well as promising physicochemical properties. Therefore, an investigation of the structure-activity relationships of **56** was initiated. Distinct groups were varied, the respective compounds synthesised, tested using a competitive fluorescence polarisation assay and the structures docked using the established docking protocol (s. 3.1.2.2). Consequently, the two moieties on the 1,2,3-triazole were varied virtually through three rounds of systematic variation, and were virtually screened. This resulted in the identification of the promising, new virtual hit compound **124**.

4.1 Synthesis of compounds based on 1,2,3-triazoles

Within our group, compound **56** (Figure 35) has been identified as an interesting lead molecule with an IC_{50} of 1.07 µM in a competitive FP assay using the Keap1 Kelch domain. The compound also showed promising activity in cellular assays (induction of the Nrf2 target gene NQO1 at low micromolar concentrations).



56

Figure 35: Structure of 1,2,3-triazole **56**

As **56** was a promising hit, it was important to explore its structure-activity relationships in order to develop an improved compound with suitable physicochemical properties for central nervous system indications. The detection of the molecule’s pharmacophores and less important functional groups would enable us to determine which groups would be suitable for truncation, elongation, replacement, etc. The importance of optimisation is illustrated by the fact that, although **56** complies with the Rule of Five[75] for estimating oral bioavailability, most parameters are unfavourable for a molecule that should penetrate the CNS (Table 13). Therefore, one goal was to explore the structure-activity relationships of the molecule.

Property	Compound 56	Properties of CNS drugs[35,36]	Complies?	Rules for oral bioavailability[37,75]	Complies?
HBD	1	≤ 2	✓	≤ 5	✓
HBA	7	≤ 2 - 4	✗	≤ 10	✓
MW	382.1 g.mol^{-1}	≤ 450 g.mol^{-1}	✓	≤ 500 g.mol^{-1}	✓
clogP	4.05 (c)	≤ 3	✗	≤ 5	✓
tPSA	83.6 Å^2	≤ 80 Å^2	✗	≤ 140 Å^2	✓
Rotatable bonds	7	≤ 3	✗	≤ 10	✓

Table 13: physicochemical properties of **56** and the compliance of these with desirable drug-like properties

Similarly, to **39**, **56** has a modular structure and can be broken down into three parts: One cyclic structure being connected by a linker to an alkyne, that would then form a 1,2,3-triazole with an azide (Figure 36).

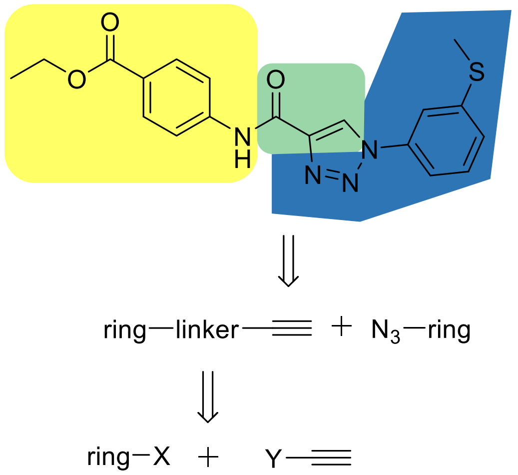


Figure 36: retrosynthetic breakdown of **56**

It was clear from previous preliminary work that the *meta*-thiomethyl substituent of **56** was preferred over other substituents. This left two main points of interest: the linker group between the triazole and the phenyl ring on the left side (in Figure 36), and the substituent on the left-handed ring. Three main question were posed:

1) Is it necessary to have a substituent on the left ring? If yes, is a smaller one sufficient?
2) Is it necessary to have a HBD in the linker group?
3) Is it necessary to have the carbonyl group in the linker group?

In order to be able to answer these questions, the following changes have been performed on **56**: the amide linkage has been changed to an ether or an ester, while the substituents have been either changed to H or to MeO or maintained as EtO_2C (in order to be able to assess the impact of the change of the linker group alone (Figure 37). Therefore, six structures have been synthesised: **57** – **62**, which are summarised in Table 14.

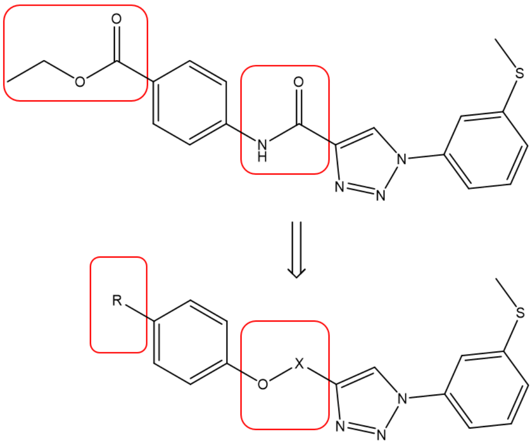


Figure 37: Compound **56** and proposed variations to the linker and R-groups.

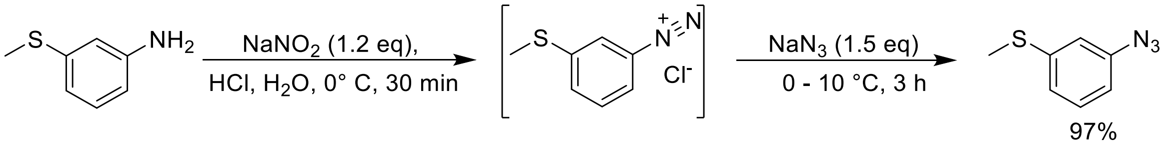
Name	X	R
57	CH_2	H
58	CH_2	MeO
59	CH_2	EtO_2C
60	C=O	H
61	C=O	MeO
62	C=O	EtO_2C

Table 14: substituent and linkers of compounds **57** – **62**

If these changes were to give rise to molecules that maintained their activity, they would have a reduced polar surface area, molecular mass and hydrogen bond acceptors or hydrogen bond donors compared to the original hit compound.

4.1.1 Azide synthesis

The azide that would be needed for all six triazoles, 1-azido-3-(methylsulfanyl)benzene **65**, was synthesised from 3-(methylsulfanyl)aniline **63** via the diazonium salt **64** in 97% yield (Scheme 6).

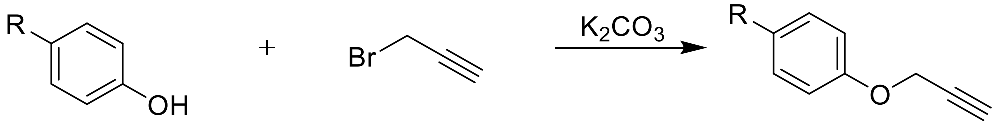


63 **64** **65**

Scheme 6: Synthesis of azide **65**

4.1.2 Ether series

In the first place, the ether series **70** – **72** were synthesised. Generally, the synthesis was straightforward. Propargyl bromide **66** was reacted with the respective phenol **67** – **69** in the presence of K_2CO_3 in a polar solvent (Scheme 7, Table 10).



67 - 69 **66** **70 - 72**

Scheme 7: Reaction scheme of ether synthesis of phenol ether **70** and *para*-substituted phenol ethers **71**, **72** with propargyl bromide **66**. For detailed conditions, see Table 15.

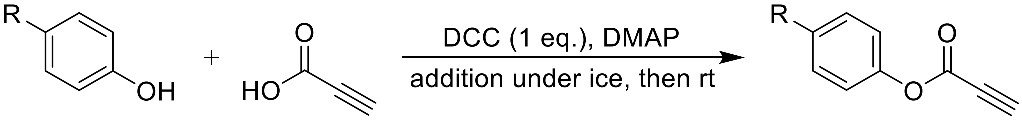
Educt	R	Eq. K_2CO_3	Solvent	Temp.	Time	Yield	Product
67	H	2.5	DMF	rt	19 h	54%	**70**
68	MeO	2.6	DMF	50 °C	24 h	96%	**71**
69	EtO_2C	1.4	Acetone (dry)	Reflux (60 °C)	3 h	92%	**72**

Table 15: Substituents of the compounds of the ether series and reaction conditions.

Upon addition of the base, the phenol is deprotonated and displaces the bromide on the propargyl group, forming the desired product. The synthesis of **67** was attempted under the same conditions as those used for **69**, however the yields were < 10%, even after increasing the amount of base to 3 equivalents and the reaction time to 16 h.

4.1.3 Ester series

Following the ethers, the esters **73** - **75** were synthesised. The respective phenols **67** – **69** were coupled with propiolic acid **43** using DCC and 4-DMAP catalysis (Scheme 8). The yields were reasonable (44% - 62%), but here again, the relatively high reactivity of **43** might affect the yields.



67 - 69 **43** **73 - 75**

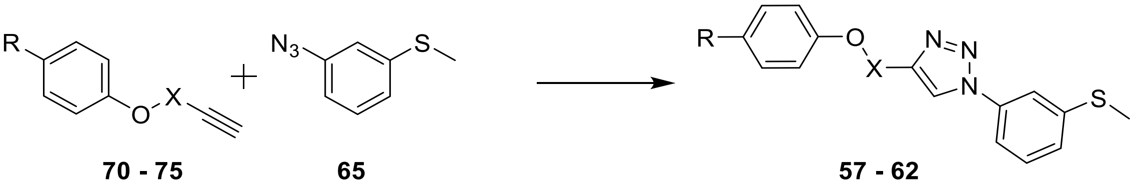
Scheme 8: Reaction scheme of ester synthesis of phenol **67** and *para*-substituted phenols **68**, **69** with propiolic acid **43**. For detailed conditions, see Table 16.

Phenol	R	Eq. acid	Eq. DMAP	Solvent	Time	Yield	Product
67	H	1.10	0.10	DCM (dry)	22 h	44%	**73**
68	MeO	1.00	0.07	DCM (dry)	15 h	62%	**74**
69	EtO_2C	0.99	0.01	THF (dry)	20 h	44%*	**75**

Table 16: Substituents of the compounds of the ester series and reaction conditions. *: calculated for propiolic acid **43**

4.1.4 1,2,3-Triazole synthesis

The triazoles were synthesis by reacting the alkynes **70** – **75** with the azide **65**. For this, standard click-chemistry reaction conditions were used under copper(I)-catalysis (Scheme 9). The added ascorbate reduces the Cu^{II} to Cu^{I} which is necessary for the reaction to take place. The yields were rather low, possibly because the alkynes **70** - **75** had only limited solubility in the reaction solvent. Table 17 gives an overview.



70 - 75 **65** **57 - 62**

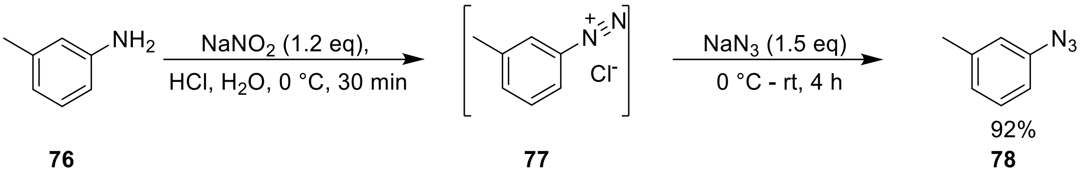
Scheme 9: Synthesis of 1,2,3-triazoles **57** – **62**. Reaction conditions: **65** (1 eq.), $CuSO_4$ (0.05 eq.), sodium ascorbate (0.2 eq.), H_2O/*t*-BuOH (0.6/0.4, V/V), 130 °C, microwave, 45 min

Name	X	R	Yield
57	CH_2	H	15%
58	CH_2	MeO	47%
59	CH_2	EtO_2C	29%
60	C=O	H	48%
61	C=O	MeO	29%
62	C=O	EtO_2C	27%

Table 17: Yields of 1,2,3-triazoles **57** – **62**

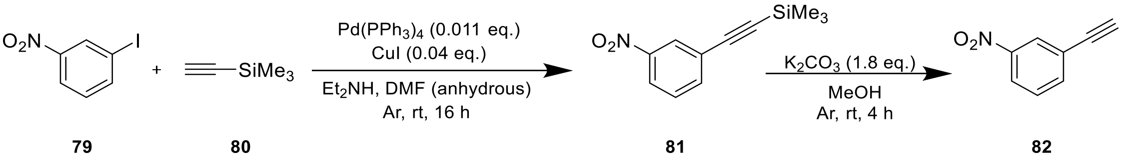
4.1.5 Synthesis of reference compound

In order to have a positive control for later stage experiments, the synthesis of the well described triazole **83** was performed.[51] For this, 3-methylanilline **76** was first reacted with $NaNO_2$/HCl and then with NaN_3 in a one pot synthesis to yield the respective azide, 3-azidotoleune **78**, with 92% yield *via* the diazonium salt **77** (Scheme 10).



Scheme 10: Formation of 3-azidotoluene **78** from 3-methylanilline **76**.

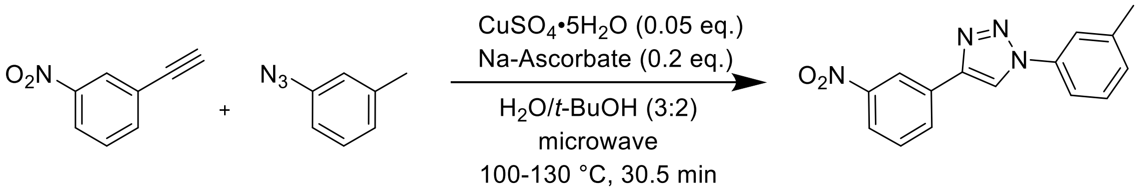
The alkyne **82** which was finally reacted with azide **78** to from triazole **83** was synthesised in a two-step procedure. The still protected alkyne **81** was synthesised by a Sonogashira-coupling, starting from 1-iodo-3-nitrobenzene **79** and using excess of trimethylsilylacetylene **80** (1.1 eq.). **81** was obtained in 42% yield. The TMS group of trimethyl((3-nitrophenyl)ethynyl)silane **81** was removed with K_2CO_3/MeOH to yield the deprotected alkyne 1-ethynyl-3-nitrobenze **82** with a crude yield of 85% (Scheme 11).



Scheme 11: Reaction sequence from 3-iodo-1-nitrobenze **79** to 1-ethynyl-3-nitrobenzene **82**.

When the Sonogashira-coupling was conducted with microwave irradiation as described by Bertrand et al.[51], the reaction failed. The NMR revealed a complex mixture of products. This is probably due to the higher reactivity of **79** compared to the respective bromo-analogue which was used by Bertrand et al.[51] The TMS-protection of the alkyne in the first step is important to prevent that it reacts twice. Additionally, it makes the handling in this case easier as trimethylsilylacetylene **80** is a liquid at room temperature, in contrast to acetylene which is a gas.

The final product, triazole **83**, was obtained using similar conditions as for triazoles **57** - **62** with a yield of 65% (Scheme 12). This was higher than for the previous structures and might be linked to differences in solubility.



82 **78** **83**

Scheme 12: Synthesis of 1,2,3-triazole **83**. Reaction conditions: **82** (1 eq.), **78** (1 eq.), $CuSO_4 \cdot 5H_2O$ (0.05 eq.), sodium ascorbate (0.2 eq.), H_2O/*t*-BuOH (0.6/0.4, V/V), microwave, 100 °C for 30 s, then 130 °C for 30 min

4.2 Biological evaluation: FP assay

In order to determine the binding potency of triazoles **57** – **62**, a competitive FP binding assay, previously described by Hancock et al.[52], was performed (Figure 38). The compounds were tested at a concentration of 10 µM, and all compounds had a purity of at least 95%, if not specified otherwise.

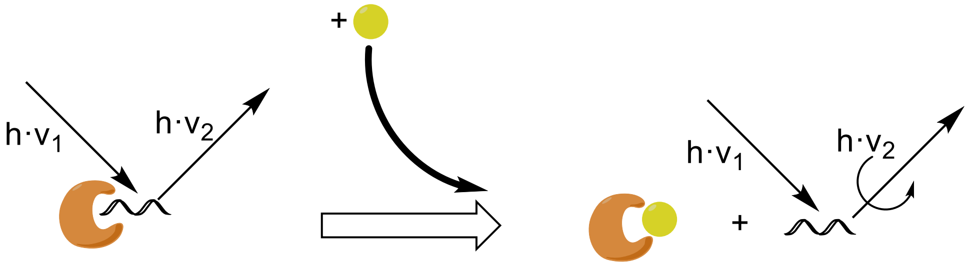


Figure 38: Schematic representation of the FP assay. The Keap1 Kelch domain (200 nM, orange crescent) was incubated with a fluorescently labelled peptide (FITC-β-DEETGEF-OH, 1 nM, wavy line) for 2.5 h. Consequently, the triazoles were added (yellow balls). The solution was irradiated with linearly polarised light with the frequency v_1 which is absorbed by the fluorophore and remitted with wavelength v_2. Due to the rotation of the molecule during the time delay between absorption and emission, the plane of the emitted light is rotated compared to the incident light. This change of angle φ between polarisation plane of irradiated and emitted light can be measured. As the fluorophore increases in size (such as upon binding of a small fluorescent molecule to a large protein), the rotation is slowed and therefore φ is reduced.

Except for **59**, the inhibition was determined to be between 26% and 39% at a compound concentration of 10 µM (Table 18, Figure 39). Not surprisingly, compound **62**, where only an NH group has been replaced by an O, showed the strongest inhibition. However, **59**, the ether analogue of **56** and **62**, only inhibited the fluorescent peptide-Keap1 interaction by 13.2%. It can be speculated that the MeO substituent seemed to be slightly preferred over the unsubstituted phenyl (**58** vs. **57** and **61** vs. **60**), however, this does not appear to affect activity strongly. Generally, all tested compounds were significantly less active than the parent structure **56** which inhibited the Nrf2-Keap1 interaction completely at a concentration of 10 µM.

Except for the two compounds with the ethoxy carbonyl substituent, **59** and **62**, the other two pairs of compounds do not suggest that the linker group plays a major role. This indicates, together with previous data that suggested that the amide of **56** could be substituted with an amine and partially retain activity, that a hydrogen bond donor in the linker position might be important for activity, in contrast to a hydrogen bond acceptor such as a carbonyl group.

Compound	% inhibition ± SD	Significance of difference in activity to the reference compound 56	Significance of difference in activity to the reference compound 62
56	107.2 ±15.4	-	Yes (P < 0.001)
57*	29.5 ± 6.8	Yes (P < 0.001)	Yes (P < 0.01)
58	30.4 ± 8.6	Yes (P < 0.001)	Yes (P < 0.05)
59	13.2 ± 10.3	Yes (P < 0.001)	Yes (P < 0.001)
60	26.2 ± 6.9	Yes (P < 0.001)	Yes (P < 0.05)
61	33.5 ± 7.2	Yes (P < 0.001)	No (P > 0.05)
62**	39.3 ± 6.4	Yes (P < 0.001)	-

Table 18: Percentage inhibition of triazoles **56** – **62** in a competitive FP assay, tested at 10 µM compound concentration. Significance level calculated using a two-tailed t-test. SD: standard deviation. *purity ≥90%, **purity ≥88%

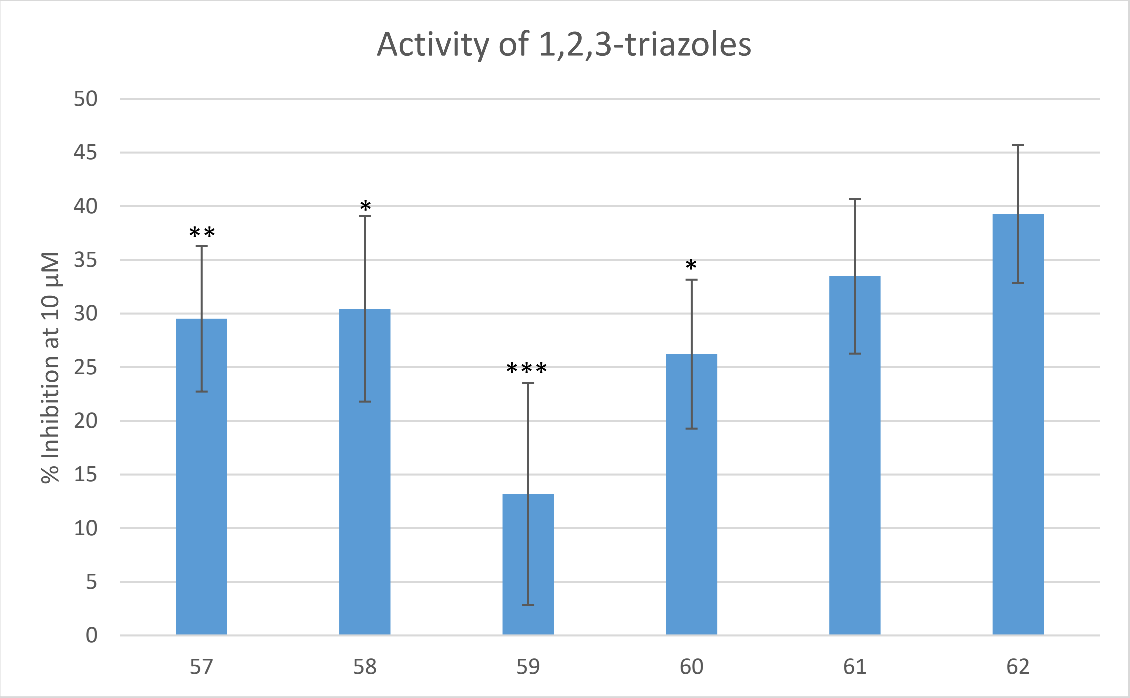


Figure 39: Percentage inhibition of triazoles **57** – **62** in a competitive FP assay, tested at a 10 µM compound concentration. (*): p < 0.05, (**): p < 0.01, (***): p < 0.001, compared to structure **62**

The experiment was performed twice and triplicate. To check that the assay was executed properly, blanks (no Keap1, peptide or inhibitor) were included and were used as a measure for the non-specific background fluorescence. Furthermore, three wells each contained only the fluorescent peptide (corresponding to 100% peptide-Keap1 interaction inhibition) and the fluorescent peptide and Keap1, but no inhibitor (corresponding to no peptide-Keap1 interaction inhibition). The values were reproducible when the same well was measured several times, but showed significant variations between different wells containing the same sample. However, as the readouts were always between the readout for 0% and 100% inhibition, this suggests that the assay itself worked well and that no systematic error was inherent to the assay, but might be a sign for random errors (such as pipetting errors). Nevertheless, as all compounds but **56** show inhibition rates close to 50% inhibition, it is not very surprising that higher error rates are observed as if we are approaching the inflection point of a dose-response curve, minute changes (in concentration *e.g.*) lead to quite a different signal. In a follow up, a multidose experiment could be performed which should show less error and gives a more complete picture of the inhibitory profile of the compounds. It would have been advantageous to include **56**, for which the inhibition rate is known, to assure that the assay would give similar results to when the inhibition rate for **56** was determined. Therefore, the result should be considered cautiously.

In conclusion, these findings suggest that the amide linker is the most potent of those investigated so far. However, if other parts of the molecule are optimised appropriately, an ether linkage may be tolerated and reduce the hydrogen bond donor count of the molecules as well improve *in vivo* stability.

4.3 Docking of synthesised triazoles

In order to be able to rationalise the observed affinities of the triazole series **57** - **62** and **56** for Keap1 and to guide further optimisation, the synthesised molecules were docked, as previously described, into the 1zgk, 3zgc and 4xmb Keap1 crystal structures using Autodock Vina[45] and Ledock[44].

First, the scores of the docked poses were normalised to the respective average scores of the 7 compounds for an individual docking algorithm and scoring function combination on an individual crystal structure, leading to a total of 21 relative scores for each compound.

Disappointingly, no strong correlation could be seen between the results from the docking compared to the experimentally observed Keap1 binding. When computing the correlation between the average relative scores and the percentage inhibition obtained from the FP assay, a negative correlation of almost -0.8 was obtained. If the compounds were ranked according to their percentage inhibition and this correlated with the average rank of the computed scores, a less bad correlation was obtained, but it was still negative (-0.31).

One problem, especially in terms of score correlations, appeared to be compound **56**. Experimental evidence suggests that this is the most potent binder to Keap1 (Table 18), but it had the lowest relative score and rank for the consensus docking and scoring (Table 19, columns 2 and 3; Table 20), and **56** was never recognised as the most potent compound in any of the investigated combinations. The observation that there were problems with the prediction of binding scores for **56** is emphasised as there is one very good combination of docking and scoring function for these compounds if **56** is excluded from the analysis: Autodock Vina[45] with its native scoring function. If **56** is excluded from the analysis, the correlation increases from -0.18 to 0.70 for the scores and from the 0.24 to 0.86 for the ranks, which correspond to large changes in correlation (Table 19). No similar change could be observed for any other combinations (including the global consensus) of docking software and scoring function, although all tended to show better correlation upon exclusion of **56** (Table 19). Due to the sample size and similar chemical structures of the docked compounds, these results cannot be easily generalised and this type of investigation can yield very different results for different chemical structures (s. 3.1.2.2). However, they emphasise the difficulty to find optimal docking and scoring function combinations that perform robustly on a wide variety of input structures.

Correlated pairs	Correlation
Relative Score (Computational) : Inhibition (Experimental)	-0.799
Rank (Computational) : Rank (Experimental)	-0.305
Relative Score (Computational) : Inhibition (Experimental) (without **56**)	-0.327
Rank (Computational) : Rank (Experimental) (without **56**)	0.176
Relative Score (Vina docking and scoring) : Inhibition (Experimental)	-0.177
Rank (Vina docking and scoring) : Rank (Experimental)	0.245
Relative Score (Vina docking and scoring) : Inhibition (Experimental) (without **56**)	0.696
Rank (Vina docking and scoring) : Rank (Experimental) (without **56**)	0.855

Table 19: Correlations between computed and experimentally determined scores/affinities and ranks of compounds **56** – **62**.

Furthermore, **59** has the worst experimental affinity, however it has the best average rank for the consensus method (Table 17). In contrast, the Vina docking and Vina scoring combination correctly assigns the worst rank to this compound (Table 17).

Compound	Relative Score (Consensus)	Rank (Consensus)	Relative Score (Vina-Vina)	Rank (Vina-Vina)	Inhibition (FP)	Rank (FP)
56	0.936±0.103	5.10±2.05	0.982±0.018	4.33±2.31	107.2 ±15.4	1
57	0.980 ± 0.059	4.67±1.88	0.986±0.007	4.67±2.08	29.5 ± 6.8	5
58	1.005 ± 0.067	3.86±1.93	0.986±0.006	4.33±1.53	30.4 ± 8.6	4
59	1.046 ± 0.069	3.00±1.82	0.982±0.009	5.00±1.73	13.2 ± 10.3	7
60	0.988 ± 0.078	4.19±2.05	1.011±0.017	2.33±0.58	26.2 ± 6.9	6
61	1.031 ± 0.048	3.29±1.49	1.012±0.018	2.33±0.58	33.5 ± 7.2	3
62	1.015 ± 0.075	3.29±2.22	1.041±0.003	1.00±0.00	39.3 ± 6.4	2

Table 20: Relative scores of the consensus docking and scoring (consensus), the relative scores after Autodock Vina[45] docking and scoring (Vina-Vina), and the experimentally observed percentage inhibition (FP) at 10 µM compound concentration and ranks of compounds **56** – **62**. Expressed as means ± standard deviation.

Overall, this shows that there is generally a poor correlation between experimental and computational results and that the best set of tools varies from case to case. Additionally, as no co-crystal structures for structures **56** – **62** together with Keap1 are available, it cannot be determined if the discrepancy arises from suboptimal docked poses or flawed scoring.

If the docked poses are examined, we can speculate that the predicted pose might be a part of the problem. For **56**, the methylthiophenyl moiety reaches outside of the binding pocket, suggesting that it may have little enthalpic or entropic (such as water displacement) contribution to interaction (Figure 38, light blue structure). However, the predicted pose of **62**, the ester analogue of amide **56**, has the methylthiophenyl moiety positioned deeper within the binding pocket. This leads to a conformation in which **62** has a high contact surface area with Keap1 (Figure 38, green structure). In terms of enthalpic contributions, the structures do not appear to largely differ as both are predicted to form two intermolecular hydrogen bonds (Figure 40).

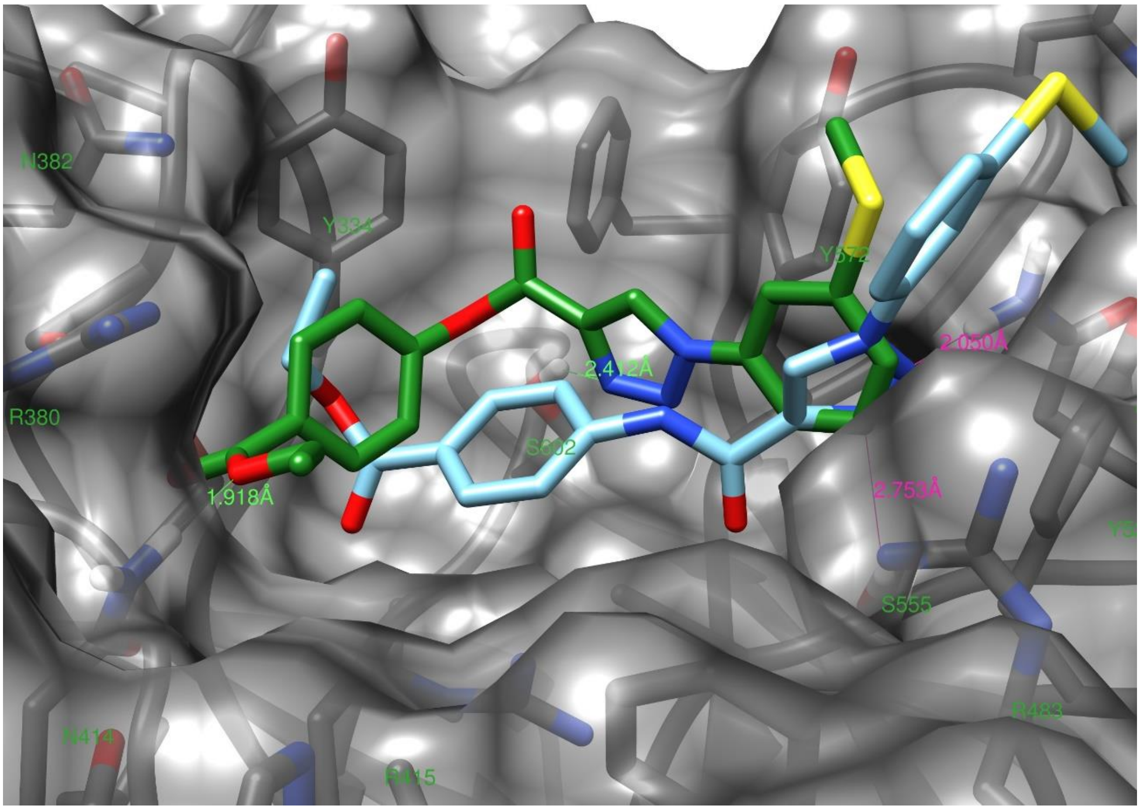


Figure 40: All shown structures were docked with Autodock Vina into the Keap1 Kelch domain from 3zgc (grey). **56** is shown in light blue and its hydrogen bonds in purple, **62** is shown in green and its hydrogen bonds in green. Both compounds engage in two hydrogen bonds (**56** via its triazole to Q530 and S555; **62** via its ester to N414 and via its triazole to 602). Compound **62** occupies more of the pocket between Y334, R380 and N414, winding along the Keap1 surface, whereas the methylthiophenyl moiety of **56** reaches out of the pocket.

Interestingly, all structures (other than **56**) with a *para*-substituent had almost the same binding conformation. The two unsubstituted analogues **57** and **60** had a binding mode in which the structures were flipped by 180° compared to **62**, exploiting the space in the pocket between

Y334 and R380 by the methylthiophenyl group. This indicates the importance of a big substituent at this position and could also explain why in other experiments, the ethyl ester showed better inhibition rates than the respective carboxylic acid, despite the usual need for an acidic group for Keap1 binding affinity. Table 21 summarises the interactions of the docked compounds and gives selected physicochemical properties.

Name	linker	R	HBA	HBD	tPSA (in Å²)	Interacting residues
56	NHC=O	EtO_2C	7	1	83.36	Q530, S555
57	CH_2O	H	4	0	37.19	Y525, S602
58	CH_2O	MeO	5	0	46.42	S602
59	CH_2O	EtO_2C	6	0	63.49	N414, S602
60	OC=O	H	5	0	54.26	Y525
61	OC=O	MeO	6	0	63.49	S602
62	OC=O	EtO_2C	7	0	80.56	N414, S602

Table 21: List of linker group between aryl group and triazole of the alkyne-bearing moiety, substituent of this aryl group, hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), total polar surface area (tPSA) and the (predicted) interacting residues of compounds **56** to **62**.

4.4 Virtual structure activity relationship

With the above-mentioned experiments, it could be established that the amide linkage can be considered being favourable for the activity of compound **56**, with the possibility to change to an amine or ether at a later point for pharmacokinetic optimisation. However, it was not known in detail how the substitution pattern affects activity.

In the previously published series of 1,2,3-triazoles Keap1-Nrf2 interaction inhibitors, different substituents and substitution patterns have been investigated.[51] In order to have a rationale for the binding mode, and to inform how this series should be manipulated in future work, a three-stepped process of virtual screening was performed. First, a series of alkynes was coupled *in silico* using AutoClickChem[71] with 3-azidobenzoic acid **84** to form the 1,4- and 1,5 substituted regioisomeric triazoles. These were subsequently docked to Keap1. Second, the same procedure was performed with one alkyne, ethyl 4-propiolamidobenzoate **85**, and a set of azides. After both experiments, promising alkynes and azides were identified. Finally, these were combined to form a new set of triazoles which were docked into the Keap1 binding pocket.

The average score was used for each of the following calculations. Crystal structures 1zgk, 3zgc, and 4xmb were used as described previously. For these screens, a new method of assessment was employed to avoid earlier discussed shortcomings. First, for every combination of docking and scoring function (Autodock Vina docking with Vina, CyScore and DSX scoring, Ledock docking with Ledock, Cyscore, DSX and Vina scoring), a relative score was calculated. This was defined as the score of a certain ligand divided by the average score of all ligands, for each specific docking and scoring function combination. Using this, the structures were ranked according to their relative scores and a curve was fitted through these data points. Most of the values were on or close to a straight line, corresponding to the large, middle group of structures with similar activity. However, at the high and low affinity ends of the scale, there was an inflection in the ranking curve associated with compounds that bound very well or very poorly, respectively. Figure 41 gives a specific example. The first group was of interest, and for every individual structure, the number of times a compound appeared in this part of the curve for each docking-/scoring function combination was counted (‘counts’) and molecules that were above a chosen threshold were considered a virtual hit. Per se, the calculation of a relative score is not necessary for this type of analysis. However, it makes the results using different software packages or crystal structures easier to compare, as the usage of certain protein structures tended to result in higher scores in general.

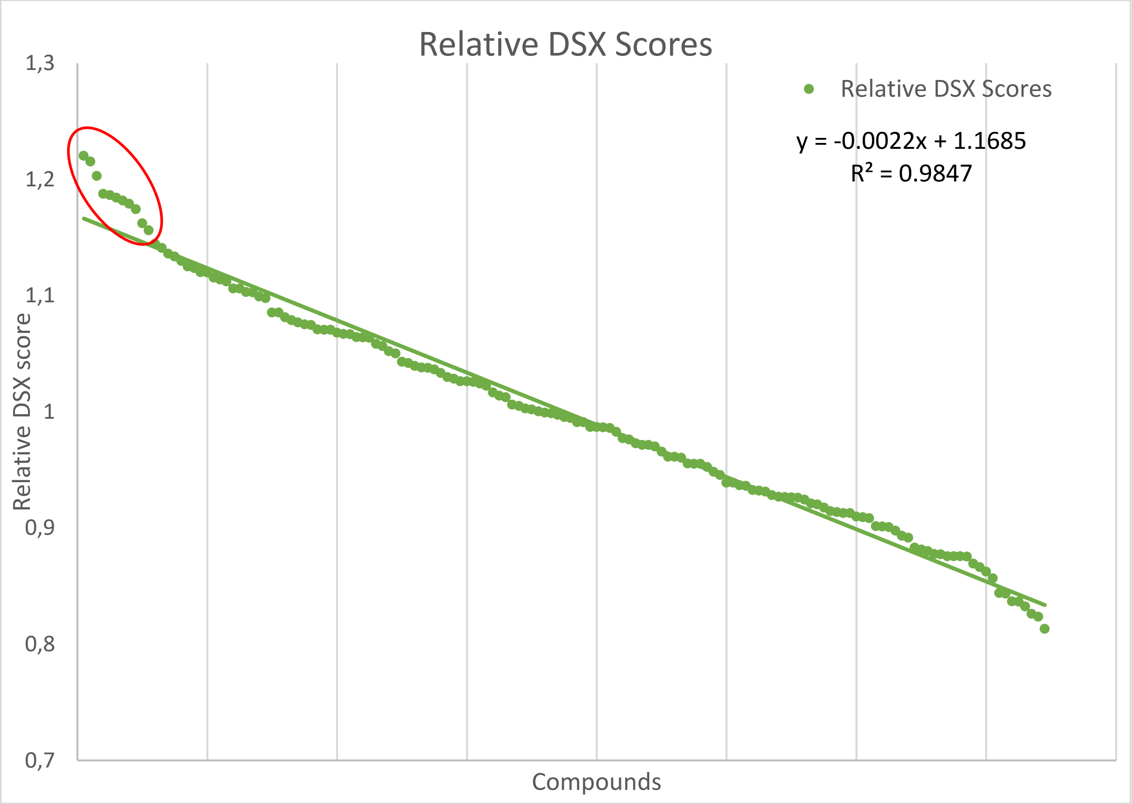
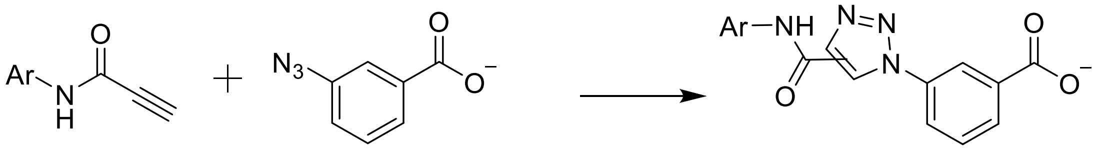


Figure 41: The structures of the first screening, described in the next paragraph, were ranked according to their relative scores, in this specific example after LeDock[44] docking and DSX[61] scoring. If a compound's relative score was superior to what would be expected from the linear relationship between rank and score (data points within red ellipse), the structure was considered a hit in this specific combination of docking and scoring.

4.4.1 First screening: Virtual structure activity relationship of the alkyne-bearing moieties

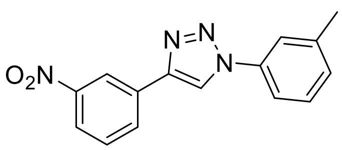
The first step was to evaluate promising substitution patterns on the aromatic ring brought into the triazole by the alkyne bearing moiety. As an azide, 3-azidobenzoic acid **84** was selected due to the known, positive contribution to binding affinity by this motif to compounds from this class, demonstrated by published and more recent, unpublished data.[51] This azide was reacted *in silico* with a library of 72 alkynes (Figure 42). Since both the 1,4- and 1,5-triazoles were created, this resulted in a library of 144 triazoles to be docked. On top of that, five triazole structures (Figure 43) which were known to bind to Keap1 were included, resulting in a total number 149 structures. This set of ligands was docked in triplicate.



86 **84** **87**

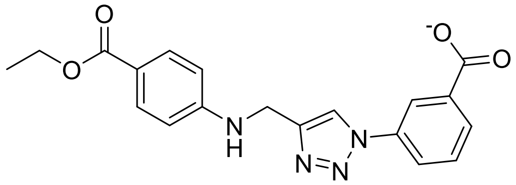
Figure 42: *in silico*-cycloaddition of a *N*-phenyl, *N*-pyridinyl or *N*-pyrimidonyl-substituted propargyl amide **86** with 3-azidobenzoic acid **84** to form the 1,4- and 1,5-substituted 1,2,3-triazoles **87**. In detail, Ar was a phenyl rest monosubstituted, with F, Cl, Br, Me, Et, *n*-Pr, *i*-Pr, *t*-Bu, CF_3, OMe, OEt, O*n*-Pr, O*i*-Pr, O*t*-Bu, OCF_3, CN, (*(Z)*-*N*-methoxyacetimidoyl cyanide)-2-yl, oxazole-2-yl, 4-ethyloxazole-2-yl, 1*H*-1,2,4-triazole-5-yl, *N,N*-dimethylaminocarboxamide, *N*-ethylcarboxamide. All these substituents were investigated in ortho-, meta- and para-position. Additionally, the following heterocycles were investigated: *N*-(4-aminopyrimidin-2(1H)-one)-yl, *N*-(4-amino-2-chloropyrimidine-5-carbonitrile)-yl, *N*-(pyridine-3-amine)-yl, *N*-(6-fluoropyridine-3-amine)-yl, *N*-(5-carboxylic acid-pyridne-3-amine)-yl, *N*-(ethyl 5-carboxylic acid-pyridne-3-amine)-yl.

The different structures of the aromatic (Ar) variable groups of **86** can be separated in broadly three sets: first, common, commercially available monosubstituted benzenes were employed. These would allow rapid synthesis of compounds for preliminary tests. This included (pseudo-)halides (F, Cl, Br, CN), simple alkyl groups (Me, Et, *n*-Pr, *i*-Pr, *t*-Bu, CF_3) and the respective ethers. The second group were ester bioisosteres, aiming at mimicking an ethyl carboxylate, which previously showed superior activity compared to the respective acid (((Z)-N-methoxyacetimidoyl cyanide)-2-yl, oxazole-2-yl, 4-ethyloxazole-2-yl, 1*H*-1,2,4-triazole-5-yl, N,N-dimethylaminocarboxamide, N-ethylcarboxamide). The substituents from both groups were placed in ortho-, meta- or para-position, relative to the amine. As to date, no heterocyclic structures for Ar in **86** have been used, it was sought to enlarge the chemical space, speculating that introducing polarity might improve interactions as polar side chains appeared to be more favourable than apolar ones. The structures included were pyridine and pyrimidine derivatives in order to maintain ring size (N-(4-aminopyrimidin-2(1H)-one)-yl, N-(4-amino-2-chloropyrimidine-5-carbonitrile)-yl, N-(pyridine-3-amine)-yl, N-(6-fluoropyridine-3-amine)-yl, N-(5-carboxylic acid-pyridne-3-amine)-yl, N-(ethyl 5-carboxylic acid-pyridne-3-amine)-yl. As reference compounds, structures **56**, **83** and **88 – 90** were included. **56** and **88** - **90** have been synthesised recently within our group and have IC_{50} values (FP) between 0.6 and 8.3 µM for interaction with the Keap1 binding pocket. Compound **83** was a compound previously synthesised by our group that also showed activity in cellular assays (Figure 44). Figure 44 gives an overview of the workflow.



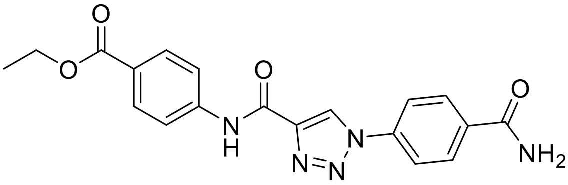
83

IC_{50} (FP) = 10.0 ± 2.7 µM[51]



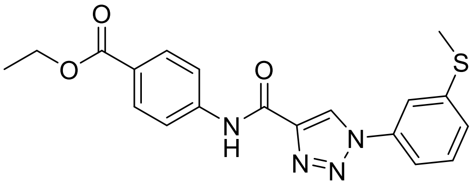
88

IC_{50} (FP) = 1.44 ± 0.34 µM



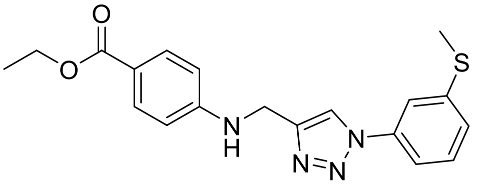
89

IC_{50} (FP) = 0.598 ± 0.061 µM



56

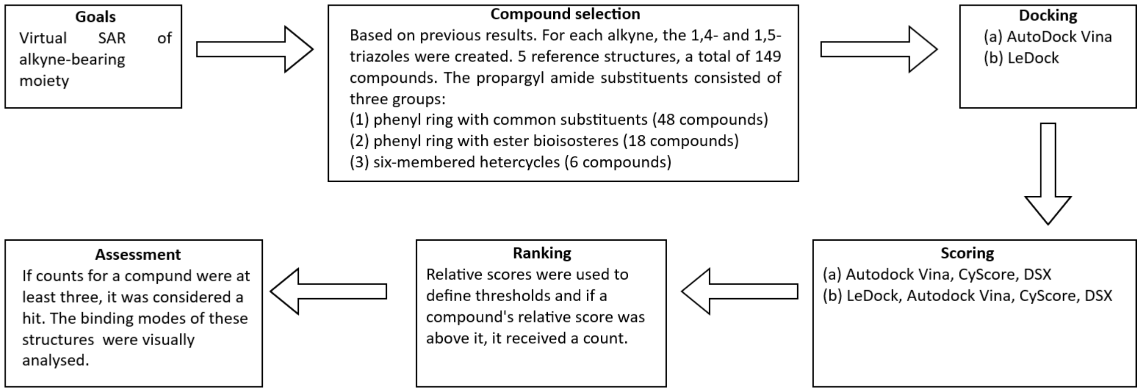
IC_{50} (FP) = 1.07 ± 0.12 µM



90

IC_{50} (FP) = 8.25 ± 1.13 µM

Figure 43: Included reference structures **56**, **83** and **89** – **90** with their respective IC_{50} values (determined by fluorescence polarisation as described in ref. 51 (**83**) or ref. 52 (**56**, **88** - **90**).



Goals
Virtual SAR of alkyne-bearing moiety

Compound selection
Based on previous results. For each alkyne, the 1,4- and 1,5-triazoles were created. 5 reference structures, a total of 149 compounds. The propargyl amide substituents consisted of three groups:
(1) phenyl ring with common substituents (48 compounds)
(2) phenyl ring with ester bioisosteres (18 compounds)
(3) six-membered hetercycles (6 compounds)

Docking
(a) AutoDock Vina
(b) LeDock

Scoring
(a) Autodock Vina, CyScore, DSX
(b) LeDock, Autodock Vina, CyScore, DSX

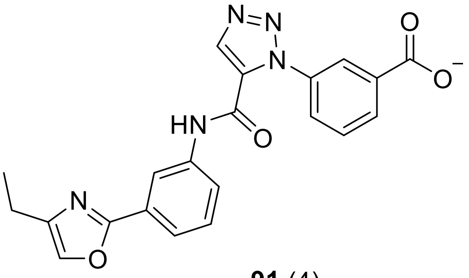
Ranking
Relative scores were used to define thresholds and if a compound's relative score was above it, it received a count.

Assessment
If counts for a compund were at least three, it was considered a hit. The binding modes of these structures were visually analysed.

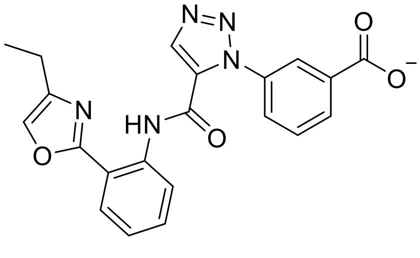
Figure 44: Workflow of the first virtual screen aimed at determining promising substituents on the alkyne. The compounds were docked using AutoDock Vina and LeDock, scored with three or four, respectively, different scoring functions and the compounds received a count if their relative score was above a threshold for an individual docking and scoring function combination. If the number of their counts was three or higher, they were considered a virtual hit in this screen and their docking pose was analysed visually.

The results were analysed visually and quantitively as described in the previous section. The counts threshold was set to three, resulting in 10 structures that would be brought forward. Out of these, five structures were ortho-, and five were meta-substituted. Half of the molecules were 1,4-, the other 1,5-substituted triazoles. There was no correlation for a preference for combinations of triazole and Ar regioisomers. The best performing reference structures were **88** and **90** with a count of two. Interestingly, the amine linked compound **90** scored better than compounds **56** and **89**, despite experimental data suggesting the opposite.

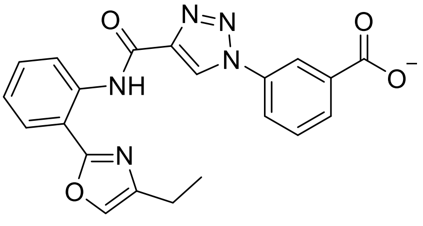
As substituents, the heterocyclic ester bioisosteres were largely preferred. Four structures contained a 4-ethyloxazole, two structures a 1,2,4-triazole and one contained an oxazole moiety. Furthermore, one structure each contained a trifluoromethoxy and an ethyl carboxamide substituent. In detail, the structures with Ar = Ph-R with the following rest were considered as virtual hits (in parentheses: central 1,4- or 1,5-triazole/counts): 3-(4-ethyloxazole-2-yl) (1,5/4) **91**, 2-(4-ethyloxazole-2-yl) (1,5/4) **92**, 2-(4- ethyloxazole-2-yl) (1,4/4) **93**, 3-(1*H*-1,2,4-triazole-5-yl) (1,5/4) **94**, 3-(4-ethyloxazole-2-yl) (1,4/3) **95**, 3-(oxazole-2-yl) (1,5/3) **96**, 2-(oxazole-2-yl) (1,4/3) **97**, 2-(1*H*-1,2,4-triazole-5-yl) (1,4/3) **98**, 3-(*N*-ethyl carboxamide) (1,5/3) **99**, and 2-trifluoromethoxy (1,4/3) **100** (Figure 45).



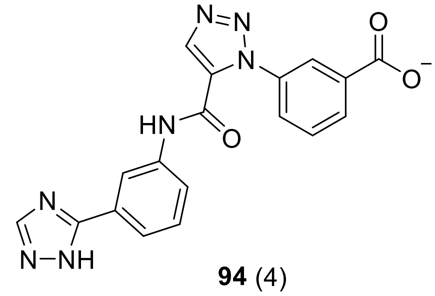
91 (4)



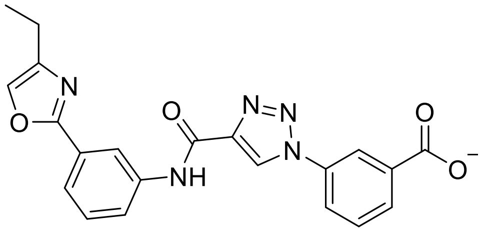
92 (4)



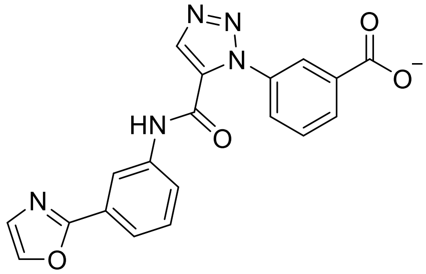
93 (4)



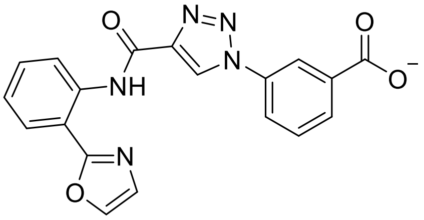
94 (4)



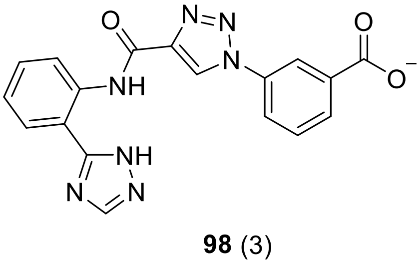
95 (3)



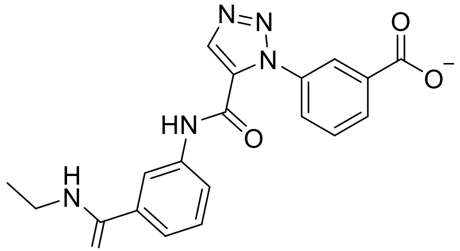
96 (3)



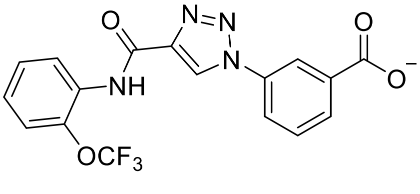
97 (3)



98 (3)



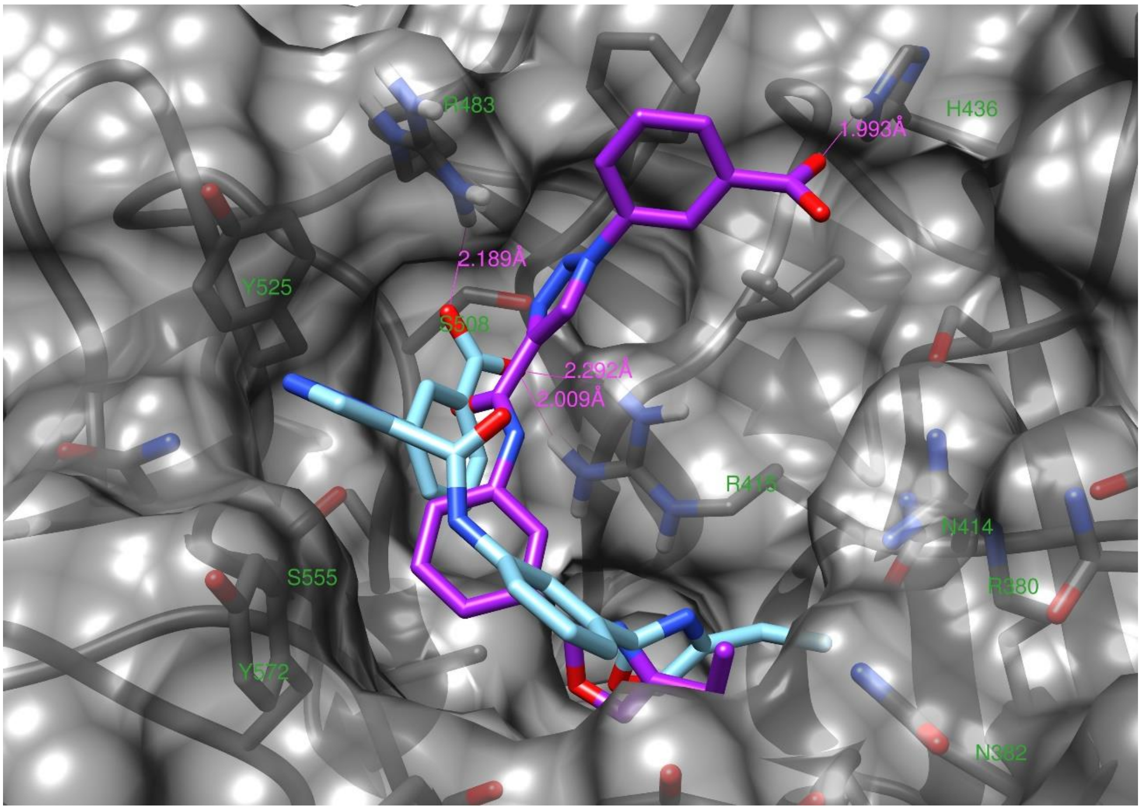
99 (3)



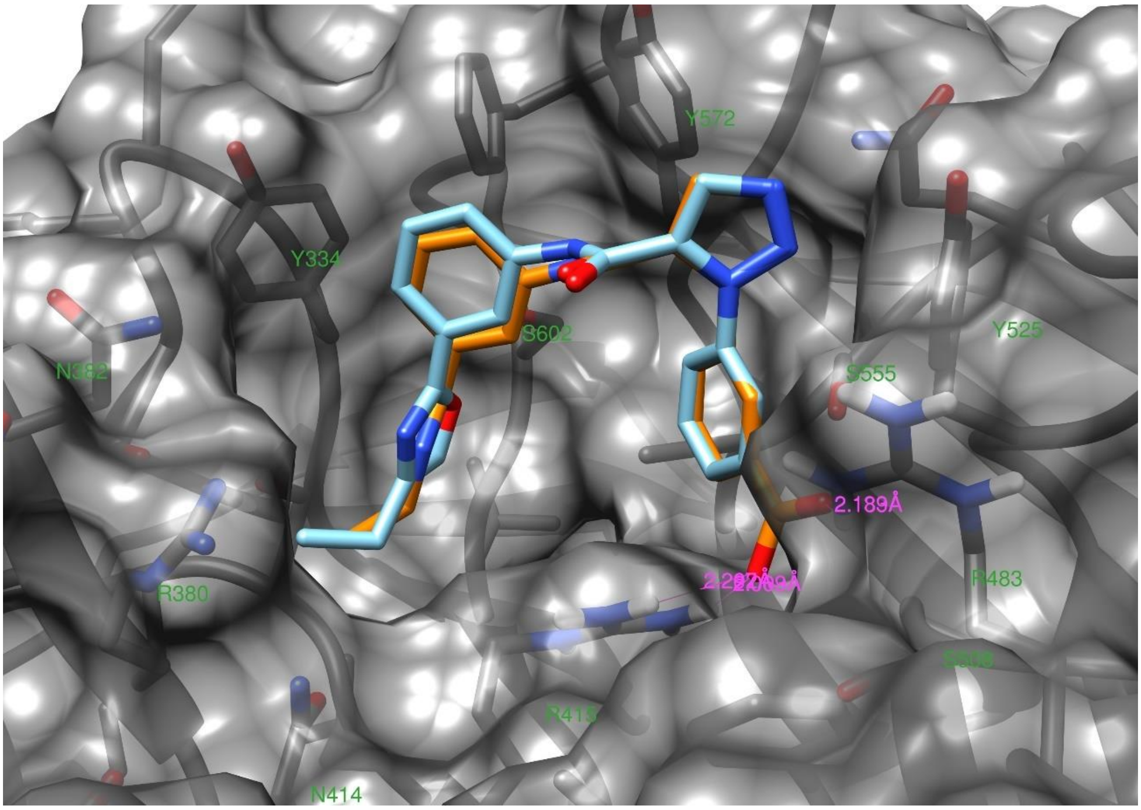
100 (3)

Figure 45: Virtual hit structures **91** – **100** from the virtual screening aiming to establish promising substituents on the alkyne bearing ring. In parentheses: number of counts.

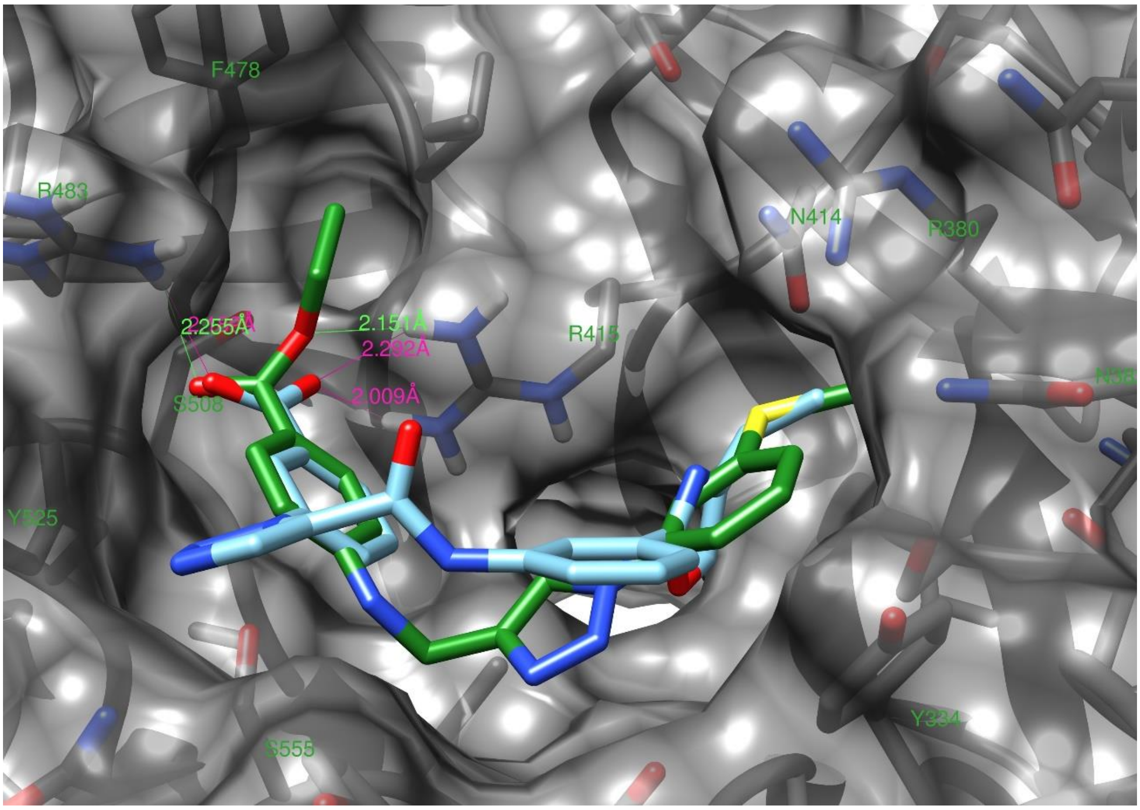
The overlay of similar structures showed that the predicted interactions with Keap1 are strongly dependent on the position and nature of the substituents. Carboxylic acid substituents usually formed at least one hydrogen bond/salt bridge interaction with Arg415 and/or Arg483, or His436 (Figure 46 A). The ethyl groups of the ethyloxazoles and *N*-ethyl amides filled the pocket between Tyr334 and Arg380, however not completely (Figure 46 B). This suggests that bulkier groups on that part of the molecules could decrease the score and gives a reason why the ethyl esters of tested compounds might have given promising results, as this binding mode could be observed for most of the docked and tested structures **56**, **88** and **89** too (Figure 46 C).


R483
H436
1.993Å
2.189Å
Y525
S508
2.292Å
2.009Å
R413
N414
R380
S555
Y572
N382

A


Y572
Y334
S602
Y525
N382
S555
2.189Å
R380
2.292Å
2.009Å
R483
S508
R413
N414

B

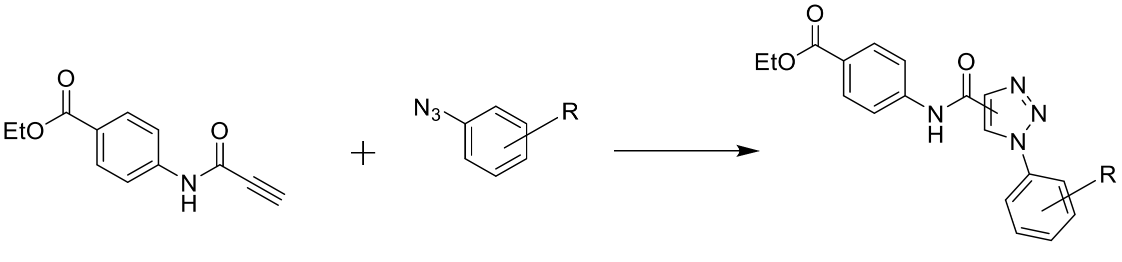


C

Figure 46: All shown structures were docked using Autodock Vina into the Keap1 Kelch domain from 3zgc (grey). Usually, polar groups such as the carboxylic acids or the triazoles formed electrostatic and hydrogen bond interactions with Arg415, Arg483, H436 or other charged groups. **A**: Superimposed structures for R = 3-(4-ethyloxazol-2-yl, 1,5-substituted triazole) (light blue) and R = 3-(4-ethyloxazol-2-yl), 1,4-substituted triazole (purple). Similar structures can have rather different positions, even if some moieties anchor the molecule at similar positions, such as the ethyl groups, filling nicely the pocket between R380 and Y334, potentially even allowing the ethyl group to be replaced by bulkier groups. The carboxylic acid of the 1,4-compound forms three hydrogen bonds with the Arg dyad and appears to be more or less optimally placed there. **B**: Superimposed structures for R = 3-(4-ethyloxazol-2-yl, 1,5-substituted triazole) (cyan) and R = 3-(*N*-ethyl carboxamide), 1,5-triazole (orange). It can be seen that the alkylated amide and the ethyloxazole are acting as bioisosteres for the ethyl ester, filling virtually the same space, with the rest of the molecules being superimposed and having the same hydrogen bond network with the Arg dyad. **C**: Superimposed structures for R = 3-(4-ethyloxazol-2-yl, 1,5-substituted triazole) (cyan) and compound **56** (green). Interestingly, the ethyl carboxylate of **56** and the carboxylate of the light blue structure occupy almost the same space. However, the ester group is rotated upwards so that it can fill the pocket between R483 and F478, although preventing by this the formation of a second hydrogen bond to R415 compared to the newly designed compound. On the opposite side of the molecules, the ethyl group on the oxazole and the methylsulfide moiety both occupy the pocket between Y335 and R380.

4.4.2 Second screening: Virtual structure activity relationship of the azide-bearing moieties

The second screening effort was aimed at exploring potential better substituents on the azide-bearing ring. Since it was known from the previous experiments, that 1,2,3-triazoles made from ethyl 4-propiolamidobenzoate **85** had activity, this was chosen as the alkyne-bearing moiety. Using again AutoClickChem[71], this alkyne was reacted *in silico* with a set of 49 azides, resulting in a library of 98 1,4- and 1,5-disubstituted triazoles. For this part of the molecule, less was known about the tolerance of different substituents. Initially, azides that could be derived from commercially available anilines were investigated. These carried simple alkyl substituents (Me, Et, *n*-Pr, *i*-Pr, *t*-Bu, CF_3) and their respective ethers, hydroxyl, nitro, carboxamide or carboxylate groups. All respective ortho-, meta- or para-regioisomers were considered. Additionally, a phenyl substituent without any further substitution was included (Figure 47). Figure 48 gives an overview of the workflow for this screen.



85 **101** **102**

Figure 47: *in silico* reaction of ethyl 4-propiolamidobenzoate **85** with azide library **101** to yield 1,4- and 1,5-substituted 1,2,3-triazoles **102**. R = Me, Et, *n*-Pr, *i*-Pr, *t*-Bu, CF_3, OMe, OEt, O*n*-Pr, O*i*-Pr, O*t*-Bu, OCF_3, NO_2, CO_2^-, $CONH_2$, OH, H, at each the ortho-, meta-, and para-position.

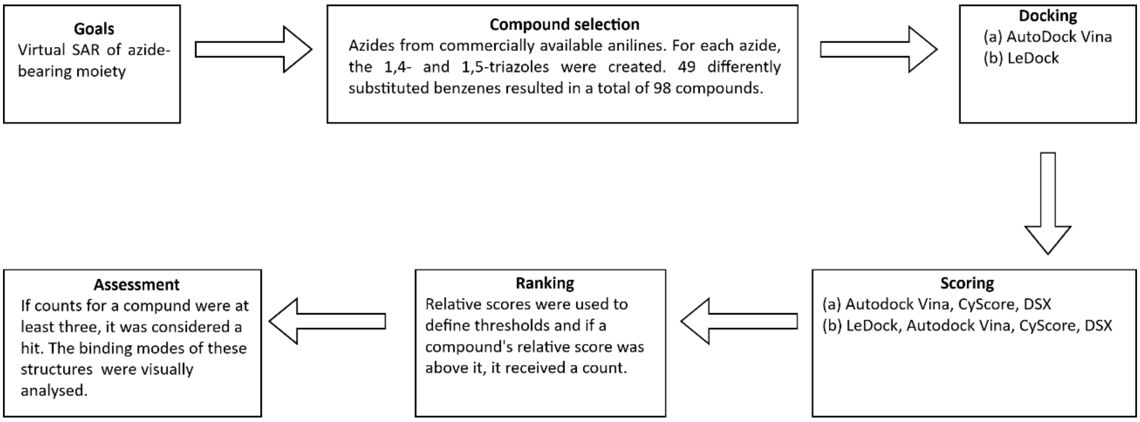
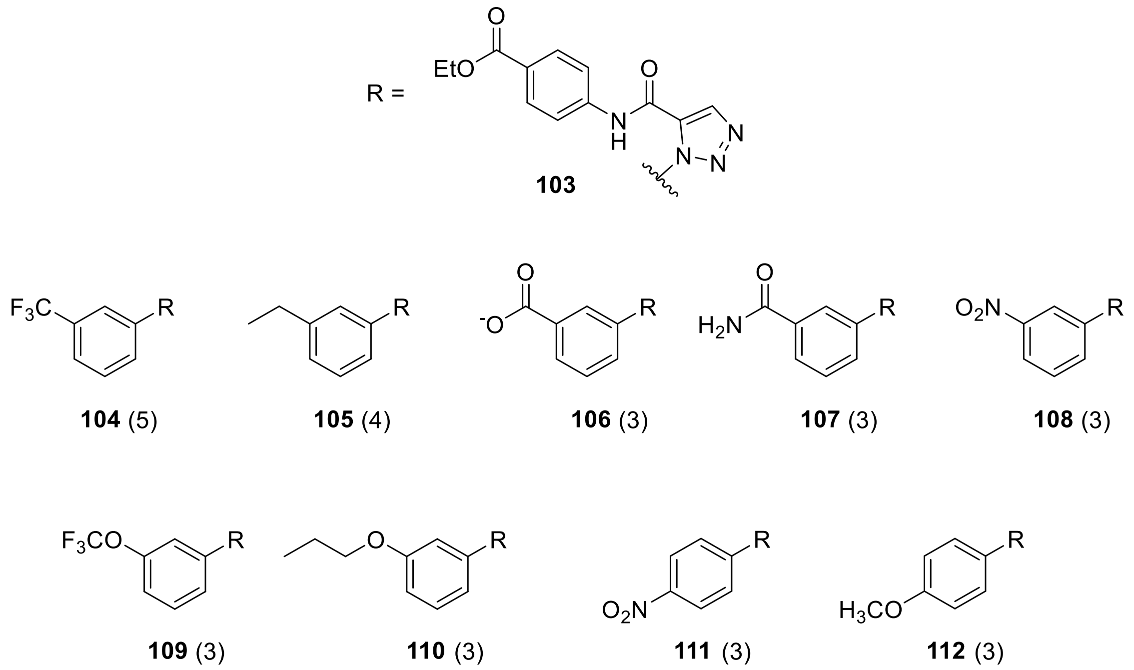


Figure 48: Workflow of the second virtual screen aimed at determining promising substituents on the azide. The compounds were docked using AutoDock Vina and LeDock, scored with three or four, respectively, different scoring functions and the compounds received a count if their relative score was above a threshold for an individual docking and scoring function combination. If the number of their counts was three or higher, they were considered a virtual hit in this screen and their docking pose was analysed visually.

The results were analysed visually and quantitively as in the previous screening. As for the previous screen the threshold for the counts for a compound to be considered a virtual hit, was set to three. Nine structures fulfilled this criterion. Seven had a meta-substituent, whereas two had a para-substituent. All nine structures were 1,5-subsituted triazoles. In detail, the structures with the following R on **103** were considered as virtual hits (in parentheses: counts): 3-CF$_3$(5) **104**, 3-Et (4) **105**, 3-CO$_2^-$ (3) **106**, 3-CONH$_2$ (3) **107**, 3-NO$_2$ (3) **108**, 3-OCF$_3$ (3) **109**, 3-O*n*-Pr (3) **110**, 4-NO$_2$ (3) **111**, 4-OMe (3) **112** (Figure 49).



103

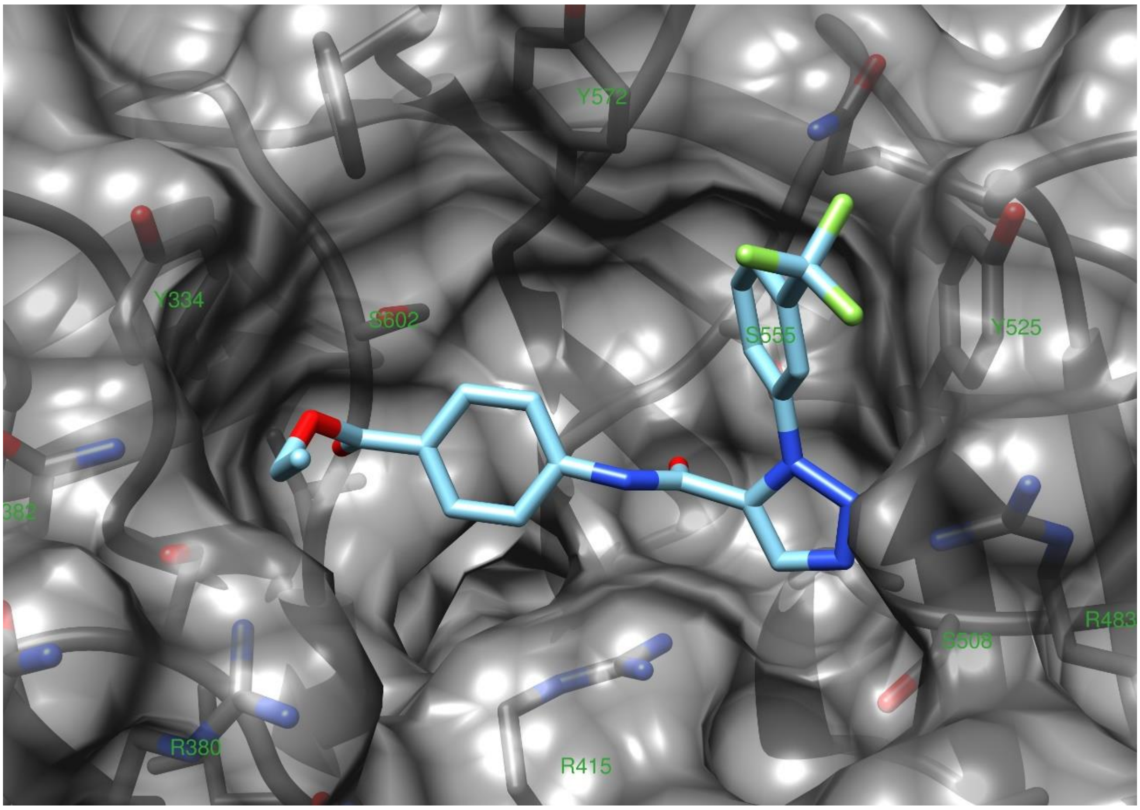
104 (5) **105** (4) **106** (3) **107** (3) **108** (3)

109 (3) **110** (3) **111** (3) **112** (3)

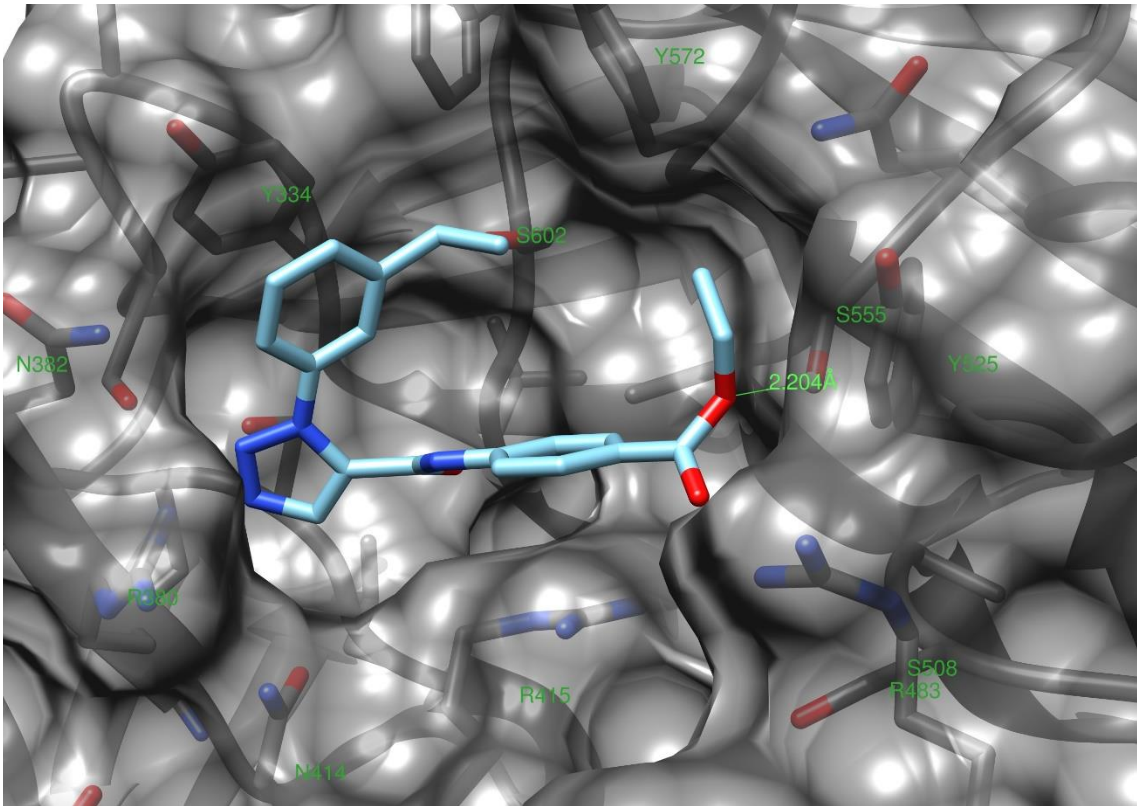
Figure 49: Virtual hit structures **104** – **112** from the virtual screening aiming to establish promising substituents on the azide bearing ring. In parentheses: number of counts.

In contrast to the previous screen, a clear preference for one type of substituent position and type of triazole could be observed. The reason for this appears to be that it is favourable for the 1,5-substituted triazole to be located in the pocket between Arg415 and Arg483, which is only possible if the substituents are located *ortho* to each other and therefore enables the molecule to adopt a shape that is less linear (Figure 50).

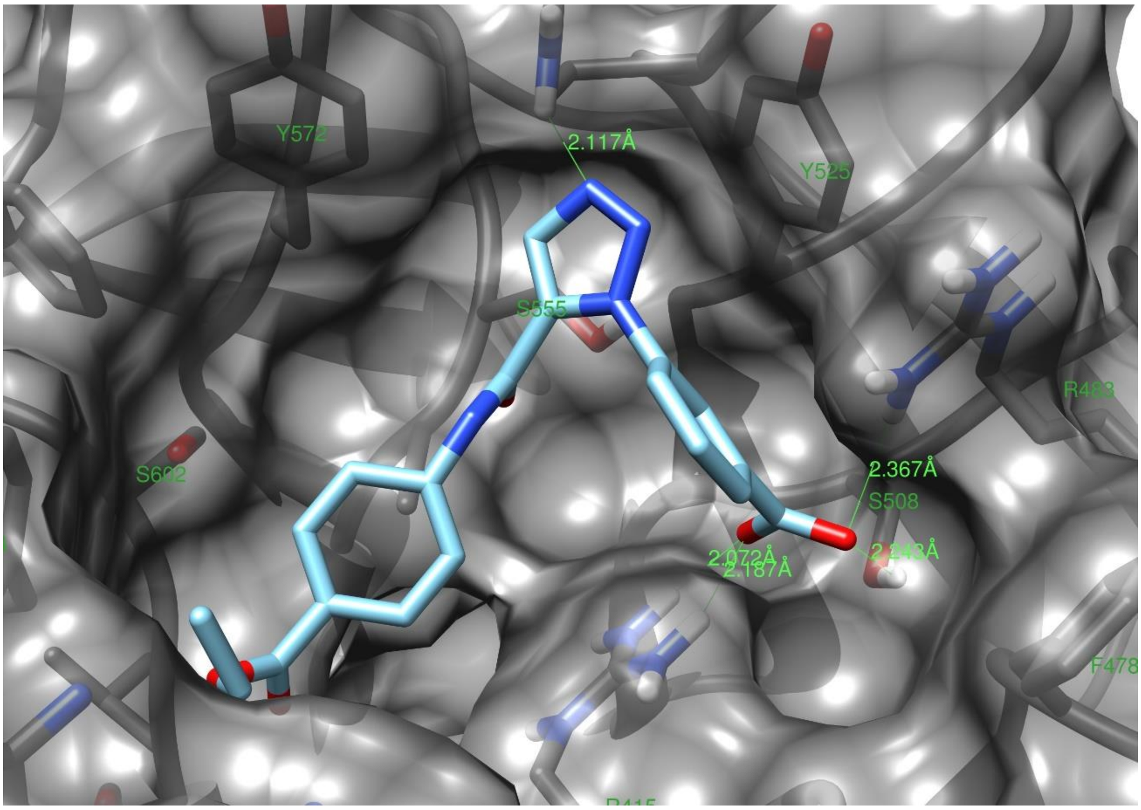
The particularly good performance of the trifluoromeyl-substituted compound **104** might be due to a strong π-π-stacking with Tyr525, due to the stronger interaction between the electron-deficient (Ar-CF_3) and the electron-rich aromatic system (Ar-OH) of tyrosine (Figure 50 A). Ethyl groups, this time of the esters though, mostly occupied the pocket between Tyr334 and Arg380 (Figure 50 A, C). However, this was not always the case (Figure 50 B).


Y572
Y334
S602
S555
Y525
382
R483
S508
R380
R415

A


Y572
Y334
S602
N382
S555
2.204Å
Y525
R380
R415
S508
R483
N414

B

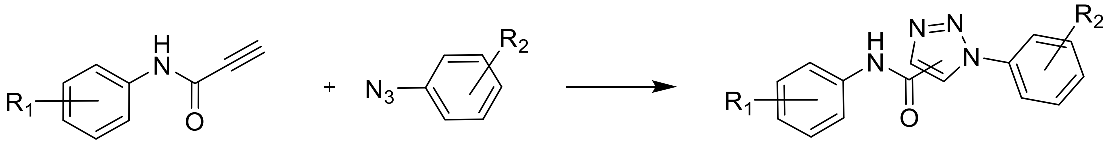


C

Figure 50: All shown structures were docked with Autodock Vina into the Keap1 Kelch domain from 3zgc (grey), and are from the 1,5-substituted 1,2,3-triazole set of scaffold **102**. **A**: Docked structure for R = 3-CF_3 **104** (cyan). The alignment of the trifluoromethylphenyl moiety with the Tyr525 side chain can be seen as well as the impossibility of this arrangement with a 1,4-substituted triazole (both right side). Furthermore, the filling of the pocket between Tyr334 and Arg380 by the ethyl group can be seen (left side). **B**: Docked structure for R = 3-Et (cyan) **105**. The orientation is rotated by ~180° compared to **102** (A) and **106** (C). A probable reason is that in this alignment, all the surface can be used for making contact. If compound **105** would be positioned as structure **104** in A, it could be expected that it would reach out of the binding pocket into the solvent. Considering the lipophilicity of an alkyl group, this should be highly unfavourable. **C**: Docked structure for R = 3-CO_2^- **106** (cyan). Here again, we can see again a different arrangement, although it resembles the one seen in A. The ethyl ester occupies the same space as in A, only the location of the phenyl ring and the triazole are swapped. This can be easily explained by the fact that the carboxylate is able to form strong hydrogen bonds the Arg415 and Arg483 (in total three), compared to the uncharged triazole or trifluoromethoxy moiety. Additionally, it forms one more hydrogen bond to Ser508. Hence, it is more favourable to place the acid near the charged arginine side chains to exploit this effect.

4.4.3 Third screening: Combination of most promising moieties from first and second screenings

The goal of this screening effort was to combine the alkyne and azide bearing groups of the selected triazoles from the two previous screening rounds. On one hand, this included eight alkynes **113**. It should be noted that although 10 structures were considered a hit in the first screen, four of them were from pairs of 1,4- and 1,5-substituion regioisomers and were derived from the same alkyne. The nine azides **114** given rise to the triazoles considered being a hit in the previous screen were also included. Using AutoClickChem[71], this resulted in a library of 144 triazoles **115** (Figure 51). Additionally, the five reference structures mentioned in section 4.4.1 were evaluated again (Figure 43). Hence, a total of 149 compounds were docked. The workflow is summarised in Figure 52.



113 **114** **115**

Figure 51: *in silico* reaction of alkyne library **113** with azide library **114** to yield 1,4- and 1,5-substituted 1,2,3-triazoles **115**. R_1 = 3-(4-ethyloxazole-2-yl), 2-(4-ethyloxazole-2-yl), 3-(1*H*-1,2,4-triazole-5-yl), 3-(oxazole-2-yl), 2-(oxazole-2-yl), 2-(1H-1,2,4-triazole-5-yl), 3-(*N*-ethyl carboxamide), 2-trifluoromethoxy. R_2 = 3-CF_3, 3-Et, 3-CO_2^-, 3-$CONH_2$, 3-NO_2, 3-OCF_3, 3-O*n*-Pr, 4-NO_2, 4-OMe.

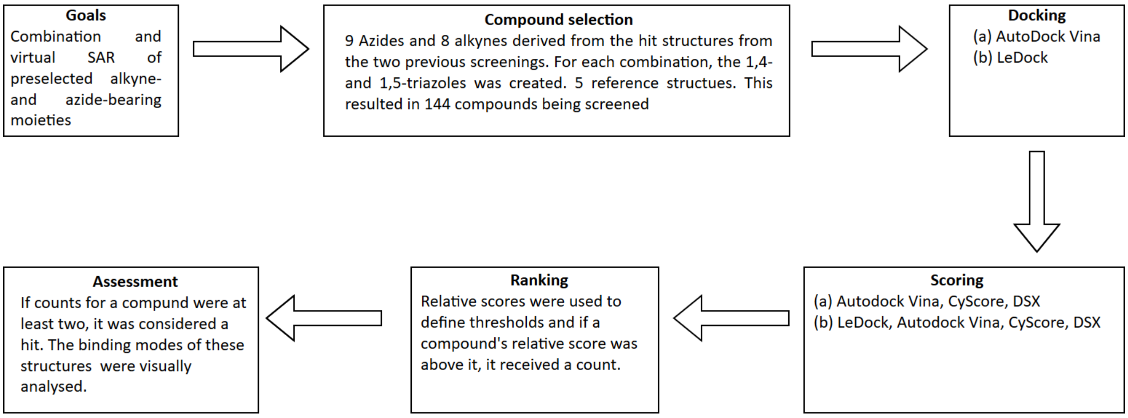
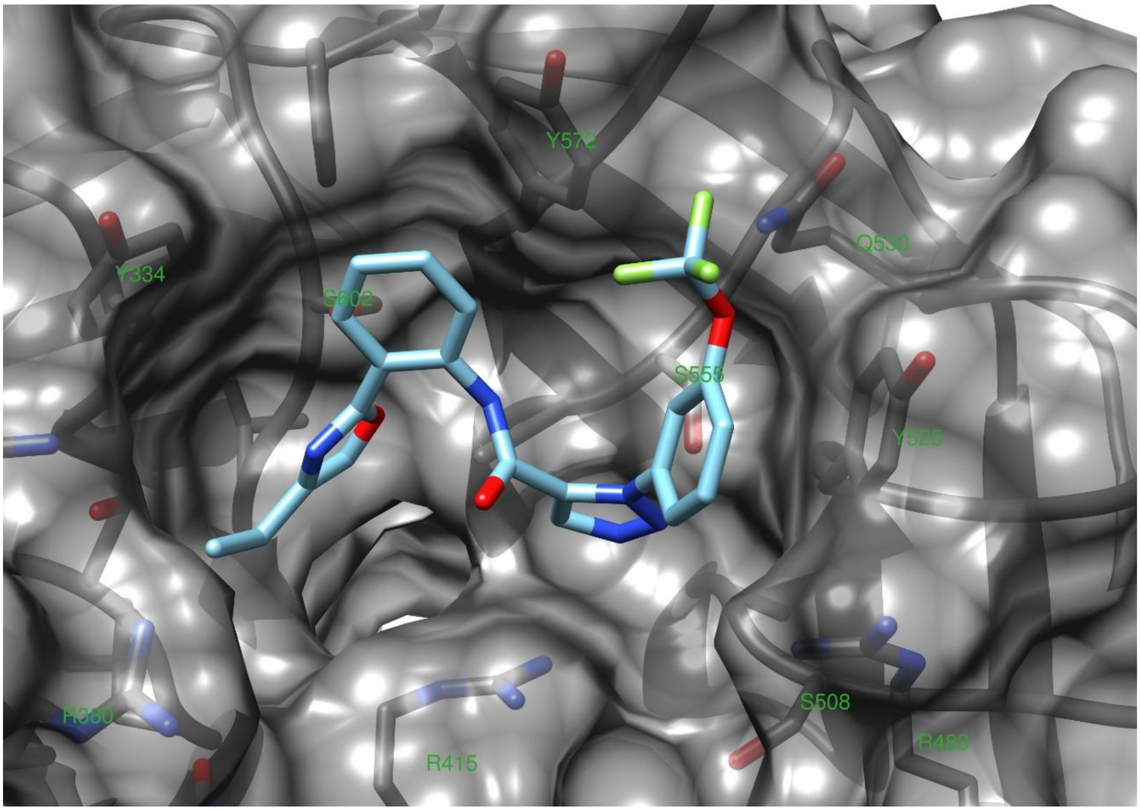


Figure 52: Workflow of the third virtual screen aimed at determining promising combinations of substituents on the alkyne- and azide-bearing moiety. The compounds were docked using AutoDock Vina and LeDock, scored with three or four, respectively, different scoring functions and the compounds received a count if their relative score was above a threshold for an individual docking and scoring function combination. If the number of their counts was two or higher, they were considered a virtual hit in this screen and their docking pose was analysed visually.

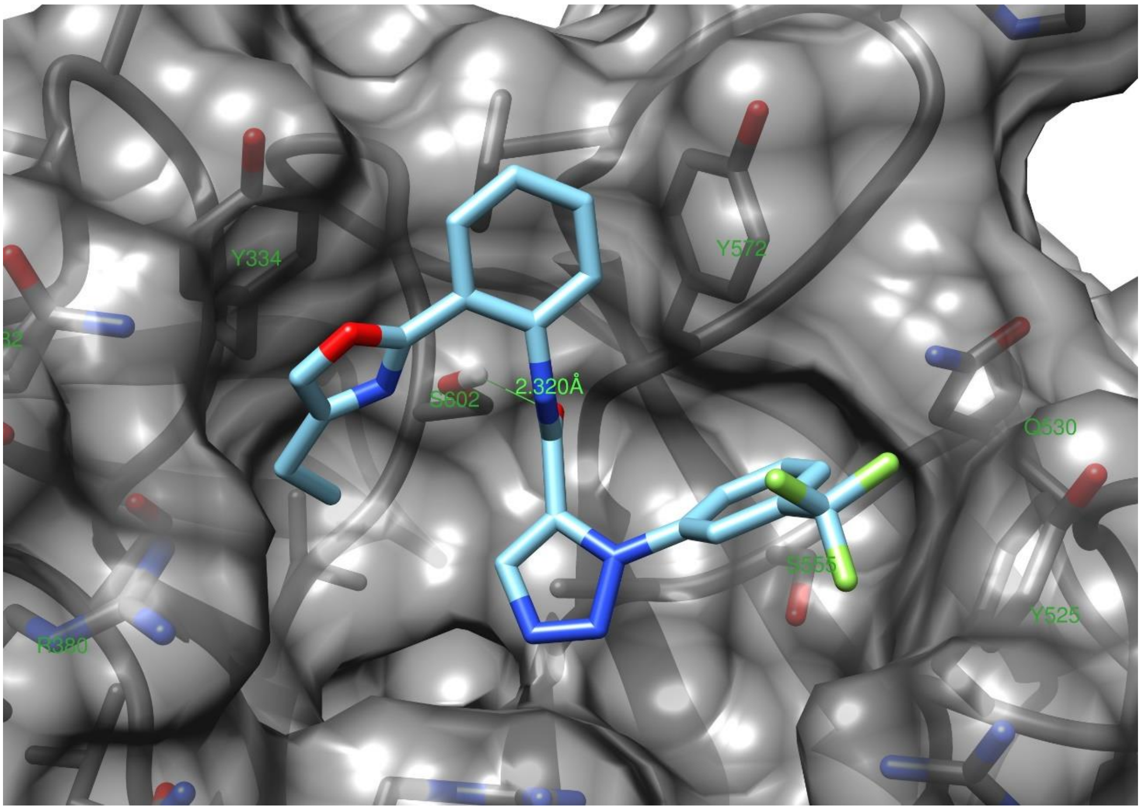
The results were analysed visually and quantitively as in the previous screenings. The threshold for the counts was set to two, as only one structure had a count of three, no reference structure had a count superior to zero. It is important though to keep in mind that the count is a relative measure and a lower count does not correlate with a lower predicted score. This library was preselected in terms of predicted binding affinity; therefore, it can be expected that the score of the different structures will be closer than in the previous libraries. As a matter of fact, the average Vina scores after AutoDock Vina docking, for example, of libraries **87** and **102** were -8.37 and -7.96, respectively, whereas it was -8.69 for library **115**, showing that the strategy of combining hit fragments was successful. In total, 10 structures were considered virtual hits. Eight had a 1,5-, whereas two had a 1,4-substituted central 1,2,3-triazole. In all 10 cases, the phenyl ring from the azide reactant **114** was 1,3-substituted, with the selected substituents being mostly the carboxamide (three times), the trifluoromethyl group (two times) or the trifluoromethyl ether (three times). Furthermore, one structure each contained an isopropyl ether and a nitro group.

In four cases, the phenyl ring of **113** had an *ortho*- and in six cases a *meta* substitution pattern. Exclusively the heterocyclic substituents were selected, especially the 4-ethyloxazole ring (five times), but as well the oxazole (three times), and the 1,2,4-triazole (twice). In detail, the following structures were considered a hit (in parentheses: central 1,4- or 1,5-triazole/counts): R_1 = 2-(4-ethyloxazole-2-yl) R_2 = 3-OCF_3 (1,5/3) **116**, R_1 = 3-(1*H*-1,2,4-triazole-5-yl) R_2 = 3-CF_3 (1,5/2) **117**, R_1 = 2-(4-ethyloxazole-2-yl) R_2 = 3-CF_3 (1,5/2) **118**, R_1 = 3-(4-ethyloxazole-2-yl) R_2 = 3-$CONH_2$ (1,5/2) **119**, R_1 = 3-(1*H*-1,2,4-triazole-5-yl) R_2 = 3-$CONH_2$ (1,5/2) **120**, R_1 = 2-(oxazole-2-

yl) R_2 = 3-CONH2 (1,5/2) **121**, R_1 = 3-(4-ethyloxazole-2-yl) R_2 = 3-NO_2 (1,5/2) **122**, R_1 = 3-(oxazole-2-yl) R_2 = 3-OCF_3 (1,5/2) **123**, R_1 = 2-(oxazole-2-yl) R_2 = 3-OCF_3 (1,4/2) **124**, R_1 = 3-(4-ethyloxazole-2-yl) R_2 = 3-O*n*-Pr (1,4/2) **125** (Figure 54). Figure 53 gives two examples of docked poses of two hits, with R_1 = 2-(4-ethyloxazole-2-yl) R_2 = 3-OCF_3 **116** (A), and R_1 = 2-(4-ethyloxazole-2-yl) R_2 = 3-CF_3 **117** (B).

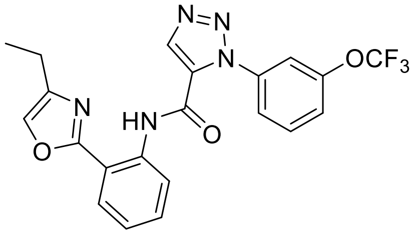


A

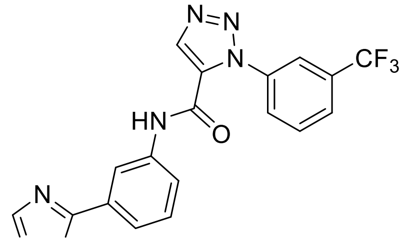


B

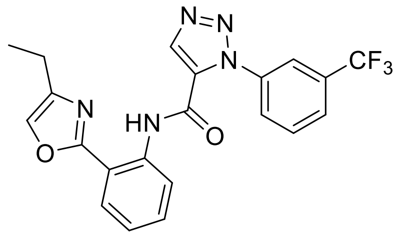
Figure 53: All shown structures were docked with Autodock Vina into the Keap1 Kelch domain from 3zgc (grey). Examples show docked virtual hit compounds with the scaffold **115** and a 1,5-substitution pattern on the central triazole ring. **A**: R_1 = 2-(4-ethyloxazol-2-yl) R_2 = 3-OCF_3 **116**. The trifluoromethoxyphenyl residue forms a T-shaped π-cation interaction with Tyr 525, the ethyl group occupies the pocket between Tyr334 and Arg380. **B**: R_1 = 2-(4-ethyloxazol-2-yl) R_2 = 3-CF_3 **117**. Similarly to A, the trifluoromethylphenyl group forms a π-cation interaction with Tyr 525 and the ethyl group of the oxazole occupies the above-mentioned pocket. Further, the linker amide forms a hydrogen bond to Ser602 which anchors the molecule within the pocket. Additionally, the oxazole ring forms a parallel π-π-stacking with Tyr334.



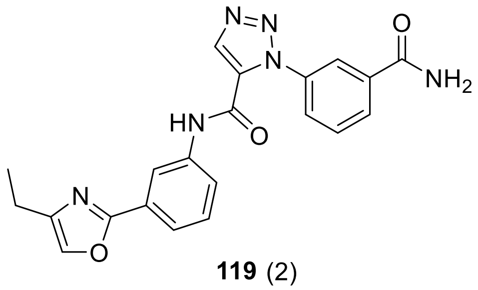
116 (3)



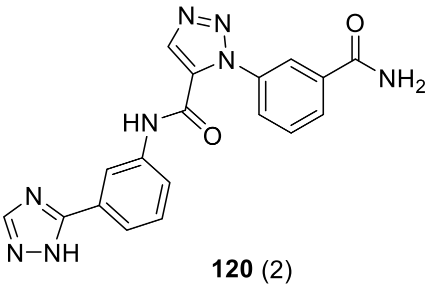
117 (2)



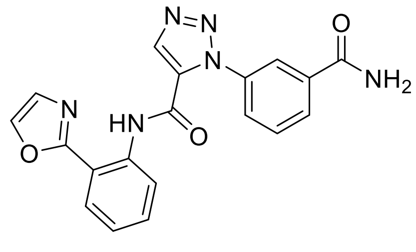
118 (2)



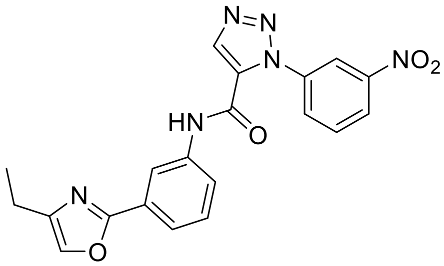
119 (2)



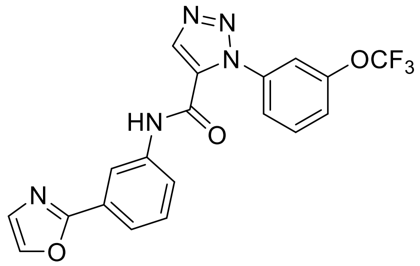
120 (2)



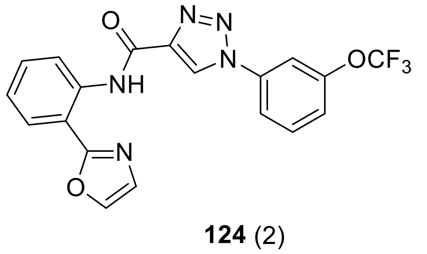
121 (2)



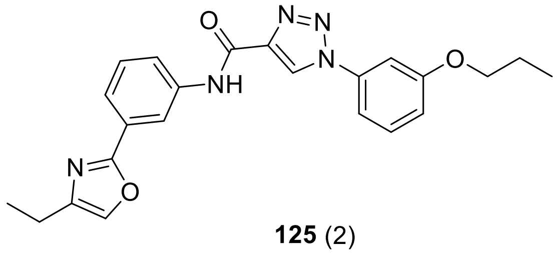
122 (2)



123 (2)



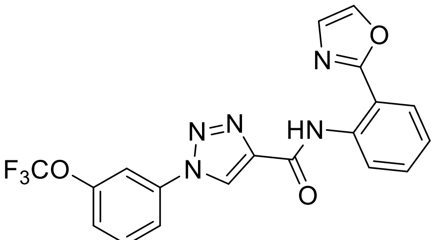
124 (2)



125 (2)

Figure 54: Virtual hit structures **116** – **125** from the virtual screening in which the hit structures from the azide and alkyne screenings were combined. In parentheses: number of counts.

In a possible continuation of the project, the next step would be the synthesis of representative examples of the virtual hits. This should include structure **124** (Figure 55) for two reasons: potential efficiency and ease of synthesis.



124

Figure 55: Potentially first structure **124** to synthesise to verify virtual screening results.

It contains an oxazole moiety, which was the second most frequent group for this position, but is easier to synthesise than the most frequent group, the oxazole with the ethyl group in position four, as this would be needed to be synthesised first, whereas oxazole **126** could be coupled using a Pd-catalysed cross-coupling to a 2-haloaniline **127**.[76] This fragment **128** could then be coupled to propiolic acid **43,** leading to alkyne **129**. Commercially available 3-fluoromethoxyaniline **130** can be easily converted to the respective azide **131** which can then be reacted with **129** to give triazole **124**, using Cu^{I}-catalysis (Figure 56).

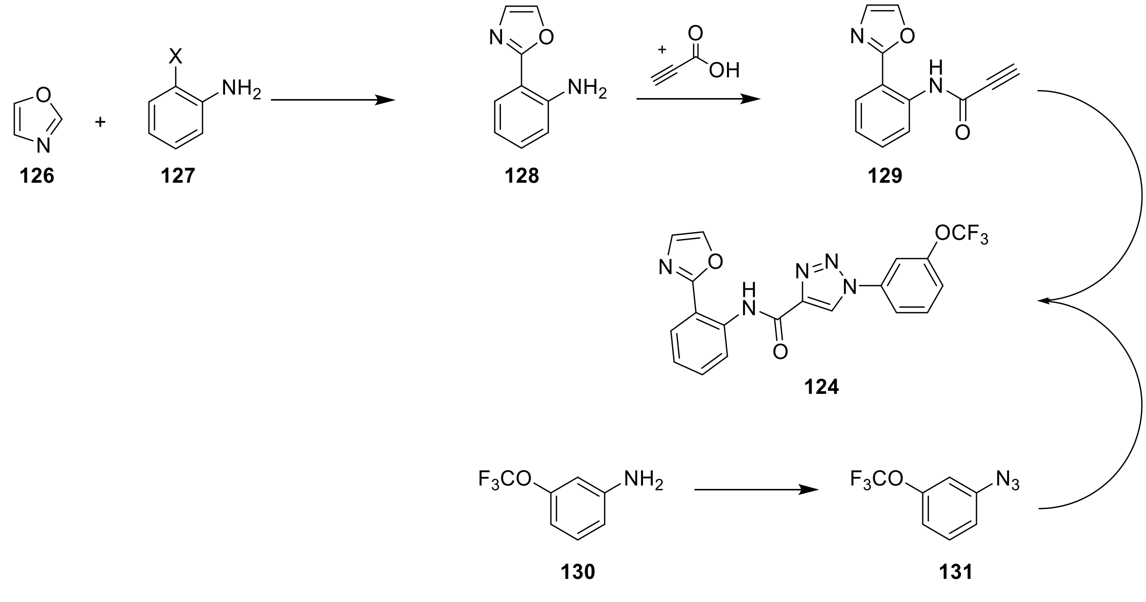


Figure 56: Proposed synthetic route for triazole **92**. X = Cl, Br, or TfO.

4.5 Conclusion

Key structure activity relationships for triazole **56** could be demonstrated. The hydrogen bond donor, such as in the amide of **56**, appears to be crucial as the respective ester analogue is much less active (39% inhibition at a 10 µM inhibitor concentration compared to 100% inhibition). The substituents do not seem to influence the activity significantly, as all of the synthesised analogues (except for **59**) have inhibition rates between 26% and 39% at 10 µM, suggesting that other groups could be added to improve activity or truncated without loss of activity. When the structures were docked, the order of activity could not be reproduced, making it challenging to rationalise the experimentally observed trends. One important aspect seems to relate to the inclusion of **56** in the analysis. Exclusion of this compound largely increased the correlation between computational and experimental rankings, particularly with specific sampling algorithm and scoring function combinations (such as docking and scoring with Autodock Vina[45]).

Additionally, to learn more about how different substituents influence the predicted affinity of this scaffold, virtual structure-activity relationship screening was performed. For this, two libraries with a variety of substituents for each of the phenyl-moieties, being either on the alkyne or the azide, were first investigated separately (with one part of the triazole being varied, the other one fixed). The most promising substituted benzenes out of these libraries were selected and combined with each other. This gave rise to the interesting triazole **124**, a promising virtual hit structure and whose synthetic route has been outlined.

5 Discussion and Future Directions

Herein, we describe two different approaches to developing small molecular inhibitors of the Keap1-Nrf2 PPI suitable for applications in neurodegenerative diseases.

First, through several iterations of VS a potential lead scaffold has been developed along with the validation of the VS protocol for this application. With further rounds of VS, SAR were studied and a desirable target compound **39** identified. Synthesis of the building blocks required to make **39** was performed successfully. Next steps will include the completion of the synthesis and testing of **39** and its component building blocks in our established competitive FP assay.[52] In parallel, virtual screening results could be verified with other computational methods as mentioned in Chapter 5.3.

Second, an extensive series of *in silico* variations to the substituent patterns on a 1-phenyl-4-*N*-phenylamido-1,2,3-triazole scaffold was generated. The compounds were docked to the Keap1 Kelch domain. Combined with the synthesis and their testing of selected analogues in our competitive FP assay[52], a (virtual) SAR of this scaffold was described and a promising candidate **124** for further work was proposed. Computational validation (Chapter 5.3), synthesis (Chapter 4.4.3) and testing it in the FP assay[52] will be the immediate next steps.

Compound **124** is particularly interesting as a probe or potential therapeutic for neurodegenerative diseases. It is based on a scaffold which shows promising Keap1 binding activity. It is relatively straightforward to synthesise, allowing experimental verification of computational results. Furthermore, although it does not possess fully optimised physicochemical properties for a CNS drug, it fulfils or almost fulfils many of the desirable criteria except for the number of hydrogen bond donors (Table 22).

Property	Compound 124	Properties of CNS drugs[35,36]	Complies?	Rules for oral bioavailability[37,75]	Complies?
HBD	1	≤ 2	✓	≤ 5	✓
HBA	8	≤ 2 - 4	✗	≤ 10	✓
MW	415.3 g.mol^{-1}	≤ 450 g.mol^{-1}	✓	≤ 500 g.mol^{-1}	✓
clogP	2.55	≤ 3	✓	≤ 5	✓
tPSA	87.9 Å^2	≤ 80 Å^2	✗	≤ 140 Å^2	✓
Rotatable bonds	4	≤ 3	✗	≤ 10	✓

Table 22: physicochemical properties of compound **124** and the compliance of these with desirable drug-like properties

The next steps of the project will include critical discussion and validation of obtained results as well as further compound optimisation and biological evaluation to move towards obtaining a blood-brain barrier-permeable molecular probe for applications in neurodegenerative diseases. Along with the computational and experimental validation of the affinity measurements, this will include testing for membrane diffusion and in particular for blood-brain barrier penetration as well as the investigation of compounds that show promise in cellular and animal disease models. Several of these aspects will be highlighted in the following paragraphs.

5.1 Water in Virtual Screening

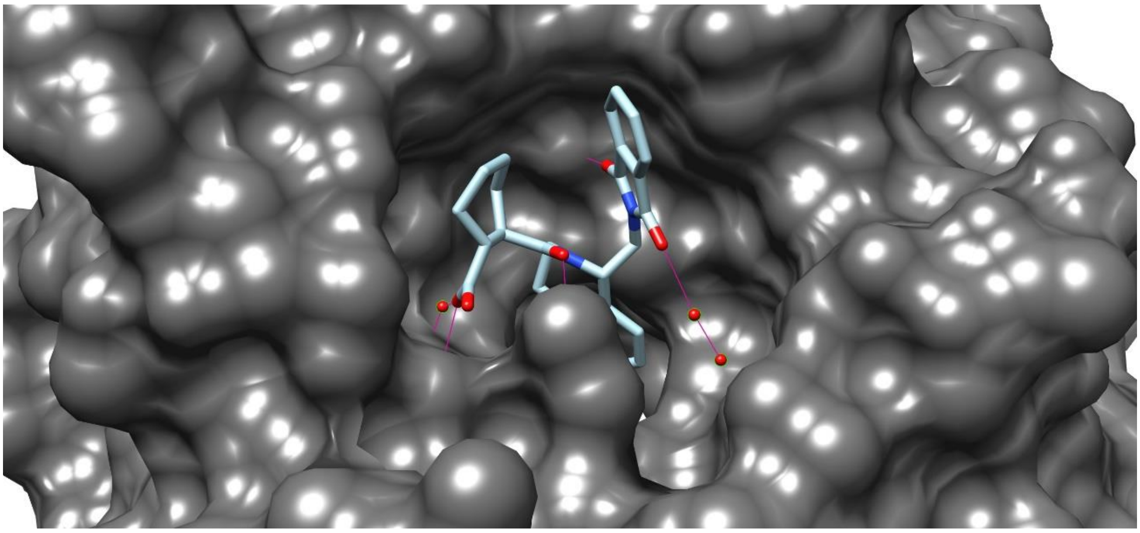
5.1.1 Water solvation

Different scoring functions consider solvation of ligands or proteins differently. For example, the UCSF DOCK[43] Grid Score does not consider water solvation effects. In contrast, Autodock Vina's scoring function contains a term for hydrophobic interactions, implicitly taking desolvation into account with a distance-dependent weighing factor for a hydrophobic term. The term decreases proportionally between 0.5 and 1.5 Å; beyond these values the term is kept constant.[45] The 1.5 Å distance corresponds roughly to the size of a water molecule. Similarly, LeDock has a term that contributes (among others) for desolvation.[44] rDock takes also the desolvation into account with a standalone, distance-dependent term combined with a solvent accessible surface-based approach.[42] As for the other programs, this only considers the effect of desolvation of the ligand and the binding site implicitly.

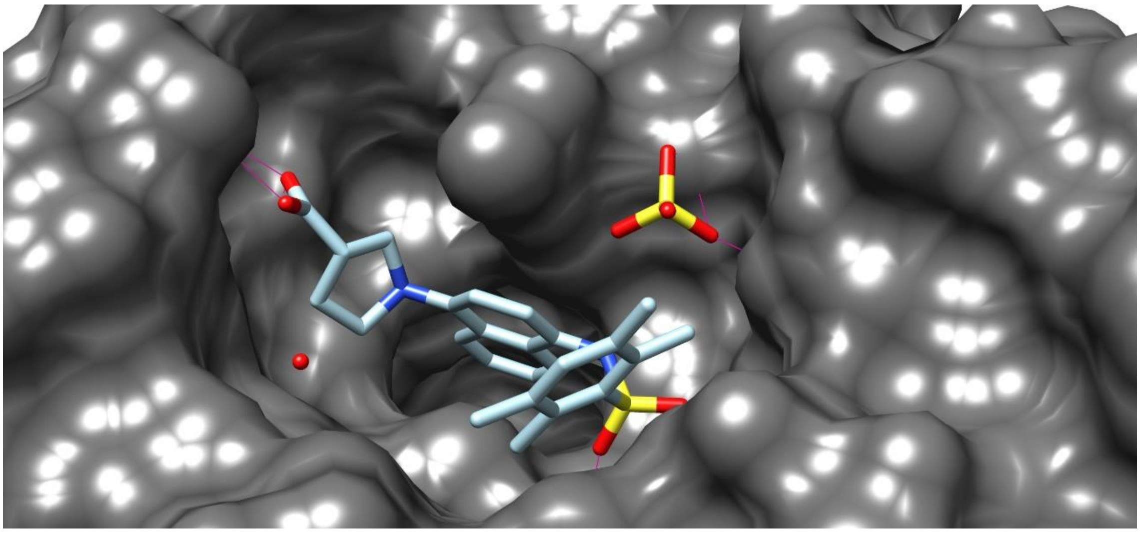
5.1.2 Water and the Keap1 binding pocket

It is recognised that water molecules can play an important role in ligand binding and can influence interactions in several ways.[89,90] First, it is possible that the water molecules stay within the binding pocket upon ligand binding, allowing polar interactions otherwise impossible by bridging atoms of the two binding partners. Second, water molecules can be displaced.[89,90] This can have two opposing effects on the enthalpy and entropy of the binding event: on one hand, this can disfavour binding by reducing the enthalpy of this interaction. That is the case if the ligand that displaced the water is not able to engage in similarly strong polar interactions with the binding partner, and the hydrogen bonds formed by the released water molecule with the bulk water molecules are weaker than the interaction the water molecule formed with the macromolecule. However, as hydrogen bonds are the most important type of interactions which water molecules form, and their strength is usually very similar within a same environment[90], the change in enthalpy is usually rather low as the broken and formed hydrogen bond have a similar strength. In contrast to that, the entropic effects of water displacement are known to be important contributors to binding affinity.[89,90] Due to the surface character and sometimes high lipophilicity of binding pockets, water molecules found in these sites are highly structured. If these water molecules are released, there is an increase in degrees of freedom and therefore a positive entropic change with a beneficial contribution to ligand binding.[90]

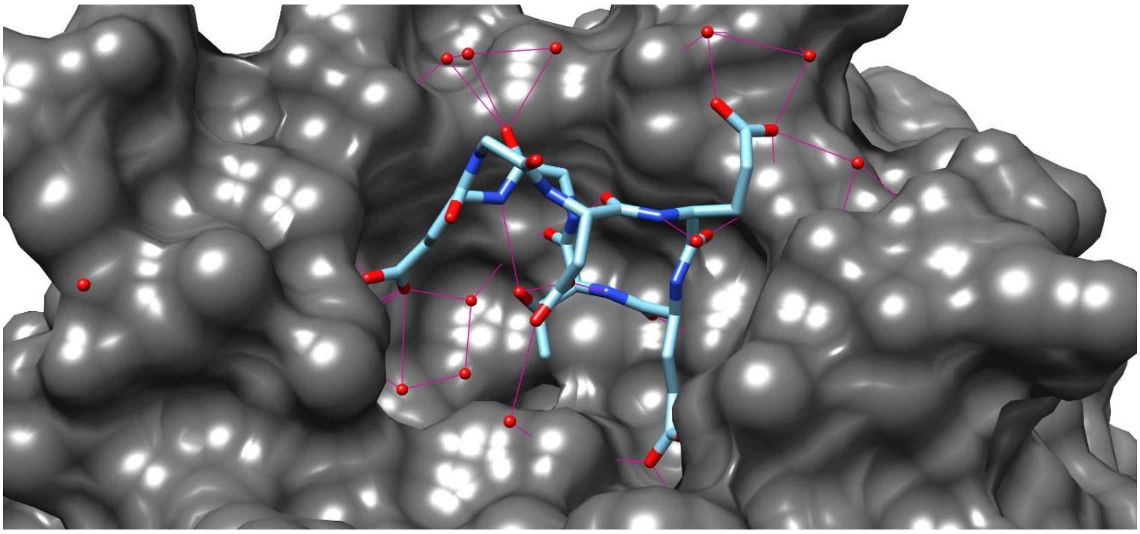
When the Keap1 crystal structures containing the Kelch domain from the PDB database[53] were investigated to determine whether they contained water molecules in the binding pocket, this was unsurprisingly the case for most of them, including some ligand-bound ones, as the pocket is not very deep, hence rather solvent-exposed (Figure 57). The positions of the water molecules are not conserved and therefore appear to be part of rather flexible bulk water rather than structurally relevant water (Figure 57). Nevertheless, some Keap1-Nrf2 PPI inhibitors interact with water in the binding pocket to form hydrogen bonds involving the ligand, water and Keap1 (Figure 57A). Other small molecular ligands, however, have no interacting water present in the binding pocket (Figure 57 B). A heptameric peptide based on Nrf2's ETGE sequence cocrystallised with the Kelch domain also shows the presence of water in and near the binding site. These do not appear to be central for interactions as the electrostatic and hydrogen bond interactions of the key Glu residues are formed directly with Keap1 and not through the intermediate of water (except for one case, where both seems to be the case; Figure 57 C).



A



B



C

Figure 57: Crystal structures of cocrystallised Keap1 Kelch domains with small molecule (**A**, pdb code 4ifn; **B**, pdb code 5cgj) and peptide (**C**, pdb code 3zgc) ligands. The protein is shown in grey and the ligands in light blue, hydrogen bonds in pink. Water molecules are shown with their oxygen atoms in red. **A**: three water molecules are found in the binding pocket, where one seems to mediate a hydrogen bond from the carboxylic acid whereas the remaining two are aligned in a linear fashion, enabling an imide oxygen to form a hydrogen bond *via* these two water molecules. **B**: one water is located in the binding pocket, although in a different place than in **A**. This molecule does not interact with the protein nor with the ligand. **C**: several water molecules can be found in the binding pocket. These do not appear to play a structural role as they are either exposed to the bulk solvent or in the wider, more open part of the pocket.

Given the different number of water molecules and their varying positions at the molecular interface, none of them appears to be important as there is not one specific water molecule which would is of general relevance for interactions. Given that, any water displacement will contribute positively to binding due to entropic contributions as long as the enthalpic effects between ligand and protein on one hand and water and protein on the other hand are similar. It would be a challenging task to model the relevance of water explicitly given that complexity, although there are efforts to develop force fields which try to take water effects into consideration and predict their contributions.[91]

5.2 Murine and Human Keap1

The human and murine Keap1 proteins are largely homologous with a sequence identity of 94.1%[88], the Kelch domains (residues 315 -598) being even more similar with a 96.1% identical sequence (Figure 58). The majority of the sequence differences involve amino acids having similar properties (*e.g.* a basic residue being replaced with another basic amino acid), such as K549R (from human to murine).

```
1   MQPDPRPSGA GACCRFLPLQ SQCPEGAGDA VMYASTECKA EVTPSQHGNR TFSYTLEDHT
1   MQPEPKLSGA PRSSQFLPLW SKCPEGAGDA VMYASTECKA EVTPSQDGNR TFSYTLEDHT

61  KQAFGIMNEL RLSQQLCDVT LQVKYQDAPA AQFMAHKVVL ASSSPVFKAM FTNGLREQGM
61  KQAFGVMNEL RLSQQLCDVT LQVKYEDIPA AQFMAHKVVL ASSSPVFKAM FTNGLREQGM

121 EVVSIEGIHP KVMERLIEFA YTASISMGEK CVLHVMNGAV MYQIDSVVRA CSDFLVQQLD
121 EVVSIEGIHP KVMERLIEFA YTASISVGEK CVLHVMNGAV MYQIDSVVRA CSDFLVQQLD

181 PSNAIGIANF AEQIGCVELH QRAREYIYMH FGEVAKQEEF FNLSHCQLVT LISRDDLNVR
181 PSNAIGIANF AEQIGCTELH QRAREYIYMH FGEVAKQEEF FNLSHCQLAT LISRDDLNVR

241 CESEVFHACI NWVKYDCEQR RFYVQALLRA VRCHSLTPNF LQMQLQKCEI LQSDSRCKDY
241 CESEVFHACI DWVKYDCPQR RFYVQALLRA VRCHALTPRF LQTQLQKCEI LQADARCKDY

301 LVKIFEELTL HKPTQVMPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSDG TWLRLADLQV
301 LVQIFQELTL HKPTQAVPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSNG SWLRLADLQV

361 PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT NQWSPCAPMS VPRNRIGVGV
361 PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT NQWSPCASMS VPRNRIGVGV

421 IDGHIYAVGG SHGCIHHNSV ERYEPERDEW HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG
421 IDGHIYAVGG SHGCIHHSSV ERYEPERDEW HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG

481 TNRLNSAECY YPERNEWRMI TAMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE
481 TNRLNSAECY YPERNEWRMI TPMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE

541 TETWTFVAPM KHRRSALGIT VHQGRIYVLG GYDGHTFLDS VECYDPDTDT WSEVTRMTSG
541 TETWTFVAPM RHHRSALGIT VHQGKIYVLG GYDGHTFLDS VECYDPDSDT WSEVTRMTSG

601 RSGVGVAVTM EPCRKQIDQQ NCTC
601 RSGVGVAVTM EPCRKQIDQQ NCTC
```

Figure 58: Multiple sequence alignment from Uniprot[88] of human (top line) and murine (bottom line) Keap1 Kelch domains. Amino acid changes that lead to a change in polarity or the introduction of a proline are highlighted in red, other changes are highlighted in yellow. The Keap1 Kelch domain is comprised of residues 315-598.

In accordance with the sequential data, an overlay of a human and a murine Keap1 Kelch domain crystal structure (3zgc and 3wn7, respectively) reveals that the three-dimensional arrangement of both protein domains is very similar with an RMSD of 0.39 Å (Figure 58). The only loops near the Nrf2 binding site that differ are 383-388 and 526-529 (Figure 58B). However, these are located outside of the ligand pocket and are not involved in binding, although the important residue R380 is located in relatively close proximity.

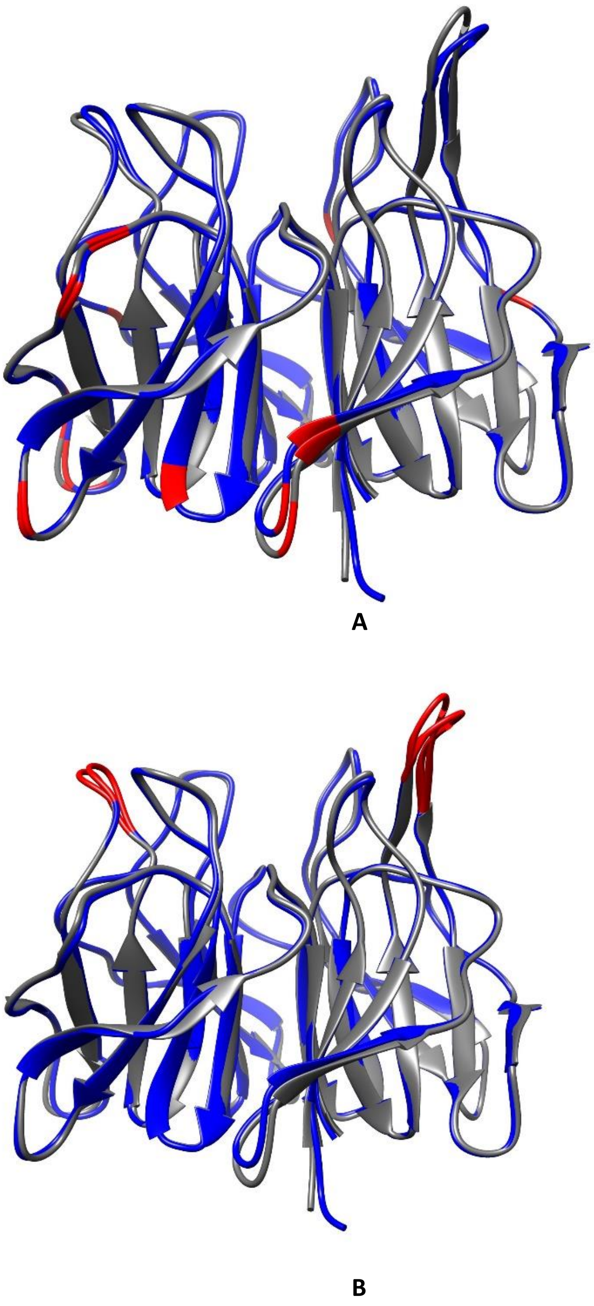


Figure 58: Overlay of crystal structures of human (3zgc, dark grey) and murine (3wn7m, blue) Keap1 Kelch domain using Chimera[41]. Differing amino acids are highlighted in red (**A**) as well as loops 383-388 and 526-529 (**B**) which differ in their spatial arrangement between both structures. The Nrf2 binding site is located at the top.

Taken together, this suggest that slight differences between the human and murine Keap1 Kelch domain exist. However, these differences do not drastically change the binding site. This is supported by the fact that ligand structures developed using the mouse protein show activity in human cells[34].

To validate the hypothesis, the highest ranked structures which have been docked on human Keap1 structures could be docked into murine Keap1 structures. This could help to establish whether compounds developed for the human form of Keap1 would be expected to show activity in mouse models, an important question when considering in vivo disease models. Alternatively, an easily set up assay, such as our competitive fluorescence polarisation assay[52], could be performed with human and murine Keap1 to evaluate if the binding affinities of the synthesised compounds differ between the target structures.

5.3 Computational hit validation

As for any assay, it is crucial to validate and cross-check the results of virtual screening with other methods. This can be achieved with computational methods as well as with wet lab-based methods.

An important way of validating a virtual screening approach is to run the protocol against a set of known ligands in order to see if the experimental ranking of compounds, based on their target affinities can be reproduced.[92] This pivotal method was successfully employed (s. Chapter 3.1.2.1) and to a certain extent, the thorough and comprehensive analysis of available human Keap1 Kelch domain crystal structures (Chapter 3.1.2) can be considered being a part of this approach.

Further computational validation methods can include molecular dynamics (MD), steered MD and free energy calculations.

5.3.1 Molecular Dynamics

Molecular Dynamics (MD) is a computational simulation method in which a system, such as a protein-ligand complex, is allowed to interact for a time in the range of femto- to milliseconds and the changes in the system are observed over this period. [111] The system's evolution can for example be due to changes in potential or kinetic energy. The force fields used to calculate the potential energy and to determine the dynamics of the system are based on Newtonian mechanics. AMBER[110], CHARMM[113] and GROMOS[114] are widely used force fields with a large number of software packages being based on them. All three force fields have the same functional form[112], similarly to force field-based scoring functions, with energy terms for bonded, nonbonded and other interactions. The bonded terms are composed of bond stretching energy, angular and dihedral contributions, whereas the nonbonded term is the sum of a Lennard-Jones (12,6) (LJ) potential accounting for van der Waals-interactions and the electrostatic Coulomb interaction. In detail, the functional form is as follows:[112]

$$E_{total} = E_{bonded} + E_{nonbonded} + E_{other} \text{ with}$$

$$E_{bonded} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi[1 + \cos(n\chi - \sigma)]^2$$

$$E_{nonbonded} = \sum_{\substack{nonbonded \\ pairs\ ij}} \left(\varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 * \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] - \frac{q_i q_j}{r_{ij}} \right)$$

One point that is different in the AMBER[110] force fields compared to the other two widely used force fields is that they account for so-called improper dihedral angles through the dihedral terms, whereas the other force fields have an explicit term for this contribution. The improper dihedral is used to maintain chirality at quaternary centres.[112] Also, AMBER[110] and CHARMM[113] use the geometric mean to determine the LJ parameter $\varepsilon_{i\ j}$, but the arithmetic mean to determine $R_{min,ij}$. In contrast, GROMOS[114] computes both parameters with the geometric mean. Furthermore, they treat differently 1,4-nonbonded interaction. There are different AMBER[110] force fields, optimised for proteins, nucleic acids, lipids, carbohydrates and small organic molecules with a similar setup for the other force fields.[111,114]

Due to the atomistic level at which conformational changes of the macromolecule or the ligand are observed, this methodology has gained interest in the drug discovery community. However, as it is a computationally expensive method, it is frequently used for hit verification and optimisation rather than screening of large compound libraries.[109] A variety of comprehensive, high performance software package are available to perform these experiments even with large biomolecules.

Nanoscale Molecular Dynamics (NAMD) is such a software package, developed by the Schulten Group at the University of Illinois, which allows efficient modelling of large-scale biomolecule interactions with millions of atoms.[100] This has been possible by scaling the required calculations to tens of thousands of processors. It contains as well a graphical molecular programme, allowing a complete modelling environment. Important setup steps such as solvating the protein can be all performed within the same environment. Furthermore, it is as well possible to perform free energy perturbations (s. 5.3.3). Its default force field is the CHARMM[113] force field, but it is as well compatible with AMBER[110] and GROMOS[114] force fields.[100]

5.3.2 Steered Molecular Dynamics

Steered Molecular Dynamics (SMD)[99] has been used for several applications to verify results obtained from other computational methods, especially virtually screening (VS), and to investigate binding modes.[96-99]

The principle of SMD is that a protein-ligand complex is perpetuated with a mechanical force, leading to a dissociation of the ligand from the complex or other change of the system.[96,100] Employed perpetuations can be a constant force or spring-like. Consequently, the strength of the forces or the position of the atoms in the system can be altered.[100] This gives information about the forces involved in the binding of the two partners, as the work employed to vary a system is dependent on the difference in free energy of two states.[96,99,100] One main advantage of SMD over VS is its improved ability to account for receptor flexibility and better consideration of conformational space in general, as a larger number of conformers are explored.[96,99] Although this is something which can be accounted for in VS, it is usually too computationally expensive to be performed on a large scale. Hence, it is a promising approach to select compounds first through VS and investigate the hits more thoroughly using SMD.[96]

5.3.3 Free energy calculations

It has been shown that free energy calculations show better agreement with experimental results than scoring functions[93], although at the expense of a higher computational cost. This would be a useful tool to verify results from the docking. Free energy calculations simulate the difference in free energy between two molecules using an MD simulation during which statistical information is generated.[93] All free energy calculations are based on the equation:

$$G_B - G_A = \Delta G = -RT \ln \langle e^{-\Delta H/RT} \rangle_A$$

where the term in brackets are the ensemble averages of a system, expressed as the Hamiltonian function H, *i.e.* the sum over all positions of a system and their respective momentum. For many applications, the so-called free energy perturbations, states A and B are two end points of a system and the system is varied incrementally in small intervals. The new free energy is then calculated as a difference from the previous state.[94] Due to its thoroughness, it is particularly useful to model small changes in a system, such as transforming one ligand into another. The resulting differences in free energy can then be used to determine which ligand is more likely to have a higher binding affinity.[93,95] In this context, docking results can be validated computationally to see if the rank of the most promising candidates could be reproduced.

Although the equations employed in free energy calculations are precise, the results are not necessarily accurate if the sampling, carried out with MD, or the Hamiltonian are impaired. However, practice has shown that this is usually the case.[93-95]

5.4 Experimental hit validation

Synthesised compounds would be tested for their Keap1 binding affinity with our competitive FP assay.[52] Before moving forward, another affinity measurement method would be employed to verify the FP results. This could be surface-based techniques such as surface plasmon resonance (SPR) or solution-based methods such as isothermal titration calorimetry (ITC) or microscale thermophoresis (MST). MST could be a promising approach as neither Keap1 nor Nrf2 are surface-bound, a fluorescently labelled probe (the Nrf2-derived peptide) is available and therefore, assay setup would be straightforward. Furthermore, as the smaller binding partner is labelled, this is expected to mean higher sensitivity compared to the opposite way of labelling. Another, very interesting assay in solution is the FRET-based assay developed in our group.[101] This would allow direct monitoring of the Keap1-Nrf2-peptide dissociation upon incubation of the complex with the inhibitors.

5.5 Permeability assays

To estimate the permeability of compounds across biological membranes, parallel artificial membrane permeability assays (PAMPA) are widely used.[102] Two compartments, usually filled with buffer, are separated by a filter membrane loaded with lipids. This mimics a cellular membrane. Depending on the artificial membrane used, different biological membranes can be mimicked. For example, the use of hexadecane membranes is widely used as an easy to handle and highly reproducible surrogate for the intestinal barrier.[102] This methodology has also been used to estimate diffusion over the blood-brain barrier using a mixture of dodecane and porcine polar brain lipid as the membrane material.[103] This approach provides an estimate for the passive diffusion, but is unable to represent active uptake or efflux. For this, cellular assays are necessary.

Caco2 cells are the most widely used cell line for the investigations of intestinal absorption.[104] They express most of the intestinal transporters and efflux pumps such as p-glycoprotein and are therefore a good representation of the situation *in vivo*.[104] Similarly, assays using brain endothelial cells, cocultured with astrocytes or alone can be used to estimate the blood-brain barrier penetration.[105]

5.6 Disease models

Finally, if the hit structures have been verified computationally and experimentally and their ability to cross cell membranes and the blood-brain barrier, their value will need to be evaluated in PD models. There are a variety of cellular[106] and animal[24,25,107,108] models described in the literature.

In a first step, the compounds would be tested in cellular models, due to their easier handling and the possibility to more easily focus on one aspect investigated. Typically, the aspects studied in cellular models for PD are cell death of dopaminergic neurons and α-synuclein aggregation.[106] The neuroblastoma cell line SH-SY5Y and the pheochromocytoma cell line PC12 are employed frequently because they are able to produce catecholamines (including dopamine), they can be induced to behave in a neuron-like manner under the right conditions and are especially useful for pharmacological experiments.[106] For obtaining high concentrations of α-synuclein which lead to its aggregation, usually cell lines such as HEK293 are transfected with strong promoters to induce high expression. Alternatively, the activity of the proteasome, which usually degrades α-synuclein, can be reduced pharmacologically. Furthermore, genes of mutant α-synculein which have a higher tendency to form aggregates, can be introduced.[106] The protective effects of compounds can be determined in these assay systems and the most promising molecules can be tested in an animal models of PD.

Animal models have advantages over cellular models in that they give a more holistic view of a disease (*e.g.* allowing observations of neuroinflammatory effects as well as cytotoxicity) and therefore they provide a more complete picture of the effects of potential drug candidates. Furthermore, they allow the study pharmacokinetic parameters, metabolic properties, toxicity and adverse effects.

Most animal models are acute models in which dopaminergic cell loss in mice is induced by the administration of chemicals such as MPP^+, rotenone or 6-hyroxadopamine which are selectively neurotoxic for dopaminergic neurons.[24,25,108] 6-Hydroxydopamine behaves *in vivo* like dopamine, but is inducing oxidative stress in the cytoplasm. It is taken up by dopamine and noradrenaline transporters. To protect noradrenergic neurons from the same damage, a noradrenaline reuptake inhibitor (such as imipramine) is coadministered.[108] Similar results can be achieved by the administration of MPTP. It is oxidised to MPP^+ by monoamine oxidase B. MPP^+ inhibits the electron transport chain, leading to an accumulation of ROS which causes cell damage. MPP^+ is taken up through the dopamine transporter.[108] Other models include inhibition of the proteasome and genetic modifications, most of them aimed at increasing α-synculein concentration or increasing its tendency to aggregate.[108] Protective effects in such models can be indicative of therapeutic efficacy in PD that could be followed up in subsequent human clinical trials.

6 References

[1] P. Moi, K. Chan, I. Asunis, A. Cao, Y. W. Kan; Isolation of NF-E2-related factor2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the β-globin locus control region; *Proceedings of the National Academy of Sciences* **99** 9926-9930 (1994)

[2] P. P. Levings, J. Bugnert; The human β-globin locus control region; *European Journal of Biochemistry* **269** 1589-1599 (2002)

[3] A. R. Leach; *Molecular Modelling – Principles and Applications*; **2nd edition** 2001

[4] B. Coupez, R. A. Lewis; Docking and Scoring – Theoretically Easy, Practically Impossible?; *Current Medicinal Chemistry* **13** 2995-3003 (2006)

[5] K. Itoh, N. Wakabayashi, Y. Katoh, T. Ishii, K. Igarashi, J. D. Engel, M. Yamamoto; Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain; *Genes & Development* **13** 76-86 (1999)

[6] I. Halperin, B. Ma, H. Wolfson, R. Nussinov; Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions; *PROTEINS: Structure, Function, and Genetics* **47** 409-483 (2002)

[7] M. Nishizawa, K. Kataoka, N. Goto, K. T. Fujiwara, S. Kawai; v-maf, a viral oncogene that encodes a “leucine zipper” motif; *Proceedings of the National Academy of Sciences* **86** 7711-7715 (1989)

[8] C. Grebner, J. Becker, S. Stepanenko, B. Engels; Efficiency of tabu-search-based conformational search algorithms; *Journal of Computational Chemistry* **32(10)** 2245-2253 (2011)

[9] P. Nioi, T. Nguyen, P. J. Sherratt, C. B. Pickett; The Carboxy-Terminal Neh3 Domain of Nrf2 Is Required for Transcriptional Activation; *Molecular and Cellular Biology* **25(24)** 10895-10906 (2005)

[10] S. K. Niture, R. Khatri, A. K. Jaiswal; Regulation of Nrf2-an update; *Free Radical Biology and Medicine* **66** 36-44 (2014)

[11] Katoh, Y., K. Itoh, E. Yoshida, M. Miyagishi, A. Fukamizu, and M. Yamamoto; Two domains of Nrf2 cooperatively bind CBP, a CREB binding protein, and synergistically activate transcription; *Genes to Cells* **6** 857– 868 (2001)

[12] M. McMahon, N. Thomas, K. Itoh, M. Yamamoto, J. D. Hayes; Redox-regulated Turnover of Nrf2 Is Determined by at Least Two Separate Protein Domains, the Redox-sensitive Neh2 Degron and the Redox-insensitive Neh6 Degron; *The Journal of Biological Chemistry* **279(30)** 31556-31567 (2004)

[13] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath; Docking and scoring in virtual screening for drug discovery: methods and applications; *Nature Reviews Drug Discovery* **3** 935-949 (2004)

[14] J. W. Kaspar, S. K. Niture, A. K. Jaiswal; Nrf2:INrf2 (Keap1) signalling in oxidative stress; *Free Radical Biology & Medicine* **47** 1304-1309 (2009)

[15] D. D. Zhang; Mechanistic Studies of the Nrf2-Keap1 Signalling Pathway; *Drug Metabolism Reviews* **38(4)** 769-789 (2006)

[16] L. M. Aleksunes, J. E. Manautou; Emerging Role of Nrf2 in Protecting Against Hepatic and Gastrointestinal Disease; *Toxicologic Pathology* **35** 459-473 (2007)

[17] J. D. Hayes, M. McMahon; NRF2 and KEAP1 mutations: permanent activation of an adaptive response in cancer; *Trends in Biochemical Sciences* **34(4)** 176-188 (2009)

[18] J. Kim, Y.-S. Keum; NRF2, a Key Regulator of Antioxidants with Two Faces towards Cancer; *Oxidative Medicine and Cellular Longevity* **2016**, article ID: 2746457

[19] A. Giudice, C. Arra, M. C. Turco; Review of Molecular Mechanisms Involved in the Activation of the Nrf2-ARE Signaling Pathway by Chemopreventive Agents; *Transcription Factors: Methods and Protocols, Methods in Molecular Biology* **647** 37-74 (2010)

[20] P. Rada, A. I. Rojo, S. Chowdhry, M. McMahon, J. D. Hayes, A. Cuadrado; SCF/β-TrCP Promotes Glycogen Synthase Kinase 3-Dependent Degradation of the Nrf2 Transcription Factor in a Keap1-Independent Manner; *Molecular and Cellular Biology* **31(6)** 1121-1133 (2011)

[21] G. Wells; Peptide and small molecule inhibitors of the Keap1–Nrf2 protein–protein interaction; *Biochemical Society Transactions* **43(4)** 674-679 (2015)

[22] G. Sita, P. Hrelia, A. Tarozzi, F. Morroni; Isothiocyanates Are Promising Compounds against Oxidative Stress, Neuroinflammation and Cell Death that May Benefit Neurodegeneration in Parkinson’s Disease; *International Journal of Molecular Sciences* **17** 1454 (2016)

[23] A. T. Dinkova-Kostova, A. Y. Abramov; The emerging role of Nrf2 in mitochondrial function; *Free Radical Biology and Medicine* **88** 179-188 (2015)

[24] F. Morroni, G. Sita, A. Tarozzi, G. Cantelli-Forti, P. Hrelia; Neuroprotection by 6-(methylsulfinyl)hexyl isothiocyanate in a 6-hydroxydopamine mouse model of Parkinson’s diseases; *Brain Research* **1589** 93-104 (2014)

[25] F. Morroni, A. Tarozzi, G. Sita, C. Bolondi, J. M. Z. Moraga, G. Cantelli-Forti, P. Hrelia; Neuroprotective effect of sulforaphane in 6-hydroxydopamine-lesioned mouse model of Parkinson’s disease; *NeuroToxicology* **36** 63-71 (2013)

[26] J.-H. Shin, S.-W. Kim, Y. Jin, I.-D. Kim, J.-K. Lee; Ethyl pyruvate-mediated Nrf2 activation and hemeoxygenase 1 induction in astrocytes confer protective effects via autocrine and paracrine mechanisms; *Neurochemistry International* **61** 89-99 (2012)

[27] I. Lastres-Becker, A. Ulusoy, N. G. Innamorato, G. Sahin, A. Rábano, D. Kirik, A. Cuadrado; α-Synuclein expression and Nrf2 deficiency cooperate to aggravate protein aggregation, neuronal death and inflammation in early-stage Parkinson's disease; *Human Molecular Genetics* **21(14)** 3173-3192 (2012)

[28] X. Li, D. Zhangs, M. Hannink, L. J. Beamer; Crystal Structure of the Kelch Domain of Human Keap1; *The Journal of Biological Chemistry* **279(50)** 54750-54758 (2004)

[29] S. Magesh, Y. Chen, L. Hu; Small Molecule Modulators of Keap1-Nrf2-ARE Pathway as Potential Preventive and Therapeutic Agents; *Medicinal Research Reviews* **32(4)** 687-726 (2012)

[30] R. Hancock, M. Schaap, H. Pfister, G. Wells; Peptide inhibitors of the Keap1–Nrf2 protein–protein interaction with improved binding and cellular activity; *Organic & Biomolecular Chemistry* **11** 3553-3557 (2013)

[31] C. Zhuang, Z. Miao, C. Sheng, W. Zhang; Updated Research and Applications of Small Molecule Inhibitors of Keap1-Nrf2 Protein-Protein Interaction: a Review; *Current Medicinal Chemistry* **21** 1861-1870 (2014)

[32] Z.-Y. Jiang, M-C- Lu, L.-L- Xu, T.-T. Yang, M.-Y. Xi, X.-L. Xu, X.-K. Guo, X.-J. Zhang, Q.-D. You, H.-P. Sun; Discovery of Potent Keap1–Nrf2 Protein–Protein Interaction Inhibitor Based on Molecular Binding Determinants Analysis; *Journal of Medicinal Chemistry* **57** 2736-3745 (2014)

[33] A. F. Winkel, C. K. Engel, D. Margerie, A. Kannt, H. Szillat, H. Glombik, C. Kallus, S. Ruf, S. Güssregen, J. Riedel, A. W. Herling, A. von Knethen, A. Weigert, B. Brüne, D. Schmoll; Characterization of RA839, a Noncovalent Small Molecule Binder to Keap1 and Selective Activator of Nrf2 Signaling; *The Journal of Biological Chemistry* **290(47)** 28446-28455 (2015)

[34] T. G. Davies, W. E. Wixted, J. E. Coyle, C. Griffiths-Jones, K. Hearn, R. McMenamin, D. Norton, S. J. Rich, C. Richardson, G. Saxty, H. M. G. Willems, A. J.-A. Woolford, J. E. Cottom, J.-P. Kou, J. G. Yonchuk, H. G. Feldser, Y. Sanchez, J. P. Foley, B. J. Bolognese, G. Logan, P. L. Podolin, H. Yan, J. F. Callahan, T. D. Heightman, J. K. Kerns; Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein−Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery; *Journal of Medicinal Chemistry* **59** 3991-4006 (2016)

[35] N. J. Abbott, A. A.K. Patabendige, D. E.M. Dolman, S. R. Yusof, D. J. Begley; Structure and Function of the blood-brain barrier; *Neurobiology of Disease* **37** 13-25 (2010)

[36] N. Bodor, P. Buchwald; Brain-Targeted Drug Delivery; *American Journal of Drug Delivery* **1(1)** 13-26 (2003)

[37] D. Veber, S. Johnson, H. Cheng, B. Smith, K. Ward, K. Kopple; Molecular properties that influence the oral bioavailability of drug candidates; *Journal of Medicinal Chemistry* **45(12)** 2615-2623 (2002)

[38] T. Erdogan; Fluorescence Polarization in Life Sciences; *Semrock White Paper Series* (https://www.semrock.com/fluorescence-polarization-in-life-sciences.aspx)

[39] Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, T. Hou; Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power; *Physical Chemistry Chemical Physics* **18** 12964-12975 (2016)

[40] www.rcsb.org; H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne; The Protein Data Bank; *Nucleic Acids Research* **28** 235-242 (2000)

[41] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE; UCSF Chimera--a visualization system for exploratory research and analysis; *J Comput Chem*. **25(13**) 1605-1612 (2004)

[42] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, AB. Garmendia-Doval, S. Juhos, et al. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids; *PLOS Computational Biology* **10(4)** e1003571 (2014)

[43] W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz, R. C. Rizzo; DOCK 6: Impact of new features and current docking performance; *Journal of Computational Chemistry* **36** 1132-1156 (2015)

[44] http://lephar.com

[45] O. Trott, A. J. Olson; AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading; *Journal of Computational Chemistry* **31** 455-461 (2010)

[46] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson; Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility; *Journal of Computational Chemistry* **16** 2785-2791 (2009)

[47] T. Sterling, J. J. Irwin; ZINC 15 – Ligand Discovery for Everyone; *Journal of Chemical Information and Modeling;* **55(11)** 2324-2337 (2015) (http://zinc15.docking.org/)

[48] L. Hu, S. Magesh, L. Chen, L. Wang, T. A. Lewis, Y. Chen, C. Khodier, D. Inoyama, L. J. Beamer,T. J. Emge, J Shen, J. E. Kerrigan, A.-N. T. Kong, S. Dandapani, M. Palmer, S. L. Schreiber, B. Munoz; Discovery of a small-molecule inhibitor and cellular probe of Keap1-Nrf2 protein-protein interaction; *Bioorganic & Medicinal Chemistry Letters* **23(10)** 3039-3043 (2013)

[49] http://www.nmsf.swan.ac.uk/documents/LTQ%20Orbitrap%20XL%20_SOPAug08.pdf

[50] D. Stokmaier, O. Khorev, B. Cutting, R. Born, D. Ricklin, T. O. G. Ernst, F. Böni, K. Schwingruber, M. Gentner, M. Wittwer, M. Spreafico, A. Vedani, S. Rabbani, O. Schwardt, B. Ernst; Design, synthesis and evaluation of monovalent ligands for the asialoglycoprotein receptor (ASGP-R); *Bioorganic & Medicinal Chemistry* **17** 7254 – 7264 (2009)

[51] H. C. Bertrand, M. Schaap, L. Baird, N. D. Georgakopoulos, A. Fowkes, C. Thiollier, H. Kachi, A. T. Dinkova-Kostova, G. Wells; Design, Synthesis, and Evaluation of Triazole Derivatives That Induce Nrf2 Dependent Gene Products and Inhibit the Keap1–Nrf2 Protein–Protein Interaction; *Journal of Medicinal Chemistry* **58** 7186-7194 (2015)

[52] R. Hancock, H. C. Bertrand, T. Tsujita, S. Naz, A. El-Bakry, J. Laoruchupong, J.D. Hayes; Peptide inhibitors of the Keap1–Nrf2 protein–protein interactio*n; Free Radical Biology & Medicine* **52** 444-451 (2012)

[53] http://wwpdb.org/validation/legacy/XrayValidationReportHelp

[54] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan; Stereochemistry of polypeptide chain configurations; *Journal of Molecular Biology* **7(1)** 95-99 (1963)

[55] A. D. Jain, H. Potteti, B. G. Richardson, L. Kingsley, J. P. Luciano, A. F. Ryuzoji, H. Lee, A. Krunic, A. D Mesecar, S. P. Reddy, T. W. Moore; Probing the structural requirements of non-electrophilic naphthalene-based Nrf2 activators; *European Journal of Medicinal Chemistry* **103** 252-268 (2015)

[56] D. Marcotte, W. Zeng, J.-C. Hus, A. McKenzie, C. Hession, P. Jin, C. Bergeron, A. Lugovskoy, I. Enyedy, H. Cuervo, D. Wang, C. Atmanene, D. Rowcklin, M. Vecci, V. Vivat, J. Kraemer, D. Winkler, V. Hong, J. Chao, M. Lukashev, L. Silvian; Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism*; Bioorganic & Medicinal Chemistry* **21** 4011-4019 (2014)

[57] N. Kunishima, T. Tanaka, M. Satoh, H. Saburi; Crystal Structure of Keap1 in Complex with Synthetic Small Molecular based on a co-crystallization; *to be published* (released on PDB: 16.01.2013)

[58] S. Hörer, D. Reinert, K. Ostmann, Y. Hoevels, H. Nar; Crystal-Contact Engineering to Obtain a Crystal Form of the Kelch Domain of Human Keap1 Suitable for Ligand-Soaking Experiments; *Acta Crystallographica Section F* **69** 592- 596 (2013)

[59] L. J. Beamer, X. Li, C. A. Botooms, M. Hannink; Conserved solvent and side-chain interactions in the 1.35 Å structure of the Kelch domain of Keap1; *Acta Crystallographica Section D* **61** 1335-1342 (2005)

[60] Y. Cao, L. Lei; Improved protein-ligand binding affinity prediction by using a curvature dependent surface area model; *Bioinformatics* **30(12)** 1674-1680 (2014)

[61] G. Neudert, G. Klebe; DSX: A Knowledge-Based Scoring Function for the Assessment of Protein–Ligand Complexes; *Journal of Chemical Information and Modeling* **51(10)** 2731-2745 (2011)

[62] Marvin 5.5.0.1, 2011, ChemAxon (www.chemaxon.de)

[63] J. P. Hughes, S. Rees, S. B. Kalindjian, K. L. Philpott; Principles of early drug discovery; *British Journal of Pharmacology* **162** 1239-1249 (2011)

[64] E. Frérot, J. Coste, A. Pantaloni, M.-N. Dufour, P. Join; PyBOP® and PyBroP: Two reagents for the difficult coupling of the α,α-dialkyl amino acid, Aib; *Tetrahedron* **47(2)** 259-270 (1991)

[65] E. Valeur, M. Bradley; Amide bond formation: beyond the myth of coupling reagents; *Chemical Society Reviews* **38** 606-631 (2009)

[66] C. A. G. N. Montalbetti, V. Falque; Amide bond formation and peptide coupling; *Tetrahedron* **61** 10827-10852 (2005)

[67] *Scifinder*, 2017; Chemical Abstracts Service: Columbus, OH, 2016; CAS Registry Number 7409-30-5 (accessed 02.02.2017). Calculated using Advanced Chemistry Development (ACD/Labs) Software V11.02

[68] P. Roose, K. Eller, E. Henkes, R. Rossbacher, H. Höke; *Ullmann's Encyclopedia of Industrial Chemistry – Amines, Aliphatic*; **7th edition** 2014

[69] V. O. T. Omuaru, N. Boisa, H. U Obuzor; Reaction of cyclic anhydrides with aromatic primary amines: Part 2 - Synthesis of novel anilinic acids from a Diels-Alder adduct; *Indian Journal of Chemistry, Section B: Organic Chemistry Including Medicinal Chemistry* **37(7)** 707-709 (1998)

[70] K. Lienkamp, A. E. Madkour, A. Musante, C. F. Nelson, K. Nüsslein, G. N. Tew; Antimicrobial Polymers Prepared by ROMP with Unprecedented Selectivity: A Molecularl Construction Kit Approach; *Journal of the American Chemical Society* **130(30)** 9836-9843 (2008)

[71] http://nbcr-222.ucsd.edu/autoclickchem/tutorial.php (accessed: April 2017)

[72] W. M. Pardridge; CNS Drug Design Based on Principles of Blood-Brain Barrier Transport; *Journal of Neurochemistry* **70** 1781-1792 (1998)

[73] J. Alam, D. Stewart, C. Touchard, S. Boinapally, A. M. K. Choi, J. L. Cook; Nrf2, a Cap'n'Collar Transcription Factor, Regulates Induction of the Heme Oxygenase-1 Gene; *The Journal of Biological Chemistry* **247(37)** 26071-26078 (1999)

[74] C. Zoja, A. Benigni, G. Remuzzi; The Nrf2 pathway in the progression of renal disease; *Nephrology Dialysis* **29** i19-i24 (2014)

[75] C. A. Lipinski; Lead- and drug-like compounds: the rule-of-five revolution; *Drug Discovery Today: Technologies* **1(4)** 337-341 (2004)

[76] N. A. Strotman, H. R. Chobanian, Y. Guo, J. He, J. E. Wilson; Highly Regioselective Palladadium-Catalyzed Direct Arylation of Oxazole at C-2 or C-5 with Aryl Bromides, Chlorides and Triflates; *Organic Letters* **12(16)** 3578-3581 (2010)

[77] K. F. Freeman-Cook, R. L. Hoffmann, T. W. Johnson; Lipophilic efficiency: the most important efficiency metric in medicinal chemistry; *Future Medicinal Chemistry* **5(2)** 113–115 (2013)

[78] S. C. Baumgartner, C. Eberle, F. Diederich, S. Lauw, F. Rohdich, W. Eisenreich, A. Bacher; Structure-Based Design and Synthesis of the First Weak Non-Phosphate Inhibitors for IspF, an Enzyme in the Non-Mevalonate Pathway of Isoprenoid Biosynthesis; *Helvetica Chimica Acta* **90** 1043-1068 (2007)

[79] N. Richy, M. Ghoraf, J. Vidal; N-Silyloxaziridines: Synthesis and Use for Electrophilic Amination; *Journal of Organic Chemistry* **77(23)** 10972-10977 (2012)

[80] S. Orbisaglia, B. Jacques, P. Braunstein, D. Hueber, P. Pale, A. Blanc, P. de Frémont; Synthesis, Characterization, and Catalytic Activity of Cationic NHC Gold(III) Pyridine Complexes; *Organometallics* **32** 4153-4164 (2013)

[81] Y.-Y. Chen, K.-L. Chen, Y.-C. Tyan, C.-F. Liang, P.-C. Lin; Synthesis of substituted 3-methylene-2,3-dihydrobenzofurans and 3-methylbenzofurans by rhodium (II)-catalyzed annulation; *Tetrahedron* **71** 6210-6218 (2015)

[82] R. Yanada, S. Obika, Y. Kobayashi, T. Inokuma, M. Oyama, K. Yanada, Y. Takemoto; Stereoselective Synthesis of 3-Alkylideneoxindoles Using Tandem Indium mediated Carbometallation and Pd-catalyzed Cross-Coupling Reaction; *Advancd Synthesis & Catalysis* **347** 1632 – 1542 (2005)

[83] M Nagel, H.-J. Hansen; Synthesis of Polyalkylphenyl Prop-2-ynoates and Their Flash Vacuum Pyrolysis to Polyalkylcyclohepta[*b*]furan-2(2*H*)-ones; *Helvetica Chimica Acta* **83** 1022 – 1048 (2000)

[84] M. D. Aparece, P. A. Vadola, Gold-Catalyzed Dearomative Spirocyclization of Aryl Alkynoate Esters; *Organic Letters* **16(22)** 6008-6011 (2014)

[85] E. Jnoff, C. Albrecht, J. J. Barker, O. Barker, E. Beaumont, S. Bromidge, F. Brookfield, M. Brooks, C. Bubert, T. Ceska, V. Corden, G. Dawson, S. Duclos, T. Fryatt, C. Genicot, E. Jigorel, J. Kwong, R. Maghames, I. Mushi, R. Pike, Z. A. Sands, M. A. Smith, C. C. Stimson, J.-P. Courade; Binding Mode and Structure–Activity Relationships around Direct Inhibitors of the Nrf2–Keap1 Complex; *ChemMedChem Communications* **9** 699-705 (2014)

[86] C. Zhang, S. Narayanapillai, W. Zhang, Y. Y. Sham, C. Xing; Rapid Identification of Keap1-Nrf2 Small-Molecule Inhibitors through Structure-Based Virtual Screening and Hit-Based Substructure Search; *Journal of Medicinal Chemistry* **57** 1121-1126 (2014)

[87] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeny; Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings; *Advanced Drug Delivery Reviews* **46** 3-26 (2001)

[88] The Uniprot Consortium; UniProt: the universal protein knowledgebase; *Nuclei Acids Research* **45(D1)** D158-D169 (2016)

[89] S. Riniker, L. J. Barandum, F. Diederich, O. Krämer, A. Steffen, W. F. van Gunsteren; Free enthalpies of replacing water molecules in protein binding pockets; *Journal of Computer-Aided Molecular Design* **26(12)** 1293-1309 (2012)

[90] C. G. Wermuth, D. Aldous, P, Raboisson, D. Rognan; *The Practice of Medicinal Chemistry*; **4th edition** 2015

[91] S. Forli, A. J. Olson; A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking; *Journal of Medicinal Chemistry* **55** 623-638 (2012)

[92] W. Czechtitzky, P. Hamley; *Small Molecule Medicinal Chemistry: Strategies and Technologies*; 2015

[93] D. A. Pearlman, P. S. Charifsonn; Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System; *Journal of Medicinal Chemistry* **44** 3417-3423 (2001)

[94] P. Kollman; Free Energy Calculations: Applications to Chemical and Biochemical Phenomena; *Chemical Reviews* **93** 2395-2417 (1993)

[95] J. W. Kaus, E. Harder, T. Lin, R. Abel, J. A. McCammon, L. Wang; How To Deal with Multiple Binding Poses in Alchemical Relative Protein–Ligand Binding Free Energy Calculations; *Journal of Chemical Theory and Computation* **11(6)** 2395-2417 (2015)

[96] S. Kalyanamoorthy, Y.-P. P. Chen; A steered molecular dynamics mediated hit discovery for histone deacetylases; *Physical Chemistry Chemical Physics* **16** 3777-3791 (2014)

[97] H. Nguyen, L. Le; Steered molecular dynamics approach for promising drugs for influenza A virus targeting M2 channel proteins; *European Biophysics Journal* **44(6)** 447-455 (2015)

[98] J. S. Patel, A. Berteotti, S. Ronsisvalle, W. Rocchia, A. Cavalli; Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5; *Journal of Chemical Information and Modelling* **54(2)** 470-480 (2014)

[99] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, K. Schulten; Molecular Dynamics Study of Unbinding of the Avidin-Biotin Complex; *Biophysical Journal* **72** 1568-1581 (1997)

[100] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten; Scalable Molecular Dynamics with NAMD; *Journal of Computational Chemistry* **26(16)** 1781-1802 (2005)

[101] M. Schaap, R. Hancock, A. Wilderspin, G. Wells; Development of a steady-state FRET-based assay to identify inhibitors of the Keap1-Nrf2 protein-protein interaction; *Protein Science* **22(12)** 1812-1819 (2013)

[102] A. Bujard, M. Sol, P.-A. Carrupt, S. Martel; Predicting both passive intestinal absorption and the dissociation constant toward albumin using the PAMPA technique; *European Journal of Pharmaceutical Sciences* **63** 36-44 (2014)

[103] L. Di, E. D. Kern, K. Fan, O. J. McConnel, G. T. Carter; High throughput artificial membrane permeability assay for blood-brain barrier; *European Journal of Medicinal Chemistry* **38** 223-232 (2003)

[104] D. A. Volpe; Drug-permeability and transporter assays in Caco-2 and MDCK cell lines; *Future Medicinal Chemistry* **3(16)** 2063-2077 (2011)

[105] C. J. Czupalla, S. Liebner, K. Devray; *Cerebral Angiogenesis - Methods in Molecular Biology* **Volume 1135** 415-437 (2014)

[106] B. H. Falkenburger, T. Saridaki, E. Dinter; Cellular models for Parkinson's disease; *Journal of Neurochemistry* **139 (Suppl. 1)** 121-130 (2016)

[107] W. Bobela, L. Zheng, B. L. Schneider; Overview of Mouse Models of Parkinson's Disease; *Current Protocols in Mouse Biology* **4(3)** 121-139 (2014)

[108] W. Le, P. Sayana, J. Jankovic; Animal Models of Parkinson's Disease: A Gateway to Therapeutics?; *Neurotherapeutics* **11** 92-110 (2014)

[109] H. Zhao, A. Caflisch; Molecular dynamics in drug design; *European Journal of Medicinal Chemistry* **91** 4-14 (2015)

[110] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N.

Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco

[111] R. Salomon-Ferrer, D. A. Case, R. C. Walker; An overview of the Amber biomolecular simulation package; *WIREs Computational Molecular Science* **3** 198-210 (2013)

[112] Lopes, P.E.M., Guvench, O., and MacKerell, A.D., Jr.; Current Status of Protein Force Fields for Molecular Dynamics; Molecular Modeling of Proteins, **2nd edition**, A. Kukol, Editor, Humana Press. Chapter 3, pp. 47 - 72, 2014

[113] http://mackerell.umaryland.edu/charmm_ff.shtml

[114] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren; A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6; *Journal of Computational Chemistry* **25(13)** 1656-1676 (2004)