RSC Medicinal Chemistry

Identification of a new class of proteasome inhibitors based on a naphthyl-azotricyclic-urea-phenyl scaffold

Duncan Allardyce, Priscilla Adu Mantey, Monika Szalecka, Robert Nkwo, Eriketi Z. Loizidou*

Middlesex University, Faculty of Science and Technology, Department of Natural Sciences, The Burroughs, London, NW4 4BT, United Kingdom

*Author to whom correspondence should be addressed:

Dr Eriketi Z. Loizidou, Middlesex University, Department of Natural Sciences, The Burroughs, NW4 4BT, London, UK; e.loizidou@mdx.ac.uk

Abstract

Proteasomes play an important role in protein degradation and regulation of many cellular pathways by maintaining protein balance. Inhibitors of the proteasome disrupt this balance affecting proteins that are key in malignancies and as such have found applications in the treatment of multiple myeloma and mantle cell lymphoma. However, resistance mechanisms have been reported for these proteasome inhibitors including mutations at the β5 site which necessitates the constant development of new inhibitors. In this work, we report the identification of a new class of proteasome inhibitors, polycyclic molecules bearing a naphtyl-azotricyclic-urea phenyl scaffold, from screening of the ZINC library of natural products. The most potent of these compounds showed evidence of dose dependency though proteasome assays with IC_{50} values in the low micromolar range and kinetic analysis revealed competitive binding at the β5c site with an estimated inhibition constant, K_i 1.15μM. Inhibition was also shown for the β5i site of the immunoproteasome at levels similar to the constitutive proteasome. Structure activity relationship studies identified the naphthyl substituent to be crucial for activity and modelling studies attributed this to enhanced hydrophobic interactions within β5c. Further to this, halogen substitution within the naphthyl ring enhanced activity and allowed for π-π interactions with Y169 in β5c and Y130 and F124 in β5i. Combined these data highlight the importance of hydrophobic and halogen interactions in β5 binding and assist in the design of next generation inhibitors of the proteasome.

Keywords

Proteasome inhibitors, virtual screening, azotricyclic-urea scaffold, competitive binding, structure activity relationships, Autodock

1. Introduction

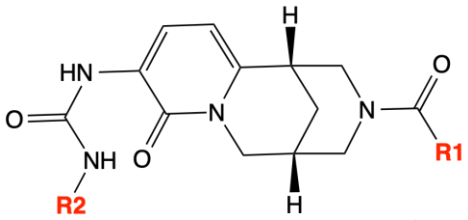
The ubiquitin-proteasome pathway (UPP) is the main system for the degradation of proteins deemed redundant, misfolded or toxic and as such has a direct involvement in the regulation of key pathways. Consequently, the UPP pathway has been an important target for drug development for many diseases including cancer, autoimmune and neurodegenerative diseases[1–3].

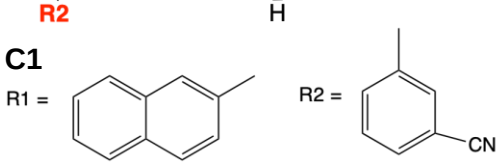
The 26S proteasome, known as the constitutive proteasome, within the UPP, is responsible for cleaving the ubiquitin-tagged proteins into oligopeptides[4,5] and this is achieved within the 20S core particle which contains the three proteolytic sites β1c, β2c and β5c. Each site exerts a different activity i.e caspase (β1c), trypsin (β2c) and chymotrypsin-like (β5c), through a mechanism that involves the action of a nucleophilic threonine at position 1. The discovery of 26S proteasome prompted the development of proteasome inhibitors most of which act at one or more active sites of the 20S core particle[6,7]. Overall, the proteasome is responsible for more than 80% of protein degradation and therefore regulates protein balance and vital processes such as, signal transduction, cell cycle control, apoptosis. Proteasome inhibitors can disrupt this balance and affect proteins that are key factors in malignancies, and this is exemplified by the clinical application of the 3 FDA approved proteasome inhibitors for the treatment of multiple myeloma[8,9]. Furthermore, since malignant cells display characteristically rapid proliferation and genetic instability, there is increased reliance upon protein degradation. Accordingly, proteasome pathway inhibition exhibits a more profound effect on cancer cells compared to normal, healthy cells.

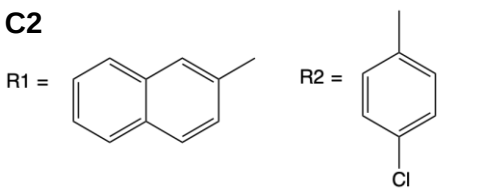
Bortezomib, is a first-generation inhibitor of the β5c subunit of the proteasome whose mechanism of action involves the formation of a slowly reversible boronate-proteasome complex with the active site threonine[9]. However, several studies indicate that bortezomib affects additional targets to the proteasome at the cellular level. Specifically, treatment of MCF7 cells with bortezomib led to changes in the expression of more than 10900 genes highlighting a wide biological effect while in a separate study bortezomib was shown to retain biological activity in cells with reduced proteasome function achieved by knockdown of the β1c, β2c and β5c active site subunits[10]. Further to this, resistance mechanisms have been reported for patients treated with bortezomib which result in part to β5c mutations and overexpression of the 26S proteasome[11,12]. Similar resistance mechanisms have also been reported for the next generation proteasome inhibitors, carfilzomib and ixazomib[9] highlighting the importance of developing new inhibitors with distinct cellular mechanisms to the current proteasome inhibitors.

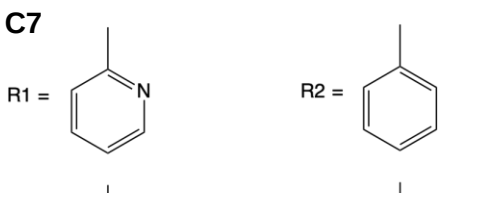
In addition to the constitutive proteasomes, cells are also capable of producing immunoproteasomes in which all of the proteolytic sites of the constitutive proteasome are replaced by their counterparts (β1i, β2i and β5i) while all other subunits remain unchanged[13,14]. Due to the preferential expression of immunoproteasomes in hematopoietic cells and other cells in disease state, recent studies have focused on the development of selective immunoproteasome inhibitors with the hopes of achieving a therapeutic outcome with less toxicity. Reports of immunoproteasome selective inhibitors include irreversible inhibitors such as peptide epoxyketone derivatives[15–19], peptidyl boronates[20] and oxathiazolones[21]. Recently, reversible noncompetitive inhibitors with enhanced activity for β5i have been reported to selectively induce cell death in malignant myeloma cells[22] and we have further reported the cyclic peptide, Argyrin B, to be a reversible noncompetitive inhibitor of the proteasome with enhanced activity for β1i and β5i[23]. Whereas some studies focus on immunoproteasome selective inhibitors others recommend a combined inhibition of β5/β5i and β2/β2i for achieving a significant clinical outcome in treating solid tumours[24].

In this project, we report a new class of proteasome inhibitors (figure 1) identified, through virtual screening of the ZINC15 library of natural products and further evaluated through in vitro assays and molecular docking. Kinetic analysis of the most active analogues (C1-4), revealed competitive binding at the β5c site of the proteasome with estimated inhibition constant at the low micromolar range while structure-activity analysis revealed the naphthyl (R1) group to be important for activity.









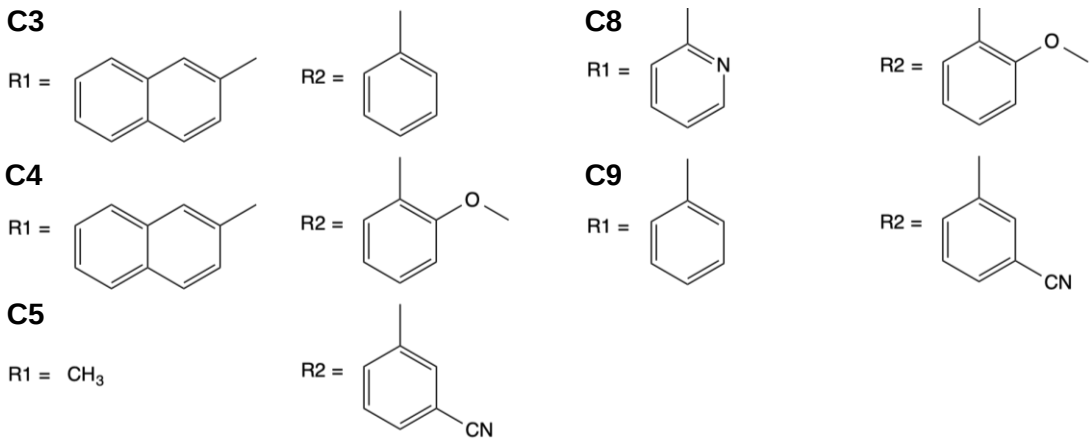


Figure 1. Structure of naphthyl-azotricyclic-urea phenyl compounds (top) and analogues (bottom).

2. Results and Discussion

Natural products have served as a valuable source of many bioactive molecules and had a major contribution to medicine. Their diverse and structurally complex structures, compared to synthetic drugs, can be advantageous in drug discovery and for this reason the natural product library was selected for virtual screening over the other ZINC libraries[25]. As natural products have typically high molecular weights, screening cutoff was set to 500-1000Da to ensure diversity in structures but also to minimize solubility problems for the subsequent biochemical assays. Another important aspect was to be able to experimentally measure the binding affinity of the selected compounds and as such, only commercially available natural products were included in the screening. The 20 best scored compounds, out of 5000 screened, as identified by Autodock vina were further analysed with Autodock 4.2 (Table S1) and the biological activity of the top scoring compound ZINC4258888 (C1) tested in vitro indicated an IC_{50} in the low micromolar region. This prompted us to investigate analogues of C1 with the aim of identifying structural features responsible for activity. The analogues share an azotricyclic-urea core structure and differ in substitutions at R1 and R2 (Figure 1). These were selected with the aim of identifying whether the naphthyl ring (R1) would have a significant contribution to binding as previous studies highlighted hydrophobic interactions to be important for β5c binding[19]. Further to this, we wished to examine the effect of substituents on benzene ring (R2) on binding. We were particularly interested in the nitrile functionality, as it is known to function as a ketone bioisostere facilitating polar interactions and acting as hydrogen bond acceptor as well as polarising aromatic rings and promoting π-stacking interactions. With regards to pharmacokinetic properties, nitriles have shown

increased biocompatibility as they are less susceptible to metabolism especially in the case of arylnitrile-based pharmaceuticals[26]. On the other hand, halogens are widely used in medicinal chemistry due to their ability to form halogen bonds with any accessible Lewis base within the binding pocket, most prominently the carbonyl oxygen but also with amino acid side chains such as hydroxyl groups, carboxylates, thiols, amines, as well as with aromatic rings[27,28].

The biochemical assay results revealed analogues C1-4 as the most potent with IC_{50} values against β5c at the low micromolar level ranging from 3.43 to 11.1μM. On the other hand, analogues C5-9 did not present any appreciable activity and IC_{50} values ranged from 200 to >500μM. The inhibition constant, Ki, was also investigated as a measure independent of enzyme concentration and as a determinant of inhibition type for the most potent analogues C1-4. This was determined for substrate concentration at the K_m value in order to avoid unrepresentative values from different inhibition modes. Table 1 shows Ki values for C1-4 at β5c where all C1-4 data best fit the competitive inhibition model as preferred according to difference in Akaike's information criterion AICc (Figure S3)[29]. Ki values are lower than IC_{50}, as expected ranging from 1.14 to 9.69μM but with similar trends showing C2 as the most potent analogue followed by C1, C4 and C3.

As predicted, these results highlight the importance of the naphthyl group (R1) for activity. Where the naphthyl group is substituted with methyl (C5) or pyridine (C6-8), activity is greatly diminished. It can be assumed that the naphthyl ring occupies a hydrophobic pocket within the active site allowing for hydrophobic interactions with the ring. This is further supported by the activity of C9 in which R1 is occupied by a benzene ring which even though is smaller in size than the naphthyl ring it can still participate in some hydrophobic interactions and therefore displaying moderate activity (IC_{50} 201μM). On the other hand, C5 that has a small methyl group in R1 is much less active (IC_{50} 491μM) compared to C9. Similar activity to C5 is observed for analogues C6-8 where a pyridine occupies R1 suggesting that polar groups are not as well tolerated as non-polar hydrophobic groups, within the β5c active site.

Furthermore, the substitution of the phenyl ring (R2) is implicated in activity as the strength of inhibition indicated by the Ki data, is slightly increased for the substituted analogues C1, C2 and C4 compared to unsubstituted benzene in C3. Interestingly the electronic nature of the substituent does not appear to have a considerable effect on binding. The most potent analogue, by narrow margins, was determined to be C2 in which a chlorine is present at the para position of the phenyl ring, implicating a role for chlorine in binding.

Table 1. Mean IC_{50} values and 95% confidence interval for analogues C1-9 determined from purified enzyme assays at β5c (n=3) at 1nM enzyme concentration. Ki values and 95% confidence interval are shown for analogues C1-4.

Compound	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**
IC_{50} (μM)	10.3	3.43	11.1	6.65	491	440	>500	560	201
95% CI (μM)	8.72-12.13	2.77-4.24	9.93-12.43	5.41-8.12	379-950	358-827	-	371-1610	156-289
Ki (μM)	3.424	1.145	9.689	4.602	-	-	-	-	-
95% CI (μM)	2.733-4.341	0.8 - 1.69	7.126-13.69	3.273-6.748	-	-	-	-	-

Molecular docking simulations at β5c show similar trends as experiments where C1-4 are predicted to bind the strongest with estimated inhibition constants Ki ranging from 4-12nM (Table S2). On the other hand, the estimated Ki for the least active analogues range from 40-500nM. As is common with many molecular docking simulations the predicted values for binding energies and inhibition constants differ from those obtained from experiment as docking simulations use simplified scoring functions to calculate energies that do not take into consideration factors such as, protein flexibility and water molecules[30,31]. Docking simulations have been successful however, in determining binding orientations and trends amongst analogues as well as exploring large databases of chemical compounds through virtual screening[32–34]. In this context, the predicted binding poses can be used to provide further insight into the active site interactions. Figure 2 shows the best docked conformations of analogues C1-4 within the β5c active site along with the interacting amino acids. Both C1 and C3 participate predominantly in Van Der Waals / dipole-dipole interactions with neighbouring amino acids, hydrogen bonding interactions were not observed neither π-π stacking with the naphthyl ring. For C4 a hydrogen bonding interaction is observed between a nitrogen amide of the compound and the backbone amide between amino acids thr21 and ala20. C2 showed most interactions exhibiting two hydrogen bonds with the backbone amide of amino acids thr21 and ala20 as well as with thr1. Further to this, π-π stacking is identified between the naphthyl ring and the amino acid tyr169 (Figure S6). Even though docking analysis cannot identify halogen bonding this can be analysed based on distance between the halogen and Lewis base of interest and the orientation of interaction i.e electrophiles have been shown to approach the halogen in a side-on fashion while nucleophiles approach in a “head-on” fashion[27]. Accordingly, a halogen bonding interaction can be presumed between the chlorine atom on the benzene ring and the carbonyl oxygen

of amino acid gly47 (head-on interaction at 3.70Å distance). Halogen bonds with carbonyl oxygens are comparable to weak hydrogen bonds and even though they may not contribute as much to binding, they are likely to suffer less from energy penalties due to polar desolvation than hydrogen bonds. An overlay of the best docked conformations of all compounds (Figure S8-S9) shows the naphthyl ring occupying various positions, indicating the variety of hydrophobic regions within the active site able to accommodate the naphthyl ring while maximising other complementary interactions.

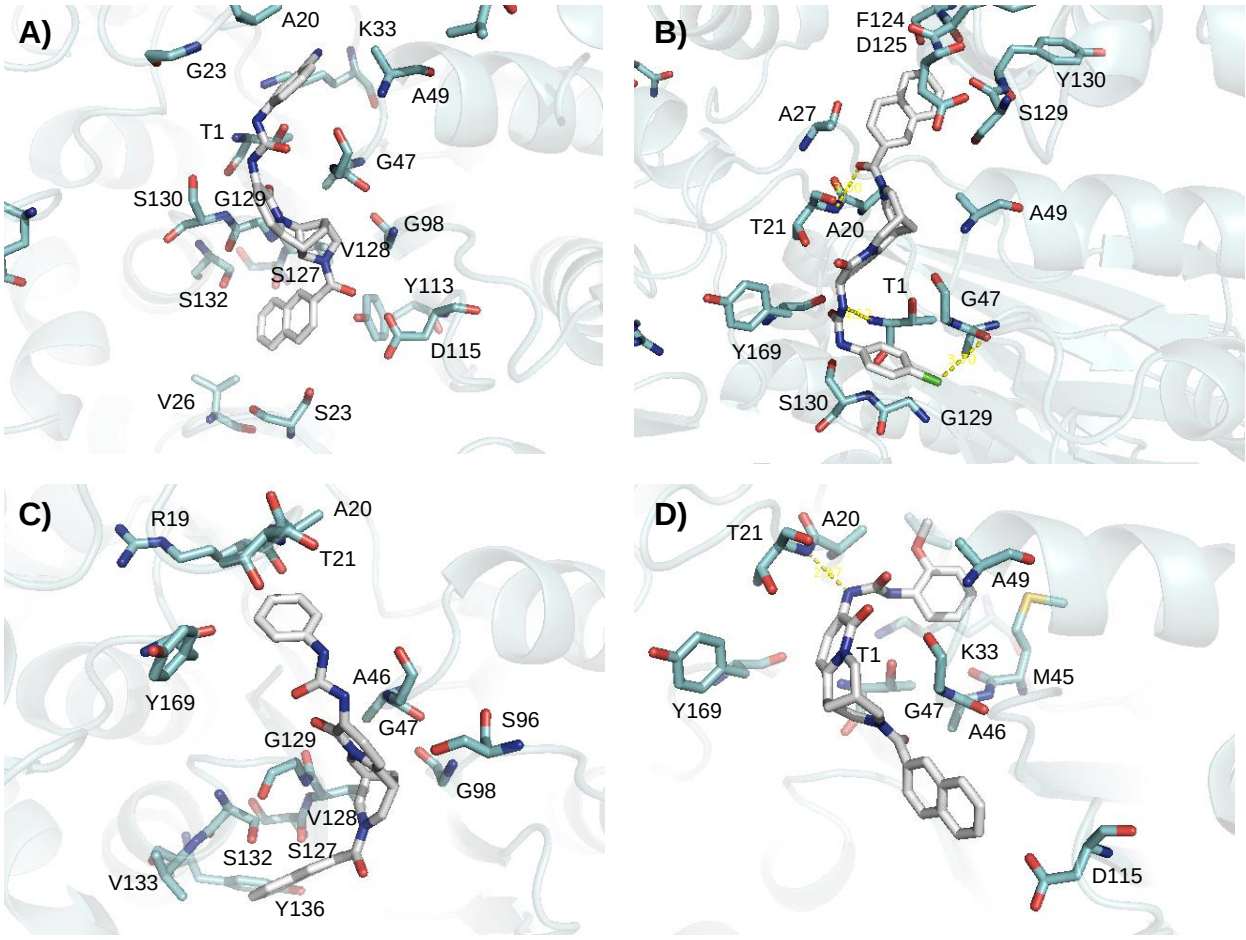


Figure 2. Best-docked conformations showing interacting amino acids with A) **C1**; B) **C2**, T21*, A20*, Y169**, G47***; C) **C3**; D) **C4**, A20*, T21*, within the β5c active site. *Denotes amino acids that participate in hydrogen bonding, ** denotes amino acids that participate in π-π interactions, ***denotes amino acids that participate in halogen bonding.

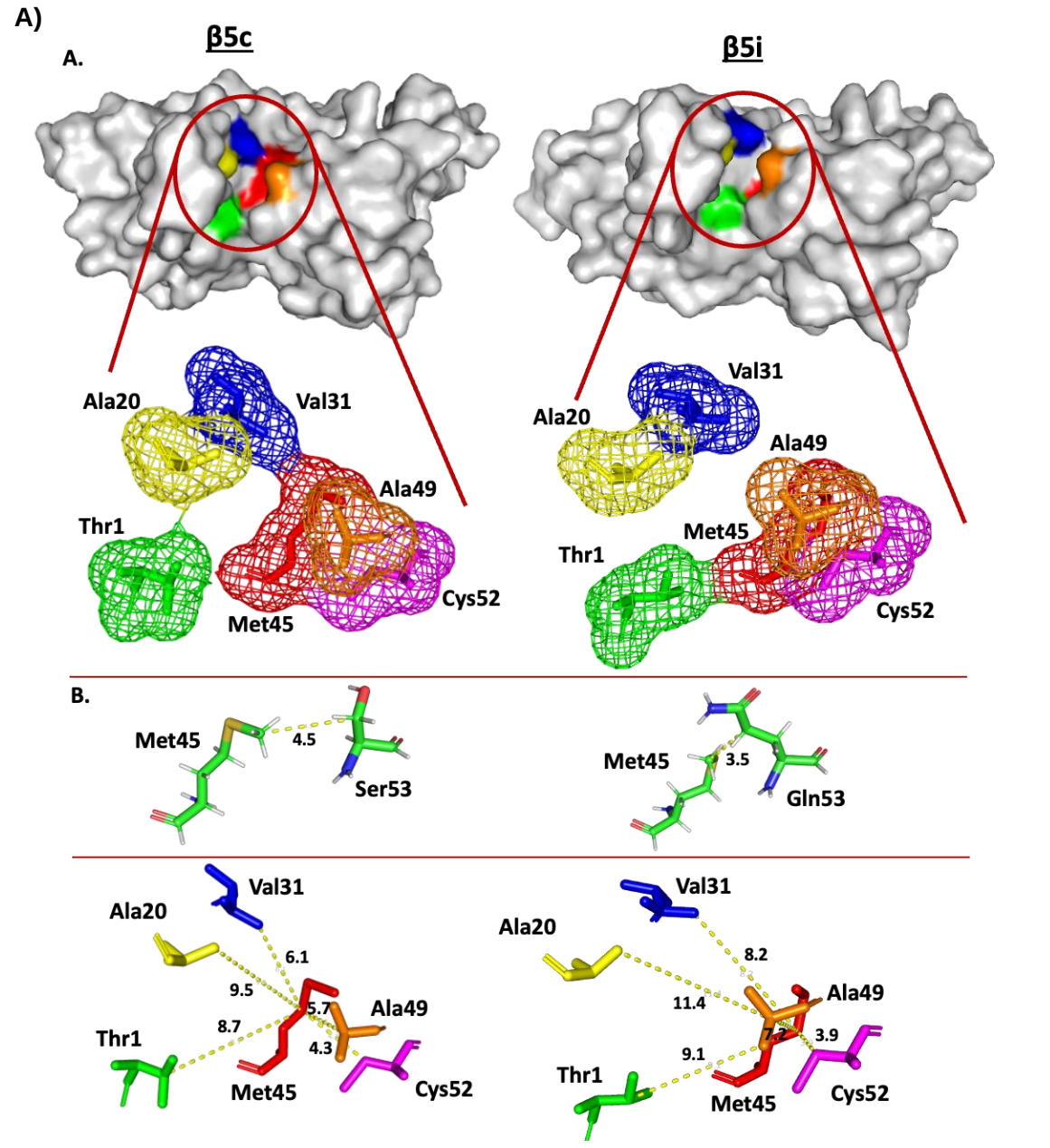
Recently inhibitors of the immunoproteasome have attracted attention as therapeutic alternatives to proteasome inhibitors. Due to the abundance of immunoproteasomes in disease cells they can potentially achieve similar therapeutic effects to proteasome inhibitors but without the accompanied toxicity that is associated with inhibiting the proteasome in healthy cells[35–37]. As such, we wished to examine the activity of the

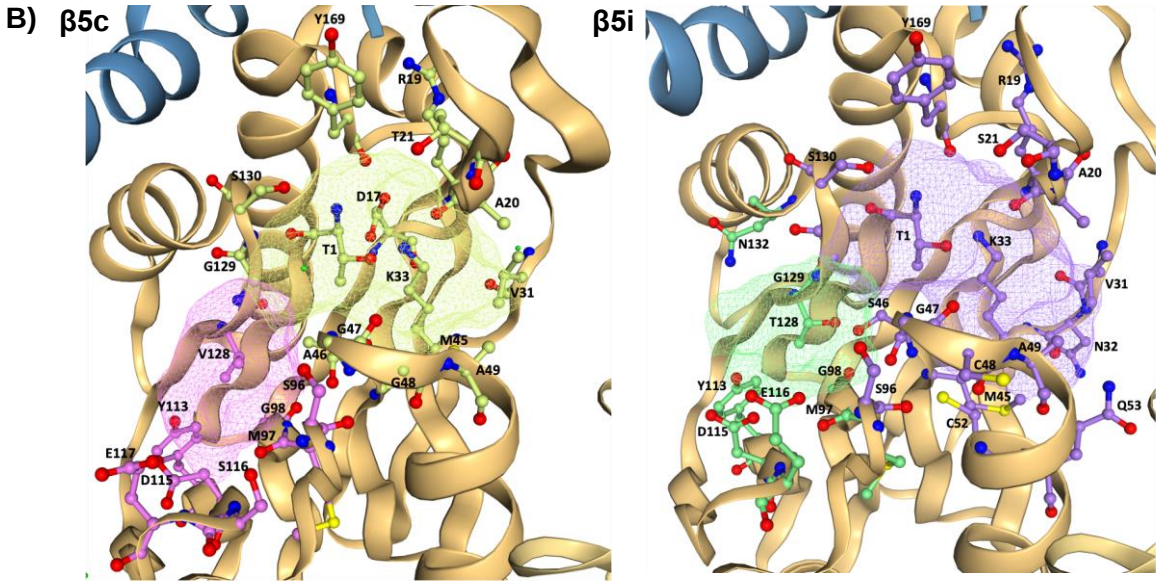
most potent analogues C1-4 against the immunoproteasome to first establish whether inhibition of the immunoproteasome is achieved and if so, the extent of selectivity towards either of the proteasomes.

As exemplified by the IC_{50} values all four compounds showed β5i inhibition at low micromolar level but with slight preference towards the proteasome for analogues C1-3 (Table 2). The slight preference towards proteasome can be attributed to the more hydrophobic nature of β5c compared to β5i, enabling for better hydrophobic interactions with the naphthyl ring. A more detailed comparison of β5c and β5i subunits, shows the key amino acids needed for activity to be conserved among the two sites, even though distances and surface topology vary (Figure 3A). The volume of β5i specificity pocket S1 is 30% greater in volume than β5c, largely accounted by a 30% increase in depth also showing a greater ratio of polar amino acid compared to β5c. Similarly, while S2 pockets between the β5 sites are of similar size, a more pronounced nonpolar character is shown at β5c vs β5i (Figure 3B-C).

Table 2. Mean IC_{50} values and 95% confidence interval for analogues C1-4 determined from purified enzyme assays at β5i at 3nM enzyme concentration (n=3).

Compound	**C1**	**C2**	**C3**	**C4**
β5i IC_{50} (μM)	22.12	8.44	31.64	5.6
95% CI (μM)	18.40-26.75	7.24-9.90	25.13-40.07	4.59-6.81


A)
β5c
β5i
A.
Ala20
Val31
Ala49
Thr1
Met45
Cys52
B.
Met45
4.5
Ser53
3.5
Gln53
6.1
9.5
5.7
8.7
4.3
8.2
11.4
7.2
3.9
9.1


B)
β5c
β5i
Y169
R19
T21
S130
D17
A20
T1
G129
K33
V31
G47
M45
V128
A46
A49
S96
G48
Y113
G98
M97
E117
S116
D115
S21
N132
S46
T128
C48
E116
Q53
C52

C)

Site	Sub pocket	Volume (Å)	Surface ($Å^2$)	Depth (Å)	H bond donors	H bond acceptors	Hydrophobic interactions	Hydrophobicity ratio	Non-polar AA ratio	Polar AA ratio
β5c	S1	269.31	375.05	10.61	19	23	23	0.47	0.33	0.47
β5i	S1	348.61	470.45	13.51	27	29	29	0.43	0.22	0.67
β5c	S2	142.46	206.13	7.65	12	12	7	0.26	0.25	0.58
β5i	S2	131.2	203.7	9.53	24	24	6	0.18	0.09	0.73

Figure 3. A) Shows distances and surface topology amongst key S1 non primed binding pocket residues in β5c/ β5i typically responsible for tight substrate interactions determining cleavage specificity. a) key residues coloured from S1 non primed binding pockets and Thr1, b) changes to architecture from nearby residue S53Q substitution leading to an increase in pocket size. Angstrom measurements showing changes to distances between key residues. **B)** Predicted ProteinPlus DoGSiteScorer pockets showing the β5c (left) S1 and S2 pocket regions in light green and pink, respectively and the β5i (right) S1 and S2 pocket regions in purple and green, respectively. **C)** Quantified physiochemical properties and architecture of the β5 sites in the proteasome and immunoproteasome (β5c PDB: 5LE5, β5i PDB: 6E5B)[38,39].

Compounds that differ only at naphthyl R1 are all statistically different ($p<0.0001$) in binding energies at both β5 sites, favouring naphthyl over methyl, pyridine or benzene (Table S2). As shown in figures 2 and 4, the naphthyl is accommodated in varied locations and does not determine a consistent pose at either site, yet additional interactions gained are key to enhancing potency. Interestingly, despite a similar hydrophobicity ratio between the β5c and β5i S1 pockets and the increased size in β5i (348Å vs 269Å), only C3 shows hydrophobic interactions at β5i S1 (between the naphthyl ring and residues ala20, met45, val31 and ala49). C3 favours β5c over β5i in both docking simulations and IC_{50} values. A likely contributing factor is that in C3 non-polar characteristics are further enhanced with a benzene at R2 that allows further interactions at β5c S1 strong enough to contribute to an overall favourable binding profile. This highlights the importance of inhibitor flexibility and that various hydrophobic regions can contribute towards valuable inhibitor interactions. The same can be argued for C1 which shows the highest ranked binding energies amongst all analogues (Table S2), with no significant preference for either site and varying binding poses maximizing interactions at each site. On the other hand, C2 is predicted to bind with similar potency at β5c and β5i and this is supported by the similar interactions that are observed at both sites i.e hydrogen

bonding with backbone amide of ser31 and ala20, the hydroxyl group of ser27 and two π-π stacking interactions between R2 chloro benzene and tyr130 phe124 (Figure S6, S7). The actual preference of C2 over β5c may thus be attributed to the halogen bonding interactions present in β5c but not in β5i. In β5i the distance between chlorine and the nearest carbonyl oxygen is over 5Å, outside the normal range for halogen bonds. Interestingly, a similar binding preference of C4 for each site is shown by both predicted binding energy values and IC_{50} data. This can be explained by a relatively consistent binding pose of C4 within both sites that enables same interactions. Specifically, the R2 methoxy benzene group in both sites is accommodated within the S1 pocket facilitating the same hydrogen bonding interaction with the backbone amide of thr21, ala20 (β5c) and ser21, ala20 (β5i) amino acids. Further to this the naphthyl group is orientated towards asp115, in both sites.

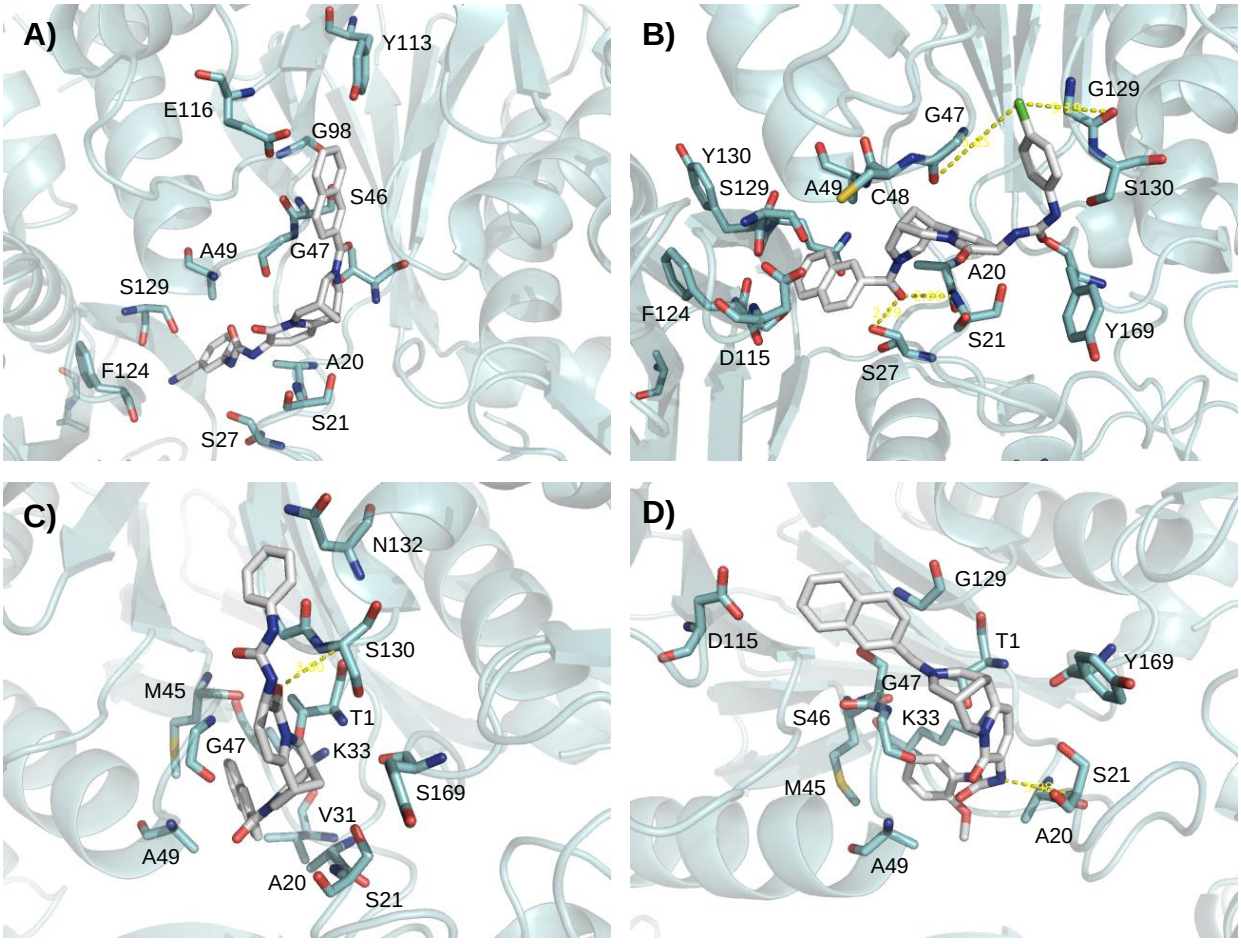


Figure 4. Best docked conformations showing interacting amino acids with A) C1; B) C2: S21*, S27*, F124**, Y130**; C) C3: S130*; D) C4: A20*, S21*; within the β5i active site. *Denotes amino acids that participate in hydrogen bonding, ** denotes amino acids that participate in π-π interactions, ***denotes amino acids that participate in halogen bonding.

Initial research into proteasome degradation in yeast *Saccharomyces cerevisiae* indicated the β5c subunit as the rate-limiting unit in protein degradation[40,41]. As such, proteasome inhibitors were predominantly designed to target β5c however, recent studies have shown that β5-selective inhibitors are not clinically effective and that inhibition of more than one active site is required in order to achieve a desirable clinical outcome[42,43]. With this in mind we further examined the binding of compounds C1-9 at the β1c/β1i and β2c/β2i sites using molecular docking. At the β2c/β2i sites slight preference is shown for C1-3 and C1-4,9, respectively though the difference in the energetics of binding to compounds C5-8 is much less pronounced as compared to β5c/β5i. This may be expected as the β2c/β2i sites have a more polar character accommodating for basic substrates[44] and substitution of the hydrophobic naphthyl ring with small groups may have less effect on binding. It must be noted that C1 is estimated to bind the strongest at β2c, K_I estimated at 0.07nM, however, in the absence of experimental data this should be regarded with care (Table S2). Best docked conformations show most of the analogues adopting an orientation whereby the naphthyl ring points away from the polar amino acids of the S1 pocket (Figure S10). In contrast to the similarity in the β2c and β2i sites, β1c is considerably different to β1i. Compared to the β1c active site, β1i is smaller in size and has a more hydrophobic character which is exemplified by the T20V, T31F, R45L and T52A substitutions[19]. Despite these differences similar binding predictions are shown for both sites showing binding preference to compounds C1-4, C6 and C9 (Table S2). Overall molecular docking predictions favour C1 binding at all active sites of the proteasome and immunoproteasome with C2 as a second close favourite, though as discussed above this may not be the case in vitro.

Conclusion

This study led to the identification of potentially a new class of proteasome inhibitors with an intriguing azotricyclic urea structure and naphthyl and substituted benzene groups as side chains. Overall purified assays indicate the analogue bearing the p-chlorobenzene substituent as the most potent, suggesting key interactions including halogen bonding and π-π stacking. An overwhelming positive impact of the naphthyl group garners valuable insight for methods to enhance potency in proteasome inhibitor drug development which is in agreement with other proteasome inhibitors utilising hydrophobic regions.

Author Contributions

DA has made a substantial contribution to the project investigation including molecular modelling and biochemical assays, formal analysis, validation and visualization; PAM has made a substantial contribution to the molecular modelling data acquisition; RK and MS have made a substantial contribution to the project investigation including the initial virtual screening data; EL has made a substantial contribution to the concept and design of the article, formal analysis and visualization, project administration and supervision and writing the original draft, review and editing.

Conflict of Interests

The are no conflicts of interest to declare.

Acknowledgements

Funding from RSC undergraduate research bursary and Middlesex University is gratefully acknowledged.

3. Experimental

3.1 Material

20S proteasome (purified human erythrocyte), abbreviated as CP, 20S immunoproteasome (purified human enzyme), abbreviated as IP, NBS 96-well microplates, 7-amino-4-methylcoumarin (AMC) standards, Suc-Leu-Leu-Val-Tyr-AMC, epoxomycin, sodium dodecyl sulfate (SDS) were purchased from Enzo Life Sciences, Exeter UK. Ac-Pro-Ala-Leu-AMC was purchased from BioTechne Abingdon, UK, dimethyl sulfoxide (DMSO) from Fisher Scientific, Loughborough, UK. Compounds ZINC4258888 (C1), ZINC4258885 (C2), ZINC4258889 (C3), ZINC4258891 (C4), ZINC4258959 (C5), ZINC4259018 (C6), ZINC4259049 (C7), ZINC4259050 (C8), ZINC4260270 (C9) were purchased from MolPort, Latvia.

3.2 Purified enzyme assays

Substrate and inhibitor reagents were dissolved in DMSO for stock solutions and subsequently diluted in proteasome assay buffer to desired concentrations. Purified enzymatic assays contained human 20S IP or CP with Suc-LLVY-AMC and inhibitor, diluted with buffer (50 mM Tris/HCl, 25 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 0.03% SDS) in a ½ volume, non-binding surface, 96-well white microplate. The plate was equilibrated for 10 minutes at 37°C prior to kinetic readings of AMC liberation using a BMG Labtech fluorescent plate reader at 355/460 nm (excitation/emission), set with an appropriate gain. Tests included: AMC standards (8 µM to 0.25 µM and

blank), positive control in the absence of inhibitor, blank in the absence of enzyme, DMSO vehicle control and negative controls using epoxomicin a potent inhibitor of β5/β5i. All conditions were performed in triplicate followed by 3 independent repeats. To determine IC_{50}, Michaelis-Menten kinetics established an enzyme concentration (1nM CP and 3nM IP, Figure S1 and S4, respectively) with associated Km values for substrate concentration then tested at a logarithmic scale of 8-12 inhibitor concentrations. In analysis, the initial rate of reaction was determined from the linear phase of the graph, at which less than 10% of substrate had been consumed. Expressed as a percentage of control, a normalised response curve with variable hill slope function non-linear regression fit was applied using GraphPad Prism v8. Using the same purified assay method, to determine type of inhibition and inhibition constants, ranges of IC_{50} x3, 1, 0.33, 0 each at Km x5, 2.5, 1.25, 0.625 and 0.3125 were tested at β5. Using GraphPad Prism v8, Ki SNLR inhibition models were compared using Akaike's information criteria (AICc), since competitive models are not nested with noncompetitive, uncompetitive or mixed. This quantified the probability a model is correct based upon sum-of-squares that measures goodness of fit and degrees of freedom[45,46]. Following this the appropriate competitive (or otherwise) inhibition simulations non-linear regression (SNLR) model was fit to report Ki with SEM and 95% CI. Competitive inhibition was the favoured model for all C1-4 at β5 (representative values shown for C1: competitive inhibition >99.99% vs noncompetitive inhibition <0.01%, difference in AiCC -21.87; competitive inhibition >99.99% vs uncompetitive inhibition <0.01%, difference in AiCC -43.59; competitive inhibition 97.97% vs mixed model inhibition 2.03%, difference in AiCC -7.76).

3.3 Computational methods

Crystal structure data was obtained from PDB, filtered by resolution and organism, then further quality assessed. Electron density score for overall structure and individual atoms was analysed using ProteinPlus StructureProfiler EDIA analysis[38]. In addition, SwissPDBViewer4.1.0 was used to compute force field energies of bonds and angles, to determine topology, bond distortion and backbone problems[47]. Human CP (PDB:5LE5) at 1.8 Å (Schrader et al., 2016b) and human IP (PDB:6E5B) at 2.77 Å were identified as the best quality macromolecules for active site molecular docking at the time of research[48]. Further software was used to investigate key features and characteristics of binding pockets. ProteinsPlus DoGSiteScorer (https://proteins.plus/) predicted binding pockets within the crystal structures, using a grid-based method with difference of gaussian filter[38]. This further quantified architecture and physiochemical properties of pockets. Within subunits, hydrophobic regions were analysed using Kyte and Doolittle scale, visualised with Chimera 1.14[49].

Finally, CryptoSite analysis (http://salilab.org/cryptosite) investigated any possible cryptic sites using sequence conservation and patchmap fragment docking[50].
Active sites β5 and β5i were prepared for docking with cuts of residues within 32Å of the catalytic Thr1 from the same β-ring on the same monomer using PyMOL as previously described[23]. Ligand structures were obtained from the online database of commercially available compounds, ZINC15[51] filtered to naturally isolated compounds, commercially available, with molecular weights between 500 and 1000 Da. The 3D structures of the first 5000 compounds to meet these criteria were downloaded and their geometry optimized using the molecular mechanics with UFF force field function of Avogadro[52]. These optimised compounds were then subjected to virtual screening using AutoDockVina[53] (grid box coordinates centred on macromolecule) and the top 20 scoring compounds (Table S1) were further analysed for bonding interactions within the β5 and β5i active sites, using AutoDock4.2[54]. The grid box was set at 70x, 70y, 70z dimensions, centred based on Thr1 co-ordinates at the β5c active subunit (-42.794, 25.372, -104.397) and β5i (-5.427, 23.332, 4.294). The ZINC ligands were allowed rotational freedom while the β5c and β5i sites were treated as rigid. A genetic algorithm with 100 runs was selected for all docking experiments. Each test was run in triplicate from which the lowest binding energy was selected for analysis of interactions. Molecular docking with Autodock 4.2 was repeated at β1c/β1i and β2c/β2i (grid box coordinates β1: 49.856, 234.315, 45.869; β1i: -34.583, -29.345, -7.232; β2: 218.845, 215, 354, 214.164; β2i: -54.895, -13.643, -12.544) using identical parameters as for β5/ β5i.
Assuming normality and equal variance, one-way ANOVA analysis tested for analogues that differ in binding energy at the same site. GraphPad Prism v8 was used to perform ANOVA and pairwise comparison with Tukey's simultaneous test for differences of means at 95% confidence interval ($p<0.05$). Multiple t-tests were used to measure differences between the same analogue binding at different sites.

Supplementary data

Supplementary information available: estimated binding energies and inhibition constants, autodock poses, Michaelis-Menten Km plots, IC_{50} and Ki plots.

References

1 E. E. Manasanch and R. Z. Orlowski, *Nat. Rev. Clin. Oncol.*, 2017, **14**, 417–433.
2 A. Ciechanover, *Cell Death Differ.*, 2005, **12**, 1178–1190.
3 K. S. Suh, T. Tanaka, S. Sarojini, G. Nightingale, R. Gharbaran, A. Pecora and A. Goy, *Crit. Rev. Oncol. Hematol.*, 2013, **87**, 306–322.
4 M. Orlowski and S. Wilk, *Arch. Biochem. Biophys.*, 2000, **383**, 1–16.
5 W. Heinemeyer, P. C. Ramos and R. J. Dohmen, *Cell. Mol. Life Sci. CMLS*, 2004, **61**, 1562–1578.

6 M. A. Gräwert and M. Groll, *Chem. Commun.*, 2012, **48**, 1364–1378.
7 G. E. Hubbell and J. J. Tepe, *RSC Chem. Biol.*, 2020, **1**, 305–332.
8 A. T. Nunes and C. M. Annunziata, *Semin. Oncol.*, 2017, **44**, 377–380.
9 S. Fogli, S. Galimberti, V. Gori, M. Del Re and R. Danesi, *Pharmacol. Res.*, 2021, **167**, 105537.
10 I. Nickeleit, S. Zender, F. Sasse, R. Geffers, G. Brandes, I. Sörensen, H. Steinmetz, S. Kubicka, T. Carlomagno, D. Menche, I. Gütgemann, J. Buer, A. Gossler, M. P. Manns, M. Kalesse, R. Frank and N. P. Malek, *Cancer Cell*, 2008, **14**, 23–35.
11 G. R. Tundo, D. Sbardella, A. M. Santoro, A. Coletta, F. Oddone, G. Grasso, D. Milardi, P. M. Lacal, S. Marini, R. Purrello, G. Graziani and M. Coletta, *Pharmacol. Ther.*, 2020, **213**, 107579.
12 S. Barrio, T. Stühmer, M. Da-Viá, C. Barrio-Garcia, N. Lehners, A. Besse, I. Cuenca, A. Garitano-Trojaola, S. Fink, E. Leich, M. Chatterjee, C. Driessen, J. Martinez-Lopez, A. Rosenwald, R. Beckmann, R. C. Bargou, E. Braggio, A. K. Stewart, M. S. Raab, H. Einsele and K. M. Kortüm, *Leukemia*, 2019, **33**, 447–456.
13 A. Kniepert and M. Groettrup, *Trends Biochem. Sci.*, 2014, **39**, 17–24.
14 J. W. Yewdell, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 9089–9090.
15 K. C. Carmony, D.-M. Lee, Y. Wu, N.-R. Lee, M. Wehenkel, J. Lee, C.-G. Zhan and K.-B. Kim, *Bioorg. Med. Chem.*, 2012, **20**, 607–613.
16 Y. K. Ho, P. Bargagna-Mohan, M. Wehenkel, R. Mohan and K.-B. Kim, *Chem. Biol.*, 2007, **14**, 419–430.
17 C. Dubiella, H. Cui, M. Gersch, A. J. Brouwer, S. A. Sieber, A. Krüger, R. M. J. Liskamp and M. Groll, *Angew. Chem. Int. Ed Engl.*, 2014, **53**, 11969–11973.
18 G. de Bruin, E. M. Huber, B.-T. Xin, E. J. van Rooden, K. Al-Ayed, K.-B. Kim, A. F. Kisselev, C. Driessen, M. van der Stelt, G. A. van der Marel, M. Groll and H. S. Overkleeft, *J. Med. Chem.*, 2014, **57**, 6197–6209.
19 E. M. Huber, M. Basler, R. Schwab, W. Heinemeyer, C. J. Kirk, M. Groettrup and M. Groll, *Cell*, 2012, **148**, 727–738.
20 M. Basler, C. Lauer, J. Moebius, R. Weber, M. Przybylski, A. F. Kisselev, C. Tsu and M. Groettrup, *J. Immunol.*, 2012, **189**, 1868–1877.
21 H. Fan, N. G. Angelo, J. D. Warren, C. F. Nathan and G. Lin, *ACS Med. Chem. Lett.*, 2014, **5**, 405–410.
22 R. de L. A. Santos, L. Bai, P. K. Singh, N. Murakami, H. Fan, W. Zhan, Y. Zhu, X. Jiang, K. Zhang, J. P. Assker, C. F. Nathan, H. Li, J. Azzi and G. Lin, *Nat. Commun.*, 2017, **8**, 1692.
23 D. J. Allardyce, C. M. Bell and E. Z. Loizidou, *Chem. Biol. Drug Des.*, 2019, **94**, 1556–1567.
24 M. Basler and M. Groettrup, *Genes Immun.*, 2020, **21**, 273–287.
25 A. G. Atanasov, S. B. Zotchev, V. M. Dirsch and C. T. Supuran, *Nat. Rev. Drug Discov.*, 2021, **20**, 200–216.
26 F. F. Fleming, L. Yao, P. C. Ravikumar, L. Funk and B. C. Shook, *J. Med. Chem.*, 2010, **53**, 7902–7917.
27 R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger and F. M. Boeckler, *J. Med. Chem.*, 2013, **56**, 1363–1388.
28 S. Jena, J. Dutta, K. D. Tulsiyan, A. K. Sahu, S. S. Choudhury and H. S. Biswal, *Chem. Soc. Rev.*, 2022, **51**, 4261–4286.
29 H. Akaike, *IEEE Transactions on Automatic Control,* 1974, **19**, 716–723.

30 S.-Y. Huang, S. Z. Grinter and X. Zou, *Phys Chem Chem Phys*, 2010, **12**, 12899–12908.
31 K. M. Elokely and R. J. Doerksen, *J Chem Inf Model*, 2013, **53**, 1934–1945.
32 D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nat. Rev. Drug Discov.*, 2004, **3**, 935–949.
33 B. K. Shoichet, *Nature*, 2004, **432**, 862–865.
34 T. F. Vieira and S. F. Sousa, *Appl. Sci.*, 2019, **9**, 4538.
35 D. J. Kuhn, S. A. Hunsucker, Q. Chen, P. M. Voorhees, M. Orlowski and R. Z. Orlowski, *Blood*, 2009, **113**, 4667–4676.
36 M. D. A. Kioon, M. Pierides, T. Pannellini, G. Lin, C. F. Nathan and F. J. Barrat, *J. Immunol.*, 2021, **206**, 1631–1641.
37 C. Zhang, H. Zhu, J. Shao, R. He, J. Xi, R. Zhuang and J. Zhang, *Future Med. Chem.*, 2020, **12**, 269–272.
38 R. Fährrolfes, S. Bietz, F. Flachsenberg, A. Meyder, E. Nittinger, T. Otto, A. Volkamer and M. Rarey, *Nucleic Acids Res.*, 2017, **45**, W337–W343.
39 A. Volkamer, D. Kuhn, T. Grombacher, F. Rippmann and M. Rarey, *J. Chem. Inf. Model.*, 2012, **52**, 360–372.
40 W. Heinemeyer, M. Fischer, T. Krimmer, U. Stachon and D. H. Wolf, *J. Biol. Chem.*, 1997, **272**, 25200–25209.
41 C. S. Arendt and M. Hochstrasser, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 7156–7161.
42 A. F. Kisselev, A. Callard and A. L. Goldberg, *J. Biol. Chem.*, 2006, **281**, 8582–8590.
43 A. Besse, L. Besse, M. Kraus, M. Mendez-Lopez, J. Bader, B.-T. Xin, G. de Bruin, E. Maurits, H. S. Overkleeft and C. Driessen, *Cell Chem. Biol.*, 2019, **26**, 340-351.e3.
44 W. Harshbarger, C. Miller, C. Diedrich and J. Sacchettini, *Structure*, 2015, **23**, 418–424.
45 Motulsky, H., *Fitting models to biological data using linear and nonlinear regression: A practical guide to curve fitting.*, Oxford University Press, Oxford, vol. 2003.
46 T. Kakkar, Y. Pak and M. Mayersohn, *J. Pharmacol. Exp. Ther.*, 2000, **293**, 861–869.
47 N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714–2723.
48 E. Ladi, C. Everett, C. E. Stivala, B. E. Daniels, M. R. Durk, S. F. Harris, M. P. Huestis, H. E. Purkey, S. T. Staben, M. Augustin, M. Blaesse, S. Steinbacher, C. Eidenschenk, R. Pappu and M. Siu, *J. Med. Chem.*, 2019, **62**, 7032–7041.
49 J. Kyte and R. F. Doolittle, *J. Mol. Biol.*, 1982, **157**, 105–132.
50 P. Cimermancic, P. Weinkam, T. J. Rettenmaier, L. Bichmann, D. A. Keedy, R. A. Woldeyes, D. Schneidman-Duhovny, O. N. Demerdash, J. C. Mitchell, J. A. Wells, J. S. Fraser and A. Sali, *J. Mol. Biol.*, 2016, **428**, 709–719.
51 T. Sterling and J. J. Irwin, *J. Chem. Inf. Model.*, 2015, **55**, 2324–2337.
52 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminformatics*, 2012, **4**, 17.
53 T. O and O. Aj, *J. Comput. Chem.*, , DOI:10.1002/jcc.21334.
54 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.