ChemComm

A targeted covalent small molecule inhibitor of HIV-1 fusion

Journal:	*ChemComm*
Manuscript ID	CC-COM-02-2021-001013.R2
Article Type:	Communication

COMMUNICATION

A targeted covalent small molecule inhibitor of HIV-1 fusion

Guangyan Zhou,[a] Li He,[a] Kathy H. Li [b], Cássio C. S. Pedroso [c] and Miriam Gochin* [a,b]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

We describe a low molecular weight covalent inhibitor targeting a conserved lysine residue within the hydrophobic pocket of HIV-1 glycoprotein-41. The inhibitor bound selectively to the hydrophobic pocket and exhibited an order of magnitude enhancement of anti-fusion activity against HIV-1 compared to its non-covalent counterpart. The findings represent a significant advance in the quest to obtain non-peptide fusion inhibitors.

Gp41, a component of the HIV-1 spike protein, has been an attractive target for drugs preventing HIV entry and infection ever since elucidation of the post-fusion ectodomain structure.[1] A deep hydrophobic pocket (HP) on the N-heptad repeat (NHR) trimer interacts with a conserved WxxWDxxI-motif in the HP binding domain (HPbd) on the C-heptad repeat (CHR) helices. The HP has been explored extensively as a potential site for inhibition of HIV fusion by small molecules.[2] It mediates a protein – protein interaction that is central to gp41 refolding, triggered by viral attachment and necessary for fusion to occur. The hydrophobic pocket is an important hotspot in the protein – protein energy landscape.[3] However, none of the inhibitors developed so far have the low nM activity needed to justify further exploration as drugs.

The HP is defined by the residues 565-LLQLTVWGIKQLQARIL-581 and is highly conserved (Supplementary Information). Conservation is likely related to the dual role of this region in stabilizing the protein – protein interaction in the fusion reaction, and as a structured RNA in the RRE - Rev interaction important in nuclear export and packaging[4]. Few mutations have been observed, even upon direct challenge with pocket binding C-peptides[5] and potent D-peptides[4]. Importantly there is an invariant lysine at position 574, which presents an opportunity for targeted covalent inhibition, in which an inhibitor with electrophilic functionality binds selectively to its target, followed by spontaneous reaction with a nucleophilic group on the protein.[6] As a ubiquitous residue in proteins, lysine is a convenient nucleophile, although its high pKa may limit its reactivity.[7] Studies have indicated that lysine-574 can form an important salt bridge or ionic interaction with some gp41 inhibitors and with an aspartate residue in the HPbd.[2c, 8] Previous literature reports showed that introducing a maleimide or thioester into a C-peptide to target Lys-574 ε-NH_2 for covalent modification stabilized and trapped the fusion intermediate and permanently blocked fusion, while retaining the potency observed for the unmodified peptide.[9]

In previous work, we identified compound **1**, a 6,6'-bisindole inhibitor containing a benzoic acid group and a benzoic acid ester attached at the two indole nitrogens (Figure 2, inset).[2f, 2g] **1** displayed sub-μM binding affinity for the HP in a competitive inhibition assay measuring displacement of the HPbd peptide. It gave 200 nM IC_{50}'s in a cell - cell fusion assay and a viral replication assay.[2g] Having

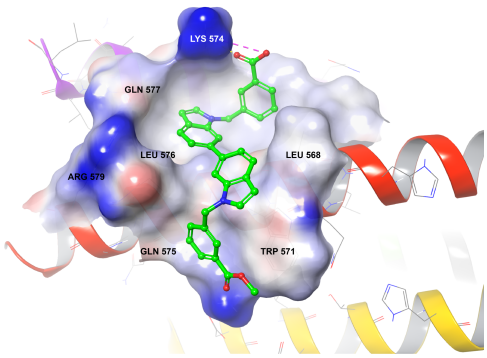

Figure 1. Docked pose of **1** in the hydrophobic pocket of gp41 (PDB *2xra*), produced by induced fitting (Schrodinger Inc.) allowing movement of side chains of Gln 575 and Trp 571. A salt bridge from one carbonyl oxygen on the ligand to Lys574-εNH_2 is shown as an orange dotted line, and a hydrogen bond is predicted from the second carbonyl oxygen to the lysine εNH_2.

[a] *Department of Basic Sciences, Touro University California, 1310 Club Drive, Mare Island, Vallejo CA 94592*
[b] *Department of Pharmaceutical Chemistry, UCSF School of Pharmacy, San Francisco, CA 94143*
[c] *Lawrence Berkeley National Laboratory, The Molecular Foundry, 1 Cyclotron Road, 67R5114, Berkeley, CA 94720*
*corresponding author, e-mail: miriam.gochin@tu.edu
†Electronic Supplementary Information (ESI) available: Supplementary figures, experimental details, compound characterisation. See DOI: 10.1039/x0xx00000x

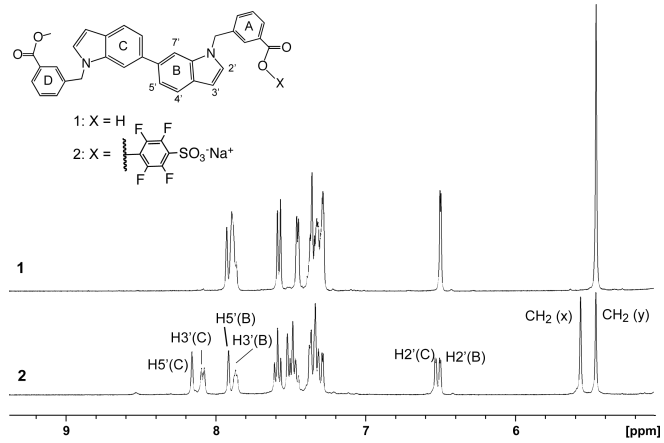

Figure 2. Aromatic ^{1}H NMR spectra of **1** and **2**. In the STP ester, downfield chemical shifts occurred for one of the indole rings, while the other was unchanged. The tentative assignment shown for well-resolved resonances was based on the predicted 3D structure of **2** (Supplementary Information).

identified both binding site and reasonable potency, we set out to add an electrophilic warhead to **1** to effect covalent association with Lys 574. Rosetta calculations predicted a small pKa perturbation of the lysine ε-NH_2 from 10.4 to the range 9.8–10.3 pH units (Supplementary Information).

Docking studies suggested an orientation of **1** in the HP in which one indole group and one benzoic acid moiety fit into the deepest part of the pocket and the second indole and benzoate extended along a groove bordered by Leu, Gln and Trp residues, with possible π-π interactions with Trp571 (Figure 1 and reference[2g]). To accommodate side chain flexibility,[10] especially along the groove, the docking was performed using an induced fit model in which Gln 575 and Trp 571 side chains were allowed to move. A large number of non-polar interactions were identified between **1** and residues forming the pocket, and both a hydrogen bond and a salt-bridge were predicted between the ligand carboxylic acid and Lys-574 ε-NH_2. We therefore prepared a derivative of **1** with the carboxylic acid activated as a sulfotetrafluorophenyl ester that can potentially react with the ε-NH_2 group of Lys-574 to form an irreversible covalent bond. The sulfotetrafluorophenyl ester has been shown to be an excellent leaving group for targeting surface lysines[7a, 11], while other electrophilic groups have been less reactive towards lysines with unperturbed pKa, or were less stable in aqueous solution.[7b, 12]

We therefore prepared and characterized compound **2** as the sulfotetrafluorophenyl (STP) ester of **1** (Supplementary Information). The nearly symmetrical **1** gave an NMR spectrum in which resonances of the B and C indole rings were almost overlapping (Figure 2). This was clear from the well-resolved peaks of protons at the indole 2', 3' and 5' positions, and in the methylene chemical shifts. Presence of the STP ester in **2** resulted in a significant downfield shift for one of the sets of indole resonances, while leaving the second set of indole chemical shifts relatively unchanged. The NMR spectrum was consistent with a solution structure in which the tetrafluorophenyl group was folded back on to the rest of the structure (Supplementary Information).

We tested reactivity of **2** with the side chain amino group of lysine by combining **2** (1.7 mM) with an excess of N-acetyl-lysine (13.3 mM) in sodium phosphate buffer, pH 8 and incubating at 37°C for 1 – 24 hours. Aliquots were withdrawn at several time points and tested by HPLC (Supplementary Information). Adduct formation was clearly observed after an hour, and grew over several hours. The reaction did not appear to go to completion even after 24 hours, suggesting either that **2** was not completely consumed or that it had degraded into **1** which had an almost identical retention time. Previous studies have shown that the STP ester is very stable over 24 hours in aqueous medium,[7a] and we have noted no degradation in DMSO at -10°C for months.

In order to assess the ability of **2** to form a covalent complex with gp41, we utilized a protein construct of the gp41 ectodomain that contains an exposed hydrophobic pocket. We have previously described a reversed hairpin protein C28(L4)N50, in which a 28-residue CHR domain preceded the 50 residue NHR in the sequence with a short 4 residue connecting loop. Truncation of the CHR compared to the full length 39 residue CHR in the six-helix bundle[13] permitted exposure of the HP in solution,[14] providing an avenue for analysing binding in the pocket. For the current study, we used a slightly modified peptide C26(L4)N50, containing some additional salt-bridges in outer residues of the CHR and only two lysines in the sequence (Figure 3 and Supplementary Information).

2 was incubated with C26(L4)N50 (225 μM) in 50 mM phosphate buffer at pH 8 for 3 hours at 37°C, using stoichiometries 1, 2, 5 and 9X of the small molecule. Maldi results (Figure 3C) clearly showed a new peak forming with a molecular weight 10179 ± 4 da, corresponding to an increase of 496 ± 4 from the base molecular weight of the protein 9703 da. Complete conversion to fully labelled protein did not occur. Despite the presence of two lysine residues in

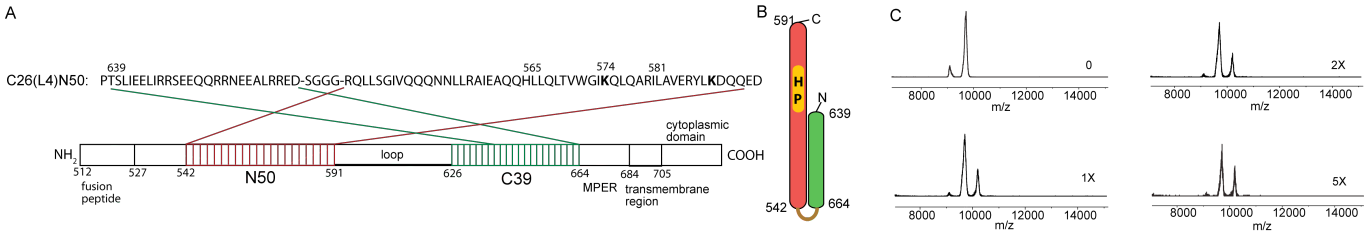

Figure 3. A. Protein construct C26(L4)N50 used to evaluate pocket binding and adduct formation. NHR and CHR domains are swapped relative to their order in gp41, shown directly below the sequence. Boundaries of the different domains of gp41 are given (HXB2 numbering) as well as the length of the NHR (50 residues) and CHR (39 residues). The CHR is truncated in the construct and the two lysine residues are bolded. B. Schematic figure of folded construct, revealing an exposed HP (residues 565-581). C. Maldi-ESI-MS spectra of reaction mixtures of **2** with C26(L4)N50 at the indicated stoichiometries., shown over the range 7 – 15 kDa. At stoichiometry 0, DMSO was used in the reaction.

the sequence, no doubly labelled protein was detected, even at 9X stoichiometry of **2** (not shown). This experiment was followed by trypsin digest and LC-MS/MS to determine the site of adduct formation (Supplementary Information). Mass modification at the Lys-574 residue was indicated, with a mass change of 496.1887. No modification at the second lysine in the protein sequence was detected. The MW increase of 496 da corresponds to formation of the covalent adduct of **2** with the protein, matching the mass of **1** (514.1893) minus a water molecule. This is the expected result of a reaction between the lysine ε-NH_2 and activated ester **2**. The suggested formula of the linkage was $C_{33}H_{24}N_2O_3$. The results unambiguously showed that the modified lysine is Lys 574 in the pocket, confirming that the specificity of equilibrium binding directed covalent bond formation.

Having established the selectivity and reactivity of **2** for the hydrophobic pocket, we examined the effect of STP esterification on anti-fusion activity. In vitro cell culture experiments were performed using TZM-bl cells, which contain an integrated reporter gene for firefly luciferase governed by the HIV-1 LTR promoter.[15] In antiviral experiments, TZM-bl cells were infected with single-round infectious pseudotyped virus particles displaying HIV envelope from CXCR4 tropic (HXB2) or CCR5-tropic (JRFL) strains, in the absence or presence of the compounds. To elucidate compound specificity for HIV-1 envelope, controls were also run using recombinant virus pseudotyped with the envelope glycoprotein from Amphotropic Murine Leukemia Virus (A-MLV), an unrelated retrovirus.[16] For cell – cell fusion experiments, HL2/3 cells[17] expressing HXB2-Env were added to TZM-bl cells in the absence or presence of the compounds. A chemiluminescent readout was used to measure the extent of fusion.

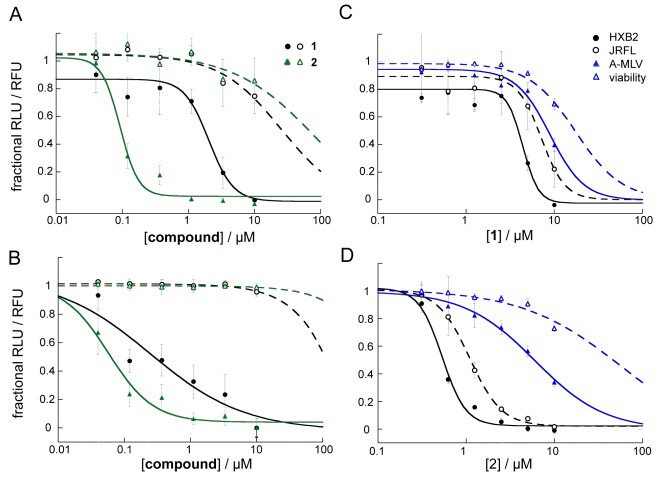

Figure 4. A, B. Dose response curves for **1** (black) and **2** (green) in an HXB2 - pseudotyped virus infection assay (A) and HXB2-Env mediated cell – cell fusion assay (B). IC_{50}'s were obtained by fitting luminescence data (solid symbols, solid lines) and cell viability was measured by fluorescence using resazurin reagent (open symbols, dashed lines). Experiments were repeated in sextuplicate and error bars shown are the standard deviation. C, D. Dose response curves for **1** (C) and **2** (CD) in infection assays using HXB2-Env (black, solid symbols, solid lines), JRFL-Env (black, open symbols, dashed lines) or A-MLV (blue, solid symbols, solid lines) pseudotyped virus. Corresponding cell viability is shown (blue, open symbols, dashed lines).

Cell viability was concomitantly measured using a resazurin reagent.[2g] Details are provided in the Supplementary Information.

2 was an order of magnitude more potent than **1** in viral infectivity assays using HXB2 pseudotyped virus. Figure 4A is a plot of data obtained in sextuplicate, giving IC_{50}'s of 2.0 ± 0.5 μM and 0.09 ± 0.02 μM for **1** and **2** respectively, with no marked toxicity. IC_{50} variations up to a factor of 3 were observed in repeat experiments, with the consistent observation of an order of magnitude difference in the IC_{50}'s of **1** and **2**. We noted that the activity of **1** was lower than was previously obtained in a viral replication assay,[2g] perhaps an indication of off-target effects by the hydrophobic inhibitor.

Dose response curves for **1** and **2** in the cell – cell fusion assay indicated a 2 - 4-fold reduction in IC_{50} as a result of STP esterification (Figure 4B). In the example shown, the IC_{50}'s obtained were 250 nM and 60 nM for **1** and **2** respectively, with no observed toxicity. In both cell – cell fusion and viral infectivity assays we noticed a sharp dose response for **2**. Since **2** is not expected to follow equilibrium binding kinetics,[18] fitting to an IC_{50} provides only an approximate evaluation of its activity. Fusion inhibition persisted as inhibitor concentration was reduced until an inflexion point where the compound appeared to lose activity and fusion rose rapidly.

Selectivity of the compounds for HIV-1 Env mediated fusion was tested by extending assays to HIV - JRFL and A-MLV pseudotyped viruses. JRFL is a faster fusing virus than HXB2 and compounds that target the fusion reaction are expected to be correspondingly less potent than against HXB2 pseudotyped virus.[19] Figures 4C and 4D compare the activity of **1** and **2** against the three pseudoviruses. A two-fold higher IC_{50} against JRFL compared to HXB2 was observed for both compounds, as well as a clear preference for HIV-1 Env compared to A-MLV Env mediated fusion. The reduction in potency against A-MLV was much more pronounced for **2** (12-fold) than for **1** (2.5-fold), suggesting an increase in specificity for the covalent inhibitor, whereas the non-covalent but highly hydrophobic inhibitor appeared to show off-target effects. The A-MLV results were also likely to be affected by toxicity, which was > 20% at concentrations ≥ IC_{50}.

We also confirmed the role of the compounds as entry inhibitors in an assay in which compound was added two hours after exposing the cells to HXB2-pseudotyped virus. Fusion occurs at a relatively early stage during infection, after attachment but before release of viral RNA into cells. Activity of a compound targeting fusion is expected to be significantly dampened with the time delay. The simple non-synchronized time-of-addition assay demonstrated complete loss of activity for the known attachment inhibitor AMD3100,[20] a slight loss of activity for **1** (< 2 fold) and an 8-fold loss of activity for **2**. The results again demonstrated higher specificity of **2** for the HIV-1 hydrophobic pocket and inhibition of fusion.

In summary, we have described an STP ester functionalized bis-indole inhibitor that formed a covalent association with the conserved lysine residue in the hydrophobic pocket of HIV-1 gp41. It formed a unique complex, was selective for HIV envelope, and conferred an order of magnitude improved anti-fusion activity compared to the corresponding reversible inhibitor. IC_{50}'s below 100 nM were obtained. This work suggests a novel paradigm for approaching low molecular weight inhibition of HIV-1 fusion. Importantly, it moves away from the "potency requires hydrophobicity" conundrum, which

has dogged fusion inhibitor development and which renders compounds susceptible to off target effects.[21]

This research was supported by NIH grant AI140904. Additional support was provided by Touro University California. KL was supported by the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation. We thank Dr. Al Burlingame for access to the UCSF Mass Spectrometry Facility. Work by MG at the Molecular Foundry, Lawrence Berkeley National Laboratory, was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

Author Contributions: GZ: Methodology, Investigation, Writing – review; LH: Investigation, Methodology; KL: Investigation, Resources; CP: Investigation, Resources; MG: Conceptualization, Formal Analysis, Funding Acquisition, Project Administration, Writing – original draft.

There are no conflicts to declare.

Notes and references

1. D. C. Chan, C. T. Chutkowski and P. S. Kim, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 15613-15617.
2. a) K. Lu, M. R. Asyifah, F. Shao and D. Zhang, *Curr Med Chem*, 2014, **21**, 1976-1996; b) G. Zhou and S. Chu, *Current Pharmaceutical Design*, 2013, **19**, 1818-1826; c) J. Qiu, T. Liang, J. Wu, F. Yu, X. He, Y. Tian, L. Xie, S. Jiang, S. Liu and L. Li, *Front Pharmacol*, 2019, **10**, 859; d) L. Lu, Q. Liu, Y. Zhu, K. H. Chan, L. Qin, Y. Li, Q. Wang, J. F. Chan, L. Du, F. Yu, C. Ma, S. Ye, K. Y. Yuen, R. Zhang and S. Jiang, *Nat Commun*, 2014, **5**, 3067; e) L. Cai and S. Jiang, *ChemMedChem*, 2010, **5**, 1813-1824; f) G. Zhou, S. Chu, A. Nemati, C. Huang, B. A. Snyder, R. G. Ptak and M. Gochin, *Eur J Med Chem*, 2019, **161**, 533-542; g) G. Zhou, V. Sofiyev, H. Kaur, B. A. Snyder, M. K. Mankowski, P. A. Hogan, R. G. Ptak and M. Gochin, *Journal of Medicinal Chemistry*, 2014, **57**, 5270-5281.
3. B. Strockbine and R. C. Rizzo, *Proteins*, 2007, **67**, 630-642.
4. A. R. Smith, M. T. Weinstock, A. E. Siglin, F. G. Whitby, J. N. Francis, C. P. Hill, D. M. Eckert, M. J. Root and M. S. Kay, *Retrovirology*, 2019, **16**, 28.
5. a) Y. Su, H. Chong, Z. Qiu, S. Xiong and Y. He, *Journal of Virology*, 2015, **89**, 5801-5811; b) W. Wang, C. J. De Feo, M. Zhuang, R. Vassell and C. D. Weiss, *Journal of Virology*, 2011, **85**, 12929-12938.
6. T. A. Baillie, *Angewandte Chemie. International Ed. In English*, 2016, **55**, 13408-13421.
7. a) J. Charoenpattarapreeda, Y. S. Tan, J. Iegre, S. J. Walsh, E. Fowler, R. S. Eapen, Y. Wu, H. F. Sore, C. S. Verma, L. Itzhaki and D. R. Spring, *Chem Commun (Camb)*, 2019, **55**, 7914-7917; b) L. Gambini, C. Baggio, P. Udompholkul, J. Jossart, A. F. Salem, J. J. P. Perry and M. Pellecchia, *J Med Chem*, 2019, **62**, 5616-5627; c) J. Pettinger, M. Carter, K. Jones and M. D. Cheeseman, *J Med Chem*, 2019, **62**, 11383-11398; d) C. Hoppmann and L. Wang, *Chem Commun (Camb)*, 2016, **52**, 5140-5143.
8. a) Y. He, S. Liu, W. Jing, H. Lu, D. Cai, D. J. Chin, A. K. Debnath, F. Kirchhoff and S. Jiang, *J Biol Chem*, 2007, **282**, 25631-25639; b) M. Gochin and L. Cai, *J Med Chem*, 2009, **52**, 4338-4344.
9. a) H. A. Yi, B. Diaz-Aguilar, D. Bridon, O. Quraishi and A. Jacobs, *Biochemistry*, 2011, **50**, 6966-6972; b) A. Jacobs, O. Quraishi, X. Huang, N. Bousquet-Gagnon, G. Nault, N. Francella, W. G. Alvord, N. Pham, C. Soucy, M. Robitaille, D. Bridon and R. Blumenthal, *J Biol Chem*, 2007, **282**, 32406-32413; c) Y. Bai, H. Xue, K. Wang, L. Cai, J. Qiu, S. Bi, L. Lai, M. Cheng, S. Liu and K. Liu, *Amino Acids*, 2013, **44**, 701-713.
10. W. J. Allen and R. C. Rizzo, *Biology*, 2012, **1**, 311-338.
11. S. M. Hacker, K. M. Backus, M. R. Lazear, S. Forli, B. E. Correia and B. F. Cravatt, *Nat Chem*, 2017, **9**, 1181-1190.
12. J. Pettinger, K. Jones and M. D. Cheeseman, *Angew Chem Int Ed Engl*, 2017, **56**, 15200-15209.
13. F. Naider and J. Anglister, *Curr Opin Struct Biol*, 2009, **19**, 473-482.
14. a) J. D. Walsh, S. Chu, S. Q. Zhang and M. Gochin, *Protein Eng Des Sel*, 2015, **28**, 107-116; b) S. Chu, H. Kaur, A. Nemati, J. D. Walsh, V. Partida, S. Q. Zhang and M. Gochin, *ACS Chem Biol*, 2015, **10**, 1247-1257.
15. X. Wei, J. M. Decker, H. Liu, Z. Zhang, R. B. Arani, J. M. Kilby, M. S. Saag, X. Wu, G. M. Shaw and J. C. Kappes, *Antimicrob Agents Chemother*, 2002, **46**, 1896-1905.
16. A. Rein, *Adv Virol*, 2011, **2011**, 403419.
17. V. Ciminale, B. K. Felber, M. Campbell and G. N. Pavlakis, *AIDS Res Hum Retroviruses*, 1990, **6**, 1281-1287.
18. J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, *Nat Rev Drug Discov*, 2011, **10**, 307-317.
19. K. Miyauchi, M. M. Kozlov and G. B. Melikyan, *PLoS Pathog*, 2009, **5**, e1000585.
20. S. Rusconi, A. Scozzafava, A. Mastrolorenzo and C. T. Supuran, *Curr Top Med Chem*, 2007, **7**, 1273-1289.
21. a) N. A. Meanwell, *Chem Res Toxicol*, 2011, **24**, 1420-1456; b) P. D. Leeson and B. Springthorpe, *Nat Rev Drug Discov*, 2007, **6**, 881-890; c) G. Zhou, S. Chu, A. Kohli, F. C. Szoka and M. Gochin, *Biochim Biophys Acta Gen Subj*, 2020, **1864**, 129724.

