2-Sulfonylpyrimidines: Reactivity Adjustable Agents for Cysteine Arylation

Maëva M. Pichon[1], Dawid Drelinkiewicz[1], David Lozano[1], Ruxandra Moraru[1], Laura J. Hayward[1], Megan Jones[1], Michael A. Mccoy[1], Samuel Allstrum-Graves[1], Dimitrios-Ilias Balourdas[2,3], Andreas C. Joerger[2,3], Richard W. Whitby[1], Stephen Goldup[1], Neil Wells[1], Graham J. Langley[1], Julie M. Herniman[1], Matthias G. J. Baud[1]*

[1]School of Chemistry, University of Southampton, Highfield, SO17 1BJ Southampton, United Kingdom

[2]Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, Max-von-Laue-Str. 9, Frankfurt am Main, 60438, Germany

[3]Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences (BMLS), Max-von-Laue-Str. 15, 60438 Frankfurt am Main, Germany

*m.baud@soton.ac.uk

Abstract

Protein arylation has attracted much attention for developing new classes of bioconjugates with improved properties. Here, we have systematically evaluated 2-sulfonylpyrimidines as covalent warheads for the mild, chemoselective and metal free cysteine *S*-arylation. 2-sulfonylpyrimidines react rapidly with cysteine, resulting in stable *S*-heteroarylated adducts at neutral pH. Fine tuning the heterocyclic core and exocyclic leaving group allowed predictable S_NAr reactivity with model tripeptide glutathione *in vitro*, covering 9 orders of magnitude. We achieved extremely fast chemo- and regio-specific arylation of a mutant p53 protein, and confirmed arylation sites by protein X-ray crystallography. Hence, we report the first example of a protein site specifically *S*-arylated with iodo-aromatic motifs. Overall, this study provides the most comprehensive structure-reactivity relationship to date on heteroaryl sulfones, and highlights 2-sulfonylpyrimidine as a synthetically tractable and protein compatible covalent motif for targeting reactive cysteines, expanding the arsenal of tunable warheads for modern covalent ligand discovery.

The last two decades have seen a tremendous expansion of the range of bioconjugation strategies for preparing increasingly complex unnatural biologicals with novel properties beyond those accessible from their canonical building blocks. These strategies rely on mild and biocompatible chemical
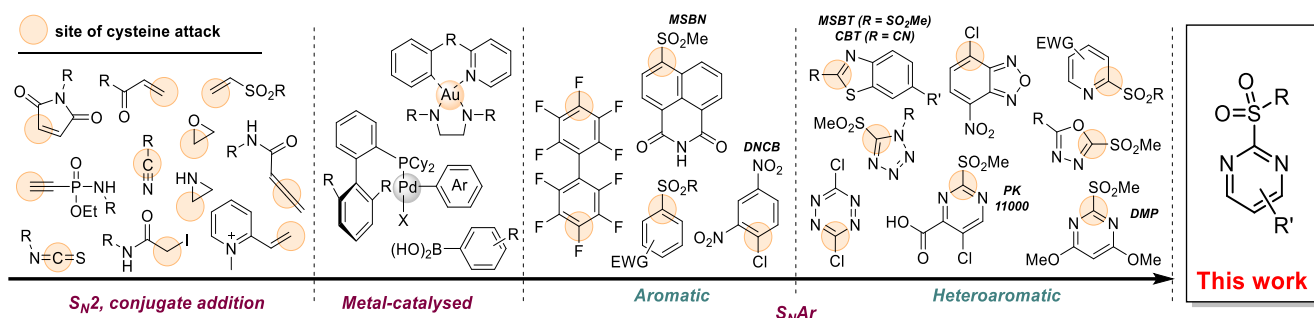
reactions, where a reactive electrophilic "warhead" creates a covalent linkage between nucleophilic sites of the biomolecule and the designed synthetic molecule. Notable examples of such reactions include a range of condensations, ligations, nucleophilic substitution, conjugate additions and substitutions, metal/light/strain promoted "click" cycloadditions, and transition metal catalysed couplings.[1] These warheads have been incorporated in a myriad of chemical labelling agents such as biochemical probes for *in cellulo* / *in vivo* mechanistic studies and characterisation of post-translational modifications (PTMs), tracers for bio-imaging, novel biomaterials, therapeutic macromolecules with enhanced metabolic stability, and small molecule targeted covalent inhibitors (TCIs) to address biomolecular targets reputed intractable.[2–5]

Mild conditions for bioconjugation are paramount to retain protein structure and functionality. Those generally require stringent conditions, including operating in aqueous buffers at pH close to neutral and minimal use of organic co-solvents, at temperature ≤ 37 °C, and with minimal stirring. Critically, such reactions must be quantitative and fast, while remaining highly chemoselective, and proceed at low substrate concentration, usually in the low micromolar range or below.[1] Cysteine bioconjugation has to date received the most attention, and still represents the cornerstone of most modern protein modification strategies. Cysteine is present in nearly all mammalian proteins, but represents only ca. 2 % of the whole proteome,[6] and has a distinctive chemical reactivity due to the superior nucleophilicity of its thiol side chain. These two features are key advantages for the development of chemoselective bioconjugation strategies.

Balancing the reactivity of the warhead is of prime importance[3,7] to allow covalent modification and minimise unspecific reaction at off-target sites at the protein surface, or inactivation through hydrolysis. Many cysteine reactive warheads have been reported, with maleimides, acrylamides, and related conjugated acceptors being the most popular (**Figure 1**). However, while they are being employed extensively, they have certain limitations. Their variable chemoselectivity,[8–10] in addition to linker cleavage *via* retro-Michael[11–14] , thiol exchange,[11,12,15–17] , hydrolysis [16,18] or aminolysis[19] are well-known historical bottlenecks,[8] leading to variable *in vivo* efficacy and toxicity due to the formation of dynamic heterogenous mixtures of conjugates.[20]

Heteroaryl sulfones have recently emerged as excellent reagents for the metal-free arylation of cysteine. The first example of such an agent is 4,6-dimethoxy-2-(methylsulfonyl)pyrimidine (DMP),

which was reported in 2005 as a cysteine "capping" agent for proteomics studies. In 2016, prototypical lead 2-methylsulfonyl pyrimidine PK11000 (**Figure 1**) and its derivatives were reported as stabiliser of a range of thermolabile p53 cancer mutants *in vitro* along with thiol (e.g. GSH) depletion, accumulation of reactive oxygen species (ROS) and selective toxicity in p53 compromised cancer cell lines.[21] Few such « thio-click » reagents based on benzothiazole, tetrazole and oxadiazole scaffolds were reported by Xian,[22] and Barbas[16] and colleagues. These reagents show preferential selectivity for cysteine over other amino-acids, and unlike maleimides, they do not react with other oxidised sulfenic acids (-SOH) and *S*-nitrosothiol (-SNO).[23,24] Importantly, the resulting thioether linked conjugates are markedly more stable than adducts of conjugate acceptors.[12,13] Heteroaryl sulfones also display diverse reaction rates towards cysteine, modulated by the nature and electrophilicity of the heterocyclic system. This was recently underscored by reports from the Bollong[25], Martin[26] and Fang groups[27] showing that functionalisation of 2-methylsulfonylpyridine, methylsulfonylbenzothiazole (MSBT) and naphtalimide (MSBN) scaffolds with electron withdrawing/donating groups (EWGs/EDGs) influence the reaction rates with cysteine (**Figure 1**).

Figure 1. Representative classes of electrophilic warheads used for cysteine modification, extracted from the literature.

The limited structure-reactivity relationship data for heteroarylsulfonyl makes it challenging to design new synthetic reagents displaying optimal stability and reactivity profiles in physiologically relevant conditions. A better understanding of the structure-reactivity relationship of 2-sulfonylpyrimidines (2-SPs) will not only be pivotal to rationalise their bioactivity profile but will also be critical for developing tuneable covalent warheads with suitable electrophilicity, aqueous stability, while maintaining chemoselectivity.

Herein we describe the first systematic structure-reactivity relationship study of 2- SP based reagents, along with straightforward and scalable synthetic routes for their preparation. We show that

2-SPs and their analogues display good aqueous stability and solubility (mM, *vide infra*) compared with more hydrophobic activated benzenes and MSBT derivatives (< 50 µM) requiring organic co-solvents (up to 20% MeCN).[26] Through the systematic UV- and NMR-based determination of *in vitro* reaction rate constants with L-glutathione (GSH), we highlight the exquisite chemoselectivity of 2-SPs and show that reactivity (*k*) can be modulated over 9 orders of magnitude by precise substitution/functionalisation of the heteroaromatic ring and exocyclic group. We provide general design principles for the controlled reactivity modulation of 2-SPs, supported by density functional theory (DFT) calculations. In full protein arylation experiments, we achieved fast and chemoselective cysteine arylation under benign buffered conditions. Finally, we could demonstrate regioselective arylation of the model 25 kDa DNA binding domain of p53 using mass spectrometry and X-ray crystallography, and conservation of protein stability using differential scanning fluorimetry.

Results

Design and synthesis of 2-sulfonylpyrimidine libraries. We assembled a library of 2-SP derivatives, through both synthesis and commercial sources (**Figure 2**). We anticipated that modulating the reactivity could be achieved by i) introducing EWGs (e.g. -CF$_3$, -NO$_2$) and EDGs (e.g. -NH$_2$, -OMe) on the pyrimidine ring (**R**) to respectively accelerate or slow down reaction rates. The generation of isomeric "matched pairs" also allowed determining whether substitution at the 4- or 5-position has the strongest effect; ii) adjusting the sterics and electronics of the leaving group (**R'**); iii) varying the heteroaromatic system (e.g. quinazoline), heteroatoms position (e.g. pyrazine) and number of heteroatoms (e.g. triazine).

Nucleophilic aromatic substitution reactions between commercially available thiols and heteroaryl chlorides **1a-y** afforded the corresponding 2-sulfanyl intermediates **2a-y** and **3a-f** generally in good to excellent isolated yields (**Figure 2A**, **Supplementary Table 1**). Further oxidation with *m*-CPBA or hydrogen peroxide afforded the corresponding 2-sulfonylated products **4a-y** and **5a-f**. 2,4-diazine, 1,3,5-triazine and quinazoline derivatives **18-20** were prepared in a similar manner (Error! Reference source not found.**B**). Oxidation of 2-mercaptopyrimidine **6** to the corresponding sulfonyl chloride and further condensation with benzylamine or pentafluorophenol afforded 2-sulfonamido and 2-sulfonate derivatives **5g-h** in a one-pot sequence, as previously reported (**Figure 2A**).[28,29] Prototypical literature heteroarylsulfones (**Figure 2C**),[16,26] 2-halo/hydroxy/amino pyrimidines (**Figure 2D**) and representative

literature electrophilic warheads commonly used for bioconjugation (**Figure 2E**) were either purchased or re-synthesised for comparison. Detailed synthetic protocols and analytical characterisation data are described in the Supporting Information section. A summary of all compound structures and associated numbering can be found in the Supplementary Information section.
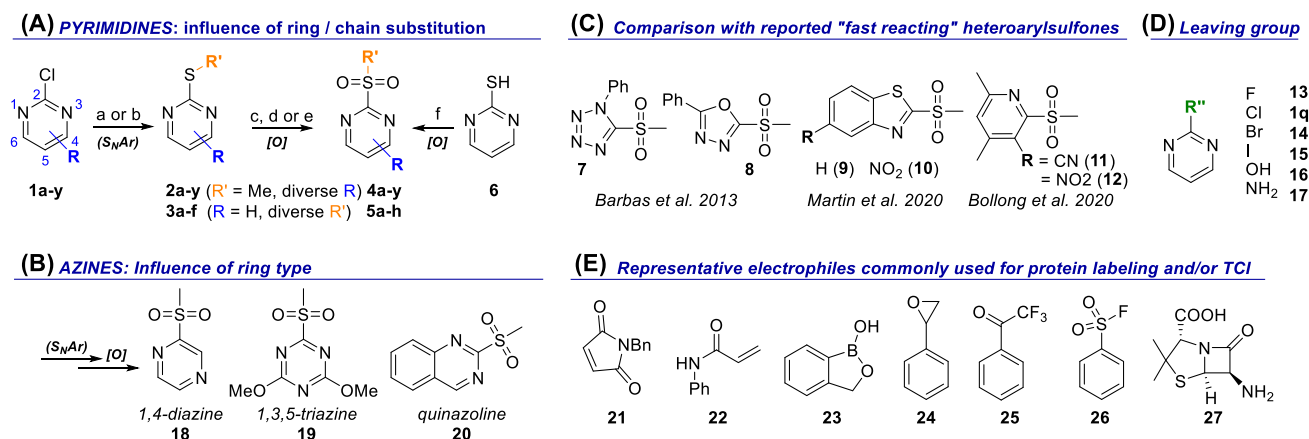
Figure 2. Assembly of electrophilic warhead libraries for reactivity assessment. General structures of R/R'-functionalised 2-SP **(A)** and azine **(B)** reagents, and representative synthetic routes for their preparation: a) R'-SNa, THF, 0 °C to rt, 15-24h; b) R'-SH, K$_2$CO$_3$, THF, 0 °C to rt, 15-24h; c) *m*-CPBA, DCM, rt, 16h-4d; d) 30 % w/w aq. H$_2$O$_2$, AcOH, rt, 16-24h ; e) (NH$_4$)Mo$_7$O$_{24}$.4H$_2$O, 30 % w/w aq. H$_2$O$_2$, EtOH, 0 °C to rt, 24-48h ; f) i) aq. HCl, NaOCl, DCM, -20 to -5 °C, 30 min, ii) BnNH$_2$ or C$_6$F$_5$OH/Et$_3$N, DCM, -20 °C to rt, 2.5h; g) MeI, R$_3$N, THF, 0 °C to rt, 3h; h) Na$_2$WO$_4$, 30 % w.H$_2$O$_2$, EtOH, 0 °C to rt, 48h; i) P(O)Cl$_3$, Δ, 3.5h; **(C-E)** Common literature electrophiles used in protein bioconjugation, for benchmarking against 2-SPs.

***In vitro* S$_N$Ar reaction rate determination and chemoselectivity, assay set up.** We employed nuclear magnetic resonance (NMR) to assess the reactivity and specificity of our 2-SP derivatives and other representative electrophiles (**Figure 2**) against cysteine. NMR allows straightforward determination of reaction rate constants through dual monitoring the consumption of the warhead and formation of the product by integration of their respective NMR signals (**Supplementary Figures 1 and 2**). It simultaneously provides a direct readout on reaction specificity and hydrolytic stability of the warhead. All measurements were carried out in KPi buffer in the presence of 5% d^6-dmso, which is generally well-tolerated by a wide range of proteins in *in vitro* studies. *N*-acetylcysteine methylester (NACME) and L-glutathione (GSH) are useful model cysteine nucleophiles for *in vitro* studies of electrophilic agents.[30] Mixing reference electrophile 2-methylsulfonylpyrimidine (**4q**) with NACME or GSH in a 1:10 ratio (pseudo-first order conditions) allowed extracting accurate and reproducible second order rate constants (*k*). At pH 7.0, quantitative arylation of NACME occurred within minutes (k ≈ 4.5 x 10^{-2} M^{-1}.s^{-1}), whereas GSH reacted approximately three times slower (k ≈ 1.6 x 10^{-2} M^{-1}.s^{-1}) (**Supplementary Figure 3**). This was unambiguously confirmed by rapid and characteristic formation

of methanesulfinic acid (δ ≈ 2.3 ppm) in each case (**Figure 3A**). This was also consistent with shielding of the pyrimidine aromatic protons from 9.1 (H4,6) - 7.9 (H5) ppm in **4q** to ~ 8.6 (H4,6) - 7.3 (H5) ppm in the arylated NACME and GSH products, characteristic of the 2-alkylthioether motif (**Supplementary figure 1A and 1B**). Not unexpectedly, the reaction was approximately 5 times faster at pH 7.0 vs 6.5 with both nucleophiles, overall consistent with a higher effective equilibrium concentration of thiolate anion. Critically, arylation was completely chemoselective and produced *S*-arylated NACME and GSH as the sole products The pK_as of the sulfhydryl groups in NACME and GSH are 8.3 and 9.2, respectively,[31] explaining the greater nucleophilic reactivity of NACME compared to GSH due to higher thiolate anion concentration, translating into higher rate constants.

Structure-reactivity relationship: reactivity of 2-SPs can be effectively modulated beyond 9 orders of magnitude. We selected GSH for our structure-reactivity study as i) the timescale on which arylation takes place is suitable for NMR studies, which is preferable for characterising faster reacting analogues; ii) cysteine is embedded within the GSH tripeptide, which provides a better reflection of the natural steric and electronic constraints of cysteine residues exposed at protein surfaces; iii) the presence of the peptidic backbone and unprotected, free *N*/*C*-termini allows for a primary assessment of 2-SP reagents chemoselectivity towards thiols and potential off-target reactivity.

To establish an accurate structure-reactivity profile of 2-SP reagents, we systematically determined their reaction rate constants for the arylation of GSH by NMR, at pH 7.0 and 6.5. This was a prerequisite to i) quantify the influence of substitution on reaction rates; ii) accurately assess the tunability and chemoselectivity of 2-SPs for cysteine arylation; iii) determine the influence of pH on the reaction rates to further inform on the precise reaction mechanism; and iv) benchmark the overall performance of 2-SPs with those of previously reported fast reacting heteroarylsulfones and other common cysteine reactive warheads (**Figure 2C-D**). NMR allowed us probing a dynamic range of approximately 10^4, with rate constants (*k*) ranging from ~ 5.0 x 10^{-5} $M^{-1}.s^{-1}$ to 0.5 $M^{-1}.s^{-1}$. The chemoselectivity of faster reacting warheads (k > 0.5 $M^{-1}.s^{-1}$, $t_{100\%}$ < 8 minutes) was also assessed by NMR, although their associated rate constants were determined by time-dependant UV-absorbance, reducing delay between mixing and data acquisition (seconds vs minutes). A representative subset of warheads were characterised in both NMR and UV-absorbance assays, and rate constants determined by both methods were generally in good agreement (**Figure 3B**/**Supplementary Table 2**). Rate constants (*k*)

presented were determined in duplicate, both at pH 7.0 and pH 6.5 (**Figure 3B**, **Supplementary Tables 2-6**). In line with our previous observation (*vide supra*, **Supplementary Figure 3**), virtually all 2-SPs reacted faster at pH 7.0, consistent with a higher equilibrium concentration of thiolate anion in solution. In comparison, the corresponding 2-chloro and 2-methylthio pyrimidines were far less reactive to completely unreactive in the same conditions. All rate constants (*k*) are summarised in **Figure 3B** and **Supplementary Tables 2-6**, providing a clear structure-activity relationship:

Substitution at the 5-position had the most important effect on reactivity. As anticipated, strong mesomeric acceptor (-M) EWGs such as $-NO_2$ and -COOMe, and inductive acceptor (-I) groups such as $-CF_3$ drastically increased reaction rate by ca. 3.5 to 6 orders of magnitude compared with the unsubstituted reference warhead (**4q**, $k_{7.0/H} \approx 1.2 \times 10^{-2}$ $M^{-1}.s^{-1}$) at both pHs. In particular, the 5-COOMe derivative **4y** ($k_{7.0/COOMe} \approx 9,900$ $M^{-1}.s^{-1}$) was the most reactive warhead and was > 800,000 times more reactive than **3q**. In contrast, strong +M EDGs such as $-NH_2$ (**4m**) and -OMe (**4n**) completely switched off reactivity. GSH arylation could not be detected ($k < 5.0 \times 10^{-5}$ $M^{-1}.s^{-1}$) even after extended reaction times of 6 hours. Weaker ± I/M representative groups such as -Me, -Ph, -Br, -Cl allowed finer reactivity adjustment within approximately one order of magnitude. Substitution at the 4-position modulated reactivity in a less pronounced but similar manner, with strong -I/M EWGs functionalised derivatives reacting faster. Trifluoromethylated derivative **1l** ($k_{7.0/CF3} \approx 21$ $M^{-1}.s^{-1}$) was the fastest reagent of the 4-series, approximately 1750 times more reactive than **4q**. Modification of the exocyclic leaving group offered additional entry points for controlling reactivity. Fine reactivity moderation could be achieved by both increasing the steric constraint around C2 using larger alkyl chains such as *n*-Bu or *t*-Bu, and reducing C2 electrophilicity by replacing the sulfone by a sulfonamide (**Supplementary Table 3**). Pleasingly, trifluoromethylation or introduction of electron-deficient aromatic systems resulted in up to 1000-fold rate acceleration, while retaining complete specificity. In control experiments, 2-halo, 2-methylthio-, 2-hydroxy and 2-amino pyrimidines (**1q** and **13-17**, **Figure 2D**) all failed to induce observable arylation of GSH in the same conditions over 6 hours, further highlighting the superior reactivity of sulfonyl based leaving groups across the pyrimidine series (**Supplementary Table 4**). Finally, alteration of the aromatic system had a profound effect on reactivity. Replacement of the pyrimidine ring by a 1,4-pyrazine (**18**) completely switched off reactivity, while quinazoline analogue **20** was marginally faster reacting ($k_{7.0/quinaz} \approx 2.8 \times 10^{-2}$ $M^{-1}.s^{-1}$) than reference pyrimidine **4q**. In contrast, replacement of the pyrimidine ring by a 1,3,5-triazine resulted in a drastically increased reactivity. We

could access milligram quantities of purified 2,4-dimethoxy-6-(methylsulfonyl)-1,3,5-triazine **19** following anhydrous flash column chromatography. However, **19** could not be assayed due to its rapid hydrolysis in aqueous buffers or by trace/ambient moisture, even upon storage at 4°C in the solid state. This evidences the fundamental importance of the ring type as a basis for reactivity modulation. Overall, our data highlight yet another opportunity to switch scaffold while maintaining reactivity in a suitable range through a judicious combination of heterocyclic system and leaving group (**Supplementary Table 5**).

In benchmarking experiments, we directly compared a set of diverse, historical Cys reactive warheads with 2-SPs. Strikingly, none of representative acrylamide **22**, boronate **23**, epoxide **24**, electrophilic ketone **25**, sulfonyl fluoride **26** and beta-lactam **27** showed any reactivity in our assay conditions (**Supplementary Table 6**). N-benzylmaleimide **21** ($k_{7.0/NBM}$ > 0.5 $M^{-1}.s^{-1}$) fully reacted within minutes but produced a heterogeneous mixture of succinimidyl products. Among reported arylating agents, lead 2-sulfonylpyridine based electrophiles 4,6-dimethyl-2-(methylsulfonyl)nicotinonitrile **11** and 4,6-dimethyl-2-(methylsulfonyl)-3-nitropyridine **12**, recently reported by Bollong,[25] were ca. 5 times less reactive than **4q**. 2-methysulfonylbenzothiazole **9** (MSBT, $k_{7.0/MSBT}$ ≈ 0.23 $M^{-1}.s^{-1}$) was ca. 20 times more reactive than **4q** while maintaining specificity. Such selectivity was conserved across other faster reacting 2-(methylsulfonyl)-6-nitrobenzo[d]thiazole **10** ($k_{7.0/NO2\text{-}MSBT}$ ≈ 1200 $M^{-1}.s^{-1}$), 1-phenyl 5-methylsulfonyl tetrazole **7** ($k_{7.0/TET}$ ≈ 4.3 $M^{-1}.s^{-1}$) and 2-methylsulfonyl-1,3,4-oxadiazole 5-phenyl **8** ($k_{7.0/oxdiaz}$ ≈ 160 $M^{-1}.s^{-1}$), with complete reaction with GSH within minutes.

In additional control experiments, we further underscored the chemoselectivity of the 2-SP scaffold by mixing 2 mM **4q** in KPi buffer with 10 equivalents of either lysine, tyrosine, proline or serine. We did not observe any reaction at room temperature and pH as high as 8.2, even after up to 6 hours. In all NMR experiments, we added a fixed concentration of 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TMSP) as a water soluble internal standard to assess both the solubility and hydrolytic stability of our reagents. With few exceptions, 2-SPs generally displayed excellent solubility at 2 mM and stability to hydrolysis in reactivity experiment (50 mM KPi, 5% v/v d^6-dmso) (**Supplementary Table 7**), leading to arylated GSH as the sole product in each case (**Figure 3A**). All arylated GSH conjugates remained stable and soluble at room temperature for up to 36h. Of note, we unsurprisingly observed slow time and pH dependant *in situ* hydrolysis of a small number of EWG-functionalised 2-SP derivatives over

extended timescales. 2-SP derivatives substituted with 5-nitro (**4w**) and trifluoromethyl (**4l** and **4x**) underwent partial, slow hydrolysis towards the corresponding unreactive pyrimidin-2-ol by-products in stability assays, which is in buffer in the absence of GSH. This was unambiguously revealed by the generation of methanesulfinate and characteristic shielding ($\Delta\delta$ ~ 1.0 ppm) of the pyrimidine aromatic ^{1}H signals. Nonetheless, in all cases hydrolysis generally occurred to a quantifiable extent (>5%) after several hours (**Supplementary Table 7**) while GSH arylation was complete within seconds to a few minutes at pH 7.0 or below. Pleasingly, 2-SPs were generally soluble at mM concentration in aqueous buffer, contrasting with fast-reacting 2-methylsulfonyl benzothiazoles **9-10** recently reported by the Martin lab, which required up to 20% organic co-solvent (MeCN) in PBS buffer to reach low µM concentrations of solubilised compounds.[26]
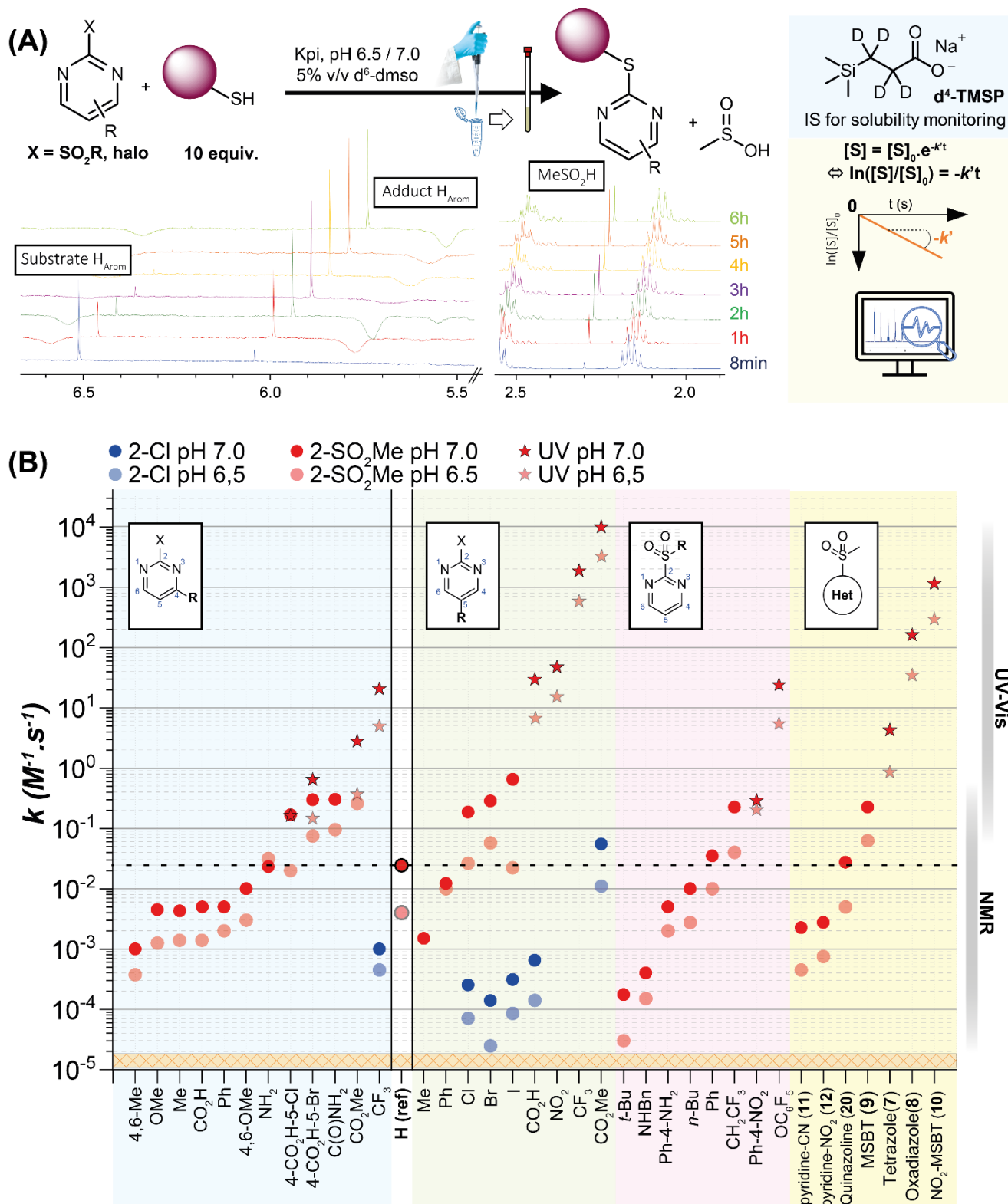
Figure 3. *in vitro* **determination of electrophilic reactivity of compounds. (A)** NMR assay set-up for warhead:GSH reaction rate constant determination and chemoselectivity monitoring, over a 6 hour timescale. Purple bead = GSH. Hydrolytic stability was determined in the same way over 36 hours. d⁴-trimethylsilylpropanoate (TMSP, blue box) was used as an internal standard for monitoring warhead solubility. Second order reaction rate constants (k) were obtained from their pseudo first order counterparts (k'), by time dependant normalised integration of the disappearing warhead signals. NMR stack for the representative reaction of 4,6-dimethoxy-2-methylsulfonylpyrimidine **4f** with GSH shows time dependant signals evolution toward quantitative formation of arylated GSH and generation of methanesulfinic acid (2.3 ppm). **(B)** Experimental second order rate constants (Y-axis, $\log_{10}$ scale) for the reaction of heteroarylsulfone library (red) and their corresponding 2-chloro- (blue) derivatives (X-axis) with GSH and pH 7.0 and 6.5, determined by NMR (circles) and/or UV (stars) titrations at 20°C. 2-methythio- synthetic intermediates were all unreactive in these conditions. Dynamic ranges probed by NMR and UV-vis are shown on the right of the plot. All rate constants were calculated as an average of at least two independent measurements. Numerical values and standard deviations, along with a full list of unreactive warheads are presented in **Supplementary Table 2**. The horizontal dashed line marks

the reaction rate of 2-methylsulfonylpyrimidine at pH 7.0, as reference when comparing with other reagents (see main text).

DFT calculations and mechanistic considerations. We performed density functional theory (DFT) calculations on a selection of heteroaromatic sulfones to i) gain a better understanding of the reaction mechanism of arylation of thiols by 2-SPs and their derivatives, in particular its rate determining step (RDS); and ii) correlate DFT calculated energies of activation for the RDS with experimentally determined kinetics. These are important steps for establishing predictive models for future rational reagent design. Few theoretical studies have discussed the precise mechanism underpinning S_NAr between heteroarylsulfones and thiols, especially in aqueous environments.[32,33] Our calculations indicated that S_NAr between 2-SPs and model methanethiolate generally proceeds in two steps and involves a stabilised Meisenheimer-Jackson complex intermediate, hence reminiscent of the "classical" model (**Figure 4A**). Unsurprisingly, the generally large negative ΔG of the overall transformation largely explains the irreversibility of the arylation. The calculated energies of activation $\Delta G^{\neq 1}_{calc}$ for nucleophilic addition towards the first transition state (TS1) were significantly greater than $\Delta G^{\neq 2}_{calc}$, supporting $\Delta G^{\neq 1}$ and Meisenheimer complex formation as the RDS of the reaction. Calculations also highlighted a significantly lower activation energy for the attack of 2-sulfonylated compounds compared with their 2-halo counterparts (**Supplementary Table 8**), in line with our experimental observation (*vide supra*). Despite a few outliers, calculated differences in activation energies ($\Delta\Delta G^{\neq 1}_{calc}$) of the diverse 2-SPs relative to reference 2-methylsulfonylpyrimidine (**4q**) were generally in good agreement with their experimental counterparts ($\Delta\Delta G^{\neq 1}_{exp}$) (**Figure 4B-D**, **Supplementary Table 8**). They also correlated similarly well with Hammett σ parameters (**Supplementary Table 8**, **Supplementary Figure 4**), consistent with previous reports on related pyridine systems.[25] We advise relying on DFT estimations for these systems in the future. It is more general and allows treating substituents not covered by Hammett parameters, such as leaving groups at the 2-position, and combinations of substituents at positions 4-6 of the pyrimidine ring.
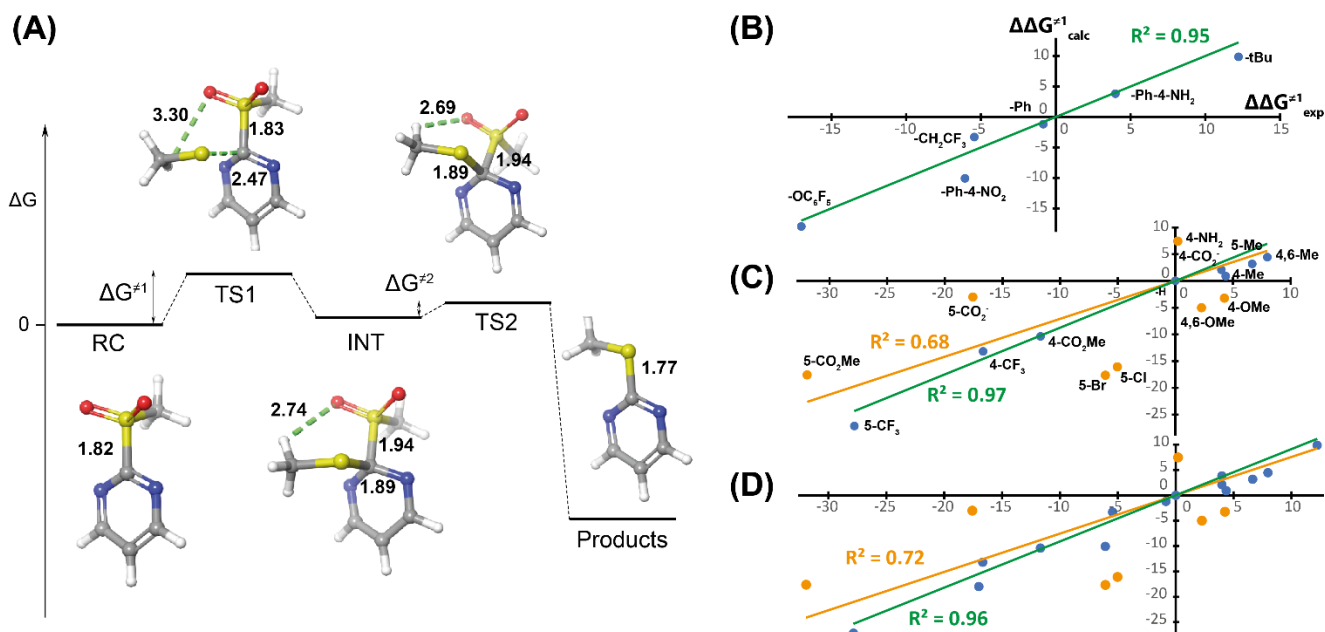
Figure 4. DFT calculations highlight nucleophilic addition as the rate determining step for the S_NAr reaction of 2-sulfonylpyrimidine derivatives with thiol nucleophiles. (A) DFT calculated general Gibbs free energy profile for S_NAr reactions of selected 2-halo- and 2-sulfonylpyrimidine derivatives with model methanethiolate. RC = reactants, TS1= transition state 1, INT = Meisenheimer intermediate, TS2 = transition state 2. Computed transition state structures and Meisenheimer intermediate for the reaction of 2-methylsulfonylpyrimidine with methanethiolate are shown; Correlation of Arrhenius derived experimental $\Delta\Delta G^{\neq 1}_{exp}$ (X-axis, kJ/mol) and DFT calculated $\Delta\Delta G^{\neq 1}_{calc}$ (Y-axis, kJ/mol) for varying leaving group **(B)**, C4/5/6-substitution **(C)**, and all derivatives **(D)**. Level of theory: ωB97XD/6-31+G(d,p) / SMD (water). Outliers are shown in orange. Best fit from linear regression and coefficient of determination (R^2) are shown with inclusion (orange) or exclusion (green) of outliers.
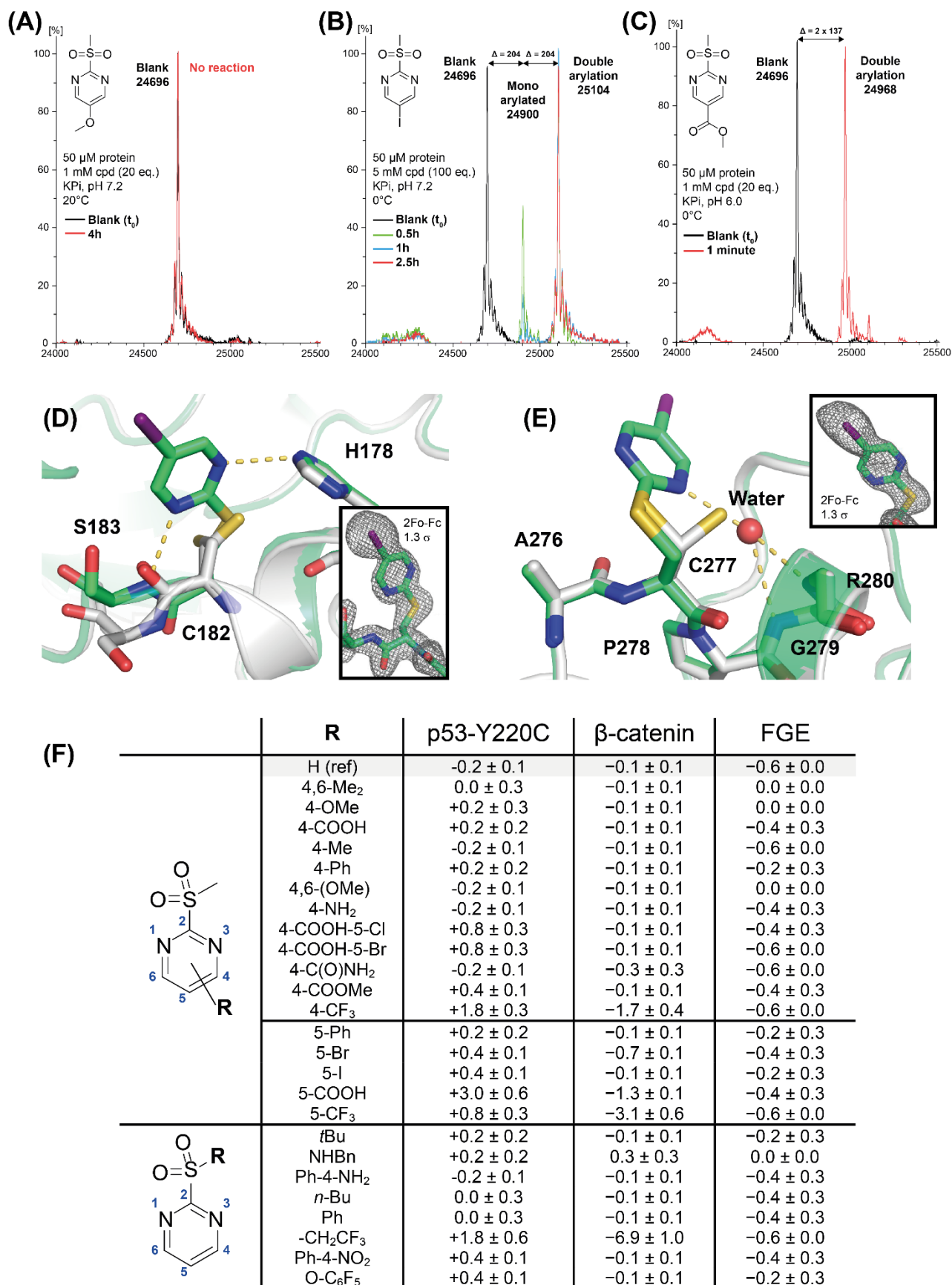
Protein cysteine arylation by 2-SP derivatives. We characterised the covalent reactivity of 2-SPs in protein arylation experiments by electrospray ionisation (ESI) mass spectrometry. The cancer associated mutant p53-Y220C is a particularly well-suited test case.[34] In the unmodified protein, cysteine residues C182/C229/C275/C277 are solvent exposed and freely accessible, while C124/135/141/176/238/242 are sterically hindered and/or involved in structural Zn(II) coordination. The cancer specific C220 lies at the bottom of a mutationally induced hydrophobic pocket at the surface of the p53 DNA-binding domain (DBD, 25kDa) and is also sterically hindered. p53-Y220C also displays relatively low intrinsic stability and is prone to aggregation, hence making it a challenging model system to evaluate the protein compatibility of our reagents. C182 and C277 are known to be intrinsically more reactive than C229 and C275. However, achieving selective modification has proven challenging. For example, Michael acceptor APR-244-MQ, currently examined as a p53 stabiliser for anticancer therapy, reacts with up to nine cysteines *in vitro*, implying partial unfolding of the protein.[35] We incubated 50 μM T-p53-Y220C DBD in KPi buffer at varying pH (6.0 - 8.0), temperature (0°C and 20°C), and equivalents of 2-SPs (20 – 100). Representative 2-SPs were selected to span a broad reactivity

range, to determine whether trends observed in GSH *in vitro* assays translate to relative rates of modification of solvent exposed cysteines in folded proteins. Consistent with their lack of reactivity towards GSH *in vitro*, we could not detect any protein modification by representative 4,6-dimethyl- (**4a**) and 5-methoxy- (**4n**) derivatives in all conditions tested, even after 4 hours at 20°C (**Figure 5A**). Pleasingly, incubation with 100 equivalents of 5-bromo or 5-iodo halogenated derivatives **4t** and **4u** at 0°C resulted in completely selective double arylation of the protein in 2.5 hours at pH 7.2 (**Figure 5B**, **Supplementary Figure 5**). The same could also be achieved with only 20 equivalents of reagent by increasing the temperature to 20°C (data not shown). This is significant because i) late-stage protein functionalisation with bromo- and iodo-(Het)aromatic motifs is challenging, and inaccessible by metal-catalysed arylation methodologies due to dehalogenation and/or off-target specificity; and ii) any off-target specificity and arylation of non-cysteine side chains would be unambiguously identifiable under high concentration soaking conditions during protein crystallography experiments (*vide infra*), due to the unmistakable electronic density around heavy halogens. Finally, fast reacting 5-methylester derivative **4y** reacted at staggering speed, with only 20 equivalents leading to complete and clean double protein arylation in under 30 seconds at pH 6.0 and 0°C (**Figure 5C**). Overall, we observed good correlation between GSH and protein experiments.

Structural characterisation of protein/2-SPs adducts. We determined the structure of the DNA-binding domain of the p53 Y220C after soaking with iodinated warhead **4u**. The high-resolution crystal structure (1.53 Å) showed modification of the solvent-exposed cysteines at position 182 and 277 (**Figure 5D-E**), consistent with their reported reactivity.[21] The modifications could be unambiguously modelled in both chains of the asymmetric unit, with an unmistakable density pattern for the iodoaromatic unit (**Supplementary Figure 6**, **Supplementary Table 9**). The pyrimidine moiety at C182 was stabilised via hydrogen bonds with the backbone amide of S183 and the imidazole group of H178. The iodine moiety protruded into the solvent and did not form specific interaction with the protein. Upon modification of C277 in the loop preceding the *C*-terminal helix, one of the two pyrimidine nitrogen atoms formed a water-mediated hydrogen bond with the backbone amides of G279 and R280. The side chain of both cysteines adopted two alternative conformations in the unmodified structure, whereas upon modification only a single conformation was observed. Conversely and despite being surface-exposed, there was no noticeable positive densities at C229 and C275, showing that subtle differences in microenvironments and reactivity can be exploited for selective targeting. C220 at the

bottom of the mutation-induced surface crevice also was not modified, despite its sulfur atom being accessible, presumably because the narrowness of the pocket prevents a productive geometry for the nucleophilic attack. Further it is interesting to note that none of the cysteines in the cluster of three neighbouring cysteines (C124, C135 and C141) that are known to be chemically reactive[36] was modified, suggesting that these more sterically hindered cysteines require more reactive agents or partial unfolding for modification. Pleasingly, we also did not observe any additional density at the protein surface. Overall, this data is very much in line with the MS data presented in **Figure 5A-C**.

Effect of 2-SPs on protein stability. We also probed the general effect of 2-SPs on protein stability using differential scanning fluorimetry (DSF). We selected the p53-Y220C cancer mutant, the folded armadillo domain (ARD) of β-catenin, and the formylglycine-generating Enzyme (FGE) as representative model proteins due to their structural and functional diversity. Mutants of the 25 kDa DNA-binding domain of the tumour suppressor p53 are notorious for their reduced thermal stability and denaturation/aggregation propensity.[37] The 56 kDa armadillo domain (ARD) of β-catenin mediates canonical Wnt signalling and plays a central role in embryogenesis and tissue homeostasis.[38] FGE (37 kDa) is the only known activator of human sulfatases and its stability is susceptible to modification of its cysteines.[39] We envisaged that the challenge imposed by the structural complexity and diversity, size, and known low intrinsic stabilities would offer a convincing demonstration of the general applicability and tolerability of 2-sulfonylpyrimidine reagents. In DSF experiments, all proteins retained wild-type (WT) like stabilities and melting temperatures (T_m) following incubation with excess reagents. With few exceptions, all compound-treated proteins displayed minor changes in melting temperatures (ΔT_m), usually within ca. 1°C of that of the non-treated proteins (**Figure 5F**).

Figure 5. Biophysical and structural characterisation of protein cysteine arylation by 2-SPs. Top: Representative deconvoluted ESI (ES+) mass spectra of 50 μM p53-Y220C incubated with arylating agents **4n** **(A)**, **4u (B)**, **4y (C)** (green/red) versus without compound (black) in KPi buffer. Stoichiometry, reaction time/temperature and pH are indicated in each case. X-axis: m/z (Da). Y-axis: normalised intensities as percentage of the most intense peak; **Middle:** Structure of the modified Y220C mutant (green) superimposed onto the structure of the unmodified protein (gray, PDB entry 6SHZ)[40] showing the region around modified C182 **(E)** and C277 **(F)**. Hydrogen bonds formed by the pyrimidine are highlighted as dashed yellow lines. 2Fo-Fc electron density maps are shown at a contour level of 1.3 σ for segments of chain B including the modified

residues C182 and C277; **Bottom (F):** Protein thermal stability (ΔT_m, °C) of p53-Y220C, β-catenin ARD and FGE, determined by DSF in the presence of 100 μM 2-SP derivatives.

Discussion

Despite their reversibility and off-target reactivity, Michael acceptors and alkylating agents still form the backbone of modern bioconjugation strategies. Comparatively, protein cysteine arylation has received far less attention. Here, we disclose a library of cysteine chemoselective 2-sulfonylpyrimidines whose reactivity can be finely adjusted over (at least) nine orders of magnitude *in vitro*, providing opportunities to match reactivity to that of specific reactive cysteines. Arylation by 2-SPs is metal-free, operates in extremely mild aqueous conditions and pH close to neutral, and forms highly stable conjugates. In full protein modification experiments, we demonstrate that 2-SPs can discriminate between several cysteine residues at a protein surface, to arylate the most reactive cysteines selectively without compromising protein stability. To the best of our knowledge, prototypical ester substituted (**1y**) is the fastest known cysteine arylating agent to date, surpassing previously reported nitro-MSBT (**10**) by an order of magnitude *in vitro*, and retaining exquisite selectivity.

It is striking that 2-sulfonylpyrimidines and other heteroarylsulfones are often absent from covalent compound libraries for screening, and rarely identified by pan assay interference (PAIN) filters, arguably due to the gap in published knowledge on their reactivity. We anticipate wide-ranging applications of 2-SPs, from the development of improved antibody-drug conjugates for selective drug delivery, to new classes of fine-tuned TCIs for therapeutic applications. The latter in particular holds promise. Michael acceptors are still being employed extensively in covalent drug development despite their limitations. It will be interesting to see how 2-SP warheads perform against e.g. maleimides and acrylamides in terms of inhibitory potency, but also selectivity for individual members from structurally related protein families, such as kinases. The recent advances in radiosynthetic methodologies for ^{18}F-trifluoromethylation of aromatics also presents interesting opportunities for developing new classes of ^{18}F-labeled arylating agents and their application in positron emission tomography (PET).[41,42] The range of synthetically tractable "exit vectors" protruding from the structurally minimalist motifs described in this study, combined with good aqueous solubility and adjustable reaction rates, make 2-SPs well-positioned for general application to next-generation protein bioconjugates.

Online content

Atomic coordinates and structure factors for the reported crystal structure have been deposited in Protein Data Bank (PDB) under accession number 8CG7. Methods, synthetic procedures, compound characterisation data, additional references, protocols for recombinant protein production, protocols for biophysical studies, X-ray crystallographic data, extended data, supplementary information, acknowledgements, details of author contributions, competing interests, and statements of data accessibility are available at https://xxx.

References

1. Lang, K. & Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chem. Biol.* **9**, 16–20 (2014).

2. Zhang, C., Vinogradova, E. V., Spokoyny, A. M., Buchwald, S. L. & Pentelute, B. L. Arylation Chemistry for Bioconjugation. *Angew. Chem. Int. Ed.* **58**, 4810–4839 (2019).

3. Lonsdale, R. *et al.* Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity. *J. Chem. Inf. Model.* **57**, 3124–3137 (2017).

4. Lonsdale, R. & Ward, R. A. Structure-based design of targeted covalent inhibitors. *Chem. Soc. Rev.* **47**, 3816–3830 (2018).

5. Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chem. Int. Ed.* **55**, 13408–13421 (2016).

6. Gupta, V. & Carroll, K. S. Sulfenic acid chemistry, detection and cellular lifetime. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1840**, 847–875 (2014).

7. Reddy, N. C., Kumar, M., Molla, R. & Rai, V. Chemical methods for modification of proteins. *Org. Biomol. Chem.* **18**, 4669–4691 (2020).

8. Ravasco, J. M. J. M., Faustino, H., Trindade, A. & Gois, P. M. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. – Eur. J.* **25**, 43–59 (2019).

9. Hoch, D. G., Abegg, D. & Adibekian, A. Cysteine-reactive probes and their use in chemical proteomics. *Chem. Commun.* **54**, 4501–4512 (2018).

10. Nakamura, T., Kawai, Y., Kitamoto, N., Osawa, T. & Kato, Y. Covalent Modification of Lysine Residues by Allyl Isothiocyanate in Physiological Conditions: Plausible Transformation of Isothiocyanate from Thiol to Amine. *Chem. Res. Toxicol.* **22**, 536–542 (2009).

11. Zhang, Y. *et al.* Thiol Specific and Tracelessly Removable Bioconjugation via Michael Addition to 5-Methylene Pyrrolones. *J. Am. Chem. Soc.* **139**, 6146–6151 (2017).

12. Serafimova, I. M. *et al.* Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat. Chem. Biol.* **8**, 471–476 (2012).

13. Krishnan, S. *et al.* Design of Reversible, Cysteine-Targeted Michael Acceptors Guided by Kinetic and Computational Analysis. *J. Am. Chem. Soc.* **136**, 12624–12630 (2014).

14. Szijj, P. A., Bahou, C. & Chudasama, V. Minireview: Addressing the retro-Michael instability of maleimide bioconjugates. *Drug Discov. Today Technol.* **30**, 27–34 (2018).

15. Baldwin, A. D. & Kiick, K. L. Tunable Degradation of Maleimide–Thiol Adducts in Reducing Environments. *Bioconjug. Chem.* **22**, 1946–1953 (2011).

16. Toda, N., Asano, S. & Barbas, C. F. Rapid, Stable, Chemoselective Labeling of Thiols with Julia-Kocieński-like Reagents: A Serum-Stable Alternative to Maleimide-Based Protein Conjugation. *Angew. Chem.* **125**, 12824–12828 (2013).

17. Ren, H. *et al.* A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chem. Int. Ed.* **48**, 9658–9662 (2009).

18. Matsui, S. & Aida, H. Hydrolysis of some N-alkylmaleimides. *J. Chem. Soc. Perkin Trans. 2* 1277 (1978).

19. Wu, C.-W., Yarbrough, L. R. & Wu, F. Y. H. N-(1-Pyrene)maleimide: a fluorescent crosslinking reagent. *Biochemistry* **15**, 2863–2868 (1976).

20. Alley, S. C. *et al.* Contribution of Linker Stability to the Activities of Anticancer Immunoconjugates. *Bioconjug. Chem.* **19**, 759–765 (2008).

21. Bauer, M. R., Joerger, A. C. & Fersht, A. R. 2-Sulfonylpyrimidines: Mild alkylating agents with anticancer activity toward p53-compromised cells. *Proc. Natl. Acad. Sci.* **113**, E5271–E5280 (2016).

22. Zhang, D., Devarie-Baez, N. O., Li, Q., Lancaster, J. R. & Xian, M. Methylsulfonyl Benzothiazole (MSBT): A Selective Protein Thiol Blocking Reagent. *Org. Lett.* **14**, 3396–3399 (2012).

23. Chen, X. *et al.* Discovery of Heteroaromatic Sulfones As a New Class of Biologically Compatible Thiol-Selective Reagents. *ACS Chem. Biol.* **12**, 2201–2208 (2017).

24. Reisz, J. A., Bechtold, E., King, S. B., Poole, L. B. & Furdui, C. M. Thiol-blocking electrophiles interfere with labeling and detection of protein sulfenic acids. *FEBS J.* **280**, 6150–6161 (2013).

25. Zambaldo, C. *et al.* 2-Sulfonylpyridines as Tunable, Cysteine-Reactive Electrophiles. *J. Am. Chem. Soc.* **142**, 8972–8979 (2020).

26. Motiwala, H. F., Kuo, Y.-H., Stinger, B. L., Palfey, B. A. & Martin, B. R. Tunable Heteroaromatic Sulfones Enhance in-Cell Cysteine Profiling. *J. Am. Chem. Soc.* **142**, 1801–1810 (2020).

27. Zhou, P., Yao, J., Hu, G. & Fang, J. Naphthalimide Scaffold Provides Versatile Platform for Selective Thiol Sensing and Protein Labeling. *ACS Chem. Biol.* **11**, 1098–1105 (2016).

28. Wright, S. W. & Hallstrom, K. N. A Convenient Preparation of Heteroaryl Sulfonamides and Sulfonyl Fluorides from Heteroaryl Thiols. *J. Org. Chem.* **71**, 1080–1084 (2006).

29. Bornholdt, J., Fjære, K. W., Felding, J. & Kristensen, J. L. Heterocyclic pentafluorophenyl sulfonate esters as shelf stable alternatives to sulfonyl chlorides. *Tetrahedron* **65**, 9280–9284 (2009).

30. Martin, J. S., MacKenzie, C. J., Fletcher, D. & Gilbert, I. H. Characterising covalent warhead reactivity. *Bioorg. Med. Chem.* **27**, 2066–2074 (2019).

31. Gustafsson, A., Pettersson, P. L., Grehn, L., Jemth, P. & Mannervik, B. Role of the Glutamyl α-Carboxylate of the Substrate Glutathione in the Catalytic Mechanism of Human Glutathione Transferase A1-1 [†]. *Biochemistry* **40**, 15835–15845 (2001).

32. Rohrbach, S. *et al.* Concerted Nucleophilic Aromatic Substitution Reactions. *Angew. Chem. Int. Ed.* **58**, 16368–16388 (2019).

33. Rohrbach, S., Murphy, J. A. & Tuttle, T. Computational Study on the Boundary Between the Concerted and Stepwise Mechanism of Bimolecular S_NAr Reactions. *J. Am. Chem. Soc.* **142**, 14871–14876 (2020).

34. Stephenson Clarke, J. R. *et al.* Discovery of Nanomolar-Affinity Pharmacological Chaperones Stabilizing the Oncogenic p53 Mutant Y220C. *ACS Pharmacol. Transl. Sci.* **5**, 1169–1180 (2022).

35. Lambert, J. M. R. *et al.* PRIMA-1 Reactivates Mutant p53 by Covalent Binding to the Core Domain. *Cancer Cell* **15**, 376–388 (2009).

36. Chen, S. *et al.* Arsenic Trioxide Rescues Structural p53 Mutations through a Cryptic Allosteric Site. *Cancer Cell* **39**, 225-239.e8 (2021).

37. Wang, G. & Fersht, A. R. Multisite aggregation of p53 and implications for drug rescue. *Proc. Natl. Acad. Sci.* **114**, E2634–E2643 (2017).

38. Klaus, A. & Birchmeier, W. Wnt signalling and its impact on development and cancer. *Nat. Rev. Cancer* **8**, 387–398 (2008).

39. Dierks, T., Schmidt, B. & von Figura, K. Conversion of cysteine to formylglycine: A protein modification in the endoplasmic reticulum. *Proc. Natl. Acad. Sci.* **94**, 11963–11968 (1997).

40. Bauer, M. R. *et al.* Targeting Cavity-Creating p53 Cancer Mutations with Small-Molecule Stabilizers: the Y220X Paradigm. *ACS Chem. Biol.* **15**, 657–668 (2020).

41. Huiban, M. *et al.* A broadly applicable [18F]trifluoromethylation of aryl and heteroaryl iodides for PET imaging. *Nat. Chem.* **5**, 941–944 (2013).

42. Kee, C. W. *et al.* ^{18}F-Trifluoromethanesulfinate Enables Direct C–H ^{18}F-Trifluoromethylation of Native Aromatic Residues in Peptides. *J. Am. Chem. Soc.* **142**, 1180–1185 (2020).