ARTICLE
PUBLISHED ONLINE: 26 OCTOBER | DOI: 10.1038/NCHEMBIO.1666

nature chemical biology

Covalent docking of large libraries for the discovery of chemical probes

Nir London[1,10], Rand M Miller[2,10], Shyam Krishnan[3,4,10], Kenji Uchida[3,4], John J Irwin[1,5,6], Oliv Eidam[1], Lucie Gibold[5–7], Peter Cimermančič[3], Richard Bonnet[7–9], Brian K Shoichet[1,5,6]* & Jack Taunton[3,4]*

Chemical probes that form a covalent bond with a protein target often show enhanced selectivity, potency and utility for biological studies. Despite these advantages, protein-reactive compounds are usually avoided in high-throughput screening campaigns. Here we describe a general method (DOCKovalent) for screening large virtual libraries of electrophilic small molecules. We apply this method prospectively to discover reversible covalent fragments that target distinct protein nucleophiles, including the catalytic serine of AmpC β-lactamase and noncatalytic cysteines in RSK2, MSK1 and JAK3 kinases. We identify submicromolar to low-nanomolar hits with high ligand efficiency, cellular activity and selectivity, including what are to our knowledge the first reported reversible covalent inhibitors of JAK3. Crystal structures of inhibitor complexes with AmpC and RSK2 confirm the docking predictions and guide further optimization. As covalent virtual screening may have broad utility for the rapid discovery of chemical probes, we have made the method freely available through an automated web server (http://covalent.docking.org/).

Small-molecule modulators of protein activity are useful as tools for investigating biology and as leads for drug discovery. Development of genuinely useful probes typically involves iterative rounds of medicinal chemistry to optimize the potency and selectivity of initial active molecules. An effective strategy for enhancing both properties is via covalent bond formation with a nucleophilic residue that is specific to a target of interest and ideally absent from off-targets. Such covalent-acting chemical probes have increasingly been used in proteome-wide target identification[1] and imaging[2] and for finding inhibitors with high specificity among related enzymes and enzyme isoforms[3,4]. Covalent drugs[5] and natural products[6] are also well known. A challenge in developing covalent probes is identifying reactive functional groups ('warheads') that do not make the molecule so reactive as to be promiscuous. Less recognized is the challenge of screening a wide variety of scaffolds for optimal presentation of such reactive functionality.

The most widely used technique for new ligand discovery is high-throughput screening (HTS), and one could potentially screen extant libraries for new small molecules that react covalently. However, protein-reactive compounds are rarely screened[7] and are typically avoided in HTS[8] or flagged as artifacts owing to concerns about promiscuous activity[9]. Whereas this is sensible for drug discovery, it removes potential starting points for covalent chemical probes[10].

Covalent ligands can target either catalytically essential nucleophiles, such as those in serine and cysteine hydrolases, or noncatalytic nucleophiles, usually cysteine, found in small-molecule binding sites on proteins, including proteins without enzymatic activity (for example, GPCRs and nuclear receptors)[11]. Depending on the electrophile and nucleophile, they can bind reversibly or irreversibly. In all cases, specific noncovalent interactions contributed by the scaffold are critical for orienting the electrophile relative to the protein nucleophile, thereby increasing the rate (and stability, in the case of reversible covalent ligands) and selectivity of covalent bond formation. A key unsolved problem in the discovery of covalent probes is how to identify a protein-binding scaffold that optimally orients the electrophile while minimizing the number of compounds that must be synthesized and tested.

In principle, structure-based docking screens[12,13] can address the gap left by HTS and its libraries. Given the structure of a protein target, docking programs computationally screen large compound libraries for molecules predicted to bind favorably within a defined binding site. The technique has been widely used for the discovery of reversible, noncovalent ligands[14,15]. To date, there have been few docking screens for covalent ligands. Key obstacles include combining classical noncovalent scoring with covalent restraints and bond energies and developing compound libraries suited to covalent modification of proteins. For the compound libraries, one would prefer not only commercially available electrophiles but also the ability to design new, readily synthesizable molecules bearing a particular electrophile. Recently, there has been encouraging progress in developing covalent docking methods, but these have been restricted to retrospective recapitulation of covalent complexes[16–18] or to screens of a few hundred compounds[19,20]; we are unaware of prospective, large-scale covalent docking screens to find new inhibitors or of any such screens against targets for which covalent ligands are unprecedented.

Here, we adapt the noncovalent docking program DOCK3.6 to large-scale, covalent virtual screening of electrophilic small molecules, including low-molecular-weight electrophilic fragments (**Fig. 1**). Nine libraries of ligands bearing different electrophiles, amounting to over 650,000 commercially available or synthetically accessible small molecules, are developed for use with the method. We used the method (DOCKovalent) to prospectively screen compound libraries against three targets of therapeutic interest: AmpC

[1]Department of Pharmaceutical Chemistry, University of California–San Francisco, San Francisco, California, USA. [2]Chemistry and Chemical Biology Graduate Program, University of California–San Francisco, San Francisco, California, USA. [3]Department of Cellular and Molecular Pharmacology, University of California–San Francisco, San Francisco, California, USA. [4]Howard Hughes Medical Institute, University of California–San Francisco, San Francisco, California, USA. [5]Faculty of Pharmacy, University of Toronto, Toronto, Canada. [6]Ontario Institute for Cancer Research, University of Toronto, Toronto, Canada. [7]Clermont Université, UMR 1071 INSERM–Université d'Auvergne, Clermont-Ferrand, France. [8]INRA, USC 2018, Clermont-Ferrand, France. [9]Service de Bactériologie, Centre Hospitalier Universitaire, Clermont-Ferrand, France. [10]These authors contributed equally to this work. *e-mail: bshoichet@gmail.com or jack.taunton@ucsf.edu

β-lactamase, p90 ribosomal protein S6 kinase-2 (RSK2) and JAK3 kinase. Multiple potent, reversible covalent inhibitors were found against all three targets. X-ray crystal structures of predicted ligands, and the occasional false negative, illuminated not only the method's ability to prospectively identify ligands and to predict their structures but also its limitations. Several of the new covalent ligands were tested in cell culture experiments that established biological efficacy and target engagement. To ensure that the method may be used by a broad community, it has been made available on an easy-to-use web server (http://covalent.docking.org/).

RESULTS

Overview of the method

We began by constructing large virtual libraries of electrophiles, either commercially available or synthetically accessible in 1 or 2 steps. We created libraries of well-studied electrophiles including α,β-unsaturated carbonyls, aldehydes, boronic acids, α-cyanoacrylamides, alkyl halides, carbamates, α-ketoamides and epoxides (**Supplementary Results**, **Supplementary Fig. 1** and **Supplementary Table 1**); other electrophilic chemotypes, such as vinyl sulfones, sulfonyl fluorides, 2-chloropyridines and cyanopyrimidines, are also imaginable. All stereoisomers, protonation states and conformations of the covalent adduct were pregenerated for each ligand, enabling rapid docking of the library to any target (**Fig. 1**).

For each ligand, DOCKovalent exhaustively samples all poses and ligand conformations with respect to the covalent bond to the target nucleophile, constrained by ideal bond lengths and angles (**Fig. 1** and **Supplementary Fig. 2**). The nucleophile is immobile during the sampling, and a separate screen is run for each likely rotamer of the nucleophile. Each sampled conformation is scored using the physics-based scoring function in DOCK3.6 (ref. 21), which evaluates the ligand's van der Waals and electrostatic interactions and corrects for its desolvation. Using this scoring function, the entire library is ranked from most to least favorable. The top 1–3% of the ranked list is inspected for misdocked ligands, which are common in a large-scale docking screen, and molecules with incorrect ionization states, tautomers or strained conformations are removed. The remaining molecules are prioritized for experimental testing on the basis of their availability or synthetic accessibility, the presence of unprecedented chemotypes and diversity of chemical structure.

Retrospective assessment of covalent docking

We first tested the method's ability to find known covalent ligands in five retrospective screens against four targets and to recapitulate geometries for a previously published benchmark of covalent ligand complexes[16]. DOCKovalent performed well in pose recapitulation (**Supplementary Tables 2** and **3**) and in four of the five retrospective screens (**Supplementary Note 1** and **Supplementary Table 4**). We thus turned to prosecuting prospective screens for new reversible covalent ligands for three enzymes.

New boronic acid inhibitors of AmpC β-lactamase

AmpC β-lactamase is the leading cause of resistance to cephalosporin antibiotics in clinical settings[22], and several new β-lactamase inhibitors are in clinical trials[23]. Boronic acids inhibit AmpC by forming a reversible covalent adduct with its active site nucleophilic serine (Ser64). We first assessed the ability of our covalent docking method to recapitulate known boronic acid complexes with AmpC. In 15 of 23 cases, the ligand pose was accurately recovered to a r.m.s. deviation of <2 Å (**Supplementary Table 5** and **Supplementary Fig. 3**). To our surprise, a relatively simple compound, *m*-aminophenylboronic acid (MAPB), failed this retrospective test. The boronic acid in the published MAPB–AmpC structure (Protein Data Bank (PDB) code 3BLS; **Fig. 2a**) adopted a pose that differed from other boronic acid complexes with the enzyme and from our other docking poses. To investigate this discrepancy, we redetermined the crystal structure of AmpC in complex with MAPB. In this 1.65-Å resolution X-ray structure, we observed unambiguous density for the ligand that corresponds to the docking prediction (0.46-Å r.m.s. deviation; **Fig. 2a**).

Encouraged by these results, we used the method to covalently dock a library of 23,000 commercially available boronic acids against AmpC. Among the top-ranked 4.5% of the library, we sought boronic acids with scaffolds that had not been tested previously against AmpC. Five such compounds ranked between 11 and 646 of the docked library (top 0.04–2%) and were purchased and tested (**1**–**5**; **Fig. 2b**). An additional, lower-ranking compound (**6**) was purchased as a proxy for structurally related, high-ranking predictions that were commercially unavailable. Five of the six compounds inhibited AmpC with K_i values ranging from 40 nM to 3.55 μM (**Table 1**), and three of these exhibited submicromolar potency (ligand efficiency 0.38–0.66; **Table 1**; docking predictions are in **Fig. 2c**,**d** and **Supplementary Fig. 4**). None of the five inhibitors resemble a known boronic acid inhibitor of AmpC (Tanimoto coefficients <0.3 to AmpC boronic acids in ChEMBL, using ECFP4-based fingerprints).

A 1.74-Å crystal structure of **3**, the most potent inhibitor from our initial set of six compounds (K_i = 40 nM), confirmed the docking pose prediction (1.38-Å r.m.s. deviation; **Fig. 2c**). The boronic acid occupies the oxyanion hole formed by the backbone amides of Ala318 and Ser64 and hydrogen bonds with Tyr150. More notably, noncovalent interactions between the scaffold and AmpC were well predicted. The pyrazole N2 accepts hydrogen bonds from Asn152 and Gln120, whereas the phenyl moiety stacks against Tyr221. The only substantial discrepancy between the docking prediction and

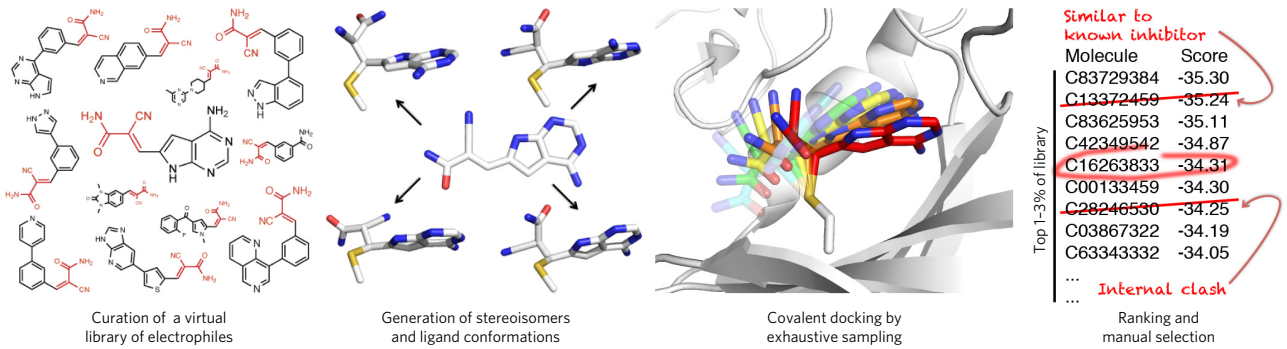

Figure 1 | Overview of the DOCKovalent methodology. A library of commercially available or easily synthesized small molecules containing a specific electrophile was constructed virtually. In this example, the cyanoacrylamide electrophile is shown in red. All stereoisomers, protonation states and conformations of each ligand are pregenerated. Conformational space is exhaustively sampled around the covalent bond for each pregenerated ligand state, and each pose is scored using a physics-based energy function. Each molecule is represented by its best scoring pose, and high-ranking candidates are manually selected for experimental validation.

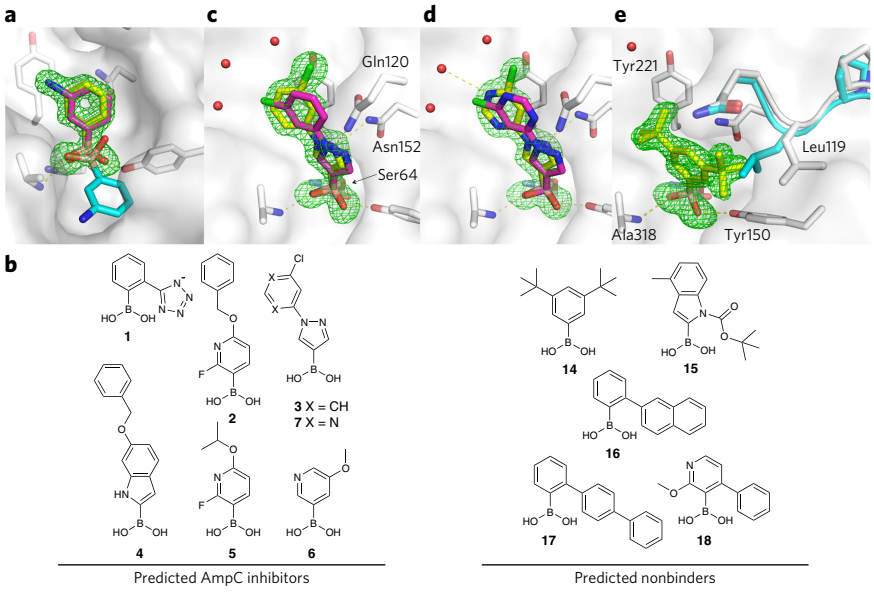

Figure 2 | Boronic acid inhibitors of AmpC predicted by virtual screening. Crystal structures of boronic acids (yellow) covalently attached to AmpC are overlaid with their respective docking predictions (magenta). The omit F_o–F_c electron map is shown in green. (**a**) Crystal structure of MAPB superposed on the docking prediction and the published structure (cyan, PDB code 3BLS) (**b**) Chemical structures of predicted binders **1**–**7** and nonbinders **14**–**18**. (**c**,**d**) X-ray structures of **3** and **7** superposed on their docking predictions. (**e**) Compound **14** induces an unanticipated rotamer change in Leu119 and a rearrangement of loop spanning residues 117–120 relative to the published structure of apo-AmpC (cyan, PDB code 1KE4).

To investigate the origins of this docking false negative, we determined the crystal structure of **14** in complex with AmpC, which revealed unambiguous ligand density in a pose different from the predicted docking model (**Fig. 2e**). To accommodate the observed geometry, an active site loop (L117-Q120) changes conformation, with Leu119 adopting a new rotamer and the loop moving by 0.7 Å (C_α r.m.s. deviation; **Fig. 2e**). This binding mode is incompatible with the AmpC structure used for docking and highlights a caveat of our approach: to enable fast screening of large libraries, we treat the receptor as fixed. The new loop conformation is unique across 23 AmpC structures (**Fig. 2e** and **Supplementary Fig. 7**).

We next tested whether the new boronic acid inhibitors could reverse antibiotic resistance in bacteria that express AmpC. We determined the minimum inhibitory concentration (MIC) of cefotaxime, alone or in combination with inhibitors, against eight clinical isolates resistant to third-generation cephalosporins (**Table 1**). Consistent with the enzymatic assays, **7** was the most potent at reversing antibiotic resistance, lowering the MIC for six strains to ≤2 μg ml^{-1}, the empirical threshold for resistance defined by the Clinical and Laboratory Standards Institute[25]. None of the compounds had substantial antibiotic activity in the absence of cefotaxime (**Supplementary Table 8**).

the crystal structure is the position of the distal chlorine atom. This may reflect the presence of a conserved water network in the active site, which was not included in the calculation (**Fig. 2c–e**).

Several compounds related to pyrazole **3** were also highly ranked by docking. We therefore purchased seven additional pyrazole boronic acids (**7**–**13**; **Supplementary Fig. 5**), one of which showed fourfold greater potency (**7**, K_i = 10 nM; **Fig. 2b**). In a crystal structure we determined, **7** binds AmpC in essentially the same manner as **3** (**Fig. 2c**,**d**). Its increased affinity may arise from a favorable interaction between the new pyrimidine ring and the conserved water network observed in both complexes or from a stronger electrostatic interaction with the carbonyl of Gln120. Ultimately, low-nanomolar inhibitors were obtained by purchasing only 13 compounds.

We characterized the selectivity of the four most potent compounds (**2**, **3**, **5** and **7**) by testing them against three common serine proteases known to bind boronic acids, trypsin, elastase and α-chymotrypsin[24], and against the yeast 20S proteasome. The new AmpC inhibitors typically showed >1,000-fold selectivity versus the serine proteases, and none inhibited the 20S proteasome by >20% at 100 μM (**Supplementary Table 6** and **Supplementary Fig. 6**). An exception was **3**, which inhibited α-chymotrypsin with a K_i of 300 nM. However, pyrimidine **7**, the most potent AmpC inhibitor, showed 10^4-fold selectivity over α-chymotrypsin and >10^5-fold selectivity over trypsin and elastase.

A concern when screening electrophilic compounds is that the electrophile will be so reactive that most compounds in the library will bind the target. To control for this, we tested five boronic acids from the bottom of the ranked docking list (**14**–**18**; **Fig. 2b**). We avoided trivial nonbinders, selecting only those molecules for which the docking program found a nonclashing pose. Four of the five predicted nonbinders showed less than 10% AmpC inhibition at 10 μM, consistent with prediction (**Supplementary Table 7**). Compound **14**, however, did have measurable activity (K_i = 3.2 μM).

New cyanoacrylamide inhibitors for RSK2 and MSK1 kinases

The C-terminal kinase domains (CTDs) of RSK2 and the closely related paralog, mitogen- and stress-activated kinase-1 (MSK1) contain a noncatalytic active-site cysteine shared by only 11 of the 518 human protein kinases. Starting with an established kinase inhibitor scaffold, we previously designed irreversible[3] and reversible[26,27] covalent inhibitors that target this cysteine (Cys436 in RSK2). To achieve reversible covalent inhibition, we exploited the atypical reactivity of cyanoacrylamide Michael acceptors, which react rapidly and reversibly with cysteine thiols at physiological pH. Both RSK2 and MSK1 are attractive therapeutic targets implicated in tumor metastasis[28,29], neurodegeneration[30] and atherosclerosis[31], among other pathological conditions. We sought new RSK2 and MSK1 inhibitors by performing covalent docking screens of thousands of cyanoacrylamide fragments.

As an initial blind test, we used the method to predict the poses of two cyanoacrylamide fragments bound to RSK2, before determining the crystal structures. The predicted binding modes anticipated the experimental structures to 1.93 Å and 1.56 Å r.m.s. deviation (**Fig. 3a**,**b**). Retrospective docking of two larger cyanoacrylamides also recapitulated their crystal structures (0.66 Å and 1.52 Å r.m.s. deviation; **Supplementary Fig. 8a**,**b**). In each prediction, the scaffold portion of the molecule, which forms critical non-covalent interactions with RSK2, closely matched the X-ray structures (0.91–1.36 Å r.m.s. deviation).

Encouraged by these results, we used covalent docking to screen for new cyanoacrylamide inhibitors. Cyanoacrylamide fragments are rare in commercial collections (602 out of 474,770 of the 'fragments in-stock' in ZINC[32]). However, β-substituted cyanoacrylamides can be synthesized in one step by condensing aldehydes with cyanoacetamide (**Supplementary Fig. 9**). We therefore assembled ~12,000 aldehyde fragments from ZINC[32] (molecular weight <250 Da). These were converted *in silico* to generate a virtual library of cyanoacrylamide fragments.

Table 1 | Docking rank, *in vitro* K_i values and MIC values of boronic acids against AmpC

				MIC (μg ml^{-1})[a]							
				AmpC overproducer						ESBL producers[c]	
Compound	Dock rank	K_i (μM)	Ligand efficiency[b]	*Enterobacter cloacae*	*Enterobacter aerogenes*	*Citrobacter freundii*	*E. coli*	*E. coli*	*E. coli*	*E. coli* (TEM-3)	*E. coli* (CTX-M-14)
Cefotaxime alone[d]				64	32	16	16	8	4	8	256
1	11	NA[e]	NA								
2	63	0.17[f]	0.50	8	8	4	8	4	4	4	64
3	95	0.04[f]	0.66	4	8	4	4	4	2	2	32
4	420	2.37[g]	0.38								
5	646	0.48[f]	0.61	8	8	8	8	4	8	4	256
6	5,240[h]	3.55[g]	0.66								
7	280	0.01[f]	0.71	1	4	2	2	1	1	1	32

[a]Compounds were dosed at a cefotaxime/inhibitor ratio of 1:4. [b]Ligand efficiency based on the calculated K_i. [c]ESBLs, extended-spectrum β-lactamase producers. [d]MICs for cefotaxime alone. [e]NA: <10% inhibition at 10 μM. [f]IC_{50} was calculated on the basis of a full dose-response curve (**Supplementary Fig. 16**). [g]IC_{50} was calculated on the basis of a single point measurement. [h]The docking hit list was dominated by larger analogs of this compound that were unavailable for purchase. Compound **6** did not rank well but was purchased as a proxy for structurally related, high-ranking predictions.

We docked this library against Cys436 of RSK2. After manually inspecting the top-ranked compounds for novelty, diversity and accessibility, we pursued eight virtual cyanoacrylamide fragments ranked between 96 and 391 (top 3%; **19**–**26**; **Fig. 3c**). The corresponding aldehydes were purchased and converted to the cyanoacrylamides, which were tested against wild-type RSK2 and the RSK2^{T493M} gatekeeper mutant (**Table 2**). We have previously used this mutant as a biochemical surrogate for MSK1, as MSK1 CTD kinase activity has yet to be reconstituted *in vitro*[27]. Five of the eight high-ranking cyanoacrylamides inhibited RSK2 with half-maximum inhibitory concentration (IC_{50}) values <10 μM in the presence of competing 0.1 mM ATP and 10 mM reduced glutathione. Pyridine **21** and pyrazole **24** (**Fig. 3d**,**e**) were the most potent against wild-type RSK2 and inhibited RSK2^{T493M} with submicromolar potency (IC_{50} 430 nM and 370 nM, respectively).

We determined the cocrystal structure of **24** bound to RSK2^{T493M} at 3.0-Å resolution. Even at this modest resolution, the electron density allowed unambiguous modeling of the phenylpyrazole fragment and the covalent bond to Cys436 (**Supplementary Fig. 10**). The crystallographic structure superposed well with the docking prediction, with an r.m.s. deviation of 0.86 Å over the phenylpyrazole fragment (**Fig. 3d**). The pyrazole forms two hydrogen bonds with the hinge region (N1 to the carbonyl of Glu494 and N2 to the amide NH of Met496). This ring also packs edgewise against the methionine gatekeeper, and the additional van der Waals contacts may explain the enhanced potency for RSK2^{T493M}.

We tested **21** and **24** for activity in mammalian cells stimulated with phorbol myristate acetate (PMA), which activates kinase cascades upstream of RSK2 and MSK1. Both compounds inhibited the activating autophosphorylation of wild-type MSK1 (half-maximum effective concentration (EC_{50}) < 5 μM; **Fig. 3f**). Although these compounds inhibit wild-type RSK2 less potently in cells, quantitation of the normalized pS380 signal reveals dose-dependent inhibition (EC_{50} ~20 μM). A cysteine to valine substitution in the CTD of MSK1 (C458V) conferred complete resistance to both inhibitors, consistent with

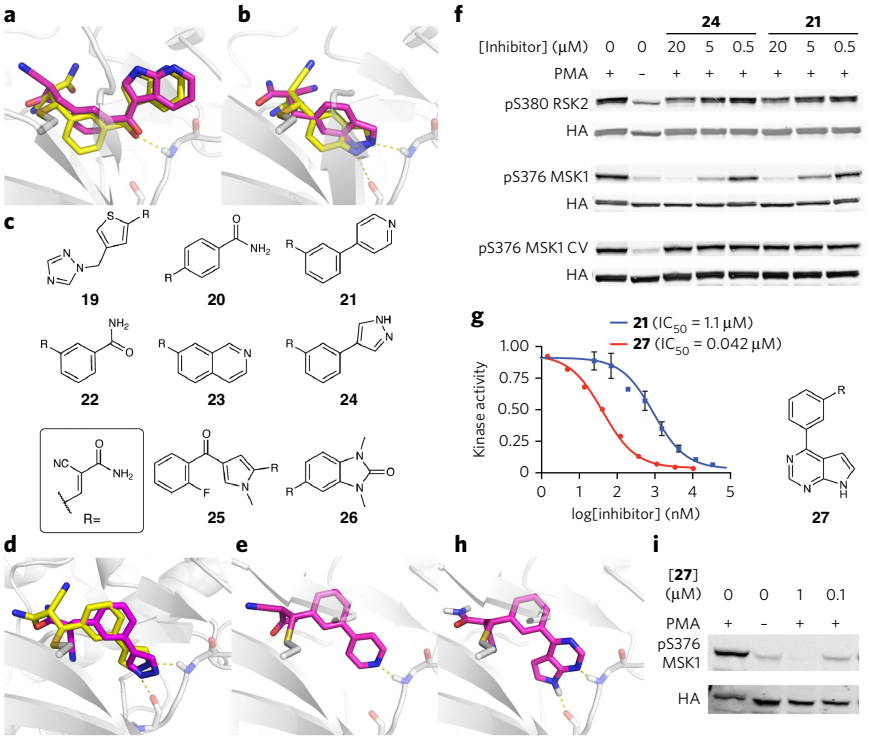

Figure 3 | Cyanoacrylamide inhibitors of RSK2 and MSK1 predicted by covalent docking. (**a**,**b**) Blind docking predictions of two cyanoacrylamide fragments covalently bound to RSK2 (magenta) recapitulate their crystallographic poses (yellow, PDB codes 4JG7 and 4JG6). (**c**) Chemical structures of cyanoacrylamide fragments selected for synthesis and testing. (**d**) Docking prediction for the most potent fragment **24** corresponds well to the experimental structure. (**e**) Docking prediction of the binding mode of compound **21**. (**f**) Compounds **24** and **21** inhibit autophosphorylation of RSK2 and MSK1 in PMA-stimulated cells. Neither compound inhibits the cysteine to valine mutant of MSK1 at concentrations up to 20 μM. Western blots are representative of duplicate biological measurements. HA, hemagglutinin. (**g**) Dose-response curves comparing pyrrolopyrimidine **27** and **21** versus wild-type RSK2. **27** was designed on the basis of the docked structure of **21** (as in panel **e**). Data are plotted as the mean of duplicate measurements ± the range. (**h**) Docked pose of **27**. (**i**) Compound **27** inhibits MSK1 autophosphorylation in PMA-stimulated cells. All western blots are representative of duplicate experiments. Full gel images can be found in **Supplementary Figure 15**.

Table 2 | Docking rank and *in vitro* IC_{50} values for cyanoacrylamides 19–26 against wild-type RSK2 CTD and mutant $RSK2^{T493M}$ CTD

Compound	DOCK rank	IC_{50} (μM)	
		RSK2	$RSK2^{T493M}$
19	66	50.4	27.9
20	96	7	5.2
21	122	1.1	0.43
22	132	3.3	6.8
23	142	12.7	6.4
24	200	1.2	0.37
25	368	>100	>100
26	391	6	7.1

on-target efficacy. Hence, the unoptimized cyanoacrylamide fragments identified by covalent docking inactivate the target kinases in cells without affecting upstream kinases such as RAF, MEK and ERK.

The docking pose of **21** (**Fig. 3e**) suggested the possibility of improving its potency by fusing a pyrrole ring, which acts as a hydrogen-bond donor, to either a pyridine hydrogen-bond acceptor, as in **21**, or a pyrimidine ring. Exemplified by pyrrolopyrimidine **27** (**Fig. 3g**), this would bury more hydrophobic surface area and form an additional hydrogen bond with the hinge region, as suggested by covalent docking to RSK2 (**Fig. 3h**). Compound **27** was not included in the original screen, as the corresponding aldehyde is not commercially available. On synthesis and testing, **27** inhibited wild-type RSK2 kinase *in vitro* with an IC_{50} of 42 nM, over 25-fold better inhibition than by **21** (**Fig. 3g**). Correspondingly, **27** was substantially more potent than **21** in cells, blocking MSK1 autophosphorylation with an EC_{50} <1 μM (**Fig. 3i**).

Selective, reversible covalent inhibitors of JAK3 kinase

Members of the Janus kinase family, comprising JAK1, JAK2, JAK3 and TYK2, are essential for signaling downstream of many cytokine receptors[33]. JAK3 is expressed predominantly in immune cells and is a potential therapeutic target for autoimmune diseases such as rheumatoid arthritis[34]. A pan-JAK inhibitor, tofacitinib[35], was recently approved for rheumatoid arthritis, but it suffers from adverse effects such as elevated liver enzymes and low-density lipoprotein cholesterol[36]. Selective JAK3 inhibitors may avoid such toxicities and, moreover, could help illuminate JAK3-specific roles in cytokine signaling. To date, development of selective JAK3 inhibitors has been hampered by the high sequence identity among JAK-family kinases[37]. JAK3 contains a solvent-exposed cysteine residue just outside the ATP binding site (Cys909) that is not found in JAK1, JAK2 or TYK2 and is present in only nine other human kinases. We used DOCKovalent in an effort to find the first reversible covalent inhibitors of JAK3, which might be expected to have specificity over closely related JAK kinases that lack Cys909.

The vector from Cys909 to the hinge differs greatly from the previously targeted Cys436 of RSK2. A preliminary screen of the virtual cyanoacrylamide fragment library developed initially for RSK2 suggested that greater diversity and perhaps larger fragments would be required to engage both Cys909 and the hinge of JAK3. Inspired by the simple two-step synthesis of **27**, we designed a combinatorial virtual library based on two synthetic transformations: a Suzuki-Miyaura cross-coupling reaction between an aryl or heteroaryl bromide and an aldehyde-containing boronic acid, followed by a Knoevenagel condensation of the aldehyde with cyanoacetamide. We selected 50 commercially available boronic acids and 4,400 aryl bromides, which were converted to their corresponding products *in silico*. This approach afforded a diverse virtual library of 220,000 heterobiaryl cyanoacrylamides (**Supplementary Table 1** and **Supplementary Fig. 11**), which was screened against eight JAK3 crystal structures.

We purchased eight arylbromides and three boronic acids common among the top 0.2% of the docked library (**Supplementary Table 9**). From these building blocks, we synthesized 15 inhibitors (**28–42**; **Fig. 4a**). Compound **42** was prepared from a commercially available aldehyde. Each compound was initially tested against JAK3 at two concentrations. Nine of the fifteen compounds inhibited the kinase by >50% at 5 μM, and five maintained >50% inhibition at 1 μM (**Fig. 4b**).

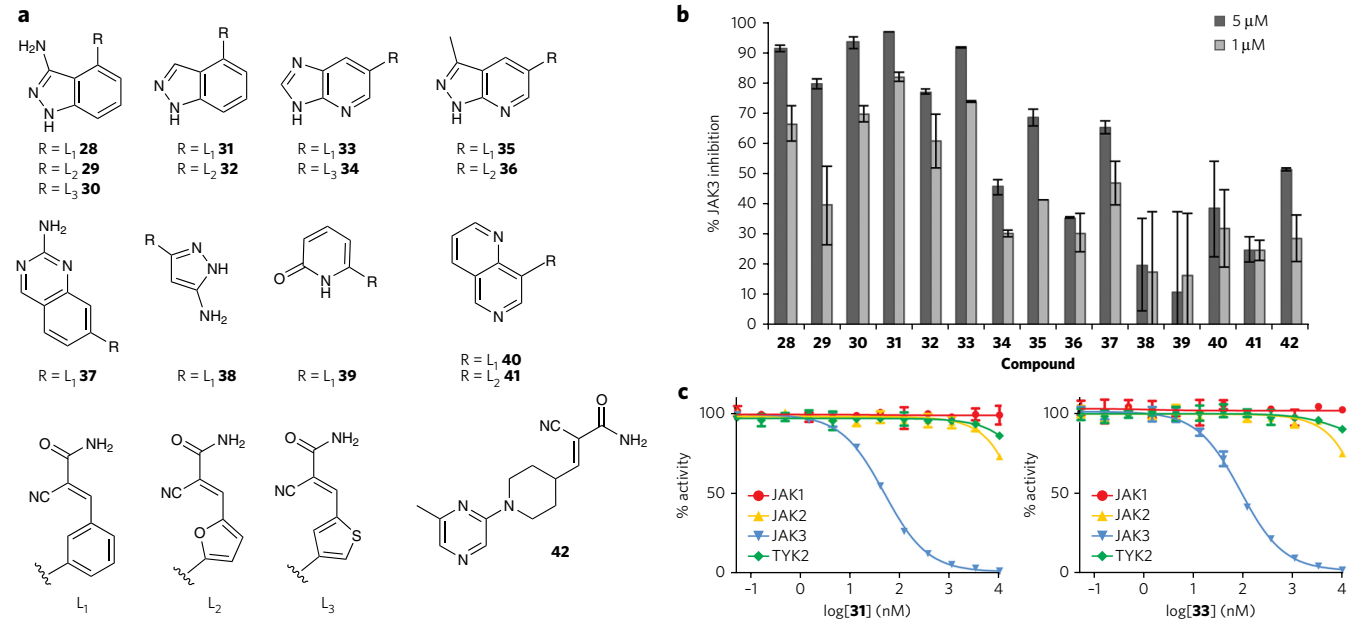

Figure 4 | Reversible covalent JAK3 inhibitors discovered via docking. (**a**) First- and second-generation virtual libraries of cyanoacrylamide fragments were screened by DOCKovalent versus JAK3. Compounds **28–42** were selected and synthesized as described in **Supplementary Note 2**. (**b**) JAK3 inhibition at 1 μM and 5 μM. (**c**) Cyanoacrylamides **31** and **33** are selective for JAK3 over JAK1, JAK2 and TYK2; for JAK3, IC_{50} = 49 nM and 93 nM, respectively. Data represent mean values of two independent experiments ± s.d.

We focused on the two most potent compounds 31 and 33 (for docking poses, see **Supplementary Fig. 12**). To assess their selectivity, we measured full dose-response curves against JAK3 and against the other three JAK-family kinases lacking Cys909. Cyanoacrylamides 31 and 33 inhibited JAK3 with IC_{50} values of 49 nM and 93 nM, respectively, but neither inhibitor affected any other JAK kinase at concentrations up to 10 μM (**Fig. 4c**). This marked selectivity for JAK3 may be largely attributed to covalent bond formation with the target cysteine. By virtue of this covalent targeting of a noncatalytic cysteine, 31 and 33 are among the most selective JAK3 inhibitors reported to date[37]. We further assessed the selectivity of 31 against nine additional human kinases that have a cysteine at the equivalent position to JAK3. 31 potently inhibited three of the nine kinases (IC_{50} = 22 nM, 44 nM and 221 nM for BLK, ERB-B4 and ITK, respectively) but had at least 30-fold specificity for JAK3 over the remaining six kinases (IC_{50} > 1 μM; **Supplementary Fig. 13**).

DISCUSSION

Covalent probes have a crucial role in chemical biology[1–4], yet electrophilic molecules that might serve as initial hits for developing such probes have largely been expunged from empirical screening libraries. Accordingly, we sought to enable large-scale screening for covalent probe molecules via structure-based docking. Although the method has limitations, its utility is supported by the discovery of new chemotypes in three prospective docking screens. For AmpC, the new inhibitors bear little topological resemblance with previously known inhibitors, with ECFP4-based Tanimoto coefficients <0.3 to the most similar previous inhibitor, despite intense study of boronic acids against this enzyme for many years. For RSK2 and MSK1, docking identified the active phenylpyridine 21, even though isomeric phenylpyridines were previously found to be inactive[27]. The predicted pose of 21 supported the design of the unique pyrrolopyrimidine variant 27. Finally, the 4-phenylindazole scaffold of the newly discovered JAK3 inhibitor 31 has little precedent among kinase inhibitors; a related scaffold found in linifanib binds in a completely different orientation to VEGFR2 (PDB code 1YWN).

For all targets, the hit rates were high: five of six molecules predicted and tested for β-lactamase (**Table 1**), five of eight for RSK2 kinase (**Table 2**) and nine of fifteen for JAK3 kinase (**Fig. 4**). These hit rates did not reflect broad promiscuity on the part of the electrophiles. The boronic acids active against β-lactamase were typically inactive against related serine proteases (**Supplementary Table 6**) and against the proteasome (**Supplementary Fig. 6**), and most poorly ranked boronic acids were inactive against the enzyme (**Supplementary Table 7**). The one exception to this was shown by crystallography to reflect an unexpected and previously unobserved conformational change. Similarly, MSK1 inhibitors were selective in cells, being inactive against the C458V substitution of MSK1 as well as the upstream kinases, MEK and ERK, which also have noncatalytic cysteines in their active sites. Finally, the most potent JAK3 inhibitors were virtually inactive against other JAK kinases, underscoring the advantages of covalently targeting poorly conserved noncatalytic cysteines. Moreover, 31 showed selectivity over six of the nine human kinases bearing a homologous cysteine. Given its low molecular weight (288 Da) and simple architecture, which is accessible in two steps from inexpensive building blocks, the selectivity of 31 among these cysteine-containing kinases is remarkably high, comparing favorably with the approved covalent drug, ibrutinib[38], which targets the same cysteine in Bruton's tyrosine kinase (BTK).

Whereas most of the initial hits do not have the potency typically associated with optimized chemical probes, the best examples from the AmpC and the RSK and MSK screens were sufficiently potent and selective to be active in cell-based assays. Hits from all screens were ligand efficient, with ligand efficiency values as high as 0.66. Moreover, the initial hits could be improved, as shown in two cases. Boronic acid 7 has fourfold higher potency relative to the initial docking hit, 3. Indeed, this inhibitor seems to have the highest ligand efficiency of any serine β-lactamase inhibitor[24]. As for RSK2 and MSK1, compound 27 was improved 25-fold versus its parent screening hit, had submicromolar cellular activity and is one of the most ligand-efficient RSK2 inhibitors known. The origins of the selectivity and affinities of these ligands are captured by their docking poses, which superpose well with the subsequently determined X-ray structures. The high fidelity of the docking to the X-ray structures suggests that the covalently docked poses alone may guide ligand optimization in cases where obtaining experimental structures is challenging, as demonstrated by 27.

An innovation introduced in this study is the development and experimental testing of truly virtual, readily synthesizable molecular libraries. The few commercially available cyanoacrylamides necessitated the design of larger and more diverse libraries, which is most likely the case for other classes of reactive molecules as well.

One of our concerns at the outset was that, in mixing covalent and noncovalent scoring terms, the former would overwhelm the latter[39]. To avoid this, we did not explicitly score the covalent bond energy of the adducts but rather imposed distance and angle constraints on bond formation and then scored the docked poses by our standard physics-based scoring function, ignoring the covalent bond. An advantage is that the scoring function is not dominated by the covalent bond to the electrophile and instead favors structural complementarity and specificity provided by noncovalent interactions with the scaffold. The approach is especially well suited to the reversible, and hence thermodynamically driven, inhibitors discovered here (our best compounds were explicitly shown to be rapidly reversible by dilution; **Supplementary Fig. 14**). Still, our docking method can be extended to model high-energy intermediate states of the covalent bond-forming reaction[40], as reflected in the high enrichment in retrospective docking screens against fatty acid amide hydrolase (FAAH) and acetylcholine esterase (**Supplementary Table 4**), where carbamates were modeled as high-energy intermediates (**Supplementary Fig. 1**). The method also may be suited to irreversible Michael acceptors, such as acrylamide-based inhibitors of the EGFR kinase, where retrospective enrichment was also high (**Supplementary Table 4**).

The failure to explicitly model the energy of covalent bond formation is one of several methodological gaps that merit mention. Without considering the covalent bond energy, we cannot directly compare molecules bearing different electrophiles. Thus, each library of different electrophiles is docked and ranked separately. Incorporating quantum-level approaches to predict the reactivity of specific scaffold-electrophile combinations may be a direction forward[41]. Even within the current constraint-based approach, sampling can be improved by relaxing around ligand dihedral angles and perhaps by accounting for bonded energies. Lastly, as demonstrated by the complex of 14 with AmpC, allowing for receptor flexibility can be essential for accurate modeling of receptor-ligand interactions. This remains an area of active research[12,42].

These caveats should not obscure the potential benefits offered by the method. We hope it will begin to address what has been a substantial gap in our ability to screen widely for useful covalent chemical probes, and we have made it and its associated libraries freely available through a general-use, public access web portal (http://covalent.docking.org/).

Received 25 May 2014; accepted 3 September 2014; published online 26 October 2014; corrected after print 22 December 2014

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. PDB. All crystal structures reported here were deposited under accession codes 4LV0, 4LV1, 4LV2, 4LV3 and 4M8T.

References

1. Weerapana, E., Simon, G.M. & Cravatt, B.F. Disparate proteome reactivity profiles of carbon electrophiles. *Nat. Chem. Biol.* **4**, 405–407 (2008).
2. Blum, G., von Degenfeld, G., Merchant, M.J., Blau, H.M. & Bogyo, M. Noninvasive optical imaging of cysteine protease activity using fluorescently quenched activity-based probes. *Nat. Chem. Biol.* **3**, 668–677 (2007).
3. Cohen, M.S., Zhang, C., Shokat, K.M. & Taunton, J. Structural bioinformatics-based design of selective, irreversible kinase inhibitors. *Science* **308**, 1318–1321 (2005).
4. Chang, J.W., Nomura, D.K. & Cravatt, B.F. A potent and selective inhibitor of KIAA1363/AADACL1 that impairs prostate cancer pathogenesis. *Chem. Biol.* **18**, 476–484 (2011).
5. Robertson, J.G. Mechanistic basis of enzyme-targeted drugs. *Biochemistry* **44**, 5561–5571 (2005).
6. Drahl, C., Cravatt, B.F. & Sorensen, E.J. Protein-reactive natural products. *Angew. Chem. Int. Edn Engl.* **44**, 5788–5809 (2005).
7. Kathman, S.G., Xu, Z. & Statsyuk, A.V. A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases. *J. Med. Chem.* **57**, 4969–4974 (2014).
8. Sirois, S., Hatzakis, G., Wei, D., Du, Q. & Chou, K.C. Assessment of chemical libraries for their druggability. *Comput. Biol. Chem.* **29**, 55–67 (2005).
9. Baell, J.B. & Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).
10. Potashman, M.H. & Duggan, M.E. Covalent modifiers: an orthogonal approach to drug design. *J. Med. Chem.* **52**, 1231–1246 (2009).
11. Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).
12. Totrov, M. & Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **18**, 178–184 (2008).
13. Shoichet, B.K. Virtual screening of chemical libraries. *Nature* **432**, 862–865 (2004).
14. Kitchen, D.B., Decornez, H., Furr, J.R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
15. Glick, M. & Jacoby, E. The role of computational methods in the identification of bioactive compounds. *Curr. Opin. Chem. Biol.* **15**, 540–546 (2011).
16. Ouyang, X. *et al.* CovalentDock: automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *J. Comput. Chem.* **34**, 326–336 (2013).
17. Toledo Warshaviak, D., Golan, G., Borrelli, K.W., Zhu, K. & Kalid, O. Structure-based virtual screening approach for discovery of covalently bound ligands. *J. Chem. Inf. Model.* **54**, 1941–1950 (2014).
18. Zhu, K. *et al.* Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* **54**, 1932–1940 (2014).
19. Schröder, J. *et al.* Docking-based virtual screening of covalently binding ligands: an orthogonal lead discovery approach. *J. Med. Chem.* **56**, 1478–1490 (2013).
20. De Cesco, S. *et al.* Virtual screening and computational optimization for the discovery of covalent prolyl oligopeptidase inhibitors with activity in human cells. *J. Med. Chem.* **55**, 6306–6315 (2012).
21. Mysinger, M.M. & Shoichet, B.K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **50**, 1561–1573 (2010).
22. Jacoby, G.A. AmpC β-lactamases. *Clin. Microbiol. Rev.* doi:10.1128/CMR.00036-08 (2009).
23. Livermore, D.M. & Mushtaq, S. Activity of biapenem (RPX2003) combined with the boronate β-lactamase inhibitor RPX7009 against carbapenem-resistant *Enterobacteriaceae*. *J. Antimicrob. Chemother.* **68**, 1825–1831 (2013).
24. Eidam, O. *et al.* Fragment-guided design of subnanomolar β-lactamase inhibitors active *in vivo*. *Proc. Natl. Acad. Sci. USA* **109**, 17448–17453 (2012).
25. Wikler, M.A. *Performance Standards for Antimicrobial Susceptibility Testing: Twentieth Informational Supplement* (Clinical and Laboratory Standards Institute, 2010).
26. Serafimova, I.M. *et al.* Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat. Chem. Biol.* **8**, 471–476 (2012).
27. Miller, R.M., Paavilainen, V.O., Krishnan, S., Serafimova, I.M. & Taunton, J. Electrophilic fragment-based design of reversible covalent kinase inhibitors. *J. Am. Chem. Soc.* **135**, 5298–5301 (2013).
28. Doehn, U. *et al.* RSK is a principal effector of the RAS-ERK pathway for eliciting a coordinate promotile/invasive gene program and phenotype in epithelial cells. *Mol. Cell* **35**, 511–522 (2009).
29. Kang, S. *et al.* p90 ribosomal S6 kinase 2 promotes invasion and metastasis of human head and neck squamous cell carcinoma cells. *J. Clin. Invest.* **120**, 1165–1177 (2010).
30. Park, J. *et al.* RAS-MAPK-MSK1 pathway modulates ataxin 1 protein levels and toxicity in SCA1. *Nature* **498**, 325–331 (2013).
31. Le, N.T. *et al.* A crucial role for p90RSK-mediated reduction of ERK5 transcriptional activity in endothelial dysfunction and atherosclerosis. *Circulation* **127**, 486–499 (2013).
32. Irwin, J.J., Sterling, T., Mysinger, M.M., Bolstad, E.S. & Coleman, R.G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **52**, 1757–1768 (2012).
33. Yamaoka, K. *et al.* The Janus kinases (Jaks). *Genome Biol.* **5**, 253 (2004).
34. Kremer, J.M. *et al.* A phase IIb dose-ranging study of the oral JAK inhibitor tofacitinib (CP-690,550) versus placebo in combination with background methotrexate in patients with active rheumatoid arthritis and an inadequate response to methotrexate alone. *Arthritis Rheum.* **64**, 970–981 (2012).
35. Jiang, J.K. *et al.* Examining the chirality, conformation and selective kinase inhibition of 3-((3R,4R)-4-methyl-3-(methyl(7H-pyrrolo[2,3-d]pyrimidin-4-yl)amino)piperidin-1-y l)-3-oxopropanenitrile (CP-690,550). *J. Med. Chem.* **51**, 8012–8018 (2008).
36. Fleischmann, R. *et al.* Placebo-controlled trial of tofacitinib monotherapy in rheumatoid arthritis. *N. Engl. J. Med.* **367**, 495–507 (2012).
37. Clark, J.D., Flanagan, M.E. & Telliez, J.B. Discovery and development of Janus kinase (JAK) inhibitors for inflammatory diseases. *J. Med. Chem.* **57**, 5023–5038 (2014).
38. Honigberg, L.A. *et al.* The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc. Natl. Acad. Sci. USA* **107**, 13075–13080 (2010).
39. Smith, A.J., Zhang, X., Leach, A.G. & Houk, K.N. Beyond picomolar affinities: quantitative aspects of noncovalent and covalent binding of drugs to proteins. *J. Med. Chem.* **52**, 225–233 (2009).
40. Hermann, J.C. *et al.* Structure-based activity prediction for an enzyme of unknown function. *Nature* **448**, 775–779 (2007).
41. Schwöbel, J.A. *et al.* Prediction of michael-type acceptor reactivity toward glutathione. *Chem. Res. Toxicol.* **23**, 1576–1585 (2010).
42. Fischer, M., Coleman, R.G., Fraser, J.S. & Shoichet, B.K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* **6**, 575–583 (2014).

Acknowledgments

Computational methods supported by US National Institutes of Health (NIH) grant GM59957, and the web portal was supported by NIH GM71896. This work was also supported by the Ministère de la Recherche et de la Technologie, the Institut national de la santé et de la recherche médicale (UMR Inserm U1071), the Institut National de la Recherche Agronomique (USC-2018) and the Centre Hospitalier Régional Universitaire de Clermont-Ferrand, France (to R.B.). We thank M. Fischer and D. Shaya for help with X-ray data collection, A. O'Donoghue (UCSF) for protease substrates, P. Coffino and S. Menant (UCSF) for the proteasome and the PA26 complex sample, X. Ouyang (Nanyang Technological University) for the r.m.s. calculation software used for the β-lactam benchmark, and S. Barelier for reading of this manuscript. N.L. was supported by an EMBO long-term fellowship (ALTF 1121-2011) and the University of California–San Francisco Program for Breakthrough Biomedical Research, which is funded in part by the Sandler Foundation. S.K. was supported by a fellowship from the California Tobacco-Related Disease Research Program (no. 19FT-0091). P.C. was supported by Howard Hughes Medical Institute Predoctoral Fellowship.

Author contributions

B.K.S. and J.T. directed the project. N.L. designed the algorithms, wrote the covalent docking code and executed the docking simulations. N.L. performed β-lactamase assays and crystallography with help from O.E. J.J.I. designed and implemented the DOCKovalent web tool. P.C. performed the proteasome experiments. L.G. and R.B. performed bacterial cell culture experiments. R.M.M. executed synthetic chemistry, kinase assays, crystallography and cell-based assays for RSK2 and MSK1. S.K. and K.U. performed synthetic chemistry for JAK3, S.K. performed JAK1 and JAK3 kinase assays. N.L., R.M.M., B.K.S. and J.T. wrote the paper. All authors contributed to the manuscript in its final form.

Competing financial interests

The authors declare competing financial interests: details accompany the online version of the paper.

Additional information

Supplementary information and chemical compound information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to B.K.S. (for correspondence relating to the docking method and to covalent inhibition of β-lactamase) or J.T. (for correspondence relating to covalent inhibition of kinases).

ONLINE METHODS

Ligand generation. Ligand flexibility was sampled by generating ligand conformations before docking. Given a SMILES string of a ligand with a specific electrophile, we used the OEChem library (OEChem TK 1.7.4; Openeye Scientific Software: Santa Fe, NM. http://www.eyesopen.com/) to convert the ligand to its final, reacted form (**Supplementary Fig. 1**). The receptor's nucleophilic atom involved in the covalent bond is represented by a dummy atom (silicon, for technical reasons). Following the generation of the 'reacted' electrophile, the ligand's 3D structures and stereoisomers are built by Corina[43] (Molecular Networks, Erlangen, Germany), and then protonated and tautomerized by EPIK (Schrodinger software, Catsville, NY). Partial atomic charges and solvation energies are calculated for each of these structures with AMSOL[44]. The electrophile serves as a starting rigid body fragment, and conformations are generated using Omega[45] (Omega parameters: EnergyWindow = 30.0; MaxConfs = 10,000; RMSThreshold = 0.5). The collection of pregenerated ligand conformations in the reacted state is saved to a DOCK-readable flexibase format file.

Sampling. Sampling of ligand poses within the protein binding site was restricted to exhaustive ligand placement with respect to the covalent bond (**Supplementary Fig. 2**). The covalent attachment point was sampled in steps of 20° around the terminal dihedral of the nucleophilic side chain. On the basis of the electrophile geometry determined during ligand generation and user-provided parameters, the vectors of the covalent bond from the ligand and receptor sides are aligned, and the ligand is rotated around this vector in 20° steps. For each placement, all of the pregenerated ligand conformations were scored, and the score for the best pose was saved. This process is repeated for different values of the covalent bond length and angles, centered on ideal values (**Supplementary Fig. 2**). The magnitude of deviation from the ideal values, as well as the step sizes, are user specified.

Scoring. Scoring is performed as previously described, using precalculated van der Waals, electrostatic and ligand solvent-excluded desolvation grids, correcting for ligand desolvation[21]. Receptor structures were prepared using an automated procedure as described in ref. 46 using DELPHI[47] for electrostatics. The ligand's electrophilic atom participating in the bond is omitted from the overall ligand score.

Availability. As noted, the method is accessible through a public web server (http://covalent.docking.org/) and for download as part of the next DOCK3.x release (http://dock.compbio.ucsf.edu/).

Virtual ligand libraries. For the curation of ligand libraries, the electrophiles were represented by SMARTS regular expressions (**Supplementary Table 1**). The Full, lead-like, and fragment-like subsets of the ZINC database of commercially available molecules were filtered using these patterns to identify electrophile-bearing molecules[32]. Known inhibitors for the five retrospective virtual screens were collected from ChEMBL14 (ref. 48) for AChE, FAAH (carbamates with <1 μM activity) and NS3 (α-ketoamides with <100 nM activity). For EGFR ligands, we used the same ligand set as in ref. 16, and FAAH boronic acid inhibitors (K_i < 10 nM) are from ref. 49. The best receptor structure was selected on the basis of its ability to enrich known ligands.

Known AmpC binders were defined as having <1 μM activity in ChEMBL16 (ref. 48). Compound similarity was calculated using ECFP4-based Tanimoto coefficients[50], as implemented in Pipeline-Pilot version 6.1 (SciTegic Inc., San Diego, CA). Boron is not parameterized in several of our ligand generation programs, and to overcome this in the construction of our boronic acids library, we searched ZINC's commercial catalogs for boronic acids, replacing the boron with a carbon during ligand generation. This does not affect ranking as the energy of the ligand's bonding atom is omitted from the docking score.

The one-step cyanoacrylamide-based library is based on filtering ZINC's fragment-like subset for fragments containing an aldehyde. The two-step, Suzuki-Miyaura–based cyanoacrylamide library is made according to (**Supplementary Fig. 11**) by combining any of 50 commercially available aldehyde-containing boronic acids with 4,397 aryl bromide fragments from ZINC's fragments-in-stock subset, containing one of the following SMARTS motifs: 'ncN', '[NX3][CX3](=[OX1])[#6]' or a pyridine, pyrimidine, pyrrole, pyrazine, pyrazole or triazole.

Covalent virtual screening. For the AmpC retrospective pose recapitulation benchmark (**Supplementary Table 5**) and prospective screen, the structure with PDB code 4E3N was used as the receptor. The flexible side chain of Gln120 was truncated at Cβ. The backbone of residues Ala318 and Ser64 and side chains of Asn152 and Tyr150 were polarized to emphasize the electrostatic interactions with the boronic acid. Covalent bond sampling parameters were set to d = 1.5 ± 0.1 Å, a = 116.0° ± 5° and b = 109.5° ± 5°. (Step size of 0.05 Å and 1°; **Supplementary Fig. 2**). **Compounds 1**, **3**, **4** and **14–18** were selected for testing on the basis of this screen. Following the determination of the crystal structure of AmpC in complex with **3**, a second screen was performed using the new structure (PDB code 4LV1; in this screen, the bond length was fixed to d = 1.6Å, and Gln120 was not truncated). Compounds **2**, **5** and **6** were selected for testing. Run time for the entire screen was 227 CPU hours (ran in parallel on an 800-CPU computer cluster; Wall time < 1 h). The RSK2 screen used PDB code 4D9T for the receptor structure. Given that nearly all kinase inhibitors form one or more hydrogen bonds with the hinge region, the backbone amides of the kinase hinge residues (Glu494 and Met496) were polarized to emphasize hinge-binding hydrogen bonds, which is a technique we have long used. Covalent bond sampling parameters were set to d = 2.0Å, a = 109.5° ± 10° and b = 109.5° ± 10° (step size of 2.5°; **Supplementary Fig. 2**). This bond length was chosen on the basis of two available RSK2 cyanoacrylamide complexes (PDB codes 4D9U and 4D9T). Though longer than a typical thioether by about 0.2 Å, this bond length helps ensure that the scores are dominated by the noncovalent terms, minimizing hard van der Waals repulsion in the region of the new bond. We have not attempted to optimize this term. Calculation time for the docking screen was 103 CPU hours, and the elapsed wall time was less than 1 h, owing to the use of a cluster. JAK3 docking was performed using eight available crystal structures with PDB codes 1YVJ, 3LXK, 3LXL, 3PJC, 4HVD, 4HVG, 4HVH and 4HVI. For each structure, we used both the native cysteine rotamer as well as an alternative rotamer (χ1 = –60°). Docking to the native rotamer produced few plausible results, most likely owing to steric clashes of the electrophile with nearby residues, and only one candidate was chosen on the basis of this rotamer (**Supplementary Table 9**). The backbone amides of the hinge residues (Glu903 and Leu905) were polarized similarly to RSK2. Covalent bond sampling parameters were set to d = 1.8 Å, a = 109.5° ± 10° and b = 109.5° ± 10° (step size of 2.5°; **Supplementary Fig. 2**). Calculation time for the entire docking screen was 11,740 CPU hours, with elapsed wall time ~14 h.

Selection criteria. As noted, following docking, the top 500 (kinases) to 1,000 (AmpC) molecules were manually inspected for exclusion criteria on the basis of considerations that are orthogonal to the docking scoring function such as structural diversity, commercial availability and/or synthetic accessibility, perceived ability to derivatize and improve hits, correct representation of the molecule and internal strain (ligand internal energy is not part of the scoring function). We remove redundant, highly similar molecules and compounds with potentially reactive or unstable functional groups. Additionally, for AmpC, we mostly selected poses in which the boronic acid was predicted to occupy the oxyanion hole. For the kinase inhibitors, we only selected poses predicted to form at least one hydrogen bond with the hinge. For JAK3 inhibitors, we usually restricted the $C\beta_{cysteine}$-$S\gamma_{cysteine}$-$C\beta_{acrylamide}$-$C\alpha_{acrylamide}$ dihedral angle to be >90° to avoid internal strain.

R.m.s. deviation calculations. R.m.s. deviation values were calculated using the Hungarian matching algorithm as implemented in DOCK6 (ref. 51). For comparison purposes, r.m.s. deviation values for the β-lactam benchmark were calculated using software generously provided by X. Ouyang (Nanyang Technological University)[16]. For AmpC, receptors were superimposed on the basis of chain A before calculations. For RSK2, receptors were superimposed on the basis of residues Cys436, Met496 and Cys560.

Enzymology. AmpC enzyme inhibition was measured from initial rates using curve fitting in the native Agilent software. Compounds were initially dissolved in DMSO at 100 mM and individually diluted from such stocks. AmpC activity on CENTA (K_m = 15 μM) was monitored by the change of absorbance at 405 nm[24]. AmpC was expressed and purified as described[24], and CENTA was purchased from Tydock Pharma (Modena, Italy). IC_{50} values were obtained by fitting percent inhibition to a sigmoidal dose-response equation using GraphPad Prism (GraphPad Software Inc.; **Supplementary Fig. 16**). K_i values (average of two biological experiments) were determined using the

Cheng-Prusoff equation assuming competitive inhibition. Reactions were performed at room temperature in 50 mM sodium-cacodylate, pH 6.5, in the presence of 0.01% Triton X-100 in 1-ml cuvettes with 50–100 μM CENTA and initiated by addition of AmpC.

Reversibility. Compounds **3** (1 μM) and **7** (276 nM) were assayed for AmpC (1 nM) inhibition with no incubation as described above or after 5-min incubation with 10 nM AmpC. In the latter, the reaction was started by 10× dilution of AmpC, and the incubated inhibitor into a reaction buffer with substrate. Reversibility experiments were performed in 50 mM potassium phosphate buffer, pH 7.0.

Selectivity. Compounds **2**, **3**, **5** and **7** were tested against porcine pancreas elastase (Sigma E-0258), porcine pancreas trypsin (Sigma T-0134) and bovine pancreas α-chymotrypsin (Sigma C-7762). The following substrates were used (Bachem Biosciences): Suc-Ala-Ala-Pro-Ala-pNA (L1775), Suc-Ala-Ala-Pro-Arg-pNA (L1720) and Suc-Ala-Ala-Pro-Phe-pNA (L1400), respectively. Serine protease activities were assayed at a concentration of 0.01 mg/ml enzyme in 50 mM Tris buffer, pH 7.0, with 0.01% v/v Triton X-100. The reaction was initiated by the addition of 200 μM substrate and monitored at 405 nm. IC_{50} values were calculated from single inhibitor concentration measurements (usually 100 μM or 1,000 μM), and K_i values were estimated for each inhibitor and protein pair using reported K_m values: 190 μM for elastase[52], 37 μM for trypsin[53] and 50 μM for α-chymotrypsin[54].

***In vitro* proteasome activity assay.** Each reaction contained the substrate Suc-LLVY-AMC (R&D Systems) at a final concentration of 150 μM; assay buffer (25 mM HEPES, pH 7.4, 100 mM KCl, 20 mM MgCl2, and 10% glycerol); 1% DMSO, 100 μM compound or 5 nM PA26; and 1 nM of the yeast 20S proteasome (except for the control reaction). Following the addition of the proteasome, fluorescence was read every 20 s, for 20 min. Experiments were conducted in duplicates at 23 °C. Activity rates were calculated on the basis of the last 10 min of a reaction. The yeast 20S proteasome and 26S proteasome activator complex were generous gifts from the laboratory of P. Coffino (UCSF).100-μl reactions were performed in 96-well plates using a SpectraMax M5 Microplate Reader.

Microbiology. Susceptibility testing was performed and interpreted following the guidelines of the Clinical and Laboratory Standards Institute[25]. The compounds were dissolved in DMSO, and dilutions were made into Muller-Hinton medium, keeping DMSO <5%. Inhibitors were tested for synergy with the third-generation β-lactam cefotaxime against clinical bacteria. The ratio of β-lactam to inhibitor was 1:4. Each value reported reflects the average of three independent experiments. The bacteria exhibited high levels of resistance to cefotaxime because of the expression of class C β-lactamases (AmpC) or class A extended-spectrum β-lactamases (ESBLs). Three *Escherichia coli* strains and one strain each of Citrobacter freundii, Enterobacter aerogenes and Enterobacter cloacae showed an AmpC-overproduction phenotype. Finally, two of the *E. coli* strains produced the plasmid-mediated class A ESBLs, TEM-3 and CTX-M-14 (**Table 1**).

Crystallography. Co-crystals of AmpC in complex with the inhibitors MAPB, **3**, **7** and **14** (PDB codes 4LV0, 4LV1, 4LV2 and 4LV3) were grown by the hanging drop vapor diffusion method equilibrated over 1.7 M potassium phosphate buffer (pH 8.6–8.8). Protein (4 mg ml^{-1}) was mixed with 1 mM inhibitors and incubated for 20–30 min. Drops were set up by mixing 2 μl of protein-inhibitor with 2 μl of well solution; and 1 μl of microseeding solution was added to promote crystal growth. Crystals appeared after 2–7 d of equilibration at 20 °C. Before data collection, crystals were immersed in a cryoprotectant solution, composed of 25% sucrose, 1.7 M potassium phosphate, pH 8.7, for about 30 s and were flash-cooled in liquid nitrogen. The cryoprotectant solution also contained the respective inhibitor at 1 mM concentration.

Diffraction was measured at beamline 8.3.1 of the Advance Light Source (ALS, Lawrence Berkeley National Laboratory). Reflections were indexed, integrated and scaled using the XDS package[55] in the space group *C*2 with two molecules in the asymmetric unit. Structure refinement was carried out using Phenix[56]. Coot[57] was used for model building, and eLBOW[58] was used to generate coordinates and ligand restraints; The Ser64 oxygen-boron bond length was set to 1.45 Å.

The initial phasing model was based on apo-AmpC (PDB code 1KE4), with water molecules and ions removed. B-factors were refined isotropically, and the protein was subjected to TLS refinement; three TLS groups were determined for each chain using the TLSMD server[59].

Similarly, co-crystals of $RSK2^{T493M}$ in complex with **24** (PDB code 4M8T) were grown by the hanging drop diffusion as described previously[27], and diffraction data were collected at the ALS on beamline 8.2.2. Diffraction images were indexed and integrated using XDS. Molecular replacement was performed using apo-RSK2 CTD as a starting model (PDB code 2QR8) using Phaser, and TLS refinement was carried out using Phenix. See **Supplementary Table 10** for crystallographic statistics.

RSK2 kinase assays. Wild-type and $RSK2^{T493M}$ kinase activity were assayed as reported previously[26]. Briefly, ERK2-activated RSK2 CTD (5 nM) was incubated with varying concentrations of each inhibitor for 30 min in the presence of 100 μM ATP and 10 mM GSH. Each reaction was initiated by the addition of 167 μM substrate peptide (RRQLFRGFSFVAK) and 0.3 μCi μl^{-1} [γ-^{32}P]ATP in a final volume of 25 μl for an additional 30 min. Reactions were spotted on phosphocellulose membranes, washed once with 10% AcOH, twice with 0.1% H_3PO_4 and once with MeOH before drying. Blots were exposed to a phosphor storage plate (Amersham Biosciences), imaged with a Typhoon scanner (GE Healthcare) and quantified using the SPOT program[60]. IC_{50} values were calculated using a sigmoidal dose response fitting in the Prism program (GraphPad) and are reported as the average of two biological experiments. Full IC_{50} curves are presented in **Supplementary Figure 17**.

JAK3 kinase assays. JAK3 kinase activity was assayed using recombinant JAK3 (Invitrogen, catalog no. PV5774). JAK3 (3.1 nM) was incubated with varying concentrations of each inhibitor for 30 min in the presence of 11.5 μM ATP. Each reaction was initiated by the addition of 17.9 μM substrate peptide and 0.3 μCi μl^{-1} [γ-^{32}P]ATP in a final volume of 25 μl for an additional 60 min. Reactions were spotted on phosphocellulose membranes, washed once with 10% AcOH, twice with 0.1% H_3PO_4 and once with MeOH before drying. Blots were exposed to a phosphor storage plate (Amersham Biosciences), imaged with a Typhoon scanner (GE Healthcare) and quantified using the SPOT program[60].

JAK kinase selectivity. Kinase selectivity dose-response experiments for **31** and **33** were performed by Nanosyn (Santa-Clara, CA). Test compounds were diluted in 100% DMSO using threefold dilution steps. Final compound concentration in assay ranged from 10 μM to 0.056 nM. Compounds were tested in a single well for each dilution, and the final concentration of DMSO in all assays was kept at 1%. All assays were performed with a substrate concentration of 1 μM and K_m ATP concentration. Enzyme concentrations ranged from 0.1 nM to 2 nM, and incubation times were 2–4 h. See **Supplementary Table 11** for exact concentrations. IC_{50} values were calculated using a sigmoidal dose response fitting in the Prism program (Graphpad) and are reported as the average of two biological experiments.

Assay demonstrating dissociation of covalent inhibitors from RSK2 CTD and JAK3. Compound **24** (5 μM) and WT RSK2 CTD (100 nM) were incubated in the presence of 10 mM GSH and 100 μM ATP for 1 h in RSK kinase buffer (20 mM HEPES buffer, pH 8.0, with 10 mM $MgCl_2$, 0.2 mg/ml BSA and 2 mM Tris(carboxyethyl)phosphine (TCEP)). The solution was then diluted 20-fold in duplicate into the RSK kinase buffer or RSK kinase buffer containing **24** (final concentration 5 μM). After 1 h, kinase activity was then assayed as previously described.

Compound **31** (2 μM) and JAK3 (62.6 nM) were incubated in Invitrogen's recommended JAK3 kinase buffer (50 mM HEPES, pH 7.5, 0.01% BRIJ-35, 10 mM $MgCl_2$ and 1mM EGTA) in the presence of 11.5 μM ATP, 2.5 mM DTT and 0.2 mg/ml BSA for 1 h. The solution was then diluted 20-fold in duplicate into the JAK3 kinase buffer or JAK3 kinase buffer containing **31** (final concentration 2 μM) in the presence of 100 μM ATP, 2.5 mM DTT and 0.2 mg/ml BSA for 1 h. After an additional hour, kinase activity was assayed as previously described.

Cell-based assay and western blotting. Confluent COS-7 cells were transfected overnight with HA-tagged full-length RSK2, MSK1 and $MSK1^{C458V}$, as reported previously[27]. Transfected cells were seeded into six-well plates at 500,000 cells/well in 2 ml DMEM supplemented with 10% FBS (Axenia),

doi:10.1038/nchembio.1666

100 units/ml penicillin and 100 µg/ml streptomycin (Gibco) and allowed to adhere for 4 h. Cells were serum starved for 18 h, then treated with inhibitor or DMSO for an additional 2 h. Cells were stimulated with PMA (100 ng/ml) for 30 min, washed with cold PBS and frozen. Cells were thawed into 60 µl of lysis buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 0.1% Triton X-100, supplemented with Roche phosphatase and protease inhibitor cocktails). Lysates were clarified by centrifugation at 14,000 r.p.m., normalized by Bradford assay, denatured in SDS and separated by 7.5% acrylamide SDS-PAGE. Gels were transferred to nitrocellulose; blocked with Odyssey LiCOR blocking buffer for 1 h; and probed with 1:1,000 HA, 1:1,000 pS380 RSK2 or 1:1,000 pS376 MSK1 antibody dilutions. After thorough washing with TBST, blots were incubated with 1:10,000 dilutions of fluorescent secondary antibodies (Odyssey) for 1 h, washed with TBST and scanned on an Odyssey LiCOR instrument. Raw gel images can be found in **Supplementary Figure 15**. Phospho-S380 RSK2 (cat. no. 9335S) and phospho-S376 MSK1 (cat. no. 9591S) antibodies were purchased from Cell Signaling. The HA antibody was 12CA5 from Roche, cat. no. 11 666 606 001. COS-7 cells were obtained from the ATCC.

Compound sources. Compounds were sourced from the following vendors: Combi-Blocks, **1**–**11** and **13**–**15**; Matrix Scientific, **12**; Alfa-Aesar, **16**; TCI, **17**; PepTech, **18**. These compounds were sourced at 95% or greater purity as described by the vendors. Compounds **19**–**42** were synthesized in-house, see **Supplementary Note 2** for synthetic chemistry details and characterization.

43. Gasteiger, J., Rudolph, C. & Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Computer Methodology* **3**, 537–547 (1990).
44. Li, J. *et al.* Extension of the platform of applicability of the SM5. 42R universal solvation model. *Theor. Chem. Acc.* **103**, 9–63 (1999).
45. Hawkins, P.C., Skillman, A.G., Warren, G.L., Ellingson, B.A. & Stahl, M.T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **50**, 572–584 (2010).
46. Mysinger, M.M., Carchia, M., Irwin, J.J. & Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
47. Gilson, M.K., Sharp, K.A. & Honig, B.H. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem.* **9**, 327–335 (1988).
48. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).
49. ADAMS, J. *et al.* Boronic acids and esters as inhibitors of fatty acid amide hydrolase. WO Patent 2,008,063,300 (2008).
50. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
51. Brozell, S.R. *et al.* Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J. Comput. Aided Mol. Des.* **26**, 749–773 (2012).
52. Del Mar, E.G., Largman, C., Brodrick, J.W., Fassett, M. & Geokas, M.C. Substrate specificity of human pancreatic elastase 2. *Biochemistry* **19**, 468–472 (1980).
53. Pouvreau, L. *et al.* Effect of pea and bovine trypsin inhibitors on wild-type and modified trypsins. *FEBS Lett.* **423**, 167–172 (1998).
54. Rodríguez-Martínez, J.A., Rivera-Rivera, I., Sola, R.J. & Griebenow, K. Enzymatic activity and thermal stability of PEG-alpha-chymotrypsin conjugates. *Biotechnol. Lett.* **31**, 883–887 (2009).
55. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
56. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
57. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
58. Moriarty, N.W., Grosse-Kunstleve, R.W. & Adams, P.D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).
59. Painter, J. & Merritt, E.A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111 (2006).
60. Knight, Z.A., Feldman, M.E., Balla, A., Balla, T. & Shokat, K.M. A membrane capture assay for lipid kinase activity. *Nat. Protoc.* **2**, 2459–2466 (2007).

ERRATUM

Covalent docking of large libraries for the discovery of chemical probes

Nir London, Rand M Miller, Shyam Krishnan, Kenji Uchida, John J Irwin, Oliv Eidam, Lucie Gibold, Peter Cimermančič, Richard Bonnet, Brian K Shoichet & Jack Taunton

Nat. Chem. Biol. **10**, 1066–1072 (2014); published online 26 October 2014; corrected after print 22 December 2014

In the version of this article initially published, the scaffold structure shown for compounds 33 and 34 was incorrect in Figure 4a, with a nitrogen atom misplaced within the five-membered ring moiety. The error has been corrected in the HTML and PDF versions of the article.