Fragment-based identification of ligands for bromodomain-containing factor 3 of Trypanosoma cruzi

Corentine M. C. Laurin,¹ Joseph P. Bluck,¹,² Anthony K. N. Chan,¹ Michelle Keller,¹ Andrew Boczek,¹ Amy R. Scorah,¹ K. F. Larissa See,¹ Laura E. Jennings,¹ David S. Hewings,¹ Fern Woodhouse,¹ Matthias Schiedel,¹ Philip G. Humphreys,³ Philip C. Biggin,² Stuart J. Conway.¹

¹Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford, OX1 3TA, UK
²Department of Biochemistry, 3 Parks Road, Oxford, OX1 3QU, UK
³GlaxoSmithKline R&D, Stevenage, Hertfordshire, SG1 2NY, UK

KEYWORDS: Chagas disease, Trypanosoma cruzi, bromodomain-containing factor 3, BI-2536, fragment-based discovery.

ABSTRACT: The Trypanosoma cruzi (T. cruzi) parasite is the cause of Chagas disease, a neglected disease endemic in South America. The life cycle of the T. cruzi parasite is complex and includes transitions between distinct life stages. This change in phenotype (without a change in genotype) could be controlled by epigenetic regulation, and might involve the bromodomain-containing factors 1-5 (TcBDF1-5). However, little is known about the function of the TcBDF1-5. Here we describe a fragment-based approach to identify ligands for T. cruzi bromodomain-containing factor 3 (TcBDF3). We expressed a soluble construct of TcBDF3 in E. coli, and used this to develop a range of biophysical assays for this protein. Fragment screening identified twelve compounds that bind to the TcBDF3 bromodomain. Based on this screen, we developed functional ligands containing a fluorescence or ¹⁹F reporter group, and a photo-crosslinking probe for TcBDF3. These tools compounds will be invaluable in future studies on the function of TcBDF3 and will provide insight into the biology of T. cruzi.

Introduction
Chagas disease is a life-threatening illness that is endemic to Latin America and affects 7-8 million people worldwide. Population mobility has resulted in a spread of the disease to the USA, Canada and Europe, making it a truly global problem.¹ While the initial acute phase of the disease can be asymptomatic, it has the potential to be fatal. In addition, between 20% and 30% of infected patients develop irreversible cardiovascular, gastrointestinal, or neurological complications. Two treatments, nifurtimox and benznidazole, are available but they are only effective if taken during the acute phase of the disease and their mechanism of action is unknown.² Chagas disease is caused by the protozoan parasite Trypanosoma cruzi (T. cruzi), which is transmitted to mammals by hemagogous triatominae insects.²

The T. cruzi life cycle is complex: in triatominae insects, the parasite exists as epimastigotes and metacyclic trypomastigotes, whereas it differentiates into amastigotes and trypomastigotes in humans. Transitions between the epimastigote, trypomastigote, and amastigote forms are required for infection and replication.³ These changes in phenotype occur without alteration to the parasite DNA sequence, suggesting that epigenetic mechanisms might play a role in these processes.⁴

Post-translational modifications (PTMs) of histone proteins are one key mechanism that mediates epigenetic processes in humans.⁵ Acetylation of lysine is a heavily studied PTM that occurs not only on histones, but at 3600 sites in 1750 proteins in MV4;11, A549 and Jurkat cells. Lysine acetylation is introduced by ‘writer’ proteins, histone/lysine acetyl transferases (HATs/KATs), and removed by ‘eraser’ proteins, histones/lysine deacetylase (HDACs/KDACs). Bromodomains are ‘readers’ of acetylated lysine (KAc) residues.⁶⁻⁷

Over the past decade, there has been intense study of the role human bromodomains⁸⁻¹⁰ play in oncology indications, leading to >30 clinical trials involving bromodomain ligands. In contrast, little is known about the role of non-human bromodomains. Here we present initial work towards the development of tool compounds to assist in the study of bromodomain function in T. cruzi.

The T. cruzi genome encodes for five established bromodomain-containing protein factors (BCFs), TcBDF1-5 (Figure 1E),¹¹ with two more recently identified: TcBDF6-7.¹² TcBDF2 binds to acetylated histones and accumulates after UV irradiation;¹³ overexpression of wild-type TcBDF1 has effects on the epimastigote, trypomastigote, and amastigote life stages.¹⁴ Interestingly, TcBDF3 was found in the cytoplasm, the flagellum, and the flagellar pocket and is thought to bind to α-tubulin. While these studies are starting to elucidate the role of these proteins in the life cycle of the parasite, there is still much to learn about the functions of these proteins. Given the fundamental role played by some bromodomain-containing proteins in humans,¹⁵⁻¹⁷ we aimed to develop small molecule tools to help elucidate the roles of bromodomains in parasite development, replication, and survival.

While the TcBDF3 bromodomain (TcBDF3₁₂₋₁₄₉; TcBDF3-BRD) does not show high sequence identity to any of the human bromodomains we considered (data not shown), it has greater identity to the second bromodomains of the human bromodomain and extra-terminal (BET) bromodomain-containing proteins than to the first BET bromodomains. Sequence analysis revealed 22-25% identity with BRD2-4(2) (Table SI-A1); TcBDF3 also has 65% identity with PP-BRD20, which is a bromodomain from Trypanosoma brucei gambiense (strain MHOM/CI/86/DAL972) (Figure 1E). In fact, TcBDF3 has higher identity to PP-BRD20 than to the other bromodomains from T. cruzi (Figure SI-A1). An X-ray crystal structure of the PP-BRD20 bromodomain bound to the human bromodomain ligand BI-2536 (1) has been solved (Figure 1-D, PDB code SC8G).
As the MD predicted (Figure 1D) the human BET proteins adjacent side of the wall surrounding L92 and comprising ZA loop residues. The one side the ZA channel is predicted to feature a flat hydrophobic residue (observed in other bromodomain Tc BRD20, and BRD2(2) occupancy during MD simulations). The KAc site water molecules were inferred from regions of high water solution (entire Tc BDF3 to have the same positions as for PP-BRD20 and HmBRD2(2).)

This provides useful structural information and indicates that ligands for human bromodomains might act as lead compounds for developing TcBDF3-BRD ligands. Taken together, these observations encouraged us to focus our investigations on developing ligands for the TcBDF3-BRD.

Results and discussion

As there is currently no X-ray crystal structure of TcBDF3 reported, we used Modeller (v. 9.2) to generate a homology model of the entire TcBDF3-BRD sequence to help guide our work. Molecular dynamics (MD) simulations predicted this model to be stable in solution (Figure 1A, see SI for details). The positions of the binding site water molecules were inferred from regions of high water occupancy during MD simulations and in analogy to T. brucei (Figure 1B). The homology model (Figure 1A) predicts TcBDF3 to have the same left-handed bundle of four α-helices observed in other bromodomain structures, and an asparagine residue (N124) that we expect to bind to KAc residues (Figure 1B). The KAc-binding site was predicted to be smaller than those found in the PP-BRD20 template structure, and the human BET BRDs. One side the ZA channel is predicted to feature a flat hydrophobic wall surrounding L92 and comprising ZA loop residues. The adjacent side of the ZA channel, which makes up the “WPF shelf” in the human BET proteins, comprises the ASF residues in TcBDF3 (Figure 1D).

As the MD predicted the bromodomain construct to be stable, the full TcBDF3 gene (GenBank ID: XP_812334.1) was used to clone 9 bromodomain constructs based on the sequence used in MD, but with varying start and end points. Expression in E. coli and purification yielded a soluble protein construct: H6-TcBDF3_148 (Figure SI-A). To the best of our knowledge, this is the first soluble construct reported for TcBDF3-BRD, although others have expressed the full length protein and reconstituted it from inclusion bodies. Analysis of the expressed protein using circular dichroism revealed a similar α-helix signature to that observed for HmBRD4(1) (Figure SI-A), providing confidence that the protein was folded appropriately.

Fragment screen

At the time this work commenced, no TcBDF3-BRD ligands had been reported. Therefore, to maximise our chances of success, we took a fragment-based approach to identify ligands for TcBDF3-BRD. Based on observations from our homology model, we assembled a 50-member focused library of fragments that contain known KAc mimics. The 1,3-dimethyl-1,3-dihydro-2H-benzo[d]imidazol-2-one KAc mimic was included as I-BET151 (2) binds to T. brucei BDF2 through this moiety, and not the 3,5-dimethylisoxazole. We also included a range of 3,5-dimethylisoxazole-based compounds based on 2 and OXFBD2 (3), which have been developed into effective BET inhibitors and CREBBP bromodomain ligands. We included a selection of 5-amino-4-chloro-2-methylpyridazin-3(2H)-one derivatives, which have been used as the basis to develop PCAF and GCN5 bromodomain ligands. Based on the X-ray crystal structure of BI-2536 (1) bound to PP-BRD20 (PDB code: 5C8G), we included 1, and three fragments based on its KAc mimic. We also included KAc mimics based on the BET bromodomain ligand BI-BET726 (4) and

Figure 1. A. The X-ray crystal structure of pp-BRD20 (carbon = blue; PDB code 5C8G) overlaid with HmBRD2(2) (carbon = green; PDB code 4UYG) and the homology model of TcBDF3 (carbon = white). This overlay shows that (B) the key residues in KAc binding site are conserved, and that (C) the water molecules in the KAc binding site are predicted to be the same positions as for PP-BRD20 and HmBRD2(2). D. The WPF shelf region of HmBRD2(2) is not conserved in PP-BRD20 or TcBDF3, with the residues replaced by ASF in the latter. E. The sequence alignment of the TcBDF3-5 bromodomains, PP-BRD20, and HmBRD2(2).
the CREBBP bromodomain ligand TPOP146 (5). During the course of our work, Serra et al.28,29 reported 6 and 7, as ligands for TcBDF3-BRD, and so these compounds were also included. The initial fragment screen was conducted using waterLOGS, which is a useful technique for identifying low affinity ligands and fragments,30 and does not require a previously identified binding partner for TcBDF3.

We were pleased to find that 12 of the 52 tested compounds bound to the TcBDF3 bromodomain (Table 1). Fragments based on the 1,3-dimethyl-1,3-dihydro-2H-benzo[d]imidazol-2-one, 3,5-dimethylisoxazole, and the dihydropteridinone KAc mimics showed binding. We were surprised to note that compound 6 did not bind to the TcBDF3 bromodomain, when assessed by waterLOGSY. NMR analysis showed that compound 7 formed soluble aggregates in solution, and so its binding could not be assessed using waterLOGSY. Based on these results we selected the 3,5-dimethylisoxazole- and the dihydropteridinone-based compounds for further investigation. While three of the 1,3-dimethyl-1,3-dihydro-2H-benzo[d]imidazol-2-one-based compounds were observed to bind to the TcBDF3 bromodomain, we were concerned that there was no scope for expanding the SAR of this fragment, and so it was not progressed for further analysis.

Affinity determination

We quantified the TcBDF3-BRD affinity of 1, 2, 3 and, 8 using 1H NMR ligand-observed protein titration.32 To obtain the $K_d$ values we analysed the changes in ligand peak chemical shifts and broadening that occurred upon protein titration (Figures SI-B). Using this approach, 2 and 3 were determined to have $K_d$ values for TcBDF3-BRD of 213 ± 80 μM and 482 ± 27 μM, respectively. 1 has a $K_d$ of 50 ± 5 μM (Table 2), while 8, which is a fragment based on the BI-2536 (1) KAc mimic, has a $K_d$ of 366 ± 4 μM. The TcBDF3-BRD affinity of 3, and the closely-related BI-6727 (S3), was also investigated using ITC (Table 2 and Figure SI-D).

### Table 1. Results of the waterLOGSY (WL) screening of 50 compound that contain a KAc mimic, or which have been reported to bind to TcBDF3. ✓ = compound binds; ✗ = compound does not bind; ? = unclear whether the compound binds; NS = compound insoluble at a concentration of 62.5 μM in phosphate buffer and 1% DMSO.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>Binds TcBDF3 by WL</th>
<th>Compound</th>
<th>Structure</th>
<th>Binds TcBDF3 by WL</th>
<th>Compound</th>
<th>Structure</th>
<th>Binds TcBDF3 by WL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure" /></td>
<td>✓</td>
<td>2</td>
<td><img src="image2.png" alt="Structure" /></td>
<td>✓</td>
<td>3</td>
<td><img src="image3.png" alt="Structure" /></td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td><img src="image4.png" alt="Structure" /></td>
<td>✗</td>
<td>7</td>
<td><img src="image5.png" alt="Structure" /></td>
<td>NS</td>
<td>8</td>
<td><img src="image6.png" alt="Structure" /></td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td><img src="image7.png" alt="Structure" /></td>
<td>✓</td>
<td>10</td>
<td><img src="image8.png" alt="Structure" /></td>
<td>✓</td>
<td>11</td>
<td><img src="image9.png" alt="Structure" /></td>
<td>✗</td>
</tr>
<tr>
<td>12</td>
<td><img src="image10.png" alt="Structure" /></td>
<td>✗</td>
<td>13</td>
<td><img src="image11.png" alt="Structure" /></td>
<td>✗</td>
<td>14</td>
<td><img src="image12.png" alt="Structure" /></td>
<td>✗</td>
</tr>
<tr>
<td>15</td>
<td><img src="image13.png" alt="Structure" /></td>
<td>✗</td>
<td>16</td>
<td><img src="image14.png" alt="Structure" /></td>
<td>✗</td>
<td>17</td>
<td><img src="image15.png" alt="Structure" /></td>
<td>✗</td>
</tr>
<tr>
<td>18</td>
<td><img src="image16.png" alt="Structure" /></td>
<td>✗</td>
<td>19</td>
<td><img src="image17.png" alt="Structure" /></td>
<td>✗</td>
<td>20</td>
<td><img src="image18.png" alt="Structure" /></td>
<td>✗</td>
</tr>
<tr>
<td>21</td>
<td><img src="image19.png" alt="Structure" /></td>
<td>✗</td>
<td>22</td>
<td><img src="image20.png" alt="Structure" /></td>
<td>✗</td>
<td>23</td>
<td><img src="image21.png" alt="Structure" /></td>
<td>✗</td>
</tr>
<tr>
<td>24</td>
<td><img src="image22.png" alt="Structure" /></td>
<td>✗</td>
<td>25</td>
<td><img src="image23.png" alt="Structure" /></td>
<td>✗</td>
<td>26</td>
<td><img src="image24.png" alt="Structure" /></td>
<td>✓</td>
</tr>
</tbody>
</table>
27 ✓ 28 ✓ 29 ✓ 30 ? 31 ✓ 32 ✕ 33 ✓ 34 ✕ 35 ✕ 36 ✕ 37 ✕ 38 ✓ 39 NS 40 ✕ 41 ✕ 42 ✕ 43 ✕ 44 ✕ 45 ✕ 46 ✕ 47 ✕ 48 ✕ 49 ✕ 50 ✕ 51 ✕ 52 ✕ 53 ✕ 54 ✕ 55 ✕ 56 ✕

Figure 2. The TcBDF3 tool compounds 54 and 56 bind to TcBDF3-BRD and can be used as reporter molecules in biophysical assays. A. The overlay of 564 MHz 19F NMR spectra showing the changes of width and chemical shift of fluorine resonances of 54 (100 µM) in the presence of TcBDF3-BRD (20, 50, or 100 µM). For clarity only one of the two fluorine resonance is show, see SI-E for full spectra. B. Overlay of mass spectra showing increasing cross-linking to the protein when increasing concentrations of 56 were irradiated at 302 nm in the presence of TcBDF3-BRD.

ITC gave a $K_d$ value of 14.6 ± 3.0 µM for 1, and 57.1 ± 4.9 µM for 53. Therefore, of the compounds we tested up to this point, 1 has the highest TcBDF3-BRD affinity, and estimations of ligand efficiency (1: LE=0.16, 2: LE=0.16, 3: LE=0.21, 8: LE=0.24) suggest that 8 provides a good starting point for developing ligands and tools to study TcBDF3-BRD. Compound 35 has one of the kinase hinge-binding nitrogen atoms removed compared to 1, and has previously been shown to have reduced PLK1 affinity (by AlphaScreen the IC\textsubscript{50} of 35 is 263 nM and 1870 nM, for BRD4 and PLK1, respectively). It is unclear whether TcBDF3 possesses a PLK1 analogue, but we...
reasoned this compound might be a useful tool to help determine which protein is responsible for any phenotypic effects displayed by 1 on T. cruzi. Using ITC, compound 35 was shown to bind to TcBDF3-BRD with a \( K_d \) value of 18.6±8.4 \( \mu \text{M} \) (vide infra). In contrast, the previously reported TcBDF3-BRD ligands 6 and 7 showed no detectable binding by ITC. The binding of these compounds was previously detected using a tryptophan fluorescence assay using W117. 29 While the homology model predicts this residue to be close to the KAc binding pocket, it is not predicted to interface directly with the inside of the pocket. It is possible that these compounds bind to the KAc-binding pocket and affect W117 fluorescence via allosteric actions.31 However, it is also possible that these compounds bind to another pocket that is closer to W117, which is present in the full construct of the protein used by Garcia et al, but which is not present in our truncated soluble construct. This would explain their lack of affinity in our hands.

The ligand efficiencies of 1 and 8 suggest that the KAc mimic makes a significant contribution to the TcBDF3-BRD affinity of both these compounds. Additionally, the tail of 1 is predicted not to form any key interactions by MD simulations. Therefore, we proposed to develop probes for TcBDF3-BRD by substituting the 2-chloro position of 8 with functional handles.

With the aim of developing higher-throughput assays for TcBDF3-BRD, we designed two fluorine-containing compounds, 54 and 55, based on the 8 headgroup, for which we investigated for use in \( ^{19} \text{F} \) NMR-based detection of TcBDF3-BRD binding. We also designed two diazirine-containing compounds, 56 and 57, based on 1 and 8, respectively, with a view to using them in mass spectrometry-based assays. Finally, based on 1 we designed a fluorescent probe 58 to use in fluorescence polarisation (FP) experiments. The syntheses of the functional probes are shown in Scheme 1, and full details are provided in the supplementary information.

\(^{19} \text{F} \) NMR-based probes for TcBDF3-BRD

To develop fluorine-containing probes for TcBDF3-BRD, compounds 54 and 55 were synthesised. These molecules contain the fragment 8 coupled to a fluorine-containing substituent. Using ITC, 54 and 55 were demonstrated to bind TcBDF3-BRD with \( K_d \) values of 62.4±96.7 \( \mu \text{M} \) and 70.8±145 \( \mu \text{M} \), respectively. Upon titration with 20-100 \( \mu \text{M} \) of TcBDF3-BRD the \( ^{19} \text{F} \) chemical shifts of both 54 and 55 both moved and broadened (Figure 2 and SI-E). The effects were more pronounced for 54, perhaps as a result of the two fluorine atoms being in closer contact with the protein. These data show that 54 is a useful \( ^{19} \text{F} \) NMR probe for use in displacement studies of putative TcBDF3-BRD ligands.

Photo-crosslinking probes for TcBDF3-BRD

To investigate photo-affinity crosslinking probes for TcBDF3-BRD, we designed two compounds, 56 and 57, which incorporate a diazirine group. While there are a range of photo-crosslinking groups available, this group generally provides efficient crosslinking when irradiated with light of ~302nm wavelength. 34 Compound 56 comprises the fragment 8 coupled directly to a diazirine-containing linker. 57 consists of the same linker attached to the piperidine ring of 1. The ability of 56 and 57 to photo-crosslink to TcBDF3-BRD was assessed by incubating increasing concentration of each compound with the protein. As the diazirine concentration was increased more crosslinking was observed until a plateau was reached, indicating that the protein has been saturated with diazirine. We have characterised the diazirines by the concentration at which they reach 50% of their maximum cross-linking effect, which we have defined as \( \text{XL}_{50} \). The \( \text{XL}_{50} \) values of 56 and 57 were determined by titrating the compounds into a fixed concentration of TcBDF3-BRD and irradiating for 10 mins at a wavelength of 305 nm. The \( \text{XL}_{50} \) value is 8.2±1.9 \( \mu \text{M} \) for 56 and 14.5±3.9 \( \mu \text{M} \) for 57 (Figure 2 and SI-G). At the highest tested concentration, 56 showed maximum cross-linking yield of 17%, whereas 57 yielded 8% of cross-linked protein. The \( \text{XL}_{50} \) value, and the percentage crosslinking yield, result from a combination of the non-covalent affinity of the ligand and its ability to covalently link to the protein upon irradiation. While 57 is based on a scaffold that has higher affinity for TcBDF3-BRD than the scaffold of 56, the smaller diazirine shows greater crosslinking to the protein. This likely arises from the position of the diazirine, which is thought to be solvent exposed in the case of 57, whereas it is predicted to be closer to the protein surface for 56. These probes could be used in a competition assay to identify TcBDF3-BRD ligands that prevent the crosslinking of 56 or 57 to the protein.

A fluorescent probe for TcBDF3-BRD

The acetylated analog of 1, compound 59 (Table 2 and Scheme 1), has a \( K_d \) value of 15.9±8.4 \( \mu \text{M} \) for TcBDF3-BRD, which is very similar to that of 1. Interestingly, the enantiomer, 60, showed no binding indicating a specific interaction with TcBDF3-BRD. These results suggest that N-acetylation is a feasible method of attaching substituents to the scaffold of 1. Addition of Alexafluor488 to the piperidine ring of 1, through acetylation, gave compound 58. The serial dilution of fluorescent probe 58 displayed suitable concentration-dependent fluorescent intensity (FI). Additionally, the probe showed stable fluorescent polarisation (FP) at concentrations above 20 nM (Figure SI-F). Using FP, 58 was shown to bind Hc-TcBDF3-BRD with a \( K_d \) value of 3.9±1.2 \( \mu \text{M} \) and have an assay window of approximately 80 mP (see SI-F). These data indicate that 58 could be used in a fluorescence polarisation (FP) assay to screen for ligands for TcBDF3. Using FP we also showed that 58 binds to HmBRD4(1), HmBRD4(2), and the human kinase PLK1 with \( K_d \) values of 9.9±2.5 nM, 14.9±1.7 nM, and 5.5±1.5 nM, respectively (see SI-F for details).

Conclusions

We have expressed a soluble construct of the TcBDF3 bromodomain and used this protein in a range of biophysical assays to identify ligands for TcBDF3-BRD. The human dual bromodomain ligand and kinase inhibitor, BI-2536 (1), was identified as a TcBDF3-BRD ligand, and the basis for the ligand efficient TcBDF3-BRD-fragment 8. We were surprised to find that compounds 6 and 7, which have been previously been reported as ligand for TcBDF3 showed no binding in our hands. Fragment 8 forms the basis for a number of functional TcBDF3-BRD ligands that underpin a range of orthogonal biophysical assays for the identification of further TcBDF3-BRD ligands. These molecules will ultimately provide useful chemical tools to study the phenotype of TcBDF-BRD3 inhibition in the T. cruzi parasite.
Table 2. Quantitative affinity measurements of key compounds on TcBDF3-BRD, HmBRD4(1) and HmPLK1 obtained using NMR, ITC or FP, as stated. N.B.: no binding detected.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>TcBDF3-BRD $K_d$ (μM)</th>
<th>HmBRD4(1) $K_d$ (μM)</th>
<th>HmPLK1 $K_d$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI-2536 (1)</td>
<td><img src="image1" alt="Structure" /></td>
<td>50 ± 5 (NMR, n=2)</td>
<td>0.037 ± 0.003 (ITC)</td>
<td>0.00083 (IC$_{50}$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.6 ± 3.0 (ITC, n=3) (LE = 0.16–0.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-BET151 (2)</td>
<td><img src="image2" alt="Structure" /></td>
<td>232 ± 80 (NMR, n=2) (LE = 0.16)</td>
<td>6.1 ± 0.22 (FP)</td>
<td></td>
</tr>
<tr>
<td>OXFBDO2 (3)</td>
<td><img src="image3" alt="Structure" /></td>
<td>482 ± 27 (NMR, n=2) (LE = 0.21)</td>
<td>0.384 (Alpha Screen)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>N.B. (ITC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>N.B. (ITC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td><img src="image4" alt="Structure" /></td>
<td>366 ± 4 (NMR, n=1) (LE = 0.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td><img src="image5" alt="Structure" /></td>
<td>18.6±8.4 (ITC, n=2) (LE = 0.18)</td>
<td>0.263 (Alpha Screen)</td>
<td>1.870 (Alpha Screen)</td>
</tr>
<tr>
<td>BI-6727 (53)</td>
<td><img src="image6" alt="Structure" /></td>
<td>57.1 ± 4.9 (ITC, n=3) (LE = 0.13)</td>
<td>0.079 ± 0.003 (ITC)</td>
<td>0.00087 (IC$_{50}$)</td>
</tr>
<tr>
<td>54</td>
<td><img src="image7" alt="Structure" /></td>
<td>62.4 ± 96.7 (ITC, n=1) (LE = 0.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td><img src="image8" alt="Structure" /></td>
<td>70.8 ± 145 (ITC, n=1) (LE = 0.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td><img src="image9" alt="Structure" /></td>
<td>253 ± 8 (NMR, n=1) (LE = 0.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td><img src="image10" alt="Structure" /></td>
<td>3.9±1.2 (FP, n=2) (LE = 0.11)</td>
<td>0.0099±0.0025 (FP, n=3)</td>
<td>0.0055±0.0015 (FP, n=2)</td>
</tr>
<tr>
<td>59</td>
<td><img src="image11" alt="Structure" /></td>
<td>15.9±8.4 (ITC, n=1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td><img src="image12" alt="Structure" /></td>
<td>N.B. (ITC)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Scheme 1. Synthesis of fragment 8 and tool compounds for TcBDF3. Reagents and conditions: A. i. NEt₃, Ga(OTf)₃, DCE, RT, 16 h, 65–78% (n=2); ii. 2,4-Dichloro-5-nitropyrimidine, NaHCO₃, cyclohexane, RT then 60 °C, 3 d, 51% (n=1); iii. Glacial acetic acid, 70 °C then Fe (activated), 70 °C, 1h, then 100 °C, 4 h, 61–81% (n=3); iv. MeI, DMF then NaH, -10 °C then RT, 1h, 90–98% (n=3) or K₂CO₃, MeI, DMF, RT, 3 d, 45% (n=1); v. (a) 4-Amino-3-methoxybenzoic acid, HCl, H₂O/EtOH 5:1, 100 °C, 24 h (b) HBTU, DIPA, 1-Boc-4-aminopiperidine, RT, 4 h, 53% (over 2 steps); vi. TFA, CH₂Cl₂, 0 °C, 1 h, 36–71%, n=3. B. vii. NH₄OH/IPA = 4:1, 160–170 °C, 6 h, MW, 80% (n=2); viii. HATU, DIPEA, CH₂Cl₂, RT, 3 h, 58% (n=1); ix. HCl, EtOH/H₂O, 2,4-difluoronaniline, 18 h, 100°C, >99% (n=1); x. HCl, EtOH/H₂O, (trifluoroacetyl)aniline, 18 h, 100°C, 28% (n=1). C. xi. R=Ac: AcCl, CH₂Cl₂, 0 °C, 1 h, 40%; R=AlexaFluor488: AlexaFluor488-NHS, DIPEA, DMF, RT, 18 h, RT, 61%; xii. 63, HATU, DIPEA, CH₂Cl₂, RT, 3 h, 53% (n=1).
Experimental section

General experimental procedures and procedures and characterisation for intermediate compounds can be found in the SI.

\((R)-2\text{-Chloro-8-cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-6(H)-one (8)}\)

\(\text{MeI (10 \, \mu L, 0.14 mmol, 2.0 eq.) was added to a solution of } (R)-2\text{-chloro-8-cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-6(H)-one (34) (20 mg, 0.071 mmol, 1.0 eq.) in dry DMF (0.5 mL). The reaction mixture was cooled to } -10^\circ \text{C using a NaCl and ice bath, then NaH (3.5 mg, 0.14 mmol, 2.0 eq.) was added. The reaction mixture was stirred at } -10^\circ \text{C for 1 h, then was warmed to room temperature and stirred for 1 h. The mixture was quenched with the addition of crushed ice, then was diluted with } H_2O (20 mL) and extracted with EtOAc (2 \times 35 mL). The combined organic components were washed with eq. LiCl (10% w/v, 6 \times 25 mL), dried over MgSO\(_4\), filtered, and concentrated in \textit{vacuo}. Purification by silica gel chromatography, eluting with petroleum ether:EtOAc (1:0 to 1:1), afforded \((R)-2\text{-chloro-8-cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-6(H)-one (8)} (21 mg, 98%) as a colourless solid: \(R_c 0.56\) (petroleum ether:EtOAc = 1:1); \([\alpha]_D^{22}\) = -80.0 \,(c 1.0, MeOH); \([\alpha]_D^{19}\) = 88.3 \,(c 1.0, MeOH); \(v_{\text{max}}\) (thin film)/cm\(^{-1}\) 2968 (w), 1682 (s) (C=O), 1578 (m), 1409 (m), 1058 (m); m.p.; 82–84 °C (EtOAc); \(\delta \text{H} NMR\) (400 MHz, CDCl\(_3\)) \(\delta 7.66\) (1 H, s), 7.68 (1 H, s), 7.41 (1 H, d, \(J = 19.1\)), 7.23 (1 H, dd, \(J = 8.5, 1.4\)), 6.00 (1 H, d, \(J = 7.8\)), 4.65–4.40 (1 H, m), 4.21 (1 H, dd, \(J = 7.9, 3.7\)), 4.16–4.02 (1 H, m), 3.96 (3 H, s), 3.31 (3 H, s), 2.92 (2 H, t, \(J = 12.6\)), 2.21–2.07 (3 H, m), 2.08–1.92 (3 H, m), 1.92–1.60 (8 H, m), 1.46 (11 H, s), 0.87 (3 H, s, \(J = 7.5, 7.5\)), LCMS (high pH) rt 10.17 mins, purity > 99.5%; Analytical HPLC (RP, method A) rt 10.17 mins, purity > 99.5%; Chiral HPLC (IB-NS (250 × 4.6 mm, 5 μm), Hexane:EtOH = 75:25 with 0.2% \(\alpha\)-cyanoisopropylamine, 1.0 mL/min, 25 mins) major rt 14.9 mins, 92.9%, minor rt 17.9 mins, 7.1%. The data are in good agreement with the literature.\(^{40}\)

\(\text{ tert-Butyl (7R)-4-(4'-(8-cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3'-methoxybenzamido)piperidine-1''''-carboxylate (60)} (736 mg, 64%) as a pale brown oil. A small amount of 60 (50 mg) was further purified by MDAP (method D), while the rest was used as such in the following reactions. The mass-directed assisted purification (MDAP) preparative HPLC fractions were collected and concentrated in \textit{vacuo} to give tert-butyl (7R)-4-(4'-(8-cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3'-methoxybenzamido)piperidine-1-carboxylate (26 mg, 51% recovery) as a colourless solid: \(R_{C} 0.55\) (CH\(_2\)Cl\(_2\):MeOH:NH\(_4\)OH = 90:9:1); \([\alpha]_D^{25}\) = -44.9 \,(c 1.0, MeOH); \([\alpha]_D^{19}\) = -112 \,(c 0.2, CHCl\(_3\))\); m.p.; 88–90 °C (CH\(_2\)Cl\(_2\)); \(\delta \text{'H} NMR\) (400 MHz, CDCl\(_3\)) \(\delta 8.51\) (1 H, d, \(J = 8.4\)), 7.66 (1 H, s), 7.41 (1 H, d, \(J = 19.1\)), 7.23 (1 H, dd, \(J = 8.5, 1.4\)), 6.00 (1 H, d, \(J = 7.8\)), 4.65–4.40 (1 H, m), 4.21 (1 H, dd, \(J = 7.9, 3.7\)), 4.16–4.02 (1 H, m), 3.96 (3 H, s), 3.31 (3 H, s), 2.92 (2 H, t, \(J = 12.6\)), 2.21–2.07 (3 H, m), 2.08–1.92 (3 H, m), 1.92–1.60 (8 H, m), 1.46 (11 H, s), 0.87 (3 H, s, \(J = 7.5, 7.5\)), LCMS (high pH) rt 10.17 mins, purity > 99.5%; Analytical HPLC (RP, method A) rt 10.17 mins, purity > 99.5%; Chiral HPLC (IB-NS (250 × 4.6 mm, 5 μm), Hexane:EtOH = 75:25 with 0.2% \(\alpha\)-cyanoisopropylamine, 1.0 mL/min, 25 mins) major rt 14.9 mins, 92.9%, minor rt 17.9 mins, 7.1%. The data are in good agreement with the literature.\(^{41}\)
Levulinic acid (4.0 g, 34 mmol, 1.0 eq.) and 7 M NH₄OH in MeOH (45 mL, 3.2 mm mol, 9.1 eq.) were stirred at 0 °C for 3 h, then hydroxylamine-O-sulfonic acid (5.8 g, 51 mmol, 1.5 eq.) was added portionwise and the mixture was left to warm to room temperature then stirred for 42 h. Nitrogen was bubbled through the reaction mixture for 1 h and the mixture was filtered. The retentate was washed with MeOH (2 × 20 mL) and the filtrate was concentrated in vacuo. The residue was dissolved in MeOH (25 mL), triethylamine (14 mL, 10 mmol, 3.0 eq.) was added and the reaction mixture was cooled to 0 °C under nitrogen in the dark with vigorous stirring until the dark colour remained. 2 M aq. HCl (ca. 30 mL) and aq. sodium thiosulfate (10% w/v; 150 mL) were added, and the mixture was extracted with EtOAc (3 × 60 mL). The combined organic components were washed with aq. sodium thiosulfate (10% w/v; 2 × 100 mL), then filtered over an hydrophobic frit and concentrated in vacuo. The crude was filtered over silica, flushing with CH₂Cl₂:MeOH (9:1, ca. 4 CV), then purified by silica gel chromatography (CombiFlash®Rf, 80 g column), eluting with CH₂Cl₂:MeOH (1:0 to 9:1) to afford 3-(3-methyl-3H-diazirin-3-yl)propanoic acid (67) (1.02 g, 22%) as a colourless liquid. Rf 0.05 (CH₂Cl₂:MeOH = 9:1); νmax (neat)/cm⁻¹ 2980 (w), 1708 (s), 1617 (m), 1542 (m), 1454 (m), 1387 (m), 1223 (m), 1198 (w); ¹H (600 MHz, CDCl₃) 2.25 (2 H, t, J 6.5), 1.73 (2 H, t, J 6.5), 1.06 (3 H, s); ¹C (176 MHz, CDCl₃) δ 178.0, 29.4, 28.4, 25.0 19.7; LCMS (formic acid) rt: 3.54 mins (70%), m/z (ES⁺) 513.3 [M+2H]⁺, 511.3 [M+H]⁺, 509.3 [M+H⁺Na⁺], 100%; HRMS (formic acid) rt: 3.54 mins (70%), m/z (ES⁺) 513.3 [M+2H]⁺, 511.3 [M+H]⁺, 509.3 [M+H⁺Na⁺], 100%; HPLC (formic acid) rt: 3.54 mins (70%), m/z (ES⁺) 513.3 [M+2H]⁺, 511.3 [M+H]⁺, 509.3 [M+H⁺Na⁺], 100%; HRMS (formic acid) rt: 3.54 mins (70%), m/z (ES⁺) 1024.2938, C₅₉H₄₁O₉N₂ requires [M+H⁺] 1024.2970; Two peaks seen in the LCMS due to partial separation of the 5- and 6- regioisomers in the 10 mins LC method. LCMS consistent with product with purity > 90%.

2-Amino-8-cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-6(5H)-one (66)

(R)-2-Chloro-8-cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-6(5H)-one (66) (0.15 g, 0.51 mmol, 1.0 eq.) was dissolved in isopropanol (1.2 mL) and ammonium hydroxide (2.6 mL, 70 mmol, 250 eq.), then submitted to microwave irradiation for 6 h at 170 °C, 6 h at 160 °C and then 4 h at 170 °C. H₂O (150 mL) was added and the mixture was extracted with EtOAc (150 mL). The organic components were removed under a flow of nitrogen, then the residue was dissolved in DMSO (0.9 mL) and purified by MDAP (method C). The solvents were removed under a flow of nitrogen to afford (R)-4-((8-cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3-methoxy-N-(1-((3-(methyl-3H-diazirin-3-yl)propanoyl)pyrrolidin-4-yl)benzamide (65) (24 mg, 0.047 mmol, 1.2 eq.) was added, and the mixture was stirred for 1.5 h. The volatile components were removed under a flow of nitrogen, then the residue was dissolved in DMSO (0.9 mL) and purified by MDAP (method C). The solvents were removed under a flow of nitrogen to afford (R)-4-((8-cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3-methoxy-N-(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)-3-oxo-3H-spiro[indolen-1,9'-xanthene]-4,5'-disulfonate, dilithium salt (58) (1.5 g, 34%) as an amorphous yellow solid. LCMS (high pH) rt: 0.74 mins (100%), m/z (ES⁺) 652.6 [M+H⁺], 542.6 [M+H⁺Na⁺], 100%; HRMS (formic acid) rt: 3.54 mins (70%), m/z (ES⁺) 1024.2938, C₅₉H₄₁O₉N₂ requires [M+H⁺] 1024.2970; Two peaks seen in the LCMS due to partial separation of the 5- and 6- regioisomers in the 10 mins LC method. LCMS consistent with product with purity > 90%.
H, m), 1.74

CDCl (m), 1428 (s), 1279 (m), 845 (m), 778 (m);

ṽ (thin film)/cm

H, m), 1.74

CDCl (m), 1428 (s), 1279 (m), 845 (m), 778 (m);

ṽ (thin film)/cm

N-(8-Cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-
tetrahydropteridin-2-yl)-3-(3-methyl-3H-diazimin-3-
yl)propanamide (56)

HATU (72 mg, 0.19 mmol, 1.2 eq.) and DIPEA (0.055 mL, 0.31
mmol, 2.0 eq.) were added to a solution of 3-(3-methyl-3H-diazimin-
3-yl)propanoic acid (67) (20 mg, 0.16 mmol, 1.0 eq.) in CHCl3 (1.0 mL). The reaction mixture was stirred for 1.5 h at room

temperature, then 2-amino-8-cyclopentyl-7-ethyl-5-methyl-7,8-
dihydropteridin-6(5H)-one (66) (52 mg, 0.19 mmol, 1.2 eq.) in

H2O (0.5 mL) was added. The reaction mixture was stirred for

2.5 h, then HATU (72 mg, 0.19 mmol, 1.2 eq.) and DIPEA (0.055
mL, 0.31 mmol, 2.0 eq.) were added. The reaction was stirred for

2 h, then left to stand for 48 h. Evaporation was observed, hence

CHCl3 (0.5 mL) and DMF (0.25 mL) were added to re-solubilize the
mixture. HATU (72 mg, 0.19 mmol, 1.2 eq.) and DIPEA (0.055
mL, 0.31 mmol, 2.0 eq.) were added and the RM was stirred at room

temperature for 18 h. The reaction mixture was diluted with H2O
(30 mL), then extracted with EtOAc (2 × 30 mL). The combined organic
components were washed with sat. aq. LiCl (50 mL), then

concentrated in vacuo. The residue was dissolved in MeOH:DMSO
(1:1, 1 mL) and purified by MDAP (method C). The fractions were
combined and concentrated under a flow of nitrogen in the dark to afford

N-(8-cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-
tetrahydropteridin-2-yl)-3-(3-methyl-3H-diazimin-3-
yl)propanamide (56) (36 mg, 57%) as an off-white powdery solid:

Rf 0.20 (petroleum ether:EtOAc = 1:1); vmax (thin film)/cm

3339 (w), 2970 (w), 2873 (w), 1665 (s), 1459 (s), 1247 (m),

954 (m); 1H NMR (600 MHz, DMSO-D6) δ 7.58 (1 H, s), 5.87
(2 H, s), 4.36–4.27 (1 H, m), 4.07 (1 H, dd, J 7.9, 3.7), 3.17 (3
H, s), 1.93–1.83 (3 H, m), 1.79–1.65 (5 H, m), 1.62–1.48 (3
H, m), 0.75 (3 H, dd, J 7.5, 7.5); 13C NMR (151 MHz, DMSO-
D6) δ 162.7, 159.1, 151.7, 139.2, 113.8, 59.6, 57.5, 28.9, 28.7, 27.6,
26.1, 23.4, 23.1, 9.0. LCMS (high pH) rt 1.08 mins (100%), m/z

ES+ 387 [(M+H)+], 100%; HR-LCMS (formic acid) rt 2.91

mins (100%), m/z (ES+) found 386.2311 (100%), C10H8N2O2 requires [M+H]+ 386.2299.

(7R)-8-Cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-6(5H)-one (55)

(7R)-2-Chloro-8-cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-
6(5H)-one (8) (20 mg, 0.068 mmol, 1.0 eq.) was suspended in H2O
(0.8 mL) and EtOH (0.2 mL). 2-4-difluoronaniline (12 μL, 0.14
mmol, 2.0 eq.) and 20 μL conc. HCl were added. The resulting
solution was stirred under reflux conditions for 18 h, then sat. aq.
NaHCO3 (20 mL) was added. The mixture was extracted with

CHCl3 (2 × 40 mL). The combined organic components were washed with sat. aq. NaHCO3 (20 mL), dried over MgSO4, filtered, and

concentrated in vacuo. Purification by silica gel chromatography, eluting with petroleum ether and EtOAc (1:0 to

0:1), afforded (7R)-8-cyclopentyl-7-ethyl-5-methyl-7,8-dihydropteridin-
6(5H)-one (54) as an amorphous brown solid (26 mg, >99%): Rf

0.79 (petroleum ether:EtOAc = 1:1); [α]D 23.6, 23.3, 9.3;

ṽ (thin film)/cm

3267 (w), 2966 (w), 1656 (s), 1506 (m), 1428 (s), 1279 (m), 778 (m); 1H NMR (500 MHz,

CDCl3) δ 8.08–7.97 (2 H, m), 7.82–7.73 (2 H, m), 7.69 (1 H, s), 7.30 (1 H, br s), 4.55–4.39 (1 H, m), 4.25 (1 H, dd, J 7.8, 3.7), 3.43

(3 H, s), 2.23–2.08 (1 H, m), 2.08–1.95 (1 H, m), 1.95–1.52 (8 H,

m), 0.88 (3 H, dd, J 7.6, 7.6); 13C NMR (126 MHz, CDCl3) δ 187.8

(q, Jc= 34.2), 163.5, 154.3, 152.39, 147.2, 137.54, 132.2, 122.6,
117.26, 117.17 (q, Jc= 290.0, CF3), 60.2, 58.8, 29.8, 29.3, 28.4, 27.4,
23.6, 23.3, 9.3; 19F NMR (471 MHz, CDCl3) δ –70.8 (3 F, brs, CF3);

LRMS m/z (ES+) 448 [M+H]+, 100%; HRMS m/z (ES+) found

448.19522 (100%), C10H8F3N2O2 requires [M+H]+ 448.19549;

Analytical HPLC (RP, method B) rt 6.75 mins, purity 99%.

ASSOCIATED CONTENT

Supporting Information

For general synthetic procedures, intermediate characterisation, assay and modelling procedures, please refer to the supporting information.

Supplementary information (biological data and synthesis) (pdf)

AUTHOR INFORMATION

Corresponding Author

stuart.conway@chem.ox.ac.uk
ACKNOWLEDGMENT

We thank Emma Grant and Dr Jacob Bush for advice on cross-linking experiments. We are grateful to Prof. Timothy Claridge for advice on NMR titration. C.M.C.L also thanks Jon Seal for helpful synthetic insight. We would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work: http://dx.doi.org/10.5281/zenodo.22558.

REFERENCES


Fragment hit LE=0.24

photocross-linking to TcBDF3

$^{19}F$-based detection of TcBDF3 binding

fluorescent probe for TcBDF3 binding