Human T-box transcription factor T (Brachyury)

A Target Enabling Package (TEP)

Gene ID / UniProt ID / EC	6862 / J3KP65
Target Nominator	The Chordoma Foundation, USA
SGC Authors	Joseph Newman[1], Angeline Gavard[1], Hazel Aitkenhead[1], David Drewry[2], Charles Lin3, Opher Gileadi[1]
Collaborating Authors	
Target PI	Opher Gileadi (SGC Oxford)
Therapeutic Area(s)	Cancer
Disease Relevance	TBXT is a driver gene for the rare cancer, Chordoma.
Date Approved by TEP Evaluation Group	November 3, 2020
Document version	Version 1
Document version date	October 2020
Citation	http://doi.org/10.5281/zenodo .4265333
Affiliations	1. SGC, University of Oxford, 2. SGC, University of North Carolina. 3. Baylor College of Medicine

USEFUL LINKS

(Please note that the inclusion of links to external sites should not be taken as an endorsement of that site by the SGC in any way)

SUMMARY OF PROJECT

Chordoma is a rare cancer occurring along the spinal cord (OMIM: 215400). Chordoma is derived from an embryonic tissue, the notochord, and over-expresses the embryonic transcription factor T-box transcription factor T, the homologue of mouse Brachyury (1-6). Chordomas are “genomicaly silent” cancers that do not carry an extensive mutation load. Recent studies indicate that expression of TBXT is essential for persistence and growth of chordoma cells (7). As TBXT is not expressed in any post-embryonic tissues, it could be an excellent target for treatment of chordoma. The long-term aim of this project is to test whether TBXT can be targeted with small molecules with sufficient affinity and specificity to be therapeutically useful. In this TEP we have determined crystal structures of the DNA-binding domain (DBD) of TBXT with and without cognate DNA oligonucleotides. The DNA-free protein crystals were used in a high-throughput fragment screen to identify 29 fragments bound in 6 clusters. The crystal structures of the bound fragments provide starting points for development of stronger binders which could be used to disrupt TBXT activity or to induce the degradation of the protein through a Proteolysis-targeting chimeric molecule (PROTAC) approach (8).

SCIENTIFIC BACKGROUND

Chordoma is a rare primary bone tumour that develops in the skull base and spine. Current treatments for chordoma are limited with no approved systemic therapy and an average lifespan from diagnosis of 7 years, which highlights a need for therapeutic development. Strong evidence supports the hypothesis that chordoma is driven by expression of the *TBXT* (Brachyury). TBXT controls the expression of genes required for mesoderm and notochord formation and is normally turned off early during vertebrate development. TBXT is a transcription factor from the T-box family. All T-box proteins bind to a similar DNA motif, seen both in biochemical selection of random DNA sequences (9) and in chromatin immunoprecipitation-sequencing (ChIP-seq) assays (10,11). Chordomas are thought to arise from remnants of the notochord that fail to turn off TBXT expression (4). The *TBXT* gene is duplicated in hereditary chordoma and there is a dramatically increased risk of chordoma associated with a mis-sense SNP, rs2305089, which changes glycine 177 to aspartate (G177D) in TBXT's DNA binding domain (12). Finally, chordoma cells require TBXT expression to survive and maintain their cancerous state (2).

From a purely genetic standpoint, TBXT is an ideal therapeutic target in chordoma. It is expressed almost uniquely in the tumour cells and is a driver of the malignant phenotype (4,13,14). However, transcription factors like TBXT have been difficult to inhibit with small molecules. Thus, there are efforts to identify more druggable targets that modulate either TBXT expression, its downstream effectors, or other contributors to the malignant phenotype (7,15-17). Such approaches, as well as immune-based therapies, may eventually form the basis of an effective chordoma-specific therapy. However, there is a concern with systemic effects that may arise from targeting a non-specific cancer driver, especially if the treatment is to be continued over months and years. Therefore, the unique position of TBXT as a chordoma-specific protein warrants an attempt to target it directly.

RESULTS – THE TEP

Proteins Purified

Brachyury is a 435 amino acid protein. The DNA-binding domain (DBD) in the N-terminal region (residues 42-219) is followed by a C-terminal region that is largely unstructured. A coding SNP (rs2305089, changing G177 to D), with an allele frequency of 42%, occurs in >90% of chordoma cases in the West, and is therefore associated with an increased risk for the disease (12).

We have expressed and purified several versions of the protein in *E. coli* cells:

Construct ID	Description	Residue range	Vector	Tag(s)
TBXTA-c005	DNA binding domain (DBD), 177G Used for protein: DNA co-crystals	E41-S224	pET28A	C-terminal His6
TBXTA-c006	DBD, 177D Used for protein: DNA co-crystals	E41-S224	pET28A	C-terminal His6
TBXTA-c020	DBD, 177G Used for fragment screens		pSUMO-LIC	N-terminal His6 and SUMO followed by a SUMO protease cleavage site
TBXTA-c021	DBD, 177D Used for fragment screens	E41-N211	pSUMO-LIC	N-terminal His6 and SUMO, followed by a SUMO protease cleavage site
TBXTA-c025	Full-length, 177D	S2-435	pGTVL2	N-terminal His6-GST-TEV cleavage site
TBXTA-c026	Full-length, 177G	S2-435	pGTVL2	N-terminal His6-GST-TEV cleavage site

Table 1: All sequences are optimized for expression in *E. coli*. Sequences, expression and purification methods are listed below. Plasmids are available in Addgene

Structures

6F58: Crystal structure of human Brachyury (T) in complex with DNA
6F59: Crystal structure of human Brachyury (T) G177D variant in complex with DNA

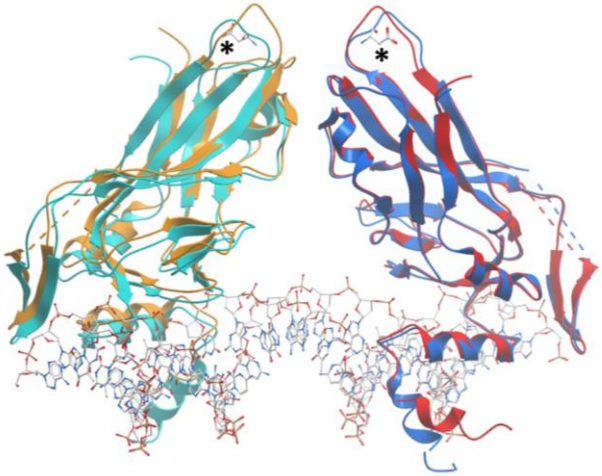


Fig. 1. Structures of WT (red/orange, PDB: 6F58) and G177D variant (blue/cyan, PDB: 6F59) in complex with DNA. The proteins bind to the palindromic DNA sequence as homodimers. The position of the variant residue G/D 177 is marked (*).

Examples of DNA-free crystals:
5QS9: Human Brachyury (TBXT) G177D variant, crystal form 1 with fragment bound
5QRM: Human Brachyury (TBXT) crystal form 2 with fragment bound

These DNA-free crystals were used in a screen of a fragment library to identify binding sites and initial chemical matter.

Fragments were found bound to crystal form 1 in five principal regions, marked A-E. (**Fig. 2A**). Two additional binding regions were identified in crystal form 2 (**Fig. 2B**).

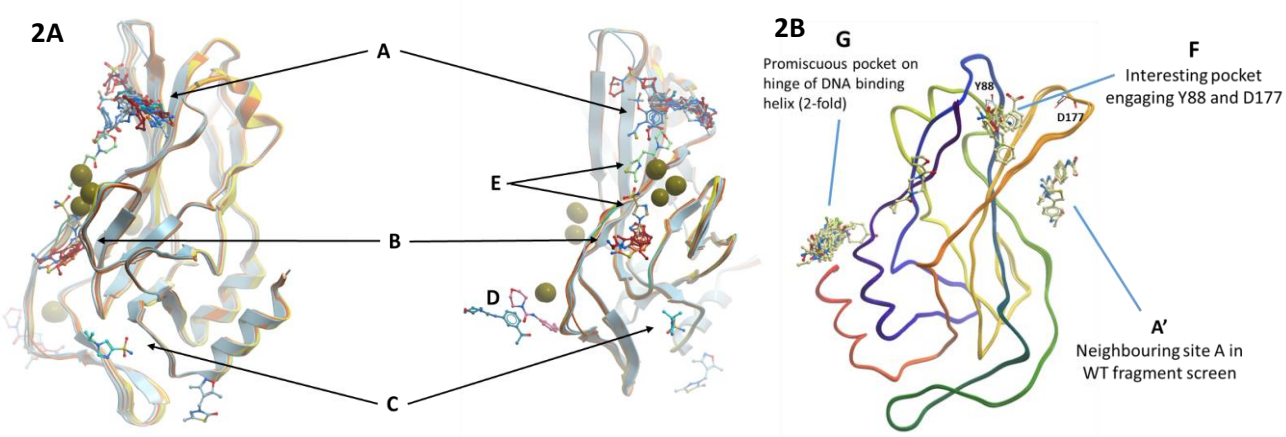


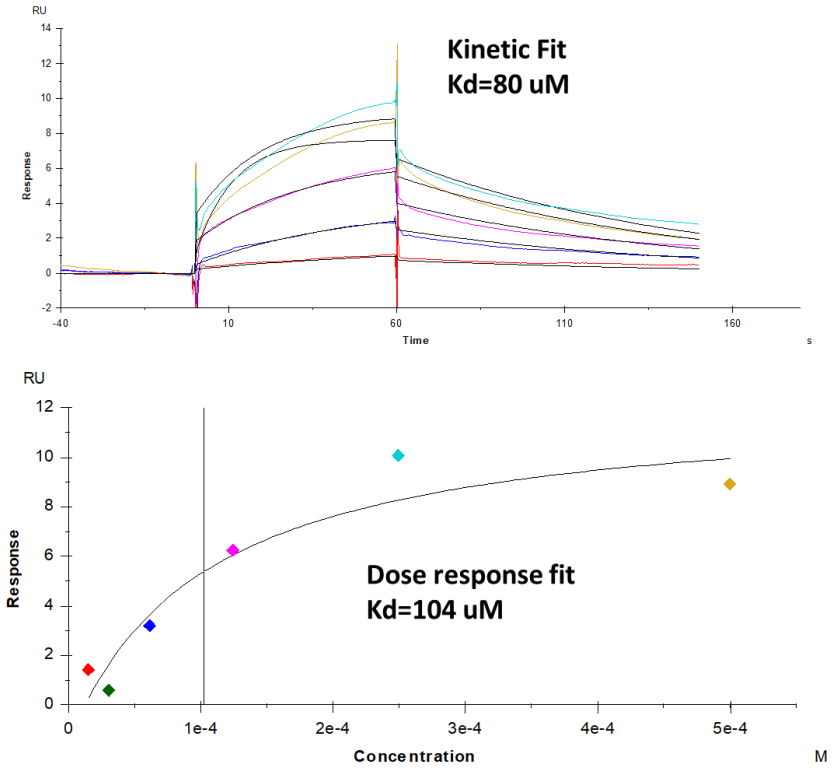
Fig. 2: (**A**) Overlay of fragment hits with crystal form 1. The compounds are depicted as sticks; the yellow spheres are Cd^{2+} ions from the crystallisation mix. (**B**) Overlay of fragment hits with crystal form 2.

Assays

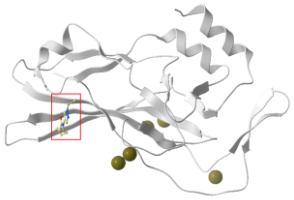
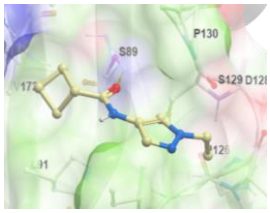
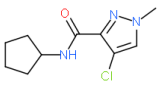
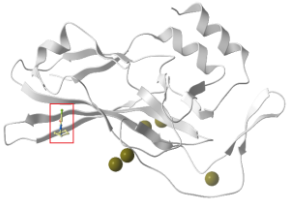
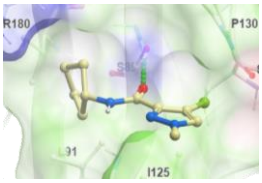
1. DNA binding: we have adapted an electrophoretic mobility shift assay (EMSA) for the binding of short T-box oligonucleotides to brachyury protein. Single shifted bands can be seen using the isolated DNA-binding domain and optimised DNA sequences; when using the full-length protein we see a smeared pattern, probably reflecting the disordered nature of the C-terminal region of the protein.

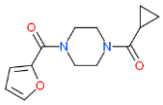
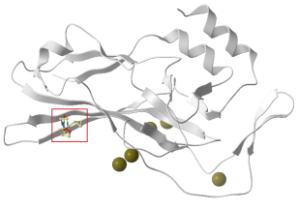
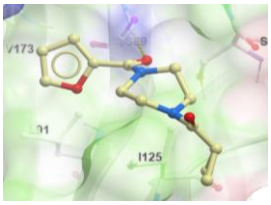
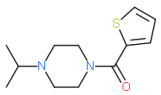
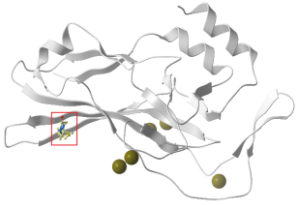
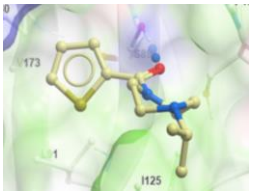
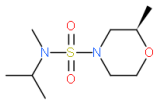
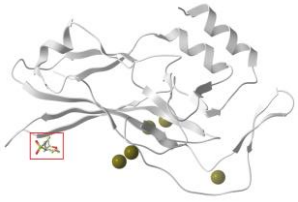
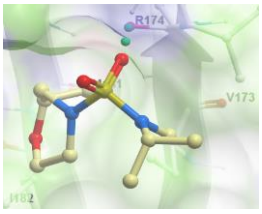
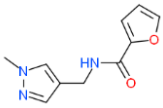
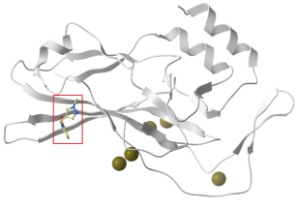
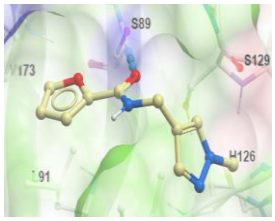
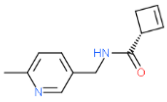
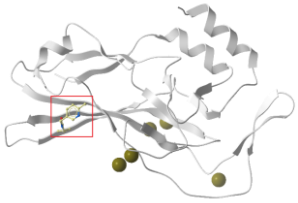
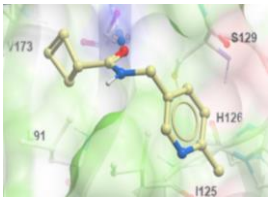
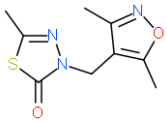
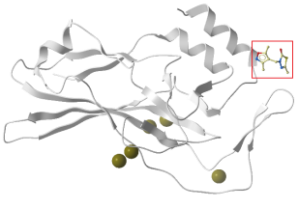
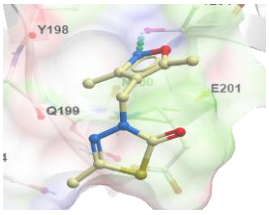
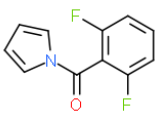
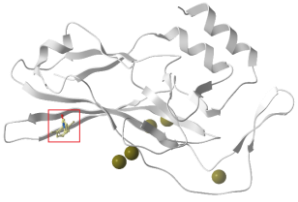
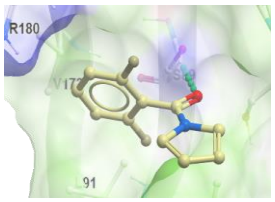
2. Reporter assays in cells: Reporter plasmids bearing a T-box-driven luciferase gene, and a control plasmid with a mutated T-box sequence, were described by the Blumberg lab (18) and are available from Addgene, plasmids 36246 and 36245.
3. Ligand binding: We have developed an SPR assay using a biotinylated form of brachyury protein bound to streptavidin chips.

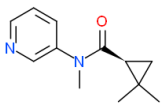
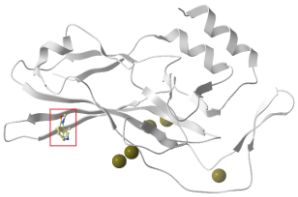
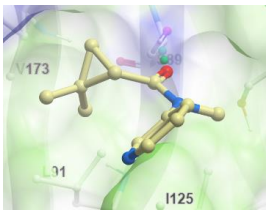
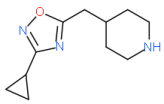
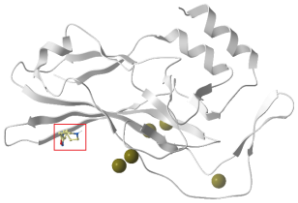
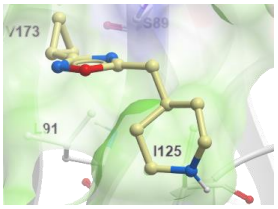
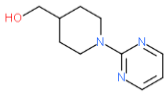
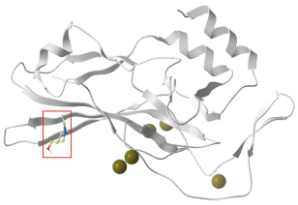
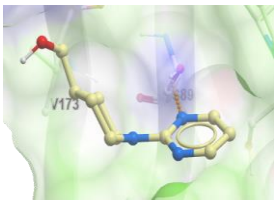
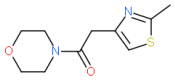
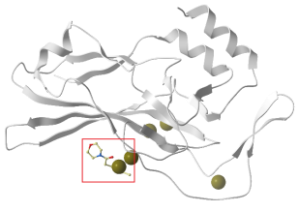
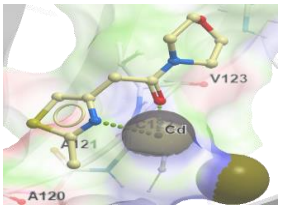
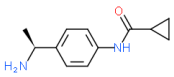
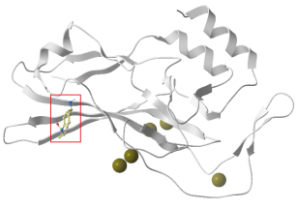
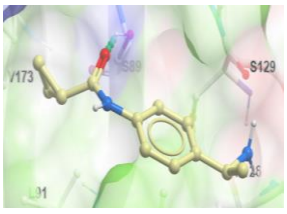
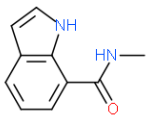
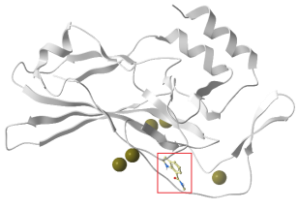
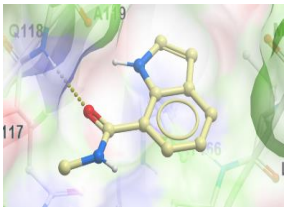
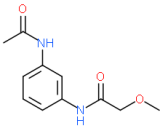
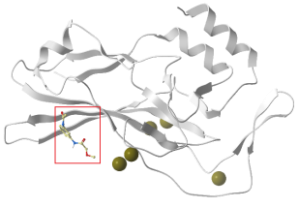
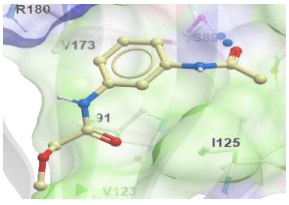
Results seen with a small molecule derived from the fragment screen:

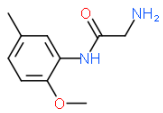
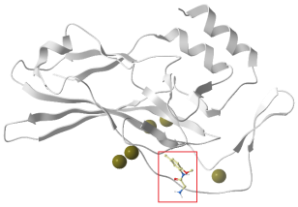
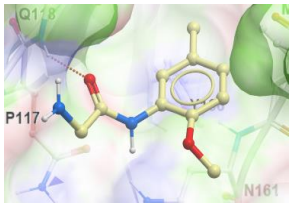
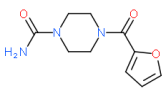
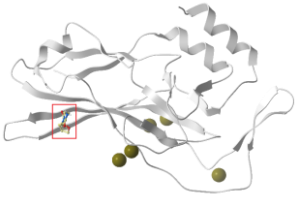
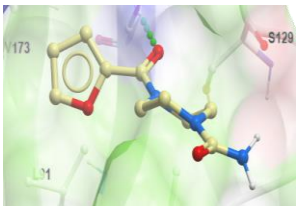
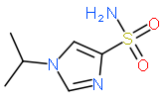
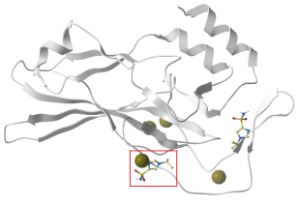
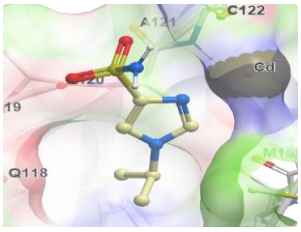
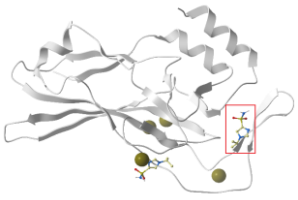
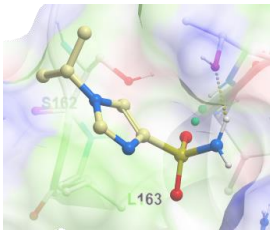
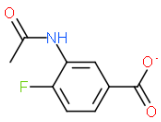
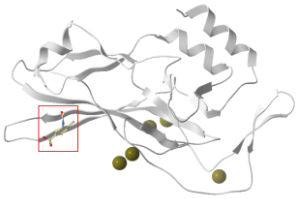
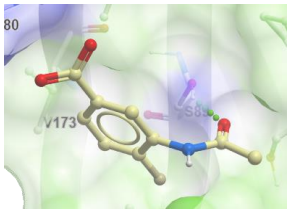
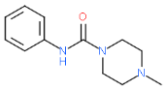
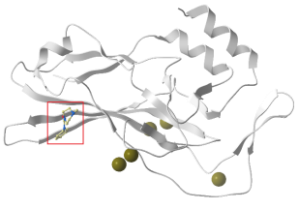
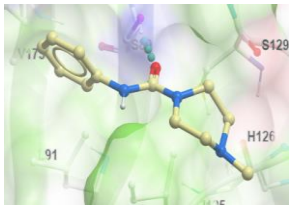
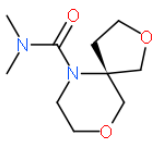
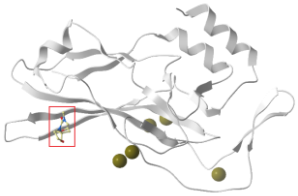
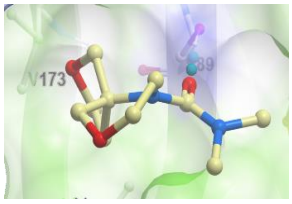


Chemical Matter

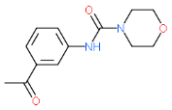
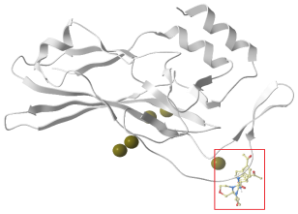
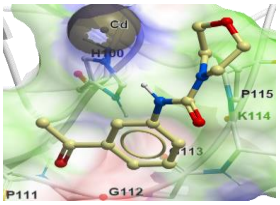
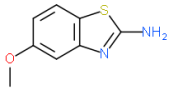
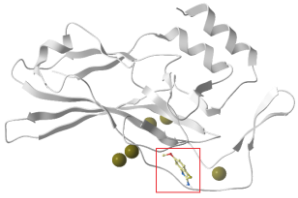
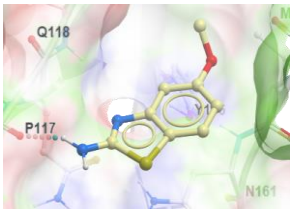
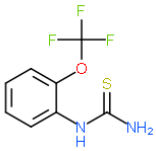
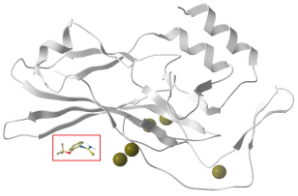
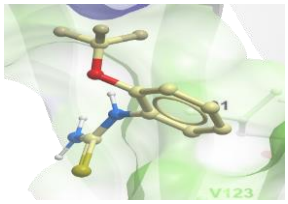
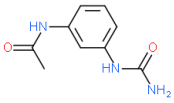
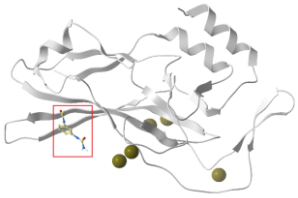
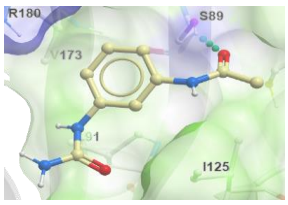
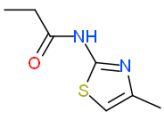
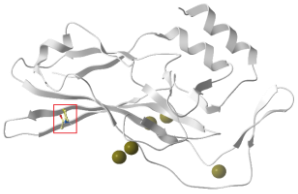
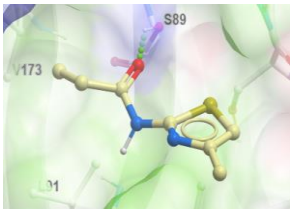
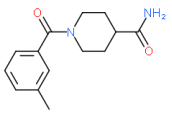
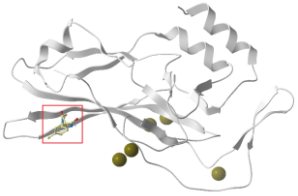
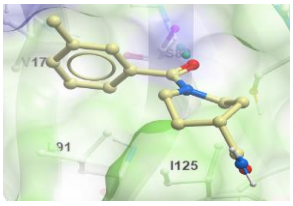
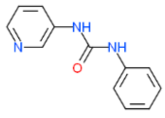
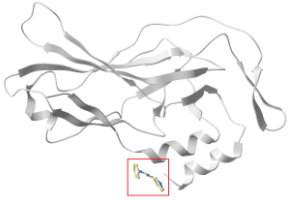
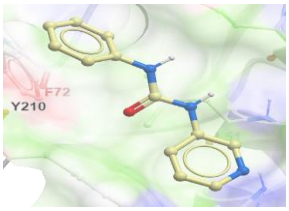
PDBID	Ligand	Binding Location	Binding Pocket	Resolution (Å)
5QRF	 F9000532	 Crystal form 1	 Pocket A	2.03
5QRG	 XS115742	 Crystal form 1	 Pocket A	1.95

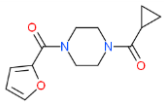
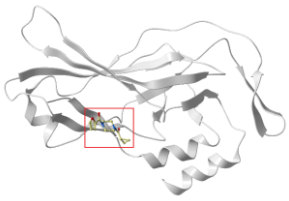
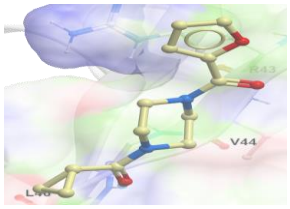
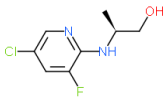
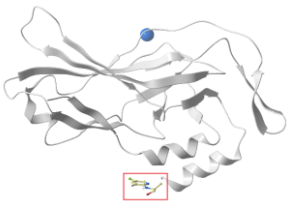
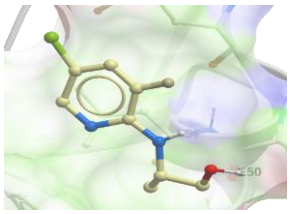
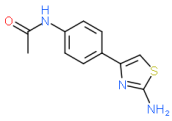
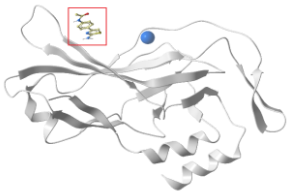
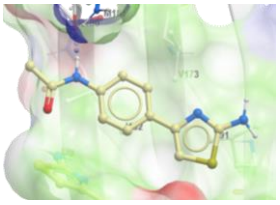
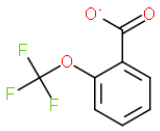
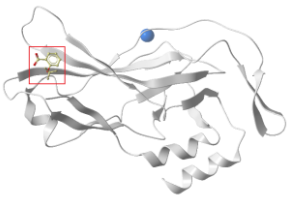
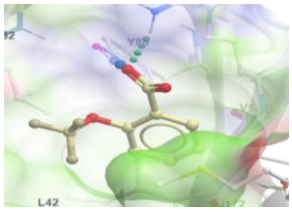
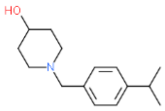
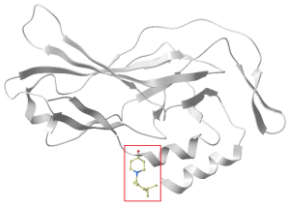
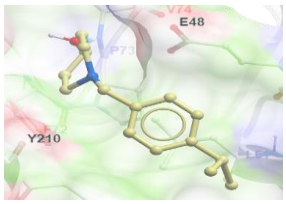
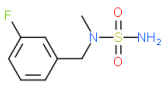
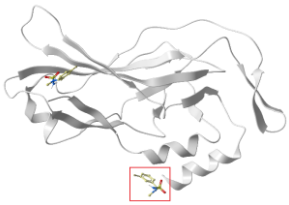
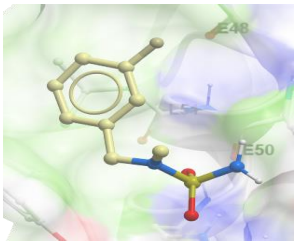
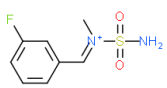
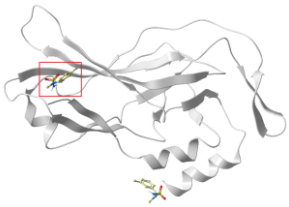
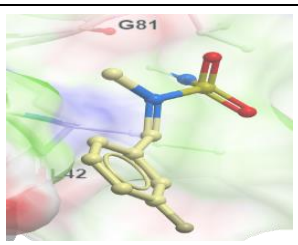
5QRH	 FM001763	 Crystal form 1	 Pocket A	1.81
5QRI	 F9000380	 Crystal form 1	 Pocket A	1.83
5QRJ	 F9000536	 Crystal form 1	 Pocket A	1.81
5QRK	 FM010104	 Crystal form 1	 Pocket A	1.63
5QRL	 F9000949	 Crystal form 1	 Pocket A	1.76
5QRM	 F9000392	 Crystal form 1	 Crystal contact	1.55
5QRN	 F9000950	 Crystal form 1	 Pocket A	1.62

5QRO	 F9000403	 Crystal form 1	 Pocket A	1.61
5QRP	 F9000414	 Crystal form 1	 Pocket A	1.67
5QRQ	 FM002138	 Crystal form 1	 Pocket A	2.1
5QRR	 F9000436	 Crystal form 1	 Pocket E	1.69
5QRS	 F9000441	 Crystal form 1	 Pocket A	2.06
5QSL	 F9000479	 Crystal form 1	 Pocket B	1.96
5QRT	 FM010020	 Crystal form 1	 Pocket A	1.77

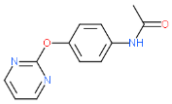
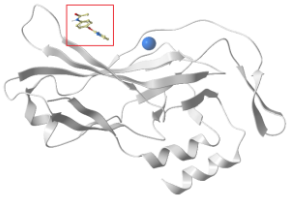
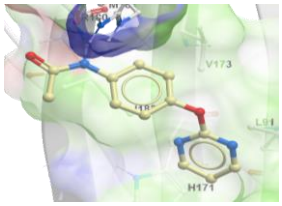
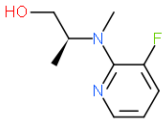
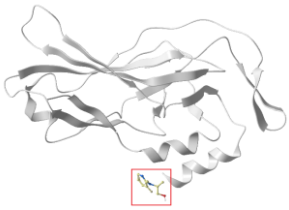
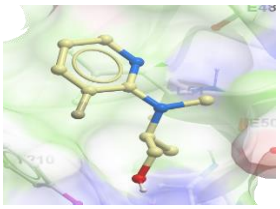
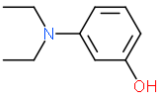
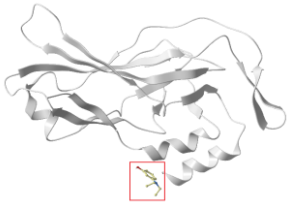
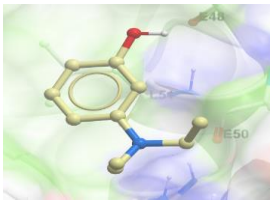
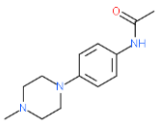
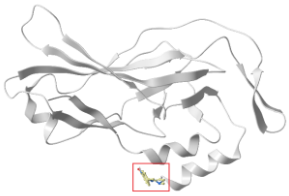
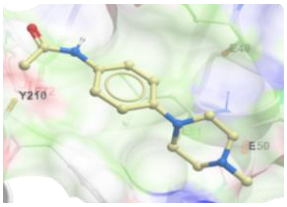
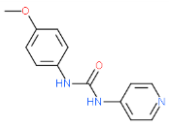
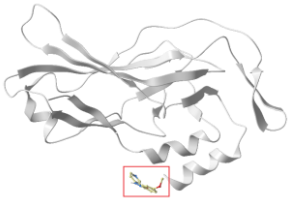
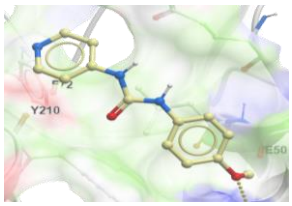
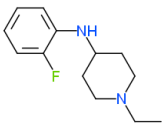
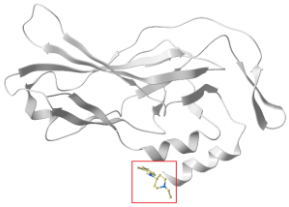
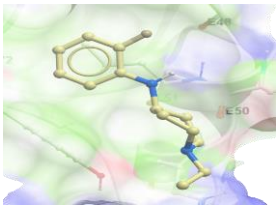
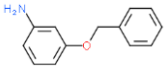
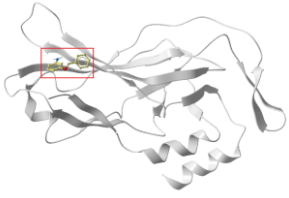
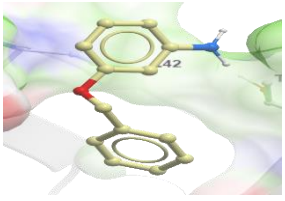
5QRU	 F9000505	 Crystal form 1	 Pocket B	1.76
5QRV	 FM002272	 Crystal form 1	 Pocket A	1.67
5QRW	 F9000547	 Crystal form 1 (Location 1)	 Pocket B	1.74
		 Crystal form 1 (Location 2)	 Pocket C	
5QRX	 F9000951	 Crystal form 1	 Pocket A	1.87
5QRY	 FM010072	 Crystal form 1	 Pocket A	1.58
5QRZ	 F9000591	 Crystal form 1	 Pocket A	1.96

For more information regarding any aspect of TEPs and the TEP programme, please contact teps@thesgc.org

5QS0	 FM010026	 Crystal form 1	 Pocket D	1.6
5QS1	 XS093338	 Crystal form 1	 Pocket B	1.66
5QS2	 FM001886	 Crystal form 1	 Pocket A	1.68
5QS3	 FM002333	 Crystal form 1	 Pocket A	1.71
5QS4	 UB000200	 Crystal form 1	 Pocket A	1.65
5QS5	 FM002032	 Crystal form 1	 Pocket A	1.81
5QS6	 FM010013	 Crystal form 2	 Pocket G	1.67

5QS7	 FM001763	 Crystal form 2	 Pocket CHECK	1.66
5QS8	 F9000511	 Crystal form 2	 Pocket G	1.47
5QS9	 FM002076	 Crystal form 2	 Pocket A’	1.43
5QSA	 FM001580	 Crystal form 2	 Pocket F	1.55
5QSB	 XS022802	 Crystal form 2	 Pocket G	1.82
5QSC	 F9000560 (location 1)	 Crystal form 2 (Location 1)	 Pocket G	1.62
	 F9000560 (location 2)	 Crystal form 2 (Location 2)	 Pocket F	

For more information regarding any aspect of TEPs and the TEP programme, please contact teps@thesgc.org

5QSD	 FM002038	 Crystal form 2	 Pocket A'	1.87
5QSE	 F9000674	 Crystal form 2	 Pocket G	2.01
5QSF	 DA000167	 Crystal form 2	 Pocket G	1.96
5QSG	 F9000710	 Crystal form 2	 Pocket G	1.87
5QT0	 XS092188	 Crystal form 2	 Pocket G	2.1
5QSH	 F9000416	 Crystal form 2	 Pocket G	1.9
5QSI	 FM001452	 Crystal form 2	 Pocket F	1.64

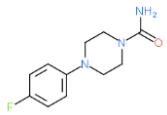
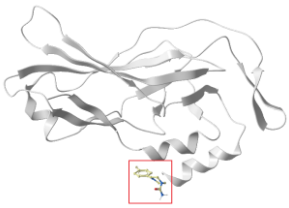
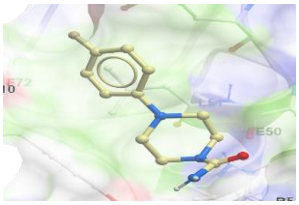
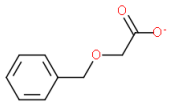
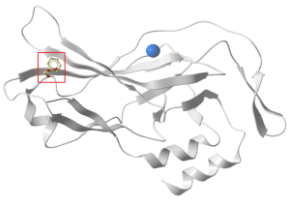
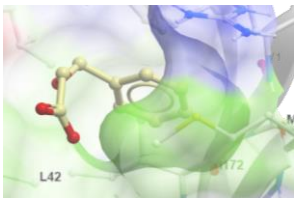
5QSJ	 FM002214	 Crystal form 2	 Pocket G	1.49
5QSK	 FM002150	 Crystal form 2	 Pocket F	1.55

Table 2: Structures and binding sites of fragment hits

IMPORTANT: Please note that the existence of small molecules within this TEP indicates only that chemical matter might bind to the protein in potentially functionally relevant locations. The small molecule ligands are intended to be used as the basis for future chemistry optimisation to increase potency and selectivity and yield a chemical probe or lead series. As such, the molecules within this TEP should not be used as tools for functional studies of the protein, unless otherwise stated, as they are not sufficiently potent or well-characterised to be used in cellular studies.

Other Components

1. The laboratory of Charles Lin (Baylor College, Houston) is generating chordoma patient-derived cell lines with endogenous TBXT tagged with GFP, HA and induced degron.
2. A widely-used polyclonal antibody (Santa Cruz) is no longer available. We are in the process of generating a nanobody in Llama.

Future Plans

In the future, we will work in collaboration with the Drewry and Lin groups to advance the fragment hits into cell-active, potent binders of TBXT.

CONCLUSION

Expression of Brachyury is a defining characteristic and a key driver of the rare cancer Chordoma. Our crystal-based fragment screen indicates several pockets within the DNA binding domain of brachyury that are promiscuous fragment binders and are close to sites of functional importance. Of particular interest are pockets A and F, the former is located close to the dimerization interface and may yield compounds that could disrupt the small interface that is formed between Brachyury subunits upon binding to palindromic DNA sites. Compounds targeting pocket F may influence the interaction of Brachyury with downstream effectors. This approach may be more tractable than attempting to directly compete with DNA binding with a small molecule.

TEP IMPACT

This TEP was generated as part of a collaborative effort funded by the Chordoma Foundation and the Mark Foundation. The co-investigators are David Drewry from UNC and Charles Lin from Baylor College. The UNC team are working on the medicinal chemistry needed to develop tight binders. The Baylor team are investigating the cell biology of TBXT in chordoma.

FUNDING INFORMATION

The work performed at the SGC has been funded by a grant from the Wellcome [106169/ZZ14/Z]. This project was further funded by the Chordoma Foundation and the Mark Foundation for Cancer research.

ADDITIONAL INFORMATION

Structure Files

PDB ID	Structure Details	Resolution (Å)
6F58	TBXT with DNA oligonucleotide	2.25
6F59	TBXT G177D variant with DNA oligonucleotide	2.15

Materials and Methods

Molecular Biology

Crystal structure 6F58
Clone Source: Michael Miley, UNC
SGC Construct ID: TBXTA-c005
Protein Region: E41-S224
Vector: pET28a
Tag: C-terminal 6HIS
Host: BL21(DE3)-R3-pRARE2

Sequence (with tag(s)):
MGELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKP
EPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNGGGQIMLNSLHKYEPRIHIVRVGGPQRMITSHCFPETQFIAV
TAYQNEEITALKIKYNPFAKAFLDAKERSHHHHHH

DNA Sequence:
CCATGGGCGAACTGCGTGTGGGTCTGGAAGAAAGCGAACTGTGGCTGCGCTTTAAAGAGCTGACCAACGAAATGAT
CGTGACCAAAAACGGCCGCCGCATGTTTCCTGTGCTGAAAGTTAACGTGAGCGGCCTGGACCCGAACGCCATGTACA
GCTTCCTGCTGGATTTCGTTGCCGCAGATAACCACCGCTGGAAGTATGTGAACGGCGAATGGGTGCCGGGTGGCAA
ACCGGAACCGCAGGCACCTAGCTGCGTGTATATCCACCCGGATAGCCCTAACTTCGGTGCCCATTGGATGAAAGCCC
CGGTGAGCTTCAGCAAGGTGAAGCTGACCAACAAACTGAACGGCGGCGGCCAAATCATGCTGAACAGCTTACATAA
ATATGAACCTCGCATCCACATTGTGCGTGTGGGCGGCCCGCAACGTATGATCACCAGCCATTGCTTCCCGGAGACAC
AGTTCATTGCCGTTACCGCCTACCAGAACGAGGAAATTACCGCACTGAAAATTAAATACAACCCTTTTGCAAAAGCCT
TCTTAGATGCCAAAGAGCGCAGCCACCACCACCATCATCACTAATGAGGATCC

Crystal structure 6F59
SGC Construct ID: TBXTA-c006
Protein Region: E41-S224
Vector: pET28a
Tag: N-6HIS;N-Thrombin;N-T7 tag;C-6HIS
Host: BL21(DE3)-R3-pRARE2

Sequence (with tag):
MGELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKP
EPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNGGGQIMLNSLHKYEPRIHIVRVGDPQRMITSHCFPETQFIAV
TAYQNEEITALKIKYNPFAKAFLDAKERSHHHHHH

DNA Sequence:
CCATGGGCGAACTGCGTGTGGGTCTGGAAGAAAGCGAACTGTGGCTGCGCTTTAAAGAGCTGACCAACGAAATGAT
CGTGACCAAAAACGGCCGCCGCATGTTTCCTGTGCTGAAAGTTAACGTGAGCGGCCTGGACCCGAACGCCATGTACA
GCTTCCTGCTGGATTTCGTTGCCGCAGATAACCACCGCTGGAAGTATGTGAACGGCGAATGGGTGCCGGGTGGCAA
ACCGGAACCGCAGGCACCTAGCTGCGTGTATATCCACCCGGATAGCCCTAACTTCGGTGCCCATTGGATGAAAGCCC
CGGTGAGCTTCAGCAAGGTGAAGCTGACCAACAAACTGAACGGCGGCGGCCAAATCATGCTGAACAGCTTACATAA
ATATGAACCTCGCATCCACATTGTGCGTGTGGGCGACCCGCAACGTATGATCACCAGCCATTGCTTCCCGGAGACAC

AGTTCATTGCCGTTACCGCCTACCAGAACGAGGAAATTACCGCACTGAAAATTAAATACAACCCTTTTGCAAAAGCCT
TCTTAGATGCCAAAGAGCGCAGCCACCACCACCATCATCACTAATGAGGATCC

DNA-binding domain, Protein-only Crystal form 1 (G177D):
SGC Construct ID: TBXTA-c020
Protein Region: E41-N211
Vector: pSUMO-LIC
Tag: N-terminal 6HIS and SUMO tags followed by a SUMO protease cleavage site
Host: BL21(DE3)-R3-pRARE2

Sequence (with tag(s)):
MCSSHHHHHHGSGSGSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLF
EGQRIADNHTPKELGMEEEDVIEVYQEQTGGGELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAM
YSFLLDFVAADNHRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNGGGQIMLNSLH
KYEPRIHIVRVGGPQRMITSHCFPETQFIAVTAYQNEEITALKIKYN

Sequence after tag cleavage:
GELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKPEP
QAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNGGGQIMLNSLHKYEPRIHIVRVGGPQRMITSHCFPETQFIAVTA
YQNEEITALKIKYN

DNA Sequence:
ATGTGCAGCAGCCATCATCATCATCATCACGGCAGCGGCAGCGGCTCTGACCAGGAGGCAAAACCTTCAACTGAGG
ACTTGGGGGATAAGAAGGAAGGTGAATATATTAAACTCAAAGTCATTGGACAGGATAGCAGTGAGATTCACTTCAA
AGTGAAAATGACAACACATCTCAAGAAACTCAAAGAATCATACTGTCAAAGACAGGGTGTTCCAATGAATTCACTCA
GGTTTCTCTTTGAGGGTCAGAGAATTGCTGATAATCATACTCCAAAAGAACTGGGAATGGAGGAAGAAGATGTGATT
GAAGTTTACCAGGAGCAAACGGGAGGTGGAGAACTGCGTGTGGGTCTGGAAGAAAGCGAACTGTGGCTGCGCTTT
AAAGAGCTGACCAACGAAATGATCGTGACCAAAAACGGCCGCCGCATGTTTCCTGTGCTGAAAGTTAACGTGAGCG
GCCTGGACCCGAACGCCATGTACAGCTTCCTGCTGGATTTCGTTGCCGCAGATAACCACCGCTGGAAGTATGTGAAC
GGCGAATGGGTGCCGGGTGGCAAACCGGAACCGCAGGCACCTAGCTGCGTGTATATCCACCCGGATAGCCCTAACT
TCGGTGCCCATTGGATGAAAGCCCCGGTGAGCTTCAGCAAGGTGAAGCTGACCAACAAACTGAACGGCGGCGGCCA
AATCATGCTGAACAGCTTACATAAATATGAACCTCGCATCCACATTGTGCGTGTGGGCGGCCCGCAACGTATGATCAC
CAGCCATTGCTTCCCGGAGACACAGTTCATTGCCGTTACCGCCTACCAGAACGAGGAAATTACCGCACTGAAAATTAA
ATACAACTGAAGCGGTGGCGGTCTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAACACCACCACCACC
ACCACTGA

DNA-binding domain, Protein-only Crystal form 2 (WT G177):
SGC Construct ID: TBXTA-c021
Protein Region: E41-N211
Vector: pSUMO-LIC
Tag: N-terminal 6HIS and SUMO tags followed by a SUMO protease cleavage site
Host: BL21(DE3)-R3-pRARE2

Sequence (with tag(s)):
MCSSHHHHHHGSGSGSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLF
EGQRIADNHTPKELGMEEEDVIEVYQEQTGGGELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAM
YSFLLDFVAADNHRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNGGGQIMLNSLH
KYEPRIHIVRVGDPQRMITSHCFPETQFIAVTAYQNEEITALKIKYN

Sequence after tag cleavage:
GELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKPEP
QAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNGGGQIMLNSLHKYEPRIHIVRVGDPQRMITSHCFPETQFIAVTA
YQNEEITALKIKYN

DNA Sequence:
ATGTGCAGCAGCCATCATCATCATCATCACGGCAGCGGCAGCGGCTCTGACCAGGAGGCAAAACCTTCAACTGAGG
ACTTGGGGGATAAGAAGGAAGGTGAATATATTAAACTCAAAGTCATTGGACAGGATAGCAGTGAGATTCACTTCAA
AGTGAAAATGACAACACATCTCAAGAAACTCAAAGAATCATACTGTCAAAGACAGGGTGTTCCAATGAATTCACTCA
GGTTTCTCTTTGAGGGTCAGAGAATTGCTGATAATCATACTCCAAAAGAACTGGGAATGGAGGAAGAAGATGTGATT
GAAGTTTACCAGGAGCAAACGGGAGGTGGAGAACTGCGTGTGGGTCTGGAAGAAAGCGAACTGTGGCTGCGCTTT
AAAGAGCTGACCAACGAAATGATCGTGACCAAAAACGGCCGCCGCATGTTTCCTGTGCTGAAAGTTAACGTGAGCG
GCCTGGACCCGAACGCCATGTACAGCTTCCTGCTGGATTTCGTTGCCGCAGATAACCACCGCTGGAAGTATGTGAAC
GGCGAATGGGTGCCGGGTGGCAAACCGGAACCGCAGGCACCTAGCTGCGTGTATATCCACCCGGATAGCCCTAACT
TCGGTGCCCATTGGATGAAAGCCCCGGTGAGCTTCAGCAAGGTGAAGCTGACCAACAAACTGAACGGCGGCGGCCA
AATCATGCTGAACAGCTTACATAAATATGAACCTCGCATCCACATTGTGCGTGTGGGCGACCCGCAACGTATGATCAC
CAGCCATTGCTTCCCGGAGACACAGTTCATTGCCGTTACCGCCTACCAGAACGAGGAAATTACCGCACTGAAAATTAA
ATACAACTGAAGCGGTGGCGGTCTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAACACCACCACCACC
ACCACTGA

Full-length, WT TBXT (G177)
SGC Construct ID: TBXTA-c026
Protein Region: S2- M435
Vector: pGTVL2
Tag: His6 and GST tag at the N-terminal fusion, followed by TEV protease cleavage site
Host: BL21(DE3)-R3-pRARE2
Note: Wild-type Full length TBXTA Codon Optimized DNA sequence

Sequence (with tag(s)):
MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAII
RYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDF
MLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGADHPPKSSSGVDLGTENLYFQ
SMSNSSPGTESAGKSLQYRVDHLLSAVENELQAGSEKGDPTERELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKV
NVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNG
GGQIMLNSLHKYEPRIHIVRVGGPQRMITSHCFPETQFIAVTAYQNEEITALKIKYNPFAKAFLDAKERSDHKEMMEEPGD
SQQPGYSQWGWLLPGTSTLCPPANPHPQFGGALSLPSTHSCDRYPTLRSHRSSPYPSPYAHRNNSPTYSDNSPACLSMLQ
SHDNWSSLGMPAHPSMLPVSHNASPPTSSSQYPSLWSVSNGAVTPGSQAAAVSNGLGAQFFRGSPAHYTPLTHPVSAP
SSSGSPLYEGAAAATDIVDSQYDAAAQGRLIASWTPVSPPSM

Sequence after tag cleavage:
SMSNSSPGTESAGKSLQYRVDHLLSAVENELQAGSEKGDPTERELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKV
NVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNG
GGQIMLNSLHKYEPRIHIVRVGGPQRMITSHCFPETQFIAVTAYQNEEITALKIKYNPFAKAFLDAKERSDHKEMMEEPGD
SQQPGYSQWGWLLPGTSTLCPPANPHPQFGGALSLPSTHSCDRYPTLRSHRSSPYPSPYAHRNNSPTYSDNSPACLSMLQ
SHDNWSSLGMPAHPSMLPVSHNASPPTSSSQYPSLWSVSNGAVTPGSQAAAVSNGLGAQFFRGSPAHYTPLTHPVSAP
SSSGSPLYEGAAAATDIVDSQYDAAAQGRLIASWTPVSPPSM

DNA Sequence:
CATATGCACCATCATCATCATCATTCTTCTATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCA
CTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGA
AACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTA
TGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATG
CTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTT
GATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGAT
CATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGT
TCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATAT
AGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGCCGACCATCCTCCAAAATCGAGCTCAGGTGTAGATC
TGGGTACCGAGAACCTGTACTTCCAATCCATGTCCAATAGCAGCCCTGGCACCGAAAGCGCCGGTAAGAGCCTACAG

TACCGTGTGGACCATTTACTGAGCGCCGTGGAGAACGAACTACAGGCAGGCAGCGAAAAAGGTGACCCGACCGAAC
GCGAACTGCGTGTGGGTCTGGAAGAAAGCGAACTGTGGCTGCGCTTTAAAGAGCTGACCAACGAAATGATCGTGAC
CAAAAACGGCCGCCGCATGTTTCCTGTGCTGAAAGTTAACGTGAGCGGCCTGGACCCGAACGCCATGTACAGCTTCC
TGCTGGATTTCGTTGCCGCAGATAACCACCGCTGGAAGTATGTGAACGGCGAATGGGTGCCGGGTGGCAAACCGGA
ACCGCAGGCACCTAGCTGCGTGTATATCCACCCGGATAGCCCTAACTTCGGTGCCCATTGGATGAAAGCCCCGGTGA
GCTTCAGCAAGGTGAAGCTGACCAACAAACTGAACGGCGGCGGCCAAATCATGCTGAACAGCTTACATAAATATGA
ACCTCGCATCCACATTGTGCGTGTGGGCGGCCCGCAACGTATGATCACCAGCCATTGCTTCCCGGAGACACAGTTCAT
TGCCGTTACCGCCTACCAGAACGAGGAAATTACCGCACTGAAAATTAAATACAACCCTTTTGCAAAAGCCTTCTTAGA
TGCCAAAGAGCGCAGCGACCACAAAGAAATGATGGAGGAACCGGGCGACAGCCAGCAGCCTGGTTATAGCCAATG
GGGCTGGCTGCTGCCGGGTACCAGTACACTGTGCCCGCCGGCAAACCCGCATCCTCAGTTTGGTGGCGCCCTGAGCT
TACCGAGCACCCACAGTTGTGATCGCTATCCTACCCTGCGTAGCCATCGTAGCAGTCCGTACCCGAGCCCGTATGCAC
ATCGTAACAACAGCCCGACATATAGCGACAACAGCCCGGCCTGTCTGAGCATGCTACAGAGCCATGATAACTGGAGC
AGCCTGGGCATGCCGGCACATCCGAGCATGCTGCCGGTGAGTCACAATGCCAGCCCGCCGACCAGTAGCAGCCAGT
ATCCTAGCCTGTGGAGCGTTAGTAACGGTGCAGTGACCCCGGGTAGCCAGGCAGCAGCAGTTAGCAATGGTCTGGG
CGCCCAGTTTTTTCGCGGCAGTCCGGCACATTACACCCCGCTGACCCATCCTGTGAGTGCCCCGAGCAGTAGTGGTAG
CCCGCTGTATGAAGGTGCCGCAGCCGCCACCGACATTGTTGACAGCCAGTATGATGCCGCCGCACAGGGTCGCCTGA
TTGCAAGCTGGACCCCTGTGAGTCCGCCGAGCATGTGACAGTAAAGGTGGATACGGATCCGAATTCGAGCTCCGTCG
ACAAGCTTGCGGCCGCACTCGAG

Full-length, G177D variant TBXT

SGC Construct ID: TBXTA-c025
Protein Region: S2- M435
Vector: pGTVL2
Tag: His6 and GST tag at the N-terminal fusion, followed by TEV protease cleavage site
Host: BL21(DE3)-R3-pRARE2
Note: G177D Full length TBXTA Codon Optimized DNA sequence

Sequence (with tag(s)):
MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAII
RYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDF
MLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGADHPPKSSSGVDLGTENLYFQ
SMSNSSPGTESAGKSLQYRVDHLLSAVENELQAGSEKGDPTERELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKV
NVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNG
GGQIMLNSLHKYEPRIHIVRVGDPQRMITSHCFPETQFIAVTAYQNEEITALKIKYNPFAKAFLDAKERSDHKEMMEEPGD
SQQPGYSQWGWLLPGTSTLCPPANPHPQFGGALSLPSTHSCDRYPTLRSHRSSPYPSPYAHRNNSPTYSDNSPACLSMLQ
SHDNWSSLGMPAHPSMLPVSHNASPPTSSSQYPSLWSVSNGAVTPGSQAAAVSNGLGAQFFRGSPAHYTPLTHPVSAP
SSSGSPLYEGAAAATDIVDSQYDAAAQGRLIASWTPVSPPSM

Sequence after tag cleavage:
SMSNSSPGTESAGKSLQYRVDHLLSAVENELQAGSEKGDPTERELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKV
NVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNG
GGQIMLNSLHKYEPRIHIVRVGDPQRMITSHCFPETQFIAVTAYQNEEITALKIKYNPFAKAFLDAKERSDHKEMMEEPGD
SQQPGYSQWGWLLPGTSTLCPPANPHPQFGGALSLPSTHSCDRYPTLRSHRSSPYPSPYAHRNNSPTYSDNSPACLSMLQ
SHDNWSSLGMPAHPSMLPVSHNASPPTSSSQYPSLWSVSNGAVTPGSQAAAVSNGLGAQFFRGSPAHYTPLTHPVSAP
SSSGSPLYEGAAAATDIVDSQYDAAAQGRLIASWTPVSPPSM

DNA Sequence:
CATATGCACCATCATCATCATCATTCTTCTATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCA
CTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGA
AACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTA
TGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATG
CTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTT
GATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGAT

CATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGT
TCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATAT
AGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGCCGACCATCCTCCAAAATCGAGCTCAGGTGTAGATC
TGGGTACCGAGAACCTGTACTTCCAATCCATGTCCAATAGCAGCCCTGGCACCGAAAGCGCCGGTAAGAGCCTACAG
TACCGTGTGGACCATTTACTGAGCGCCGTGGAGAACGAACTACAGGCAGGCAGCGAAAAAGGTGACCCGACCGAAC
GCGAACTGCGTGTGGGTCTGGAAGAAAGCGAACTGTGGCTGCGCTTTAAAGAGCTGACCAACGAAATGATCGTGAC
CAAAAACGGCCGCCGCATGTTTCCTGTGCTGAAAGTTAACGTGAGCGGCCTGGACCCGAACGCCATGTACAGCTTCC
TGCTGGATTTCGTTGCCGCAGATAACCACCGCTGGAAGTATGTGAACGGCGAATGGGTGCCGGGTGGCAAACCGGA
ACCGCAGGCACCTAGCTGCGTGTATATCCACCCGGATAGCCCTAACTTCGGTGCCCATTGGATGAAAGCCCCGGTGA
GCTTCAGCAAGGTGAAGCTGACCAACAAACTGAACGGCGGCGGCCAAATCATGCTGAACAGCTTACATAAATATGA
ACCTCGCATCCACATTGTGCGTGTGGGCGACCCGCAACGTATGATCACCAGCCATTGCTTCCCGGAGACACAGTTCAT
TGCCGTTACCGCCTACCAGAACGAGGAAATTACCGCACTGAAAATTAAATACAACCCTTTTGCAAAAGCCTTCTTAGA
TGCCAAAGAGCGCAGCGACCACAAAGAAATGATGGAGGAACCGGGCGACAGCCAGCAGCCTGGTTATAGCCAATG
GGGCTGGCTGCTGCCGGGTACCAGTACACTGTGCCCGCCGGCAAACCCGCATCCTCAGTTTGGTGGCGCCCTGAGCT
TACCGAGCACCCACAGTTGTGATCGCTATCCTACCCTGCGTAGCCATCGTAGCAGTCCGTACCCGAGCCCGTATGCAC
ATCGTAACAACAGCCCGACATATAGCGACAACAGCCCGGCCTGTCTGAGCATGCTACAGAGCCATGATAACTGGAGC
AGCCTGGGCATGCCGGCACATCCGAGCATGCTGCCGGTGAGTCACAATGCCAGCCCGCCGACCAGTAGCAGCCAGT
ATCCTAGCCTGTGGAGCGTTAGTAACGGTGCAGTGACCCCGGGTAGCCAGGCAGCAGCAGTTAGCAATGGTCTGGG
CGCCCAGTTTTTTCGCGGCAGTCCGGCACATTACACCCCGCTGACCCATCCTGTGAGTGCCCCGAGCAGTAGTGGTAG
CCCGCTGTATGAAGGTGCCGCAGCCGCCACCGACATTGTTGACAGCCAGTATGATGCCGCCGCACAGGGTCGCCTGA
TTGCAAGCTGGACCCCTGTGAGTCCGCCGAGCATGTGACAGTAAAGGTGGATACGGATCCGAATTCGAGCTCCGTCG
ACAAGCTTGCGGCCGCACTCGAG

Plasmid depositions in Addgene

construct ID	Domain	variant	Vector	tags	Crystallized	Boundaries	Addgene
TBXTA-c005	DBD	WT	pET28a	C-His6	Yes/6F58	E41-S224	139754
TBXTA-c006	DBD	G177D	pET28a	C-His6	Yes/6F59	E41-S224	139755
TBXTA-c019	DBD	WT	pSUMO-LIC	N-His-SUMO		E41-A214	139756
TBXTA-c020	DBD	WT	pSUMO-LIC	N-His-SUMO	Yes/ Fragment screening	E41-N211	139757
TBXTA-c021	DBD	G177D	pSUMO-LIC	N-His-SUMO	Yes/ Fragment screening	E41-N211	139758
TBXTA-c022	DBD	G177D	pSUMO-LIC	N-His-SUMO		E41-A214	139759
TBXTA-c023	DBD	WT	pGTVL2	N-His-GST-TEV		E41-D225	139760
TBXTA-c024	DBD	G177D	pGTVL2	N-His-GST-TEV		E41-D225	139761
TBXTA-c025	FL	G177D	pGTVL2	N-His-GST-TEV		S2- M435	139762
TBXTA-c026	FL	WT	pGTVL2	N-His-GST-TEV		S2- M435	139763
TBXTA-c027	FL	G177D	pHGT-Bio	N-His-GST-TEV, C-Bio		S2- M435	139764
TBXTA-c028	FL	WT	pHGT-Bio	N-His-GST-TEV, C-Bio		S2- M435	139765
TBXTA-c043	DBD	WT	pNIC-Bio3	N-His-TEV, C-Bio		E41- D225	139766

TBXTA-c044	DBD	G177D	pNIC-Bio3	N-His-TEV, C-Bio		E41- D225	139767

Protein Expression and Purification
Medium: Terrific Broth (TB) Merck with 4 ml of glycerol
Antibiotics: Kanamycin, 50 µg/ml

From the glycerol stock, bacteria were inoculated in 15 ml of 1 x TB in a 50 ml tube with Kanamycin 0.05 mg/ml and 0.034 mg/ml of chloramphenicol and grown overnight in a shaker at 37°C, 250rpm. The following day, 4 ml of the overnight culture were inoculated in 1L of TB. The bacteria grew in an incubator at 37°C, shaking 180 rpm. Once the OD reached 2-3, IPTG (300uM) was added to the media and left overnight at 18°C, shaking 180 rpm. The pellets were harvested the next following day.

Protein Purification: DNA-binding domain constructs
The pellet was re-suspended in 250 ml of Lysis Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 5% Glycerol and 1 mM TCEP). The cells, on ice, were sonicated for 20 minutes with 5 seconds pulse ON and 10 seconds pulse OFF with 35% of amplitude and centrifuged for 25 minutes at 66700 x g. The supernatant was incubated for an hour at 4°C, with Nickel beads pre-washed with Lysis buffer. After one hour of batch-binding, the tubes containing the lysate centrifuged at 700 x g at 4°C for 5 minutes and the supernatant discarded. This step was repeated twice with, respectively, 100 ml and 50 ml of Lysis Buffer.

Beads were loaded on a gravity column with 20 ml of Wash Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM Imidazole, 5% Glycerol and 1 mM TCEP) and, followed by two elution of 10 ml each with Elution Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 300 mM Imidazole, 5% Glycerol and 1 mM TCEP). After an SDS-PAGE gel, the elution containing the protein was concentrated with an Amicon 10kDa concentrator and loaded on a Hi Load 16/600 Superdex 75 pg column at 1 ml/min, collecting 2-ml fractions. The fractions containing the protein were pooled together and concentrated with an Amicon 10kDa concentrator to 10 mg/ml was reached. Protein aliquots were flash frozen in Liquid Nitrogen and stored at -80°C. The protein was confirmed by ESI-TOF intact mass spectrometry:

6F58: Predicted: 21987; observed: 21856 – corresponding to expected mass with loss of N-terminal methionine
6F59: Predicted: 22045; observed: 21914 – corresponding to expected mass with loss of N-terminal methionine.

Protein Purification for DNA-free Crystal forms 1 and 2
The pellet was re-suspended in 250 ml of Lysis Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 5% Glycerol and 1 mM TCEP). The cells, on ice, were sonicated for 20 minutes with 5 seconds pulse ON and 10 seconds pulse OFF with 35% of amplitude and centrifuged for 25 minutes at 66700 x g. The supernatant was incubated for an hour at 4°C, with Nickel beads pre-washed with Lysis buffer. After one hour of batch-binding, the tubes containing the lysate centrifuged at 700 x g at 4°C for 5 minutes and the supernatant discarded. This step was repeated twice with, respectively, 100 ml and 50 ml of Lysis Buffer.

Beads were loaded on a gravity column with 20 ml of Wash Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM Imidazole, 5% Glycerol and 1 mM TCEP) and, followed by two elution of 10 ml each with Elution Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 300 mM Imidazole, 5% Glycerol and 1 mM TCEP). Fractions containing TBXT were pooled and SUMO protease was added to a final mass ratio 1:150. Cleavage was performed overnight during dialysis into dialysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 % Glycerol, 1 mM TCEP) using 3500 MWCO snakeskin dialysis tubing.

After dialysis the protein was concentrated with an Amicon 10kDa concentrator and loaded on a Hi Load 16/600 Superdex 75 pg column. The flow-rate of the gel filtration was 1 ml/min and the volume of the fractions collected was 2 ml. The fractions containing the protein were pulled together and concentrated with an Amicon 10kDa concentrator, until the concentration 12 mg/ml was reached. Protein aliquots were stored at -80°C after being flash frozen in Liquid Nitrogen. The protein was confirmed by ESI-TOF intact mass spectrometry

Crystal form 1, TBXTA-c020: Predicted: 19571.7; observed: 19571.8
Crystal form 2, TBXTA-c021: Predicted: 19629.7; observed: 19629.8

Protein Purification: Full-length protein constructs TBXTA-c025, TBXTA-c026
The cell pellet was re-suspended in 250 ml of Lysis Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 5% Glycerol and 1 mM TCEP). The cells were sonicated on ice for 20 minutes with 5 seconds pulse ON and 10 seconds pulse OFF with 35% of amplitude and centrifuged for 25 minutes at 66700 x g. The supernatant was incubated for an hour at 4°C with Nickel-sepharose beads pre-washed with Lysis buffer. After one hour of batch-binding, the beads were collected by centrifugation (700 x g at 4°C for 5 minutes) and the supernatant was discarded. The beads were washed with 100 ml and 50 ml of Lysis Buffer.

The beads were then transferred to a gravity column with 20 ml of Wash Buffer 1 (50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM Imidazole, 5% Glycerol and 1 mM TCEP). The beads were washed with 20 ml of a Wash Buffer 2 (50 mM HEPES pH 7.5, 500 mM NaCl, 60 mM Imidazole, 5% Glycerol and 1 mM TCEP). Washes were followed by 4 elution of 10 ml each with Elution Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol and 1 mM TCEP) with increasing concentrations of Imidazole: 100 mM, 200 mM Imidazole, 300 mM and 500 mM Imidazole.

The 4 elutions were dialysed overnight in Dialysis Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol and 1 mM TCEP) with 0.5 mg of TEV for 20 mg of protein in a Cold Room.

The beads were washed with 20 ml of Lysis Buffer. The dialysed elution was loaded on the beads and washed with 20 ml of Lysis Buffer. The imidazole gradient was repeated 10 ml of Wash 1, 10 ml of Wash 2, followed by the same 10 ml elutions prior to the dialysis (100 mM Imidazole, 200 mM Imidazole, 300 mM Imidazole and 500 mM Imidazole). The fractions were analysed by SDS-PAGE; the cleaved protein was found in wash1 and wash 2.

The combined fractions were concentrated with an Amicon 30kDa concentrator and loaded on a Hi Load 16/600 Superdex 200 pg column with Gel Filtration Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol and 1 mM TCEP) at 1.2 ml/min, collecting 2 ml fractions. The fractions containing the protein were pooled together and concentrated with an Amicon 30kDa concentrator, to 10 mg/ml. Protein aliquots were stored at -80°C after being flash frozen in Liquid Nitrogen. The protein was confirmed by ESI-TOF intact mass spectrometry

TBXTA-c025: Predicted: 47789; observed: 47790.
TBXTA-c026: Predicted: 47731; observed: 47732

Structure Determination

6F58:

Crystallisation: A self-complementary DNA oligonucleotide 5'- AATTTCACACCTAGGTGTGAAATT was dissolved to 1mM, heated to 95°C on a heat block and allowed to cool slowly over 2hrs. The protein and DNA were mixed in a 1:1.1 molar ratio (assuming a duplex DNA molecule) and sitting drop vapour diffusion crystallisation trials were set up with a Mosquito (TTP Labtech) crystallisation robot at a final concentration of 6.6 mg/ml. TBXT crystallised at 4°C in conditions containing 40% PEG300 -- 0.1M citrate pH 4.2. Crystals were loop mounted and cryo-cooled by plunging directly into liquid nitrogen.

Data Collection: Data were collected to 2.2Å resolution at Diamond light source beamline I04-1 and processed using XDS

Data Processing: The structure was solved by molecular replacement using the program PHASER and the structure of *Xenopus laevis* brachyury (1XBR) as a search model. Refinement was performed using PHENIX REFINE to a final Rfactor = 24.2%, Rfree = 28.8%.

6F59 (G177D variant):
Crystallisation: A self-complementary DNA oligonucleotide 5'- GAATTTCACACCTAGGTGTGAAATTC was dissolved to 1mM, heated to 95°C on a heat block and allowed to cool slowly over 2hrs. The protein and DNA were mixed in a 1:1.1 molar ratio (assuming a duplex DNA molecule) and sitting drop vapor diffusion crystallsation trials were set up with a Mosquito (TTP Labtech) crystallization robot at a final concentration of 8 mg/ml. TBXT crystallised at 4°C in conditions containing 56% MPD, 0.1 M SPG pH 6.0. Crystals were loop mounted and cryo-cooled by plunging directly into liquid nitrogen.

Data Collection: Data were collected to 2.1Å resolution at Diamond light source beamline I04-1 and processed using DIALS.

Data Processing: The structure was solved by molecular replacement using the program PHASER and the structure of *Xenopus laevis* brachyury (1XBR) as a search model. Refinement was performed using PHENIX REFINE to a final Rfactor = 22.1%, Rfree = 25.2%.

Crystal form 1 for fragment screens (G177D, no DNA)
Crystallisation: The protein was adjusted to 7.5 mg/ml and sitting drop vapor diffusion crystallisation trials were set up with a Mosquito (TTP Labtech) crystallisation. TBXT crystallised at 4°C in conditions containing 32% PEG400, 0.1M acetate pH 4.5, 0.1 M cadmium chloride. Crystals were loop mounted and cryo-cooled by plunging directly into liquid nitrogen.

Fragment soaking: Fragments from the DSI-Poised library were added to the crystallisation drops by acoustic dispensing using an ECHO acoustic liquid handler from a 500 mM stock concentration dissolved in DMSO to a final concentration of 10%. Soaking times varied from 1.5 to 4 hours.

Data Collection: Data were collected at Diamond light source beamline I04-1 and processed using the XChem Explorer pipeline.

Data Processing: Structures were solved by difference Fourier synthesis using the XChem Explorer pipeline. Fragment hits were identified using the PanDDA program. Refinement was performed using REFMAC.

Crystal form 2 for fragment screens (WT, no DNA)
Crystallisation: The protein was adjusted to 16 mg/ml and sitting drop vapor diffusion crystallisation trials were set up with a Mosquito (TTP Labtech) crystallisation robot. TBXT crystallised at 4°C in conditions containing 30% PEG1000, 0.1M SPG pH 7.0. Crystals were loop mounted and cryo-cooled by plunging directly into liquid nitrogen.

Fragment soaking: Fragments from the DSI-Poised library were added to the crystallisation drops by acoustic dispensing using an ECHO acoustic liquid handler from a 500 mM stock concentration dissolved in DMSO to a final concentration of 10%. Soaking times varied from 1.5 to 4 hours.

Data Collection: Data were collected at Diamond light source beamline I04-1 and processed using the XChem Explorer pipeline.

Data Processing: Structures were solved by difference Fourier synthesis using the XChem Explorer pipeline. Fragment hits were identified using the PanDDA program. Refinement was performed using REFMAC.

Assays

Electrophoretic mobility shift assays (EMSA)

The following oligonucleotide pair was used:

TA50-F CATGCATGCAGGGAATTTCACACCTAGGTGTGAAATTCCCATTCGTGCGA

TA50-R TCGCACGAATGGGAATTTCACACCTAGGTGTGAAATTCCCTGCATGCATG

To reduce the formation of hairpin structures, the oligos were annealed at a high concentration (200 µM each) in 10 mM tris-HCl, pH 7.5, 50 mM NaCl by heating to 95°C in a dry block and leaving to cool to room temperature. The dsDNA was subsequently labelled with T4 polynucleotide kinase (NEB) and γ-^{32}P-ATP. The labelled DNA was separated from the remaining ATP/ADP using a BioRad MicroBiospin P-6 column equilibrated in annealing buffer.

EMSA buffer was: 25 mM HEPES, pH 7.4, 10% glycerol, 75 mM NaCl, 0.1% tween 20, 1 mM TCEP.

Protein was diluted serially in this buffer and mixed with 1-5 nM of DNA diluted in the same buffer. After 10-minute incubation on ice, the samples (5 µl) were loaded on a pre-run 8% polyacrylamide gel (40:1 acrylamide/bis) in chilled TAE buffer (40 mM TRIS base, 20 mM acetic acid, 1 mM EDTA). The gel tanks were placed in an ice bucket and run for 75 minutes at 150V. The dried gels were exposed overnight using a BioRad phosphorimager screen.

Ligand binding by Surface Plasmon Resonance (SPR)

Machine

Biacore S200

Sensor surface

Series S SA sensor, Immobilized with ~1500 RU of biotinylated G177D TBXTA full length protein (immobilization at approximately 20 nM for 100 seconds)

Running buffer

10 mM Hepes pH 7.5, 150 mM NaCl, 1 mM DTT, 1% DMSO.

Protein:

Protein sequence

SMSNSSPGTESAGKSLQYRVDHLLSAVENELQAGSEKGDPTERELRVGLEESELWLRFKELTNEMIVTKNGRRMFPVLKV
NVSGLDPNAMYSFLLDFVAADNHRWKYVNGEWVPGGKPEPQAPSCVYIHPDSPNFGAHWMKAPVSFSKVKLTNKLNG
GGQIMLNSLHKYEPRIHIVRVGDPQRMITSHCFPETQFIAVTAYQNEEITALKIKYNPFAKAFLDAKERSDHKEMMEEPGD
SQQPGYSQWGWLLPGTSTLCPPANPHPQFGGALSLPSTHSCDRYPTLRSHRSSPYPSPYAHRNNSPTYSDNSPACLSMLQ
SHDNWSSLGMPAHPSMLPVSHNASPPTSSSQYPSLWSVSNGAVTPGSQAAAVSNGLGAQFFRGSPAHYTPLTHPVSAP
SSSGSPLYEGAAAATDIVDSQYDAAAQGRLIASWTPVSPPSMSSKGGY**GLNDIFEAQKIEWHE***

*AviTag biotin conjugation sequence

The protein was expressed from construct TBXTA-c027, encoding a C-terminal Avitag which is biotinylated in E. coli when co-expressed with BirA enzyme in the presence of biotin. (Keates et al. PMID: 22027370**).** The protein was purified as described for TBXTA-c026.

References

1. Vujovic, S., Henderson, S., Presneau, N., Odell, E., Jacques, T. S., Tirabosco, R., Boshoff, C., and Flanagan, A. M. (2006) Brachyury, a crucial regulator of notochordal development, is a novel biomarker for chordomas. *J Pathol* **209**, 157-165
2. Presneau, N., Shalaby, A., Ye, H., Pillay, N., Halai, D., Idowu, B., Tirabosco, R., Whitwell, D., Jacques, T. S., Kindblom, L. G., Bruderlein, S., Moller, P., Leithner, A., Liegl, B., Amary, F. M., Athanasou, N. N.,

Hogendoorn, P. C., Mertens, F., Szuhai, K., and Flanagan, A. M. (2011) Role of the transcription factor T (brachyury) in the pathogenesis of sporadic chordoma: a genetic and functional-based study. *J Pathol* **223**, 327-335
3. Sa, J. K., Lee, I. H., Hong, S. D., Kong, D. S., and Nam, D. H. (2017) Genomic and transcriptomic characterization of skull base chordoma. *Oncotarget* **8**, 1321-1328
4. Tarpey, P. S., Behjati, S., Young, M. D., Martincorena, I., Alexandrov, L. B., Farndon, S. J., Guzzo, C., Hardy, C., Latimer, C., Butler, A. P., Teague, J. W., Shlien, A., Futreal, P. A., Shah, S., Bashashati, A., Jamshidi, F., Nielsen, T. O., Huntsman, D., Baumhoer, D., Brandner, S., Wunder, J., Dickson, B., Cogswell, P., Sommer, J., Phillips, J. J., Amary, M. F., Tirabosco, R., Pillay, N., Yip, S., Stratton, M. R., Flanagan, A. M., and Campbell, P. J. (2017) The driver landscape of sporadic chordoma. *Nat Commun* **8**, 890
5. Wansleben, S., Peres, J., Hare, S., Goding, C. R., and Prince, S. (2014) T-box transcription factors in cancer biology. *Biochim Biophys Acta* **1846**, 380-391
6. Yang, X. R., Ng, D., Alcorta, D. A., Liebsch, N. J., Sheridan, E., Li, S., Goldstein, A. M., Parry, D. M., and Kelley, M. J. (2009) T (brachyury) gene duplication confers major susceptibility to familial chordoma. *Nat Genet* **41**, 1176-1178
7. Sharifnia, T., Wawer, M. J., Chen, T., Huang, Q. Y., Weir, B. A., Sizemore, A., Lawlor, M. A., Goodale, A., Cowley, G. S., Vazquez, F., Ott, C. J., Francis, J. M., Sassi, S., Cogswell, P., Sheppard, H. E., Zhang, T., Gray, N. S., Clarke, P. A., Blagg, J., Workman, P., Sommer, J., Hornicek, F., Root, D. E., Hahn, W. C., Bradner, J. E., Wong, K. K., Clemons, P. A., Lin, C. Y., Kotz, J. D., and Schreiber, S. L. (2019) Small-molecule targeting of brachyury transcription factor addiction in chordoma. *Nat Med* **25**, 292-300
8. Sakamoto, K. M., Kim, K. B., Kumagai, A., Mercurio, F., Crews, C. M., and Deshaies, R. J. (2001) Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc Natl Acad Sci U S A* **98**, 8554-8559
9. Kispert, A., and Herrmann, B. G. (1993) The Brachyury gene encodes a novel DNA binding protein. *EMBO J* **12**, 3211-3220
10. Lolas, M., Valenzuela, P. D., Tjian, R., and Liu, Z. (2014) Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. *Proc Natl Acad Sci U S A* **111**, 4478-4483
11. Nelson, A. C., Pillay, N., Henderson, S., Presneau, N., Tirabosco, R., Halai, D., Berisha, F., Flicek, P., Stemple, D. L., Stern, C. D., Wardle, F. C., and Flanagan, A. M. (2012) An integrated functional genomics approach identifies the regulatory network directed by brachyury (T) in chordoma. *J Pathol* **228**, 274-285
12. Pillay, N., Plagnol, V., Tarpey, P. S., Lobo, S. B., Presneau, N., Szuhai, K., Halai, D., Berisha, F., Cannon, S. R., Mead, S., Kasperaviciute, D., Palmen, J., Talmud, P. J., Kindblom, L. G., Amary, M. F., Tirabosco, R., and Flanagan, A. M. (2012) A common single-nucleotide variant in T is strongly associated with chordoma. *Nat Genet* **44**, 1185-1187
13. Heery, C. R. (2016) Chordoma: The Quest for Better Treatment Options. *Oncol Ther* **4**, 35-51
14. Papaioannou, V. E. (2014) The T-box gene family: emerging roles in development, stem cells and cancer. *Development* **141**, 3819-3833
15. Aleksic, T., Browning, L., Woodward, M., Phillips, R., Page, S., Henderson, S., Athanasou, N., Ansorge, O., Whitwell, D., Pratap, S., Hassan, A. B., Middleton, M. R., and Macaulay, V. M. (2016) Durable Response of Spinal Chordoma to Combined Inhibition of IGF-1R and EGFR. *Front Oncol* **6**, 98
16. Beisaw, A., Tsaytler, P., Koch, F., Schmitz, S. U., Melissari, M. T., Senft, A. D., Wittler, L., Pennimpede, T., Macura, K., Herrmann, B. G., and Grote, P. (2018) BRACHYURY directs histone acetylation to target loci during mesoderm development. *EMBO Rep* **19**, 118-134
17. Xia, M., Huang, R., Sakamuru, S., Alcorta, D., Cho, M. H., Lee, D. H., Park, D. M., Kelley, M. J., Sommer, J., and Austin, C. P. (2013) Identification of repurposed small molecule drugs for chordoma therapy. *Cancer Biol Ther* **14**, 638-647
18. Janesick, A., Shiotsugu, J., Taketani, M., and Blumberg, B. (2012) RIPPLY3 is a retinoic acid-inducible repressor required for setting the borders of the pre-placodal ectoderm. *Development* **139**, 1213-1224

We respectfully request that this document is cited using the DOI value as given above if the content is used in your work.