

# Anthrax toxin lethal factor domain 3 is highly mobile and responsive to ligand binding

Kimberly M. Maize,<sup>‡</sup> Elbek K. Kurbanov,<sup>‡</sup> Teresa De La Mora-Rey, Todd W. Geders, Dong-Jin Hwang, Michael A. Walters, Rodney L. Johnson, Elizabeth A. Amin\* and Barry C. Finzel\*

Department of Medicinal Chemistry and Minnesota Supercomputing Institute, University of Minnesota, 8-101 Weaver-Densford Hall, 308 Harvard Street SE, Minneapolis, MN 55455, USA

<sup>‡</sup> These authors contributed equally to this work.

Correspondence e-mail: eamin@umn.edu, finze007@umn.edu

Received 11 June 2014  
Accepted 7 August 2014

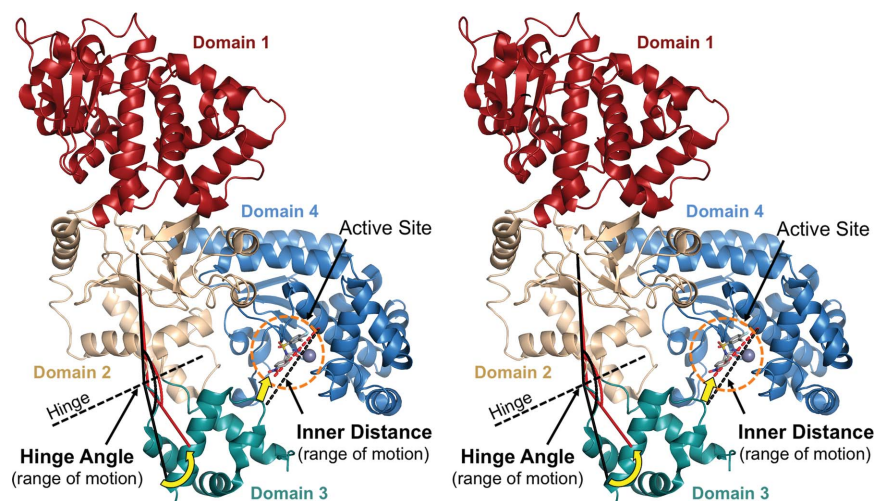
**PDB references:** anthrax toxin lethal factor, 4pkt; 4pkv; 4pkr; 4pkq; 4pku; 4pks; 4pkw

The secreted anthrax toxin consists of three components: the protective antigen (PA), edema factor (EF) and lethal factor (LF). LF, a zinc metalloproteinase, compromises the host immune system primarily by targeting mitogen-activated protein kinase kinases in macrophages. Peptide substrates and small-molecule inhibitors bind LF in the space between domains 3 and 4 of the hydrolase. Domain 3 is attached on a hinge to domain 2 *via* residues Ile300 and Pro385, and can move through an angular arc of greater than 35° in response to the binding of different ligands. Here, multiple LF structures including five new complexes with co-crystallized inhibitors are compared and three frequently populated LF conformational states termed 'bioactive', 'open' and 'tight' are identified. The bioactive position is observed with large substrate peptides and leaves all peptide-recognition subsites open and accessible. The tight state is seen in unliganded and small-molecule complex structures. In this state, domain 3 is clamped over certain substrate subsites, blocking access. The open position appears to be an intermediate state between these extremes and is observed owing to steric constraints imposed by specific bound ligands. The tight conformation may be the lowest-energy conformation among the reported structures, as it is the position observed with no bound ligand, while the open and bioactive conformations are likely to be ligand-induced.

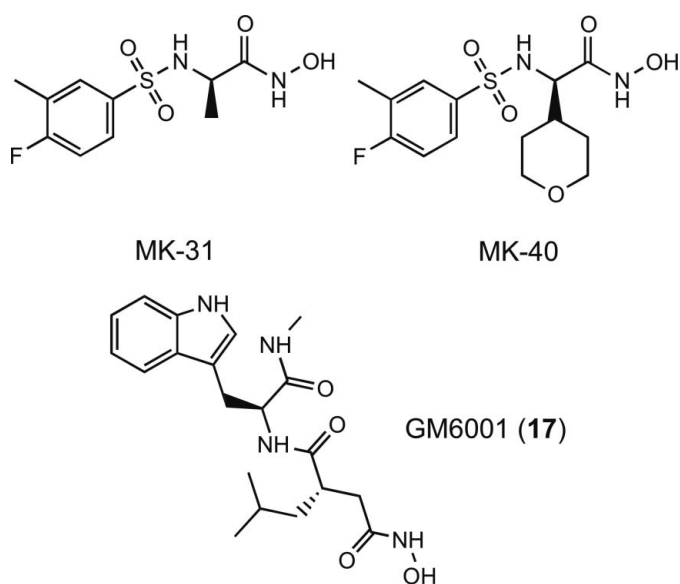
## 1. Introduction

Anthrax toxin lethal factor (LF) is one element of a tripartite exotoxin produced by *Bacillus anthracis*. Together with protective antigen (PA) and edema factor (EF), this toxin reportedly compromises host immune responses, although the exact mechanism is not fully understood (Erwin *et al.*, 2001; Collier & Young, 2003; Kim *et al.*, 2003; Moayeri *et al.*, 2003; Moayeri & Leppla, 2004). From the perspective of biodefense, the comprehensive characterization and development of toxin inhibitors such as those of LF are of significant interest, as the secreted toxins may cause cytotoxicity even after the bacterial infection has been resolved by antibiotics such as ciprofloxacin and/or doxycycline (Dixon *et al.*, 1999; Hughes & Gerberding, 2002). An unmet medical need for treatment options has led to a number of efforts to identify effective LF inhibitors (Pannifer *et al.*, 2001; Panchal *et al.*, 2004; Turk *et al.*, 2004; Forino *et al.*, 2005; Shoop *et al.*, 2005; Xiong *et al.*, 2006; Chiu *et al.*, 2009; Jiao *et al.*, 2012).

The LF protein is a 94 kDa zinc metalloproteinase that cleaves the N-terminal proline-rich portion of mitogen-activated protein kinase kinases (MAPKKs), with high selectivity for a P1 proline and basic residues at P4, P5 and P6. The lethal factor consists of four domains (Fig. 1): domain 1


**Figure 1**

Overview of the lethal factor structure. Stereoview. Domain 1 (red), domain 2 (tan), domain 3 (green) and domain 4 (blue) are shown as cartoons based on structure 1pwq. The active site (circled in orange) displays the catalytic zinc (gray sphere) and the hydroxamic acid ligand of 4pkv (sticks). The hinge angle and inner distance measurements are depicted with maximum (1pwq, black) and minimum (4pkq, red) values.


**Figure 2**

The sulfonamide hydroxamate LF inhibitors MK-31 and MK-40 (Xiong *et al.*, 2006).

(residues 1–263), which interacts with PA; domain 2 (residues 264–299 and 386–550), which may play a role in recognition of the C-terminal portion of the substrate peptide beyond P5'; domain 3 (residues 300–385), which is utilized in substrate recognition of P1–P5' and is inserted into domain 2; and domain 4 (residues 551–776), which contains the catalytic zinc site and the key substrate-recognition sites S1'–S6 (Pannifer *et al.*, 2001; Turk *et al.*, 2004).

Two constructs of LF have previously been crystallized: the full-length sequence (mature protein residues 1–776) lacking only a 33-amino-acid signaling peptide from the immature transcript (Bragg & Robertson, 1989; Pannifer *et al.*, 2001; Turk *et al.*, 2004; Forino *et al.*, 2005) and an N-terminal truncation that eliminates domain 1 of LF (the PA-binding

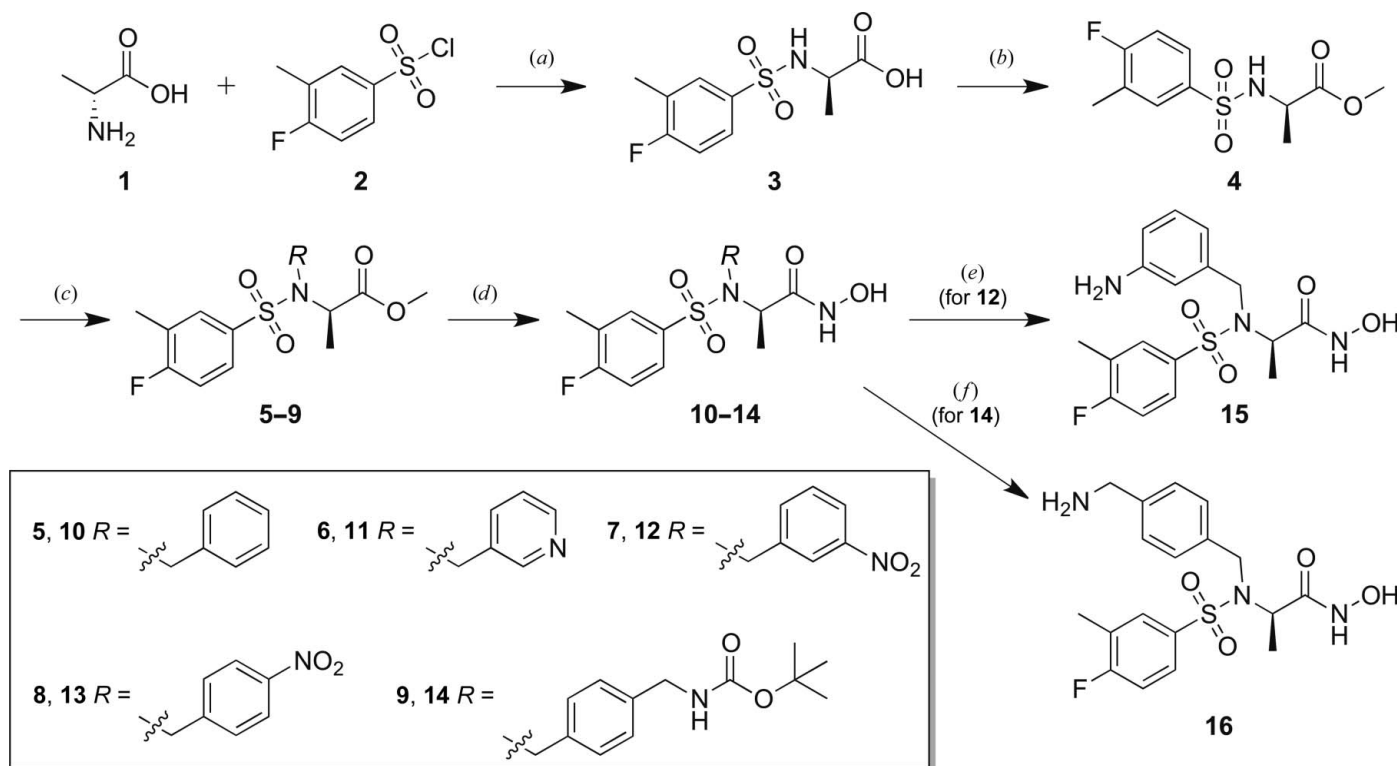
domain; residues 1–264), referred to as LF<sub>NT</sub> (Shoop *et al.*, 2005; Jiao *et al.*, 2012).

In this work, we describe and characterize several complexes with a series of small-molecule inhibitors of LF. Included are five new analogs of the potent sulfonamide hydroxamate inhibitors (Xiong *et al.*, 2006; MK-31 and MK-40; Fig. 2) that include benzylic additions at the sulfonamide N atom. The first apo structure of LF<sub>NT</sub> is also reported, as well as the complex of this construct with the well studied metalloprotease inhibitor GM6001 (Grobelyny *et al.*, 1992; Turk *et al.*, 2004; Fig. 2). These structures allow the characterization of dramatic conformational changes that occur in LF domain 3 upon ligand binding to reshape the LF active site. We have identified at least three distinct structural states for LF that may be specifically targeted in the design of novel LF inhibitors.

## 2. Materials and methods

### 2.1. Synthesis of compounds

Synthetic modifications to the sulfonamide of MK-31 were accomplished as outlined in Fig. 3. A generalized synthetic route was fashioned for all novel analogs as follows. The intermediate sulfonamide **3** was readily synthesized from commercially available D-alanine and 4-fluoro-3-methylphenylsulfonfylchloride *via* nucleophilic substitution. Protection of the carboxylic acid was accomplished under Fischer esterification conditions to give **4** in a 78% yield over the first two steps. The alkylation of sulfonamide **4** with the appropriate aryl bromides and chlorides was carried out under basic conditions to afford the tertiary sulfonamides **5–9** in 52–83% yields. The penultimate esters were converted to the hydroxamic acids **10–14** using hydroxylamine hydrochloride and sodium methoxide in 28–73% yields. Further modification to **12** was pursued by reducing the *m*-nitrobenzyl substituent of

**Figure 3**

Reagents and conditions: (a)  $K_2CO_3$ , dioxane/ $H_2O$  (1:1), room temperature (rt); (b) concentrated  $H_2SO_4$ , methanol reflux (78% over two steps); (c)  $R-X$ ,  $K_2CO_3$ , DMF, rt (**5**, 82%; **6**, 79%; **7**, 83%; **8**, 52%; **9**, 71%); (d)  $NH_2OH.HCl$ ,  $NaOMe$ , methanol, 273 K to rt (**10**, 54%; **11**, 28%; **12**, 73%; **13**, 45%; **14**, 50%); (e) 10% (w/w) Pd/C, DCM, rt (**15**, 82%); (f) 4 N HCl in dioxane (**16**, 97%)

the sulfonamide. The hydrogenation of **12**, to afford **15**, was accomplished in an 82% yield under a hydrogen atmosphere in the presence of palladium on activated carbon (Pd/C). Moreover, *tert*-butoxycarbonyl deprotection of analogue **14** was achieved with 4 N HCl in dioxane to yield **16** as its HCl salt in 97% yield. See the Supporting Information<sup>1</sup> for detailed synthetic information.

Compounds **10**, **11**, **13**, **15**, **16** and **17** were tested for inhibitory activity against  $LF_{NT}$  in a Förster resonance energy transfer (FRET) assay as described previously (Chiu *et al.*, 2009). Results from triplicate measurements are reported in Table 2.

## 2.2. Protein purification

DNA encoding residues 265–776 (A266S) of *Bacillus anthracis* lethal factor ( $LF_{NT}$ ) was cloned into pMCSG10 (Stols *et al.*, 2007; Eschenfeldt *et al.*, 2009; Cormier *et al.*, 2010, 2011; Seiler *et al.*, 2014) to produce a TEV-cleavable N-terminal GST fusion bearing a His<sub>6</sub> tag.  $LF_{NT}$  was expressed using *Escherichia coli* BL21(DE3) Rosetta2 pLysS cells. On a 10 l scale, the cells were grown to an  $OD_{600}$  of 0.6–0.8 at 310 K, cooled to 303 K, induced with 0.2 mM IPTG for 6–8 h and then harvested by centrifugation (15 min at 8200g). Cell pellets were frozen at 253 K. Cell pellets were resuspended in 145 ml 50 mM Tris pH 7.6, 500 mM NaCl, 10% glycerol, 1 mM

DTT and lysed by sonication on ice. Lysozyme (1 mg ml<sup>-1</sup>), benzonase (1 mU ml<sup>-1</sup>) and  $MgCl_2$  (1 mM) were added and stirred for 30 min on ice. The lysate was cleared by centrifugation at 40 000g for 45 min at 277 K and the supernatant was clarified using a 0.45  $\mu$ m syringe filter prior to loading onto a 50 ml Ni-NTA column and elution with lysis buffer containing 500 mM imidazole. Histidine-tagged *Tobacco etch virus* (TEV) protease was added at 0.8% (w/w) and incubated at ambient temperature for 45 min followed by extensive dialysis overnight at 277 K against lysis buffer with 0.5 mM TCEP instead of 1 mM DTT. The dialyzed material was passed through the Ni-NTA column, and untagged  $LF_{NT}$  in the flowthrough was dialyzed extensively against 25 mM HEPES pH 7.5 at 277 K. Light, flocculent white precipitate was isolated by centrifugation (15 min at 5000g) and resuspended in 50 mM Tris pH 7.6, 500 mM NaCl, 10% glycerol. The redissolved  $LF_{NT}$  was applied onto a HiPrep 26/60 Sephacryl S-200 HR column (GE Healthcare) equilibrated with 25 mM HEPES pH 7.5, 150 mM NaCl and eluted as a single peak.  $LF_{NT}$  was concentrated to an  $A_{280}$  of 25.7 and stored at 193 K. The yield was 25 mg from a 10 l batch.

## 2.3. Crystallization

Prior to crystallization, the protein was incubated with each compound of interest. In brief, the incubation solution (500  $\mu$ l) consisted of 200  $\mu$ M compound, 2  $\mu$ M protein and 10% DMSO in 25 mM HEPES pH 7.5, 150 mM NaCl. After incu-

<sup>1</sup> Supporting information has been deposited in the IUCr electronic archive (Reference: DW5106).

**Table 1**

Crystallographic refinement and summary statistics.

Values in parentheses are for the highest resolution shell.

PDB code	4pkq	4pkr	4pks	4pkt	4pku	4pkv	4pkw
Resolution (Å)	2.20	2.20	2.30	2.40	2.40	2.50	1.75
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters							
<i>a</i> (Å)	56.897	52.200	51.920	57.530	61.337	57.480	51.427
<i>b</i> (Å)	574.544	78.000	77.890	76.810	67.895	78.729	82.420
<i>c</i> (Å)	139.461	134.700	134.770	139.330	143.358	139.090	130.550
$\alpha = \beta = \gamma$ (°)	90	90	90	90	90	90	90
Data processing							
Resolution range (Å)	139.46–2.20 (2.21–2.20)	44.89–2.20 (2.28–2.20)	44.92–2.30 (2.38–2.30)	139.33–2.40 (2.41–2.40)	143.36–2.40 (2.41–2.40)	139.09–2.50 (2.51–2.50)	130.55–1.75 (1.76–1.75)
Observations measured	201109 (1871)	181140 (n/a)	172653 (n/a)	159595 (1543)	153739 (1633)	143774 (1320)	364950 (3667)
Unique reflections	30912 (273)	60393 (588)	25017	24878 (241)	24160 (243)	22541 (195)	56847 (563)
Average multiplicity	6.5 (6.9)	6.62 (6.78)	6.9 (6.87)	6.4 (6.4)	6.4 (6.7)	6.4 (6.8)	6.4 (6.5)
Completeness (%)	100.0 (100.0)	95.4 (91.5)	99.9 (99.6)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
$R_{\text{merge}}$	0.068 (0.334)	0.143 (0.488)	0.159 (0.477)	0.100 (0.308)	0.059 (0.408)	0.071 (0.362)	0.041 (0.357)
$\langle I/\sigma(I) \rangle$	20.8 (6.5)	8.7 (2.9)	7.1 (2.8)	13.4 (6.0)	19.9 (6.7)	18.2 (6.1)	26.8 (4.8)
Refinement statistics							
Resolution range (Å)	52.68–2.20 (2.27–2.20)	39.00–2.20 (2.28–2.20)	38.95–2.30 (2.39–2.30)	44.37–2.40 (2.50–2.40)	71.68–2.40 (2.50–2.40)	44.04–2.50 (2.59–2.50)	41.38–1.75 (1.81–1.75)
Working-set reflections	29302 (2601)	27291 (2400)	24966 (2577)	24815 (2568)	24103 (1224)	21336 (2060)	56769 (2878)
$R_{\text{free}}$ reflections	1551 (130)	1377 (129)	1271 (142)	1262 (125)	1224 (128)	1151 (119)	2552 (103)
<i>R</i>	0.1756 (0.2011)	0.2042 (0.3821)	0.2090 (0.2350)	0.1782 (0.2062)	0.1973 (0.2316)	0.1812 (0.2306)	0.1772 (0.2072)
$R_{\text{free}}$	0.2171 (0.2970)	0.2534 (0.4152)	0.2604 (0.3067)	0.2399 (0.2902)	0.2477 (0.2935)	0.2528 (0.3386)	0.2067 (0.2833)
No. of non-H atoms	4362	4422	4265	4209	4044	4301	4458
No. of solvent waters	192	308	220	115	81	108	386
No. of missing atoms	121	211	291	218	357	102	302
Mean <i>B</i> factors (Å <sup>2</sup> )							
Protein atoms	30.48	30.33	35.54	43.45	52.20	45.20	27.00
Solvent atoms	32.06	31.25	35.80	40.82	52.21	42.00	34.12
Ligand atoms	31.91	31.55	36.82	50.36	55.63	63.11	23.90
R.m.s. deviations from ideal geometry							
Bond lengths (Å)	0.008	0.002	0.004	0.008	0.003	0.010	0.007
Bond angles (°)	1.070	0.645	0.725	1.056	0.752	1.300	1.056
Ramachandran plot outliers (%)	0.2	0.0	0.0	0.2	0.0	0.8	0.0
<i>MolProbity</i> score	1.3	1.2	1.5	1.3	1.0	1.8	1.0

bation at room temperature for 30–45 min, the solution was filtered (0.22 µm) and concentrated to greater than 5 mg ml<sup>-1</sup>.

Crystals were grown at 286 K using the hanging-drop vapor-diffusion method and microseeding to encourage the growth of fewer larger crystals. Crystallization drops consisted of post-incubation protein solution (2.0 µl) and 2.0 µl of either well solution or well solution (1.5 µl) plus microseeding solution (0.5 µl). The well solutions that yielded crystals consisted of 50 mM bis-tris pH 6.8, 100 mM magnesium acetate, polyethylene glycol 8000 (PEG 8K; 11–16%). A microseeding solution was prepared by crushing crystals grown without seeding with a micropestle. Crystals appeared and grew to full size within a month. To harvest samples for data collection, crystals were quickly dipped in a 25% ethylene glycol-supplemented well solution, followed by flash-vitrification in liquid nitrogen.

#### 2.4. Crystallographic data collection and processing

Diffraction data for the structures deposited as PDB entries 4pkq, 4pkt, 4pku, 4pkv and 4pkw were collected from crystals at 100 K on beamline 17-ID-B (IMCA-CAT) using a Dectris PILATUS 6M pixel-array detector at the Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois,

USA. The data were processed using *XDS* (Kabsch, 2010) and scaled with *SCALA* (Evans, 2006).

For the structures deposited as PDB entries 4pkr and 4pks, diffraction data were collected from crystals at 100 K using a NOIR-1 MBC detector on beamline 4.2.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California, USA. The data were processed using *d\*TREK* (Pflugrath, 1999).

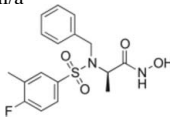
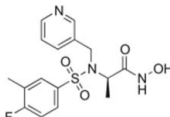
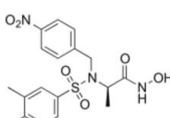
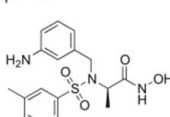
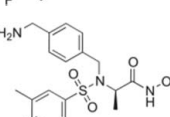
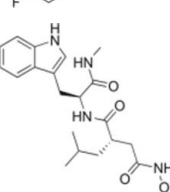
The structures were solved using molecular replacement with the atomic coordinates of PDB entry 1yqy (Shoop *et al.*, 2005) using *Phaser* (McCoy *et al.*, 2007) in the *CCP4* suite (Winn *et al.*, 2011). Both *REFMAC5* (Murshudov *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010; Afonine *et al.*, 2012) were utilized for data refinement, along with the *Coot* modelling and visualization software (Emsley & Cowtan, 2004).

Data-collection and refinement statistics are summarized in Table 1.

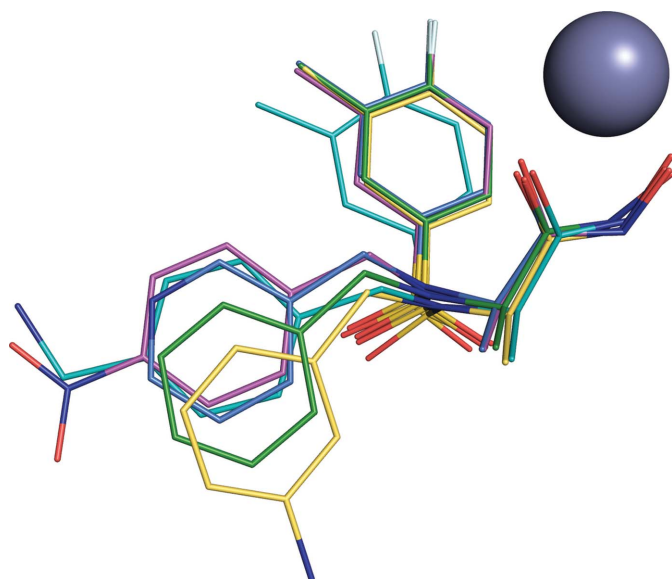
#### 2.5. Protein superposition

Non-isomorphous protein structures were aligned onto a common frame of reference using only a conserved core substructure comprised of two helical segments (residues 686–692 and 735–740) from reference structure 1yqy (Shoop *et al.*,

**Table 2**  
Summary of crystallographic complexes.

PDB code	Resolution (Å)	Compound No.	IC <sub>50</sub> (μM)	Compound
4pkq	2.20	None	n/a	n/a
4pkr	2.20	<b>10</b>	15.2 ± 0.5	
4pks	2.30	<b>11</b>	15.2 ± 0.9	
4pkt	2.40	<b>13</b>	14.9 ± 2.4	
4pku	2.40	<b>15</b>	30.0 ± 1.0	
4pkv	2.50	<b>16</b>	5.6 ± 3.0	
4pkw	1.75	<b>17</b> (GM6001)	22.5 ± 1.6	

2005). The segments include the Zn-coordinating histidines and glutamate. Locally centralized superposition of only this core substructure gives rise to a better alignment of the ligands



**Figure 4**  
Overlaid compounds. **10**, green; **11**, blue; **13**, purple; **15**, yellow; **16**, cyan.

(Finzel *et al.*, 2011) and simplifies the recognition of changes to the protein quaternary structure relative to the fixed active site. The 'ATLF' overlay method has been shared at <https://drugsite.msi.umn.edu/>, where web-based services exist to overlay any structures that share this core (Finzel *et al.*, 2011).

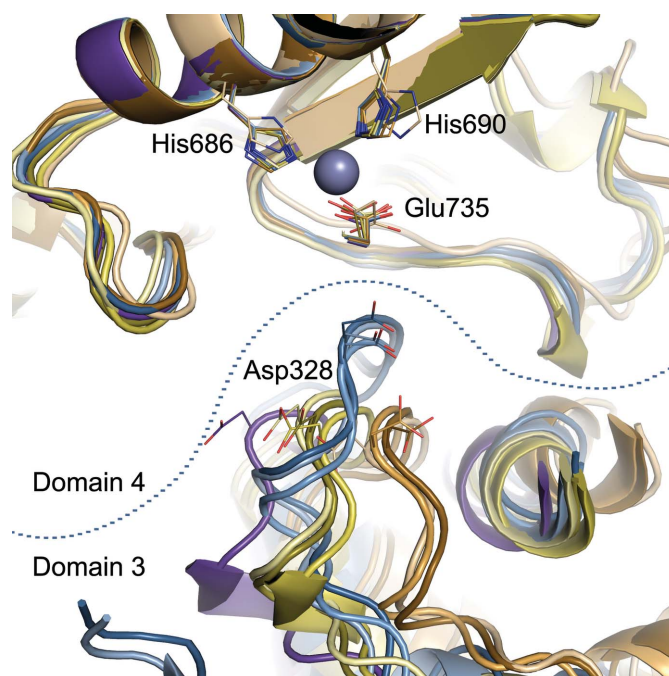
## 3. Results and discussion

### 3.1. Crystallographic complex overview

Seven new crystal structures are reported: five with novel ligands, one with the established inhibitor GM6001 and one apo structure of the LF<sub>NT</sub> construct with resolutions from 1.75 to 2.5 Å (Table 2). While they are all orthorhombic crystals, the unit cells are non-isomorphous, with a range of 10 Å or more on each edge. They are all the result of co-crystallization efforts; apo crystals are not amenable to soaking. The structure and position of domain 3 also differs in these complexes (see the next section for an in-depth discussion) and it is rarely completely ordered, especially in the region of residues 345–370. Overall, the ligands conform to a standard binding mode driven by the presence of a hydroxamate zinc-chelating group, and common features align well (Fig. 4).

### 3.2. Classification of the domain 3 position

The differences in the LF domain 3 position are quite pronounced between the full-length structures 1jky, 1j7n (Pannifer *et al.*, 2001), 1pwp, 1pwq, 1pwu, 1pwv, 1pww (Panchal *et al.*, 2004; Turk *et al.*, 2004) and 1zxv (Forino *et al.*,



**Figure 5**  
Three classes of loop states seen in LF structures: bioactive (1zxv, 1pwu and 1jky, dark, medium and light orange), open (4pks, 4pkr and 4pkw, dark, medium and light yellow; outlier 4pku, purple) and tight (4pkt, 4pkq and 4pkv, dark, medium and light blue). The catalytic Zn is shown as a sphere, with Zn-coordinating residues and Asp328 side chains displayed as sticks.

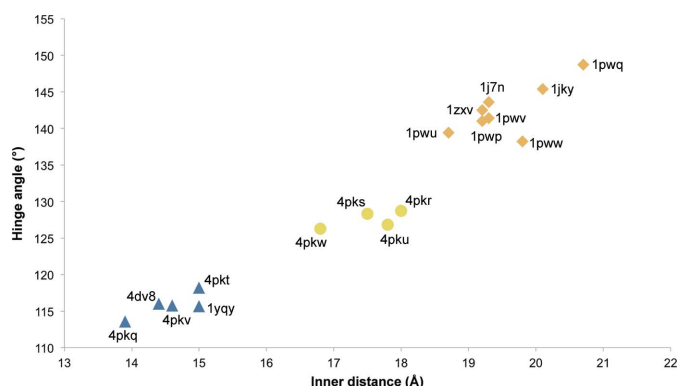


2005), and the N-terminally truncated structures 1yqy (Shoop *et al.*, 2005) and 4dv8 (Jiao *et al.*, 2012) and those reported here. Lethal factor domain 3 consists of residues 303–382 and is connected to domain 2 by a hinge at Ile300 and Pro385 (as identified using *HingeProt*: <http://www.prc.boun.edu.tr/appserv/prc/hingeprot/hingeprot.html>). This domain is observed crystallographically in three main positional states; for clarity, in the following discussion we have designated these states as ‘bioactive’, ‘open’ and ‘tight’ (Fig. 5). A fourth possible state is observed in structure 4pku (a complex with compound **15**), but has not been replicated in any other structure to date and is considered to be an outlier at this time. Two parameters can be used to quantify the degree of movement of domain 3. The first is the distance between the C $\alpha$  atom of Asp328, a residue in domain 3 that comes into close contact with ligands and substrates, and the C $\alpha$  atom of His686, a zinc-binding residue that remains stationary; this measure quantifies the distance between domain 3 and the active site and is called the ‘inner distance’. The angle between the C $\alpha$  atoms of Asp344, a residue on the outer portion of domain 3, Ile300, a hinge residue, and Tyr438, a stationary

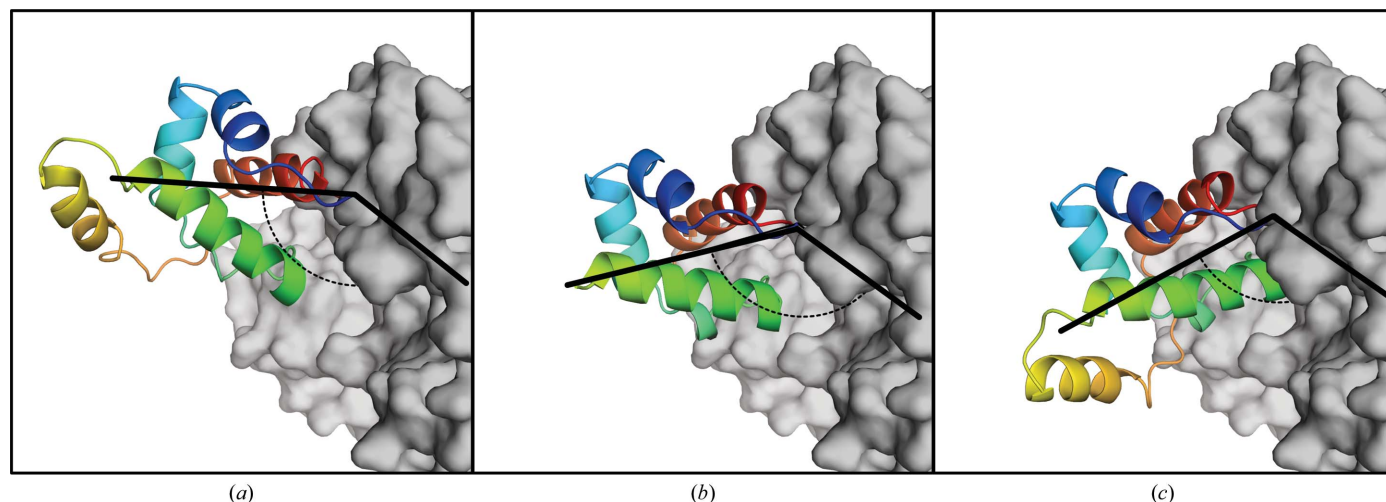
residue in domain 4 which completes a plane perpendicular to the hinge axis, is utilized as a measure of domain 3 movement as a whole; this is called the ‘hinge angle’ (see Fig. 1 for an overview). A plot of hinge angle *versus* inner distance (Fig. 6) reveals the clustering of available structures into three distinct conformational states. The range of motion in domain 3, as defined by the available crystal structures, encompasses a 35.1° change in hinge angle and a 6.8 Å change in the inner distance. Fig. 7 illustrates the magnitude of this domain movement.

The bioactive state is seen exclusively and exhaustively in the full-length LF structures (PDB entries 1jky, 1j7n, 1pwq, 1pwu, 1pww, 1pww and 1zxv). This state is characterized by the largest substrate-binding cleft, hinge angle and inner distance, which maximize the contact between an extended peptide substrate and domains 3 and 4. Structures 1pwv and 1pww contain substrate peptides that may be used to define the LF binding subsites S5–S5’ often used to describe protease-binding clefts. Of the three enzyme states defined here, the bioactive state has the largest variance in the position of domain 3, with an average inner distance of  $19.6 \pm 0.7$  Å and an average hinge angle of  $142.7 \pm 3.6^\circ$ . One might argue that the existence of this conformational state is owing to crystal contacts that are unique to the full-length monoclinic crystal form, but the same bioactive state is also seen in 1jky, where the cubic-form packing constraints are different (discussed below).

Examples of the open state include structures 4pkr (compound **10**), 4pks (compound **11**) and, of interest for purposes of comparison, 4pkw, which contains compound **17** (GM6001). In the open conformation, domain 3 is slightly closer to the active site than in the bioactive conformation (inner distances of  $17.8 \pm 0.3$  Å *versus*  $19.6 \pm 0.3$  Å on average). The standard deviations of this position are also smaller. It appears that the open state of domain 3 would not alone preclude the binding of peptidic substrates as in 1pwv or 1pww, but adaptations to the Gly674–Glu676 loop in domain 4 block the S4’ subsite and close contacts occur between the peptide ligand of 1jky and the Asp328–Phe329 loop in the



**Figure 6** Plot of hinge angle *versus* inner distance. Note the clustering of the bioactive (orange diamonds), open (yellow circles) and tight (blue triangles) conformations.



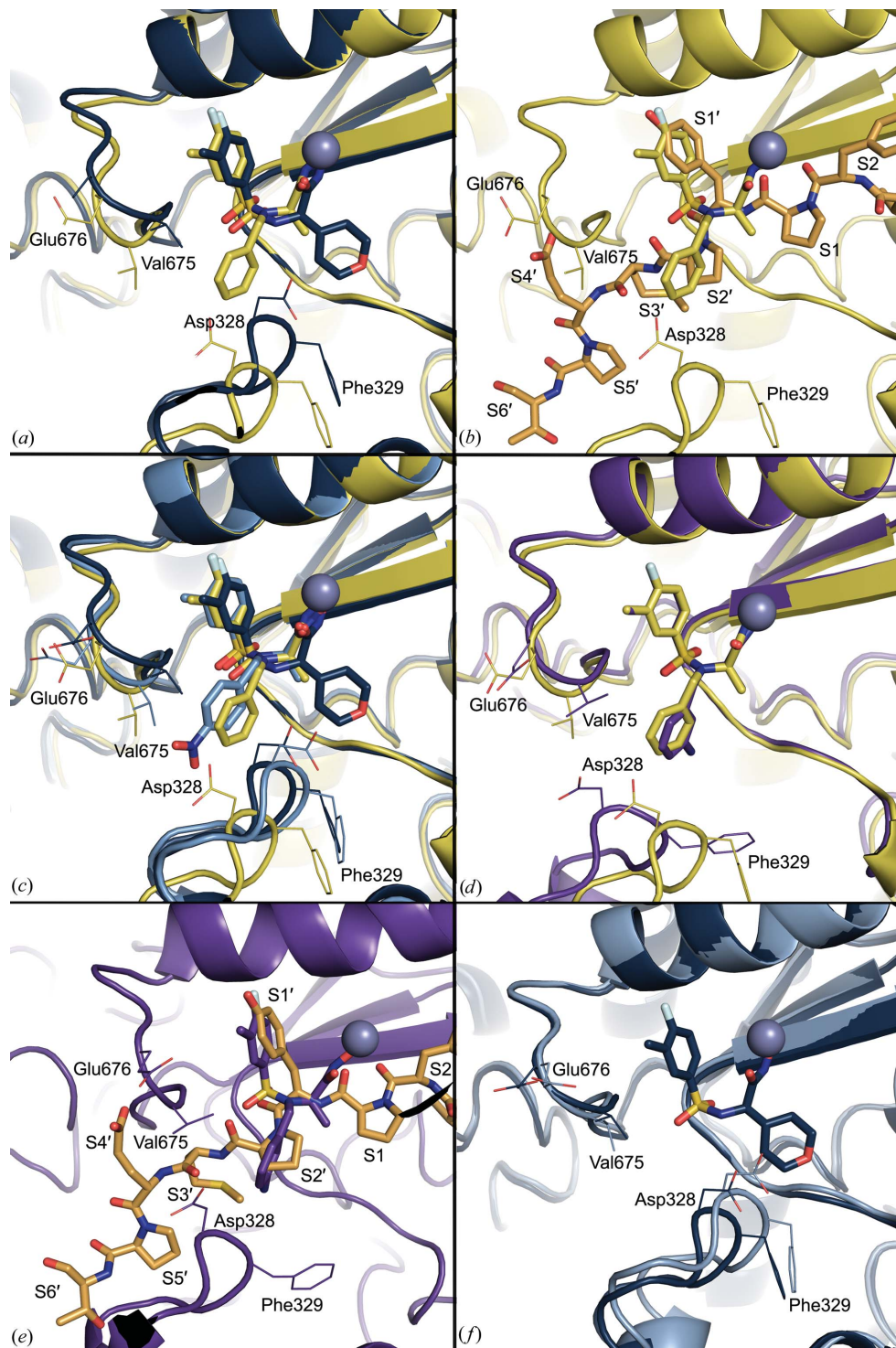
**Figure 7** Range of motion of domain 3. (a) The bioactive position as seen in structure 1jky. (b) The open position of 4pkw. (c) The tight position seen in 4pkv. The hinge angle is also illustrated.

open state. Subsite S2' may also be occluded in 4pks owing to the conformation of Phe329. Structures 4pkr, 4pks and 4pkv are likely to be observed with this domain 3 position owing to steric incompatibilities of compounds **10**, **11** and **17** with the tight state.

Based on the hinge angle and inner distance alone ( $126.8^\circ$  and  $17.8 \text{ \AA}$ , respectively), complex 4pku (compound **15**) could also be assigned to the open category; however, visual inspection of the structure indicates that this structure may be significantly different. The major differences arise from

changes in the position of the loop containing Asp328. There is a shift of approximately  $4 \text{ \AA}$  along the vector described by the  $C^\alpha$  atoms of Asp328 in structures in the open and bioactive positions. In this unique outlier state, the only peptide-binding subsite that seems to be significantly altered by the domain 3 movement is the S5' subsite, which is practically ablated owing to the shift in the Asp328–Phe329 loop.

The tight positional state, as seen in the structures 4pkq, 4pkt, 4pkv, 4dv8 and 1yqy, exhibits the smallest hinge angle and inner distance ( $115.9 \pm 2.7^\circ$  and  $14.6 \pm 0.2 \text{ \AA}$  on average). When domain 3 is in the tight conformation, the subsites S2, S2', S5' and S6' are obstructed (see Supplementary Animation S2). In the tight conformation three internal helices of domain 3 have rotated, collectively burying hydrophobic residues that are more solvent-exposed in the bioactive or open states, indicating that the tight state has distinct energetic advantages. The apo structure is the most extreme version of this conformation, as the hinge angle is  $113.6^\circ$  and the inner distance is  $13.9 \text{ \AA}$ . The tight state may represent an energy minimum for an unliganded structure (4pkq); in addition to the helix rotation that buries hydrophobic residues, two hydrogen bonds are formed between domain 3 and domain 4. The hydrogen bond between the backbone carbonyl of Phe329 and the amide N atom of His654 is shared by the structures 4pkv (compound **16**) and 4dv8. These structures represent the closest that domain 3 comes into contact with domain 4 (as measured by a small hinge angle and inner distance). The other hydrogen bond between side-chain atoms



**Figure 8**

Ligand binding. (a) 1yqy (dark blue) and 4pkr (complex with compound **10**, yellow). (b) Peptide substrate from 1pww (orange) and 4pkr (yellow). (c) 1yqy (dark blue), 4pkr (yellow) and 4pkt (complex with compound **13**, light blue). (d) 4pku (complex with compound **15**, purple) and 4pkr (yellow). (e) 4pku (purple) and peptide substrate from 1pww (orange). (f) 4pkq (medium blue) and 1yqy (dark blue).



of Asp328 and Tyr728 is a shared feature of all structures in which domain 3 is in the tight position.

### 3.3. Ligand-induced conformational changes

Selected structures from among the series of seven included here will be discussed in detail as representative examples with important structural features.

In the complex with compound **10** (PDB entry 4pkr), the hydroxamate inhibitor conforms to the expected conformation of MK-40 (Fig. 2; Xiong *et al.*, 2006) from 1yqy (Shoop *et al.*, 2005), with the hydroxamate coordinated to the Zn cofactor, the 4-fluoro-3-methylphenyl group in S1' and the chiral C atom directed along a vector toward S1. This complex, and all complexes with benzylic sulfonamide analogs (**10**, **11**, **13**, **15** and **16**) predictably lose a key hydrogen bond between the sulfonamide NH of MK-40 and Tyr728. Tyr728 remains held in the same position, however, by a hydrogen bond to the ligand hydroxamate carbonyl O atom. In order to accommodate the benzyl extension of the inhibitor from the sulfonamide N atom, domain 3 in these complexes is moved outwards on a hinge relative to structure 1yqy (Fig. 8*a*). This movement opens a portion of the S2' subsite that is closed in previously described N-terminally truncated structures (1yqy and 4dv8). A change in the Gly674–Glu676 loop conformation of domain 4 further expands the available space between Val675 and Asp328; the conformational change of the Gly674–Glu676 loop is conserved in all of the complexes with sulfonamide hydroxamate inhibitors reported here. Without these changes, the benzyl group would have steric clashes with domain 3 as present in 1yqy. The benzyl aromatic ring, however, is generally located perpendicular to the observed conformation of the P2' proline that occupies S2' in peptide-complex structures (1pwv and 1pww) (Fig. 8*b*). There appears to be favorable hydrophobic contact with Tyr728, although the

interaction angle of  $\sim 50^\circ$  is not suited to  $\pi$ -stacking. Compound **11** induces a similar domain shift (PDB entry 4pks).

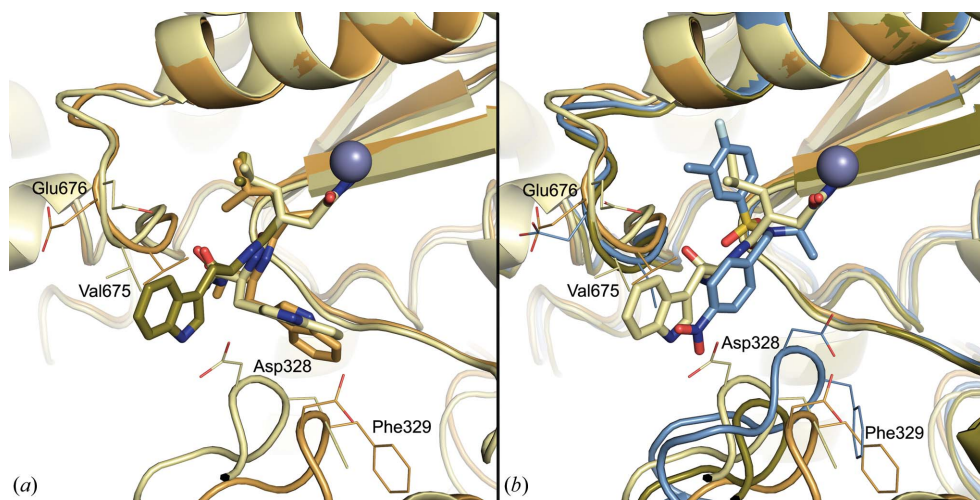
The longer nitrobenzyl and benzyl methylamine extensions from the sulfonamide N atom in compounds **13** and **16** (PDB entries 4pkt and 4pkv, respectively) affect the protein position differently than does the simple benzyl substituent. In these structures the tight state of domain 3 is observed rather than the open state. If the ligand-binding mode were the same as in the complex with **10** and the enzyme were in the open state of domain 3 there would be significant steric clashes between the ligand and the Asp328–Phe329 loop. The tight state is accommodated in these structures by a change in the conformation of the ligand: the internal ligand torsion angle S–N–C $^\alpha$ –C $^\beta$  is reduced from  $73.5^\circ$  in **10** to  $59.7^\circ$  in **13** and  $34.2^\circ$  in **16**. This tilts and shifts the ring of the benzylic substituent towards domain 4, mostly vacating the S2' subsite and allowing the tight state of domain 3 (Fig. 8*c*).

The domain 3 position in the complex with compound **15** (PDB entry 4pku) is not easily categorized. The conformation of the Asp328–Phe329 loop is unique to this complex (Fig. 8*d*); the loop is tilted towards higher-numbered substrate subsites, so that the S5' site is largely closed and the S2' site is opened (Fig. 8*e*). The *meta*-aniline **15** more fully occupies the S2' site and the amine appears to be oriented towards domain 4, resulting in a hydrogen bond to the backbone carbonyl of His654. (Admittedly, there is some ambiguity in the position of the amine, as little electron density is seen outside of the plane of the ring; see the OMIT map in Supplementary Table S1.)

The unliganded structure (PDB entry 4pkq) has the most extreme domain 3 movement compared with the structures containing substrate peptides. The Asp328–Phe329 loop conformation is similar to structures 4pkt, 4pkv, 4dv8 and 1yqy, but it is shifted towards domain 4; such a shift precludes the binding of MK-40 because close contacts would exist

between the tetrahydropyran ring and both the carboxylic acid and the backbone carbonyl of Asp328 (Fig. 8*f*).

A 1.75 Å resolution structure of the LF<sub>NT</sub> protein construct in a complex with the established metalloprotein inhibitor **17** (GM6001; PDB entry 4pkw) exhibits two conformations of the ligand. The tryptophan residue in the inhibitor is observed in two discrete conformations with equal occupancy in this structure (see Supplementary Table S1 for an OMIT map). Conformation *A* is highly similar to that observed in structure 1pwu, where the indole plane is orthogonal to both the P2' proline and the aromatic rings of sulfonamide hydroxamate inhibitors (Fig. 9*a*). The alternate



**Figure 9**

GM6001 binding. (*a*) Two conformations of **17** (*A*, light yellow; *B*, dark yellow) in contrast to the conformation from structure 1pwu. (*b*) One conformation of **17** (*B*, light yellow) compared with the conformation of sulfonamide hydroxamate **13** (blue) and the unique Gly674–Glu676 loop conformation of the complex with **17** (4pkw) relative to the complex with **13** (4pkt) and 1pwu.



conformation places the indole in a position similar to that of the benzylic substituents of the sulfonamide hydroxamates, but the Gly674–Glu676 loop must adopt a unique conformation in order to accommodate the large indole in proximity to Val675 (Fig. 9*b*). The Asp328–Phe329 loop is similar to that observed in other structures with domain 3 in the open position.

### 3.4. Crystal packing

Structures of LF have been published in three crystal forms: monoclinic  $P2_1$  (1j7n, 1pwp, 1pww, 1pwu, 1pww and 1pww), cubic  $I4_132$  (1jky) and orthorhombic  $P2_12_12_1$  (1yqy, 4dv8, 4pkq, 4pkr, 4pks, 4pkt, 4pku, 4pkv and 4pkw), although many of the orthorhombic structures are not truly isomorphous. The cubic and monoclinic forms appear from crystallization of the full-length LF (LF<sub>FL</sub>), while the orthorhombic forms arise from LF<sub>NT</sub>. Apart from the absence of the N-terminal domain (residues 1–263), the primary difference in the structure in these forms is the position of domain 3 and its helix 3 $\alpha$ 3 (residues 350–364), as identified by Pannifer *et al.* (2001). As it is possible that crystal packing influences the position of domain 3 in the different ligand-bound complexes, a discussion of the intermolecular interactions in the different crystal forms is warranted.

The orthorhombic form includes at least three non-isomorphous subforms that vary in the lengths of the cell edges, which correlates with the movement of domain 3; complexes in the tight state (1yqy, 4dv8, 4pkq, 4pkv and 4pkt) have longer *a* and *c* cell edges (~57 and 139 Å, respectively), while complexes in the open state (4pkw, 4pkr and 4pks) have shorter *a* and *c* (~52 and 134 Å). Structure 4pku is the lone example of a third subform with the longest *a* (61 Å) and *c* (143 Å) but a significantly shorter *b* (68 Å). In each of these forms the packing of molecules in the crystals is similar; space for alternate domain 3 orientations is produced by a modest rotation of the entire LF molecule and cell axis length changes that largely leave dominant crystal contacts between domains 2 and 4 intact. Although domain 3 is involved in crystal contacts in each of these subforms, the crystals grow only from inhibited protein solutions by co-crystallization and appear under the same solution conditions. It seems likely that the conformational equilibrium between inhibitor and LF is established prior to incorporation into growing crystals and that the domain 3 position that we observe is not an artifact of packing but is also a prominent state in solution.

The major intermolecular interface in the cubic crystal form exists where the flat face formed of domains 2 and 4 of one LF molecule packs against the same face of another molecule related by a crystallographic twofold axis. Contacts between these LF<sub>FL</sub> 'dimers' are mediated primarily by domain 1. Domain 3 makes no intermolecular contacts in this packing arrangement, and it appears that a shift from bioactive to open to tight states could occur in this crystal form without changes to the overall packing. Cubic-form crystals might provide a good platform for the study of ligand-induced dynamics. Unfortunately, they diffract poorly and tend toward twinning,

and have not been extensively used. The one example of a complex in the cubic form (1jky) is a complex with a peptide ligand in the bioactive state.

Monoclinic crystals also grow only from LF<sub>FL</sub>. Two LF molecules that occupy the asymmetric unit assemble to form an arch-shaped dimer with the two molecules meeting at the top of the arch, with interactions primarily between domains 1 and 2 (Supplementary Fig. S3). Domains 4 form principal contacts with other dimers at the foot of the arch to give the crystal three-dimensional stability. Substrate-binding clefts face the concave interior of the arch in this dimer and domain 3 of each monomer lies on either side of the plane of the arch, making minimal intermolecular crystal contacts. Within this packing arrangement, a transition from the bioactive to the open state would result in a shift of domains 3 from opposite sides of this arch towards a position directly beneath the keystone, where they would overlap. For domain 3 to adopt the tight state, the 3 $\alpha$ 3 helix from the two molecules would lie in exactly the same position: a clear impossibility (Supplementary Fig. S2). It therefore appears that the monoclinic packing prevents LF from adopting either the tight or open state prevalent in orthorhombic crystals.

Complexes with GM6001 provide a basis for comparison of the crystal forms, as it is the only compound to be captured in binding with both full-length and truncated LF protein constructs. The monoclinic structure with GM6001, 1pww, also happens to be the only complex structure of the monoclinic class to result from co-crystallization rather than soaking. The orthorhombic structure 4pkw is also a product of co-crystallization. In relation to zinc coordination and the conformationally rigid parts of the binding site, GM6001 binding is very similar in both crystal forms. In both molecules of the 1pww asymmetric unit, however, domain 3 is in the bioactive state, while in the orthorhombic 4pkw domain 3 is in the open position. The observed domain 3 position in 1pww may be attributed to a lack of adaptive capacity of the monoclinic crystal form: 1pww is isomorphous with the other monoclinic structures, where the open state cannot exist without creating intolerable steric clashes. In 1pww, domain 3 is consequently constrained by crystal packing to occupy a more energetically unfavorable bioactive position.

## 4. Conclusions

There are two primary changes that occur in the anthrax toxin lethal factor structure in response to ligand binding, and these are uniquely observable *via* co-crystallization with LF<sub>NT</sub>. On the large scale, domain 3 moves as a unit into three frequently populated conformational states categorized here as bioactive, open and tight. In relation to the bioactive position, both the open and tight states have a smaller hinge angle, indicating that domain 3 is closer to domain 4, yet some peptide substrate binding does not appear to be precluded in the open position. In the tight state, the S2, S2', S5' and S6' sites are significantly altered so that peptide substrates are unable to bind. More localized changes in the Gly674–Glu676 loop conformation are also observed in all open and tight structures with ligands,

although these conformational changes may alter the recognition site for the P4' substituent of a substrate peptide. This change accommodates the large extensions of the aryl sulfonamides beyond the S2' pocket, and the conformational difference in this loop is even more extreme in the case of GM6001 (conformation *B*) because the tryptophan side chain is a larger substituent than the benzylic extensions of the sulfonamides.

The tight conformation of the apo structure (4pkq) can likely be viewed as the low-energy conformation; variations from this are ligand-induced. While the ligands examined in this paper (with the exception of GM6001) are very closely related, the effect of small changes in the benzyl extension from the sulfonamide N atom can produce pronounced but localized responses, including domain 3 movement. The open state may normally be a sparsely occupied intermediate between the bioactive (substrate-bound) and tight (no substrate) states that is stabilized by specific inhibitors (see Supplementary Animation S2). These movements could not be observed in the full-length crystal structure owing to constraints imposed by the crystal packing. Nevertheless, all of the conformations that we have observed are likely to represent biologically valid protein states, and it is possible that the packing of the full-length crystals artificially constrains this dynamic system.

There is no clear correlation between inhibitor potency and domain 3 position. Domain 3 is necessary for peptide substrate recognition, but the position of this domain appears to be more of a reaction to small-molecule ligand binding than a driver of ligand selection. It might be that molecules can be designed to target interactions with specific structural features of each of these conformational states, thereby leading to inhibitors of higher potency and therapeutic value.

This work was supported in part by the Minnesota Department of Employment and Economic Development (SPAP-06-0014-P-FY07 to BCF) and the National Institutes of Health (R01 AI083234 to EAA). Support was also provided by NIH 5T32-GM008700 (Chemistry–Biology Interface Training Grant), the Amherst College Forris Jewett Moore Fellowship and the University of Minnesota Bighley Graduate Fellowship. The authors would also like to thank Jonathan Solberg and Jon Hawkinson of the Institute for Therapeutics Discovery and Development for compound inhibition data. Computing resources of the Minnesota Supercomputing Institute for Advanced Computational Research (MSI) are also gratefully acknowledged. Use of the IMCA-CAT beamline 17-ID at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Hauptman–Woodward Medical Research Institute.

## References

Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Bragg, T. S. & Robertson, D. L. (1989). *Gene*, **81**, 45–54.
- Chiu, T.-L., Solberg, J., Patil, S., Geders, T. W., Zhang, X., Rangarajan, S., Francis, R., Finzel, B. C., Walters, M. A., Hook, D. J. & Amin, E. A. (2009). *J. Chem. Inf. Model.* **49**, 2726–2734.
- Collier, R. J. & Young, J. A. T. (2003). *Annu. Rev. Cell Dev. Biol.* **19**, 45–70.
- Cormier, C. Y., Mohr, S. E., Zuo, D., Hu, Y., Rolfs, A., Kramer, J., Taycher, E., Kelley, F., Fiacco, M., Turnbull, G. & LaBaer, J. (2010). *Nucleic Acids Res.* **38**, D743–D749.
- Cormier, C. Y., Park, J. G., Fiacco, M., Steel, J., Hunter, P., Kramer, J., Singla, R. & LaBaer, J. (2011). *J. Struct. Funct. Genomics*, **12**, 55–62.
- Dixon, T. C., Meselson, M., Guillemin, J. & Hanna, P. C. (1999). *N. Engl. J. Med.* **341**, 815–826.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Erwin, J. L., DaSilva, L. M., Bavari, S., Little, S. F., Friedlander, A. M. & Chanh, T. C. (2001). *Infect. Immun.* **69**, 1175–1177.
- Eschenfeldt, W. H., Stols, S., Millard, C. S., Joachimiak, A. & Donnelly, M. I. (2009). *Methods Mol. Biol.* **498**, 105–115.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Finzel, B. C., Akavaram, R., Ragipindi, A., Van Voorst, J. R., Cahn, M., Davis, M. E., Pokross, M. E., Sheriff, S. & Baldwin, E. T. (2011). *J. Chem. Inf. Model.* **51**, 1931–1941.
- Forino, M. *et al.* (2005). *Proc. Natl Acad. Sci. USA*, **102**, 9499–9504.
- Grobelny, D., Poncz, L. & Galaray, R. E. (1992). *Biochemistry*, **31**, 7152–7154.
- Hughes, J. M. & Gerberding, J. L. (2002). *Emerg. Infect. Dis.* **8**, 1013–1014.
- Jiao, G.-S., Kim, S., Moayeri, M., Crown, D., Thai, A., Cregar-Hernandez, L., McKasson, L., Sankaran, B., Lehrer, A., Wong, T., Johns, L., Margosiak, S. A., Leppla, S. H. & Johnson, A. T. (2012). *Bioorg. Med. Chem. Lett.* **22**, 2242–2246.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kim, S. O., Jing, Q., Hoebe, K., Beutler, B., Duesbery, N. S. & Han, J. (2003). *J. Biol. Chem.* **278**, 7413–7421.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Moayeri, M., Haines, D., Young, H. A. & Leppla, S. H. (2003). *J. Clin. Invest.* **112**, 670–682.
- Moayeri, M. & Leppla, S. H. (2004). *Curr. Opin. Microbiol.* **7**, 19–24.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Panchal, R. G. *et al.* (2004). *Nature Struct. Mol. Biol.* **11**, 67–72.
- Pannifer, A. D., Wong, T.-Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, D. B., Collier, R. J., Park, S., Leppla, S. H., Hanna, P. & Liddington, R. C. (2001). *Nature (London)*, **414**, 229–233.
- Pflugrath, J. W. (1999). *Acta Cryst.* **D55**, 1718–1725.
- Seiler, C. Y., Park, J. G., Sharma, A., Hunter, P., Surapaneni, P., Sedillo, C., Field, J., Algar, R., Price, A., Steel, J., Throop, A., Fiacco, M. & LaBaer, J. (2014). *Nucleic Acids Res.* **42**, D1253–D1260.
- Shoop, W. L. *et al.* (2005). *Proc. Natl Acad. Sci. USA*, **102**, 7958–7963.
- Stols, L., Zhou, M., Eschenfeldt, W. H., Millard, C. S., Abdullah, J., Collart, F. R., Kim, Y. & Donnelly, M. I. (2007). *Protein Expr. Purif.* **53**, 396–403.
- Turk, B. E., Wong, T.-Y., Schwarzenbacher, R., Jarrell, E. T., Leppla, S. H., Collier, R. J., Liddington, R. C. & Cantley, L. C. (2004). *Nature Struct. Mol. Biol.* **11**, 60–66.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Xiong, Y. *et al.* (2006). *Bioorg. Med. Chem. Lett.* **16**, 964–968.