Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the N-terminal domain of *Yersinia pestis* YopH at 2.0 Å resolution

Artem G. Evdokimov,[a] Joseph E. Tropea,[b] Karen M. Routzahn,[a] Terry D. Copeland[c] and David S. Waugh[a,b]*

[a]Protein Engineering Section, Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, PO Box B, Frederick, MD 21702-1201, USA, [b]Structural Biology Core Facility, National Cancer Institute at Frederick, PO Box B, Frederick, MD 21702-1201, USA, and [c]Protein Chemistry Core Facility, National Cancer Institute at Frederick, PO Box B, Frederick, MD 21702-1201, USA

Correspondence e-mail: waughd@ncifcrf.gov

Received 10 January 2001
Accepted 19 March 2001

PDB Reference: YopH, 1huf.

Yersinia pestis, the causative agent of bubonic plague, injects effector proteins into the cytosol of mammalian cells that enable the bacterium to evade the immune response of the infected organism by interfering with eukaryotic signal transduction pathways. YopH is a modular effector composed of a C-terminal protein tyrosine phosphatase (PTPase) domain and a multifunctional N-terminal domain that not only orchestrates the secretion and translocation of YopH into eukaryotic cells but also binds tyrosine-phosphorylated target proteins to mediate substrate recognition. The crystal structure of the N-terminal domain of YopH ($YopH_N$; residues 1–130) has been determined at 2.0 Å resolution. The amino-acid sequences that target YopH for secretion from the bacterium and translocation into eukaryotic cells form integral parts of this compactly folded domain. The structure of $YopH_N$ bears no resemblance to eukaryotic phosphotyrosine-binding domains, nor is it reminiscent of any known fold. Residues that have been implicated in phosphotyrosine-dependent protein binding are clustered together on one face of $YopH_N$, but the structure does not suggest a mechanism for protein–phosphotyrosine recognition.

1. Introduction

Like many Gram-negative pathogens of plants and animals, *Y. pestis* employs a type III (contact-dependent) secretion system to inject specific proteins, termed effectors, directly into the cytosol of eukaryotic cells (Galan & Collmer, 1999; Cheng & Schneewind, 2000). Collectively, these effectors enable the bacterium to evade the immune response of the infected organism by interfering with specific signaling pathways. Studies of *Yersinia* effectors during the last few years have begun to illuminate their functions and targets in mammalian cells (Cornelis, 2000). YopH, in particular, has been the focus of intense scrutiny. The C-terminal domain of this modular effector is a potent protein-tyrosine phosphatase (PTPase) that is structurally similar to eukaryotic PTPases (Stuckey *et al.*, 1994). YopH dephosphorylates $p130^{Cas}$, paxillin and focal adhesion kinase, leading to disruption of the focal adhesion and inhibition of phagocytosis (Black & Bliska, 1997; Persson *et al.*, 1997; Black *et al.*, 1998). The PTPase activity of YopH is indispensable for virulence (Bliska *et al.*, 1991).

It has recently come to light that the phosphotyrosyl peptide-binding activity of YopH is not confined to its C-terminal PTPase domain. The N-terminal domain of YopH (residues 1–130; $YopH_N$) can also recognize $p130^{Cas}$ and paxillin in a phosphotyrosine-dependent manner (Black *et al.*, 1998; Montagna *et al.*, 2001). In this respect, YopH is reminiscent of the eukaryotic signaling proteins SHP-1 and SHP-2,

which also contain separate PTPase and phosphotyrosine-binding (SH2) domains (Tamir *et al.*, 2000). However, there is no detectable sequence similarity between $YopH_N$ and eukaryotic phosphotyrosine-binding domains (*e.g.* SH2 and PTB).

$YopH_N$ also contains the signals that target the protein for secretion from the bacterium and translocation into eukaryotic cells. Unlike proteins that transit *sec*-dependent (type II) secretory pathways, no residues are cleaved from YopH during secretion and translocation. Analysis of the secretion of truncated YopH polypeptides fused to various reporters revealed that the secretion signal is located at the N-terminus of the coding sequence, between residues 1 and 17 (Sory *et al.*, 1995). Similar deletion analyses localized the translocation signal in YopH to residues 17–70 (Sory *et al.*, 1995), which corresponds to the binding site for a cytosolic protein (SycH) that functions as an accessory factor for translocation (Wattiau *et al.*, 1996; Woestyn *et al.*, 1996).

To gain further insight into the mode of substrate recognition by YopH and to illuminate the structural basis for recognition of effector proteins by the type III secretion apparatus in *Y. pestis*, we have crystallized $YopH_N$ and determined its structure at 2.0 Å resolution.

2. Materials and methods

2.1. Protein expression and purification

The open reading frame encoding $YopH_N$ (residues 1–130 of YopH) was amplified from *Y. pestis* genomic DNA (strain 195/P) by the polymerase chain reaction (PCR) using the following oligonucleotide primers: 5′-GAG AAC CTG TAC TTC CAG GGT GGC GGT GGT GGT ATG AAC TTA TCA TTA AGC GAT CTT C-3′ and 5′-ATT AGT GAT GAT GGT GGT GAT GCC CCC TCG CTC CCG ACT CTT GT-3′. This PCR amplicon was subsequently used as the template for a second PCR with the following primers: 5′-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC GGA GAA CCT GTA CTT CCA G-3′ and 5′-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT AGT GAT GAT GGT GGT GAT G-3′. The amplicon from the second PCR was inserted by recombinational cloning into the entry vector pDONR201 (Life Technologies) to create pKM902 and the nucleotide sequence of the entire insert was confirmed experimentally. The open reading frame encoding $YopH_N$, now bracketed by a hexahistidine tag on its C-terminus and a recognition site (ENLYFQ) for tobacco etch virus (TEV) protease on its N-terminus, was moved by recombinational cloning from pKM902 into the destination vector pKM596 (Evdokimov *et al.*, 2000) to construct pKM904, which directed the expression of $YopH_N$ in the form of an 'affinity sandwich' with MBP fused to its N-terminus and a hexahistidine tag joined to its C-terminus. The MBP fusion protein was expressed in *E. coli* BL21(DE3)-RIL cells (Stratagene) as described in Evdokimov *et al.* (2000).

E. coli cell paste was suspended in 50 m*M* sodium phosphate pH 8, 300 m*M* NaCl (buffer *A*), 1 m*M* ethylenediaminetetraacetatic acid (EDTA) and disrupted with an APV Gaulin Model G1000 homogenizer. After pelleting the insoluble debris by centrifugation, the supernatant was filtered through a 0.45 µm cellulose acetate membrane and applied to an amylose affinity column (New England Biolabs) equilibrated in buffer *A*. After extensive washing, the MBP-$YopH_N$-His_6 fusion protein was eluted with buffer *A* containing 30 m*M* D-maltose. Fractions containing the fusion protein were pooled and dithiothreitol (DTT) was added to a final concentration of 0.3 m*M*. The fusion protein was cleaved with MBP-TEV(S219D)-Arg protease (Kapust *et al.*, 2001) for 17 h at room temperature. The digest was dialyzed into buffer *A* and applied to a Ni-NTA Superflow affinity column (Qiagen) equilibrated with buffer *A*. The column was washed extensively with buffer *A* containing 25 m*M* imidazole and elution was then carried out with a linear gradient of 25–250 m*M* imidazole in buffer *A*. Fractions containing $YopH_N$-His_6 were pooled and EDTA was added to a final concentration of 1 m*M*. The pooled fractions were concentrated by diafiltration and fractionated on a Sephacryl S-100 column (Amersham Pharmacia Biotech) equilibrated in buffer *A*. Fractions containing monomeric $YopH_N$-His_6 were pooled, concentrated as above and dialyzed against 25 m*M* HEPES pH 7, 50 m*M* NaCl. Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was added to the dialysate to a final concentration of 0.1 m*M* and the protein was concentrated to 16 mg ml^{-1} (determined spectrophotometrically using a molar extinction coefficient of 1280 M^{-1} cm^{-1}). The final product was judged to be >95% pure, on the basis of silver staining after sodium dodecyl sulfate–polyacrylamide gel electrophoresis (data not shown). The molecular weight of $YopH_N$ was confirmed by electrospray mass spectrometry. The protein solution was flash-frozen in liquid nitrogen and stored at 193 K.

2.2. Peptide synthesis and purification

The phosphotyrosyl peptide SP-3462 (biotin-Leu-Asp-Val-pTyr-Asp-Val-Pro-Pro-Ser-NH_2) was assembled on an amide resin using 9-fluorenylmethoxycarbonyl chemistry. All residues were added on a model 430A automated peptide synthesizer (Applied Biosystems) except for the phosphorylated tyrosine and biotin, which were added manually. Following deprotection, the peptide was purified by reversed-phase high-performance liquid chromatography with an increasing gradient of acetonitrile in water. The purified peptide had the correct amino-acid composition and electrospray mass spectrometry yielded a molecular weight of 1308.89 Da (theoretical value, 1308.37 Da).

2.3. Crystallization and data collection

All crystals were grown by vapor diffusion in VDX 24-well plates containing 1 ml of precipitant solution per well. Crystallization trials were performed with sparse-matrix kits obtained from Hampton Research (Jancarik & Kim, 1991) and Emerald Biostructures. $YopH_N$ crystallized under several neutral and alkaline conditions with polyethylene glycol 6000

and 4000 (PEG 6000 and PEG 4000) as the precipitant. Crystals were obtained in the absence and in the presence of the phosphotyrosyl peptide, but addition of the peptide was clearly beneficial as virtually every crystal grown in the absence of the peptide was twinned and exhibited high mosaicity. Accordingly, all crystals used in this study were obtained in the presence of the phosphotyrosyl peptide. The best conditions were 11–13% or 19–22% PEG 6000 and 100 m*M* Tris–acetate or Na–bicine pH 8 or 9. Crystals growing from 12% PEG were orthorhombic (arrow-like prisms), whereas those growing from 20% PEG were monoclinic (clusters of irregular 'leaves').

The rate at which the protein crystals grew could be controlled by the reservoir to protein ratio of the drops. A 6 µl:2 µl (reservoir:protein) mixture yielded multiple crystals after 1 d, whereas ratios between 3:3 and 2:6 produced a few large crystals over a period of 3–5 d. We noticed that freshly thawed protein solution always yielded excellent crystals, whereas solutions that had been stored for a few days at 277 K invariably produced crystals of inferior quality. This behavior is likely to be a consequence of the fact that $YopH_N$ is prone to oxidative dimerization even in the presence of modest concentrations of DTT. It is also worth mentioning that DTT, or perhaps its oxidized form, seemed to have an unfavorable influence on the formation of $YopH_N$ crystals: if the final concentration of DTT in the drop was higher than 1 m*M*, then the crystals were visibly flawed and uneven or did not form at all. To scavenge the oxygen from the crystallization experiment without adding DTT to the drop, 10 m*M* DTT was added to the well solution after the crystallization experiment was set up.

Typically, a crystal of $YopH_N$ measuring 0.1 × 0.2 × 0.2 mm was dipped into artificial mother liquor containing 20% ethylene glycol, 10% PEG 6000, 100 m*M* Na–bicine pH 9.0 and then mounted immediately in a monofilament loop and flash-frozen in a cryogenic nitrogen stream (Oxford Cryosystems Cryostream) at 100 K. X-ray diffraction data were recorded using a MAR 345 image plate mounted on a Rigaku X-ray generator equipped with Osmic multilayer focusing mirrors. Orthorhombic crystals generally diffracted to higher resolution and had lower mosaicity and more uniform peak shape than monoclinic crystals. To obtain maximum completeness and good redundancy, usually 80–110° of data for the orthorhombic form of $YopH_N$ crystals were collected in 1–2° oscillations with exposure times between 3 and 8 min per degree using a crystal-to-detector distance of 120 mm. The data were reduced and scaled with *HKL* software (Otwinowski & Minor, 1997). For heavy-atom derivatization, crystals were soaked in artificial mother liquor containing 1–20 m*M* heavy-atom reagent in 10% PEG 6000, 100 m*M* Na–bicine pH 9.0. All Hg^{2+} derivatives as well as most derivatives of Au^{3+} and Pt^{4+} resulted in immediate crystal destruction. Interestingly, soaking the crystals in artificial mother liquor containing 1–2 m*M* $KAu(CN)_2$ resulted in an unacceptable reduction of diffraction quality, whereas crystallization of $YopH_N$ in the presence of the same concentration of $KAu(CN)_2$ yielded crystals that in spite of their high mosaicity and poor crystal shape (multiple leaves) were isomorphous with the orthorhombic crystal form and useful for the solution of the structure. A summary of parameters for data collection and structure solution is given in Table 1.

Table 1
Statistics of data collection and structure solution.

	Native orthorhombic	$KAu(CN)_2$	$K_2Pt(NO_2)_4$	Native monoclinic
Unit-cell parameters				
a (Å)	48.06	47.88	47.08	48.65
b (Å)	120.65	120.23	120.23	48.69
c (Å)	49.04	48.88	48.90	105.61
β (°)				98.48
Space group	$C222_1$	$C222_1$	$C222_1$	$P2_1$
Resolution (Å)	100–2.01	100–2.75	100–2.1	100–2.3
Completeness† (%)	97.4 (85.0)	91.3 (78.2)	97.2 (93.0)	93.2 (95.1)
Completeness >2σ† (%)	83.3 (55.0)	78.3 (66.2)	83.1 (54.1)	79.2 (48.5)
Redundancy	4.01	3.92	4.11	2.9
Unique reflections	9659	10353‡	15555‡	10321
$I/\sigma(I)$	15.3	11.0	14.1	11.3
Used for phasing§	3546	3147	3465	—
R_{merge}† (%)	7.2 (34.2)	12.1 (45.3)	8.3 (44.2)	—
ΔF_{iso} (%)	—	27.2	15.8	—
ΔF_{ano} (%)	—	7.7	4.5	—
No. of sites	—	2	2	—
Peak heights (σ)	—	16, 7	11, 2	—
Centric *R*	—	0.57	0.68	—
Figure of merit	0.47 (0.84)¶	—	—	—

† Numbers in parentheses represent values for the highest resolution shell. ‡ Friedel equivalents not merged. § Reflections in the resolution range 100–2.8 Å and higher than 2σ. ¶ Calculated using the data in the range 100–2.8 Å phased with isomorphous and anomalous differences for both derivatives. The number in parentheses corresponds to the figure of merit after density modification.

Owing to much higher mosaicity and larger unit-cell dimensions, the data for the monoclinic form of the protein had to be collected in 0.5° oscillations using a crystal-to-detector distance of 160 mm. The peaks were rather diffuse and uneven in shape, which is not surprising in view of the fact that the crystals had to be extensively manipulated in order to separate individual leaves from clusters.

2.4. Structure determination

Despite the worse than average data quality of the gold derivative prepared by cocrystallization, peaks corresponding to two heavy atoms were clearly visible on Patterson maps. An automated search using the program *SOLVE* (Terwilliger & Berendzen, 1999) located two Au atoms in the unit cell. Phasing with *SHARP* (de La Fortelle & Bricogne, 1997) followed by density modification in *SOLOMON* (Abrahams & Leslie, 1996) using the Au atoms as the sole source of phasing resulted in electron-density maps with discernible secondary-structure elements, but the tracing was difficult. Consequently, the search for heavy-atom derivatives continued until the platinum derivative [6 m*M* $K_2Pt(NO_2)_4$ soaked for 3 h] was found. Phasing with both derivatives at 2.8 Å resolution using *SOLVE* and *RESOLVE* (Terwilliger, 2000) followed by *SHARP* and *SOLOMON* resulted in interpretable electron-density maps (Fig. 1*a*) that were manually traced using the program *O* (Jones *et al.*, 1991). Essential details of the structure solution are listed in Table 1.

2.5. Refinement

Initially, residues 8–69 and 81–120 were built into the experimental electron density; residues 70–80 were omitted because they were poorly resolved at this stage. After several cycles of conjugated-gradient least-squares refinement with *SHELXL*97 (Sheldrick & Schneider, 1997) using data in the resolution range 30–2.5 Å, the density for residues 70–80 was tentatively interpreted as an α-helix. Simulated annealing was performed in *CNS* (Brunger *et al.*, 1998), after which a total omit map was calculated (also in *CNS*) to check for model inconsistencies. Residues 70–80 appeared as a somewhat disordered α-helix even when they were completely excluded from the model and they were thus kept and refined as a single helix. Additional residues were built into the electron density at the termini of the protein; the final model included residues 4–126. Some density was visible for residues 127–129 in the simulated-annealing omit maps, but when these residues were introduced into the model they refined to maximum temperature-displacement factors and no density was observed for them in the final map. Consequently, they were excluded from the final model. The hexahistidine tag and the pentaglycine N-terminal extension were not located in the electron-density maps, most likely because they were disordered.

At no point did we observe any sign of the peptide in or near the tentative binding site (residues 11, 30, 33 and 34), even though the density for the polypeptide chain and water molecules in that area was clear and unambiguous (Fig. 1*b*). After several cycles of manual rebuilding followed by least-squares minimization in *SHELX*97, 94 solvent molecules were added to the structure on the basis of difference density (peaks >3σ) and standard proximity criteria. The final refinement (using data in the range 30–2.0 Å) was performed using 50 conjugated-gradient least-squares cycles with isotropic temperature-displacement factors for all the atoms. Validation of the structure was performed with *WHATIF* (Vriend, 1990). Molecular graphics were generated with *BOBSCRIPT* (Esnouf, 1997), *MOLSCRIPT* (Merrit & Murphy, 1994) and *RASTER3D* (Kraulis, 1991). The essential details of the refinement are provided in Table 2. The coordinates and structure factors were deposited with the Protein Data Bank (Berman *et al.*, 2000) under accession code 1huf.

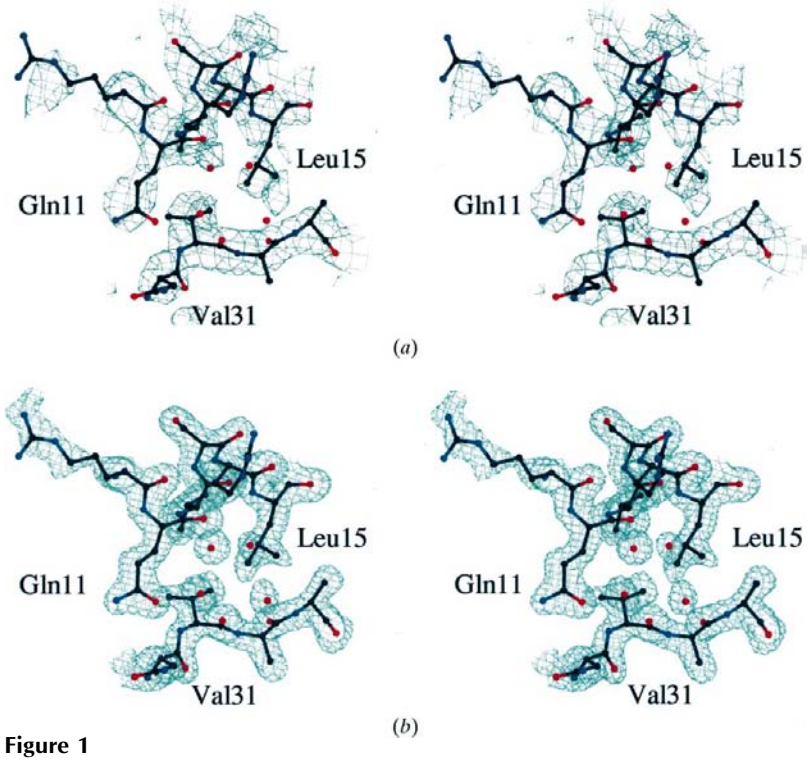

Figure 1
Stereoviews of the electron density in the vicinity of residues Gln11 and Val31. (*a*) Experimental map contoured at 1.2σ. (*b*) $3|F_o| - 2|F_c|$ map contoured at 1.5σ.

3. Results and discussion

3.1. Overall description of the structure

The N-terminal domain of YopH is a compact structure composed of four α-helices and two β-hairpins (Fig. 2). Its fold does not resemble that of any known structure, according to the results of a structure-similarity search conducted using the Dali server (Holm & Sander, 1993). The crystal structure is in perfect accord with the partial secondary-structure assignments derived from NMR experiments (Pellecchia *et al.*, 2000). Helices α1 (residues 4–18) and α3 (residues 69–84) are parallel to each other and antiparallel to helices α2 (residues 49–63) and α4 (residues 10–125). The main axis of helix α1 is rotated approximately 30° with respect to the main axis of the bundle formed by α2, α3 and α4. Between helices α1 and α2, the backbone meanders around one end of the helical bundle, forming a short helical twist (residues 43–46) and a small β-hairpin (β1–2, residues 24–33). Residues 19–23 and 34–42 adopt an extended conformation. Helices α2 and α3 are connected by a short linker that is also in an extended conformation. Helix α3 participates in relatively few interactions with the rest of the protein and consequently is the least ordered structural element ($\langle B_{C^\alpha} \rangle$ = 44.5 Å^2; Fig. 2). A long β-hairpin (β3–4) is positioned between helices α3 and α4. The end of β3–4 is also somewhat disordered ($\langle B_{C^\alpha} \rangle$ = 46.3 Å^2).

3.2. The secretion and translocation signals

The signal that targets YopH for secretion from *Y. pestis* has been localized to codons 1–17 (Sory *et al.*, 1995), which correspond to helix α1 in the structure of YopH$_N$ (Fig. 2). Helix α1 packs tightly against helices α2 and α4, and also makes close contacts with the short β-hairpin between helices α1 and α2. Thus, in contrast to the *Yersinia* effector YopM, in which the N-terminal secretion signal is evidently disordered (Evdokimov *et al.*, 2001), in YopH the secretion signal adopts a well defined conformation that appears to be an integral part of the structure of the N-terminal domain. The notion that helix α1 contributes to

Table 2
Statistics of structure refinement.

Protein atoms	989
Solvent atoms	94
Unique reflections	9659
Restraints	3797
wR^2	0.45
Crystallographic R	
All data (%)	0.22
$F > 4\sigma F$ (%)	0.19
R_{free}	
All data (%)	0.26
$F > 4\sigma F$ (%)	0.24
R.m.s. deviations	
Bond distances (Å)	0.007
Bond angles (°)	2.2
Dihedral angles (°)	5.2
Planarity (Å)	0.003
Ramachandran plot statistics	
Most favored regions (%)	91.2
Allowed regions (%)	8.8
Average atomic displacement factors (Å^2)	
All atoms	32.7
Side-chain atoms	36.6
Main-chain atoms	28.5
C_α atoms	27.8

the stability of the fold is supported by the observation that whereas $YopH_N$ (residues 1–130) is highly soluble when overproduced in *E. coli*, a fragment lacking this helix (residues 20–130) accumulates in the form of insoluble aggregates (data not shown).

The idea that type III secretion signals reside in the secreted proteins themselves has recently been challenged. Anderson & Schneewind (1997, 1999) conducted a systematic mutational analysis of the N-terminal secretion signals in three *Yersinia* proteins: YopE, YopN and YopQ. Surprisingly, they could not find any point mutations that specifically abolished secretion. Even frameshift mutations that completely altered the peptide sequences of the putative secretion signals failed to prevent secretion. These findings led them to conclude that the secretion signals are encoded in the mRNA rather than in the peptide sequence. The structure of $YopH_N$ seems partly at odds with this proposal because, as discussed above, the secretion signal corresponds to an α-helix that forms an integral part of the folded protein. Accordingly, it seems likely that frameshift mutations in this region of YopH would destabilize the structure of the N-terminal domain.

The translocation signal in YopH, which enables the protein to cross the eukaryotic cell membrane, has been localized by deletion analysis to within residues 20–70 (Sory *et al.*, 1995). YopH is one of the effectors that must associate with a cognate Syc (specific Yop chaperone) protein in order to be translocated. The binding site for SycH is coincident with the translocation signal in YopH (Woestyn *et al.*, 1996). This part of the structure includes the segment between helices α1 and α2, which forms most of the surface at one end of the helical bundle, and all of helix α2 (Fig. 2). Like the secretion signal, the translocation signal in YopH does not appear to be an independently folding unit, because this region of the protein is intimately associated with the remainder of the N-terminal domain. Consequently, it is not clear how residues 1–70 of YopH, when fused to an adenylyl cyclase reporter (Sory *et al.*, 1995), could adopt a conformation that resembles the one we observe in the context of the complete N-terminal domain.

3.3. Putative structural similarity between YopH, YscM1 (LcrQ) and YscM2

The amino-acid sequences of two other secreted *Yersinia* proteins, YscM1 (also called LcrQ) and YscM2, are similar to that of $YopH_N$ (Fig. 3). *Y. pestis*, *Y. enterocolitica* and *Y. pseudotuberculosis* all produce YscM1, whereas YscM2 evidently is present only in *Y. enterocolitica* (Stainier *et al.*, 1997). Although its precise mechanism of action remains elusive, YscM1 is thought to function as a global transcriptional regulator that represses the expression of virulence genes prior to physical contact with eukaryotic cells (Cambronne *et al.*, 2000). YscM1 is injected into eukaryotic cells, whereas YscM2 appears to remain associated with bacterial cells (Cambronne *et al.*, 2000).

The greatest degree of similarity exists between residues 52–129 of YopH and the corresponding segments of YscM1 and YscM2. Although the N-terminal residues of both YscM proteins have the potential to form amphipathic α-helices analogous to helix α1 in $YopH_N$, their amino-acid sequences are quite different from that of YopH. In the YscM proteins, the segment between helices α1 and α2 is also shorter than the corresponding segment in YopH. As discussed above, residues 20–50 in $YopH_N$ form the surface that 'caps' one end of the helical bundle. Thus, the YscM proteins could conceivably adopt the same fold as $YopH_N$ but with a different cap at one end of the bundle.

YscM1 and YscM2 also bind to SycH (Cambronne *et al.*, 2000). The region of sequence similarity between YopH and YscM (residues 52–129 in YopH) overlaps with part

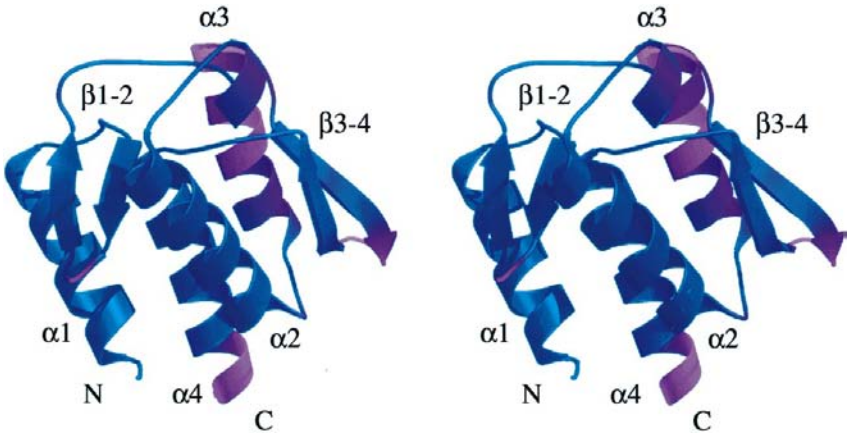

Figure 2
A stereoview of the secondary structure of $YopH_N$, colored in accordance with atomic displacement factors. Blue corresponds to low values, red to high values.

of the putative translocation signal in YopH (residues 20–70), which has been shown to include the binding site for SycH (Woestyn *et al.*, 1996). Thus, it seems likely that SycH binding occurs in this overlap region (residues 52–70), which corresponds to helix $\alpha2$ and the interval between helices $\alpha2$ and $\alpha3$ in $YopH_N$. There are five solvent-accessible side chains in this region of $YopH_N$ that are either identical or conservatively substituted in both YscM proteins: Lys53, Leu60, His62, Asn65 and Val67. Collectively, this constellation of conserved residues forms a patch on the surface of $YopH_N$ (Fig. 4). Site-directed mutagenesis could be used to determine whether these conserved residues are involved in SycH binding.

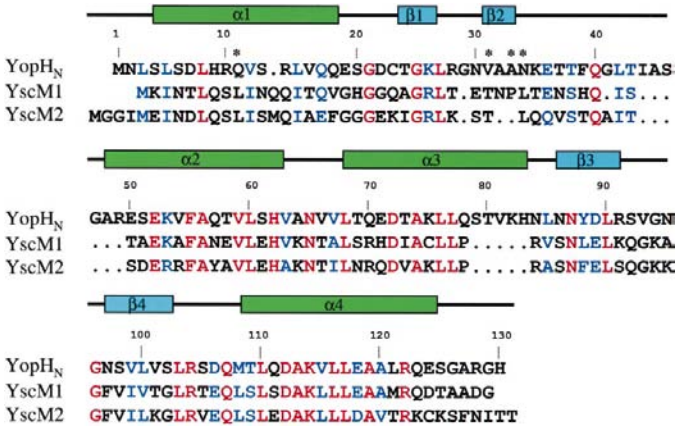

Figure 3
Sequence alignment of $YopH_N$ (*Y. pestis*, residues 1–130), YscM1 (*Y. pestis*) and YscM2 (*Y. enterocolitica*). Amino-acid sequences (in single-letter code) were aligned with the program *PILEUP* (University of Wisconsin Genetics Computer Group) using default parameters. The numbering scheme corresponds to the sequence of $YopH_N$. Residues that are identical in all three protein sequences are shown in red, conservative substitutions in blue. The locations of α-helices and β-strands in $YopH_N$ are indicated by green and blue boxes, respectively. Amino-acid residues in $YopH_N$ that have been implicated in phosphotyrosine-dependent binding of $p130^{Cas}$ are marked by asterisks.

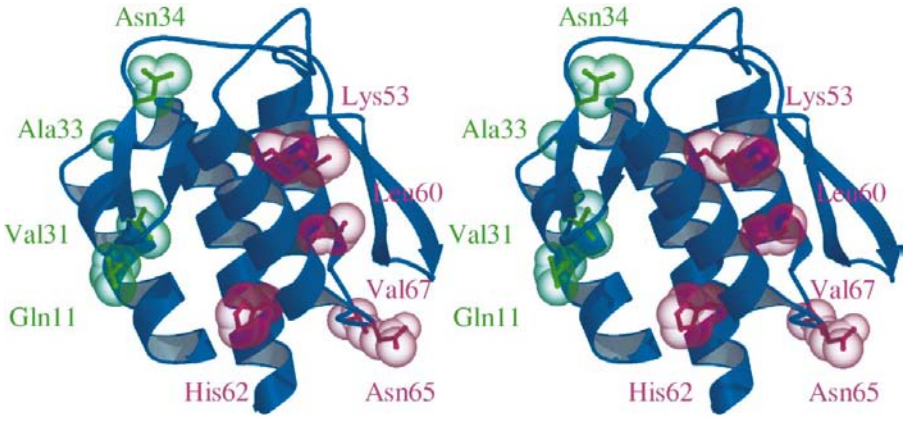

Figure 4
A stereoview of the secondary structure of $YopH_N$ (blue), showing residues that have been implicated in phosphotyrosine-dependent binding of $p130^{Cas}$ (green) and residues that may be involved in binding of SycH (red). The van der Waals surfaces of the amino-acid side chains are represented by transparent spheres.

3.4. Amino-acid residues implicated in phosphotyrosyl peptide binding

Using random mutagenesis in conjunction with a yeast two-hybrid assay, Montagna *et al.* (2001) identified mutations at several positions within the N-terminal domain of YopH that interfered with its ability to bind tyrosine-phosphorylated $p130^{Cas}$. As shown in Fig. 4, these residues are clustered together on one face of the folded protein. Most of the affected residues (Val31, Ala33 and Asn34) are contained within the short hairpin $\beta1$–2, between helices $\alpha1$ and $\alpha2$, whereas Gln11 is located in helix $\alpha1$. None of these residues is conserved in YscM1 (Fig. 3), which does not bind to phosphotyrosyl peptides (Black *et al.*, 1998). Close inspection of the area around these positions does not reveal an obvious cavity that might accommodate a phosphotyrosine residue, as exists in SH2 domains (Kuriyan & Cowburn, 1997). Nevertheless, it is conceivable that phosphotyrosyl peptide binding by $YopH_N$ is accompanied by a conformational change such as occurs when the PTPase domain of YopH binds its substrates (Stuckey *et al.*, 1994).

To investigate the mode of phosphotyrosyl-peptide recognition by the N-terminal domain of YopH, we endeavored to crystallize the protein in complex with a phosphorylated peptide derived from the sequence of $p130^{Cas}$, one of the targets of YopH in mammalian cells. $YopH_N$ has been shown to interact directly and in a phosphotyrosine-dependent manner with $p130^{Cas}$ *in vitro* and *in vivo* (Black *et al.*, 1998; Montagna *et al.*, 2001). Because the addition of the phosphotyrosyl peptide markedly improved the quality of the $YopH_N$ crystals, we expected to observe well defined $|F_o| - |F_c|$ electron density in the vicinity of the putative binding site on $YopH_N$ (residues 11, 31, 33 and 34), but this was not the case. Although there are a few islands of unexplained $|F_o| - |F_c|$ electron density in this area, this density is too weak and ill defined to conclude that it corresponds to the phosphotyrosyl peptide; it could be equally well described as several disordered water molecules. Electron-density maps calculated on the basis of data collected from crystals that had been soaked in 20 m*M* phosphotyrosine (also in the presence of 1 m*M* peptide) and in 20 m*M* peptide did not exhibit significant differences when compared with the native maps. Interestingly, crystals that were grown in the presence of 1 m*M* peptide tended to deteriorate slowly upon prolonged exposure to artificial mother liquor containing no peptide. In contrast, prolonged exposure of the crystals from the same drop to artificial mother liquors containing 1 m*M* peptide did not adversely affect either the crystal shape or diffraction quality. It appears that the peptide improves the packing of $YopH_N$ in the crystal lattice, but it is not clear exactly how it does so.

We thank Suzanne Specht for assistance with peptide synthesis and purification, Sergey Tarasov (Biophysics Resource) for the use of the MS equipment, Alex Wlodawer for critical reading of the manuscript and Anne Arthur for expert editorial advice. We also wish to acknowledge Zbigniew Dauter for providing the synchrotron beam time (beamline X9B of the National Synchrotron Light Source) used to study some of the $YopH_N$ crystals.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Anderson, D. M. & Schneewind, O. (1997). *Science*, **278**, 1140–1142.
Anderson, D. M. & Schneewind, O. (1999). *Mol. Microbiol.* **31**, 1139–1148.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Black, D. S. & Bliska, J. B. (1997). *EMBO J.* **16**, 2730–2744.
Black, D. S., Montagna, L. G., Zitsmann, S. & Bliska, J. B. (1998). *Mol. Microbiol.* **29**, 1263–1274.
Bliska, J. B., Guan, K. L., Dixon, J. E. & Falkow, S. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 1187–1191.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Cambronne, E. D., Cheng, L. W. & Schneewind, O. (2000). *Mol. Microbiol.* **37**, 263–273.
Cheng, L. W. & Schneewind, O. (2000). *Trends Microbiol.* **8**, 214–220.
Cornelis, G. R. (2000). *Proc. Natl Acad. Sci. USA*, **9**7, 8778–8783.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. (2000). *Acta Cryst.* D**56**, 1676–1679.
Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. (2001). Submitted.
Galan, J. E. & Collmer, A. (1999). *Science*, **284**, 1322–1328.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). Submitted.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kuriyan, J. & Cowburn, D. (1997). *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259–288.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Merrit, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Montagna, L. G., Ivanov, M. I. & Bliska, J. B. (2001). *J. Biol. Chem.* In the press.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pellecchia, M., Vander Kooi, C. W., Keliikuli, K. & Zuiderweg, E. R. P. (2000). *J. Mag. Res.* **143**, 435–439.
Persson, C., Carballeira, N., Wolf-Watz, H. & Fallman, M. (1997). *EMBO J.* **16**, 2307–2318.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sory, M.-P., Boland, A., Lambermont, I. & Cornelis, G. R. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 11998–12002.
Stainier, I., Iriarte, M. & Cornelis, G. R. (1997). *Mol. Microbiol.* **26**, 833–843.
Stuckey, J. A., Schubert, H. L., Fauman, E. B., Zhang, Z.-Y., Dixon, J. E. & Saper, M. A. (1994). *Nature (London)*, **370**, 571–575.
Tamir, I., Dal Porto, J. M. & Cambier, J. C. (2000). *Curr. Opin. Immunol.* **12**, 307–315.
Terwilliger, T. C. (2000). *Acta Cryst.* D**55**, 1863–1871.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
Wattiau, P., Woestyn, S. & Cornelis, G. R. (1996). *Mol. Microbiol.* **20**, 255–262.
Woestyn, S., Allaoui, A., Wattiau, P. & Cornelis, G. R. (1996). *Mol. Microbiol.* **20**, 1261–1271.